



Title	平等主義の進化的起源と規範形成のゲーム・モデル
Author(s)	町野, 和夫
Citation	経済學研究, 58(4), 199-215
Issue Date	2009-03-12
Doc URL	https://hdl.handle.net/2115/37309
Type	departmental bulletin paper
File Information	58-4_016.pdf



平等主義の進化的起源と規範形成のゲーム・モデル

町 野 和 夫

1. はじめに

アダム・スミスの時代は倫理的規範も経済学者の主要な関心領域の一つであったが、その後数理的経済理論が発展する中でこうした問題を扱うのは社会的選択論など厚生経済学の特定の分野のみであった。しかし、最近では経済学者も実験経済学や行動経済学によっていわゆるホモ・エコノミクス（経済人）的合理性の論理とは違う人間の行動に再び着目するようになった。限定合理的プレーヤによる協力的行動の進化あるいは慣習や規範の形成を進化・学習ゲームを使ってモデル化しようとする研究も少なくない。ただし、ある評価軸を決めて公共的政策を評価しようとする厚生経済学に対して、現実の均衡状態をより正確に説明するモデルを考案しようとするのが進化ゲームなどの研究である。

倫理的規範は社会科学あるいは哲学では中心的な概念の一つであり、政治哲学・法哲学の分野では、秩序や法の正当性を議論するために、善悪を判断する基準は何か、あるいは正義とは何かを考えることは重要な研究テーマである。この問題に対する考え方としては、倫理的規範を公理的・演繹的に組み立てていくアプローチと人間社会の現実や歴史から経験的・帰納的に導出するアプローチが形を変えて長年対抗し続けているが、両者の違いは上述の経済学の中での厚生経済学と進化ゲームの分析視点の違いと平行である。しかし、両者は経済学の主流派と違い、経済主体を目的合理的に行動するだけの主体と考えないことで向社会的行動を説明しようとする点では共通である。

本稿は進化ゲーム的アプローチを取るものの、町野（2003）と同じく、既存研究に欠けていた、規範に対する価値観の獲得のプロセスのモデル化を試みる。とくに、従来の進化ゲーム・モデルが、一部のプレーヤの選好の一部として利他主義、互酬性の存在を仮定していたのに対して、本稿ではそのような選好がどのように生じてきたかをモデル化することに焦点を絞って考察を進める。町野（2003）では、社会のルールとしての倫理規範を大人が子供に躰ける社会化過程と、ある行為が社会のルールであるべきだと提案者が社会（共同体）に説得し、それが倫理的規範に変化する過程をモデル化した。前者では、とくに大人のインセンティブを厳密に考えることで、従来の学習モデルとは異なるゲーム構造を提示した。後者では提案者のイニシアティブの重要性を明らかにし、従来の動学的進化ゲームに欠けていた倫理性発生のモデル化を試みた。

後者の社会的ルールを提案する側の動機について考えると、利己的なものもあるが、倫理的なものもある。提案する側の動機が倫理的なものだとすると、それを教えた者の動機は何か、そしてそれを教えた者の動機は... と、因果関係に関する議論は無限に遡る社会化過程に戻ってしまう可能性がある。従って、倫理規範形成過程モデル化の第一段階として、そうした無限の連鎖を断ち、倫理規範の合理的形成の論理を提供しなければならない。そこで、何らかの自己犠牲的行動を他の共同体メンバーに説得する、合理的な提案者の存在が必要となる。説得された行為が共同体にとって有益であることが実証されれば、この行為が次第に「倫理的行為」と

して認識され、その認識が共同体に浸透していく。このようなある種の倫理的行為が集団の中で一定の割合で存在すれば、それが慣習・規範化することは、これまでの多くの進化ゲームの研究成果から明らかである。

プレーヤがなぜ利他主義的、互酬的、あるいは集団主義的選好を持つかという問いは、言い換えると、人が何故ある行為を「善い」ものと考え、ある行為を「悪い」と考えるのかという問いである。これは上述のように決着のつかない政治哲学的問いでもあるが、ここではそうした善悪に関する選好を、人間の社会的動物としての、あるいは民族や個人としての経験が積み重ねられて形成されたものと考えている。社会的動物としての歴史とは、自然淘汰が結果として生物学的な意味で「利他的」行為へと向かわせるよう人間社会に組み込まれた規則であり、我々の感情は、生き残りに役立った行為が「正しく」、役立たなかった行為は「悪い」という感覚によって裏打ちされている。(内井, 1996)

人間の社会的動物としての本能については、動物行動学、とりわけ類人猿の研究で近年多くのことが確かめられてきた。類人猿が利他的、互酬的、あるいは集団主義的行動を取るのは、親子間の愛着というような一般の動物としての本能からだけでなく、社会的動物として生存のために集団に所属することが不可欠なため、集団の様々な規則（食物の部の分配や交配に関する規則など）に従うためである事が明らかにされている。(de Waal, 1996) 集団生活を行う類人猿の観察から、集団間の序列を守らせるために、規則を守らないメンバーに対する攻撃・制裁行動が「感情」を引き金にして引き起こされることが観察されている。さらに、類人猿と猿の集団の比較からは、自意識を持つことの出来る類人猿が、そののできない猿の固定的な階層社会と異なり、党派行動をとることによって、より複雑で変動的な社会構造を構成することも分かってきた。こうした類人猿の行動は、彼らが自分と他者を区別して、お互いの行動を予測

し、ある種の契約・信頼関係を築けることを意味する。そして他者の行動に対する予測が裏切られたときに、裏切ったものへの「怒り」が生じ、それが攻撃へと繋がる。しかし一方で、集団を維持することの高い必要性から、対立の和解への過程も用意されており、個々の対立が集団を崩壊させないような仕組みも内蔵されている。この和解に至る過程も、制裁を受けたことの「痛み」「悲しみ」を和らげるという効果があり、それは対立する当事者以外が和解を取り持ったり、和解の直後に集団の他のメンバーたちが「喜ぶ」ことによって、彼らが他者の気持ちを予測して「共感」したり「同情」したりする能力があることを示している。

こうした動物行動学の知見を受け入れると、人間の善悪の感情も、他者との力関係から自己利益に反することを受け入れなければならない状況で、自身の行動を正当化するための感情と言えるかもしれない。あるいは、他者を喜ばせることが自身の利益になる状況で、(幼児と親、生徒と先生、家来と主人など) 他者を喜ばせる自分の行為を正当化するためかも知れない。しかしこの類人猿の行動を規制するルールは人間の倫理規範とはかなり距離がある。そこで次節では、その違いを、人間の主要な倫理的規範である平等の観点から考察する。具体的には、動物行動学・人類学の平等主義の進化的起源についての一つの有力な見方を紹介し、それを踏まえた上で、第3節で倫理的規範が集団に植えつけられるメカニズム(社会化)のゲーム・モデル化を試みる。最後に第4節でモデルから得られた含意を検討する。

2. 平等主義の進化的起源

Boehm (2004) は類人猿と非定住型狩猟・採集民との比較によって、平等主義的行動における両者の共通点と相違点を明らかにし人間社会の倫理性の進化的起源について考察している。彼が注目するのは大きな獲物を、その猟に貢献

したかどうかに関係なく共同体全員に平等に分配する慣習である。チンパンジーも集団で狩をし、その獲物を分けるが、分配の仕方は食糧の豊富さなど様々な状況によって異なる。とくに人間と異なるところは、集団で獲物を分配するといっても基本的には集団のボスや獲物を捕獲した者にせびるようにして貰うことである。これに対して狩猟・採集民の大きな獲物の分配はもっと制度的である。Boehmによると後者には共有財産という概念が当てはまるが、前者の共同の狩で得た獲物はあくまでも私的所有物としての性格を持っている。チンパンジーの集団では獲物を殺したのが雄の大人であれば、ボスであっても分けてくれるように頼まなければ（頼むような態度をとらなければ）ならない。ただし、獲物を殺したのが雌や若者であればボスは取り上げる。

