

第2部

数理モデルとシミュレーション

移動可能な N 人囚人のジレンマにおける協力の進化

— 移動コストと関係の長期性の効果 —

金井 雅之・小林 盾・大浦 宏邦
(山形大学)・(シカゴ大学)・(帝京大学)

【要旨】

選択的相互作用のある繰り返し N 人囚人のジレンマゲームにおいて、プレイヤーの移動戦略を明示的に考えたときに、協力的戦略が進化するための条件を分析した。

まず、非協力的戦略が支配的であるときに協力的戦略が侵入できるためには、移動コストが一定以上あることが本質的に重要である。これは、協力的戦略同士の恣意的なペアリングや集団の不均衡分裂といったこれまでの解決策とはちがった、移動するかどうかの選択という行為者のマイクロなプロセスから導出された新しい解決策と位置づけることができる。

つぎに、協力的戦略が支配的であるときに非協力的戦略の侵入を防ぐためには、第一義的には関係の長期性が重要であるが、その条件が十分満たされない場合でも、移動コストがある程度あれば非協力戦略の侵入を防げる。さらに、この移動コストの効果は、社会的ジレンマ状況における協力のメリットが小さいほど相対的に重要性を増す。これは、従来知られていた条件に付け加えることによって、協力の進化をさらに容易にする新しい条件として、移動コストが位置づけられることを意味する。

キーワード: 社会的ジレンマ, N 人囚人のジレンマ, 選択的相互作用, 移動戦略, 移動コスト

1. はじめに

社会的ジレンマ、すなわち個人が合理的に相互行為した結果が社会的にみて必ずしも最適なものではない状況、をめぐっては、理論・実験・調査等の方法によって多くの研究が積み重ねられてきた¹⁾。

このうち理論的研究において注目されてきた論点のひとつに、選択的相互作用の問題がある。選択的相互作用とは、ゲームにおける相互作用が、集団のすべての構成員のあいだで均等におこなわれるわけではない状況一般を指す。相互作用範囲の偏りを定式化するモデルは理論的には多様に考えうる。たとえば、2人囚人のジレンマゲームにおける社会的に望ましい結果の実現可能性について、ゲームの繰り返し、および各ゲームにおける利得の大小によって集団における戦略シェアが時間とともに変化していくという進化論的な観点を取り入れて分析した R. Axelrod (1984) は、相互作用の偏りをつぎのように定式化した。すなわち、協力的戦略が協力的戦略と対戦する確率が、その時点での集団全体での2つの戦略のシェアに基づくのではなく、特別に設定した高い確率をとる、という仮定である。この仮定をおくことにより、彼は非協力的戦略が支配的な場合に協力的戦略が侵入できることを理論的に示した。

選択的相互作用の発生メカニズムをもう少し具体的に定式化したモデルが、多水準選択モデルである²⁾。これは、戦略シェアの時間変化を考える全体集団（以下では「人口集団」と

よぶ)が、いくつかの部分集団(以下ではたんに「集団」とよぶ)に分かれていて、個々のプレイヤーの相互作用範囲は各集団内に限られるという、部分集合構造を仮定するものである。この場合、何らかの理由で協力的戦略の比率の高い集団が存在すれば、協力的戦略同士の相互作用の機会が増える。非協力的戦略同士が相互作用するよりも協力戦略同士が相互作用したほうが高い利得を得るという社会的ジレンマ状況の特性より、人口集団全体で見たとときに協力的戦略の平均的利得が非協力的戦略の平均的利得を上回ることがありうる。

ここで、集団を比較的恒常的な実体と考え、集団同士の選択過程を進化的に考える議論が、進化生物学におけるいわゆる群淘汰論である。群淘汰論をめぐっては進化生物学内部で一連の論争があり、現在では集団を単純に選択の単位と考える素朴な群淘汰論は否定されている。これに対して、個体の集団への帰属関係を時間とともに変化しうるものと考え、集団構造による相互作用の偏りを個体の適応度に影響を与える重要な要因のひとつとして考慮するモデルが、D. S. Wilson らの離合集散モデルである(Wilson and Sober (1994), Wilson (1998)など)。これは、集団が定期的に解散して新しい集団が再編成されると想定するもので、これによって集団の再編成がない場合よりは集団間での協力的戦略のシェアのばらつきが維持されるため、選択的相互作用による協力的戦略の優位性が人口集団全体で見ても確保されるという議論である。たとえば Wilson (1975) は、この新しい集団への再編成の仕方として、たとえ個体の集団への帰属がまったくランダムに決まる場合でも、集団間の適応度の差の影響が個体間の適応度の差の影響を上回り、協力的戦略が進化できる可能性があることを示した。また金井(2000)は、厳密な離合集散モデルではないが、集団が定期的に分裂するという仮定をおいたときに、分裂後の集団における協力的戦略のシェアがばらつくほど協力的戦略が進化しやすいことを示した。

しかし、これらのモデルの問題点は、集団の再編成をもっぱらマクロな視点からとらえていて、個々のプレイヤーのマイクロな選択の結果とは考えていない点である。実際の社会現象、たとえば企業組織における個々の構成員の協力的行動の成立メカニズムを考える場合は、人びとの意思と独立に企業が定期的に解散したり分裂したりすると想定するよりも、転職するかしないかという個々人の意思決定の集積として、企業の構成員の変化や存亡を考えるほうが有益であろう³⁾。

このように、集団の(漸次的)再編成というマクロな現象を個々のプレイヤーの所属集団から移動するかしないかというマイクロな選択の集積として理解するためには、移動するかしないかという選択を個々人の戦略の要素として明示的に組み込むことが必要となる。すでに繰り返し2人囚人のジレンマゲームにおいては、TIT-FOR-TATを基本としつつ非協力的な相手とそれ以降の相互作用を避けるOUT-FOR-TAT戦略の有効性が知られている(林(1993), Hayashi and Yamagishi (1998))。

そこで本稿では、選択的相互作用のある繰り返し N 人囚人のジレンマゲームにおいて、移動戦略を明示的に考えたときに、協力的戦略が進化するための条件を探っていくことにする。この問題を考察するにあたっては、本来戦略空間全体での戦略の時間変化を分析するのがよいことは言うまでもないが、戦略空間の次元が大きくなることや集団の再構成のメカニズム(あるいは移動先の選択)が確率的であること等により、モデルの自由度が大きくなりすぎ

て解析的研究が容易ではない．そこで本稿では進化的ダイナミクスの出発点としての戦略の局所的安定性，すなわちある戦略が支配的であるときの他の戦略の侵入可能性に絞って，考察を進める．

2. モデル

2.1 N 人囚人のジレンマ

本稿で分析するのは，各集団内で社会的ジレンマが起こっている状況である．個々人の採りうる戦略を協力戦略（C），非協力戦略（D）の2つとし，集団の人数を n (≥ 1)，戦略 C を採用している人数を n_c ($0 \leq n_c \leq n$) とする．戦略 C，D の利得は，いずれも集団内で戦略 C を採用している人の数によって決まるとし，利得関数をそれぞれ $C(n_c)$ ， $D(n_c)$ とする．このとき，集団内で社会的ジレンマが起こっている，すなわちこのゲームが N 人囚人のジレンマゲームであるのは，つぎの2条件が満たされるときである（Dawes（1980））．

$$(1) \quad D(n_c) > C(n_c + 1),$$

$$(2) \quad C(n) > D(0).$$

この条件を満たすゲームとして，本稿ではつぎのような「資源提供ゲーム」 G_1 を考えよう．集団の個人はある量 b (> 0) の資源を提供する（戦略 C）かしない（戦略 D）かの意思決定をする．提供された資源は集団中で合計され，それにある倍率 λ ($1 < \lambda < n$) をかけたものが集団の構成員に均等に分配されるというゲームである⁴⁾．このとき，各戦略の利得関数は

$$C(n_c) = \frac{\lambda b n_c}{n}, \quad D(n_c) = \frac{\lambda b n_c}{n} + b \tag{1}$$

となる．

2.2 選択的相互作用と移動に関する仮定

つぎに，本稿では選択的相互作用モデルを考えるために， n 人から構成される集団が全部で m (≥ 1) 個存在する人口集団を仮定する（図 1）．ここで人口集団は戦略シェアの変動を考える際の基準となるものであるが，資源提供ゲーム G_1 自体は各集団内でおこなわれる．

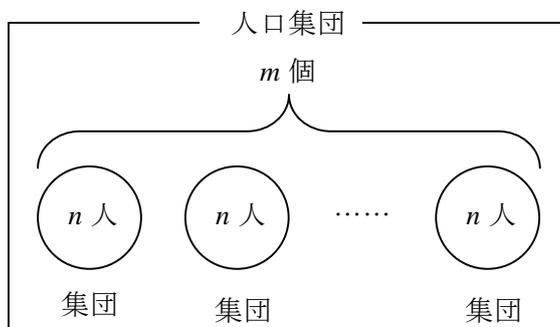


図 1 集団と人口集団

さらに、本稿では各個人の自発的移動による集団構造の漸次的再編成を考慮するため、移動に関するつぎの仮定をおく。すなわち、 G_1 を $T(\geq 1)$ 回繰り返すごとに、すべての集団のすべての構成員に一律に移動の機会が訪れるとする。逆に言うと、 T 回分の間は各集団の構成員は固定されているので、個人の利得計算においては T 回を単位とみることができる。

移動機会に実際に移動するかどうかは各個人の判断によるが、一般に移動にはコストがともなうと考えるのが自然である。そこでこのコストを $\xi(\geq 0)$ としよう。資源提供ゲーム G_1 を T 回繰り返し、移動機会に移動した場合には移動コスト ξ を負担する、というゲームを、「繰り返し資源提供ゲーム」 G_2 とよぶことにする。以後本稿では、各集団内でおこなわれる G_2 における利得の相対的大小に応じて人口集団における戦略のシェアが変動していく、という進化ゲームを考える。

2.3 戦略

繰り返し資源提供ゲーム G_2 における戦略は、2つの要素の組み合わせから定義できる。

第一に、各ステージの G_1 における行為選択を決める戦略（行動戦略）である。ここでは繰り返し囚人のジレンマゲームにおけるこれまでの研究に依拠して、「トリガー戦略」(T)および「非協力戦略」(D)の2つを考えよう。「トリガー戦略」とは、各 G_2 における初回の G_1 においては協力戦略C、2回目以降の G_1 においては、それ以前の任意の回の G_1 において(G_1 における)非協力戦略Dをとった個人が1人でもいる場合はD、そうでなければCをとる戦略である。「非協力戦略」とは、すべての G_1 においてDをとる戦略である。

第二に、移動機会に移動するかどうかを決める戦略（移動戦略）である。ここでは、「移動戦略」(M)および「固定戦略」(F)の2つを考える。「移動戦略」とは、 G_2 を構成する任意の G_1 においてDをとった個人が1人でもいれば移動し、そうでなければ移動しない戦略である。「固定戦略」とは、いかなる場合でも移動しない戦略である。

各個人において、行動戦略と移動戦略は独立に考えることができるので、 G_2 における戦略は、行動戦略と移動戦略の組み合わせからなる計4つを想定できる。すなわち、「移動型トリガー戦略」(MT)、「移動型非協力戦略」(MD)、「固定型トリガー戦略」(FT)、「固定型非協力戦略」(FD)である。

表1 繰り返し資源提供ゲーム G_2 における4つの戦略

	トリガー	非協力
移動	移動型トリガー (MT)	移動型非協力 (MD)
固定	固定型トリガー (FT)	固定型非協力 (FD)

3. 戦略の局所的安定性

繰り返し資源提供ゲーム G_2 における4つの戦略の局所的安定性を分析しよう。

戦略の局所的安定性分析とは、ある戦略（既存戦略）が人口集団中のシェアの全体を占めている中にきわめて少数の他の戦略（侵入戦略）が生じたときに、その戦略が人口集団中でシェアを増やしていけるかどうかを判定するものである。一般に既存戦略を x 、侵入戦略を

y , 人口集団全体での戦略 s の平均利得を $u(s)$ としたとき, $u(x) > u(y)$ であれば y は x に侵入できず (x による均衡は安定), $u(y) > u(x)$ であれば y は x に侵入できる (x による均衡は不安定). なお, ここでは侵入戦略は単一の種類であるとする.

G_2 には 4 つの戦略があるため, 既存戦略と侵入戦略の組み合わせは全部で 12 通りある. ここではいくつかの例を示す.

たとえば, 既存戦略が移動型トリガー (MT), 侵入戦略が移動型非協力 (MD) の場合を考えよう. 侵入戦略がごく少数出現するという仮定を具体的に定式化するために, ある 1 つの集団に 1 人の移動型非協力戦略が現れたとしよう.

このとき, 侵入戦略である移動型非協力戦略の人口集団中での平均利得は, この戦略が現れた集団におけるこの戦略の平均利得に等しいから,

$$\begin{aligned} u(MD) &= D(n-1) + (T-1)D(0) - \xi \\ &= \lambda b \frac{n-1}{n} + Tb - \xi \end{aligned}$$

である.

一方, 既存戦略である移動型トリガー戦略については, 侵入戦略が現れた集団における平均利得

$$\begin{aligned} u_1(MT) &= C(n-1) + (T-1)D(0) - \xi \\ &= \lambda b \frac{n-1}{n} + Tb - b - \xi \end{aligned}$$

および, それ以外の集団における平均利得

$$u_2(MT) = T \cdot C(n) = \lambda b T$$

より, 人口集団全体での平均利得は

$$u(MT) = \frac{(n-1)u_1(MT) + n(m-1)u_2(MT)}{nm-1}$$

となる.

よって, 移動型非協力戦略が移動型トリガー戦略に侵入できないための条件 $u(MT) > u(MD)$ は

$$(\lambda-1)nbT + n\xi > (n-1)\lambda b + \frac{n-1}{m-1}b \quad (2)$$

となることである.

侵入戦略が極めて少数であるという仮定より, 集団の人数 n および集団の数 m が十分に大きい場合には ($n \rightarrow \infty, m \rightarrow \infty$), 式(2)は

$$\begin{aligned} &(\lambda-1)bT + \xi > \lambda b \\ \Leftrightarrow T &> \frac{\lambda}{\lambda-1} - \frac{\xi}{(\lambda-1)b} \end{aligned} \quad (3)$$

$$\Leftrightarrow \xi > \lambda b - (\lambda-1)bT \quad (4)$$

となる (記号 \Leftrightarrow は必要十分条件を表す). すなわち, 侵入戦略が極めて少数である場合, (3)もしくは(4)が成り立つときかつそのときに限り, 移動型非協力戦略は移動型トリガー戦

略に侵入できない。つまり、移動型トリガー戦略が移動型非協力戦略に対して安定である。

つぎに、既存戦略が移動型非協力 (MD)，侵入戦略が移動型トリガー (MT) の場合を考えよう。先と同様に、ある1つの集団に1人の移動型トリガー戦略が現れたとすると、 $u(MD) > u(MT)$ となるのは

$$\lambda - m\lambda + nm - 1 > 0 \tag{5}$$

のときである。先と同様に侵入戦略が極めて少数であると仮定すると、式(5)の左辺は1となり、式(5)は常に成り立つ。すなわち、常に $u(MD) > u(MT)$ が成り立つことになり、移動型トリガー戦略はいかなる場合でも移動型非協力戦略に侵入できない。

さらに、既存戦略が移動型トリガー (MT)，侵入戦略が固定型トリガー (FT) の場合を考えてみよう。集団中のすべての個人がトリガー戦略を採用している場合、移動戦略を採用している個人がいても実際には移動は起こらないから、常に $u(MT) = u(FT)$ 。すなわち、移動型トリガー戦略は固定型トリガー戦略に対して中立安定である。

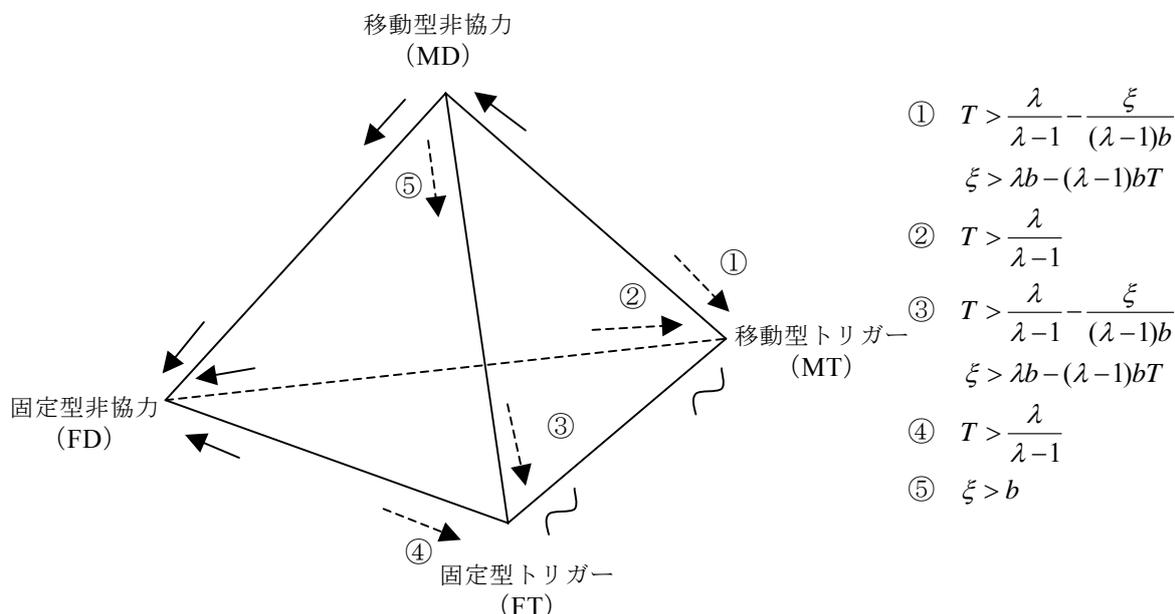


図2 繰り返し資源提供ゲーム G_2 における戦略の局所的安定性

以上のような計算を、12通りの既存戦略と侵入戦略の組み合わせについておこなった結果をまとめたのが図2である。

実線の矢印は、無条件で状態がその方向に遷移していくことを示す。たとえば、移動型非協力 (MD) には移動型トリガー (MT) は常に侵入できないので、図の MD と MT を結ぶ辺の MD 近傍では、MT から MD に向かう方向に状態が遷移する。逆に、移動型非協力 (MD) には固定型非協力 (FD) が常に侵入できるので、図の MD と FD を結ぶ辺の MD 近傍では、MD から FD に向かう方向に状態が遷移する。

波線は、その戦略がその方向に中立安定であることを示す。具体的には、移動型トリガー (MT) と固定型トリガー (FT) は、相互に中立安定である。

破線の矢印は、ある条件を満たした場合に状態がその方向に遷移していくことを示す。た

たとえば、移動型トリガー（MT）に移動型非協力（MD）が侵入できないのは、先の式(3)もしくは(4)を満たすときであるから、図の MT と MD を結ぶ辺の MT 近傍では、この条件を満たすときに限り状態が MT 方向に遷移する (①)。こうした条件付きの状態遷移は全部で 5 つあり、図 2 ではそれぞれの条件を右に示した。

4. 協力的戦略の進化における移動の効果

図 2 で示した各戦略の局所的安定性より、この 4 つの戦略によって構成される戦略空間における状態の時間的推移がある程度推測できる。

まず目につく特徴としては、固定型非協力戦略に向かう力がやはり大きいことが挙げられる。これは Axcelrod 以来指摘されている、非協力的戦略が支配的なときに協力的戦略が侵入することの困難さを、改めて再認識させる結果である。しかし、本稿のモデルの目的は、移動という戦略要素を加えることによって、この非協力へ向かう傾向に何らかの歯止め、もしくは（たとえ一時的にであれ）擾乱を引き起こすことができるかどうかを検討することにある。

そこで以下では、移動に関する 2 つのパラメータ、すなわち移動機会 (T) と移動コスト (ξ) に着目して、これらのパラメータの存在もしくは値の大小が戦略空間の状態遷移にどのような影響を与えるかを分析していく。

なお、移動機会 T の大小は、集団における関係の長期性、もしくは固定性を示すことに注意しよう。もちろん本稿のモデルでは、 T はあくまでも移動機会であるにすぎず、実際に移動するかどうかは各個人の選択に任されている。しかし、少なくとも T 回は移動したくてもできないという意味で、 T は同じ相手との関係を強制する最低限の制約である。その意味でパラメータ T は、繰り返し 2 人囚人のジレンマモデルにおける未来係数に相当するものとみなすことができる⁵⁾。

4.1 非協力が支配的なときに協力が侵入する際の移動の効果

まず、人口集団内で非協力的戦略が支配的であるときに協力的戦略が侵入しようとした場合を考えよう。

先に確認したように、固定型非協力戦略に対しては、移動型・固定型を問わずトリガー戦略は侵入できない。しかし、移動型非協力戦略 (MD) に対しては、固定型トリガー戦略 (FT) が侵入できる可能性がある (図 2 の破線の矢印⑤)。その条件は

$$u(FT) > u(MD) \Leftrightarrow \xi > b \quad (6)$$

すなわち移動コストが提供資源よりも大きいことである。このことの意味は、つぎの 2 つの点にまとめられる。

第一に、もしこの条件が満たされなければ、すなわち移動コストが提供資源よりも小さいか、そもそも存在しなければ、移動型非協力戦略に固定型トリガー戦略は侵入できない。条件の如何にかかわらず、移動型非協力戦略に移動型トリガー戦略は侵入できないし、固定型非協力戦略にトリガー戦略は侵入できないから、条件(6)が満たされないときには、非協力的

戦略に協力的戦略が侵入できる可能性はまったくない。つまり、移動コストがある程度以上存在することは、囚人のジレンマ状況において非協力均衡から脱出する際にきわめて重要な役割を果たしうる。

第二に、式(6)に T が含まれていないことからわかるように、移動型非協力戦略に固定型トリガー戦略が侵入するにあたって、移動機会 T すなわち関係の長期性は影響を及ぼさない。

このことは、第一にこれまでの繰り返し 2 人囚人のジレンマゲームにおける知見と整合性をもつと解釈することができる。すなわち、全面裏切り戦略と TIT-FOR-TAT 戦略からなる繰り返し 2 人囚人のジレンマにおいて、全面裏切りが支配的であるときに TIT-FOR-TAT 戦略は未来係数の値にかかわらず侵入できないという結果 (Axelrod (1984)) と相応している。つまり、協力的戦略が非協力的戦略の侵入を防ぐ際には関係の長期性が有効に作用するが、非協力的戦略に協力的戦略が侵入しようとする際には関係の長期性は有効な手段となりえないということである。

しかし、式(6)に T が含まれていないことの意味はこれだけではない。本稿のモデルにおいては、そのように関係の長期性が非協力均衡からの脱出に直接効果をもたないことを確認しつつも、別の方法による脱出の可能性を明らかにしているからである。すなわち式(6)は、たとえ $T=1$ 、すなわち同じ集団でのゲームの繰り返しがなく、1 回ごとに移動の機会が訪れる場合であっても、移動コストが提供資源よりも大きくさえあれば、固定型の協力的戦略が移動型の非協力的戦略に侵入できるということを積極的に意味するのである⁶⁾。

ちなみに、この場合移動コストが効果をもつのは、比較されているのが移動型非協力戦略と固定型トリガー戦略であるから、すなわち移動コストを負担するのは非協力戦略のみだからである。今の場合固定型トリガー戦略はごく少数の侵入戦略と仮定されているから、そのようにほぼ全員が移動をおこなっているときに、トリガー戦略だけ移動コストがかからないと想定することの是非をめぐっては若干の疑義があることは否めまい。しかし、選択的相互作用と移動コストの存在による非協力均衡からの脱出の可能性を理論的に示したことは、協力的戦略同士が対戦する確率の人為的操作という Axelrod の解決や、集団の離合集散や不均等な分裂というマクロなメカニズムのみを想定する従来の選択的相互作用モデルとは異なる、重要な論点として位置づけられるだろう。

4.2 協力が支配的などときに非協力の侵入を防ぐ際の移動の効果

つぎに、人口集団内で協力的戦略が支配的であるときに非協力的戦略が侵入しようとした場合を考えよう。

協力的戦略に非協力的戦略が侵入するのは全部で 4 通り考えられるが、それらにおいて非協力的戦略の侵入を防ぐための条件は、つぎの 2 つに集約される。

$$u(MT), u(FT) > u(MD) \Leftrightarrow T > \frac{\lambda}{\lambda-1} - \frac{\xi}{(\lambda-1)b} \quad (7)$$

$$u(MT), u(FT) > u(FD) \Leftrightarrow T > \frac{\lambda}{\lambda-1} \quad (8)$$

すなわち、移動型非協力戦略の侵入を防ぐための条件(7)と、固定型非協力戦略の侵入を防ぐ

ための条件(8)である。この2つの条件から考えられる移動の効果は、つぎの3点にまとめられる。

第一に、4.1 でみた非協力的戦略に協力的戦略が侵入する場合と対照的に、協力的戦略が非協力的戦略の侵入を防ぐ際には、関係の長期性 (T) が重要な役割を果たす。すなわち、

