規範の受諾

――説諭としてのゲーム理論――

長久領壱

1. はじめに

ゲーム理論を用いて，正義や道徳の規範理論を構成しようとする試みは，ハーサニ（Harsanyi［1955］）の独創的な研究を嚆矢として，近年特に盛んになってきている。これらの研究のテーマは人びとの合意による社会秩序形成の可能性を探求することである。すなわち「ルールや制度を，私的な利益や目的を追求する個人間の合意のみで，どの程度まで説明できるか。人びとの間での協力関係の形成を，共同体理念や利他的道徳観に訴えることなく，各人の私的な選好の充足のみによって規範的に正当化することは可能であるか。」（小林［1991］）がここでの主題なのである。

本論文の目的も，基本的にはこれらの研究と同一の系譜に属するのであるが，われわれの意図はハーサニ，ロールズ（Rawls［1971］），あるいはゴティエ（Gauthier［1986］）のように規範的正義の１つのモデルを提出しようという点にあるのではない。規範的正義論の研究へゲーム理論を適用する際に，理論家が嵌まりやすい「ある陥穽」を指摘すること，これが本稿での主題である。わかりやすく言えば，「ゲーム理論使用上の注意事項」を述べることなのである。では，その陥穽とは何か，詳しくは次節に譲ることにするが，考察の出発点は盛山［1995］の囚人のジレンマ繰返しゲームに対する批判である。以下，簡潔に盛山の議論を素描しよう。

ゲーム理論家の多くは「一回限りでは協調が合理的にはならないゲームでも繰り返し行われれば，協調を生むのではないか。それが社会規範の定着を説明できるのではないか」（川浜［1999］223頁）という直観をもとに，社会規範の説明原理として繰返しゲームの研究に期待をかけてきた。しかし盛山によるとこの直観（厳密には先の括弧内の後半部）は誤りであるという。ここで，社会規範をどう定義するかが問題になるが，ここでは何らかの normative な観点から人の行為を拘束する規則ないしルールとして機能すると定義してみよう。すると，繰返しゲームでプレイヤー達がある時点以降，協調解を選び続けるという行動パターンが観察されたとしても，そのことから「彼らは協調解を選ばなければならない」という規範が生成されたとはいえない。例えば，「地下鉄のホームで，乗客は乗り口の左側で列をなして待っている」という行為が繰り返し観察されたからといって，「彼らはそうすべきである」という規範が成立したわけではない。「行動パターンの定着＝規範の成立」とする図式は実は誤りなのである。これが盛山の批判の要諦である。

本稿の目的は，盛山の批判を検討し，正義や規範に関する理論的研究においてゲーム理論をどう利用すべきか，を考察することにある。われわれが得た結論は，以下のとおりである。確かに盛山のいうように，ゲーム理論は「規範の生成」（人はいかにして規範を受け入れるのか）――規範の成立に関する経緯と過程――を説明する経験科学上の説明原理(1)としては失敗している（第２節）。しかし，規範の受諾（人はなぜ規範を受け入れるか）――規範の受諾に関する理由と根拠――を正当化するための道徳哲学・倫理学上のモデルとしてなら機能しうる（第３節），と。比喩的にいうならば，ゲーム理論は規範を否定する人びとに対

＊ 関西大学経済学部教授

し，もし規範を受け入れなかったらどのような悲惨な運命になるかを示し，規範の受諾を迫る「説諭」として機能するのである。囚人のジレンマ繰返しゲームは，この説諭のために用いられる寓話の１つである。そして，この「説諭としてのゲーム理論」は，ホッブズ（Hobbus［1651］）やロールズ Rawls［1971］などゲーム論的な考えを応用して，演繹的に道徳原理を導きだそうとした人びとが実際に採った戦略なのである（第３，４節）。

いくつかの留意点がある。これらの諸点はいずれも，本稿を最後まで辛抱強く読んでいただければ氷解する類のものだが，予防線を張る意味で前もって簡潔に述べておこう。第１は，先にゲーム理論家達の囚人のジレンマ繰返しゲームに関する解釈に関してである。盛山は彼らの１部は「規範の生成」を説明するモデルとして解釈している（節がある），と批判している(2)。本稿では，盛山のこの批判は文献的に正しいかといった学説史的な主題にはタッチしない。ゲーム理論家の誰が実際にそのような解釈をとっているのか，はさしあたり我々の関心事ではない。場合によっては本稿を，「もし仮にそのような解釈をとるとしたら……」と想定した場合での議論である，と解釈されても結構である。ただ，私見を言わせていただければ，「規範の生成」のモデルとして解釈したゲーム理論家はごく少数ではないか，とは思う。

読者諸氏の中には，わけてもゲーム理論家は，私と同ような見解で持って，盛山の批判を一蹴する方もおられるかもしれない。「我々はそのような解釈をしていない。故に批判は的外れである」と。しかし，それでは盛山のせっかくの教訓的な議論を台無しにしてしまうことになるだろう。仮に，このような反批判が文献的に正しいとしても，なお盛山の批判は傾聴に値するのだ。私はそう判断している。本稿の以下の議論を読めば，納得していただけると思う。

第２の留意点はコンヴェンション（Convention）に関してである。コンヴェンションも社会規範と同よう，人の行為を律する規則ではあるが，その規則は便宜的に決まっており，なんら規範的要素をも含んでいない。コンヴェンションの例としてはテーブルマナーや商慣行など，広く社会慣習とでもいわれているものがそれに該当する。第２の留意点は，繰返し囚人のジレンマゲームはコンヴェンションの生成を説明するモデルとしてなら，適格である，ということである（第２節）。実際，多くのゲーム理論家はこの解釈をとっている。

第３の留意点は，盛山［1995］の議論との関連である。「説諭としてのゲーム理論」という我々の見解は盛山の主張とは矛盾しない。彼は「ゲーム論的モデルには制度現象を取り扱う上での根本的限界がある……」（87 頁）とは述べているが，ではその限界内でゲーム理論は制度現象をどのように取り扱えるか，に関しては答えを控えている。本稿はその答えを出したといえる。規範をあくまで先に定義したように，その言語表現が「……すべし」という命令法の形をとり，何らかの行為の規範的な妥当性に関わると仮定した場合での話であるが。それでもやはりわれわれの結論は盛山の議論から出てくる論理的帰結の１つである。この点に関しては最終節（第５節）で詳しく論じる。

本稿の目的はゲーム理論の社会規範研究に与える意義を正しく位置づけることであるが，そこだけにはとどまらない。これが第４の留意点である。問題は更に一般化できるのである。われわれが主張したいことは，規範経済学が取り扱う主題の多くは実証経済学の方法のみでは十分明らかにすることはできない，ということである。人がある規範や正義の原則に従っているからといって，あるいはその原則が一定の条件の下で支持されやすいという理由で持って，その規範なり正義なりが「よい」とはいえない。この点を明らかにするため，先の地下鉄の例を振り返ってみよう。すぐに分かるように，この例では「……である」という事実命題から「……であるべきである」という価値命題を導く誤りを犯している。ヒューム（Hume［1739］）の原則，事実命題から価値命題を導くことはできないという原則，に明らかに違反しているからである。われわれは，ヒュームのみならずムア（Moore［1903］）など科学哲学上の議論を参考にしながら，第５節で詳しく論じることにする。

社会規範を説明できる可能性のある議論として，繰返しゲームの他に，進化ゲームやシグナリングのアプローチがある。これらの理論を用いても本稿の結論は同じである。その意味で本稿の結論は

普遍的妥当性を持つ。付録では進化ゲームを例にとって，このことを論証することにしたい。

論文の構成について簡潔に触れておく。第２節では囚人のジレンマ繰返しゲームとそれに対する盛山の批判を検討する。第３節が核心部で，本稿でのアイディア「説論としてのゲーム理論」を論じる。第４節は実際に道徳哲学ではゲーム理論の説論としての使用例を１つあげる。ロールズの原初状態と反照的均衡の概念がそれである。第５節は結論である。付録では各セミナーで出された疑問や批判に対する回答である。

筆者は，規範経済学に強い関心を抱く者であるが，この分野はその重要性にもかかわらず，世界的に見ても決して振るっているとはいえない。その理由は規範経済学が価値の問題を扱わざるをえず，経験科学としては成り立ち難い点に求められよう。その点に関していえば，本稿は，規範経済学研究の１つの方向性を示し得たといえるかもしれない。本稿を読んで，多くの人が規範経済学の面白さをわかっていただけたら，筆者にとってこれに優る喜びはない。

2. ゲーム理論と規範の生成

人びとの間で協力関係を築き上げるのは難しい主題である。たとえ人びとが協力した方がそうでない場合よりも多くの成果をあげうるとは理解はしていても，である。ゲーム理論でつとに有名な「囚人のジレンマ」は，この事実を印象的に示している。