人間も類人猿も大きな獲物を共同体のメンバー全員で平等に分ける、人間はそれが制度化されているところが類人猿とは異なるものの、平等主義という人間の主要な倫理的性質が進化的起源を持つことが推測される。この分配行動の生物学的理由は以下のとおりである。大きな獲物を得たときに、狩に加わったものだけで食べたり、いつもボスや有力者だけが分け前に預かったりするのではなく、共同体皆で分配することにすれば、個人や個々の家族の食糧（タンパク質）の摂取が著しく変動する確率は低くなる。なぜなら家族単位ではたまにしか獲れない大きな獲物を、自分たちが獲った時に他の家族に分ける代わりに他の家族が獲った時には自分たちにも分けてもらえるからである。（これは変動削減（variance-reduction）理論と呼ばれる）。この、大きな獲物をめぐる平等主義が（観察された150の非定住型狩猟・採集民の）様々な共同体で例外なく観察されるのは、食料に乏しく、大きな動物はなかなか捕えられないが不可欠の食糧源であった（12.8万年前から7.2万年前にかけての）氷河期を生き抜いてきた人類（Homo Sapiens）の進化の結果と考えられ

る¹⁾。その過程で脳が（従って社会的な関係を築く能力も）発達し、ボスの力をその他のメンバーが共同して抑え込むという平等主義を実現する政治的構造が可能になったと推測される。

Boehm (1999) は、こうした（平等主義という）倫理観を持つ道徳的共同体が成立するためのいくつかの前提条件（進化の準備段階）を挙げている。

- ① 支配したい、支配されたくないという性格
- ② 武器の発見・発明（これによって身体の大きさや強さの重要性が低下した。犬歯の退化や体毛の減少、二足歩行などとも関連すると思われ、50万年（2万世代）前までに変化したと考えられる。）
- ③ 知性の発達（多数の提携を形成する政治的知性と、合理的な（資源に余裕があれば障害を受けた者も世話するが限りがあれば見捨てる）セーフティーネットを構築する保険数理的知性。これらは10万年前から20万年前に脳の拡大に伴って発達したと思われる。）
- ④ コミュニケーション能力（言語、情報網、については類人猿も一定の能力はあったが、彼らの行動の選好は道徳規範として明確にできるほどのものではなく、言語と言えるものは約10万年前に人類によって完成したと考えられる。言語がなければ協力、寛大さ、利他性などの複雑な概念を繰り返し説いたり教えたりできない。）

彼によれば平等主義的な共同体を維持するには、それに反するメンバーが出てきてから対処するのでは遅く、独裁者が現れないような不断の警戒が必要であり、共同体のメンバーが「道

1) 大きな獲物の平等的分配に関しては、狩猟のうまさ誇示することによって性的な魅力になるという評判あるいはシグナリングの効果を重視する研究もある。(Hawks and Bird, 2002)

徳」的で相互に監視しあうことが有効である。①のように、メンバーが支配されたくないという性格を持っていれば、独裁的な行動に対する憤りは共有されやすい。それが集団での政治的非協力となり、独裁的行動は不適切であるという道徳的感情が定着し、さらにそうした行動への制裁へと向かうこともある。道徳的感情の存在が先か、支配への対抗としての制裁のための提携が先で、それが他の「反社会的」行動へも拡大して道徳になったのかは不明である。独裁的な行動を取る可能性のある力の強いボスに対しては、②の武器の使用によって力の弱いメンバーでも反撃しやすくなり、またそのことによって遺伝的にも身体の高さや力の差が小さくなり、ボスと他のメンバーとの体力差も縮まる。さらに、支配される側が提携してボスに対抗するには③のような政治的な知性が必要である。

上述のように、氷河期は食糧難という平等主義的傾向を促進する環境であった。しかし氷河期を過ぎても、一度支配者からの政治的自由を享受すると、その政治構造は共同体のメンバーから支持され維持される。他の独裁的な共同体で支配されているメンバーにも平等主義の利点は明らかなので、平等主義は他のグループにも広がっていく。食糧難の氷河期には各共同体の移動も激しくなるので共同体間の接触も増え、平等主義はさらに広がったと推測される。

進化生物学者や人類学者は、従来、自然淘汰は縁故者を助ける性質を残すように働くが、利他的な性質を残すようには働かないと考えてきた。個体が集団の生存競争のための犠牲になるという集団選択（群淘汰）の理論が広く受け入れられていた時代もあるが、現在は否定的な評価が一般的（例えば、内田（2007））である。集団間の利他的行為は血縁選択（血縁淘汰）や強制、長期的利益との交換（利己的互惠性）によってなされると考えられている²⁾。

こうした集団選択に対する否定的な議論の多くは、ハチやアリなどを含む人間以外の社会的生物について行われており、また、上述の集団

メンバーの倫理的行動は、リスクを冒して他のメンバーと提携して強力なボスに対抗し、成功すれば自分も利益を得られるということから自己犠牲ではなく利己的互惠性と言えるかもしれない。したがってこの集団選択に対する批判はBoehmの議論と矛盾するとは限らないが、彼は他者を助ける行為を血縁選択や利己的互惠性に限る議論は、人間が一般に協力的性向を持つという直観や多くの非定住型狩猟・採集民の共同体で困っているメンバーへの一方的な援助や集団内で利他性が常に熱心に説かれ教えられているという観察とは整合的でないと主張する。とくに、平等主義的倫理観が定着してきた旧石器時代の後期の人類にとっては、集団内の個体同士の競争による淘汰の方が集団間の淘汰より重要だという通説は妥当しないと考え、その理由として次の4つをあげている。

まず、上述の進化生物学者などの用いる数学モデルでは、ほとんどの生物で遺伝子型と表現型の違いは小さいのでイコールとみなしているというのが通説であるが、人間の道徳的な（＝平等主義的な）共同体では異なる³⁾。集団の倫理観やそれに基づく制裁によって支えられている肉の平等な分配などの表現型は、遺伝子型である自然な支配欲（あるいは支配されたくないという欲求）と異なっている。しかし遺伝子型と異なるその表現型が、結果的に食糧や生殖機会の平均化をもたらした集団としての再生産能力を高めた。

次に、狩猟・採集民の共同体は、狩猟や防衛のため集団で行動しなければならないが、平等

2) 親が自分を犠牲にして子を救うような行為は、進化生物学などでは利他的とは呼ばず血縁選択（血縁淘汰）と呼ぶ。遺伝的繋がりがあれば親子に限らない。

3) 生物の個体が持っている形態や行動上の特徴のことを形質と呼び、形質が遺伝子に支配されているとき、その形質のそれぞれの変異を「表現型」と呼ぶ。また、そのような表現型を作り出している遺伝子のタイプを「遺伝子型」と呼ぶ。（長谷川・長谷川、2000、p27 参照）

主義社会では合意形成が必要になる。合意形成がされなければ集団は弱体化ないし離散し、メンバーの生存確率は低下する。合意形成が成功するとメンバー間の協調行動が増え、個人間の違いは減少する。即ち個人間の競争による淘汰は緩和される。

第三に、こうした合意形成による集団行動は、移住場所の決定など自然淘汰における成功と失敗の決定的要因になり、そのため集団間の違いが集団内の個人差より重要になった。

最後に現在の狩猟・採集民を観察すると、フリーライダーをコントロールするために、ゴシップ、からかい、村八分など多様な制裁が使われる。利他主義を「立法（村の掟）化」する場合もある。このように、人類は子供の社会化、利他的行為の賞賛とフリーライダーへの非難、によって社会的に利他主義を強制してきたと考えられる。

このように Boehm の議論は従来の集団選択論（群淘汰）とは異なる論理で集団間の淘汰が平等主義という倫理観を社会に浸透させるというシナリオを提唱している。次節ではこの Boehm のシナリオに基づいて倫理的規範が社会的に形成されるメカニズムをゲーム・モデル化する。