$T > \frac{\lambda}{\lambda-1}$ ならば、移動コストの有無にかかわらず、非協力的戦略は協力的戦略に侵入できない。

このことは、協力的戦略が均衡になるためには未来係数の大きさが重要であるというこれまでの繰り返し2人囚人のジレンマゲームにおける知見と整合的である。

第二に、にもかかわらず移動コストが意味をもつのは、関係の長期性が十分でないときである。 $T < \frac{\lambda}{\lambda-1}$ のとき、移動コストの存在は移動型非協力戦略の侵入を阻止することに効果を発揮する。このように、コストをとまなう移動を仮定することは、協力的均衡の頑健性を多少強める働きをすといえる。

第三に、移動コストが効果を発揮するかどうかを決める $\frac{\lambda}{\lambda-1}$ という値は、ある意味をもっている。 λ ($1 < \lambda < n$) とは、資源提供ゲーム G_1 において集団の構成員が提供した資源の総量の価値を高める倍率であった。 λ が大きくなるほど、資源を提供することの(集団全体での)メリットは高くなる。すなわち協力のメリットが増す。 $\frac{\lambda}{\lambda-1}$ は λ が小さいほど大きくなるから、 λ が小さい、すなわち協力のメリットが小さいほど、移動コストの重要性が相対的に大きくなるといえる。

5. まとめと今後の課題

本稿では、集団間で移動可能な N 人囚人のジレンマにおいて協力的戦略が進化していくための条件を、戦略の局所的安定性の観点から分析してきた。その際、移動に関する仮定がどのように影響を与えるかに特に注目した。主な知見は以下の通りである。

まず、非協力的戦略が支配的であるときに協力的戦略が侵入できるためには、移動コストが一定以上あることが本質的に重要である。この場合、先行研究でも示唆されていたように、たんに選択的相互作用(人口集団内の集団構造による関係の固定化)があるだけでは不十分なのであって、移動コストという形で非協力的戦略に選択的に不利益を課すことが、協力的戦略の侵入の可能性を与えていた。これは、協力的戦略同士の恣意的なペアリングや集団の不均等分裂といったこれまでの解決策とはちがった、移動するかどうかの選択という行為者のマイクロなプロセスから導出された新しい解決策と位置づけることができる。

つぎに、協力的戦略が支配的であるときに非協力的戦略の侵入を防ぐためには、第一義的にはこれまでの研究で指摘されていたように関係の長期性が重要であるが、その条件が十分満たされない場合でも、移動コストがある程度あれば、(移動型)非協力戦略の侵入を防げる。さらに、この移動コストの効果は、社会的ジレンマ状況における協力のメリットが小さ

いほど相対的に重要性を増す。このように、移動を個人の選択として明示的に考慮し、さらにそれには一定のコストがかかると仮定することによって、従来の研究で指摘されていたよりも協力的戦略の進化にとって有利に働くメカニズムを見つける糸口を見出せたといえよう。

なお、今回のモデルにおいては、非協力的戦略に協力的戦略が侵入する場面においても、協力的戦略が非協力的戦略の侵入を阻止する場面においても、固定型非協力戦略と比較して移動型非協力戦略のパフォーマンスが悪くなる傾向が見られた。一般に、移動を許すモデルにおいて、移動型非協力戦略は移動型トリガー戦略への対抗戦略とみなすことができる。なぜなら、ある集団から移動型トリガー戦略が離脱したときに非協力戦略がその集団に残るよりも（固定型）、他の協力的戦略の比率の高い集団に移動したほうが（移動型）平均利得が高くなるかもしれないからである。今回の結果は、局所的安定性という限られた状況に限定されているとはいえ、移動型非協力戦略のこうした優位性が必ずしも成り立たない場合があることを示しており、移動することのメリットが協力的戦略と非協力的戦略では異なる可能性があることを示唆していると考えられる。

最後に、今回の研究をさらに発展させるためには、つぎのような方向性が重要になるだろう。第一に、今回のモデルでは各集団の人数 (n) と人口集団における集団の数 (m) が十分に大きいと仮定した。しかし、実際の社会現象ではこれらはそんなに大きくない有限な値をとることが普通であるから、そうした場合の分析が必要である。第二に、それとも関係するが、局所的安定性の理論は一般にプレイヤーや集団の数が十分大きい場合を想定したものであって、その条件が満たされない場合は局所でない状態の遷移、つまり状態空間全体のダイナミクスを細かく見ていくことが必要になる。その際は、少なくとも移動先の選択に関する仮定を明示的に考慮に入れなければならないであろうし、遺伝型よりは学習型のダイナミクスを仮定したほうがより適切であろう。第三に、今回のモデルでは資源提供ゲーム G_1 の利得関数 $C(n_c)$ および $D(n_c)$ を、集団内で戦略 C を採用している人の数 n_c について線形と仮定した（式 (1)）。しかし、（Dawes の 2 条件を満たす範囲で）利得関数が非線形である場合に今回の知見が成り立つかどうか、検討する必要があるだろう。

【注】

- (1) 平成 14 年度～16 年度科学研究費補助金（基盤研究 B）「秩序問題への進化ゲーム理論的アプローチ」（研究代表者：大浦宏邦）では、社会的ジレンマについてこうしたさまざまな角度から学際的に研究を積み重ねてきた。プロジェクトの概要と主要な成果については、たとえば小山他（2004）を参照。
- (2) 進化ゲーム理論における多水準選択モデルの概要と理論的課題については、たとえば金井（2003）を参照。
- (3) もちろん、現実の企業の再編が従業員の意思と独立に（つまり経営陣の判断等によって）おこなわれることがないわけではなく、むしろそういう事例は多い。しかし、ここで問題にしているのは、企業組織内部における構成員の自発的協力的行動を誘発するための条件であり、そのためには企業の再編をフリーライダーに対する対抗策としての自発的移動という個々人のマイクロな戦略に関連づけたほうが分析上有意義である、という趣旨である。
- (4) ここで $\lambda > 1$ なる条件は Dawes の条件(1)に、 $\lambda < n$ なる条件は条件(2)にそれぞれ対応する。
- (5) 繰り返し 2 人囚人のジレンマの利得の大小に依拠して人口集団中での戦略シェアの変動を考えていくという Axelrod 以来のランダムマッチングモデルは、各集団の構成員が 2 人に限られる特殊な離合集散モデルと解釈することもできる。ランダムマッチングモデルでは、2 人のプレイヤーが人口集団中で

ランダムにペアを組んで、一定回数（もしくは確率的に定められる回数） N のゲームを繰り返す。ゲームの回数が N に達するまで、プレイヤーは特定の（1人の）プレイヤーを相手にゲームを繰り返し、それ以外のプレイヤーと相互作用することはない（=2人の構成員からなる集団内に限定された選択的相互作用）。そしてゲームが N 回に達したとき、すべてのプレイヤーはランダムに新しい相手を探して新しいペアを組む。これは、一旦すべての群が解散してランダムに新しい群を決めるという離合集散の特殊な場合に他ならない。

- (6) もちろん $T > 1$ （関係が長期）であっても、移動コストがその条件を満たせば固定型トリガー戦略は移動型非協力戦略に侵入できる。

【文献】

- Axelrod, R. 1984. *The Evolution of Cooperation*. New York: Basic Books. =1987. 松田裕之（訳）『つきあい方の科学』HBJ 出版局。
- Dawes, R. M. 1980. "Social Dilemmas." *Annual Review of Psychology* 31:169-93.
- 林直保子. 1993. 「TIT-FOR-TAT から OUT-FOR-TAT へーネットワーク型囚人のジレンマにおける戦略選手権一」『理論と方法』8(1):19-32.
- Hayashi, Nahoko and Toshio Yamagishi. 1998. "Selective play: Choosing partners in an uncertain world," *Personality and Social Psychology Review*. 2, 276-89.
- 金井雅之. 2000. 「多水準淘汰による協力の進化」東京大学大学院博士論文。
- 金井雅之. 2003. 「進化ゲームにおける選択的相互作用モデルの意義と課題」『理論と方法』18(2):153-67.
- 小山友介・小林盾・藤山英樹・針原素子・谷口尚子・大浦宏邦. 2004. 「社会的ジレンマ問題への学際的接近」『オペレーションズ・リサーチ』49(12):733-40
- Wilson, D. S. 1975. "A Theory of Group Selection." *Proceedings of the National Academy of Sciences of the United States of America* 72:143-46.
- . 1998. "Hunting, Sharing, and Multilevel Selection." *Current Anthropology* 31:73-97.
- Wilson, D. S., and E. Sober. 1994. "Reintroducing Group Selection to the Human Behavioral Science." *Behavioral and Brain Sciences* 17:585-654.

【謝辞】

本稿のモデルの作成および改良にあたっては、平成14年度～16年度科学研究費補助金（基盤研究B）「秩序問題への進化ゲーム理論的アプローチ」（研究代表者：大浦宏邦）においておこなわれた研究会や研究合宿における議論が大いに役立ちました。特に小山友介氏および七條達弘氏には多くのことを学ばせていただきました。

囚人のジレンマゲームにおける移動コストと協力の進化

七條 達弘
(大阪府立大学)

【要旨】

囚人のジレンマゲームにおいて、パートナーの変更という要素を取り入れたモデルを検討する。パートナー変更に伴うコストが、ある程度高い場合には、協力状態が維持できることを示す。さらに、学習の進化が早い場合は、非協力状態を初期状態としても、協力が進化する事を示す。

キーワード：囚人のジレンマゲーム、社会的ジレンマ

1. 囚人のジレンマゲームの先行研究

囚人のジレンマゲームは、社会的ジレンマを表現する最も単純なゲームとして、多くの研究がなされてきた。ここでは、簡単に、囚人のジレンマゲームの先行研究について紹介する。

C と D の二つの選択肢があり、以下の利得行列に従う 2 人ゲームを考える。

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

ここで、利得は、 $T > R > P > S$ 、 $2R > T + S$ であるとする。この条件をみたすとき、このゲームは囚人のジレンマゲームとよばれる。このゲームでは、選択肢 D を出すのが（強）支配戦略であり、互いに D を出す状態が唯一のナッシュ均衡であるが、この均衡状態パレート効率的でないことから、社会的ジレンマを表現していると考えられる。

Axelrod (1984) は、このゲームを繰り返し同じ相手と行う、繰り返し囚人のジレンマゲームを考察し、割引率 δ が十分に大きければ、協力状態が維持可能であることを示した。具体的には、「相手が C を取り続ける限り C を選択するが、一度でも相手が D を出すと、D を出し続ける」というトリガー戦略を考え、これが、 $\delta > (T - R) / (T - P)$ であれば、安定戦略であることを示した。D をとり続ける非協力戦略は、常に進化的に安定な戦略であるため、このことは、協力状態の発生を示しているとはいえないが、協力状態の維持について興味深い示唆を与えている。

2. 移動があるモデル

前節の先行研究では、同じ相手と常に対戦を繰り返す状況を考えていた。しかし、選択の余地なく、同じ相手と対戦し続ける状態は、かならずしも現実的ではない。たとえば、相手が D を出した場合には、パートナーを変えて、別の相手とゲームを行うとしても不思議ではない。そこで、本論文では、パートナー変更という形で移動が行われる場合について考察する。この節のモデルでは、次の仮定をおく。

- 1、 パートナーの変更は常に行う事ができる
- 2、 パートナーの片方がパートナーの変更を希望した場合は、両者が新しいパートナーを見つけてグループを作らなければならない
- 3、 母集団のサイズが大きいために、新しいパートナーをすぐに発見できる
- 4、 新しいパートナーの形成は、ランダムに行われ、人口に比例して、新しいパートナーのタイプが決定する。

戦略としては次の4つのみを考察する。

FD 固定非協力戦略（移動を行わず常に D を出す戦略）

FT 固定トリガー戦略（移動を行わず通常のトリガー戦略を行う戦略）

MD 移動非協力戦略（毎回移動を行い、常に D を出す戦略）

MT 移動トリガー戦略（相手が D を出すとすぐに移動し、常に C を出す戦略）

移動非協力戦略(MD)は、「毎回移動する」と定義されているが、これとは、ことなる移動戦略の定義として、「相手が D を出した場合に移動を行う」という戦略を考えることも可能である。常に D を出す非協力戦略の場合、どちらの定義の移動戦略の利得が高いか考察する。固定非協力戦略(FD)と対戦した場合、1期目に互いに D を出すため、どちらの移動の定義でも2期目に移動をするので、利得の差はない。また、移動トリガー戦略(MT)との対戦においても、1期目に D を出すために2期目に MT 戦略が移動をし、どちらの移動の定義でも、移動を余儀なくされるので、利得の差はない。移動の定義の差が現れるのは、固定トリガー(FT)戦略との対戦のみである。固定トリガーと対戦した場合、1期目には、FT 側が C を出し、MD 側は D を出す。2期目には、1期目に MD が D を出したために FT 側は、D を出す。したがって、MD 側は移動を選択した方が得策である。つまり、移動の定義として、「毎回移動」と「相手が D を出した場合に移動」という二つの選択肢があるが、常に D を選択する戦略としては、「毎回移動」の方が、利得が高い戦略であるということができる。

移動トリガー戦略は、これとは逆に「相手が D を出した場合に移動」と定義されている。これは、上記と同様に、こちらの移動の定義の方が、利得が高いからである。たとえば、MT 同士の対戦の場合、「相手が D を出した場合に移動」であれば、互いに C を出し合う状態を続けることができる。そのため、「毎回移動」より高い利得をえることができるのである。

上記4つのみの戦略空間を考えたときに、固定トリガー戦略(FT)が ESS であるための条件を考察する。戦略 Y が全体を占めている状況化で戦略 X が得る利得を $U(X|Y)$ と表記する。

固定トリガーが全体を占めているときには、対戦相手は、固定トリガー戦略であると考えられる。固定トリガー戦略(FT)は、互いに C を出し合い R の利得をえることができるため、以下の利得をえる。

$$U(FT | FT) = R/(1 - \delta)$$

一方、移動非協力戦略は、相手を変えながら、D を出し続けることになる。仮定 3 より、新しいパートナーとすぐにグループを作ることができ、集団に FT が占めているので仮定 4 より、新しいパートナーは、FT であると考えることができる。このとき、FT は最初に C を出すため、一番高い利得である T をえることができる。したがって、MD の利得は、 $U(MD | FT) = T/(1 - \delta)$ となる。つまり、 δ の値によらず、 $U(FT | FT) < U(MD | FT)$ となっている。そのために、FT の集団に MD が進入可能であり、FT は進化的に安定ではない。

次に MT が進化的に安定な戦略で否かどうかにについて検討する。 $U(MT | MT) = R/(1 - \delta)$ であり、 $U(MD | MT) = T/(1 - \delta)$ であるため、 $U(MT | MT) < U(MD | MT)$ である。したがって、MT の集団に MD が進入可能であり、MT は進化的に安定ではない。このように、協力状態を達成する二つのトリガー戦略が進化的に安定ではなく協力状態が達成されることがないことがわかる。これは、パートナーの変更が行われなかった場合に δ が 1 に近ければトリガー戦略によって協力状態が達成できる事とは対照的である。すなわち、パートナーの変更という、移動可能性は、基本的には協力を困難にするものであることが分かる。

3. 移動コストによる協力の進化

前節では、パートナーの変更を行うには、なんらのコストも発生しないと仮定した。しかし、現実には、パートナーの変更をする場合には、新しいパートナーを探したり、新しいパートナーに対する情報を入手したりするために一定のコストがかかると考えられる。そこで、この節では、パートナー変更を行った場合、移動に伴うコスト $\xi (> 0)$ が発生すると仮定する。そして、移動のコストが適度な値である場合に、移動トリガー戦略(MT)が進化的に安定であり、協力状態を維持することが可能であることを示す。

まず、移動トリガー戦略(MT)戦略が母集団を占めている場合に、MT があげる利得について考察する。このとき、MT は、相手が D を出すまで移動を行わず C を出し続けるため、MT 同士が対戦した場合には、移動をせずに、互いに C を出し続けることになる。したがって、MT の利得は以下ようになる。

$$U(MT | MT) = R/(1 - \delta)$$

一方、MD は、移動をし続けるために¹⁾、毎回移動コストを支払う必要があり、利得は以下ようになる。

$$U(MD | MT) = (T - \xi)/(1 - \delta)$$

したがって、 $\xi > T - R$ のとき、MT の集団に MD が進入することができない。

FD と MT が対戦した場合も MT がすぐに移動するため、FD もパートナーの変更を余儀な

くされ、移動コストを支払う必要があり、 $U(FD|MT) = (T - \xi)/(1 - \delta)$ となり、同様に、 $\xi > T - R$ であれば、FDはMTの集団に進入することができない。

最後に、MTの集団の中にFTが進入できるか否かについて検討する。MTとFTの対戦では、互いに移動をせずCを出し合うために、 $U(MT|MT) = U(FT|MT) = R/(1 - \delta)$ であり、両者の利得は等しくなる。そこで、さらに高次の近似を考察する。まず、高次の利得を記述するため、集団がほぼZで占められているときに、最初の対戦相手がYであった場合に戦略Xがあげる利得を $U(X, Y|Z)$ と表記する。MTの集団の中に、わずかにFTが進入した場合、小さな確率で、最初の対戦相手がFTとなる。このFTとの対戦の利得を計算すると、安定性の2次条件を調べることができるが、先ほどと同様に、 $U(MT, FT|MT) = U(FT, FT|MT) = R/(1 - \delta)$ となり、2次条件においてもMTとFTの利得は等しくなる。したがって、現在、考察している状況では、FTはMTの集団に中立的に進入可能となり、Maynard-Smith (1982)の意味では、MTは進化的安定な戦略ではない。

そこで、より高次の近似を考え、より広義の進化的安定性を考察する。MTの集団の中にFTとFDの二つの戦略が同時に進入した場合を想定すると、小さな確率ではあるが、最初にFDと対戦する可能性がある。MTが最初にFDと対戦した場合には、MTは、Cを出し、FDがDを出すために、MTは低い利得のSを得る。しかし、次の期には、移動をし、MTと対戦することができるため、2期以降は、Rの利得をえることができる。したがって、 $U(MT, FD|MT) = S - \xi + R\delta/(1 - \delta)$ である。一方、FTはFDが対戦した場合、1期目は、MTのときと同様に片方がC、片方がDという状態が発生するが、2期以降は、移動をせず互いにDを出し続けるため $U(FT, FD|MT) = S + P\delta/(1 - \delta)$ である。この二つの利得から $\xi < (R - P)/(1 - \delta)$ であれば、 $U(MT, FD|MT) > U(FT, FD|MT)$ となる。MDも同時に進入する場合を考えると、 $U(MT, MD|MT) = U(FT, MD|MT)$ であり、こちらは、両者に差がない。以上から、 $\xi < (R - P)/(1 - \delta)$ のとき、FTと共にFDなど他の戦略が同時に進入する場合は、FTはMTの集団に進入することができない。突然変異的な試行錯誤によって、小さい確率ながらさまざまな種類の戦略が実行される状況を想定すると、純粹にFTのみが進入する状況は考えにくく、一般に、移動コストが一定の条件を満たすときFTがMTの集団に進入できないと考えてもよいだろう。つまり、移動のコストが非常に大きい場合には、MTより移動しないFTの方が有利となり、MTは進化的に安定ではなくなるが、移動コストが適度に低い場合は、FTが進入することができず、MTが進化的に安定な可能性がある。

MTが進化的に安定であるためには、MDやFDからの進入をうけない条件 $\xi > T - R$ とFTに進入されないための条件 $\xi < (R - P)/(1 - \delta)$ の両方が満たされる必要がある。この両方を満たす移動コスト ξ が存在するか否かについて検討する。両方の条件を満たす、 ξ が存在するためには、 $(R - P)/(1 - \delta) - (T - R) > 0$ である必要があるが、囚人のジレンマの利得の条件より、 $(R - P)/(1 - \delta) - (T - R) \geq (R - P) - (T - R) = 2R - P - T > 0$ となり、両方の条件をみたす ξ が存在することが分かる。ここで、着目すべき点は、 δ に依存せずにこの条件が求まっている点である。すなわち、 δ が極めて0に近い場合であっても、適切な移動コストがあれば、協力状態が維持されるのである。これは、ある意味で、パートナーの変更がない場合よりも広い範囲で、協力状態の維持が可能であることを意味する。

次に、FTの安定性条件について検討する。FTの集団の中にMDが進入したときに得る利得は $U(MD|FT) = (T - \xi)/(1 - \delta)$ であり、そのときFTが得る利得が $U(FT|FT) = R/(1 - \delta)$ であるため、MDからの進入を防ぐには、 $\xi > T - R$ である必要がある。一方、FDからの進入を防ぐためには、通常のパートナーの変更がない場合と同様に、 $\delta > (T - R)/(T - P)$ が必要となる。さらに、MTの進入を防ぐためには、先ほどのMTの集団がFTに進入できない条件の逆をとって、 $\xi > (R - P)/(1 - \delta)$ が必要となる。さきほどの考察より、 $(R - P)/(1 - \delta) > T - R$ なので、 $\xi > (R - P)/(1 - \delta)$ であれば、 $\xi > T - R$ は常に満たされる。したがって、移動コストが十分に高くかつ、 δ が大きい場合には、FTが(拡張された意味で)進化的に安定な戦略となる。

4. 学習による進化

4.1 基本モデル

前節までの議論では、「人口に比例して、新しいパートナーのタイプが決定する」(仮定4)という仮定がおかれていた。これは、出生死滅モデルであり、パートナーの変更を行ったとき、人口に比例して、生まれる新しい個体とパートナーを組むことを想定していると考えられることができる。

しかし、学習によって、行動の変化が起こる場合は、比較的短い時間で、戦略の変化(戦略の進化)が行われる。そのため、新しく生まれる個体を考慮せず、現在の世代内のみでの対戦としてモデルを考えることができる。そこで、次のよう状況を想定する。MTの集団にMDが進入したとし、MTの割合を $1 - \varepsilon$ 、MDの割合を ε であるとする。MT同士のグループでは、パートナーの変更がなく、MTとMDあるいは、MD同士のグループのみがパートナーの変更を行う。つまり、パートナーの変更を行う人の割合は、 $2(1 - \varepsilon)\varepsilon$ の割合を占めるMTとMDのグループと ε^2 の割合を占めるMD同士のグループである。MDの進入があった後、グループの変更を行う人のうちMTがしめる割合は、 $(1 - \varepsilon)\varepsilon / \{2(1 - \varepsilon)\varepsilon + \varepsilon^2\} = (1 - \varepsilon)/(2 - \varepsilon)$ となる。この値は、 ε が小さいときには、 $1/2$ となる。よって、パートナーの変更をして、新しくグループを作ったときに、MTが新しいパートナーとなる確率は、 $1/2$ となる。このような状況を考えたときには、MDが比較的進入しにくくなる事が分かる。

この状況について、詳しく計算を行う。MTの集団に、 $\rho(n)$ をMDが進入後、 n 期目にパートナーを変更した場合に、新しいパートナーがMDである確率とする。上記の例から $\rho(1) = 1/(2 - \varepsilon)$ であり、 $\rho(n+1) = 1/(2 - \rho(n))$ である。ここで、 ε がゼロの極限を考えると $\rho(1) = 1/2$ であり、さらに、 $\rho(n) = (n)/(n+1)$ とすると、漸化式の条件を満たすことが分かる。これから $\rho(n) \geq 1/2$ である。以降では、モデルの簡単化のため $\rho(n) \geq 1/2$ の条件のみで議論を進める。

4.2 トリガー戦略の安定性

この節では、協力状態の維持が可能であるかどうかを調べるため、協力を生み出す、MT 戦略と FT 戦略の安定性について考察する。

まず、MT の進化的安定性について考察する。MT の集団の中に MD が進入した場合得る利得は、次のようになる。

$U(MD | MT) = T - \xi + \delta(1 - \rho(1))(T - \xi) + \delta\rho(1)(P - \xi) + \delta^2(1 - \rho(2))(T - \xi) + \delta^2\rho(2)(P - \xi) + \dots$
 ここで、 $T - \xi > P - \xi$ からこの利得は、各 n について $\rho(n)$ の減少関数である。そこで、 $\rho(1) = 1/2$ で、 $\rho(n) > 1/2$ であるので、 $\rho(n) = 1/2$ を代入して、以下の式を得る。

$$U(MD | MT) < T - \xi + \frac{\delta T - \xi}{2(1 - \delta)} + \frac{\delta P - \xi}{2(1 - \delta)}$$

