

# 合理性のジレンマ

池 田 信 夫

## Dilemma of Rationality

Nobuo Ikeda

The concept of rationality is often quoted in the literature of social sciences, but its implication is not fully understood until recently. Traditionally, Hume, Smith, and many economists after the "Marginalist Revolution" has built their theories upon the *instrumental rationality*, i.e. utility maximization independent of other social contexts. It was an attempt at imitating classical mechanics that is based on the deterministic differential equations describing the motion of independent matters.

But human behaviour cannot be exactly described by the mechanical model of neoclassical economics. Hayek rejected the concept of Walrasian equilibrium as erroneous "scientism" and Keynes referred to *fallacy of composition* due to the psychological interaction, which was confirmed logically by the game theory.

In contrast to the neoclassical premise that adding up the individual best lead to the social best, game-theoretic solution concept, Nash equilibrium, does not necessarily coincide with the Pareto optimum (social efficiency). The most famous example is the Prisoner's Dilemma in which cooperation is better than defection for each player but cannot be attained spontaneously. Ironically, the players cooperate *if and only if they are not rational*. This paradox is rooted in the logical structure of deductive decision-making, because player A's strategy depends on B's that depends on A's... and so on. As a corollary of Gödel's famous theorem, this kind of self-referential reasoning cannot constitute self-contained system of "right" decision.

In 1980's many scientific enterprise such as artificial intelligence, generative grammar, and Bayesian decision theory tried to formulate human intellect as a logically consistent system, but failed at last. Now many social scientists are re-examining Simon's concept of *bounded rationality* that take into account the complexity of the interaction among people. It corresponds to the new findings in the natural sciences such as deterministic chaos, self-organization and complex adaptive systems.

情報や意思決定の問題を考える場合、その情報を処理する人間の行動をどう想定するかはすべての理論のよって立つもっとも基本的な問題である。これについて新古典派経済学などの在来の社会科学は「合理性」という公準をおいて、いわゆる経済人 (homo economicus) によるメカニカルな情報処理を仮定することでこの問題を迂回してきた。しかし、この合理性とはいったい何なのか、そしてそれは社会全体に拡張した場合に論理的な整合性を持つものなのか、という点についての検討は、実はごく最近になって始まったばかりである。

ここでは合理性の概念を歴史的な視野の中で再検討し、それを経済現象の分析に応用して大きな成果を収めた新古典派経済学と、その鬼子ともいべきゲーム理論を手がかりにして、この概念のきわめて特殊なコンテキストとその論理的な帰結を明らかにする。近代イギリスの個人主義的な風土の中で生まれた手段の合理主義は、近代科学の基礎となった古典力学的な世界像とあいまってモダンな社会の枠組を形成してきたが、今あらわになりつつあるのは、この一見、明快な概念がその基礎に致命的なパラドックスをはらんでいるということである。以下では合理性の概念の持つ宿命的なジレンマを明らかにし、それが社会科学にとって、また社会そのものにとってどういう意味を持つかを考える。

## 1. 手段的合理性

キリスト教においては自己犠牲 (Self-denial) がその神の国に近づく道であると考えられていたのに対して、近代社会の倫理の出発点となったのは、バーナード・マンデヴィルの『蜂の寓話』に始まる自己愛 (Self-liking) の肯定の思想であった。彼はこの「悪徳」を「取引や雇用の基礎」であるとして肯定し、当時の社会に大きなスキャンダルを巻き起こしたが、その思想は後にデイヴィッド・ヒュームによって「理性だけではどんな意志的な行動の動機にも決してなりえない」のだから、「理性は感情の奴隷であり、またそうでなければならない」(『人性論』) という手段的合理性の思想として体系化され、近代社会における合理性の概念の基礎となった。ここでは合理性の概念と目的の概念は切り離され、後者は前者からみちびくことはできず、合理性はその与えられた目的をどう達成するか、という手段についてのみ問題になるに過ぎない。こうしたイギリス経験論の伝統における合理性と目的の分離は、カントの定言命題に代表されるように後者を前者からみちびこうとする大陸の合理主義的な倫理観とは対照的なものであるが、現代において主流となったのは前者である。