3. モデル全体の枠組み

我々は既に社会化あるいは教育によって、その社会や時代の人々に共有された価値観と、個人の独自の経験による価値観とが内面化されて、共通性と独自性を備えた価値観を持っている。したがって、何が倫理的かあるいは「善いことか」を考察するときに、自分の価値観を前提にしないように注意しつつ、しかし自分の価値観も一つの材料にしながら、価値観はどのように形成されてきたのかを客観的に考えていかなければならない。

前節で紹介したように、Boehm が分析した狩猟・採集民の共同体では、平等主義的倫理観

をメンバーに浸透させるために各種の社会化が行われていた。人間の社会ならどこでも、子供の頃の家庭教育、学校教育、その他の社会化によって、社会の価値観が個々のメンバーに教え込まれる。もちろん、その後、所属する社会集団の影響や個人的経験・思索を経て個人の倫理観が変化することは大いにあり得る。しかし、成人するまでの教育や社会化によって、家族（養育者）、地域社会、学校、職場、宗教組織などによって形成された倫理観はその人の倫理観の基盤となることは確かであろう。では、教育・社会化する側は、何を何故どうやって教えるのだろうか。

狩猟・採集民の共同体での独裁的ボスをメンバーが提携して抑えるように、グループの個々のメンバーの行動がグループ全体の利益に影響するとき、個々のメンバーの行動をコントロール（抑制や促進）することは結果的に（全員ではないかもしれないが）他のメンバーの利益になるはずである。「自己犠牲」を伴うことが倫理的行動であるとする、親の本能的な養育の欲求や防衛反応である血縁選択以外では、そのような行為が「教育」による「納得」無しに自然に身に付いているとは考えられない。したがって、「善いこと」であると納得するメカニズムとはどのようなものを明らかにしなければならない。

最初にだれかが、（おそらく自分も含めた）集団メンバーの「自己犠牲的」行為が、自分の利益を増加させることに気づくと、その行為は自分に高い利得をもたらす「良いこと」だと気づく⁴⁾。しかし、その行動は、個々のメンバーにとっては短期的に損失をもたらす。長期的に利益をもたらすかどうかは不確実なので、選択させるには、それが「善いこと」だと説得して

4) この「気づき」は、試行錯誤の学習ゲームや調整ゲームで均衡を導く突出性（salience）を発見する過程と解釈できる。（例えば Sugden, 2004 参照。）

利得を増加させる（と思込ませる）か、その行動をとらない場合のコストを高める（強制する）しかない。いずれにしても多大のコストが必要である。

独裁的なボスを皆で抑えて、狩で得た大きな獲物を皆で平等に分配することを考えてみよう。ここでは、この習慣がまだない、つまり倫理的規範が無い世界を考えているので、初めは、他者に「自己犠牲的」行動を取らせるコスト（例えば自分もその行動を取ってみせる）をかけても、そのことで大きな利益を得る人がいなければ、原則として倫理的行動は広がらない⁵⁾。このイニシアティブのインセンティブを持つのは、例えば、ボスと腕力が拮抗しているメンバー、皆をまとめる政治的知性を持つメンバー（ムラの長老）、などであろう。しかし、その結果提携が成功して皆に利益があり、自己犠牲を強いられた（説得された）個人も長期的には利益を受ければ、その行為を正当化する論理が受け入れられ、受け入れる人がある程度増えれば、その規範が維持される、という好循環を生み出せる。

最初にその行動を取らせるようにする説得が可能であるためには、説得されるメンバーに記憶力、想像力、共感能力がなければならない。また、その規範的行動を継続して行ううちに、納得した論理を忘れることで、その行為を当然の行為と思ひこまなければならない。そうすると、自分がそれに反する場合は罪悪感が生じ、

他人がそれに反する場合は憤りを感じるようになる⁶⁾。この納得した規範が、個人に内面化された倫理規範に転換する過程は数年を要するかもしれないし、何世代かを要するかもしれない。また、思想家や宗教家などによって既存の倫理体系にうまく取り入れられたり、政治的リーダーやオピニオン・リーダーに取り上げられたりすることによって、社会への浸透速度は速まることもある。

要約すると、社会的には望ましいが個人には負担のかかるある行動が、倫理規範としてその社会に定着するには、メンバーの誰かのイニシアティブ、その行動の社会的有用性のメンバーからの認知、その行動が善いものであることの、メンバーへの内面化という3段階のステップが必要である。また、それが長期的にその社会で維持され続けるには、教育・社会化という次世代へ引継がれるメカニズムが必要である。

3-1 ステージ・ゲーム

本項では、まず倫理的行動を「社会的には望ましいが個人には負担のかかる」行動と定義する。「社会的に望ましい」というのは、個々のメンバーが、それぞれの費用を上回る便益を享受でき、さらにイニシアティブをとるリーダーによる説得、モニター、非協力者の便益享受阻止などの費用も賄えることとする。この倫理的行動をリーダーがメンバーに提案し説得する過程をモデル化するのだが、リーダーはこの行動を論理的あるいは強制的に「説得」するのである、この行動を、義務感を伴う倫理規範とは呼べない。リーダーの諸コストを考えると、

5) 多くの人が同時に「自己犠牲的」な行為のメリットに気づけば、少ないコストで倫理的行動が広がる可能性は論理的にはある。むしろその方が、社会的動物である人類の古くからの共有感覚の説明として正しいかもしれない。人間がそもそも集団で生活する動物であり、個々の成員が協力しなければ生存できなかったことから、集団的行動を乱すことは悪いこと（他のメンバーの期待（expectation）を裏切り、怒りを呼ぶ行為）であるという認識は、無意識のうちに人類に埋め込まれた常識であるというところから議論を始めてもいいかもしれない。

6) Sugden (2004) はこれを規範的予想（当然の行為と思う相手の行為予想）とのギャップとして捉え、それを含んだ主観的利得を基にした均衡を規範的予想均衡と呼ぶ。期待を裏切られたときの憤りを避けようとして、自分の（物質的）短期的利益に反する行動を取る。他方で、憤りのためにコストを伴う非利己的行動である報復的行動をとることもある。

リーダーの提案・説得を受け入れるメンバーの人数が少ないと、協同行為としての、この倫理的行動は成功する可能性はほとんど無く、受け入れる人数が多いほど成功する確率が高くなると想定できる。当初は、リーダーの提案が、そもそも個々のメンバーにとって有益かどうか分からないという非対称情報と、提案された倫理的行動が実際にうまくいかないかもしれないという不確実性がある中で、提案された行為を実行してもらえ水準まで納得してもらわなければならない。非対称情報については、それを少しでも高めるための説得が必要である⁷⁾。不確実性については、説得する側の提示した因果関係、具体的にはその協同行為が成功する確率や利得についての情報を納得してもらわなければならない。通常のモデルではこのような情報については共有知識と仮定されるが、それらが共有知識になるためには、提案者の使う論理が、説得される側の信じている因果関係と整合的でなければならない。前節のシナリオを念頭に置くと、独裁的ボスに支配されたくないという感情はほぼ共有されているとみなせる。

また、前節の狩猟・採集民の共同体や氏族などの共同体が成立していることを前提にすると、共同体メンバーの間で、それぞれのメンバーが正直である事前確率も、ある程度メンバー間に共有されていると思われる。各メンバーの提案者としての信頼性の高さも、共同体の中でほぼコンセンサスがあるであろう。また、提案者の使う論理も、共同体のメンバーが共有している因果関係の種類から大きく外れるものは少ないと考えてよいだろう。即ち、提案者の信頼性、彼(女)の提案する倫理的行動の主観的成功確率、費用便益効果の期待値は、概ね共有知識であり、その協同行為が成功する事前確率や期待利得も、