よって、

$$U(MT | MT) - U(MD | MT) > \frac{1}{1 - \delta} \left\{ R - \xi - (1 - \delta)(T - \xi) - \frac{\delta(T + P - 2\xi)}{2} \right\}$$

よって $\delta > \frac{2(T - R - \xi)}{T - P}$ のとき、 $U(MT | MT) > U(MD | MT)$ となり、MT の集団に MD が進入できない。さらに、 $U(FD | MT) = U(MD | MT)$ なので、同じ条件で、FD が MT の集団に進入できない。また、FT の進入については、3 節の議論と同様の議論を行うことができ、 $\xi < (R - P)/(1 - \delta)$ のとき、FT は MT に進入できない。ここで、興味深いのは、 $\delta > \frac{2(T - R)}{T - P}$ のときには、 $\xi = 0$ であっても MD の進入を防ぐことである。つまり、3 節までのモデルでは、移動のコストがあることが協力の維持に不可欠であったが、この節のモデルでは、移動コストは必要条件ではなくなっている。

次に、FT 戦略の安定性について考察する。MD 戦略との対戦においては、FT 戦略と MT 戦略は、同じ状況となり $U(MD | MT) = U(MD | FT)$ である。さらに、 $U(MT | MT) = U(FT, FT)$ であるため、 $\delta > 2(T - R - \xi)/(T - P)$ のとき、FT は MD 戦略の進入を防ぐことができる。さらに、FT と FD の対戦は、1 節と同じ状況であり、 $\delta > (T - R)/(T - P)$ であれば、FT は FD から進入されない。さらに、 $\xi > (R - P)/(1 - \delta)$ のとき、FT は、MT に進入されない。

以上の事から、 δ が十分に大きい場合には、協力が維持されることがわかる。

4.3 初期進化

非協力的な D を皆が選択する非協力状態を初期状態として、そこからの協力の初期進化について検討する。非協力的な初期状態として FD ばかりの集団と MD ばかりの集団の二通りがありえる。しかし、FD と MD の対戦においては、わずかでも移動コストがあると考えれば、移動コストの分だけ MD は不利であり、MD の集団の中に FD は進入可能であるが、FD の集

団に MD は進入できない ($U(FD|MD) > U(MD|MD)$ $U(FD|FD) > U(MD|FD)$)。そこで、この節では、非協力的な初期状態として、FD ばかりの集団を想定する。

FD ばかりの集団の中に MT が進入した場合の利得について考察する。t 回移動後に FD と対戦する確率を $\rho(t)$ とおく²⁾。さらに、 $r(t) = \rho(t)r(t-1)$ 、 $r(0) = 1$ と定義する。このとき、移動後も t 回連続して MD と対戦する確率は、 $r(t)$ であり、ちょうど t 回目に MT と対戦する確率は、 $r(t-1) - r(t)$ である。一度 MT と対戦した場合は、それ以降、同じ相手と対戦し続けることを考慮すると MT の利得は次のようになる。

$$U(MT|FD) = S - \xi + \sum_{t=1}^{\infty} \delta^t r(t)(S - \xi) + \sum_{t=1}^{\infty} (r(t-1) - r(t)) \frac{\delta^t R}{1 - \delta}$$

よって、平均利得は、以下のようになる。

$$(1 - \delta)U(MT|FD) = (1 - \delta)(S - \xi) + \sum_{t=1}^{\infty} \delta^t \{(1 - \delta)r(t)(S - \xi) + (r(t-1) - r(t))R\}$$

ここで、 δ を 1 に近づけた極限は、 R となる³⁾。

一方、FD の利得は、 $U(FD|FD) = P/(1 - \delta)$ で平均利得は $(1 - \delta)U(FD|FD) = P$ となる。従って、 δ が十分に大きい場合には、 $U(MT|FD) > U(FD|FD)$ であり、FD ばかりの集団に MT が進入可能である事がわかる。ここでも、移動コストの大きさはさほど重要ではない。MD ではなく FD の集団を初期状態と仮定するために、わずかな移動コストがあれば十分であり、 δ が十分に 1 に近ければ、協力の初期進化が可能であることがわかる。

【注】

- (7) あるいは、MT が移動するために、パートナーの変更を余儀なくされ。
 (8) $\rho(t)$ は、t 回移動を行ったという条件の下で、t 回移動後に FD と対戦する確率である。
 (9) ここで、 $\lim_{t \rightarrow \infty} r(t) = 0$ なので $\sum_{t=1}^{\infty} (r(t-1) - r(t)) = r(0) = 1$ であることを考慮した。

【文献】

- Axelrod. 1984. *The Evolution of Cooperation* Basic Books (邦訳 『つきあい方の科学』松田裕介 訳)
 Maynard Smith. 1982. *Evolution and the Theory of Games* Cambridge University Press (邦訳『進化とゲーム理論』寺本 英, 梯 正之訳)

所属集団を変更できる社会的ジレンマモデルの計算機実験

小山 友介 大浦 宏邦
(東京工業大学)(帝京大学)

【要旨】

社会的ジレンマゲームにおいて、集団内に非協力型戦略が多いときには集団を離脱する戦略の有効性が議論されている。また、被験者実験によって普段の行動は TFT 戦略にしたがい、非協力者が増えると集団を移る戦略（離脱型 TFT 戦略）の存在が確認されている。本論文では、離脱型 TFT 戦略 (ET) 戦略の有効性を調べるために、社会的ジレンマ状態での選択を 5 回行うごとに所属集団を変更移動可能な繰り返し社会的ジレンマゲーム（「ビジネスゲーム」）のシミュレータを構築し、シミュレーションを行った。シミュレーションには、離脱型 TFT (ET)、固定型 TFT (FT)、移動型 ALLD (MD)、固定型 ALLD (FD) の 4 戦略による生態系シミュレーションを行い、様々な初期分布において協力型戦略と非協力型戦略の勝利する範囲を調べた。その結果、協力型戦略が十分多ければ非協力型戦略を排除できること、協力型戦略が勝つ領域と非協力型戦略の勝つ領域の間では 4 戦略間の優越関係が循環することが示せた。

キーワード: 社会的ジレンマ, 離脱型 TFT 戦略, 計算機実験

1. 問題意識:社会的ジレンマ

1.1 社会的ジレンマ

環境問題やグループ内協力などを典型的な例として、利害関係者全員の利益を最大化させる方法（社会的合理性）と個人の利益を最大化させる方法（個人的合理性）が一致しない事例は社会に多数存在する。この状態を「社会的ジレンマ状態」と呼ぶ（盛山ら, 1991）。

Dawes(1980)に従って社会的ジレンマ状態を定式化しよう。ある N 人ゲームにおいて、各プレイヤーは二つの選択肢、 C （協力）と D （非協力）を持っている。1 回のゲームで各プレイヤーが得る利得は、自分の選択（ C か D ）と、他のプレイヤーのうちで C を選択した人数（残りは D を選択）に依存する。協力者が x 人のときの協力を選択したプレイヤーの利得を $C(x)$ 、非協力を選択したプレイヤーの利得を $D(x)$ とする。以下の二つの不等式を満たすとき、ゲームが社会的ジレンマ状態であるとする。

1. $D(x) > C(x+1)$: $x=0, 1, \dots, N-1$ とするとき、常に非協力の利得が高い。
2. $D(0) < C(N)$: 全員が協力した場合、各プレイヤーの利益は、全員が非協力の場合の利益より高い。

1 から、「他のプレイヤーの選択にかかわらず、自分は非協力を選択する」ことが支配戦

略となる。2 から、社会的により望ましい（パレート優位な）状態があるにもかかわらず、全員が自己利益最大化行動をとった結果、その状態が達成されないことがわかる。

1.2 囚人のジレンマからの知見

社会的ジレンマ問題は、2 人ゲームの囚人のジレンマの多人数版、と見ることができる。囚人のジレンマで有名な解決策（「解決」は、双方が協力を選択する状態を示す）のうち、著者たちは次の3 つに注目する。

1. 繰り返しゲームのフォーク定理：無限回繰り返しゲームで将来利益の割引率が十分大きいとき、1 回目に協力を選択する戦略の組み合わせがナッシュ均衡となる（松島, 1994）。
2. TFT(Tit-For-Tat)戦略（Axelrod, 1984）：繰り返し囚人のジレンマにおける戦略の一つで、1)初回は必ず協力を選択する、2) 次回以降は前回に相手をとった行動を選択する、というアルゴリズムである。TFT は協力を選択する戦略とは協力しあい、非協力的な戦略に対しては最初の1 回だけ協力してあとは非協力をとるため、被害を最小限にとどめることができる。このため、幅広い戦略に高いパフォーマンスを示す。
3. Out-For-Tat（林, 1993）：繰り返し囚人のジレンマでプレイする相手を選べる状況では、「D をとる相手とは次回から対戦をしないという戦略(Out-for-Tat)」が有利である。

繰り返し2 人囚人のジレンマではこのような解決策があるが、社会的ジレンマゲームではどうだろうか。

1.3 選択的プレイと ET 戦略

社会的ジレンマ問題の文脈で Out-For-Tat 戦略に該当する概念は、Orbell & Daws (1991, 1993)が示した「選択的プレイ」状況(現在の関係から離れて新しい相手とプレイできる状況)である。N 人囚人のジレンマ状況では、協力行動をとる者（C プレイヤー）は他人も C プレイヤーだと認識する傾向が、非協力行動をとるプレイヤー（D プレイヤー）は他人も D プレイヤーだと考える傾向があることが知られている。そこで、プレイヤーに1)「ゲームをしない」ことを選択すること、2)自分の相手となるプレイヤーを探すこと、の2 つを認めるなら、C プレイヤーは協力し合える相手を積極的に探すようになり、D プレイヤーは相互作用に消極的になる結果、C プレイヤー同士がグループを作りプレイする機会が増えることになる。その結果、C プレイヤーの利得が D プレイヤーを上回り、C プレイヤーが淘汰されずに社会内に存続可能となることを Orbell & Dawes(1991,1993)は示唆した。

しかし、「選択的プレイ」に問題がないわけではない。例えば D プレイヤーは、C プレイヤーのグループに侵入して大きな利益を得るという戦略が可能だからである。C 行動の進化には、この D プレイヤーの侵入戦略への対抗戦略が必要となる。

対抗戦略として、「離脱型トリガー戦略」もしくは「離脱型 TFT 戦略 (Exiting Tit -For-Tat : ET 戦略)」が考えられている（大浦ら, 2003ab）。この戦略のアルゴリズムは以下の通りである。

1. 最初は C を選ぶ
2. グループ内で D を選ぶものが現れると D を選ぶ
3. グループ内にある程度以上 D が増えた場合、機会があれば他のグループに移ろうとする
4. 新しいグループでは、また C からはじめる

ET 戦略の採用者は、D プレイヤーによる搾取を TFT 戦略で防ぎつつ、機をみて移動することで同じ戦略が集まり、協力の利益を享受できる可能性がある。そのため移動を行わない単純な非協力者(固定型 D 戦略)や、協力者の多い集団に移動して利益を得ようとする非協力者(移動型 D 戦略)より高い利得をあげる可能性が考えられる。

インターネット内のコミュニティや、ある種の社会ネットワークなど、集団からの離脱と別の集団への移動が容易な状況では、D プレイヤーに対するサンクションが存在しなくても、ET 戦略によって高い協力率が保たれる可能性はないだろうか。

大浦らはこの予想にもとづき、所属集団を変更できる社会的ジレンマ状況で ET 戦略を採用する人間がいるかを次のような被験者実験で調べることを試みた。

2. 実験と結果

大浦らは次のような「ビジネスゲーム¹⁾」を考案して被験者実験を行った。

2.1 ゲーム全体の流れ

N 人の被験者が繰り返し社会的ジレンマゲームを行なう。社会には G 個の企業があり、被験者はいずれかの企業に所属して行動を選択する。被験者は一定回数の選択を行うたびに、所属企業を変更(移動)できる。

1 回のゲーム内で、被験者が社会的ジレンマ環境で行動選択する場面を「通常フェイズ 2)」と呼び、プレイヤーがどのグループに移動するかを選択する場面を「移動フェイズ」と呼ぶ(図 1)。通常フェイズの行動選択は「毎月の働き方」、移動フェイズの行動選択は「半年に一度の転職チャンスにおける転職先の選択」、1 回のビジネスゲームで得る利得は、「数年間の出稼ぎ期間で稼いだ金額」と被験者には教示された。

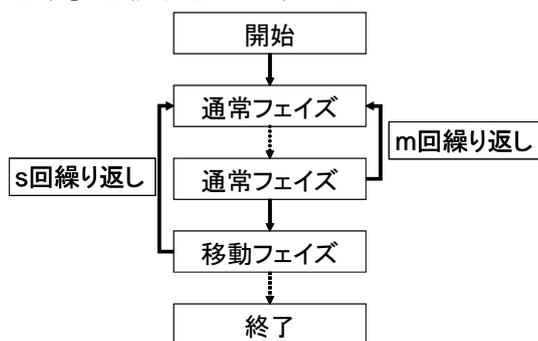


図 1 ビジネスゲームの流れ

2.2 通常フェイズ

通常フェイズでは、プレイヤーは自分の所属する企業で熱心に働く(C)/熱心に働かない(D)を選択する。仕事の成果と報酬は社会的ジレンマゲームと同じ利得構造を持っている。各企業内で熱心に働いた被験者数を N_c 、熱心に働かなかった被験者数を N_d としたとき、企業から支払われる労働への報酬を $\Pi = 4 N_c / (N_c + N_d)$ とし、このときの被験者の選択と利得の関係を以下のように設定する。

- ・ 熱心に働いた(C)プレイヤーの利得： $U_c = \Pi - 2 = 4 N_c / (N_c + N_d) - 2$
- ・ 熱心に働かなかった(D)プレイヤーの利得： $U_d = \Pi = 4 N_c / (N_c + N_d)$

各プレイヤーの利得は、C を選択するより D を選択する方が常に大きい ($D(N_c) > C(N_c + 1)$)。また、自分の所属する企業の全プレイヤーが D を選択したときのときの利得 (=0) は、全プレイヤーが C を選択したときの利得 (=2) より小さい。よって、このゲームは社会的ジレンマ状況を表している。

選択時にプレイヤーが参照できる情報は、(1)各企業の所属人数、(2)直前の自分の利得、(3)自分の企業での前回の C 選択者数と D 選択者数、(4)前回の各企業の Π 、(5)前回の移動フェイズから前回の通常フェイズまでの各企業の Π 、である。

2.3 移動フェイズ

移動フェイズは通常フェイズ m 回ごとに発生し、被験者は任意の企業に移る(「転職」)機会が与えられる。移動した場合には、被験者がこれまで得た利得から移動コスト (≥ 0) を支払う 3)。参照できる情報は、通常フェイズと同じである。

移動フェイズの結果、誰もいない企業が発生することもある。この場合、その企業の所属人数、平均労働報酬をともにゼロとして扱う。

2.4 被験者実験の結果 4)

東京都と山形県の大学生を被験者として、プレイヤー数を 16 人から 20 人、企業数を 4、移動コストをゼロ、通常フェイズ 6 回(半年)ごとに 1 回の移動フェイズ(転職チャンス)という条件で実験を行なった。被験者には 2 年(通常フェイズ 24 回)が経過した後にはその時点で実験を終了する可能性があること、実験後にビジネスゲームで得られた利得に比例した金額の金銭的報酬が得られることを明示した 5)。被験者数は 111 人で、6 グループである。

被験者実験の結果は、(1) 通常フェイズを繰り返すごとに、協力選択率は移動フェイズ直後の約 50%から 30%程度まで徐々に減少した(平均協力選択率は約 40%)、(2) 移動フェイズ直後はまた 50%程度まで協力選択率が上昇した、という既存の繰り返し社会的ジレンマ実験(Thaler, 1992)と類似の結果が得られた。既存の実験と想定は異なるが、被験者の行動は従来の実験の場合と大きく異なっていないといえる。

被験者の行動原理を検討した結果、「ET 戦略」を採用していると考えられる被験者が

全体の3分の1程度存在していることがわかった。また、「移動をしない TFT 戦略」を採用していると考えられる被験者が10分の1程度、ほとんど非協力を採る被験者も10分の1程度存在していた。ET 戦略は常に非協力を選択する戦略より平均利得は低かったが、協力的な戦略の中で平均利得は高い方である、という結果が得られた。

2.5 ET 戦略はどのような場合に有効か?

大浦らの実験結果は、「ET 戦略的な行動は見られたが、必ずしも高い利得を得ていない」ことを示していた。利得が低いにもかかわらず ET 戦略を採る被験者がかなりの割合で存在していたのはなぜだろうか。一つの可能性として、現実社会では ET 戦略の利得が非協力戦略の利得を上回る場合が存在して、被験者の行動はそのような状況で学習／進化してきたことが考えられる。その行動が実験条件下で（実際には不利であるにもかかわらず）発現したのかもしれない。

そうであるなら、ET 戦略の利得が非協力戦略の利得をどのような場合に上回るのかを調べることによって、現実社会で社会的ジレンマ回避に役立つ戦略がどのようにして学習／進化してきたかを明らかにできる可能性がある。

Axelrod (1984)の TFT 戦略と同様、ET 戦略は直接相互作用した他の戦略に直接勝つ（利得で上回る）ことはない。ET 戦略が勝つ可能性があるとするれば、それは「周囲のプレイヤーとの間で互惠関係を育んだ結果、平均的に高い利得を得る」ことで非協力戦略に「間接的に勝つ」場合である。

そのような場合があるのかないのか、あるとすればどのような場合なのか。このことを明らかにするために、本研究では ET 戦略や非協力的な戦略の利得に比例した確率で子孫を残せる生態学的シミュレーションによって、ET 戦略が集団中に定着できる場合があるかどうかを調べることを試みた。

3. シミュレーションモデル

3.1 シミュレーションで用いる戦略

ET 戦略と比較する対象として、対抗する戦略をいくつか用意する。戦略の次元としては、通常フェイズに関わる戦略（行動戦略）と、移動フェイズに関わる戦略（移動戦略）がありうる。ここでは行動戦略として TFT 戦略と ALLD 戦略（常に非協力を選択する戦略）を考え、移動戦略としては D 行動者がいる場合に移動する条件付移動戦略と、常に移動しない固定戦略を考えることにする。

これらの行動戦略と移動戦略を組み合わせると、次の4つの戦略が考えられる。

- ・ TFT 戦略 + 条件付移動戦略 (ET 戦略)

- ・ TFT 戦略+固定戦略 (FT 戦略)
- ・ ALLD 戦略+条件付移動戦略 (MD 戦略)
- ・ ALLD 戦略+固定戦略 (FD 戦略)

もちろん、これらの他にも無数の戦略を考えることは可能であるが、ここで考えた戦略についてもどの戦略が有利であるかを簡単に判定することはできない。分析の出発点としてこれらの4戦略について考えることは意味があると考えられる。

各戦略のアルゴリズムを以下のように定める。2つの協力型戦略 (ET 戦略と FT 戦略、以下 ET, MT と表記) の通常フェイズでの行動は、(1)移動フェイズ直後には協力を選択し、(2)直前の通常フェイズで1体でも非協力を選択したプレイヤーが存在した場合非協力を選択する、という最も厳格な TFT 型行動である⁶⁾。移動フェイズでは、ET は直前の通常フェイズで1体でも非協力を選択していれば移動を決断する。この場合、移動先企業の選択は、(1)直前の通常フェイズの II が最も高い企業、(2) Π が最も高い企業が複数存在するときには、利益が最も高い企業の中からランダムに選択、とする。

2つの非協力型戦略 (MD 戦略と FD 戦略、以下 MD, FD と表記) は、通常フェイズで常に非協力を選択する。MD は移動フェイズで常に移動し、移動先企業決定原理は ET と同じとする。

3.2 戦略分布と各戦略の利得

被験者実験の手続きに準じてシミュレーションを行なった。シミュレーションの設定を表 1 に示す。

表 1 シミュレーションで用いた設定

エージェント総数	100
企業数	10
1回のビジネスゲームの通常フェイズのターン数	100
移動フェイズ	通常フェイズ5ターンごと
移動コスト	0
シミュレーション開始時の戦略分布	初期値として定める (本文参照)
1シミュレーションの繰り返し回数	30回

シミュレーション開始時の各エージェント数は、それぞれが0もしくは5の倍数とし、シミュレーション中のエージェント総数を100に固定する。すなわち、各戦略のエージェント数を ET_n , FT_n , MD_n , FD_n としたとき、常に $ET_n + FT_n + MD_n + FD_n = 100$ とする。

それぞれの初期分布につき30回シミュレーションを行い、戦略ごとの平均利得と全戦略の平均利得⁷⁾を求め、初期分布の全組み合わせ(1771通り)のうち、ETおよびFTの平均利得がMDとFDの平均利得を上回る組み合わせの存在を調べた。

ETの平均利得が非協力型戦略の平均利得を上回ったのは $MD_n \leq 10$ の場合である。特に $MD_n = 0$ のときには、 $FD_n \leq 30$ となるすべての組み合わせで上回り、 $FD_n \leq 50$ となる一部の組み合わせで上回った。

FTに関しては、 $FD_n=0$ の場合には $MD_n \leq 10$ となるすべての組み合わせ、 $FD_n=5$ の場合には $ET_n=0$ 以外のすべての組み合わせで非協力型戦略の利得を上回った。また、 $FD_n \geq 10$ の場合でも、いくつかの組み合わせで非協力型戦略の利得を上回った。

非協力型戦略が多くの場合で協力型戦略の利得を上回るが、ET および FT の利得が非協力型戦略の利得を上回る領域も存在することが明らかとなった。これは、協力型戦略が多数を占める集団に非協力型戦略が少数侵入してきても、侵入を阻止できる可能性が存在することを示している。そこで、戦略シェアの変化をモデルに組み込んだ生態系シミュレーションを行ない、実際に協力型戦略のみの状態に復帰することができるかどうかを調べた。

3.3 生態系シミュレーション

シミュレーションの手続きを以下に示す（図 2）。

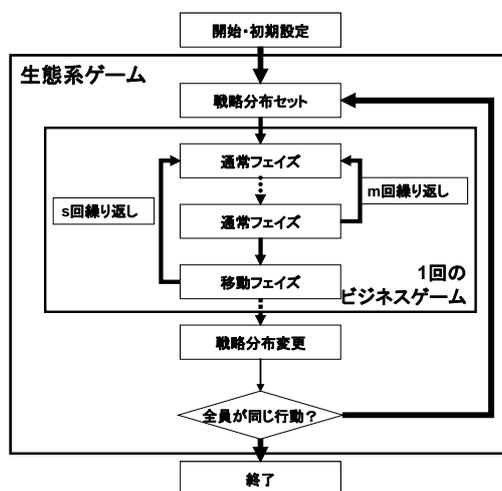


図 2 生態系ゲームの流れ

1. ゲーム開始の初期状態として、プレイヤーが採用する戦略の組み合わせ（初期分布）をセットする。
2. ビジネスゲームを 1 回行う。
3. ビジネスゲームの結果に応じて、戦略の組み合わせ（分布）を変化させる。相対的に高い利得をあげた戦略の採用者数が増加し、利得の低い戦略の採用者数が減少する。プレイヤー数は変化しない。
4. 戦略分布を変化させた結果、全員が常に協力を選択する状態か、常に非協力を選択する状態に到達すれば、そこでシミュレーションは終了し、残った側の勝利数に 1 を加えて 5 にすすむ。そうでない場合には 2 に戻る。2 から 4 を 100 回繰り返した場合、「引き分け」に 1 を加えて次にすすむ⁸⁾。
5. 同じ初期分布で 1 から 4 を 30 回繰り返す。
6. 初期分布を様々に変更して 1 から 5 を繰り返す。

初期分布を設定してから、協力型の勝利／非協力型の勝利／引き分けが決まるまで（1 が

始まって4が終了するまで)を、1回の生態系ゲームと定義する。

1回の生態系ゲームは、ビジネスゲームの繰り返しで構成されている。生態系ゲームには4戦略のみが加わり、各ゲームに参加するプレイヤー数は一定に保たれる。各プレイヤーはビジネスゲームで得た利得に比例した確率で、次のゲームに参加する子孫を残せる(利得の高い戦略は複数の子孫を残すことも可能だが、利得の低い戦略は全く子孫を残せないこともある)。ゲームが繰り返される毎に4戦略の比率は変化する。