「囚人のジレンマ」は２人２戦略の標準形ゲームである。 オリジナルなストーリーは以下のとおりである(3)。２人の容疑者が別々に留置所で拘禁されて互いに連絡不可能な状態でいる。検事は，２人がある犯罪の共犯者であることを確信しているのだが，裁判で有罪を立証するだけの十分な証拠がない。そこで検事は２人に次のような取り引きをもちかける。「自白するか，しないか，２つの選択肢がある。もしも２人とも自白しなければ，窃盗や武器の不法所持など適当な罪をでっち上げてやるが，２人共に一年の懲役で済む。もしも２人ともに自白しなければ，２人共に八年の懲役だ。しかしもし一方が自白し，他方が自白しない場合は，自白した方は捜査協力を理由に寛大な処置（懲役三ヶ月）がとられるが，自白しない方は極刑（懲役十年）になるだろう。」

容疑者達が，自白しないという約束で共闘すれば，共に懲役八年で済む。この状態は，２人共に自白してしまった場合よりもパレート優位にあり，２人にとってはより望ましい状態である。しかし，２人は別々に拘禁されており，コミュニケーションがとれず，約束を交わすことはできない。仮に２人がコミュニケーションのとれる状態にあっても事態は変わらない。自白しないという約束が履行される保証はどこにもないからである。２人は互いに相手が自白しないという確信の下に，自らは約束を破る誘惑に駆られるであろう。またこのことから２人は次のように相手の心理を読むだろう。「自分がこのような誘惑に駆られるということは相手も同じ誘惑に駆られているに違いない。」と。２人は互いに相手が約束を破るのではないか，という疑念を抱く。こうして容疑者達は互いに相手の出方を読み合う，という戦略的状況に直面せざるをえない。容疑者達は，まず「相手がどう出るか」を読まなければならない。しかし考察はここで終わらない。「自分が『相手がどう出るか』をどう読んでいるか」もまた相手は読んでいるわけだから，相手が「自分が『相手がどう出るか』をどう読んでいるか」をどう読んでいるか，も読まなければならない。読みが自分の読みの中に組み込まれ，更にそれが相手の読みに組み込まれ，更にそれも自分の読みに組み込まれる……。こうして推論のプロセスは原理的には終わりがないのであるが，今の場合は，〈自白する，自白する〉という読みに至った後，２人の選択は変化しなくなる。２人は推論の結果，自白を選ぶ。そしてここがナッシュ均衡にもなっている(4)。しかし，この帰結は，先に述べたように〈自白しない，自白しない〉よりもパレート劣位にある。かくして，２人の間に協調関係は成立しないのである。この「囚人のジレンマ」は，合理的個人の間での秩序形成に潜む，最も基本的な病理を描き出した例としてよく知られている。

しかし，この囚人のジレンマのような，１回限りでは協力が合理的にはならないゲームでも，ゲームが繰返される状況を想定すれば，協力関係を

生むことは可能になる。これが繰返しゲームとして知られている議論である。1回限りのゲームでは，協調するよりも裏切った方が得にはなるとしよう。しかしゲームが繰返し行われる状況では，裏切ったプレイヤーは次の期のゲームで報復される。長期的な利益を考えれば，相手に協調する方が合理的であることをプレイヤー達は悟り，協調の遵守が得られるという訳である。繰返しゲームで帰結する利得ベクトルが個人合理的であれば，そしてプレイヤー達が将来に獲得すると予想する利得に十分関心を抱いている（割り引き因子の値が十分大きい）とき，それをナッシュ均衡とする戦略が必ず存在する（フォーク定理（Furdenberg and Maskin［1986］）。囚人のジレンマゲームでは〈告白しない，告白しない〉は個人合理的であるから，2人が将来利得にそれなりの重きを置けば，ナッシュ均衡として実現できることになる。つまりこの場合，〈告白しない，告白しない〉からの逸脱は，当の本人の利益にならないのである。この意味で協調は2人の間で遵守され続けるのである。

ただ，フォーク定理だけからでは2人が〈自白しない，自白しない〉を確実に選ぶ，という保証はない。〈自白する，自白する〉も同じく個人合理的であるからで，このどちらが帰結するかはフォーク定理からはわからない。ここで，アクセルロッド（Axelrod［1984］）のコンピューターシュミレーションによる実験が参考になるだろう。彼は，囚人のジレンマ繰返しゲームで，戦略のプログラムを幾つか作成し，プログラム同士を対戦させてどのプログラムが最も高い利得を得るかを計算した。戦略のプログラムは，繰返しゲームの各ステージゲームでの戦略を指定する。例えば，全面的協調プログラムは，毎回「告白しない」を選択するようなプログラムである。どのプログラムと対戦しても比較的良好なスコアを残したのが，しっぺ返し戦略と呼ばれるプログラムである。これは第1回目のゲームでは，「告白しない」を選ぶ，それ以後は前回のゲームで相手が何を採ったかによって選択を決める。前回相手が告白した（裏切った）場合，自分も今回は告白する（裏切りかえす）し，反対に，前回相手が告白しなかった（協調した）場合，自分も今回は告白しない（協調する），という方式である。わかり易くいえば，しっぺ返し戦略は「目には目を，歯には歯を」という応報原則の考えを具現化したもの，といえよう。アクセルロッドの実験は，囚人のジレンマゲームで協調解〈自白しない，自白しない〉をプレイヤー達が遵守する可能性は少なくともゼロではない，いやむしろかなり高い確率で可能性がある，ということを示している。コンピューター同様，プレイヤー達がさまざまな戦略を用いて，試行錯誤し，最終的にしっぺ返し戦略に行き着くことは十分にありうるからである。

さて，ようやくここで本稿の主題は入る地点に着た。繰返しゲームでみられるような，この「協調の遵守」が社会規範（単純に本稿では規範と呼ぶことにする）の生成と解釈してよいか，である。まずこの場合，規範の定義が問題であるが，仮に規範を「すべし ought」の言明の形をとるとして概念化してみよう。すなわち「規範とは，人が与えられた状況のもとで，何を考えるか，考えるべきでないか，いうべきか，いうべきでないか，行うか，行うべきでないか，等の行為に関しての指針を与えるルールないし原則であり，言葉で表せば，「……すべし，あるいは……すべきでない」といった命令法の形をとると定義しよう(5)。つまり，規範は，何らかの意味である種の当為的な拘束性に関わっている，とするのである。規範のこのような定義は，われわれの直観にも合致しており，社会学等で広くみられる定義でもある。

規範が「……すべし」言明で表現できるとした場合，繰返しゲームでの「協調の遵守」は規範の生成とは見做すことはできない。これが盛山の出したユニークな批判である。彼の批判の骨子は次のとおりである。繰返しゲームでみられる協調解の出現は，言い換えればプレイヤー達がある決まった戦略を選択している，ということに他ならない。戦略の選び方にある種の規則性が観察できるといってもよいだろう。ここでは行動パターンの定着と呼んでおこう。しかし，行動パターンの定着は「規範の生成」を意味するのであろうか。仮にそうであるならば，例えば「ある学生は決まって講義では教室の左側の最前列に座る」という行動パターンが観察されれば，「その学生はその席に座らなければならない」という規範が成立したことになる。これは明らかに不自然である。この種の例はいくらでも我々の日常で観察できる。第

１節で挙げた地下鉄を待つ乗客の列の例などもその１つである(6)。

もちろん，現実には，規範は行為者の意識の中に内面化されている場合が多いので，行動パターンは規範の遵守の顕れとみてよいかもしれない。しかしゲームに登場するプレイヤー達はそのような観念を持っていない。プレイヤーだけでなく，利得，選択肢とその間でのプレイヤーの選択行為，など，ゲームを構成している概念は全て，なんら規範的要素を帯びてはいないのである。それに仮にゲームに登場するプレイヤーが規範に関する何らかの観念を持っていると仮定しても，われわれは繰返しゲームの議論から規範の生成に関して何の知見も得られないだろう。なぜなら，それは「プレイヤーが規範に従って行動すると仮定したときに，プレイヤーがなぜ規範に従うのか」と問うているに等しいからである。

また，行動パターンの定着がいつのまにか規範性を帯びるケースもある。例えば「電話をきるのは，かけた側である」とか「ナイフやフォークを使う際には音を立ててはいけない」等の，マナーやエチケットの類がこれに該当する。このケースでは繰返しゲームでの行動パターンの定着を規範の生成と解釈することも許されるかもしれない。しかし問題は，繰返しゲームの論理構成に先の「教室に座る学生」と「電話やナイフ」のケースを分けて議論することが可能になるような要素が存在するかである。残念ながら，そのような要素は見当たらない。「電話やナイフ」のケースを規範の生成と解釈するならば，「教室に座る学生」のケースもそう解釈するしかない。繰返しゲームは極めて一般的・抽象的なモデルであり，２つのケースを分かつ線引きをすることはできない。

以上，規範の生成に関する説明原理として，そして経験科学のモデルとして，ゲーム理論は不適切である，という盛山の主張をみてきた。ただしゲーム理論家の立場を擁護するべく，２つばかりいっておかねばならない留意点がある。公正を期すためにも，これらは述べておかねばならない。１つはゲーム理論はコンヴェンション（convention）の生成に関する説明モデルに関してなら適切なものである，ということである。例えば，囚人のジレンマ繰返しゲームなどで，行動パターンの定着をコンヴェンション生成の説明として解釈しているゲーム理論家は多くいる。実際，こちらの方が多数派である。