説得が始まる前に、共同体の歴史の中でほぼ決まっている。

このモデルで誰がどのように共同体のリーダーになるかという点はもう少し説明が必要であろう。上述の動物行動学の研究から明らかなように社会的動物である人間は、生存のために共同で生活することが必要であり、その中で力や知恵のあるものが、食料分配や異性の獲得をめぐる争いに勝ち、ボスの地位を得る。その共同体全体の利益が長期的にはボスの利益にもなることは明らかであるが、ボスには自分の短期的直接的な利益や支配欲のために、共同体全体の利益(例えば氷河期の肉の分配)を犠牲にするインセンティブがある。他方ボス以外の共同体のメンバーは、ボスの専横に対して反感を持つが、反抗するコスト(痛めつけられる)と便益(成功して食や性の満足度が上がる)と反抗の成功確率、服従するコスト(不十分な食や性、心理的不満など)と便益(とりあえずの安全)を(意識的にではないとしても)考える。もしボスが独裁的に振る舞い他のメンバーの不満が高まれば、反抗する便益は増え、服従するコストは増え、便益は減る。また、多くのメンバーの反感が高まれば、反抗に参加するメンバーの割合も増え、反抗が成功する確率も高まる。ここで、最初に反抗のイニシアティブを取るのは、上述したようにボスと腕力が拮抗しているメンバー、皆をまとめる政治的知性を持つムラの長老などであろう。彼らは服従のコストが高い(不満が大きい)、成功確率を高める方法に気付く、反抗勢力をまとめるコミュニケーション能力が高い、などの理由で反抗勢力のリーダーとなる⁸⁾。

他のメンバーがいったん「説得」を受け入れて「倫理的」行動(ここでは独裁的ボスへの反抗)を行った後は、結果として利益が費用を上

7) 事前に何の先入観も無いとすれば、有益なものである事前確率は当初は二分の一という考え方もある。しかし通常、新しい提案には提案者、提案理由など、何らかの情報を伴っており、成功確率が中立的であることは少ない。

8) 勝者の呪い(winner's curse)と同様、誤って成功確率を高く予測して失敗する可能性もあるが、ここではそれはないと仮定する。

回るほど高ければ、その「倫理的」行動の良さに対する他のメンバーの信頼が高まってくる。不確実性もあるので、常に共同体にとっての利益を実現できるとは限らないが、長期的に共同体の利益を増進できれば、信頼されるようになる⁹⁾。つまり、歴史的な実績の積み重ねがあれば、当初の意図に拘らず信頼は醸成される。

この過程は形式的にはベイジアン・ゲームとしてモデル化できる。ここでは、現実の結果を見て、個人の自己犠牲の費用を上回る利益が実現すれば、その行為が「良いこと」である事後確率を、より高い数値へ修正するし、逆の場合はより低い数値へ修正する。

以上の状況を素直にゲーム・モデル化すると、提案者が一人、その他の共同体メンバーが多数のゲームになる。まず、このゲームの構成ゲームは以下のようなゲームである。

プレーヤ：{一人の共同体リーダー、(独裁的ボスを除く) M 人の共同体メンバー、自然}
 {成功, 失敗}, {成功, 失敗} の戦略集合：
 {善い提案, 悪い提案}

共同体リーダーの t 期の戦略集合：{説得, 説得せず}

共同体メンバーの t 期の戦略集合：{受容, 拒否}

共同体リーダーの利得関数：以下で説明

共同体メンバーの利得関数：以下で説明

リーダーの提案のタイプが共同体にとって善いものであるという、共同体メンバーの信念： $\alpha (0 \leq \alpha \leq 1)$

提案者と個々のメンバーによる次のような流れと利得を持つ二人ゲームが、ステージ・ゲームとして長期的に繰り返される。以下の過程

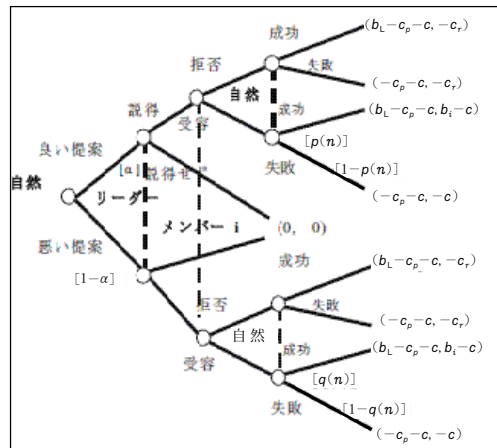


図1 ステージ・ゲーム：リーダー vs. メンバー

が、まず言葉による説得によって行われ、そこで納得されれば、以後は実績による再評価も加わって行われる。最初の言葉による説得の際には、共同体メンバーは、利得関数のパラメータ、提案の善さについての事前確率、倫理的行動の成功確率を、過去の類似の経験や提案者の説得を勘案しながら設定し、受容したときと拒否したときの期待利得を比較して、諾否を決定する。そこで受容されれば、実際に下記の④～⑥のステップが実行される。説得を受け入れた人数 n_t が多いほど提案された倫理的行動の成功確率は上昇し、とくにある水準を上回るとほぼ成功するが、リーダーの説得を受け入れたメンバーの人数がその水準に達しないとほぼ失敗する。

ステージ・ゲームの流れ

- ① 自然 (N) が、ある倫理的行動のタイプを、共同体利益の実現可能性の高いタイプ (善い提案) か、そうでないかの二つから一つ選ぶ。
- ② リーダーは、いずれのタイプの行動であっても、(説得しないときの利得は0なので) 自己の期待利得が正であれば、その行動を取るように他の共同体メンバーを「説得」する。このときの説得コストを c_p と

9) 「共同体の利益」は防衛や大きな獲物の獲得のように客観的に成否が区別しやすいものもあるし、成否の評価が主観的で個人ごとに異なるような行為もあるが、前者にも多くの要因があるので当該行動の評価は主観的である。

する。

- ③ 説得を受けた共同体メンバーは、説得を受け入れるか拒否するか決める。拒否した場合、メンバーには費用 ($-c_r$) がかかる。
- ④ 提案者の説得を聞き入れたメンバーは自己犠牲 ($-c$) を伴う行動をとる。提案者自身も同じ行動をとる。
- ⑤ 説得された倫理的行動が共同体利益を実現するかどうかは、自然 (N) に依存して決まる。ただし共同体利益実現の可能性は、善い提案の場合 ($p(n_t)$) が、悪い提案の場合 ($q(n_t)$) より高い。
($0 < q < p < 1$)
- ⑥ 共同体利益が実現した場合は、リーダーもメンバーも共同体利益 (それぞれ b_L, b_i) を得る。リーダーと説得を受け入れたメンバーの純利得 (便益から費用を引いたもの) はその倫理的行動がうまく行ったときは正だが、説得を拒否したメンバーの利得は常に非正 ($-c_r$) である¹⁰⁾。

このステージ・ゲームでは図1のように描ける。均衡解を求めるために、まず共同体メンバーが提案者の説得を受容するときの t 期の期待利得 $E\pi_{at}$ を求める。

$$\begin{aligned} E\pi_{at} &= \alpha_t \{p(n_{t-1})(b_i - c) - (1 - p(n_{t-1}))c\} + \\ &\quad (1 - \alpha_t) \{q(n_{t-1})(b_i - c) - (1 - q(n_{t-1}))c\} \\ &= \alpha_t (p(n_{t-1}) - q(n_{t-1}))b_i + q(n_{t-1})b_i - c \end{aligned}$$