価値から自由になった手段的合理性は、アダム・スミスによって経済学の基礎とされ、エゴイズムは「自分の利益を追求することによって、社会の利益を増進せんと思ひこんでいる場合よりも、もっと有効に社会の利益を増進することがしばしばある」(『国富論』) という理由で全面的に肯定された。このいささか便宜的な合理性の概念は限界効用理論によって論理的に「純化」され、人間の意思決定は利己的な目的関数 (効用・利潤) の条件つき最大化問題として定式化され、これを数学的に彫琢することによって、今世紀の新古典派経済学は見かけ上は自然科学に似た厳密さを持つようになった。ここでは合理性は選好の完全性と無矛盾性として定義され、両者はあいまって経済主体による目的関数の最大化を意味する。

しかしこの手段的合理性の概念には、その誕生の時からつきまとっている論理的な弱点がある。それは、この目的が合理性によって基礎づけえないとすれば、それが正しいかどうかは、いったい何によって基礎づけられるのか、という問題である。またヒュームやスミスが強調した、利己的な利益の追求が結果的に社会的に有益であるという命題も、何らかの論理的根拠から出てくる

ものでない限り、当時のイギリスにのみ実現した幸運な偶然である可能性を排除できない。この問題に答えようとしたのが有名なスミスの「見えざる手」であるが、これは実際には、『道徳感情論』と『国富論』の中でそれぞれたった1回、謎のように出てくるメタファーに過ぎず、見えざる手がどうして望ましい結果をもたらすのか、という問題については彼はまったく答えていない。

見えざる手は、彼にとって証明すべき仮説というよりも、すべての基礎になる公理のようなものであった。彼は、個人をチェスの駒のように考えて権力によってあやつろうという考えに反対して、「人間社会という偉大なチェス・ボードにおいては、それぞれひとつずつの駒がそれ自体の運動法則を持っており、それは為政者が押しつけようとして選ぶであろうものとはまったく違うのである」（『道徳感情論』）とのべる。彼の生きた18世紀はニュートン力学が圧倒的な成功を収めた時代であり、ここには社会を力学の運動法則との類比で考え、天体が万有引力による運動の結果として整然と運行するように、人間が自己愛にみちびかれながらも秩序や制度を作り出していく、という古典力学的な世界像が見られる。

この力学的な社会モデルはスミスの場合、ニュートンと同様、本来の理神論的な信仰にもとづくものであったが、その後の科学の世俗化の中で力学的世界像は最初の神という「最初の一撃」を必要としないものとなり、ベンサムやミルによって作り出された「効用」の概念と限界革命をへて19世紀末のワルラスに至って形式的には古典力学に似た自己完結した連立方程式の体系として記述された。それ以後の新古典派経済学はワルラスの一般均衡理論を数学的に彫琢する形で発展し、1950年代のアロウ＝デブブルーらによる一般均衡解の存在証明によって完成したが、そこでわかったのは、均衡が存在するためには、すべての経済主体が永遠の未来にわたる市場の条件を正確に予測して需要や供給を提示し、それを見て「せり人」が両者が一致するように価格を調整する、きわめて特異な調整過程を想定しなければならないということであった。個人が経済全体についての完全な情報を持っており、その行動がすべて巨大なせり人に申告される、というモデルは、新古典派経済学の前提としているはずの分権的な経済の記述とはみなしえず、アロウ自身も指摘するように、むしろ全知全能の計画経済当局の「見える手」による調整過程のモデルである。<sup>1</sup>事実、ワルラスやパレートといった初期の新古典派の理論家は政治的には社会主義者であり、均衡理論はむしろ社会を合理的にコントロールする手段として研究されたのである。

## 2. 合成の誤り

市場メカニズムが多くの場合にうまく分権的な個人をコーディネートしているように見えることは経験的事実であるが、その根拠は前節でみたような力学的なアナロジーとはことなる論理に求められねばならない。そうした代替的な考え方として、新古典派理論のように数学的なモデルとして洗練されることはなかったものの、それに対抗してながく主張され続けてきたオーストリア学派の新自由主義がある。その代表者であるハイエクは、新古典派のような力学的なアナロジーを誤てる「科学主義」であるとしてしりぞけ、個人の主観によって形づくられる自生的秩序（spontaneous order）をその理論の基礎に置いた。この思想の源流を彼はマンデヴィルからヒューム、スミスに至るイギリス啓蒙思想の社会観に求め、他方、大陸のデカルトやルソーなどの合理主義的な社会観を社会主義やファシズムなどの設計主義（constructivism）の温床であるとして、理性や権威によって社会秩序を維持しようとするすべての思想を自由社会の敵と断じた。ハイエクによれば、市場機構の本質は、新古典派の想定するような「希少な資源の効率的な配