コンヴェンションの概念はヒューム（Hume [1739]）に発する。コンヴェンションも規範と同様，人びとの行動を律する規則の１つであるが，これは成文法のように計画的意図的に作られたのではない。コンヴェンションは，いわば慣習的規則の１つであり，人びとがそれに従う理由は，権威や倫理的妥当性をその中に見出しているからではなく，それに従うことによって利益を得ることができることを経験的に知っているからである。

コンヴェンションは一応規則ではあるが，同時にそれが規則でならなければならない理由が「便宜的」に決まっている，そういう規則でもある。これは形容詞 Conventional に「便宜的」という意味があることからも推察できよう。便宜的にでも規則を決めておかなければ，人びとの間で協力関係を作り出すことは困難である。そういう例はわれわれの日常で幾つも観察できる。例えば，恋人との待ち合わせ場所の決定の問題を考えよう。この場合でも，百貨店の１階の入り口か，２階の紳士服売り場か，どれかに決めなければ，２人はうまく落ち合うことができないだろう。コンベンションは，社会規範とは異なり，「それに従うことがなぜ正しいのか」という問いに答える必要はない。先の例でも１階の入り口でも，２階の紳士服売り場でも，どこでも構わない。大事なのはどの階で待ち合わせるかを決めておかねば２人のデートはうまくゆかない，ということである。

ゲーム理論はコンヴェンションの生成に関する説明モデルとして解釈するなら，何ら問題はないと思える。繰返しゲームにおける協力解の定着は１つのコンヴェンションが形成されたとして解釈できるだろう。この意味を，先の恋人の待ち合い場所に関するゲームで考えてみよう。男女は１階で待つか，２階で待つかの２つの選択があり，同じ階でうまく落ち合えれば２人とも１の利得を得て，そうでなければ双方０の利得を得るとしよう。デートが何回も続く，繰返しゲームを考えよう。このゲームで，２人が取る戦略として次のものを考えよう。２人は(1)前回のデートで２人が同じ階で落ち合えたら，次のデートでは選択を変更しない，(2)前回のデートの時に，同じ階で落ち合うことができなかった場合，次回はどの階で待つ

かはランダムに選ぶ（サイコロを振って決めると考えてもよい），(3)初回のデートでの選択は任意でよい。この戦略の組は経験的にも尤もらしく見えるだろう。容易に証明できるように，このゲームにおいて，男女は確率1で何回目かのデートにおいて同じ場所で落ち合うことができ，以後2人はその選択を変更することはない。なぜならそこがナッシュ均衡だからである。同じ階で落ち合うという，コンヴェンションの成立がゲームのナッシュ均衡として説明できるのである。

第2に留意すべき点は，規範への随順に関してである。囚人のジレンマ繰返しゲームは，規範はすでに何か与えられており，それに対する人びとの随順行動を説明するモデルとしてなら説得力を持つと思える。この見解には盛山［1995］も同意している。この場合，規範に従うとは，ナッシュ均衡をもたらす戦略をプレイヤー達がとることであると解釈できよう。プレイヤーは，そこからの逸脱行動をとることはない。なぜなら逸脱はプレイヤーにとって不利益でしかないからである。これは逸脱行為に対して罰が科されている，と解釈できる。

しかし，このモデルが現実の規範への随順行動をどの程度うまく説明しているかに関しては，疑問が残る。フォーク定理が主張するように，あらゆる個人合理的な利得ベクトルに対してそれを実現するナッシュ均衡戦略が存在するからである。複数の規範が論理的には考えられるのにも拘らず，なぜ人びとがある1つの，現実に存在している，規範に従っているのか，に関しては何も語ってくれないのである。せいぜい言えることは，規範は一旦成立してしまえば，そこからの逸脱行為は容易に起りえず，その意味で極めて安定したものである，と言うことであろう。

3. 説論としてのゲーム理論

まず，用語の説明から始めたい。本論文のタイトルにある「規範の受諾」という用語に関してである。これは前節までに論じてきた「規範の生成」とは別の概念である。規範の受諾とは，人びとがなぜ規範を受け入れるのか，という理由と根拠に関わっている。これに対して規範の生成とは，人びとがどのような契機を経て規範を受け入れたのか，という規範形成のプロセスと経緯に関係している。簡単に言えば，前者での問いは Why であり，後者のそれは How to である，と言うことである。

2つの概念の違いを説明するため例を出そう。ある大学で，学生の自動車での通学許可の問題を考えよう。大学関係者は全員，駐車スペースが殆どないこと，車通学を許可すれば近隣住民に多大な迷惑が及ぶことを知っている，としよう。他の色々な諸条件を慎重に吟味した結果，彼らは誰しも車での通学は認められない，という結論に達せざるをえないとしよう。その時彼らは「自動車通学はしてはならない」という規範を受諾している，と考えてよいだろう。しかし，このことから直ちに自動車通学禁止が規範として，つまり学生が従うべき規則として機能するわけではない。学部教授会や評議会など関係部局で審議し承認されるという，学則上の諸手続きが必要となる。このプロセスが，すなわち，「規範の生成」に該当する部分である(7)。

この例からわかるように，「規範の受諾」は「規範の生成」よりも論理的には弱い概念である。後者は前者を前提とするが，逆は真ではない。ある規範が生成されている，ということは人びとがその規範に従っているわけであり，それはすでに人びとが「その規範に従うことが正しいことである」と認めているのである。しかし人びとが「その規範に従うことが正しいことである」と認めたからといって，直ちにそれが規範になるわけではないのである。

以上の概念の区別は本稿でのオリジナルであり，社会学や哲学でこのような区別が一般的である，というわけではない。この点は予めお断りしていく。

さて，結論から述べよう。ゲーム理論は「規範の生成」を説明する経験科学的なモデルとしては，不適格である。これは前節で論じたとおりである。しかし「規範の受諾」を説明する道徳哲学・倫理学上のモデルとしてなら，適格である。これが結論である。いわばそれは，規範を受諾しようとしない人に規範の受け入れがなぜ必要なのかと説法する際に使われる（よくできた？）寓話なのであ

る。これが本稿の副題「説諭としてのゲーム理論」に込められた意味である。具体的には，この説諭は次のような筋書きで進むであろう。場面は，規範を守護する者（アポロ）が規範を否定する者（ディオ）に対し説法を試みている，そういうくだりである(8)。

> アポロ「規範を受け入れなければ，この世は弱肉強食の世界になる。弱者が不憫だとは思わないかね。あなたにも人を思いやる心がある筈だ。」
> ディオ「否！人間は全て利己的であり，他者への配慮など考える筈がない。あなたの主張は人間の本質を見失った間違った考えに基づいている。」
> アポロ「……。宜しいでしょう。人間が利己的か否か，この問題に関しては議論しても結論は出そうもありません。ここではあなたがいうように，人間はその本性において全て利己的であり，隣人への愛や同情の感情は一切持たないと仮定しましょう。あくまで仮定ですが。そのように仮定しても尚，人は規範を受け入れざるをえないことを私が論証して差し上げる。ただ念を押しておきますが，人間本性に関するこの仮定は，あなたにとって最も有利な足場です。そのような有利な足場に立っても，なお規範を受諾せざるをえないという結論が出れば，もはやあなたは私の議論の正当性を承認せざるをえないことになります。宜しいですね。」
> ディオ「……」

ここでアポロは囚人のジレンマ繰返しゲームの寓話を語る。利己的に行動しあえば，両者共倒れになること，そして人びとは規範に従う方が長期的には有利になるという論理的帰結をディオに説くのである。この説諭が成功するか否かは，ひとえに囚人のジレンマ繰返しゲームに説得力があるか否かに係っているのだが，それは当面の問題ではない。大事なことは次の２点である。第１に，ここでは囚人のジレンマ繰返しゲームは，多くのゲーム理論家たちが見做しているような，人間社会の対立と協調の関係を説明する経験科学上のモデルとして利用されているのではないという点である。そして第２に，それはむしろ協調しなければいかなる論理的帰結が生まれるかを示した，１つの思考実験として機能しているという点である。

ゲーム理論の人間行動に関する合理性の仮定は，経験的事実に照らし現実妥当性を欠くとして，しばしば批判されてきた。人間は利己的動機によってのみ行動するというのは，人間行動の多くの要素を捨象しており，確かに経験科学としては問題ある仮定である。しかし，先のアポロの台詞にあるように，説諭としてならこの仮定はむしろ強みを持つ。利己的に行動し，他者への配慮は一切行わない人びとのみからなる社会でさえ，規範に従うことに合意するのであるから，どの社会でも，その社会の構成員がどのような人間本性を持とうとも，規範は合意される筈である。