なお、 $p(n)$ と $q(n)$ は、常に $p(n) > q(n)$ であるが、 n が低い水準ではほぼ0であり、ある程度以上高い水準ではそれぞれの上限、 p^* ,

q^* に近づく S 字型の関数と仮定する。この形状は、ここで議論している協同作業が、参加者の人数が増えると成功確率が高まるようなものであると考えていることからきている¹¹⁾。提案の受容者数を推定するのは個々のメンバーには困難なので、 $E n_t = n_{t-1}$ と仮定する。また、第1期の受容者の割合を α_1 と仮定すると、

$$E n_1 = \alpha_1 M.$$

共同体メンバーが提案者の説得を拒絶するときの t 期の利得 π_{rt} は、⑥から

$$\pi_{rt} = -c_r.$$

上の $E\pi_{at}$ と比べ、 $E\pi_{at} > \pi_{rt}$ のとき、即ち

$$\alpha_t (p(n_{t-1}) - q(n_{t-1}))b_i + q(n_{t-1})b_i - c > -c_r$$

のときに、共同体メンバーはリーダーの説得に納得し、その提案を受け入れる。この式を整理すると次の(1)式となり、期待便益が提案の受容と拒絶のコスト差より大きければ説得が成功することがわかる。

$$\{\alpha_t p(n_{t-1}) + (1 - \alpha_t) q(n_{t-1})\} b_i > c - c_r \quad (1)$$

他方、リーダー側の t 期の期待利得は、提案するとき、

$$\begin{aligned} E\pi_{pt}^L &= \alpha_t \{p(n_{t-1})(b_L - c_p - c) \\ &\quad - (1 - p(n_{t-1}))(c_p + c)\} \\ &\quad + (1 - \alpha_t) \{q(n_{t-1})(b_L - c_p - c) \\ &\quad - (1 - q(n_{t-1}))(c_p + c)\} \\ &= \alpha_t (p(n_{t-1}) - q(n_{t-1}))b_L + q(n_{t-1})b_L - c_p - c \end{aligned}$$

であり、提案をしないときには、

$$\pi_{nt}^L = 0.$$

10) 二つの提案の利得の違いは多様でありうるが、町野 (2003) では、提案者の利得関数として、公共的提案のときは公共的利益が実現したときに正になり、私的提案のときは公共的利益が実現しないときの正になるような項を想定して、私的提案が公共的提案と利害が相反するゲームを考えた。

11) 参加者の人数と成功確率が正の相関をもつことについては異論はないであろうが、この関数の具体的な形状については別の動学モデルやシミュレーションによる研究が必要であろう。今後の課題の一つである。

したがって、 $E\pi_{pt}^L > \pi_{nt}^L$ のとき、即ち

$$\alpha_i(p(n_{t-1}) - q(n_{t-1}))b_L + q(n_{t-1})b_L - c_p - c > 0$$

のとき、リーダーは提案する。この式を整理すると次の(2)式となり、期待便益が提案の説得と実行のコストの合計より大きければ提案の説得を試みることがわかる。

$$\{\alpha_i p(n_{t-1}) + (1 - \alpha_i)q(n_{t-1})\}b_L > c + c_p \quad (2)$$

(1), (2)の両式を変形して次の二式, (1'), (2')を得る。

$$\alpha_i p(n_{t-1}) + (1 - \alpha_i)q(n_{t-1}) > (c - c_r)/b_i \quad (1')$$

$$\alpha_i p(n_{t-1}) + (1 - \alpha_i)q(n_{t-1}) > (c + c_p)/b_L \quad (2')$$

なお、 $1 \geq \alpha_i \geq 0$, $1 > p(n_{t-1}) > q(n_{t-1}) > 0$ より、 $1 > \alpha_i p(n_{t-1}) + (1 - \alpha_i)q(n_{t-1}) > 0$ である。協同作業の費用が拒否する費用より小さい ($c < c_r$) と必ず(1')が成立してしまうので、ここでは $c \geq c_r$ であるときのみを考える。この二式から、

$$(c + c_p)/b_L \geq (c - c_r)/b_i$$

であるとき、リーダーが提案すると((2')式が成立すると)、(1')式が常に成り立つので、元々の提案の成功確率がある程度高いと思われる ($\alpha_i p(n_{t-1}) + (1 - \alpha_i)q(n_{t-1})$ の値があまり低くなく)、メンバーは必ず説得されることがわかる。逆に、

$$(c + c_p)/b_L < (c - c_r)/b_i$$

であれば、リーダーが提案しても((2')式が成立しても)、(1')式が成り立たない場合があり、したがって、メンバーはリーダーの説得を拒絶する場合もある。この式を整理すると、

$$(b_L - b_i)(c - c_r) - b_i(c_p + c_r) > 0 \quad (3)$$

となり、 c が c_r に比べてかなり大きいか、 b_L が b_i に比べてかなり大きい、あるいはその両方の場合、言い換えると、提案が比較的断りや

すく、メンバーの便益がリーダーの便益に比べてかなり小さいときに、(2')式が成立しても(1')式が成立しない場合があり、そのときメンバーはリーダーの説得を拒絶する。もちろんその他に、 $\alpha_i p(n_{t-1}) + (1 - \alpha_i)q(n_{t-1})$ の値がかなり低い、即ちその提案があまりよいものとは思えない (α_i の値が低い) か、他のメンバーが賛成しように無い (n_{t-1} が小さく、その結果 $p(n_{t-1})$, $q(n_{t-1})$ が小さい)、あるいはその両方であるために、リーダーが提案しないか、リーダーは提案するがメンバーが拒絶するという可能性もある。

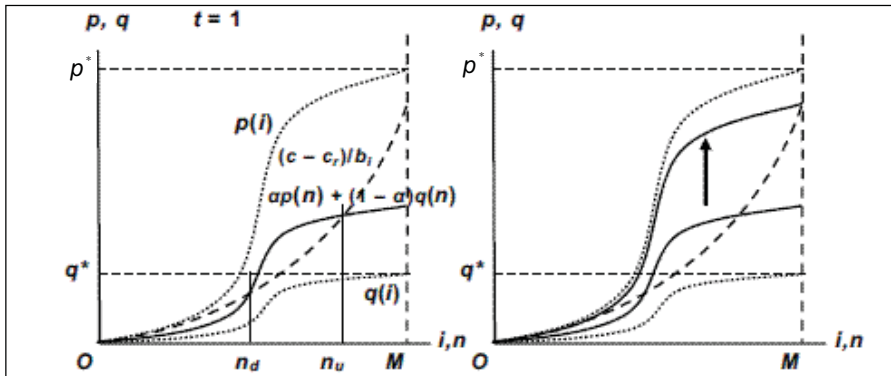
提案が比較的断り難いか、メンバーの便益がリーダーの便益に比べてそれほど小さくない場合、メンバーがリーダーの説得を受け入れ易いことは直感的にも納得できよう。より厳密には、(3)式が成り立たないか、成り立っても(1'), (2')の両式が成り立つとき、メンバーがリーダーの説得を受け入れる。

3-2 スーパー・ゲーム

次に、このステージ・ゲームが繰り返されるスーパー・ゲームを考える。所与の α_i , b_L に対して(2')式が成立する最低の n を $n_i^L(\alpha_i, b_L)$ 、リーダーの提案を最も高く評価する人を、メンバー $i = 1$ とし、 i は評価の高い順、即ち b_i の大きい順に並んでいるものとする。簡単化のために、 $i \neq j$ のとき $b_i \neq b_j$ と仮定する。しかし、 b_i の値が大きくて(1')式の右辺が小さくても、提案を受け入れるメンバーの人数が少なければ(1')式の左辺も小さいので、必ずしもリーダーの提案を最も高く評価する人の方が、提案を受け入れやすいわけではない。むしろ、上述のような S 字型の $p(n)$ と $q(n)$ を想定すると、図2のように b_i の値が大きくても n_{t-1} が小さければ、

$$\alpha_i p(n_{t-1}) + (1 - \alpha_i)q(n_{t-1}) < (c - c_r)/b_i$$

となり(1')式が成立しない。従って、第1期に図2で

図2 $\alpha p(n) + (1 - \alpha)q(n)$ & $(c - c_r)/b_i$ (一例)

$$\alpha_t p(n_{t-1}) + (1 - \alpha_t)q(n_{t-1}) = (c - c_r)/b_i$$

となる二つの n_t の点を n_d, n_u , ($n_d < n_u$) とすると, $\alpha_t M < n_d$ のときは第1期にリーダーの提案を受容する人はいないので, $n_1 = 0$ となる。しかも, $n_t^L(\alpha_t, b_L) > 0$ なので, この案が2度と提案されることはない。また, $\alpha_t M \geq n_u$ のときは図2(左)からわかるように n_u 以上の人数が提案を受容することはないので, $n_1 = n_u$ となり, $n_u \geq n_t^L(\alpha_t, b_L)$ ならば, リーダーがこの協同作業を提案・説得し, 成功すれば α_t が上昇して n_u も上昇し, 受容者も増加する。