3.4 戦略分布変更ルール

1回のビジネスゲーム終了後、次のビジネスゲームに参加するプレイヤーの戦略分布は以下のプロセスで決定される。

1. ビジネスゲームでは最終的に手にした利得がマイナスになることがあるので、最低利得プレイヤーの利得が $\epsilon=0.001$ になるように、全プレイヤーの利得をシフトさせる。
2. シフト後のプレイヤーの利得和で自分の利得を割った値を各プレイヤーの再生産確率とする。
3. 総プレイヤー数と同じ回数だけ乱数くじを引き、次のゲームに参加するプレイヤーを決定する。

戦略分布の変更ルールとしては、相対的に利得の低い戦略を利得の高い戦略に置き換えるなどのルールも考えられるが、利得の高い戦略同士や低い戦略同士の利得の差が過小評価されるという問題点がある。今回用いた方法では、利得の違いを戦略分布の変化に適切に反映させることができる。

3.5 生態系シミュレーションの結果

- 1) 結果の見方

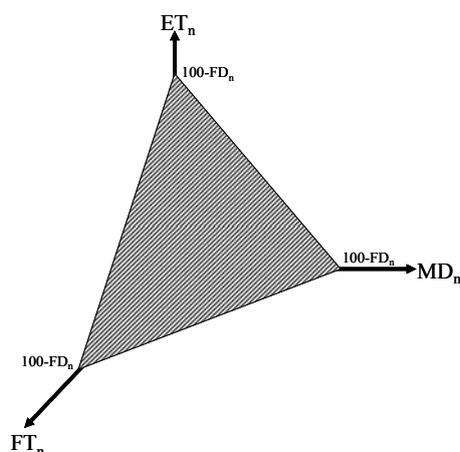


図3 FD_n を固定したときの単体

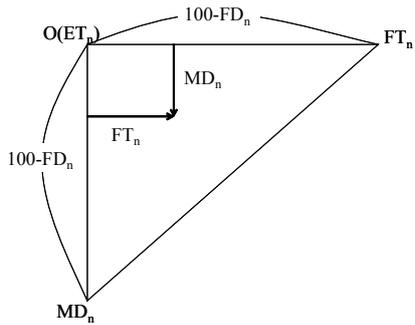


図 4 単体内での FT_n , MD_n , ET_n の表示

エージェントの初期配置空間は 3 次元単体となる．そのまま紙面に表示することは不可能なため，3 次元単体の切断面の 2 次元単体を表示する．図 3 の斜線部分は FD_n を固定したときの他の 3 戦略の組み合わせを表現したものである．この単体（正三角形）を直角二等辺三角形に射影したものが図 4 である．

直角二等辺三角形の原点 O から下にすすんだ距離が MD_n となり，左にすすんだ距離が FT_n となる． ET_n は， $100-FD_n$ から MD_n と FD_n を引いた残りである．

直角三角形の各点でシミュレーションでの戦略の初期配置を示し，30 回のうち協力型戦略が勝利した回数によって白（勝利回数 20 回以上），灰（勝利回数 11 回以上 19 回以下），黒（勝利回数 10 回以下）に色分け⁹⁾した．同じ形式で， MD_n 固定時の結果も議論する．

2) 結果

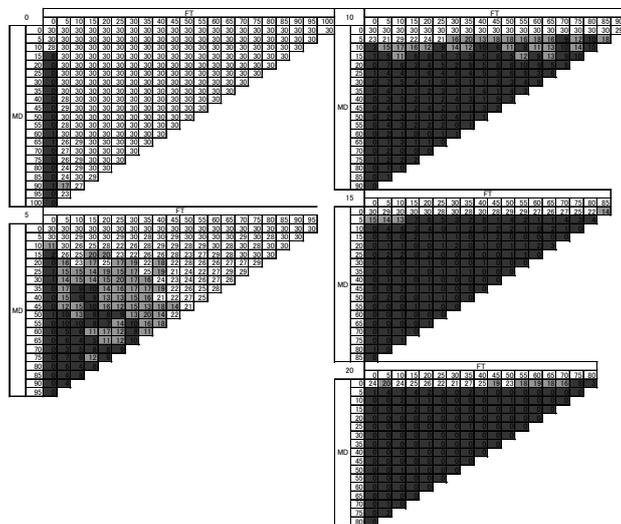


図 5 FD_n の変化による結果の変化

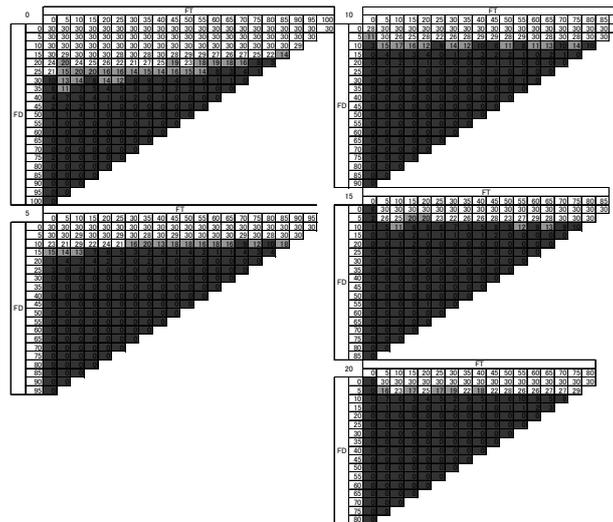


図 6 MD_n の変化による結果の変化

図 5は $FD_n=0,5,10,15,20$ のとき，図 6は $MD_n=0,5,10,15,20$ のときの単体である．左側の上から順に $FD_n(MD_n)=0,5$ のとき，右側の上から順に $FD_n(MD_n)=10,15,20$ のときの結果である．横軸は ET_n (左から右へ 0 から $100-FD_n(MD_n)$ まで)，縦軸は $MD_n(FD_n)$ (上から下へ 0 から $100-FD_n(MD_n)$ まで)を示している． FD_n もしくは MD_n が 25 以上の場合の結果は，非協力型戦略がほぼ勝利するので省略する．

- ・ 大まかな傾向は，非協力型戦略の合計が 10 以下で協力型戦略が勝利，15 および 20 の間では 4 戦略の組み合わせで結果が変化し，非協力型戦略の合計が 25 以上で非協力型戦略が勝利する
- ・ ただし， $FD_n \leq 5$ もしくは $MD_n \leq 5$ 以下の場合，合計が 25 を超えていても協力型戦略が勝利する初期状態の組み合わせが多数存在する
- ・ 特に $FD_n=0$ の場合， $FT_n=0$ かつ $MD_n \geq 15$ の領域を除くほとんどで，協力型戦略が勝利する
- ・ FD_n と MD_n が増えると，左下の協力型戦略が少ない領域から徐々に非協力型戦略の勝利領域が広がるが，FD は左側(FT_n が少なく ET_n が多い)から，MD は右側(FT_n が多く ET_n が少ない)から広がる．
- ・ 勝敗が決した状態での 100 個体中の 2 戦略の内訳は試行ごとの分散が非常に大きい，一方の戦略が絶滅している試行はほとんど無い．

4. 考察

今回シミュレーションした戦略の初期配置は，「協力型戦略が占める空間に，非協力型戦略が侵入した状態」と考えることが可能である．シミュレーション結果は，約 20 程度(協力型戦略戦略の約 4 分の 1 程度)の侵入まで，非協力型戦略を排除可能であることを示してい

る。すなわち、協力型戦略が集まって社会が一度形成されれば、非協力型戦略の侵入に対してかなり頑健である。

ここでは、非協力型の戦略が侵入したときに何が起きているのかを考察することで、なぜ協力型戦略が非協力型戦略の侵入を阻止できるのか考えてみる。さらに、今回の結果の意義について考察を行なう。

4.1 モデルの挙動

移動フェイズと次の移動フェイズの間に何が起きているかを考察する。このとき、移動フェイズ直後の通常フェイズとそれ以降の通常フェイズで結果が異なるので、これらを分けて考える必要がある。

企業内が協力型戦略のみで占められている場合・非協力型戦略だけで占められている場合・双方が混在している場合について、企業内で何が起こるかを追いかけてみる。

1) 移動フェイズ直後の通常フェイズ

協力型戦略のみで占められている企業の場合、全員が協力を選択する。よって、労働への報酬は $\Pi=4$ となり、各プレイヤーの利得は 2 となる。

非協力型戦略のみで占められている企業の場合、全員が非協力を選択する。よって、労働への報酬は $\Pi=0$ となり、各プレイヤーの利得はゼロとなる。

協力型戦略・非協力戦略が混在している企業においては、協力型戦略プレイヤーは協力を選択し、非協力型戦略プレイヤーは非協力を選択する。よって、 $\Pi=4 N_c / (N_c + N_d)$ となる。協力型戦略プレイヤーの利得は $\Pi-2$ 、非協力型戦略プレイヤーの利得は Π となる。

2) それ以降の通常フェイズ

協力型戦略のみで占められている企業では、引き続き全員が協力を選択する。よって、労働への報酬は $\Pi=4$ となり、各プレイヤーの利得は 2 となる。

非協力型戦略のみで占められている企業の場合、引き続き全員が非協力を選択する。よって、労働への報酬は $\Pi=0$ となり、各プレイヤーの利得はゼロとなる。

双方が混在している企業では、非協力型戦略プレイヤーだけでなく協力型戦略プレイヤーも非協力を選択する。よって、労働への報酬は $\Pi=0$ となり、各プレイヤーの利得はゼロとなる。

3) 通常フェイズでの利得の合計

この結果から、通常フェイズで手にする利得の合計が計算できる。通常フェイズが m 回繰り返されるとして、通常フェイズでの利得の合計を計算しよう。

協力型戦略のみで占められている企業では、各プレイヤーは通常フェイズ 1 回あたり 2 の利得を手にしたので、合計 $2m$ の利得を手にする。

非協力型戦略のみで占められている企業では、各プレイヤーの通常フェイズ 1 回あたりの利得はゼロなので、合計利得もゼロである。

協力型戦略・非協力戦略が混在している企業では、協力型戦略プレイヤーは 1 回目の通常フェイズで $4 N_c / (N_c + N_d) - 2$ の利得を手にし、それ以降は 0 である。よって合計利得も $4 N_c$

$/(Nc + Nd)-2$ となる。非協力型戦略プレイヤーは 1 回目の通常フェイズで $4 Nc / (Nc + Nd)$ の利得を手にし、それ以降は 0 である。よって合計利得は $4 Nc / (Nc + Nd)$ となる。

4) 移動フェイズ

移動フェイズを分析するときには、移動する 2 戦略 (ET と MD) のみ考えればよい。

ET は現在所属する企業に所属するプレイヤーが全員協力型戦略プレイヤーなら移動しないが、協力型戦略と非協力型戦略が混在している企業に所属している場合には移動することになる。移動する場合には、直前の通常フェイズの II が最も高い企業に移動する。非協力型戦略のみの企業と双方が混在する企業の II はゼロなので、協力型戦略のみで構成されている企業があれば、そこに移動する (複数ある場合にはその中からランダム)。協力型戦略のみで構成されている企業が存在しない場合には、今所属している企業を除くすべての企業の中からランダムで移動する。MD は必ず移動する。移動先は ET と同じである。

5) 利得の比較

ここまでの考察をもとに、各戦略の利得がどうなるかをまとめてみよう。

協力型戦略が高い利得をあげるチャンスは、同じ企業内に非協力型戦略が存在しないときに限られる。しかし、移動フェイズが終わるとその企業には高い確率で MD がやってくるため、 MD_n が大きいときは連続して $2m$ の利得をあげることは難しい。

非協力型戦略は、同一企業に協力型戦略が存在するときの初めの通常フェイズでの利得しか利得獲得機会はない。このときの利得は、協力型戦略より 2 高い。

協力型戦略が集まったときの $2m$ の利得と、非協力型戦略が協力型戦略と同じ企業に所属したときに手にする協力型戦略より 2 だけ高い利得のうち、最終的にいずれが大きくなるかが分かれ道となる。

非協力型戦略が十分少ないときは、協力型のみの企業の割合が多く、協力型戦略が利得 $2m$ を獲得できる機会が多いため、協力型戦略が有利になる。

4.2 移動コストの影響

紙数の関係で省略したが、移動コストを 1 にしたときは、協力型戦略の勝利領域がより広くなる、という結果となった。ゼロから移動コストを増やす限りにおいて、協力型戦略に有利に働いた。一般に、移動型戦略への移動コストの影響は以下のようなになる。

移動直後の通常フェイズで MD が得る利益は $2+II < 4$ なので (MD がいるため必ず $II < 2$) 移動コストが 4 以上では協力型戦略がいる企業へ移動する利益を移動コストが上回り、MD は移動しない戦略より損になる。

4 未満の移動コストが各戦略に与える影響は、協力型戦略のみで占められる企業数に依存する。協力型戦略のみの企業が多いと、ET は移動しないケースが多くなる。一方、MD は必ず移動する (MD に対して協力型戦略が TFT で対応するため) ため、ET より 1 回のビジネスゲームで支払う移動コストが大きくなる。このことが、協力型戦略に有利に働く。

4.3 戦略間の優越性の循環

非協力型戦略の合計が 10 以下、25 以上の場合では「どちらの戦略が有利か」についてかなり明白な結果が得られた。しかし、中間の区間では、初期分布の組み合わせで戦略の有利・不利が微妙に変化している。

FD は FT_n が少ない(ET_n が多い)領域から、MD は FT_n が多い(ET_n が少ない)領域から黒色の領域が広がったことから、 $ET > FD > FT > MD > ET > \dots$ と、「弱者」→「強者」の関係が循環していることがわかる。

4.4 ET 戦略の有効性

それでは、ET をどのように評価すべきだろうか。(1)協力型戦略が勝利した場合、初期分布で存在すれば ET は必ず最終状態でも存在すること、(2)非協力型戦略の侵入に対して、 MD_n が多い場合には図の左側 (ET_n が多い領域)から、 FD_n が多い場合には図の右側 (FT_n が多い領域)から、非協力型戦略の勝利領域が広がること、(3) $ET_n=0$ では MD と FD が FT の利益を上回るが、 $ET_n \geq 5$ では FT が MD と FD の利益を上回るような 4 戦略の組み合わせが多数存在することから、「ET が一定数存在することで FT の非協力型戦略への有効性が増す」というべきだろう。

4.5 結論と今後の課題

本論文では ET・FT・MD・FD の 4 戦略が加わる「ビジネスゲーム」モデル（移動可能な社会的ジレンマモデル）を構築し、100 ターンからなるゲームの利得と、それに応じて次のゲームの戦略分布が変化する生態系ゲームを議論した。そこで得られた結果は(1) 協力型戦略の集まりに非協力型戦略が侵入した状態を想定したとき、少数であれば排除可能、(2)侵入した非協力型戦略数の中程度のとき、協力型戦略の組み合わせと非協力型戦略の組み合わせによって勝敗が大きく変化する。(3)協力型戦略が勝つ領域と非協力型戦略の勝つ領域の間では 4 戦略間の優越関係が循環する、というものであった。

特定プレイヤーにサンクションを与えない社会的ジレンマモデルでも、協力型戦略が非協力型戦略の侵入を排除できる領域の存在を示せたことは大きな意義を持つ。社会的ジレンマに関する従来の研究は、外部からのサンクションや、プレイヤーに社会的動機などの協力をポジティブに評価する動機を仮定することでジレンマが回避されていると考えてきた（大浦 2003）。しかし、外部サンクションには誰がサンクションコストを引き受けるのかという問題があり、社会的動機には一般的に損をする協力をなぜそのような動機で行動するのか、という問題が存在する。今回の結果は、これらの問題点を含まないジレンマ回避モデルが構築可能であることを示している。

次に、協力型戦略を圧倒しない程度に非協力型戦略が増えた場合に戦略間に優越関係の循環があり、非協力型戦略の侵入を防げる状態が 2 戦略の混合で示せることは、「実際の社会でどのように非協力型戦略が排除されているか」を知る一つの示唆となる。

今後は、プレイヤーが採りうる戦略を多様化して分析する必要がある。今回は4戦略のみを仮定する分析であったが、突然変異によって新しい戦略が発生する場合を想定したシミュレーションを行なうことで、今回の結果の頑健さを調べる必要がある。

また、今回の結論から、複数の種類の非協力型戦略が存在しているときには、「ある同質な戦略によって協力が維持されるのではなく、複数の戦略の組み合わせと分布によって維持される」場合があることを示した。この視点から、「非協力型戦略を排除しつつ、自分たちが協力の果実をできるだけ享受可能な戦略群」を調べることも、重要である。

【注】

- (10) 通常、「ビジネスゲーム」は学生が経営学的な諸概念を容易に理解・体得するために作られたゲームを指す。実験では、被験者のイメージがわかりやすいためにこの名前を用いた。
- (11) 以降本論文では繰り返し社会的ジレンマゲーム全体で「1回のビジネスゲーム」と数えることにし、以下で示すように1回の社会的ジレンマゲームは「通常フェイズ」と呼ぶことにする。
- (12) 被験者実験では、被験者の意志決定コストを低減させるために移動コストをゼロに設定した。
- (13) 被験者実験の結果について、ここでは必要最小限の結果だけ述べる。詳細は大浦・石原・小林(2003), Kobayashi et. al(2003), 針原・辻(2003)などを参照してほしい。
- (14) 被験者に支払う金額の予算総額をあらかじめ決めておき、被験者の利得に応じて比例配分する、という方法とった。
- (15) ここでの平均利得とは、30回のシミュレーション全体での、各戦略の平均利得である。すなわち、ある戦略分布での*i*回目のシミュレーションでの戦略ETの平均利得を $P(ET^i)$ としたとき、この戦略分布での戦略ET平均利得は、 $\{\sum_i P(ET^i)\}/30$ となる。
- (16) 社会的ジレンマモデルのTFT戦略には、「グループ内協力率がどこまで下がれば自分の選択を非協力に変更するか」についての基準がない。分析を容易にするため、本稿ではトリガー戦略に近い基準を採用する。
- (17) 全プレイヤーが協力型戦略だけ、もしくは非協力型戦略だけになった状態で終了するのは、この状態では全プレイヤーの通常フェイズでの選択が同じになり、分析上無意味であるためである。また、初期分布次第ではなかなか勝負がつかない（ビジネスゲームを数百回繰り返しても勝負がつかない）場合があるため、引き分けを用意している。
- (18) 境界はそれぞれ、「協力型戦略と非協力型戦略の勝利確率は等しい」という帰無仮説の片側5%棄却点となっている。

【文献】

- Axelrod, R. (1984), *The Evolution of Cooperation*, Basic Books, New York (松田 裕之訳 (1998), 『つきあい方の科学—バクテリアから国際関係まで』, ミネルヴァ書房) .
- Axelrod, R. (1997), *The Complexity of Cooperation*, Princeton University Press, Princeton, New Jersey, (寺野隆雄 監訳 (2003), 『対立と協調の科学—エージェント・ベース・モデルによる複雑系の解明』, ダイヤモンド社) .
- Dawes, R., (1980), *Social Dilemmas*, *Annual Reviews of Psychology*, 31, 169-93.
- 針原素子, 辻竜平 (2003), 「所属集団の変更可能なSDゲームにおける協力行動の規定因: 規制の小さい状況でも発動する他者からの疎外回避動機」, 第35回数理社会学会大会, 大分大学
- 林直保子(1993), 「TIT-FOR-TAT から OUT-FOR-TAT へ: ネットワーク型囚人のジレンマによる戦略選手権」, 『理論と方法』, 8, 19-32
- Jun Kobayashi, Hirokuni Ooura, and Hideki Ishihara (2003), "Coevolution of Cooperation and Selective Interaction: An Experimental Study," Paper presented at International Social Network Conference. February 16. Cancun, Mexico.
- 松島斉(1994), 「過去、現在、未来: 繰り返しゲームと経済学」, 伊藤 元重 岩井 克人 (編) 『現代の経済理論』, 東京大学出版会, 1994.

Orbell, J. and Dawes, R. (1991), A cognitive miser theory of cooperators' advantage, *American Political Science Review*, 85, 515-528.

Orbell J. and Dawes, R.(1993), Social welfare cooperator's advantage, and the option of not playing the game, *American Sociological Review*, 58, 787-800.

大浦 宏邦(2003), 「秩序問題への進化ゲーム理論的アプローチ」『理論と方法』, 34,133-151

大浦博邦 , 石原秀樹 , 小林盾(2003a), 「所属集団を変更できる社会的ジレンマ実験(1): その意義と実験手続き」,第 35 回数理学学会大会,大分大学

大浦博邦 , 石原秀樹 , 小林盾(2003b), 「所属集団を変更できる社会的ジレンマ実験(2): その意義と実験手続き」,第 35 回数理学学会大会,大分大学

盛山和夫 , 海野道郎(編)(1991),『秩序問題と社会的ジレンマ』,ハーベスト社

Thaler, R. (1992), *The Winner's Curse ---Paradoxes and Anomalies of Economic Life*, Princeton University Press,
(篠原勝訳 (1998), 『市場と感情の経済学—「勝者の呪い」はなぜ起こるのか』,ダイヤモンド社).

【謝辞】

本原稿は「シミュレーション&ゲーミング」, Vol.13 No.2 169-178,2003 に掲載されたものを, 研究会での合意をもとに一部用語(実験中の各場面の用語, 戦略の用語)を統一したものである. 掲載を許可して頂いた, 日本シミュレーション&ゲーミング学会に感謝する.