慧眼の読者諸氏は，先のアポロの説諭がホッブズ（Hobbus）が主著リヴァイアサン［1651］で展開した議論と同型であることに気づかれたかと思う。これが先に私が「道徳哲学・倫理学上のモデルである」と述べた意味でもある。ホッブズは，人が国家権力による秩序を認めざるをえないことを，有名な自然状態の概念を用いて論証しようとする。自然状態において，人びとはあらゆる行為を行う権利（自然権）を持っている。他者の身体や財産を自由にする権利さえもである。この状態では，人びとは互いに相手を攻撃しあう，「万人による万人の闘争」へと導かれていく。そして最終的に人びとは，このような闘争状態の愚を悟り，互いに自己の自然権の一部を放棄する契約を結ぶ。この契約の履行と維持のために国家が必要とされる。以上がリヴァイアサンの骨子である。

自然状態は，あくまで架空上の想定であり，歴史上かつて実在したというわけではない。それはホッブズが国家権力の正当性を人びとに説くために捻出した概念的な装置である。したがってホッブズの議論は，規範ないし秩序の歴史的生成のプロセスを論じた社会科学として評価するのは不適切であり，規範や秩序の正当性を説く道徳哲学ないし倫理学の系列に属すると考えるべきである。実際，学説史の上でもリヴァイアサンでの言説は，カントやミルら道徳哲学と同様，正義や道徳に関する社会思想の１つとして位置づけられている。

さて，先ほどアポロの説諭が成功するかどうかは，ひとえに囚人のジレンマ繰返しゲームに説得力があるか否かに懸っている，と述べた。もし仮

に説得が成功したらどうなるであろうか？　まず彼は「規範に従う方が結局は得なのだ」という論理的帰結を理解したことになる。これは囚人のジレンマ繰返しゲームの議論を理解することに他ならない。この論理的帰結を前提にして「規範に従うことは正しいことである」という結論を支持したことになるから，そのためには「得になることは正しいことだ（得は徳なり）」という価値命題をもう１つの前提にすることが必要になる。こうすれば三段論法で２つの前提から先の結論「規範に従うことは正しいことである」が導ける。するとこれは道徳哲学上の１つの立場である「正善一致」の考えをディオは支持したと解釈してよいであろう。たとえばディオはミル（J.S. Mill [1863]）流の功利主義的正義論などを支持しているのかもしれない。

しかし，説論が常にミルらの「正善一致」の立場からばかり行われるわけではない。非功利主義的な立場に立って，規範の受け入れに関する説論も（その説論が成功するかどうかは別として）可能なのである。次節で我々はロールズの正義論 [1971] がその代表的なものであることを見るだろう。

ここで前節の議論を振り返ってみよう。一部のゲーム理論家達は繰返し囚人のジレンマゲームを規範の生成モデルとして誤って解釈した。彼らはどこで躓いたのであろうか？　私の診断は以下のとおりである。ゲームに登場するプレイヤー達は，自己の利得のみに関心を抱いて行動する合理的主体である。そしてそれだけの存在である。彼らの頭には，規範や規則といった観念はない。そもそも社会という観念すらない。そういうプレイヤー達の行動を「規範に従ったものである」と判断しているのは，ゲームを外側から観察している，他ならぬ研究者（一般には論文を読んでいる読者）なのだ。つまり，ゲーム理論家達は外から観察している自分達の視点をゲームの内側にいるプレイヤー達も持っている，と短絡的に思い込んでいるのである。研究者の視点（これをモデル外視点と呼んでおく）とプレイヤーの視点（モデル内視点と呼んでおく）の混同，これが躓きの原因であったのだ。

ゲームの構成概念，プレイヤー，利得，戦略など，はすでに予めその意味が限定されている。これは，ゲーム理論に限らず，数学など演繹的に構成された学の体系全てに共通していえることである。そこでは公理や前提が予め定まっており，これらを組み合わせて命題を導いていく。したがって当たり前のことだが，そこから出てくる諸命題の解釈は，――それをあくまで経験世界の有りようを説明する目的で行う限り――，モデルを構成する諸概念に予め備わった意味のみを組み合わせて遂行されねばならない。残念ながら，規範の生成という解釈は，この方法では導出できない。それはモデルを外から観察している研究者が勝手に持ち込んだ解釈である。これに対して，前節で出てきたコンヴェンションは，プレイヤーの利得や合理性に直接結びついた概念であるから，コンヴェンションの生成という解釈は，モデルに内属するプレイヤーの視点から導けるのである。この場合はモデル外とモデル内の２つの視点から得られた解釈が一致しているのである。

ゲーム理論を「説論」と見なすわれわれの立場は，モデル外視点とモデル内視点をうまく分離し，両視点からの解釈が食い違っても問題が起きないように工夫しているといえよう。（もっとも，この場合は経験科学としての役割をゲーム理論は放棄しなければならないが。）説論としての囚人のジレンマ繰返しゲームからいかなる教訓を引き出すかはモデル外視点に立つ観察者の自由である。先の劇では，アポロとディオはモデルの外で寓話（囚人のジレンマ繰返しゲーム）を鑑賞している。彼らが仮に「規範への受諾は正当なもの」と結論したとしても，この判断が，もとよりモデル内視点と一致する必要はない。これは，映画鑑賞における観客と作り手の関係に例えていえば，以下のようになろう。「観客がどんな感想を抱こうとも自由であり，それが作り手の映画に込めた意図と一致する必要はない」と。

以上が「説論としてのゲーム理論」に関する説明であるが，まだ触れておかねばならない留意事項が２つ残っている。１つは規範の行動主義的理解であり，もう１つは事実命題と価値命題に関するヒュームの原則そしてムアの自然主義批判である。

ゲーム理論家たちの誤った解釈は規範生成に関する行動主義的理解に端を発している。規範の行

動主義的な定義は，例えば「規範とは，諸行動を発する頻度の変化からなっており，それは社会的な再強化刺激（サンクション）による再強化刺激によって影響を受ける。」(Scott［1971］）等が代表例である。このような理解に対しては次のような批判が可能だろう。「『サンクションとして』という意味は，行為者の行為に対する正ないし負の評価と報奨ということである。そして，他者のある反応が行為者の行為に対する『サンクション』であるという命題は，きわめて理論的な負荷のかかった解釈であって，現象そのものの直接的な記述のレベルを超えている。ある他者の反応的行為がある行為者の行為に対するサンクションであるためには，前者が後者に対応しているという結合が行為主体の中に成立しなければならない。動物の場合，この結合は，反応的行為が初発行為の直後に提供されることによってある程度可能になる。ただし，この場合でも初発行為に対する反応的行為が，前者に対する『サンクション』だという認識は，あくまで人間の側のものであって，我々は，動物の行為と我々の反応的行為との相互作用的連結を行為とそれに対するサンクションとして解釈しているのである。」(盛山［1995］131-32 頁）要するに盛山は，サンクションという概念は，モデル（動物実験）の外にいる研究者が知らずに持ち込んでしまった解釈であり，モデル内にいる行為者（動物）の反応的行為に関する直接の記述ではない，という点を批判しているのである(9)。その批判の要諦は，実はわれわれがつい先に述べたモデル内視点とモデル外視点の混同と全く同じである。

第 2 の留意点に移ろう。ゲーム理論家は例えば「ある教授は常に講義に 10 分遅れて来る」という行動パターンが観察されることから「その教授は講義に 10 分遅れて来るべきである」という規範が生成したものと，誤って解釈したのだった。この例では「……である」という事実命題から「……であるべきである」という価値命題を導くという誤りを犯している。ヒューム［1739］の原則，事実命題から価値命題を導くことはできないという原則，に違反しているのである。この意味でゲーム理論家達の誤った解釈は，ヒュームの原則を侵犯した 1 つの具体例である，と言えよう。このような誤りは，価値の問題に取り組まざるをえない規範経済学研究では特に注意しなければならない点である。人がある規範や正義の原則に従っているからといって，その規範なり正義なりが「よい」とは言えないのである。経験科学の方法をストレートに規範経済学の場に持ち込むことは慎まねばならない。

この点で，想起すべきはムア［1903］の道徳哲学での自然主義批判である。ムアの所説を掻い摘んで説明しよう。ムアによれば「善とは何か」は定義できないし，定義しようという試みはことごとく失敗するという。「善とは最大多数の最大幸福である」，「善とは進歩である」，「善とは律法の教えを守ることである」等々，これらは全て善の概念を自然的実在物に還元しようとする試みであり，彼はこれらを自然主義的誤謬として退けた。これに対して「このりんごはよい」という場合の「良さ」は自然的性質によって定義できるだろう。例えば，香り，甘さ，光沢などでりんごの良さは定義できると考えてよい。しかし，この論法は，「ある人のとる行為の良さ」を判定する場合には適用できない。ムアは，従来の倫理学説は「行為の良さ」を「りんごの良さ」と同じ論理で判定できる，とする過ちを犯してきたと批判したわけである。

ムアのこの批判は，規範生成に関するゲーム理論家の誤った解釈にも妥当する。彼らは規範概念を行動パターンの定着という観察可能な現象と同値であるという，1 つの自然主義的誤謬を犯したのであるから。観察可能な事象のみに科学の対象を絞る，という方法は社会科学の客観性に敏感な研究者ならば誰しもすぐに思いつく方法ではある。先に批判した行動主義はこの一形態であり，観察可能な現象として行動に注目したのである。しかしこの方法論の社会科学への適用には限界がある。社会科学は価値の問題にコミットせざるをえず，観察対象の性質や機能，個々の対象間に成り立つ関係のみを分析しても，価値の問題は解けないのである。