その後も, 図2で $\alpha_t p(n_{t-1}) + (1 - \alpha_t)q(n_{t-1})$ が $(c - c_r)/b_i$ を上回っているメンバー i が n_d 人以上で, かつ(2')式が成立する最低の人数以上の場合, 即ち $n_t \geq n_t^L(\alpha_t, b_L)$ の場合にリーダーがこの協同作業を提案し, それがい提案であれば, $p(n_t)$ の確率で成功する。逆に $n_t < n_t^L(\alpha_t, b_L)$ の場合は, リーダーは提案しない。リーダーが提案しなければ協同作業は始まらないので, いったん $n_t < n_t^L(\alpha_t, b_L)$ となると, (2')式が成立するチャンスもなくなり, そこでゲームが終了する。

基本的なバイジアン・ゲームと同様, t 期に成功すれば, この提案が「善い」提案だという信念は高まる ($\alpha_{t+1} > \alpha_t$) し, 失敗すれば

「善い」提案だという信念は低下する ($\alpha_{t+1} < \alpha_t$)。具体的な信念の改定は, 成功して共同体利益が実現したときには,

$$\begin{aligned} \alpha_{t+1} &= \alpha_t p(n_{t-1}) / \{\alpha_t p(n_{t-1}) + (1 - \alpha_t)q(n_{t-1})\} \\ &= \alpha_t [p(n_{t-1}) / \{\alpha_t p(n_{t-1}) + (1 - \alpha_t)q(n_{t-1})\}] \\ &> \alpha_t \end{aligned}$$

である。逆に失敗して共同体利益が実現しなければ,

$$\begin{aligned} \alpha_{t+1} &= \alpha_t (1 - p(n_{t-1})) / \{\alpha_t (1 - p(n_{t-1})) \\ &\quad + (1 - \alpha_t)(1 - q(n_{t-1}))\} \\ &= \alpha_t [(1 - p(n_{t-1})) / \{\alpha_t (1 - p(n_{t-1})) \\ &\quad + (1 - \alpha_t)(1 - q(n_{t-1}))\}] \\ &< \alpha_t \end{aligned}$$

に低下する。 α_{t+1} が α_t より上昇すれば, (1')式の左辺が増加し, 図2の右の図のように, (1')式が成立する人数が増加し, さらに成功確率が高まるという好循環にもなりうる。もちろん, 逆に α_{t+1} が α_t より低下すれば, 図2の右の図と逆方向の変化が起こり, (1')式が成立する人数が減少する。

以上のことから, リーダーが提案をするだけの期待利得を見込める人数が, 前の期にリーダーの提案を受容する限り, この「善い」協同作業は繰り返し提案され, 繰り返す回数が増えるに従って「善い」提案だという信念 (α_t) は1に近づく。図2に即して言えば, $\alpha p + (1 - \alpha)q$

が p に近づく。しかしそれは、この協同作業の提案を受容する場合の期待利得が、拒否する場合より高いというだけで、行動自体に価値を見出す「倫理的規範」とは言えないであろう。即ち、その個人の価値観の中に内面化されて初めて、「倫理的規範」と呼べる。前節の平等主義的共同体の例で言えば、共同体内部で肉を平等に分配することは変動削減理論で説明できる（長期的な自己利益に適う）互惠的行動ではあるが、利他的（倫理的）行動とは言えない。そこで次項では、「客観的」に「善い」提案であることが明らかになり（ α_i が 1 に近づく）、それが内面化されて、主観的倫理規範となる過程のモデル化を試みる。

3-3 「倫理的」行動の内面化

前項の考察から、 t 期までのリーダーの戦略は、それまでのどこかの期で提案を止めるか、ずっと続けているかのどちらかである。それに対して、メンバーの戦略は、いったん提案を拒否しても、リーダーが提案を続けて、他のメンバーが受容し、かつ成功が続いて、自分が受け入れられる期待利益の水準に達してから受容することも可能である。したがって、様々な受容と拒否のパターンがあり得る。

町野（2005）の蟻ゲームでは、「自然」による成功と失敗の不確実性が無いので、徐々に忘却するというメカニズムを組み入れると、親が教育的行動を取り続けた場合、ある時点以降、親が教育的かどうかという信念に拘わらず、子供が倫理的行動を取り続けるような均衡としての倫理の内面化をモデル化できた。本稿のゲームでは不確実性が存在するため、「失敗」と「善い」提案であるという信念（ α_i ）が低下する。しかし、 α_i が一定期間以上不等式(1')を満たす水準を維持すれば、限定的記憶を持つプレーヤーは「倫理的」行動をとり続けるようになり、内面化が達成される。

ステージ・ゲームを長期的に繰り返すと、長期的には「善い」提案であれば成功頻度は

$p(n_i)$ に近づき、同時に α_i は 1 に近づく。「悪い」提案であれば成功頻度は $q(n_i)$ に近づき α_i は 0 に近づく。具体的には、仮に n_i を所与とすると、もしよい提案であれば、 t 期までの成功回数を s とすると、

$$\begin{aligned}\alpha_i &= \alpha_i p^s (1-p)^{t-s} / \{ \alpha_i p^s (1-p)^{t-s} + (1-\alpha_i) q^s (1-q)^{t-s} \} \\ &= 1 / [1 + (1-\alpha_i) (q/p)^s \{ (1-q)/(1-p) \}^{t-s}]\end{aligned}$$

となる。成功回数 s は t を無限大に近づけると pt になるので、

$$\begin{aligned}\alpha_i &= 1 / [1 + (1-\alpha_i) (q/p)^{pt} \{ (1-q)/(1-p) \}^{t-pt}] \\ &= 1 / \{ 1 + (1-\alpha_i) [(q/p)^p \{ (1-q)/(1-p) \}^{1-p}]^t \}\end{aligned}$$

と書ける。

$$(q/p)^p \{ (1-q)/(1-p) \}^{1-p}$$

は、 $p = q = 1/2$ のとき最大値 1 となるが、 $p > q$ なので、

$$(q/p)^p \{ (1-q)/(1-p) \}^{1-p} < 1$$

であり、

$$\lim_{t \rightarrow \infty} [(q/p)^p \{ (1-q)/(1-p) \}^{1-p}]^t = 0.$$

したがって、

$$\alpha_i = 1$$

となる。

このように、その提案が「善い」提案であれば、ステージ・ゲームを長期的に繰り返すと、そのことが確信される。逆にその提案が「悪い」提案であれば、やはりステージ・ゲームを長期的に繰り返すと、そのことが明らかになる。何が「善い」ことであるのかを再確認すると、自己犠牲的行為（費用 c を払うこと）が十分な共同体利益を生むことである。それが実現するには、多くのメンバーが繰り返し提案を受け入れなければならない。その上で成功が実績として証明され続けられれば、リーダーの提案が望ましいものであるという信念が社会的に形成されてい

く。さらに、この行動が慣習化すれば、人々は一々その行動を説得されなくても、それは「当然すべき行動」だと感じ、倫理規範となる。このように規範が内面化するのには、メンバーが提案拒否という選択肢を忘れており、期待利得の計算という因果関係も忘れる「限定合理的」なプレーヤだからである¹²⁾。