社会的ジレンマと協力概念

——フリーライドの視点による N 人ゲームの構造分析 I ——

武藤 正義
(東京工業大学)

【要旨】

本稿の目的は、Dawes (1980) の後、Taylor (1987) や Elster (1989) らによって拡張された種々の社会的ジレンマ概念を、フリーライドの視点から整理することにある。具体的にはつぎの 2 つの課題にとりくむ。第 1 に、種々の社会的ジレンマを協力行為が定義される状況 (= 協力状況) によって包括する。協力行為 (協力状況) は、「協力行為のコスト」と「共通の利益の存在」という 2 条件によって定義される。第 2 に、多元的な個人的合理性にもとづいて、協力状況の下位にある種々の社会的ジレンマを整理する。すなわち、 N 人囚人のジレンマ、 N 人円満ゲーム、 N 人保証ゲーム、 N 人チキンゲームなどの諸状況を「均衡の配置」から統一的に定義し、それらを個人的合理性の異なる社会的ジレンマとして位置づける。結果として、つぎの 4 点が明らかになった。①Dawes (1980) の「社会的ジレンマ」と「協力状況」は独立である。②Taylor (1987=1995) の先手コミットメントやナッシュ均衡戦略などの多元的な個人的合理性を考慮することで社会的ジレンマ概念を拡張・精緻化すると、社会的ジレンマは協力状況に含まれる。③さらに、進化的な合理性を個人的合理性にカウントすると、社会的ジレンマは協力状況に一致する (N 人円満ゲームでさえ社会的ジレンマといえる)。④協力行為は、個人にとってリスクがあるだけでなく、全体にとってもリスクがある「過剰協力」をひきおこすこともある。重度の過剰協力がおこる状況では、集団淘汰 (group selection) によっても協力状態は実現できない。本稿は N 人ゲームという複雑な社会状況の理解を刷新するものとして位置づけられよう。

キーワード 協力のコスト、共通の利益の存在、多元的な個人的合理性、過剰協力

1. はじめに

社会的ジレンマの数理構造の分析には、すでに多くの先行研究がある。しかしそれらはフリーライドという視点からみて、十分に整理されているとはいえない (フリーライドの問題性は明らかだろう)。そこで、本稿はこの視点から古典的なゲーム理論を用いて、社会的ジレンマの基礎概念の整理を行う。

本稿と進化ゲーム理論との関係を述べておこう。進化ゲーム理論的研究では 2 人ゲームのランダムマッチングが多く、 N 人ゲームの分析はあまりみられない。これには古典的なゲーム理論の範囲でも N 人ゲームの分析がすすんでいないということが一因にあるだろう。本稿はこれを補うものである。つまり、社会的ジレンマへの進化ゲーム理論的アプローチの理論的基礎として、本稿は (基本的には) N 人ゲームを古典的なゲーム理論の範囲で精査する¹。

¹ 私見では、日本の数理社会学における社会的ジレンマ研究は、盛山・海野 (1991) において一応の区切

用語上の注意として、本稿では単に「協力」と書く場合には行為を表し、行為の組である社会状態（協力状態）を表すものではないとする。また、本稿における均衡とは、純粋戦略ナッシュ均衡を意味する。

1.1 諸問題

社会的ジレンマ研究に用いられる諸概念には、すくなくとも3つの問題点がある。

第1に、**協力的行為**概念の曖昧さがあげられる。社会的ジレンマ研究では、行為選択肢（戦略）に協力(C)/非協力(D)というラベルが使われる（Cは Cooperate, Dは Defect の略）。このラベルはもともと囚人のジレンマゲーム（Dawes 1975 など）で使われていたものだが（以下「ゲーム」を省略）、チキンゲーム（Taylor 1987=1995: 30）や保証ゲーム（山岸 1991: 240）でもこのラベルが使われるようになった（表 1, 2, 3）。そしてこのことはこれまで特に違和感なく受けとめられてきた。

表 1 囚人のジレンマ

	C	D
C	3, 3	1, 4
D	4, 1	2, 2

表 2 チキンゲーム

	C	D
C	3, 3	2, 4
D	4, 2	1, 1

表 3 保証ゲーム

	C	D
C	4, 4	1, 3
D	3, 1	2, 2

注. 太字は均衡を示す。他表も同様。

しかし、ゲームが拡張されれば、協力や非協力の意味合いも元の囚人のジレンマでのものとは変わってくる。すくなくとも、囚人のジレンマ、チキンゲーム、保証ゲームの3つにみられる共通の構造に依拠した形で、協力/非協力の意味が考えられるべきである。そうでなければこのラベルは類推でつけられたままになってしまいうだろう。

第2に、**協力状態**概念の曖昧さがあげられる。表 1, 2, 3 の3つの2人ゲームでは、両行為者が協力をとった社会状態(C, C)が、ふつう協力状態とよばれる。一方、非協力状態はふつう(D, D)とされるが、この意味が非・協力状態であるならば、(D, C)も(C, D)も含まれるべきだろう。

第3に、特にN人ゲームにおいて、社会的ジレンマの**分類カテゴリー**の曖昧さがあげられる。うへの表 1, 2, 3 は2人ゲームの利得表であったが、ふつう社会的ジレンマはN人ゲーム（すなわち3人以上のゲーム）で定義される。しかし、社会的ジレンマ研究で用いられるN人チキンゲームとかN人保証ゲームなどの諸ゲームのクラス概念は、2人ゲームからの類

りを迎え、その後、共有知識、社会的動機、進化ゲーム理論、ネットワークなどが数理社会学におけるゲーム理論的研究に導入された。このとき、N人ゲームは構造が難しいので、2人ゲームにおいてもつばらこれらの新しいデバイスが使われた。このために、N人ゲームの分析は、木村（2002）など小数の例外の例外を除くと、理論的にはいったん下火になったものと考えられる。とはいえ、ゴミ問題等、社会的ジレンマの経験例は2人ゲーム（囚人のジレンマ）ではなく、N人ゲームのほうがはるかに多いのであり、調査、実験等の実証的研究ではやはりN人ゲームが用いられている。近年では、理論的にも大浦（2002）が集団淘汰（group selection）の再検討により、N人ゲームを再び問題にしている。

推で、全体として未整理な形で用いられている。くわえて、 N 人ゲームでは、なにを協力状態とよび、なにを非協力状態とよぶのがさらに曖昧になってくる。

以上、3点の問題を解消すべく本稿では、協力行為、協力状態、諸 N 人ゲームをシステムティックに定義する。

1.2 先行研究と本稿の目的

先行研究としては、Dawes (1975, 1980), Taylor (1987=1995), Raub (1988), 永田 (1988), Elster (1989=1997: Ch13), 瀧川 (1989), 海野 (1988), 三隅 (1993), 木村 (2002), 鈴木 (2003) などがある。しかしまず、Dawes (1975, 1980) の社会的ジレンマの定義は狭すぎる。たとえば、Taylor は N 人チキンゲームを社会的ジレンマとみなしており、これは Dawes の定義には入らない。とはいえ、Taylor (1987=1995: 56-60) も N 人チキンゲームの定義をきちんと数理化してはいない。数理化の曖昧さは、Elster (1989=1997: Ch13) にもいえる。一方、海野 (1988), 三隅 (1993), 木村 (2002) は、経験的な資源の性質を数理化するという方向性ゆえに、くわえて Dawes の社会的ジレンマ定義に依拠しているゆえに、社会的ジレンマ一般のゲーム構造そのものは十分分析してはいない。

Dawes とは対照的に、Raub (1988) および永田 (1988) の社会的ジレンマ (問題状況) の定義は広すぎる (永田と Raub は独立にほぼ同じ内容の定義をしている)。Raub は Harsanyi (1977) を参考に「問題状況」を包括的に定義しているが、特にこれは $m \times n$ 行列の一般のゲームについてのものであって、実験・調査研究でよく用いられる2選択肢ゲームとしての社会的ジレンマに特有の性質を明らかにしてはいない。一方、瀧川 (1989) は2選択肢ゲームについて構造分析を試みているが、線形の場合のみであるにもかかわらず分析が複雑ゆえ実用上使いにくい (これは利得和の概念を用いているためでもある)。また、鈴木 (2003) は比較的明確に社会的ジレンマを定式化してはいるが、ステップ財に限ったものであり、限定的過ぎる。

このように、先行研究では N 人2選択肢ゲームとしての社会的ジレンマ一般の構造はきちんと分析してはいない。よって本稿の目的は、協力行為と協力状態を定義し、そして社会的ジレンマのサブカテゴリーにあたる N 人チキンゲーム、 N 人保証ゲームなどをシステムティックに定義することによって、社会的ジレンマ研究の基盤となる諸概念を整理することにある。ただし、本稿における協力概念は、フリーライドの視点から構成する。だから本稿の協力概念が唯一絶対というわけではない。

本稿の構成を述べる。まず、第2節では、協力/非協力という行為ラベルが使われるにふさわしいゲームのクラス (協力状況) を定義する (つまり協力行為を定義する)。また、この状況において、協力状態、非協力状態なる用語を定義する。つぎに第3節では、社会的ジレンマと協力状況の関係を多元的な個人的合理性の観点から議論する。さらに、 N 人チキンゲームや N 人保証ゲームを定義する。第4節はまとめである。

2. 協力状況と協力状態

この節では、協力行為が定義される状況として、「協力状況」なるものを定義する。協力という行為の意味は、ゲーム理論的には利得構造（利得関数の組）によってきまる。

2.1 協力行為の定義

Elster (1989=1997: 152) は、協力行為 (cooperate) を「何人かが（あるいは全員が）ある行為をすればすべての人が便益を受けるような行為を、自分個人の利害に逆らって行うこと」として定義している。同著のほかの部分と照らし合わせると、「自分個人の利害に逆らう」は、明らかに被支配戦略を意味している。しかし、Taylor (1987=1995: 26, 30) は、支配戦略のない N 人チキンゲームにおいても協力行為 (cooperate) という用語を用いる。そこで本稿では基本的に、Elster の協力行為定義の拡張を試みる。

付録にあるように、“cooperate”の意味を英語辞書で調べると、“To work or act together toward a common end or purpose.”とか“**To form an association for common, usually economic, benefit**”などとある（太字は引用者による）。このように多くの辞書が“cooperate”の意味に、“a common purpose”, “end”, “aim”, “benefit”すなわち「共通の目的」や、より弱い意味での**共通の利益**を認めている。これは“cooperate”が、たんに“work together”「いっしょに働く」だけでなく、共通の利益にむかって方向づけられた行為であることを意味する²。

また、work や form an association には、明らかに労働や資金等のコストがかかる。「もっと家事に協力して」とか「募金にご協力ください」などというように、日本語の協力という言葉にもそのコストが前提にある。そこで、つぎのように「協力」を定義する。

定義 1 協力とは、（他者ととともに）コストを払うことで共通の利益をもたらしうる行為のことをいう。

「他者ととともに」に括弧をつけたのは、他者の行為を選択することはできないからである。この定義のポイントを述べよう。

1. 「コストを払う」ことは、それじたいとしては自分自身にとって不利益である。つまり、協力行為は、個人的な不利益によって、共通の利益を生みうる行為である。
2. 「共通の利益」について。たとえ自分だけしか協力しなくても、それは共通の利益を促進しうるから、協力者たる自分が利益を受け取る時はかならず、非協力者も同様の利益を受け取る。つまり、協力はこの行為者たちにとっての「公共財」を生みうる行為である。

2.2 協力行為のための必要条件

² 共通の利益は、社会にとって根幹的である。社会とは、人と人との関わりあいの総体であるが、この関わりあいは、一般形として、互いに正の外部性（利益）と負の外部性（害）をもたらさう（いうまでもなくそのモデルがゲームである）。その特殊形として、互いに正の外部性をもたらさうことが、典型的な共通の利益である。ただし、意味を広くとれば、共通の利益は、害がなくなることをも含む。つまり、共通の利益の本質は、「パレート改善」である。

協力行為をゲーム理論的に定式化してみよう。行為者を均質的とする。このときゲームは対称である。また行為選択枝は、協力、非協力の2つのみとする。つぎのように記号を定める。 N, m は整数とする。

- N 行為者の人数
- m 社会状態における自分自身を含めた協力者の人数, ただし $m \in [0, N]$
- $v(m)$ m 人協力しているときの, 1人あたりの共通の利益
- $c(m)$ m 人協力しているときの, 1人あたりの協力行為のコスト, $c(m) > 0$

注意点が2つある。第1に, 協力の利益 $v(m)$ も協力のコスト $c(m)$ もともに協力者の人数 m によって変化しうる。第2に, $v(m)$ は負の値, つまり共通の不利益(害)をも表すことができる。利益, 不利益の絶対量ではなく, 相対的な利益の差だけが問題である。

うへの記号を用いて, 協力行為を表現する式が満たすべき必要条件を考えよう。

第1に, 協力者だけがコスト $c(m)$ を払い, 協力者も非協力者も等しく共通の利益 $v(m)$ を受け取るから,

$$\text{協力者の利得} = v(m) - c(m), \quad \text{非協力者の利得} = v(m)$$

が成立する。したがって, 非協力者は協力者の貢献につねに「フリーライド」する,

I) 非協力者の利得 > 協力者の利得

なる関係(協力行為のコスト)が成立する(支配戦略関係ではない)。

第2に, 共通の利益について考えよう。協力行為は, コストを払っているにもかかわらず, つねに共通の利益を生むとは一般にはいえない。コストを払って協力しても, かえってだれもに不利益を与えることもある。たとえば, Elster (1989=1997: 156-8) は,

- 「一方的な軍縮は, 軍事バランスを崩すのでかえって緊張を高める」
(「過少協力」状態のパレート非効率性),
- 「パーティのあとの皿洗いは, 全員でやるとむしろ非効率」
(「過剰協力」状態のパレート非効率性)

という例をあげている。このような状況では, 協力行為は単にひとりだけが損をするというリスクだけでなく, だれもが損をするというリスクをもつ。

しかし, 協力行為がまったく共通の利益を生む可能性がなければ, それはもはや協力行為とは定義上よばれない。したがって, 協力行為が共通の利益を生むためには, すくなくとも全く協力者のいない社会状態よりはだれにとってもよいような, 協力者がいる社会状態が存在しなければならない。つまり, 協力行為には,

II) 全員非協力の状態 $<_p$ 協力者がいるある状態 ($<_p$ は狭義パレート優越関係)

という条件 (**共通の利益の存在**) が必要である。結局、この条件は「全員非協力がパレート非効率」ということである。そしてこれは、

$$v(0) < v(m_1) - c(m_1) \quad \exists m_1 \in [0, N]$$

という式と同値である。というのは、 $v(m_1) - c(m_1)$ は非協力者の利得である一方、協力者の利得は $v(m_1)$ であり、 $c(m_1) > 0$ より、 $v(0) < v(m_1)$ も成り立つので、このとき協力者、非協力者ともに、だれもが非協力者である状態 $m=0$ よりよいからだ。

2.3 協利行為のための十分条件

I) II) は、ある2つの行為選択肢が協力/非協力であるための必要条件であった。しかしじつは、これは十分条件でもある。このことを示そう。ある2つの行為選択肢 A, B において、

I) B 選択者の利得 $>$ A 選択者の利得

II) 全員 B を選択したときの結果 $<_p$ A 選択者が存在するある結果

という2つの関係が成立していれば、I) から A 選択者の人数を m として、

$v(m)$: = B 選択者の利得

$c(m)$: = B 選択者の利得 $- A$ 選択者の利得 (> 0)

が定義できる。このとき、 A 選択者の利得 $= v(m) - c(m)$ 、 B 選択者の利得 $= v(m)$ となり、かつ II) から選択 A が共通の利益をもたらさうから、 $c(m)$ は選択 A のコスト、 $v(m)$ は共通の利益として解釈できる (B 選択者の利得を共通の利益とよべるのは、 A 選択者のコスト c が操作的に定義されているからである)。よって、 A が「協力」であり、 B が「非協力」であることがわかる。よって、I) II) は協利行為が成立するための十分条件である。

2.4 協利行為を定義できる状況としての協利状況

結局、I) II) はうえで定義した協利行為概念の必要十分条件と考えられる。もちろん、本稿の協利行為の定義はフリーライドを念頭においたものなので、協利行為の定義がこれではなくてはいけないということではない。しかし、うえの協利行為の定義は、協利のコスト $c(m)$ が協力者 m の関数であるためにかなり広いものである。以上の議論から、つぎの「協利状況」を定義する。

定義 2 I) II) をみたす対称ゲームを**協利状況**とよぶ。

協力状況では協利行為、非協利行為が定義できる。協力状況は当該の行為者たちにとって、彼らに公共財の供給が社会的に求められる状況である。

2.5 2人協力状況および協力の過剰と過少

状況の理解のために、2人の場合を考えておく(表4)。この2×2対称ゲームが協力状況であるためには、

$$v(2)-c(2) > v(0) \quad \text{または} \quad v(1)-c(1) > v(0)$$

が成り立つことが必要十分である。ここで簡単のため、

$$\begin{aligned} x &:= v(2)-c(2) & y &:= v(1)-c(1) \\ z &:= v(1) & 0 &:= v(0) \end{aligned}$$

表4 2人ゲームでの協力状況

	協力	非協力
協力	$v(2)-c(2), v(2)-c(2)$	$v(1)-c(1), v(1)$
非協力	$v(1), v(1)-c(1)$	$v(0), v(0)$

表5 表4の略記

	協力	非協力
協力	x, x	y, z
非協力	z, y	$0, 0$

とする(表5)。このときうへの不等式は、 $(x > 0 \text{ または } y > 0)$ かつ $(z > y)$ とかける。このような対称ゲームを選好順序だけでかいてみると、つぎのようになる。

① $x > z > y > 0$

② $x > z > 0 > y$

③ $x > 0 > z > y$

④ $z > x > y > 0$

⑤ $z > y > x > 0$

⑥ $z > x > 0 > y$

⑦ $z > y > 0 > x$

表6 2人ゲームでの具体的な協力状況

	C	D	①円満	②保証	③調整
C	x, x	y, z	<u>3, 3</u>	<u>2, 2</u>	<u>1, 1</u>
D	z, y	$0, 0$	$2, 1$	$1, -1$	$-1, -2$
	④チキン	⑤指導者	⑥囚人	⑦英雄	
	<u>2, 2</u>	<u>1, 3</u>	$1, 1$	<u>2, 3</u>	$1, 1$
	<u>3, 1</u>	$0, 0$	<u>3, 2</u>	$0, 0$	$-1, -1$
			$2, -1$	<u>0, 0</u>	<u>1, 2</u>
					$2, 1$
					$0, 0$

このような7つのゲームを表6にまとめておく。2×2対称ゲームは12種類あるので(武藤2004)、その半数以上が協力状況ということになる。表6の太字は均衡、下線部は後に定義する「協力状態」である。利得の数字は、順序のみに意味があり、基数的な意味はない。いずれも第1の選択Cが協力であり、第2の選択Dが非協力である。協力状況は、これら有名なゲームに共通の構造なのである。以下の4点について簡単に分析しておく。

- 個人的に最善な状態とナッシュ均衡について。①②③は全員協力(C, C)が個人的に最善の状態である(全員協力はナッシュ均衡でもある)。④⑤⑥⑦では、自分是非協力で相

手から一方的に協力してもらおうのが個人的に最善の状態である(よって全員協力はナッシュ均衡でない)。このうち④⑤⑦は多型状態(C, D), (D, C)が均衡であり, これらは広義のチキンゲームといえる(チキンゲームの厳密な定義は後述)。

- 支配戦略について. ①では協力が支配戦略となっており, ⑥では非協力が支配戦略になっている。ほかのゲームでは支配戦略はない。
- 協力の個人的なリスクについて. ②③⑥⑦では協力が個人的に最悪の結果を起こしうる。ここでは, 非協力がマクスマニ戦略でもある。
- 協力の集団的なリスクについて. 協力はいつも望ましいとは限らず, 集団的なリスクを含む場合がある。これは, 協力の過少・過剰の問題である。
 - ③のある種の「調整ゲーム」では, 一方的な協力が, かえって全員非協力に対してパレート劣位になる**重度の過少協力**状態をひきおこす。
 - ⑦「英雄ゲーム」では, 全員協力が, かえって全員非協力に対してパレート劣位になる**重度の過剰協力**状態をひきおこす³。

当然, 重度の過少協力・過剰協力状態はパレート非効率である。一方, 軽度の協力の過少, 過剰をつぎのように考える。

- ①②では一方的な協力が, 全員非協力に対してパレート劣位になるわけではないが, パレート非効率的になる**軽度の過少協力**状態をひきおこす。
- ⑤では全員協力が, 全員非協力に対してパレート劣位になるわけではないが, パレート非効率的になる**軽度の過剰協力**をひきおこす。
- ④⑥では過少協力, 過剰協力といえる社会状態はない。

2.6 協力状態の定義

全員非協力が望ましくないのは, パレート非効率だからである。その反対概念である協力状態は, 当然, パレート効率的であるべきだろう。これは協力状況にみられる共通の利益の一側面である。

しかし, 社会状態が単にパレート効率的であるだけでは, 協力状態とはふつうよばれない。たとえば, 表 6⑥の囚人のジレンマでは, (C, C)だけが協力状態とよばれ, (D, C)と(C, D)はパレート効率的であるにもかかわらず, 協力状態とはよばれない。つまり, 協力状態は, 全員非

³ たとえば, だれもが協力的だと息苦しいということはよくある。適度にだれかがサボっているときのほうがうまくいくことは多い。だれもがあまりに協力的ゆえすすんで仕事をやろうとして, 仕事の奪い合いや路線の対立などが起こり, いわゆる「内ゲバ」が起こって集団そのものが崩壊してしまうというのが英雄ゲームである。

協力に対してパレート改善されていることが必要だろう。これは共通の利益のもうひとつの側面である。以上から、協力状態をつぎのように定義する。

定義 3 協力状態とは、全員非協力に対してパレート優位、かつパレート効率的な社会状態のことをいう。

協力状態以外の社会状態を「非協力状態」とよぶ。したがって全員非協力以外の非協力状態も一般に存在する。表 6 において協力状態を下線で示した（それ以外が非協力状態）。

2人協力状況の①～④,⑥は、全員協力(C, C)がパレート効率的かつ(D, D)にパレート優位であり、協力状態といえる。一方、⑤,⑦では、(C, C)がパレート非効率である。特に、⑦の英雄ゲームは重度の過剰協力が起こりうる（全員が英雄であろうとして争いが起こる）。結局、⑤,⑦の場合には、(D, C)と(C, D)のみを協力状態とよぶべきだろう。

ただし、④のチキンゲームでは、やや直感に反し、(D, C)と(C, D)はパレート効率的かつパレート改善されたものであり、これらも協力状態とよぶことになる。しかし、⑤や⑦において(D, C)と(C, D)が協力状態とよばれるのであれば、チキンゲームにおいてもこれらを協力状態とよんでもおかしくはない。

N人ゲームにおける協力状態の基本的な性質を述べよう。

定理（協力状態の存在） 協力状況では、協力状態がつねに存在する。

証明) II) より、協力状況では、全員非協力に対してパレート優位な社会状態 α が存在する。 α がもしもパレート非効率であれば、その α に対してパレート優位な社会状態 β が存在する。行為者は N 人であり、社会状態は 2^N 個で有限であるから、この操作は有限回で終わり、パレート効率的な社会状態 ω が存在して、全員非協力に対してパレート優位となる。この ω が協力状態である（証明終）。

協力状態は、かならずしも均衡ではない（たとえば表 6 の⑥「囚人のジレンマ」）。

協力状態は、複数存在しうる（たとえば表 6 の④「チキンゲーム」）。

2.7 他の協力行為定義の可能性

他の協力行為（協力状況）の定義の可能性を考えておこう。まず、I) に対して、協力行為にコストがかからない場合とコストがむしろマイナスである場合が考えられる。この場合、非協力者の利得より協力者の利得のほうが大きくなりうる。これは、他者を出し抜けば全体的にもよくなりうる、という状況である。ここでは、個人的な行為と社会との間に緊張関係はない。つまり、ここではフリーライドが存在しない可能性がある。この状況を数理モデル化することもできるが、本稿では、協力という概念をフリーライドと関連させて定義したいゆえ I) という条件を課した。

なお、もしも I) を仮定しないならば、2つの選択を協力と非協別に区別するメルクマール

には、どのようなものがあるだろうか。たとえば、

I') 全員が選択 1 をとった状態と、全員が選択 2 をとった状態という両極を考えたとき、パレート優位になるほうの選択を協力、もう一方を非協力とよぶ

というものがあえよう。このとき II) の条件は自然に満たされる。しかし I') だけでは、両極のうちのパレート優位な「全員協力」状態でもパレート非効率である可能性がある（つまり軽度の過剰協力）。

そこで、いま、選択 A を全員がとっている状態が、選択 B を全員がとっている状態に対してパレート優位であるとし（すなわち、選択 A=協力、選択 B=非協力）、かつ、どちらの状態もパレート非効率であるとしよう。この状況では、選択 A と選択 B が混在する「協力状態」が存在して両極に対してパレート優位であり、かつパレート効率的である。このとき、この協力状態において 2 つの選択肢の一方を協力とよび、もう一方を非協力とよぶべき積極的な理由はない。なぜなら、どちらの選択肢もこの協力状態には欠かせず、どちらかを非協力とはいえないからである（このことはじつは I) II) によって定義される協力状況においてもいえるのだが、I) II) では非協力のほうが協力よりつねに利得が高いという非対称性があるから問題はない）。

よって、I') の仮定のもとでは、II') 全員協力がパレート効率的である、という第 2 の仮定が必要になる。しかし、この場合には Elster (1989=1997: 156-8) が問題にした「過剰協力」という興味深い問題が扱えなくなる。具体的には、表 6 の⑤指導者ゲームと⑦英雄ゲームを協力状況から排除することになる。しかし本稿では、これら 2 つのゲームも協力状況（社会的ジレンマ）として扱いたい。後述するように、これが制度の存立と関係するからである。

なお、協力ゲーム理論における協力概念は、「提携」coalition を意味する。協力ゲームでは、協力的行為とそれに伴うコストは考えていない。ただし提携概念は、本稿における協力状態（パレート効率性とパレート改善）を前提にしている。

3. 社会的ジレンマと協力状況

この節では、先行研究において考えられてきた多くの社会的ジレンマの定式化が、協力状況であることを示す。なお、この節における選択 C, D は前節において定義した協力、非協力とはさしあたり別物とする。

3.1 既存の社会的ジレンマ定義の問題点

社会的ジレンマの本質は、「個人的合理性と社会的合理性の乖離」にあるとされる（木村 2002: 20）。社会的合理性は、平等性などもありえるが、ふつうはパレート効率性とされる。一方、個人的合理性には幅がある。たとえば、Dawes (1980) は個人的合理性を支配戦略としてとらえ、社会的ジレンマをわかりやすい形で定式化している⁴。

⁴ Dawes (1975) では、(1)(2)以外に「 $C(m)$, $D(m)$ は単調増加」、 「協力者が増加すると、全行為者の総

定義 4 Dawes による社会的ジレンマとは、次の条件(1)(2)をみたす行為選択肢 C, D をもつ N 人 2 選択肢対称ゲーム (= 2^N 対称ゲーム) である.