4. 道徳哲学とゲーム理論 ——ロールズにおける原初状態と反照的均衡

ロールズの主著『正義論』の刊行は，20 世紀

の倫理学をメタ倫理学から規範的正義論へ大きく転回させたという意味で画期的なできごとであった。内容に対するさまざまな批判はともかく，「正義論」はこの分野の研究に携わる全ての人びとが拠って立つ１つのパラダイムを提供してきたのであり，この点は今日大いに評価されるべきことである。本節では，正義論の論理構成も前節で見たホッブズのリヴァイアサン同様，正義の原理の受け入れに関する１つの説論であることを示したい。しかもそれは，ゲーム理論やホッブズが正善一致の立場を説く説論であったのに対し，非功利主義的な説論なのである。

まず正義論の概要を説明しよう。ただし，その全容ではなく本稿に関係する限りにおいて，であるが。例で説明したい。今，一定額の富があり，社会の人びとの間でこれをどう分配すべきか，話し合っているとする。社会の構成員たちはさまざまな立場から分配方法を提案するであろう。極めて資質と能力に恵まれた人は，能力に応じた分配を主張するだろう。彼は「できるだけ多くの富を高い能力を持つ人に与えるべきである。彼はそれを使って事業を起こし，結果的により多くの富を我々にもたらし，社会を繁栄と進歩へ導いてくれるからである」と主張する。一方，資質と能力に恵まれない人は，平等分配を主張するかもしれない。彼はこう反論する。「能力に応じた分配は結果的に貧富の差を拡大し，社会不安を引き起こすことになる。進歩や繁栄も結構だが，人びと同士の連帯と友愛こそ人間にとって最も価値あるものなのだ。そのためには，たとえ貧しくとも，人びとが平等に扱われることが必要なのである」と。

しかし，２つの判断ともに正義感情に基づいた判断とはいえない。どちらも自分の立場からしか発言していないからだ。ある分配方法が正義に適うものとして認められるためにはそれが人びとの現在の立場から離れて，すなわち不偏的な立場から，提案されねばならない。ロールズはこう考える。では，どうすれば人は自分の立場を離れることができるのか？　そのためにロールズが出したアイディアが原初状態（the original position）であり，これは公正かつ不偏的な正義判断を人がとることを可能にする仮想的な場である。原初状態において，人は現在の自分の立場や境遇，自分に関する個人的な情報，および自分が現在住んでいる社会，これらに関する記憶と知識を一切剝奪されている（無知のヴェール［veil of ignorance］）。自分が社長なのか従業員なのか，男なのか女なのか，どんな趣味を持っているのか，日本にいるのかアメリカにいるのか，全くわからない。人が知っていることは以下の３つである。第１に，人は論理的に起りうる，このような自分自身の境遇と環境の組み合わせの集合が何であるかは，完全に知っている（その中でどれが現実に生起しているのかは知らない）。第２に，人間・自然・社会に関する一般事実と一般法則も知っている。例えば，アラブには石油が豊富にあるとか，人はどのような状況で嫉妬や恐怖に駆られるのか，とか収穫逓減の法則とか，に関する知識と理論をである。最後に人は自分が合理的存在であることも知っている。

さて，原初状態において人は自分が最悪の状況に陥ることを懸念する結果，次のような分配方法を正義に適った原理として提案する筈だとロールズは主張する。すなわち「分配の不平等は，それが最も恵まれない立場にいる人の境遇を改善する限りにおいて許容される」と。この原理は格差原理と呼ばれる。格差原理は功利主義的な分配方法，例えばベンサム的な効用和の最大化を必ずしも支持しないし，平等主義的な分配方法とも違っている。

原初状態や格差原理に関するこれ以上詳細な記述は，本稿では必要ない。本稿で主張したい点は，このロールズの原初状態での正義判断も，格差原理の必要性を説くための寓話である，ということである。ではこの寓話を説諭するものは誰か？　原初状態にいる人びとなのか，ロールズなのか，それとも他の誰かなのか，といった疑問が新たに発生する。

この問いに答える鍵は反照的均衡（reflective equilibrium）の概念にある。ロールズは原初状態から演繹された正義の原理がそのままストレートに現実社会に適用されるとは考えていない。人びとは原初状態で正義の原理に合意した後に，無知のヴェールが取り払われ自分の境遇と置かれた環境を知る。そして改めて正義の原理を適用し，不都合がないかどうかを調べてみる。人びとは十分に熟慮を経た道徳判断を用いてこの調査を行う。仮に何か不都合があった場合，考えられる原因は

3つである。原初状態における正義の原理の演繹に誤りがあったか，原初状態での知識に関する想定に誤りがあったか，十分に熟慮を経た道徳判断が実は歪められていたかである。一応，原初状態での正義判断というロールズの構想そのものは正しいと仮定し，また無知のヴェールのもとでの人びとの推論も全く誤りがないとすると，考えられる原因は後の2つということになろう。したがって，演繹された正義の原理がうまく機能しない場合，一般事実や一般法則の理解に間違いがあったか，論理的に生起可能な境遇と環境の集合の記述に誤りがあったか，または直感的道徳判断に不備があったかである。誤りを改訂した後に，再び原初状態に戻り正義判断を行い，新しい正義の原理を演繹する。この原理もまた，現実世界で十分に熟慮を経た道徳判断を用いて吟味される。この帰納と演繹のプロセスを繰返し，人びとは十分な熟慮を経た道徳判断に完全にマッチしてくれる正義の原理を演繹できるような，そういう原初状態の完全な記述を発見できる。この状態が反照的均衡であり，そこでの正義の原理は格差原理になる，とロールズは主張する。

反照的均衡は演繹と帰納によって適切な正義の原理を発見する方法であるが，演繹と帰納の担い手は違っている。反照的均衡の担い手は直感的道徳判断を行い，正義の原理の現実妥当性を判断している人びとである。わかり易くいえば，現実に存在する「我々」自身である。一方，原初状態で正義の原理を演繹する人びとは，直観的道徳判断を行う「我々」が頭の中で想定した架空の人びとである[10]。「我々」は原初状態の条件を設定し，そこに住む住民を無知のヴェールで覆ったうえで，彼らに論証させる。彼らが導き出した正義の原理を今度は「我々」が現実と照合する。十分に熟慮した上で，正義の原理がわれわれの健全なる正義感覚と適合するならば，プロセスはここで終了する。そうでなければ原初状態の記述の誤りを見つけ，訂正した後で再び原初状態の人びとに論証させる……。反照的均衡はこのようなプロセスを経て導かれるのである。

もはや明らかであると思うが，原初状態という寓話を説いているのは，正義の原理と熟慮を経た道徳判断とを照合させている「我々」である。ただし，この場合の説論は，前節のアポロとディオのように対立する2人の間での緊迫したダイアローグの形式を取るのではなく，自分自身に語り掛け納得するという意味で独語的な形式で進んでいくが。

原初状態で人が格差原理を正義の原理として採択するという命題を導出するため，ロールズはゲーム理論のマクシミン原理を利用する。原初状態において人は極端にリスク回避的行動をとり，それが人をして格差原理を採択せしめることになる，とロールズは説明するのである。この議論は，後にハーサニ［1955］らゲーム理論家から激しい批判を浴びたが，規範的正義の研究でゲーム理論を意識的に適用した最初の試みとしては，――成功したかどうかは別にして――，評価してよいと思う。ハーサニらの批判があるとしても，少なくともロールズは説論としてゲーム理論を利用したという点では「正しかった」のだから。

5. 結　論

再び，盛山［1995］に戻ろう。彼は最終章の最終節でこう述べている。「……秩序問題という問題は本来的に経験的な問題というよりはむしろ理論的に構成されたモデルが自ら作り出した問題だったからである。経験的な問題という表現をここでは経験的に存在する現象を理論的に説明する問題という意味で用いている。この意味では，秩序問題はこれまで一度も経験的な問題ではなかった。それはむしろ，理論的に構成された社会のモデルが自ら作り出した問題である」（盛山［1995］267頁）本稿をここまで読まれた方々には，この文章の意味は説明するまででもないであろう。ただ，引用の最後の一文にあるように，「自ら作り出した問題」が何なのか，そして何の役に立つのか，に関してまでは論及していないが。本稿で明らかにしたように，それは「説論」として，規範や秩序を否定する者に対する説法として，道徳哲学・倫理学上のモデルとして役割を果たすのである。

さて，第1節で私は「我々の結論は盛山の議論から出てくる論理的帰結の一つである」と述べた。しかし，同時に「規範をあくまで先に定義したように，その言語表現が「……すべし」という命令

法の形をとり，何らかの行為の規範的妥当性に関わると仮定した場合での話であるが,」という但し書きもつけている。この但し書きは盛山の批判を意識してのことである。というのも，盛山［1995］の第 5 章第 3 節を読む限り，彼はわれわれの定義に反対しているように読めるからである。以下盛山の議論を簡単に検討しよう。