慣習化の過程は以下のように進行すると考える。まず、「善い」提案の場合は長期的には α_i が1に近づくので、(1')式が

$$p(n_{t-1}) > (c - c_r)/b_i \quad (1'')$$

となる。このときの n_{t-1} は明らかに $n^L_i(\alpha_i, b_L)$ より大きい値であるから（そうでなければ、長期的に続けられていない）提案され、この行為は多くのメンバーにとって、受容するか拒絶するかを選択ではなく、ある状況において必然的に取る行為となっている¹³⁾。これは、例えば Young(1993)のモデルのように、限定的記憶によって、過去の限られた期間以前に選択されなかった戦略は忘れてしまうと考えるべきよい。

このように、善い行為であるという確信($\alpha_i \approx 1$)と慣習化によって、この行為は「倫理的」「規範」に転化することになる。即ち、この行為を行わないという選択肢を含んだゲームを考えると、形式的には図3の第1段階とほぼ同じだが、利得構造はまったく異なる（例え

ば）図4のようなゲームに転化する。規範からの逸脱には負の効用が伴い、また他人の行動が規範から逸脱している場合は、それは社会に損失を与える（「悪い」）行為だと判断される¹⁴⁾。この場合、逸脱行為は負の効用(-g)を伴い、逸脱しない他のメンバーの費用は、リーダーと同様、逸脱を罰するコストも含む費用(- c_p)に変化する。即ち、各プレーヤの利得構造(=効用)が変化する。そのステージ・ゲームは以下のとおり。

ただし、共同体リーダーの利得と共同体メンバーの利得は図4のとおりで、各変数は正の値を持つ。

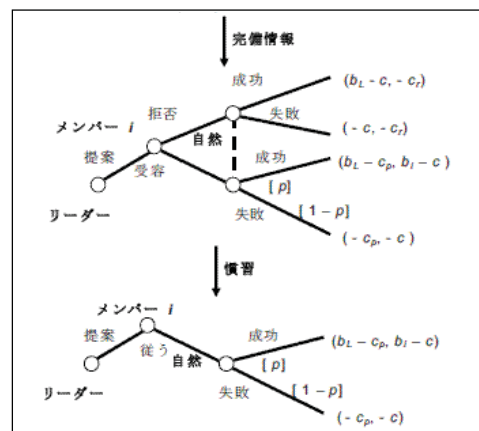


図3 完備情報化と慣習化

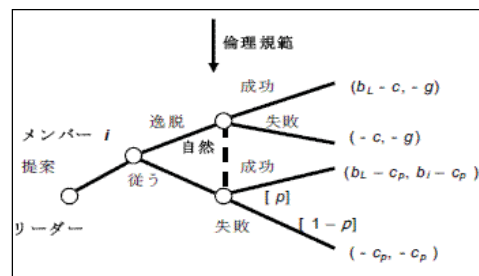


図4 倫理規範化

プレーヤ：{共同体リーダー，共同体メンバー}

共同体リーダーの戦略：{提案，提案せず}

共同体メンバーの戦略：{(従う(制裁))，逸脱}

12) ただし、人間の記憶能力の限定性を前提とすると、慣習化され得る行為の因果関係を一々考えないのは限られた資源である記憶能力の有効利用という進化的に合理的な反応だとも言える。

13) 図2のようなケースでは、 α_i が1に近づくとき全ての i に対して(1'')式が成り立つがそうでない場合は n_u の数のメンバーにとって慣習化が進むことになる。

14) Knight (1992 第3章) はゲームの均衡が慣習化して「制度」となると、そこからの逸脱は他のプレーヤの期待利得の実現を妨げるので、彼らに制裁するインセンティブを与えると指摘している。この論理のゲーム・モデル化の例として脚注6の Sugden の期待予想均衡がある。

このような過程を経てこの行為が倫理的行為だという価値観を持った人の数が、一定の数まで増えれば、その倫理が社会的に維持できる好循環を生み出せることは、たとえば Bowls and Gintis (2004) によって示されている。

本節の最後に、町野 (2005) で試みた集団内の多世代にわたる倫理規範維持過程の簡単なモデル化について、その考え方を要約しておく。この過程は社会化、または家庭での躾のモデル化と考えればよい。社会の既存の構成メンバーが、新たなメンバーにその社会のルールを教えることで、その社会あるいは過程の倫理規範が、次の世代へ伝えられていく過程である。標準的な不完備情報ゲームであれば、子供は親のタイプは分からないが、タイプが分かれば利得はわかる。しかし躾の場では、子供には養育者の利得と戦略集合が (完全には) わからない。したがって通常のモデルは使えない。そこで限定合理性を前提とした学習ゲーム、例えば最適反応のモデルを使えば、叱られる不効用がその行為から得られる効用を上回るときに、子供が利己的な動機から叱られないようにするにはどうすればいいかを徐々に学んでいくメカニズムを表現できる可能性はある。それでも、そのゲームにおける子供は、倫理的に行動すること自体から効用を得るという意味で倫理的になることはない。

子供が倫理的になる状況をゲーム・モデルで表現するには、当初負の効用をもたらす自己犠牲を伴うような倫理的行動が正になる過程を説明できなくてはならない。そこで、それに似た過程である条件反射の類推で考え、子供がある正の効用をもたらす反倫理的行動を取ったときに、常にその正の効用を打ち消す以上の制裁を与える (叱る) ことを繰り返すモデルを考案した。これは、(反倫理的行動、叱る) という戦略プロファイルと負の効用という結果が繰り返されるたびに、その戦略プロファイルと効用の組合せの記憶が堆積し、人間の限定記憶 (あるいは記憶の節約、あるいは防衛本能) の結果、

マイナスの利得が相手の意図的行動の結果であるという部分の記憶が薄れ、反倫理的行動をとると罪悪感を抱くという、より効率的な反応メカニズムとしての倫理感という感情が形成される、というものである。養育者のように相手側に既に倫理感があることを仮定すると、相手側の反応も半ば条件反射的であると考えられる。ただし、この倫理感形成過程は、制裁の繰り返しによって実現するとは限らず、信頼する他者の模倣、説得と納得という過程を経て実現するかもしれない。

いずれにしても、自分の行動に対する善悪の区別をつけ、善い行動=褒められるべき行動、悪い行動=制裁されるべき行動、という倫理感という感情にまで成熟するには、「自分の行為に対する他者からの評価及びそれに伴う正もしくは負の心理的刺激 (賞賛や叱責) の記憶」が、自分の行為と心理的刺激を直結させるまで繰り返される必要がある。換言すれば、自分の行為と他者の評価の因果関係 (おそらく無意識のうちに) を納得して、自分の行為に対して他者から受ける心理的刺激を「当然のもの」と考えるようになることが必要である。

社会学で「内面化」と呼ばれる過程がこれにあたる。「内面化」の社会学的説明としては、自分が全面的に依存している養育者との葛藤を避けるために、養育者の価値観を自分のものとする過程 (作田, 1972) というものがある。「内面化」されるには、他者による「社会的」評価と反応の記憶の蓄積がある臨界値を超えなければならないのであろう。もちろん一回ごとの刺激の強さ、次の刺激までの間隔の長さ、類似行為が内面化されているかどうかによって、臨界値に達するのに必要な繰り返しの回数は異なるはずである。

多様な学習の概念があり、それらの分類基準も様々であるが、認知的 (意識的) かどうか、習慣化されたものかどうかで大きく分けると、子供の社会化は、当初は非認知的な強化学習であり、養育者に怒られたり褒められたりしなが

ら、徐々に倫理的に振舞うことによる効用が高くなっていくという、利得が変化する効用関数をモデル化しなければならない。その後認知的能力が高くなれば、「倫理的」行動のパターンを自分なりに身に付ける学習ルーティンを獲得し、さらにそのパターンの現実適合性によっては学習ルーティンを修正することもできるようになるが、そのような高度な学習のモデル化は今後の課題である。