- (1) $D(m-1) > C(m) \quad \forall m \in \{1, \dots, N\}$
 (2) $D(0) < C(N)$

ここで、 m は自分自身を含む C 選択者の人数であり、 $D(m)$ は D 選択者の利得、 $C(m)$ は C 選択者の利得である。よって $C(m)$ の定義域は $\{1, \dots, N\}$ であり、 $D(m)$ の定義域は $\{0, \dots, N-1\}$ である。なお、 $D(m), C(m)$ は減少しうるから、この状況は意外に一般的ではある。

(1)(2) を満たす状況は、「 N 人囚人のジレンマ」とよばれる ($N=2$ の場合に通常の囚人のジレンマに一致する)。(1) は D が **強支配戦略** であることを意味し、(2) は全員 D ($m=0$) が全員 C ($m=N$) に対してパレート劣位であることを意味する。つまり(1)(2)は、だれもが個人的に合理的な強支配戦略 (D) を選択すれば、社会的に非合理的なパレート非効率状態を帰結してしまう、ということを主張する。

このような状況の構造から、ふつう C は「協力」、 D は「非協力」とよばれるが、これらははたして前節の協力行為の条件 I) II) を満たすのだろうか。このことの吟味も含めて、既存の社会的ジレンマの定義に対して 4 つの問題点 (①パレート劣位条件(2)の限定性、②支配戦略条件(1)の限定性、③パレート非効率均衡の限定性、④逆転現象) をあげる。

問題点 1 パレート劣位条件(2)の限定性 (2)の本質は、個人的に合理的な選択 D をだれもがとるとパレート非効率になる点にあるので、特に全員 C が全員 D に対してパレート優位である必要はない。したがってこの条件は、

$$(2)' \quad D(0) < C(m_1), D(m_1) \quad \exists m_1 \in \{1, \dots, N\} \quad \text{ただし } D(N) \text{ は定義できない}$$

におきかえられるべきである⁵。じっさい、Elster (1989=1997: 152) は社会的ジレンマを条件(1)と(2)'で定式化している。条件(2)'は、 $D(0) > C(N)$ がありうること、すなわち過剰協力の可能性を暗示している。

問題点 2 支配戦略条件(1)の限定性 個人的な合理性は、(1)が意味する強支配戦略に還元さ

利得をも増加する」という 2 つの条件をつけて社会的ジレンマをさらに狭く定義している。海野 (1988) に詳しい。

⁵ ただし、 $C(m)$ が非減少であること (= 選択 C が負の外部性を他の C 選択者に与えない) を仮定すれば、(2) ⇔ (2)' がいえる。証明) (→) は明らかなので、(←) を証明しておこう (これもほぼ自明だが)。 $C(m) \leq C(m+1)$ とする。したがって任意の m に対して、 $C(m) \leq C(N)$ が成立する。よって、(2)' からある m_1 に対して $D(0) < C(m_1)$ が成立すれば、(2) $D(0) < C(N)$ が成立する (証明終)。

れるものではない。それはナッシュ均衡として表現されてもよい⁶。したがって支配戦略が存在しない状況でも、「すべてのナッシュ均衡がパレート非効率」[Raubおよび永田による社会的ジレンマ定義]であれば社会的ジレンマとして考えられる (Raub 1988, 永田 1988)。(1)は、社会的ジレンマ一般の定義には狭すぎるのである。

問題点 3 パレート非効率均衡の限定性 そうかといって(1)を「すべてのナッシュ均衡がパレート非効率」におきかえるだけでは社会的ジレンマを包括できない。ナッシュ均衡がパレート効率的でありうる *N* 人チキンゲームも社会的ジレンマに含めるべきという議論があるからだ (Taylor 1987=1995: 2 章, 木村 2002: 22)。このとき条件(1)をどのような条件におきかえるべきかは次節で議論する。

問題点 4 逆転現象 以上の 3 つの問題点が基本的には社会的ジレンマの Dawes 定義が狭すぎることを問題にする一方、第 4 の問題点は、それが広すぎることを問題にする。この問題点は本稿独自のものである。Dawes 定義の問題条件(1)(2)だけだと、 $D(m)$ と $C(m)$ がともに減少することがある場合に $D(m) < C(m)$ となることがある⁷。したがって I) により、この状況は協力状況ではない。すなわち、条件(1)(2)で定まる *C*を「協力」、*D*を「非協力」とよぶことは一般にはできない。たとえば、 $D(0)=0, C(1)=-1, D(1)=4, C(2)=3, D(2)=2, C(3)=1$ となる 3 人ゲームを考えよう (表 7)。この表では条件(1)として、任意の $D(m-1)$ より、その右下のセルの $C(m)$ は小さくなる。また条件(2)として、最も左上の $D(0)$ より最も右下の $C(3)$ のほうが大きい。よって、このゲームは(1)(2)をみたす。その一方、 $m=2$ では $D(m) < C(m)$ である。したがって Dawes の社会的ジレンマは協力状況に包含されない。

表 7 Dawes 定義での逆転現象

<i>C</i> の人数 <i>m</i>	0	1	2	3
<i>D</i> (<i>m</i>)	0	4	<u>2</u>	-
<i>C</i> (<i>m</i>)	-	-1	<u>3</u>	1

注. 太線部は均衡を示す。他表も同様。

その一方で、協力状況では支配戦略の組がパレート効率的となる場合があり (たとえば円満ゲーム (表 6①, 表 8))、この場合は古典的なゲーム理論の範囲では明らかに社会的ジレンマではない。よって、協力状況と Dawes の社会的ジレンマは、論理的には独立なのである。しかし多くの社会的ジレンマ研究では、

$$(3) \quad D(m) > C(m) \quad \forall m \in \{1, \dots, N-1\}$$

を事実上仮定している。この式の必要性を次節で考える。

⁶ 支配戦略が他者の戦略を顧みない「個別的」合理性を意味するのに対して、ナッシュ均衡戦略は他者の戦略と相関的にとられるから個別的ではない (丸田 1991)。しかし、それは自身の利得を最大化するという点で「個人的」合理性にかなう。

⁷ この点、Dawes (1975) では、このようなことは起こらない。 $C(m)$ が非減少関数であれば、 $m=1, 2, \dots, N$ に対して $C(m) \leq C(m+1) < D(m)$ だからつねに $D(m) > C(m)$ が成り立つ。

3.2 社会的ジレンマは協力状況に含まれる

問題点2「支配戦略条件(1)の限定性」に対するRaubおよび永田の代替案は、行為選択肢が3個以上のゲームをも対象にしていた。しかし本稿ではこの方向は考えず、簡潔のため最も基本的な2選択肢の場合のみを考える⁸。

問題点3「パレート非効率的均衡の限定性」は、 N 人チキンゲームをも社会的ジレンマに含めようという議論から出てきた。 N 人チキンゲームの本質は、Taylor (1987=1995: 56-60) によれば**先手コミットメント**のインセンティブの存在にある。これは**全員 C がナッシュ均衡ではなく、かつ $D(m^*) > C(m^*) (> D(0))$ をみたく多型ナッシュ均衡が存在すること**を意味する⁹[Taylorによる N 人チキンゲームの定義]。この状況は、多型均衡の実現を意図して「先に」 D を選択すれば他者に C を強制できるけれども(=先手コミットメント)、だれもがそのように考えて行為選択すれば全員 D (パレート非効率)が帰結してしまう、というものである。先手コミットメントを個人的合理性として考えることで、この状況もやはり社会的ジレンマ(個人的合理性と社会的合理性の乖離)として考えられるだろう。このときいうまでもなく、先手コミットメントは、フリーライダー問題のキー概念である。

問題点4「逆転現象」の観点から、さらに N 人チキンゲームを考えてみよう。Taylorは、 D 選択者が有利な多型均衡と全員 D (均衡でなくてもよい)のみを実現可能な社会状態として考えている。しかし、現実にはもちろん理論上でさえかならずしもそれらの社会状態が実現するとは予測できない。なぜなら、支配戦略の場合とは異なり、ゲーム理論の個人合理的な推論は、「ひとたびナッシュ均衡が実現すればだれもそこから動かない」と予測するだけで、「**ナッシュ均衡戦略を行為者が選択する**」とまでは予測しないからだ。よって、 $D(m) < C(m)$ なる均衡外の社会状態がありえれば、 D が個人合理的な選択であるとは断定しにくくなる。

すくなくとも進化的過程を考えれば、 $D(m) < C(m)$ なるある状態が一定期間実現すれば D は C に淘汰される(利得単調性)。あるいはこのことを合理的に考えられるとすれば、行為者は C をとりうるので D の個人的な合理性は弱まる。このような**自分の利得^{マイナス} - 他者の利得**という相対的利得を最大化する原理を**進化的合理性**とよぶ¹⁰。進化的合理性は個人的合理性の一種としてカウントできよう。したがって、 N 人チキンゲームにおいて D が確実に個人合理的な選択であるためには、条件(3)が必要になる。

⁸ もっと積極的に2選択肢社会的ジレンマを位置づけることもできる。2選択肢はなにかをする/しない、という根本的な二項対立にもとづくことで「ゲームの外部がない」ということと「相互行為のルールから逸脱する/しない」という特別な意味をもつからだ(武藤 2005a)。

⁹ 多型とは異なる行為の選択者がいる状態のこと。

¹⁰ 相対的利得の最大化は、社会的動機としては競争主義とよばれる(土場 1995)。競争主義あるいは進化的合理性は、いわゆる嫌がらせ(スパイト)行動に象徴的である。戦略や社会的動機や遺伝子を抽象した進化の単位をタイプとよぼう。進化するタイプ(=シェアを拡大するタイプ)は、他者に対して相対的に利得が高いタイプである(利得単調性)。したがって、スパイト行動や競争主義はあらゆる N 人ゲームの協力状況[I], II]で進化する。一方、状況が2人ゲームのランダムマッチングの場合、これらの進化はゲームによって条件づけられている。たとえばチキンゲームでは競争主義(非協力戦略)は一定の範囲までしかシェアを拡大できない。なお、利得単調性についてはWeibull (1995=1998: 182) 参照。

まったく同様に、個人的合理性としての進化的合理性を考えると、2 選択肢社会的ジレンマ一般においても(3)は必要になる。じっさい、山岸 (1989) や木村 (2002) に登場するほとんどの社会的ジレンマの具体例では、この条件が仮定されている。結局、先手コミットメントは、進化的過程の考慮によって拡張される。このときもちろん、(3)はフリーライドが実行されるための十分条件である。

翻って問題点 1「パレート劣位条件(1)の限定性」にかんして、前述したように Elser (1989=1997) は社会的ジレンマを(1)と(2)'で定式化しているが、条件(3)を明示してはいない。そこで(3)を仮定すれば、条件(2)'は、さらに簡潔に、

$$[2] \quad D(0) < C(m_1) \quad \exists m_1 \in \{1, \dots, N\}$$

とかくことができる。

結局、ナッシュ均衡戦略、先手コミットメント、進化的合理性という多元的な個人的合理性を考えていくと、条件(1)「支配戦略条件」は社会的ジレンマには不要であり、かわりに条件(3)が必要であることがわかった。そこで(1)を条件(3)におきかえ、これを条件[1]と改めよう。これと上の[2]を併せると、社会的ジレンマはつぎの 2 条件をみたす必要がある。

社会的ジレンマの必要条件

$$[1] \quad D(m) > C(m) \quad \forall m \in \{1, \dots, N-1\} \quad \text{協利行為のコスト}$$

$$[2] \quad D(0) < C(m_1) \quad \exists m_1 \in \{1, \dots, N\} \quad \text{共通の利益の存在}$$

明らかに、[1]=I), [2]=II) がいえる。つまり、個人的合理性が多様な意味を容れることで、社会的ジレンマは協力状況(公共財供給状況)に包含される(社会的ジレンマ \subseteq 協力状況)。よって、[1][2]の成立をもって C を協利行為、 D を非協利行為とよぶ。第 2 節の記号を用いれば、共通の利益 $v(m)$ とコスト $c(m) > 0$ を用いて、協力者の利得 $C(m)$ と非協力者の利得 $D(m)$ をつぎのようにかける。

$$C(m) = v(m) - c(m), \quad D(m) = v(m)$$

これまで公共財供給状況は社会的ジレンマの一種として考えられてきたから(木村 2002: 20-30)、これはすこし奇妙にもみえる。しかし、もともと公共財供給状況は表 6①円満ゲームの例にみたように、個人的合理性と社会的合理性の「乖離」のない状況をも含むのである¹¹。

結局、Dawes 定義においては社会的ジレンマと協力状況は独立だったが、個人の多元的合理性に照らして社会的ジレンマを拡張かつ精緻化していくと、社会的ジレンマは、協力状況に包

¹¹ 「乖離のない状況」は公共財の価値 $v(m)$ の傾き $v(m+1) - v(m)$ が大きい場合、すなわち

$D(m) < C(m+1) \Leftrightarrow v(m) < v(m+1) - c(m+1) \Leftrightarrow v(m+1) - v(m) > c(m+1)$ が任意の m で成り立つ場合である。つまり 1 人の協利のコストよりもそれによる 1 人あたりの利益のほうが大きい場合には、「乖離」はないのである。しかしこのような状況は人数が多いほど稀だろう(大集団のジレンマ)。

含まれる。

ただし、[1][2]はこのままでは社会的ジレンマ（二水準の合理性の「乖離」）に対する十分条件ではない。しかし、次節にみるように、進化的視点ではこれは異なってくる。

4. 協力状況のサブカテゴリー

この節では、協力状況の下位のゲームのクラスである N 人囚人のジレンマ、 N 人チキンゲーム、 N 人保障ゲーム、 N 人円満ゲームを「均衡の配置」という視点から統一的に定義する。そしてこの4つのゲームがそれぞれ4つの個人的合理性すなわち、支配戦略、先手コミットメント、マクスミニ戦略、進化的合理性に対応した社会的ジレンマであることを明らかにする。

4.1 種々の N 人ゲームの定義

ゲーム理論的な分析軸は、当然ながら均衡による。ここで、条件[1][2]を満たす状況には均衡が存在することが証明できる（武藤 2005b）。したがって、均衡が存在すればそこで協力者数 m^* が $0 \sim N$ をとるので、図 1 が得られる。協力状況の条件[1][2]の下で、純粋型として、つぎの N 人ゲームのクラスを均衡の配置によって操作的に定義する。

定義 5

N 人囚人のジレンマ	⇔ 均衡が全員非協力 ($m^*=0$) のみからなる
N 人円満ゲーム	⇔ 均衡が全員協力 ($m^*=N$) のみからなる
N 人チキンゲーム	⇔ 均衡が多型状態 ($m^* \in [1, N-1]$) のみからなる

言い換えれば、チキンゲームとは、全員協力および全員非協力がいずれも均衡になっていないゲームとして定義される。すなわちチキンゲームとは、すべての均衡が協力者と非協力者が混在する多型状態からなるゲームのことである。

定義5の3つのゲームは典型的なもので、これらだけで協力状況を埋め尽くすわけではなく、たとえば N 人囚人のジレンマと N 人円満ゲームの「あいこの」のような、これらの中間的な状況が存在する。さらに各ゲームのクラスについていくつか注意点がある。

- N 人囚人のジレンマでは**非協力が強支配戦略**であることが証明できる (ibid)。したがって、この N 人囚人のジレンマ定義は、Elster (1989=1997: 152) の社会的ジレンマ定義=条件(1)(2)'に一致する。
- N 人円満ゲームでは**協力が強支配戦略**であることが証明できる。また、その組である**全員協力 $m=N$ が唯一のパレート効率状態**であることが証明できる（武藤 2005b）。つまり N 人円満ゲームでは、全員協力が唯一の協力状態でありかつ唯一の均衡である。なお、 N 人円満ゲームは、非協力者のほうが協力者より利得が高いが（条件[1]）、だれかが協力したほうがよいとだれもが思っており（条件[2]）、しかもだれもが協力をしたいと思っている（協力が支配戦略）、という状況である。

- N 人チキンゲームの定義では、Taylor (1987) のものと異なり、 $m=0$ が均衡ではないことが要請される。しかしこの定義は Taylor の先手コミットメントの考え方を継承している。これは純粋型という都合である。また、協力者数 m を異にする複数の均衡が存在する場合もある。いずれにしても、**全員非協力に対してパレート優位な多型均衡が存在**することが証明できる (武藤 2005b)。

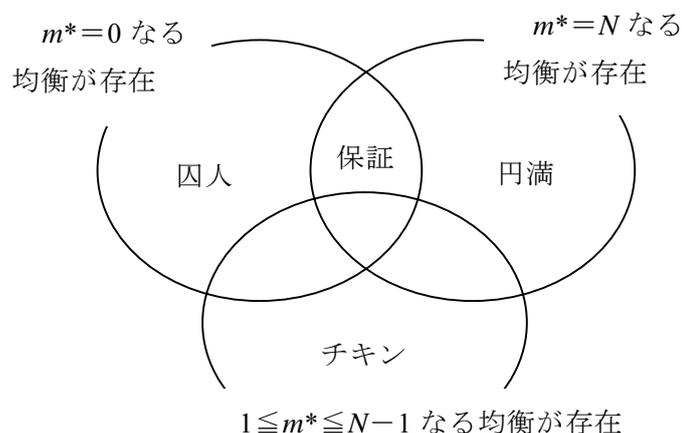


図 1 協力状況のサブカテゴリ

N 人保証ゲームを、 N 人囚人のジレンマと N 人円満ゲームの「あいのこ」として定義する。

定義 6 N 人保証ゲーム \Leftrightarrow 全員非協力 ($m^*=0$) と全員協力 ($m^*=N$) が均衡で、かつ均衡はこれらのみである。

- N 人保証ゲームでは、非協力が**マクスミニ戦略**であることが証明できる (ibid)。そして、マクスミニ戦略を個人的合理性として考えることもできるので、 N 人保証ゲームも社会的ジレンマとして考えられる (次項)。なお、 N 人保証ゲームでは、よく知られた**分水嶺** (critical mass) が存在することが証明できる (ibid)。

N 人囚人のジレンマ、 N 人円満ゲーム、 N 人保証ゲームがかなり特殊な協力状況である一方、 N 人チキンゲームは、かなり一般的な協力状況である。このような違いは、先行研究ではほとんど意識されていなかった。 N 人チキンゲームの内訳は多様で複雑である。

4.2 N 人保証ゲームとマクスミニ戦略

Taylor (1987) は N 人保証ゲームそのものを社会的ジレンマとは考えていない。 N 人保証ゲームでは全員 D だけでなく、全員 C も均衡になるからだ。一方、森村 (1987) は、社会的ジレンマというわけではないが、ホップズの自然状態が囚人のジレンマだけでなく、保証ゲームとしても考えられると論じている (ただし 2 人ゲームでの議論だが)。なぜなら表 3 にみるように、2 人保証ゲームでは、相手が協力しないときに自分が協力すれば、その結果は最悪とな

るからだ（したがってこのとき非協力 D はマクスミニ戦略である）。

このようにチキンゲームとは異なり、保証ゲームでは協力 C がリスクを伴う。この点を敷衍して、 N 人保証ゲームを社会的ジレンマとみなすことができる。このとき、個人的合理性にあたるものが「マクスミニ戦略」である。**マクスミニ戦略**とは、それをとったときに起こりうる最悪の結果が最もましになるような戦略のことである。

マクスミニ戦略は、ゼロサムゲームのときのみ有効で、非ゼロサムゲームでは意味がないと一般には考えられている。しかし、**現実の相互行為状況では、かならずしも他者が合理的であるとは限らない**。そのような相手がナッシュ均衡戦略をとる保証はどこにもない。このような場合には、マクスミニ戦略を用いて最悪の結果を回避するという手は、ひとつの個人的合理性の表現である。

もっとも、相手に合理性を仮定できないときには、マクスミニ戦略だけではなく、似た構造をもつ「ラプラス戦略」と「マクシマクス戦略」にも個人的合理性があるといえよう。**ラプラス戦略**とは、それをとったときに起こりうる結果の平均が最もよくなるような戦略のことである。**マクシマクス戦略**とは、それをとったときに起こりうる最高の結果が最もよくなるような戦略のことである。

しかし、この3つのなかではマクスミニ戦略だけが「保証水準」を明確にすることができる。**保証水準**とは、当該の状況で相手がどんなことをしてきても、マクスミニ戦略をとることによって最低限確保できる利得である。このように、マクスミニ戦略は、 N 人ゲームという幾重にも不確定な状況に、保証水準という確定的な利得を保証する。この「確定性」、簡単にいえば「確実さ」という点で、マクスミニ戦略は、ラプラス戦略やマクシマクス戦略に比べて個人合理的であると考えることができる。

つまり、非ゼロサムゲームでは無視されてきたマクスミニ戦略も、他者がかならずしも合理的に振舞うとは限らない現実の相互行為では、確実さという点で考慮に値する個人的合理性なのである。よって、**マクスミニ戦略の組がパレート非効率になる**状況を、社会的ジレンマの一種として考えることができる。

結局、保障ゲームでは、協力行為は個人的に最悪の結果を招きうる。そのため、だれもがそれを回避しようとするれば全員非協力＝パレート非効率が帰結するのである¹²（表8参照）。

¹² なお、保証ゲームの議論でよく用いられるリスク支配戦略は、その前提にナッシュ均衡戦略をとることを想定しているため、すでに他者の合理性を前提にしている（Harsanyi and Selten 1988, 岡田 1995: 377）。そのうえで立場を交換して期待利得計算を比較することが必要であり、それじたいとしては、社会学的には合理性を要求しすぎる（本稿では、経済学におけるゲーム理論とは異なり、確率や利得和を行為者が計算するという仮定は入れない）。ただし、Kandori et al. (1993) のように、リスク支配戦略を進化論的に限定合理性の下で解釈する議論もあるので、リスク支配戦略にかんする議論すべてを棄却することはできない。しかし、本稿のような古典的なゲーム理論の枠組みでは、このリスク支配戦略は合理的すぎるゆえ、考えない。

表 8 N 人保証ゲーム

C の人数 m	0	1	2	3
$D(m)$	0	0	0	-
$C(m)$	-	-10	-10	1

4.3 N 人円満ゲームと進化的合理性

繰り返しになるが、 N 人円満ゲーム (表 9) は、社会的ジレンマとは通常考えられていない。ところが、進化的合理性を考慮すると、 N 人円満ゲームも一種の社会的ジレンマとして考えられる。なぜなら、進化的合理性の下では、協力が強支配戦略であるにもかかわらず、全員協力は均衡ではあっても安定ではないからだ (進化ゲームの解概念は均衡ではなく安定性である)。すなわち、だれかが全員協力から逸脱すれば、すべての者の利得は下がるけれども、あらゆる社会状態において非協力をとる逸脱者は協力者より利得が高いために、利得単調性 (模倣や学習や教育) にもとづいて増加していく。このとき、協力者は淘汰され、全員非協力が実現する。これは N 人円満ゲームと 2 人円満ゲームのランダムマッチングとの大きな違いである¹³。

このことは、 N 人円満ゲームだけでなく、すべての協力状況においていえる。したがって、進化的合理性を考慮するならば、条件[1][2]は、社会的ジレンマであるための十分条件でもあることになる (4 つの個人的合理性: 進化的合理性, マクスマニ戦略, 強支配戦略, 先手コミットメントは互いに打ち消しあうものではない)。つまり、進化的にみれば, [1][2] は社会的ジレンマの必要十分条件なのである。

表 9 N 人円満ゲーム

C の人数 m	0	1	2	3
$D(m)$	0	2	4	-
$C(m)$	-	1	3	5

4.4 過剰協力・集団淘汰・制度

2.5 節で述べたように、協力はある状況において全員非協力 $m^*=0$ よりも悪い結果をもたらすことがある。たとえば、野外パーティのあと、全員で芝生のうへのゴミを拾ったならば、芝生がダメになってしまい、だれにとっても最悪の結果となる (Elster 1989=1997: 157)。これは重度の過剰協力である。重度の過剰協力は、 N 人チキンゲームの一種である N 人英雄ゲーム (表 10)、 N 人囚人のジレンマ (表 11)、 N 人保証ゲーム (表

表 10 過剰協力 (N 人英雄ゲーム)

C の人数 m	0	1	2	3
$D(m)$	0	2	1	-
$C(m)$	-	1	0	-1

表 11 過剰協力 (N 人囚人のジレンマ)

C の人数 m	0	1	2	3
$D(m)$	0	2	4	-
$C(m)$	-	-1	1	-1

¹³ 強支配戦略の組は、ランダムマッチング下のレプリケータダイナミクスにおいて漸近安定である。なぜなら、強支配戦略以外の強支配される戦略はすべて淘汰されるからである。Weibull (1995=1998: 103) 参照のこと。

12) など円満ゲーム以外の協力状況で起こりうる。

表 12 過剰協力 (N 人保証ゲーム)

C の人数 m	0	1	2	3	4
$D(m)$	0	2	2	-2	-
$C(m)$	-	-1	1	-3	-1

進化的合理性を考えると、このような重度の過剰協力のある社会的ジレンマでは、集団淘汰 (group selection) による協力状態の実現も不可能になる¹⁴。なぜなら、集団淘汰では、当の集団が協力者のみからなる場合にだけ進化する (シェアを拡大する) が、重度の過剰協力がある場合では、協力者のみの集団は非協力者のみの集団に淘汰されるからである。

したがって、重度の過剰協力が存在する場合には、協力者と非協力者が順番に交代するような制度や協力者に報奨を与えるような制度を導入しなければ、社会的ジレンマは解決できない。しかし、これは上位下位あるいは指導者を要する組織や民主主義のような制度を考えていくうえできわめて示唆に富むのではないだろうか。これは非対称性としての権力の問題とももちろん関係してくる。また、集団は国家や企業やNPOなど、社会学の中心的概念でもある¹⁵。従来の研究では、進化は慣習と関係づけられてきた一方、集団や制度にはあまり接点がなかった。この重度の「過剰協力」を介して、進化・集団・制度の関係を議論していくことができるだろう。

5. おわりに

本稿の内容をまとめておこう。本稿では、まず、フリーライダー問題に準拠して、「協力行為」が定義される状況としての「協力状況」を、I) 協力のコストとII) 共通の利益の存在 (= [1][2]) によって定義した。つぎに、これを起点として、協力状態を定義し、さらに協力状況のサブカテゴリーとして、 N 人囚人のジレンマ、 N 人円満ゲーム、 N 人保証ゲーム、 N 人チキンゲームを均衡の配置によって操作的に定義した。これらはそれぞれ異質な個人的合理性によって特徴づけられることにより、社会的ジレンマとして考えることができた (表 13 にまとめた)。

¹⁴ 社会学における集団淘汰のモデルには、大浦 (2002)、金井 (2003) などがある。

¹⁵ 佐藤 (1995) は、ウェーバーのプロテスタンティズムの議論を発展させ、ゼクテという集団が会社の母体となることが、近代資本主義社会の成立にとって重要であったことを明らかにしている。「会社によって社会をつくる」のである。ゼクテ (会社) では、集団間を比較的自由に移動できる。本報告書に本稿とは別の論文として収められている諸々の実験的研究は、このような文脈から解釈できるだろう。

表 13 協力状況のサブカテゴリーと個人的合理性

ジレンマ分類	操作的定義	個人的合理性
N 人囚人のジレンマ	均衡が $m^*=0$ のみ	強支配戦略
N 人円満ゲーム	均衡が $m^*=N$ のみ	進化的合理性
N 人チキンゲーム	均衡が $\neq 0, N$	先手コミットメント
N 人保証ゲーム	均衡が $m^*=0$ と N のみ	マクスミニ戦略

注. うえの定義は協力状況を前提する. 社会的合理性はパレート効率性である.