彼は規範を ought to「……すべし」の言明ではない，と言い切っている。その理由は，私の見る限り 2 つある。第 1 に，彼は「……すべし」言明の中には人が従うべき規範を表現したものとは必ずしもいえない例があることを指摘する。例えば，ルパンのように紳士的に「君が手に持っている宝石を私によこすべきだと思うがね」（盛山［1995］，137 頁）という表現は規範ではなく，個人の願望や強要を表現したものである。第 2 の批判は，当たり前のことだが，規範は言明ではない。規範は文章で表現したもの，そのものではない，ということである。

しかし，以上の批判は私の定義には当てはまらない。まず第 2 の批判から見ていくが，私は規範を言明そのものとは定義していない。慎重に「その言語表現が『……すべし』という命令法の形をとる……」といっているのは，盛山の批判に配慮した結果である。第 1 の批判については次のように答えることができよう。盛山が挙げている例は全て「……すべし表現を取りつつも規範を表現したものとはいえない言明」である。これと逆「規範の表現でありながら，……すべし表現を取り得ない言明」の例は 1 つも挙げていない。つまりこういうことだ。盛山の展開する議論は「ある言明が規範の表現になっているためには，それが『……すべし表現』であれば十分である」という主張に対する批判にはなり得ても，その逆「ある言明が規範の表現になっているためには，それが『……すべし表現』であることが必要である」という主張，これが本稿での立場だが，に対する批判にはなっていないのである。

ともあれ，規範概念の定式化について，われわれと盛山の違いをこれ以上詮索しても得るところは少ないであろう。私は本稿を演繹的な理論体系として，つまり仮に規範を「……すべし」言明で表現されるとしたら（どうなるか），というシナリオで組み立てたつもりである。この場合，定義は何を分析するかの目的に応じて適切に立てればよい。われわれの意図は秩序形成の問題にゲーム理論を正しく適用するに際しての注意事項を述べることにあり，その意図に照らせば，われわれの規範の定義はそれなりに適切なものである。この点に関しては盛山も同意するであろう。同じくまた，盛山は彼独自の意味論的社会理論，一次・二次理論と呼んでいるが，の構築の必要性を訴えるための例としてゲーム理論を取り上げているのであり，その場合における彼の規範概念の把握もそれなりに納得のいくものである。われわれと盛山とではゲーム理論の制度現象研究での位置づけという点を除いては，共有している関心は小さいと見るべきである。

付録　セミナーを終えて――質問と返答

この論文は 21 世紀 COE-GLOPE「開かれた政治経済制度の構築」主催でのコンファレンスを始め，多くの研究会で発表された。反応は予想したとおり，賛否両論さまざまなものであった。しかし本稿の目的は 1 つの問題提起にあることを考えれば，これはこれで良かったのかもしれない。目的は達せられたと筆者は判断している。

当初は，これらコメントに従って改訂しようと考えたのだが，やはりオリジナルな原稿を大部分そのまま残すことにした。敢えて「隙」のある構えを見せた方が，読み手側も突っ込み甲斐があるというものだ。更なる議論を巻き起こせるかもしれない。

ここではセミナーの席上で出されたさまざまな疑問や批判のうち特に重要と思えるものを 3 つ挙げ，Q&A 形式で答えていくことにしたい。これを持って改訂の代わりにしたいと思う。

批判 1　本稿での規範の定義は妥当ではない。規範とは「……べきである」という言明の形をとるのではなく，慣行的に決まった約束事である。規範に従っている者に「何故そうしなければならないのか」と聞いても「ただそうしていることになっているから」としか，答えようがないもの，それが規範である。

回答：この他，規範とは同調することへの自分自身と他者を含めての期待である，とする批判などもあった。規範とは何なのか，という問は確かに重要である。しかし，私の論文は規範の定義を巡る議論ではない。規範を「……すべし」あるいはそれに類似する，人びとが服すべき義務や正義の観念を具備した言葉で表現可能なもの，と仮定して，での話である。規範という呼び名が相応しくなければ別の名をつけても，例えば規範*でも構わない。このように定義した規範*が成立していくさまをゲーム理論がどこまで説明できるか，その説明を規範経済学研究にどう生かすか，これが本論文での主題である。批判１は，この主題の展開に何ら打撃を与えるものではない。

ところで呼び名であるが，規範が批判１で述べてあるとおりであるとすると，本稿での規範はスタンダード（standards あるいは standard model）と呼ぶ方がいいかもしれない。英語の使い方であるが，norm と standards は微妙に意味が違うようにも感じる。例えば「ここの会社では皆 9 時に出勤している」というのが norm，「しかし理想はもう少し早く 8 時半ぐらいに出勤すべきなのだ」が standards である。ただこの分け方も確実なものとはいえないが。

批判２　規範の生成とは何か。その定義が今ひとつ明確ではない。

回答：規範の生成という概念は「囚人のジレンマ」繰り返しゲームの説明で登場した。そこで筆者（及び盛山）は「行動パターンの定着＝規範の生成」という図式は誤りである，と主張した。一方，繰り返しゲームにおいて，ゲーム理論家たちはこの図式を使って「規範の生成」を説明しており，ゆえにその説明は失敗している，と結論したのである。

少なくともわれわれは規範の生成は行動パターンの定着と同じではないと考える(11)。ではわれわれが考える規範の生成とは何か。筆者は本稿の第３節で「規範の生成とは，人びとがどのような契機を経て規範を受け入れたのか，という規範形成のプロセスと経緯に関係している」と書いた。規範の形成とは，次の２つのステップからなる。

①　規範形成に関わっている人びと全てが「規範に従うことは正しいことである」という認識と自覚を持つようになる。これを正義感覚の発生と呼ぼう。

②　規範そのものを暗黙の了解の次元を超えて，制度として機能させる必要がある場合，規範の制度化に関する形式的手続きも必要となる。これを規範の公共的承認のプロセスと呼ぼう。

本稿において大事なのは①である。①での「規範形成に関わっている人びと全て」とは規範に従わねばならない人びとのことである。したがって，規範の生成という場合，この人びとの内的心理構造の変化が必要とされる。ゲームのプレイヤーには合理的主体という以外，何の心理的構造も与えられていない。ゲーム理論を用いて，①が説明できないのは至極当然なのだ。確かに「囚人のジレンマ」繰り返しゲームでは，協調の遵守が見られた。しかし，そのような「行動」をプレイヤーがとった理由は，「そうすることが自分にとって利益になる」からであっても，「[そうすることが自分にとって利益になる］からそれを規範として認め，従っている」からではない。この２つの言明には大きな違いがある。前者でのプレイヤーは単なる合理的主体でしかないが，後者でのプレイヤーは少なくとも「[自分が合理的主体である］ことを自分自身で認識できていなければならない」。

われわれが規範の生成という場合，人びとの行為レベルでの変化のみならず，その変化の基礎にある人びとの正義感覚の発生を説明できねばならない。したがってゲーム理論で規範の生成を説明するには，世界を認識し，理解し，善悪の判断を行う主体として，プレイヤーを構成しなおさなければならない。このためには合理的経済人仮説から離れ，新しい人間モデルを探求せねばならないだろう。これはかなりの難問である。しかし，もし成功すれば非常に価値ある研究となろう。

最後に一点だけ，申し添えておく。規範の生成のためには「規範形成に関わる人びとが正義感覚を持つようになる」ことを必要であるとする見解は私のオリジナルであり，盛山の見解ではない。彼は「規範の生成＝行動パターンの定着」の図式は誤りである，としか述べていない。彼自身の規範の生成概念は提示していない。彼の議論にとっては，行動主義的理解とでもいうべき，この図式の誤りを指摘するだけで十分だからである。

批判３　規範の生成モデルを議論するならば，繰り返しゲームよりも進化ゲームを焦点に据えるべきではないのか。進化ゲームによる制度の生成にも同様な問題があると考えるのか。

回答：繰り返しゲームを例として使った理由は２つ。１つは本稿の起点となった盛山の議論が繰り返しゲームに基づいていること。もう１つは繰り返しゲームの方が進化ゲームに比べ構造がより簡単な分だけ私の論理がよりわかりやすいものになるだろうという，配慮である。

さて質問に対する回答を述べよう。答えは「同様な問題がある」だ。進化ゲームにアップグレードしたところで私の批判を打ち返すことは不可能である。批判２で回答したように，進化ゲームにも規範の生成に必要な「正義感覚の発生」の契機となるものがどこにもないからだ。

進化ゲームによる制度生成に関して，具体例を挙げて説明していこう。この例は Aoki［2001］の議論を整理単純化した清水［2003］に依拠している(12)。労働者がある技能選択に直面しているとする。技能は２つある。１つは文脈的技能であり，OJT で身に付くような幅の広い可塑的な技能のことである。もう１つは機能的技能であり，より専門的知識が必要とされる細分化された技能のことである。彼らはこのうちひとつの技能を身に付けるものとする。２人一組でチームを作り，作業するとする。チームの組み方はランダムマッチングで，何回も繰り返すとする。同じ技能を持つもの同士が出会えばより高い生産性をあげ，より高い報酬をもらえるとしよう。文脈的技能を持つもの同士だと，互いに利得（報酬）は６，機能的技能を持つもの同士だと，互いに利得は４であるとする。違った技能を持つもの同士では生産性はあがらず，したがって報酬は少ない。利得は互いに１であるとする。