4. 考察：モデルの解釈と倫理的規範形成の現代的課題

本稿では、平等主義的な倫理規範の進化的起源に関する動物行動学と人類学の研究を踏まえ、町野（2003）の倫理規範形成のゲーム・モデルを発展させた倫理規範形成過程のモデル化の枠組みを提示した。具体的には、社会的には望ましいが個人的には負担のかかる行動が、倫理規範としてその社会に定着する過程を、リーダーのイニシアティブ、社会的有用性のメンバーからの認知、メンバーへの内面化、という3段階のステップでモデル化できることを示した。

倫理的行動を社会的には望ましいが個人には負担のかかる行為と定義したので、利己的な個人が初めから進んでその行為を行うことはない。しかし、ゲームのプレーヤがすでに社会ないし集団で生活していることを前提にすると、その社会の存続、繁栄はメンバー全員の利益になり得る。中でもその共同体の有力者は、通常、共同体全体の利益の拡大から得られる便益が他のメンバーより大きい。従って彼らは、費用をかけてでも他のメンバーを「説得する」インセンティブを持ちやすい。こうした有力者主導の倫理規範であれば、寛容な支配者が主導する階級的な倫理規範が形成されるかもしれない。しかし本稿で見た動物行動学と人類学の研究のように、有力者の独裁的支配の可能性を抑えるために共同体のメンバーによる平等主義的な政治的連携が形成され、彼らの支持する（平等主義的）

規則が規範として広がっていくというケースも存在し得る。この政治的連携は必ず成功するわけではなく、失敗する可能性もあるので、連携のイニシアティブをとるリーダーの意思決定が重要になる。反支配者連携のリーダーの費用・便益の内容は寛容なリーダーのものとは異なるが、ゲームの構造は（第3節で示したように）同じである。

そのような、不確実でしかも個々のメンバーやリーダーの思惑や実力がわからないという不完備情報の状況をベイジアン・ゲームとしてモデル化すると、連携した反抗がうまくいけば、メンバーたちもそれが有益であることに確信を持てるようになってくる。さらに、限定的記憶を持つプレーヤたちが、肉の平等な分配のような「倫理的」行動を繰り返すうちに、それ以外の選択肢があることを忘れ、元々の効用の因果関係も忘れて（即ち効用関数が変化し）この行動を行う「べき」行動だと感じるようになり逸脱すると罪悪感を持つようになる。以上のモデルは、町野（2003）のモデルと共通の発想のもとにあるため、そこでの以下のような考察（要約）は、本稿のモデルとも整合的である。

- * 社会の巨大化、複雑化が、人々の所属する共同体と公共サービスを提供する共同体のズレを生じさせ、公共的義務に対する共同体からの心理的圧力を著しく退化させ、モラルハザードやフリーライダー（あるいは社会的ジレンマ）の問題を拡大させた。もちろん現代の個人といえども、教育や社会化は、依然として家族や学校など狭い仲間社会で行われる。しかし、このレベルでの社会化、即ち倫理的規範の修正・維持がうまく機能しなければ、規範維持の循環が断ち切られる。
- * 最近の日本における倫理的規範喪失への懸念の高まりは、倫理的規範形成・維持のメカニズムに問題が生じているという認識が社会の中で共有されつつあることを意味し

ている。全く倫理的規範の無い時点からのモデル化を考えると、公理的に倫理性を考えるべきだという主張の方が、受け入れ易いかもしれない¹⁵⁾。しかし、潜在的な新しい倫理的規範は、現在の共同体・社会における実践的な取組みの中で発見され提案されていくはずである。個々の共同体が新たな（あるいは修正された）共同体倫理（ルール）を発見し浸透させ、それが社会全体に広がっていくというメカニズムがモデルから演繹される倫理的規範形成過程の現実的姿であろう¹⁶⁾。

本稿の第2節で、後者の指摘は人類学などでの観察から帰納されることを示した。また、ある行為が倫理的規範になる過程を改めて考えると、次のような点も新たに指摘できる。「倫理的規範喪失」は、単に古い規範へ回帰することによって解決できるものではない。客観的に現在の社会で必要とされる公共的協同行為の便益と、社会の個々のメンバーにとってのコストを計算し、さらにメンバーを説得して図2の n_d を超える人数を継続的に確保できるかどうかを考えなければならない。しかし上述の第一の指摘のように個人にとっての倫理的行動の（少なくとも短期的直接的な）便益は見えにくく、倫理的行動を取らないことの費用は（監視や制裁の難しさから）極めて低い。したがって、一定数のメンバーを説得するのは容易なことではない。氷河期の肉の分配のようにぎりぎりの生存条件をめぐる分かり易い問題ではないので、氷河期の人類が平等主義に到達するために必要としていた知性やコミュニケーション能力の進化についても、現在の問題を個々の人間の能力の

進化のスピードでは解決できそうにない。社会のシステムとしての知性やコミュニケーション能力の一層の意識的向上が必要なのではないだろうか。

付記

本稿は文部科学省補助金（萌芽研究「社会規範のネットワーク - ゲーム・モデル」）及びグローバルCOE『多元分散型統御を目指す新世代法政策学』に基づく研究成果の一部である。

参考文献

- 1) Boehm, Christopher (2004) "What Makes Humans Economically Distinctive? A Three-Species Evolutionary Comparison and Historical Analysis." *Journal of Bioeconomics*, 6 2, pp.109-35.
- 2) Boehm, Christopher (1999) *Hierarchy in the forest: the evolution of egalitarian behavior*, Harvard University Press.
- 3) Binmore, Ken (1994, 98) *Game Theory and Social Contracts I, II*, The MIT Press.
- 4) Bowls, Samuel and Gintis, Herbert (2004) "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations," *Theoretical Population Biology*, 65(1), pp.17-28.
- 5) de Waal, Frans (1996) *Good Natured* - The Origins of Right and Wrong in Humans and Other Animals* -, Harvard University Press. (西田利貞・藤井留美訳 (1998) 『利己的なサル、他人を思いやるサル』草思社)
- 6) 長谷川寿一、長谷川真理子 (2000) 『進化と行動』東京大学出版会
- 7) Hawkes, Kristen and Bird, Rebecca B. (2002) "Showing Off, Handicap Signaling, and the Evolution of Men's Work" *Evolutionary Anthropology*, pp.58-67.

15) この初期時点は Rawls (1971) の veil of ignorance の original position と解釈できる。

16) 同様の議論として、Binmore (1994, 98) は、上記 Rawls の original position の代わりに status quo を規範形成の出発点として考えることを提唱している。

- 8) Knight, Jack (1992) *Institutions and Social Conflict*, Cambridge University Press.
- 9) 町野和夫 (2005) 「倫理的規範の形成と内面化過程：不完備情報・進化ゲーム・モデル」北海道大学 Discussion Paper Series B-52.
- 10) 町野和夫 (2003) 「倫理的規範形成のゲーム論的分析」『経済学研究』(北海道大学) 53 巻 3 号 283-297 頁.
- 11) Rawls, John (1971) *A Theory of Justice*, Harvard University Press. (矢島鈞次 監訳 (1979) 『正義論』紀伊国屋書店)
- 12) Sugden, Robert (2004) *The Economics of Rights, Co-operation and Welfare* 2nd ed., Palgrave Macmillan. (ロバート・ザグデン (2008) 『慣習と秩序の経済学：進化ゲーム理論アプローチ』日本評論社).
- 13) 作田啓一 (1972) 『価値の社会学』岩波書店.
- 14) 内田亮子 (2007) 『人類はどのように進化したか：生物人類学の現在』勁草書房.
- 15) 内井惣七 (1996) 『進化論と倫理』世界思想社.
- 16) Young, Peyton, H. (1993) "The Evolution of Convention," *Econometrica*, Vol. 61, pp.57-84.