この枠組みから結果的に、社会的ジレンマと協力状況についてつぎの3点がわかった.

1. Dawes (1980) の「社会的ジレンマ」と「協力状況」は独立である.
2. Taylor (1987=1995) の先手コミットメントやナッシュ均衡戦略などの多元的な個人的合理性を考慮することで社会的ジレンマ概念を拡張・精緻化すると、社会的ジレンマは協力状況に含まれる.
3. さらに、進化的な合理性を個人の多元的な合理性にカウントすると、社会的ジレンマは協力状況に一致する (N 人円満ゲームでさえ社会的ジレンマといえる).

また、過剰協力と集団淘汰の関係についてつぎの点がわかった.

4. 協力的行為は、個人にとってリスクがあるだけでなく、全体にとってもリスクがある重度の「過剰協力」をひきおこすこともある. 重度の過剰協力が起こる状況では、集団選択 (group selection) によっても協力状態は実現できない.

本稿は、これまで曖昧になっていた社会的ジレンマ研究における協力的行為概念、協力状態概念、諸 N 人ゲームをシステムティックに定義し、多元的な個人的合理性という視点からこれらの N 人ゲームを社会的ジレンマとして位置づけた. これにより、社会的ジレンマ研究の見通しがよくなれば幸いである.

今後の課題は、本稿の基礎概念と木村 (2002) などの現実の現象に注目した研究とを融合しながら、経験的な研究に生かしていくことである. なお、個々の協力状況のサブカテゴリーについての定理の証明は、武藤 (2005b) を参照してほしい.

付録 協力 (cooperate) の辞書的な意味

いくつかの英語辞典で、協力 (cooperate) の意味を調べた. 共通の目的、共通の利益などを太字にしてある.

① The American Heritage® Dictionary of the English Language

1. To work or act together toward a **common end or purpose**.

2. To acquiesce willingly; be compliant:
3. To form an association for **common, usually economic, benefit**.

② Encarta® World English Dictionary, North American Edition

1. work together: to work or act together to achieve **a common aim**.
2. comply: to do what is asked or required

③ The Wordsmyth English Dictionary-Thesaurus

To work with others, esp. for **a common purpose**.

④ Infoplease Dictionary

1. To work or act together or jointly for **a common purpose or benefit**.
2. To work or act with another or other persons willingly and agreeably.
3. To practice economic cooperation.

⑤ Cambridge Dictionary of American English

To act or work together for **a shared purpose**, or to help willingly when asked

⑥ Rhymezone

Work together on **a common enterprise** or project

⑦ Cambridge International Dictionary of English

To act or work together for **a particular purpose**, or to help someone willingly when help is requested:

⑧ Longman Dictionary of Contemporary English

1. If two people or two groups cooperate, they work together in order to achieve **a result that is good for both of them**.
2. To help someone willingly when they ask for your help.

⑨ Merriam-Webster's Online Dictionary, 10th Edition

1. To act or work with another or others : act together
2. To associate with another or others for **mutual benefit**

また、「共通の利益」を明示しないほかの辞書の定義も参考にあげておく（下線は引用者）。

● UltraLingua English Dictionary

To work together in harmony; to collaborate; “To succeed at this task, all members of the group must cooperate.”

● Online Plain Text English Dictionary

To act or operate jointly with another or others; to concur in action, effort, or effect.

【文献】

- Dawes, R. M., 1980, “Social Dilemmas”, *Annual Review of Psychology*, 31: 169-193.
- 土場学, 1995, 「社会的価値と合理性：社会的ジレンマに関する実験社会心理学的研究とゲーム理論的研究の対話の試み」『理論と方法』10(2): 115-132.
- Elster, J., 1989., *The Nuts and Bolts for the Social Sciences*, Cambridge University Press. (=1997, 海野道郎訳『社会科学の工具箱』ハーベスト社.)
- Harsanyi, J. C., 1977, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.
- Harsanyi, John C. and Reinhard Selten, 1988, *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge, Massachusetts.
- 金井雅之, 2003, 「進化ゲームにおける選択的相互作用モデルの意義と課題」『理論と方法』18(2): 153-167.
- Kandori, Michihiro, George J. Mailath, and Rafael Rob, 1993. “Learning, mutation, and long-run equilibria in games”, *Econometrica*, 61: 29-56.
- 木村邦博, 2002, 『大集団のジレンマ』ミネルヴァ書房.
- 永田えり子, 1988, 「自由と効率」『理論と方法』3(1): 43-56.
- 丸田利昌, 1991, 「個人主義的社会理論と Nash 均衡概念」『理論と方法』6(1): 21-38.
- 三隅一人, 1993, 「社会的ジレンマモデルの一般化」『理論と方法』8(1): 89-92.
- 武藤正義, 2004. 「社会的動機による 2×2 対称ゲームの変換」数土直紀編『社会理論の実践的可能性の探求』平成 13～15 年度科学研究費補助金研究成果報告書, 学習院大学, 103-130.
- 武藤正義, 2005a, 「秩序問題としての社会的ジレンマの一般性：《二重の選択性》の視点から」大浦宏邦編『秩序問題への進化ゲーム理論的アプローチ』平成 14～16 年度科学研究費補助金・研究成果報告書, 帝京大学.
- 武藤正義, 2005b, 「社会的ジレンマの諸定理：フリーライドの視点による N 人ゲームの構造分析 II」大浦宏邦編『秩序問題への進化ゲーム理論的アプローチ』平成 14～16 年度科学研究費補助金・研究成果報告書, 帝京大学.
- 岡田章, 1996, 『ゲーム理論』有斐閣.
- 大浦宏邦, 2002, 「農耕戦略の進化モデル：非ランダムモデルによるジレンマの回避」『帝京社会学』16: 97-120.
- Raub, W. 1988. “Problematic Social Situation and the Large Number of Dilemma: A Game-theoretical Analysis.” *Journal of Mathematical Sociology*. 13(4): 311-357.
- 佐藤俊樹, 1993, 『近代・組織・資本主義』ミネルヴァ書房.
- 盛山和夫・海野道朗編, 1991, 『秩序問題と社会的ジレンマ』ハーベスト社.
- 鈴木鉄忠, 2003, 「集合行為問題のゲーム理論的分析及び社会学的考察」平成 14 年度東京工業大学社会理工学研究科修士論文.
- 瀧川哲夫, 1989, 「N 人ゲーム状況の構造解析」『北海道大学文学部紀要』37: 77-146.
- Taylor, M. 1987. *Possibility of Cooperation*. Cambridge University Press. (=1995 松原望訳『協力の可能性』木鐸社.)

Weibull, J., 1995, *Evolutionary Game Theory*, M.I.T. Press. (大和瀬達治監訳. 1998. 『進化ゲームの理論』文化書房博文社.)

海野道郎, 1988, 「社会的ジレンマの数理モデル: 共有物の性質とモデルの対応関係」『社会学研究』52.

山岸俊男, 1989, 「社会的ジレンマ研究の主要な理論的アプローチ」『心理学評論』32(3):262-294.

**Social Dilemmas and Cooperation Concept:
N-person Game Structure Analysis from the viewpoint of Free Ride I**

Masayoshi MUTO

Graduate School of Decision Science and Technology

Tokyo Institute of Technology

2-12-1 Oookayama, Meguro, Tokyo 152-8552, JAPAN

社会的ジレンマの諸定理

——フリーライドの視点による N 人ゲームの構造分析 II ——

武藤 正義
(東京工業大学)

【要旨】

本稿の目的は、社会的ジレンマの数理構造をフリーライドと均衡の配置という視点から公理系によって整理することにある。具体的には、武藤 (2005) が均衡の配置によって定義した N 人囚人のジレンマ、 N 人円満ゲーム、 N 人チキンゲーム、 N 人保証ゲームのよく知られた性質を、定理として示す。本稿は集合行為問題や組織の分析の一助となるだろう。

キーワード：均衡の配置，強支配戦略，パレート優位，分水嶺

1. はじめに

社会的ジレンマは個人的合理性と社会的合理性の乖離を本質とするが、じっさいの研究では Dawes (1980) の社会的ジレンマ定義がよく使われる。Dawes の社会的ジレンマとは、協力と非協力という 2 つの選択肢があり、①非協力は強支配戦略であるが、②全員非協力は全員協力に対しパレート劣位になる、という N 人対称ゲームのことである。

Dawes (1980) の定義は単純でわかりやすい。しかしその反面、非常に限定的ではある。そのため、Elster (1989=1997) は過剰協力や過少協力の問題を、Taylor (1987=1995) は N 人チキンゲームを、それぞれ独立に社会的ジレンマに加えている。また、ほかの論者もそれぞれ別様に社会的ジレンマを定式化している。その結果、これまでの社会的ジレンマ定義を整理し、統一的な視点からその全体像を再構成することにより、社会的ジレンマ全体の見通しをよくすることが必要になってきた。

そこで武藤 (2005) は、フリーライドという観点から、彼らの主張をともに含むような社会的ジレンマの必要条件としての「協力状況」——「協力のコスト」と「共通の利益の存在」によって特徴づけられる——を定義し、さらにその基本的な下位のゲームのクラスとして、 N 人囚人のジレンマ、 N 人円満ゲーム、 N 人チキンゲーム、 N 人保証ゲームを「均衡の配置」によって定義した。しかし武藤 (2005) は、定義にとどまり、そこでなにかいえるのかという定理をほとんど示してはいない。しかるに本稿の目的は、協力状況とその下位のゲームのクラスにおいて、いくつかの基本的な定理を数学的に厳密に導くことにある。

本稿はつぎの 2 つの分析概念を主に用いる。第 1 に、瀧川 (1989) などが用いている利得和ではなく「パレート性」を用いる。また第 2 に、Dawes (1980) が用いている支配戦略ではなく「ナッシュ均衡」を用いる。この 2 つの視点は、Taylor (1987=1995)、木村 (2002)、鈴木

(2003)でも用いられており、利得和や支配戦略よりも一般性が高い。ただし、彼らの分析は包括的ではないため、この点に本稿の意義がある。

本稿の構成を述べる。第2節では、協力状況の分析の準備として、もっと一般的なゲームを分析する。第3節では、協力状況の基本的なサブカテゴリーである N 人円満ゲーム、 N 人囚人のジレンマ、 N 人チキンゲーム、 N 人保証ゲームについて基本的な定理を示す。第4節はまとめである。

2. 2^N 対称ゲーム

この節では、協力状況分析の準備として、一般的な2選択枝 N 人対称ゲーム (= 2^N 対称ゲームとよぶ) において成り立つ定理を示す。以下では、「混合戦略」を仮定しない。本稿における「均衡」とは純粋戦略ナッシュ均衡のことである。

2.1 記号の定義

定義 1 2^N 対称ゲームとは、 N 人の行為者 (プレイヤー) が、2つの行為選択枝 (純粋戦略) をもつ対称ゲームとする。

行為選択枝集合を $\{C, D\}$ とする。この第2節では C と D は協力、非協力という意味はなく、単なるインデックスである。以下のように、記号を導入する。

- a_i i の行為, $a_i \in \{C, D\}$
- a_{-i} i 以外の行為者の行為, $a_{-i} \in \{C, D\}^{N-1}$
- $a := (a_1, \dots, a_N)$ 全行為者の行為の組 (= 社会状態), $a \in \{C, D\}^N$
- $\pi_i(a)$ 行為者 i の利得関数, ただし $\pi_i(a) = \pi_i(a_i; a_{-i})$ ともかく
- $\#(\cdot)$ C 選択者の人数 (m で表す) を示す関数 (例. $\#(C, D, C, C) = 3$)

本稿において対称ゲームとは、 C 選択者が同数の社会状態 a, b (組成が同じということ) について、同じ行為を選択している者は、同じ利得を得るようなゲームのことである。これを記号でかけばつぎのようになる。

定義 2 (対称性) $a, b \in \{C, D\}^N$ について、 $\#(a) = \#(b)$, $a_i = b_j$ ならば $\pi_i(a) = \pi_j(b)$

たとえば、 $N=3$, $a=(C, D, D)$, $b=(D, D, C)$ では C 選択者 $m = \#(a) = \#(b) = 1$ だから、 C 選択者について $\pi_1(a) = \pi_3(b)$ であり、かつ D 選択者について $\pi_2(a) = \pi_3(a) = \pi_1(b) = \pi_2(b)$ が成り立つ ($\pi_2(a) = \pi_3(a)$ などは、対称性の仮定で $a=b$ としたときに成り立つ)。

ひきつづき基礎概念を導入する。 C 選択者が m 人となる社会状態の集合

- $S(m) := \{a \in \{C, D\}^N \mid \#(a) = m\}$

を**状態集合**とよぶ。 $S(x)$ を「状態集合 $m=x$ 」とよぶことがある。明らかに $S(m)$ は $S(0), S(1), \dots, S(N)$ までをとり、これらは $\{C, D\}^N$ を分割している。ここで状態集合 $S(m)$ は、 ${}_N C_m$ 個の社会状態からなることに注意されたい。たとえば、 $N=3$ において、 $S(2) = \{(C, C, D), (C, D, C), (D, C, C)\}$ である。

対称性の下で、利得に影響を与えるのは、自身のとる選択を除くと、 C 選択者の人数のみである。つまり $a, b \in S(m)$ ($\Leftrightarrow \#(a) = \#(b) = m$) のとき、任意の i, j に対して、

$$\begin{aligned} a_i = b_j = C \quad \text{ならば} \quad \pi_i(C; a_{-i}) &= \pi_j(C; b_{-j}) \\ a_i = b_j = D \quad \text{ならば} \quad \pi_i(D; a_{-i}) &= \pi_j(D; b_{-j}) \end{aligned}$$

が定義から明らかに成り立つ。よって、対称性の下ではつぎのような簡潔な利得関数表記ができる。自分自身を含めて C 選択者が m 人の任意の社会状態 $a \in S(m)$ ($\Leftrightarrow \#(a) = m$) に対して、 C 選択者の利得 $C(m)$ 、 D 選択者の利得 $D(m)$ を、

- $C(m) := \pi_i(C; a_{-i})$ ただし $\#(a_{-i}) = m-1$
- $D(m) := \pi_i(D; a_{-i})$ ただし $\#(a_{-i}) = m$

のように定義できる（表1は状況理解のための2人ゲームでの利得表）。ただし $C(m)$ の定義域は $\{1, \dots, N\}$ であり、 $D(m)$ の定義域は $\{0, \dots, N-1\}$ である。なお、 $C(m), D(m)$ は、状態集合の族 $\{S(0), S(1), \dots, S(N)\}$ を定義域とする実数値関数ともいえる。

表1 2×2 対称ゲーム

	C	D
C	C(2), C(2)	C(1), D(1)
D	D(1), C(1)	D(0), D(0)

2.2 ナッシュ均衡

周知のように、社会状態 $a^* \in \{C, D\}^N$ がナッシュ均衡であるとは、

$$\forall i \quad \forall a_i \quad \pi_i(a^*_i; a^*_{-i}) \geq \pi_i(a_i; a^*_{-i})$$

が成り立つことである。このとき、 2^N 対称ゲームでは、 $\#(a^*) = m^*$ として、社会状態 a^* に対して、上記の $C(m)$ と $D(m)$ の定義から明らかに

$$\begin{aligned} a^*_i = C \text{ なる } i \text{ について} \quad \pi_i(C; a^*_{-i}) &\geq \pi_i(D; a^*_{-i}) \quad \#(a^*_{-i}) = m^* - 1 \quad \Rightarrow \quad C(m^*) \geq D(m^* - 1) \\ a^*_i = D \text{ なる } i \text{ について} \quad \pi_i(D; a^*_{-i}) &\geq \pi_i(C; a^*_{-i}) \quad \#(a^*_{-i}) = m^* \quad \Rightarrow \quad D(m^*) \geq C(m^* + 1) \end{aligned}$$

がいえる。ここから以下の補題を得る。

補題 1 状態集合 $S(m^*)$ が均衡をもつ $\Leftrightarrow D(m^*-1) \leq C(m^*), D(m^*) \geq C(m^*+1)$

証明) 上段の (→) はすでに明らかにしたので, (←) を背理法で示す。

$C(m^*) \geq D(m^*-1), D(m^*) \geq C(m^*+1)$ が成り立つとき, どんな $a \in S(m^*)$ についても,

$$\begin{aligned} \pi_i(C; a_{-i}) &< \pi_i(D; a_{-i}) & \#(a_{-i}) &= m^* - 1 \\ \text{または} \quad \pi_i(D; a_{-i}) &< \pi_i(C; a_{-i}) & \#(a_{-i}) &= m^* \end{aligned}$$

が, ある i について成り立つとする。このとき, $C(m^*) < D(m^*-1)$ または $D(m^*) < C(m^*+1)$ が成立し, 明らかに矛盾。■

補題 2 状態集合 $S(m^*)$ が均衡をもてば, $S(m^*)$ のすべての社会状態は均衡である。

証明) 社会状態 $a^* \in S(m^*)$ が均衡であるとすれば, C, D をとる任意の i に対して

$$\begin{aligned} \pi_i(C; a^*_{-i}) &\geq \pi_i(D; a^*_{-i}) & \#(a^*_{-i}) &= m^* - 1 \\ \text{かつ} \quad \pi_i(D; a^*_{-i}) &\geq \pi_i(C; a^*_{-i}) & \#(a^*_{-i}) &= m^* \end{aligned}$$

が成り立つ。これと対称性により, 任意の社会状態 $b \in S(m^*)$ に対して, $\#(b) = \#(a^*)$ より

$$\begin{aligned} b_j = C \text{ なる任意の } j \text{ について} \quad \pi_j(C; b_{-j}) &= \pi_i(C; a^*_{-i}) \geq \pi_i(D; a^*_{-i}) = \pi_j(D; b_{-j}), \\ b_j = D \text{ なる任意の } j \text{ について} \quad \pi_j(D; b_{-j}) &= \pi_i(D; a^*_{-i}) \geq \pi_i(C; a^*_{-i}) = \pi_j(C; b_{-j}). \end{aligned}$$

したがって, b は均衡。■

定理 1 C 選択者 m^* 人のすべての社会状態が均衡 $\Leftrightarrow D(m^*-1) \leq C(m^*), D(m^*) \geq C(m^*+1)$

補題 1, 2 から, 明らかにうえの定理が成り立つ。

この定理から, 均衡をもつ状態集合 $S(m^*)$ を**均衡集合**とよんでさしつかえないことがわかる。また, 均衡集合 $S(x)$ を「均衡集合 $m^*=x$ 」とよぶことがある。なお, $S(0)$ と $S(N)$ は, 均衡集合が唯一の社会状態からなるので, 均衡集合と均衡を同一視して, たんに「均衡 $m=0$ 」, 「均衡 $N=0$ 」とよぶ。

さて, 本稿の分析は純粋戦略ナッシュ均衡に限定するが, ジャンケン (表 2) をみればわかるように, 純粋戦略ナッシュ均衡は 3 選択肢以上の一般の対称ゲームでは存在するとはいえない。しかし, 2 選択肢の対称ゲーム (2^N 対称ゲーム) ではそれが存在することがいえる。

表2 ジャンケン

	グー	チョキ	パー
グー	0, 0	1, -1	-1, 1
チョキ	-1, 1	0, 0	1, -1
パー	1, -1	-1, 1	0, 0

定理2 (均衡の存在) 2^N 対称ゲームでは、(純粋ナッシュ) 均衡が存在する.

証明) 均衡が存在しないと仮定する. このとき, $m=0$ は均衡ではないので, $D(0) < C(1)$ である. つぎに, $m=k \leq N-2$ において, $D(k) < C(k+1)$ を仮定する. このとき, $m=k+1$ において $D(k+1) \geq C(k+2)$ とすれば, 状態集合 $m=k+1$ は, 均衡集合である. しかし仮定より均衡は存在しないので, $D(k+1) < C(k+2)$ とならねばならない. このとき, 数学的帰納法により, $\forall m \in \{0, 1, \dots, N-1\}$ において, $D(m) < C(m+1)$ が成立する. このとき, $m=N-1$ において, $D(N-1) < C(N)$ となるので, $C(N)$ は均衡である. これは仮定に反する. したがって, 均衡は存在する.

■

2^N 対称ゲームでの均衡の唯一性は, 強支配戦略の存在を意味する.

定理3 (支配戦略均衡) 2^N 対称ゲームにおいて a^* が唯一の均衡 $\Leftrightarrow a^*$ は強支配戦略の組

証明) (\Leftarrow) は明らかなので (\Rightarrow) を証明する. 多型均衡すなわち C 選択者, D 選択者がともに存在するような均衡は, 対称性 (補題 2) によりあれば必ず複数存在する (たとえば $m=2$ において均衡が存在すれば, そこでは 2 人の C 選択者の選び方の数だけ, つまり ${}_N C_2$ 個の均衡が存在する). したがって, 均衡が唯一であれば, それは単型均衡であるから, 全員 D か全員 C つまり $m^*=0$ か $m^*=N$ である.

$m^*=0$ が唯一の均衡であるとする. このとき, $D(0) \geq C(1)$ が成り立つ. いま, $\exists k \in \{0, 1, \dots, N-1\}$ に対して, $D(k) \leq C(k+1)$ であったとする. このとき, $D(k+1) \geq C(k+2)$ が成立すれば $m=k+1$ が均衡集合となるが, $m=0$ 以外では均衡は存在しないので $D(k+1) < C(k+2)$ でなければならない. このとき同様に $D(k+2) \geq C(k+3)$ が成立すれば $m=k+2$ が均衡集合になるので, $D(k+2) < C(k+3)$ となる. よって帰納的に $\forall m \in \{k+1, \dots, N-1\}$ について, $D(m) < C(m+1)$ が成立するが, このとき $D(N-1) < C(N)$ なので $m=N$ が均衡となって矛盾. よって, $\forall m \in \{0, 1, \dots, N-1\}; D(m) > C(m+1)$ が成立する. したがって, D は強支配戦略である.