このランダムマッチングの過程で労働者がどれだけの期待利得が得られるかは，文脈的技能を持つ労働者と機能的技能を持つ労働者の人口比に依存して決まる。例えば文脈的技能を持つ人の割合がP，機能的技能を持つ人の割合が$1-P$であれば，各労働者は確率Pで文脈的技能を持つ人とチームを組み，確率$1-P$で機能的技能を持つ人とチームを組むことになる。したがって文脈的技能を持つ労働者の期待利得は$6P+(1-P)=5P+1$，機能的技能を持つ労働者の期待利得は$P+4(1-P)=4-3P$である。Pが3/8よりも大きければ，文脈的技能を持つ方が機能的技能を身に付けるよりも期待利得が高い。この場合，この集団内では文脈的技能の労働者の方が機能的技能の労働者より有利である。文脈的技能を持つ労働者の数が増え，逆に機能的技能を持つ労働者の数が減っていく。母集団の人口比が変っていくのだが，この調整過程は次のように解釈すればいいだろう。文脈的技能を持つ者が多すぎるため，機能的技能を持っても高い報酬が得られない。そこで労働者は機能的技能の習得を捨て，文脈的技能の習得へ切り替えるのである。ともかく，この人口比の調整は集団内が全て文脈的技能を持つ者ばかりになるまで続く。そこが１つの均衡，動学的均衡と呼ばれる，である。ここから人口比は動かない。Pが3/8よりも小さければ，この関係は逆になる。文脈的技能を持つ労働者の数が減り，逆に機能的技能を持つ労働者の数が増えていく。この人口比の調整は集団内が全て機能的技能を持つ者ばかりになるまで続く。そこももう１つの動学的均衡である。やはり人口比はこれ以上変動しない。Pが3/8であれば，どちらの技能を身につけても無差別なので人口比は増減しない。ここが第３の動学的均衡である。以上のように淘汰の過程を経て人口比が「進化」していくのである。

さて，３つの動学的均衡のうち進化的に安定なのは，最初の２つのみである。ここで進化的に安定であるとは，異種の侵入（突然変異）に対して均衡がロバストであることをいう。最初の均衡を見てみよう。ここでは文脈的技能を持つものばかりで集団が成り立っている。いま機能的技能をもつ集団がεの比率だけ侵入し，人口比において文脈的技能集団が$1-\varepsilon$，機能的技能集団がεとなったとする。両者が受ける利得は各々，$6(1-\varepsilon)+\varepsilon=6-5\varepsilon$，$(1-\varepsilon)+4\varepsilon=1+3\varepsilon$であり，$\varepsilon$の値が小さい限り，以前，文脈的技能を持つ方が有利である。いずれこの異種は集団から駆逐され，元の動学的均衡に戻る筈である。このように侵入する異種の人口比に対する割合がごく僅かである限り，均衡が攪乱されない場合，その均衡を進化的に安定な均衡と呼ぶ。同ようにして第２の均衡も進化的に安定である。一方第３の均衡は進化的に安定ではない。先と同様，機能的技能をも

つ集団がεの比率で侵入したとする。人口比は文脈的技能集団3/8−ε，機能的技能集団5/8＋εとなる。先の議論での通り，文脈的技能の集団は減り続け，機能的技能集団は増え続ける。元の均衡には戻らない。これはεがどんなに小さくてもいえる。文脈的技能をもつ集団がεだけ侵入してくるケースも同様である。

社会においてこの2つの進化的に安定な均衡のどちらが成立するかはアプリオリには決定できない。その社会での人口分布という初期条件に依存する。この場合は文脈的技能を選んでいる人が人口の3/8を超えていれば文脈的技能が普遍化する。そうでなければ機能的技能が普遍化する。この依存性を「歴史的経路依存性」と呼ぶ。このような議論を経て，日本では文脈的技能の習得が「制度」化し，アングロ=アメリカン諸国においては機能的技能の習得が「制度」化した，と説明するのである。以上が進化ゲームによる制度の生成の説明である。

さていささか進化ゲームに説明が長くなったが，問題の核心に迫ろう。技能習得のケースでは進化ゲームは制度の生成に関する尤もらしい説明に見える。しかしここでも繰り返しゲームで指摘したのと同じ問題がある。淘汰の過程を経て，全員が文脈的技能を持つことになったとしても，「文脈的技能を持たなければならない」という制度が生成されたといえるのだろうか？　せいぜいいえることは「慣行的に文脈的技能の方を身に付けるようになっている」ということぐらいである。

例を変えて，コンピューターのOS選択の問題に変えてみよう。文脈的技能をWindows，機能的技能をMacとして，ランダムマッチした2人が自分のコンピューターを使って共同で仕事をするという風に。同機種ならば仕事ははかどるが，そうでなければ互換性の問題で面倒になるとしよう。先の議論がそのまま使える。このとき全員がWindowsを使う均衡に落ち着いたとしても，「Windowsを使わねばならない」という制度ができたといえるのだろうか？　ここでも，せいぜいいえることは「慣行的にWindowsを使うことになった（なっている）」と言うことぐらいだ。こう考えれば進化ゲームによる制度生成の説明がいかに問題を孕んでいるかがわかるだろう。

労働者たちは進化の過程において，ここで言う「制度」に従って技能の習得をするのは何故だろうか？　その答えは明白である。そうすることが得だから，そうしているだけだ。彼らは与えられた進化ゲームの枠組みの中で，「反応」しているだけに過ぎない。その「反応」の様子を眺め，制度の生成と解釈しているのは，モデルの外側から見ている研究者である。モデル内のプレイヤーたちには制度に従っている，生成している，といった観念なり意識なりは全く持ち合わせていないのである。ここにもモデル外視点とモデル内視点の混同がある。（あるセミナーが終わったあとで，懇親会の席上，ある法哲学者から「しかし，何故制度の生成に関する進化論的説明が説得力を持つのでしょうかね？」と問われた。この場合，誰に対して説得力を持つのか，によって答えは変わってくる。進化論的モデルでの説明を聞いている人（モデル外にいる人）には説得力あるように聞こえるのであって，モデルに参加しているプレイヤー達にはそうは聞こえてこないのだ。）

もちろん，動学的均衡への収束は，制度ができていく過程を記述したかのように見える。しかし，先に見たように，これは人口比の変動に対してプレイヤーが適応すべく「反応」した結果であって，われわれの言う正義感覚の発生の証左などではない。ましてや制度への公共的承認の契機となるものはどこにも存在しない。

無論制度にも色々ある。制度生成に関係する人びと全てが正義感覚を持ち，それに基づいて公共的な承認を得る必要はない場合もあろう。慣行的なルールや習慣などである。その場合には進化ゲーム的な説明に対して異議を唱えるつもりは全くない。しかし，4節でのロールズの正義原理や，立法機関が制定する法律などはいかがであろうか。十分な説明ができないことは先に見たとおりである。

制度分析に関し，進化ゲームはいかなる知見をわれわれにもたらすであろうか。2つあると思う。ひとつは，やはり制度の受け入れに関する説諭として機能するということである。先の進化的に安定な均衡をもう一度考えよう。全ての労働者が文脈的技能を持っている場合である。このモデルを見ている人びとに対し，「あなたがたも文脈的技能を身につけるべきだ。そうしないと，業績が上がらず報酬面で不利になりますよ」と説諭するの

である。説諭を受け入れた人びとが「文脈的技能を身につけるべきである」と判断すれば，制度として制定される段階の手前までくることができるだろう。

もう１つは制度分析に関する経験科学的な説明理論の立場を固持した場合である。これには更に２つあると思う。１つはすでに制度が存在し，なぜその制度が定着しているのか，を説明する理論としては進化ゲームはよくできていると思う(13)。その制度が進化的に安定しているからだ，という説明は文字通り進化論的な説明である。おそらくゲーム理論家が「制度が生成される」という場合，この意味で用いているのではないか，と推察する。しかし，それは制度が安定している，異種侵入に対してロバストである，という説明であっても，制度が０から立ち上がり，生成されていく過程を説明しているわけではない。

もう１つは以下のとおりである。われわれは先に進化ゲームは「何故制度が生成されるか」に関し，十分な説明ができない，と論じた。しかし，その否定「何故制度が生成されないのか」に関しては説明できているように思える。先の技能の例で説明しよう。日本では「機能的技能を身に付けさせる」ことが制度として定着しなかった。これはなぜなのか。「それは皆文脈的技能の労働者ばかりであり，その中に，ごく一部機能的技能を持つ者が入っても，仕事ははかどらず，報酬も低いため」であると答えればよい。これは，要するにそこが進化的に安定な均衡であるということを日常的な言葉で述べ換えただけだが，説明になっている。否定言明「制度が生成されない」を論証することは「制度が生成される」ことの論証よりはるかに簡単である。なぜなら生成される場合に要求される「人びとが制度に従うべきである」という正義感覚の発生と，それの公共的承認プロセスの存在を証明する必要がないからである。