以上の証明は, D と C を置き換えても成り立つ ($m=N$ も同じということ). ■

定理3 は, 3 選択肢以上では成り立たない. つまり, 唯一の純粋戦略ナッシュ均衡が強支配戦略の組ではないことがある. たとえば, 表3のような反例がある (表中の太字は均衡を示す). 表3の(1, 1)は唯一の均衡であり, 戦略に支配関係はない. したがって, この定理3は 2^N 対称

ゲームに特徴的な性質である。

なお、この定理は後にみる N 人円満ゲームと N 人囚人のジレンマで重要な意味をもつ。ただし、特定の $m \neq 0, N$ が唯一の均衡集合であるような状況は存在する。表 4 の 3 人チキンゲームでは $m=1$ が唯一の均衡集合であるが、この均衡は 3 つある。

表 3 唯一の均衡

1, 1	1, 0	0, 1
0, 1	0, 0	3, 0
1, 0	0, 3	2, 2

表 4 N 人チキンゲーム

C の人数 m	0	1	2	3
$D(m)$	0	3	4	-
$C(m)$	-	2	2	3

3. 2^N 対称協力状況

第 1 節で述べたように、武藤 (2005) は、「協力状況」なる概念を定義し、社会的ジレンマが「協力状況」に含まれることを示した。それゆえ、協力状況进行分析することは、社会的ジレンマ进行分析することになる。しかるにこの節では、協力状況进行分析。具体的には、協力状況の基本的なサブカテゴリーである N 人円満ゲーム、 N 人囚人のジレンマ、 N 人チキンゲーム、 N 人保証ゲームについて基本的な定理を示す。ただし、本稿では、数学的に厳密に議論するために、「協力状況」を「 2^N 対称協力状況」とよぶことがある。以下の諸定義は武藤 (2005) を踏襲している。

3.1 諸定義

定義 3 つぎの 2 式が成り立つような 2^N 対称ゲームを **2^N 対称協力状況** (略して協力状況) とよび、 C を「協力」行為、 D を「非協力」行為とよぶ。

$$\text{協力行為のコスト} \quad \forall m \quad D(m) > C(m)$$

$$\text{協力行為による共通の利益の存在} \quad \exists m \quad D(0) < C(m)$$

$\exists m$ について $D(0) < C(m) < D(m)$ から、「全員非協力のパレート非効率性」がいえる。

定義 4 2^N 対称協力状況において、「パレート効率的」かつ「全員非協力に対してパレート優位」となる社会状態を **協力状態** とよぶ。

じつは、 2^N 対称ゲームにおいてパレート効率性を $C(m)$ 、 $D(m)$ で表現するのは多少厄介である。そこで協力状態より緩い概念としてつぎのものを定義しておく。

定義 5 2^N 対称協力状況において、全員非協力に対してパレート優位となる社会状態 ($D(0) < C(m)$ なる $a \in S(m)$) を **準協力状態** とよぶ。

定理 2 により均衡は常に存在するので、つぎの各ゲームのクラスの定義が可能になる。

定義 6 2^N 対称協力状況において、各ゲームのクラスを以下のように定義する。

N 人円満ゲーム \Leftrightarrow 均衡が「全員協力」 ($m^*=N$) のみからなる

N 人囚人のジレンマ \Leftrightarrow 均衡が「全員非協力」 ($m^*=0$) のみからなる

N 人チキンゲーム \Leftrightarrow 均衡が「全員協力」と「全員非協力」以外からなる

N 人保証ゲーム \Leftrightarrow 均衡が「全員協力」と「全員非協力」のみからなる

これらの定義によって、 2^N 対称協力状況すべてを分類しきれぬわけではない。たとえば N 人囚人と N 人チキンゲームの「あいのこ」であるような中間的な状況が存在する (図 1 参照)。このタイプのゲームは、均衡として全員非協力 $m^*=0$ と多型状態集合 $m^* \in [1, N-1]$ をもっている。このような中間的な状況は論理的に 4 つあるが、これらについては記述が複雑になるので本稿では扱わない。今後の課題とする。

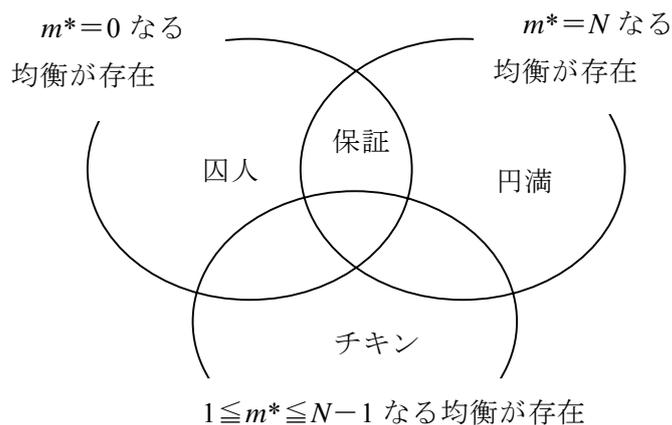


図 1 協力状況の下位分類

3.2 N 人囚人のジレンマと N 人円満ゲームの定理

2^N 対称協力状況において、均衡が唯一なのは N 人円満ゲームと N 人囚人のジレンマだけなので、定理 3 からつぎの系が明らかに成り立つ。

系 3 (支配戦略均衡)

N 人円満ゲームまたは N 人囚人のジレンマ $\Leftrightarrow 2^N$ 対称協力状況において強支配戦略が存在

N 人円満ゲームは、その名のとおり、つぎの望ましい性質をもつ。

定理 4 (円満) N 人円満ゲームの均衡 $m^*=N$ は、唯一のパレート効率的な社会状態である。

証明) $m=N$ が強支配戦略均衡 (C が強支配戦略) であることと, $\forall m; D(m) > C(m)$ から, $C(N) > D(N-1) > C(N-1) > D(N-2) > C(N-2) > \dots > D(0)$. ■

強支配戦略が存在するのは N 人囚人のジレンマか N 人円満ゲームであり, 前者の均衡 (全員非効率) は協力状況の仮定によりパレート非効率だから, パレート効率的となるのは, 後者のみである. したがって, 定理 4 から以下の系が成り立つ.

系 4A N 人円満ゲーム $\Leftrightarrow 2^N$ 対称協力状況において唯一の均衡がパレート効率的

系 4B N 人円満ゲーム $\Leftrightarrow 2^N$ 対称協力状況において強支配戦略の組がパレート効率的

系 4C N 人囚人のジレンマ $\Leftrightarrow 2^N$ 対称協力状況において唯一の均衡がパレート非効率

系 4D N 人囚人のジレンマ $\Leftrightarrow 2^N$ 対称協力状況において強支配戦略の組がパレート非効率

強支配戦略そのものは, 明らかに個人的合理性の一種であり, パレート効率性は社会的合理性とされる. そしてこの 2 水準の合理性の乖離が社会的ジレンマであった. したがって, 古典的なゲーム理論の範囲では, N 人囚人のジレンマは社会的ジレンマであるが, N 人円満ゲームは社会的ジレンマではない (ただしは進化ゲーム理論的には N 人円満ゲームも社会的ジレンマとして考えられる (武藤 2005)).

なお, 当たり前に見えるかもしれないが, N 人囚人のジレンマと N 人円満ゲームの定義は, あくまで均衡における協力者の人数のみで機械的に定義されており, そこからこれらの基本的な性質が導けるのである.

定理 4 により発見もある. 協利行為が強支配戦略であり, かつ全員協力がパレート非効率であるという可能性およびほかにもパレート効率的な社会状態が存在する可能性は, この定理 4 によってはじめて排除できる.

3.3 N 人チキンゲームと先手コミットメント

定義 6 から, 多くの協力状況は N 人チキンゲームであることがわかる. N 人チキンゲームでは協力者と非協力者が混在する多型均衡が定義上存在するが, これは全員非協力より常に「まし」であることがつぎの定理によりわかる.

定理 5 (準協力的均衡) 2^N 対称協力状況において, 全員非協力が均衡でなければ, それに対して狭義パレート優位な均衡 (準協力的均衡) が存在する.

証明) 2^N 対称ゲームでは均衡が存在するから, そのうちで C 選択者の人数 m が最小となる均衡集合を $S(m^*)$, $m^* \geq 1$ としよう (全員 D は均衡ではないため $m^* \geq 1$ となる).

以下, $C(m^*) > D(0)$ を示す. まず, 全員 D が均衡でないから, $D(0) < C(1)$ である.

いま, $m \leq m^* - 1; D(m-1) < C(m)$ だとすると, $S(m)$ は均衡集合ではないので $D(m) < C(m+1)$ でなければならない. これと $C(m) < D(m)$ とから

$m \leq m^* - 1; D(m-1) < C(m) < D(m) < C(m+1)$ がいえる.

帰納的に, $D(0) < C(1) < D(1) < C(2) < \dots < D(m^* - 1) < C(m^*)$ がいえる. ■

定理 5 の該当対象は, N 人円満ゲームと N 人チキンゲームとそれらの「あいのこ」である. したがって, N 人チキンゲームに限定すれば, つぎの系がいえる. これは, N 人チキンゲームでは準協力状態となる多型均衡が存在するということである.

系 5 (準協力的多型均衡) N 人チキンゲームでは, 全員非協力に対して狭義パレート優位な多型均衡が存在する.

ところで Taylor (1987=1995: 56-60) は, 「先手コミットメント」のインセンティブを N 人チキンゲームの本質と考えている. すなわち, Taylor の N 人チキンゲームとは, 全員非協力に対してパレート優位な多型均衡 (準協力状態) の実現を意図して「先に」非協力を選択すれば他者に協力を強制できるけれども (=先手コミットメント), だれもがそのように考えて行為を選択すると全員非協力 (パレート非効率) が帰結してしまうという社会状況のことである. 先手コミットメントは, 明らかにある種の個人的な合理性である. したがって, この意味で Taylor の N 人チキンゲームも二水準の合理性の乖離としての社会的ジレンマに該当する.

一方, 本稿の N 人チキンゲームの定義も系 5 によって, 先手コミットメントのインセンティブが存在するので, Taylor の N 人チキンゲームとほぼ同じ定義といえる. ただし, Taylor の定義では全員非協力が均衡であることもあるので, この点で本稿のチキンゲームのほうが狭い. しかし, このようなゲームは, N 人チキンゲームと N 人囚人のジレンマとの「あいのこ」とみるほうがすっきりするだろう.

3.4 N 人保証ゲームとマクスミニ戦略

マクスミニ戦略は, 他者が必ずしも合理的に振舞うとは限らない現実の相互行為では, 確実さという点で考慮に値する個人的合理性である (武藤 2005). そこで「マクスミニ戦略の組がパレート非効率になる状況」をマクスミニジレンマとよぶ. これは社会的ジレンマの一種である. なお, 本稿でのマクスミニ戦略とは, 厳密にはレクシミン戦略のことである. レクシミン戦略とは, 最悪の結果がタイのときは, 2 番目に悪い結果を比較したときのよいほうの戦略, それがタイなら 3 番目に悪い結果を比較し… (以下同様) という戦略のことである (金井 1998).

N 人保証ゲームはマクスミニジレンマであることがいえる (つぎの定理 6 の系).

定理 6 (マクスミニジレンマ) 2^N 対称協力状況において, 全員非協力が均衡であれば, 非協力はマクスミニ戦略である.

証明) 全員 D が均衡であることにより, $\min C(m) \leq C(1) \leq D(0)$.

これと $\forall m \in [1, N-1]; D(m) > C(m)$ により, $\min\{D(m), C(m)\} = \min\{C(m), D(0)\} = \min C(m)$.

また, もしも $\min C(m) = C(1) = D(0)$ であっても ($D(m)$ と $C(m)$ の最小値が同じ), 二番目に小さ

い値においては, $\forall m \in [1, N-1]; D(m) > C(m)$. よって C は最悪の結果を引き起こしうるので, D がマクスマニ戦略である. ■

N 人囚人のジレンマもマクスマニジレンマであることに注意されたい. また, N 人保証ゲームでは, 全員非協力が均衡であるから, つぎが成り立つ.

系 6 N 人保証ゲームは, マクスマニジレンマである.

3.5 N 人保証ゲームの分水嶺

N 人保証ゲームでは分水嶺 (critical mass) がよく問題にされる (山岸 1991). 分水嶺にまで協力者が達しないならば, 全員非協力 (パレート非効率) が帰結するからである. ここでいう分水嶺とは, つぎの 2 条件をみたす協力者の人数 r のことである. ①協力者が r 人以上いれば, 最適応答の繰り返しによって全員協力を到達する. ②協力者が r 人未満であれば, 最適応答の繰り返しによって全員非協力を到達する. なお, 保証ゲームでは一般に全員協力がパレート効率的であるわけではない. したがって正則な保証ゲームにはこの仮定をつけ加えるべきだが, 複雑になるのでこの問題には立ち入らない.

以下では分水嶺を厳密に定義したい. そのためにまず, 「協力者数の状態」を定義する. 協力者数 r の状態 (状態集合 $S(r)$ の状態) は, 2 つの式 $D(r-1) \square C(r), D(r) \square C(r+1)$ の 2 つの \square に入る不等号および等号 $\{<, >, =\}$ によって決定される. したがって, この状態は $3 \times 3 = 9$ つあるが, それはつぎの 3 つに分類される. 均衡概念は重複記述となるが, 許されたい.

定義 7 協力者数 $r \in [1, N-1]$ が **推移点** であるとは,

「非協力推移点」 $D(r-1) > C(r), D(r) > C(r+1)$ あるいは
「協力推移点」 $D(r-1) < C(r), D(r) < C(r+1)$

が成り立つことをいう (略号 $(>, >)$ と $(<, <)$).

定義 8 協力者数 r が **均衡** であるとは, 次式が成り立つことをいう.

$$D(r-1) \leq C(r), D(r) \geq C(r+1)$$

ただし, $D(-1), C(0), D(N), C(N+1)$ は定義されないのでこの場合は 2 つの式の一方をもって均衡を定義する (略号 $(<, >), (<, =), (=, >), (=, =)$).

定義 9 協力者数 $r \in [1, N-1]$ が **転回点** であるとは, 次式が成り立つことをいう.

$$D(r-1) \geq C(r), D(r) \leq C(r+1)$$

ただし同時に 2 式の等号は成立しないとする (略号(>, <), (>, =), (=, <)).

非協力推移が断ち切られる点は, 均衡または転回点になっている. さらに, つぎのように, 分水嶺を, 転回点の外部が推移点となっていることとして定義する (ただし両端は均衡).

定義 10 協力者数 r が**分水嶺** (critical mass) であるとは, r が転回点であり, かつ

$$\begin{aligned} m \leq r-1 \quad D(m-1) > C(m) \\ m \geq r+2 \quad D(m-1) < C(m) \end{aligned}$$

が成り立つこと (イメージとして $\gg \geq \cdot \leq \ll$).

つまり r が分水嶺であるとは, 転回点 r に対してその周辺で,

$$D(r-2) > C(r-1), D(r-1) \geq C(r), D(r) \leq C(r+1), D(r+1) < C(r+2)$$

となっており, かつ $m (< r)$ が非協力推移点かつ $m (> r)$ が協力推移点になっていること. たとえば, $N=5$ として, 次式では $m=2$ が分水嶺である.

$$D(0) > C(1), D(1) > C(2), D(2) < C(3), D(3) < C(4), D(4) < C(5)$$

うえのように, 協力者 $r=(>, <)$ すなわち

$$D(r-1) > C(r), D(r) < C(r+1)$$

で定義される協力者数 r を**孤立的な**転回点とよぶ (たとえば表 5 の $m=2$).

一方, 転回点である協力者数のペア $(r, r+1)=((>, =), (=, <))$ はあたかも 1 つの転回点のように扱うことができる. そこで,

$$D(r-1) > C(r), D(r) = C(r+1), D(r+1) < C(r+2)$$

で定義される協力者数のペア $(r, r+1)$ を**連結的な**転回点とよぶ. さらに,

$$m \leq r; D(m-1) > C(m), D(r) = C(r+1), m \geq r+2; D(m-1) < C(m)$$

が成り立つとき, 連結的な転回点 $r, r+1$ は分水嶺の定義をみたす (イメージとして $\gg \gg \cdot = \cdot \ll \ll$). たとえば, $N=5$ として, 次式では $m=2, 3$ が分水嶺である.

$$D(0) > C(1), D(1) > C(2), D(2) = C(3), D(3) < C(4), D(4) < C(5)$$

なお、注意として、 $(> \cdot > \cdot = \cdot > \cdot >)$ のような場合には、右から 2 番目の協力者数を意味する \cdot は均衡であるから、分水嶺が存在する場合、転回点の外部では同じ向きの不等号が並ばなければならない。

定理 7 (分水嶺) N 人保証ゲームでは、分水嶺が存在する。

証明) 仮定として、保証ゲームでは、 $D(0) \geq C(1)$, $D(N-1) \leq C(N)$, かつ $m=0, N$ 以外に均衡集合は存在しない。まず、転回点が存在することを示す。転回点が存在しないとすれば、任意の協力者数 m は推移点か均衡である。いま、均衡は $m=0, N$ しかないから、 $m=1, \dots, N-1$ は推移点でなければならないが、転回点、均衡がともに存在しないので、すべての点が非協力推移点 $(>>\dots>)$ かすべての点が協力推移点 $(\ll\dots\ll)$ であるはずだ。しかし、これは D または C が強支配戦略であることを意味するから明らかに矛盾。よって、転回点は存在する。

つぎに孤立的な転回点は唯一であることを示す。もし孤立的な転回点が 2 つあるとし、それを r, t ($r < r+1 < t$) とすれば $D(m-1)$ と $C(m)$ の大小関係の列 ($m=r \sim t+1$) は、

$$>_r, <_{r+1}, \square_{r+2}, \dots, \square_{t-1}, >_t, <_{t+1} \quad (\text{ただし } >_m \text{ は, } D(m-1) > C(m) \text{ を意味する})$$

となるはずだ (\square には $<$ または $>$)。 $m=r+2$ から $t-1$ では転回点、均衡点が存在しないから、単調に協力推移点 $(\ll\dots\ll)$ か非協力推移点 $(>>\dots>)$ でなければならないが、このとき $(>>\dots>)$ ならば $m=r+1$ が均衡になり ($\because D(r) < C(r+1), D(r+1) > C(r+2)$)、 $(\ll\dots\ll)$ ならば $m=t-1$ が均衡になる ($\because D(t-2) < C(t-1), D(t-1) > C(t)$)。いずれも保証ゲームの仮定に反す。したがって、孤立的な転回点は唯一である。転回点 r と $m=0, N$ 以外は推移点であるから、明らかに r が分水嶺となる。

つぎに、転回点が連結している場合を考える。それは、 $>_r =_{r+1}, <_{r+2}$ となっている場合である (3 つの点を連結することはできない)。この場合は、2 つの転回点があたかも 1 つの孤立的な転回点と同じ機能をなすので、うへの証明と同様にして連結的な転回点のペアは唯一である。転回点 $r, r+1$ と $m=0, N$ 以外は推移点であるから明らかに $r, r+1$ のペアが分水嶺となる。

■

うへの証明では、特に協力状況の仮定は用いていないので、これは均衡が 2 つの単型均衡からなる 2^N 対称ゲーム (2 選択肢対称 N 人調整ゲーム) 一般について成り立つ。

なお、この節で用いた協力者数の状態に対する 3 つの概念「推移点」「均衡」「転回点」は 2^N 対称ゲーム一般で用いることができる分析用具である。つまり、これら 3 つの組合せで、自在に 2 選択肢 N 人対称ゲームを構成することができる。

4. おわりに

本稿では、均衡の配置によって定義される N 人囚人のジレンマ、 N 人円満ゲーム、 N 人チキンゲーム、 N 人保証ゲームの各状況において成り立つ基本的な性質を、定理として示した。これらの定理は、協力状況と均衡の配置という公理から演繹したものである。まとめておこう。

1. N 人円満ゲームでは協力が強支配戦略であり、その組はパレート効率的である。
2. N 人囚人のジレンマでは非協力が強支配戦略であり、その組はパレート非効率である。
3. N 人チキンゲームは、全員非協力に対して狭義パレート優位な多型均衡が存在する。
4. N 人保証ゲームでは、非協力が常にマクスミニ戦略である。
5. N 人保証ゲームでは、分水嶺が存在する。

N 人保証ゲームにおいて非協力がマクスミニ戦略である、ということを除けば、本稿が証明した定理は、その性質としては目新しいものではない。しかし、本稿のように 2 選択肢の N 人ゲームを、均衡の配置という視点から演繹的かつ統一的に整理した研究はこれまでほとんどなかったものであり、うえにあげたような基本的な性質は、互いに関係づけられることなく断片的に知られていたにすぎない。

社会的ジレンマとは個人的合理性と社会的合理性の乖離のことであるが、武藤（2005）は、パレート効率性としての社会的合理性に対して、 N 人囚人のジレンマ、 N 人チキンゲーム、 N 人保証ゲームの各ゲームの個人的合理性を、それぞれ強支配戦略、先手コミットメント、マクスミニ戦略として位置づけていた（表 5）。そして本稿は、均衡の配置による機械的な各ゲームのクラスの定義から、質的に異なる個人的合理性を演繹的に導き出したのである。

表 5 3つの典型的な社会的ジレンマの定義と性質

ジレンマ分類	定義	個人的合理性
N 人囚人のジレンマ	均衡が $m^*=0$ のみ	強支配戦略
N 人チキンゲーム	均衡が $\neq 0, N$	先手コミットメント
N 人保証ゲーム	均衡が $m^*=0$ と N のみ	マクスミニ戦略

注. うえの定義は協力状況を前提し、社会的合理性はパレート効率性である。

今後の課題を 2 つあげておく。第 1 に、本稿は、Harsanyi (1977) において先鞭がつけられ、Raub (1988) と永田 (1988) が明確に示した「すべての均衡がパレート非効率」という社会的ジレンマの定式化を無視してきた。これはパレート効率性を定式化しなかったためである。しかし、じつはパレート効率性を $C(m)$ と $D(m)$ で定式化するのは場合分けがあり、すこし複雑である。また、 N 人チキンゲームの多型均衡と保証ゲームの全員協力均衡は、パレート非効率であることがある。このような複雑な議論をするためには、さらにいくつかの概念を定義しなければならず、諸社会的ジレンマの基本性質を定理として示すという目的をもつ本稿では割愛した。しかしこれは、もちろん今後の課題である。

第2に、本稿では N 人囚人のジレンマ、 N 人円満ゲーム、 N 人チキンゲームのありうる4つの中間的な状況のうち、 N 人囚人のジレンマと N 人円満ゲームの中間的な状況である N 人保証ゲームについては分析したが、ほかの3つの中間的な状況は分析していない。つまり、① N 人囚人のジレンマと N 人チキンゲームの中間的な状況、② N 人円満ゲームと N 人チキンゲームの中間的な状況、そして③ N 人囚人のジレンマ、 N 人円満ゲーム、 N 人チキンゲームすべての性質をもつ中間的な状況 (= N 人保証ゲームと N 人チキンゲームの中間的な状況) の分析が課題として残されている。

本稿によって、組織や集合行為に必須の N 人ゲームという複雑で不透明な状況が、すこしでも見通しのよいものになれば幸いである。

【文献】

- Dawes, R. M., 1980, "Social Dilemmas", *Annual Review of Psychology*, 31: 169-193.
- Elster, J., 1989., *The Nuts and Bolts for the Social Sciences*, Cambridge University Press. (=1997, 海野道郎訳『社会科学の道具箱』ハーベスト社.)
- 金井雅之, 1998, 「功利主義とレクシミン」『数理科学』3月号.
- Harsanyi, J. C., 1977, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.
- 木村邦博, 2002, 『大集団のジレンマ』ミネルヴァ書房.
- 永田えり子, 1988, 「自由と効率」『理論と方法』3(1): 43-56.
- 森村進, 1987, 「「リヴァイアサン」の自然状態」『理想』637: 85-93.
- 武藤正義, 2005, 「社会的ジレンマと協力概念: フリーライドの視点による N 人ゲームの構造分析 I」大浦宏邦編『秩序問題への進化ゲーム理論的アプローチ』平成 14~16 年度科学研究費補助金・研究成果報告書, 帝京大学.
- Raub, W. 1988. "Problematic Social Situation and the Large Number of Dilemma: A Game-theoretical Analysis." *Journal of Mathematical Sociology*. 13(4): 311-357.
- 鈴木鉄忠, 2003, 「集合行為問題のゲーム理論的分析及び社会学的考察」平成 14 年度東京工業大学社会理工学研究科修士論文.
- 瀧川哲夫, 1989, 「 N 人ゲーム状況の構造解析」『北海道大学文学部紀要』37: 77-146.
- Taylor, M. 1987. *Possibility of Cooperation*. Cambridge University Press. (=1995 松原望訳『協力の可能性』木鐸社.)
- 山岸俊男, 1991, 「社会的交換と社会的ジレンマ」盛山和夫・海野道朗編『秩序問題と社会的ジレンマ』ハーベスト社.

Theorems of Social Dilemmas:

N-person Game Structure Analysis from the viewpoint of Free Ride II

Masayoshi MUTO

Graduate School of Decision Science and Technology

Tokyo Institute of Technology

2-12-1 Oookayama, Meguro, Tokyo 152-8552, JAPAN