以上である。本稿では２つの主張があった。

①ゲーム理論は，規範の生成（人はいかにして規範を受け入れるのか）――規範の成立に関する経緯と過程――に関する経験科学上の説明原理としては失敗している，ということ，そして

②しかし，規範の受諾（人はなぜ規範を受け入れるか）――規範の受諾に関する理由と根拠――に関する道徳哲学・倫理学上のモデルとしてなら機能しうる。いわばそれは規範を否定しようとする人びとに対する説諭の１つなのである，ということ，

の２つである。多くのセミナーでは①に対する疑問と批判が集中した。①は私のオリジナルではなく，盛山［1995］の主張を大部分敷衍したに過ぎない。もっとも批判２に対する回答には若干盛山の主張を超えた部分があるが。規範経済学研究の戦略上の方法論としては，むしろ②の方が重要である。

ゲーム理論を始め，合理的選択モデルの説明力は大変素晴らしいものだと思う。それは経済学のみならず社会科学の多くの分野で利用されつつある状況を考えれば首肯できよう。しかし，所詮経験科学的な方法である以上，価値の問題，――なにが善く・何が悪いのか――といった，を取り扱うにはおのずと限界がある。合理的選択モデルによる規範経済学研究には限界があるのだ。このことは一見するとわれわれにとって大きな悲劇であるように見える。しかし，筆者はそうは考えない。むしろこの限界は規範経済学研究の領域の巨大さを物語るものではないだろうか。そして合理的選択モデルとは別の新たな方法論の探求へとわれわれを誘っているのだ。筆者はそう確信している。

［謝 辞］

本稿は平成９年度関西大学重点領域研究，規制緩和の総合的研究――望ましい社会システムの観点から――，における盛山和夫氏（東京大学）の報告に刺激を受け執筆した次第である。本稿は2004年３月，早稲田大学にて開かれたコンファレンス，テーマ「脱国境化時代における社会形成理念：公共性の可能性――公平・福祉・効率性をめぐる法学・政治学・経済学の対話（――）」にて発表された。討論者の矢澤正嗣氏（早稲田大学）を始め，多くの参加者の方々から有益なコメントを頂いた。その他，一橋大学経済研究所（2004年１月）大阪大学経済学部（2002年７月）関西大学経済学部（2001年６月）及び神戸大学（2001年９月）でのセミナーでも発表した。本稿は起草してから公刊まで長い歳月を要したが，その分多くの方々から貴重な助言を頂くことができた。これらの方々へ深く感謝したい。

［注］

(1) 経験科学という用語を私はここで「現実の現象を理論的に説明するための科学的な知の営み」という意味で使用している。

(2) 具体的にはアクセルロッド（Axerlod［1984］）とTaylor［1987］の名前が挙げられている。

⑶ 説明は Luce and Raiffa［1957］に従った。

⑷ プレイヤーはなぜナッシュ均衡を選ぶか，に関する理論は幾つものタイプがあるが，ここでは Binmore［1990］らの合理的推論によるナッシュ均衡の選択の考えを援用した。

⑸ この定義はパーソンズ（Parsons［1937］）および Blake and Davis［1964］を参考に，私が練り上げた。おおむね，彼らの定義と同じであるが，１つだけ相違点，特にパーソンズとの相違点，がある。Parsons は規範を「……すべし」という言明そのものである，としている（ように読める）。一方，私はここまで強い定義は採用していない。規範は，仮にそれを言語で表現するならば，「……すべし」という命令法の形をとる，というだけある。些細な違いのように見えるが，規範の定義に関する盛山の議論と比較する際に重要なポイントとなる。この点は第５節で改めて述べる。

⑹ 反対に，規範が成立したからといって，それが遵守されるとは限らない。例えば，「差別はいけないことだ」は，たとえ現実の社会でいかに差別が横行していようと，規範であることに変わりない。要するに行動パターンの定着は，規範の生成にとっての必要条件でもなければ，十分条件でもないのである。

⑺ 規範の「生成」と「成立」は，本稿ではほぼ同じ意味で使っている。両者の違いを挙げるとすれば，前者は現在進行形，後者は現在完了形のニュアンスがあるということぐらいである。規範の「受諾」と「受け入れ」もほぼ同じ意味である。これに対して規範の「合意」は「受諾」よりも意味が狭い。「合意」は２人以上の関係者が規範の受け入れに対し合意している，という意味だが，「受諾」はそれに加えて，自分自身に独語的に語り掛け，納得するという意味も含んでいる。「合意」は他者を必要とするが，「受諾」は必ずしもそうではない。

⑻ 役者のネーミングはギリシャ神話で著名な２人の神の名からとった。アポロは秩序を象徴する。ディオはディオニソスの略で，混沌，カオスを意味する。

⑼ このことは，また行動主義の方法論的破綻を意味しているともいえよう。行動主義は現象を観察可能な要素やパラメーターのみで記述し，説明することを主張するが，当の行動主義者自身は，規範を説明する段になって，サンクションという直接には観察可能とはいえない概念に訴えてしまうという反則を犯しているからである。

⑽ この議論は渡辺［1998］を参考にした。

⑾ 正確には，行動パターンの定着は規範の生成にとって必要でもなければ十分でもない，と論じた。注⑹参照。

⑿ 清水は本稿とは異なった視点から進化ゲームによる制度生成の可能性を吟味している。

⒀ これは繰り返しゲームが「規範への随順行動」を説明する理論としてなら適格である，とした２節での議論と同じである。

参考文献

川浜　登（1999）「法の経済学と限界と可能性」井上・島津・松浦編『法の臨界』第２巻所収，東京大学出版会，209-34 頁。

小林　公（1991）『合理的選択と契約』弘文堂。

清水　和巳（2003）「合理的経済人仮説の終焉：進化と制度生成の視点から」佐藤良一編『市場経済の神話とその変革〈社会的なこと〉の復権』所収，比較経済研究所研究シリーズ 18　法政大学出版局。

盛山和夫（1995）『制度論の構図』創文社。

渡辺幹雄（1998）『ロールズ正義論の行方』春秋社。

Aoki, M.（2001）, *Towards a Comparative Institutional Analysis* MIT Press（瀧澤・谷口訳『比較制度分析に向けて』NHK 出版）.

Axelrod, R.（1984）, *The Evolution of Cooperation*, Basic Books.（松田裕之訳（1987）「つきあい方の科学」HBJ 出版局）.

Blake, J. and K. Davis（1964）, “Norms, values and sanctions” pp.456-84 in R.E.L. Faris ed. Handbook of Modern Sociology. Chicago: Rand McNally.

Binmore, K.（1990）, *Essays on the Foundations of Game Theory*, Blackwell: Oxford.

Fudenberg, D. and E. Maskin（1986）, The folk theorem in repeated games with discounting or with incomplete information *Econometrica*, 54: 533-54.

Gauthier, D.（1986）, “Morals by Agreement”, Clarendon Press, Oxford.

Harsanyi, J.（1955）, Cardinal welfare, individual ethics, and the interpersonal comparison of utility. *Journal of Political Economy*, 63: 309-21.

Harsanyi, J.（1975）, Can the maxmin principle serve as a base for morality? A critique of John Rawls’ theory *American Political Science Review*, 69 pp.594-606.

Hobbus, T.（1651）, “Leviathan” Oxford Clarendon Press（永井・宗片訳（1979）「リヴァイアサン」『ホッブズ』（世界の名著 23）中央公論社）.

Hume, D.（1739）, “A Treatise of Human Nature” Clarendon Press Oxford（大槻訳『人性論』岩波書店）.

Luce, R. D. and H. Raiffa（1957）, *Games and Decisions: Introduction and Critical Survey* New York Wiley.

Mill, J. S.（1863）, “Utilitarianism” In “Utilitarianism and Other Essays” ed. by A. Ryan Penguin, Harmondsworth 1987（伊原訳（1975）「功利主義論」『ベンサム J・S・ミル』（世界の名著 38）中央公論社）.

Moore, G. E.（1903）“Principia Ethica”（深田訳（1973）『倫理学原理』三和書房）.

Parsons, T.（1937）, “The Structure of Social Action. MacGraw-Hill.（稲上・厚東・溝部監訳（1974–89）『社会的行為の構造』全5巻 木鐸社）.

Rawls, J.（1971）, “A Theory of Justice” Cambridge: Harvard Univ. Press（矢島訳（1977）「正義論」紀伊国屋書店）.

Robbins, L.（1949）, “An Essay on the Nature and Significance of Economic Science”, Macmillan Place of Publication（辻六兵衛訳（1957）『経済学の本質と意義』（全2巻）岩波書店）.

Scott, J. F.（1971）, “Internalization of Norms” Prentice-Hall.

Taylor, M.（1987）, “The Possibility of Cooperation” Cambridge Cambridge Univ. Press.