分」などというメカニカルな問題にあるのではなく、きわめて限られた情報しか持っていない分権化された個人の利害がどうやって巨視的に調整されてゆくのか、という問題を解くためのシステムであるという点にある。<sup>2</sup>ここでは、経済システムは力学的な運動法則ではなく、いわば分散型の情報ネットワークとしてとらえなおされ、社会に散在する知識が市場メカニズムによって自発的にコーディネートされる、というきわめて現代的な発想が見られる。

この考え方が新古典派よりもはるかに正確に市場メカニズムの本質をとらえていることは、ハイエクとランゲらとの間で行われた「分権的社会主義」をめぐる論争で明らかになった。経済的な問題を新古典派的に「資源の制約のもとでの目的関数の最大化」という最適化問題として定式化した場合、実はそれは必ずしも市場によらなくても実現できるし、むしろもっとすぐれたシステムがありうる。<sup>3</sup>しかし、ハイエクやミーゼスがこうした新古典派的な数理計画理論について指摘した最大の欠陥は、それが情報の効率性という問題を考慮にいていない、という点であった。刻々と変化する経済状態の中で膨大な財についてこうした計算を行うためには、収集しなければならない情報や計算の量は天文学的な大きさとなって実行は不可能であり、分権的社会主義は、資源配分の観点からは市場と同等だが情報伝達の機能において致命的な欠陥を持ち、実現不可能なのである。

ハイエクのいう自生的秩序の概念は、自然科学において最近ようやく注目されはじめた自己組織化の理論の先駆ともいえる。しかし、そのような自生的秩序としての市場が、政府の介入なしでもカオスに転落せず、ずっと秩序を保ちうるという論理的な保証はあるのだろうか？ハイエクの経済学者としての仕事は、もっぱらこういう問題意識にもとづく景気循環論や貨幣論の分野にあるが、残念ながらその貢献は、今となってはさほど重要なものではない。彼の問題意識はあまりに新しすぎて、当時の経済学においてはそれを十分に理論化する用具がそろっていなかったともいえよう。むしろ、現代の資本主義においてはこうした個と全体の予定調和は持続しえず、政府の介入こそがマクロ的なコーディネーションを実現するのだ、というケインズの「見える手」を支持する合理主義的な理論の方がはるかに大きな説得力と現実的な有効性を持ったために、ハイエクの経済理論は忘れられてしまった。ケインズは、基本的にはロックやヒューム以来の個人主義の思想を受け継ぎながら、それがつねに望ましい結果をもたらすとは限らない、という留保条件をつける：

世界は、私的利害と社会的利害とがつねに一致するように天上から統治されているわけではない。世界は、現実の上でも、両者が一致するように、この地上で管理されているわけではない。啓発された利己心は、つねに社会全体の利益になるようにはたらくというのは、経済学の原理からの正確な演繹ではない。<sup>4</sup>

このように、社会を単なる個人の集計と区別する考え方は、「合成の誤り」として『一般理論』の中でも随所で見られる。たとえば不況の時に賃金を切り下げるとは、個々の企業にとっては合理的だが、それが社会全体に集計されると、労働者の所得が落ち、消費が落ち、そして売り上げが落ちて、景気はかえって悪化してしまうわけである。しかし、このような誤りがなぜ均衡理論の想定するように合理的行動によって是正されないのかは論理的には必ずしも明らかではない。

### 3. 囚人のジレンマ

合成の誤りはどうして生じるのだろうか？これをゲーム理論によって定式化したのが、「囚人のジレンマ」と呼ばれる有名なパラドックスである。図の行はプレイヤー A の戦略を、列は B の戦略を示し、欄の中の数字は左から順にそれによる A, B の利益を示す。図のような利得行列はさまざまなケースにおいて見られるが、一般的にいえば「裏切りによる利益が協力による利益より大きい場合には、全員が裏切ることになる」というものである。このように個人の利益の追求の集計が社会にとっての不利益をもたらす合成の誤りが生じるには、各プレイヤーの行動が相手の行動に影響するという相互作用があるために、全体の結果が個々の行動に還元できないからである。一般に個人の最適化の結果であるナッシュ均衡は、それ以上だれの利益も改善できないという意味でのパレート最善とは一致しない。

		B	
		協力	裏切
A	協力	3,3	0,5
	裏切	5,0	1,1

しかし、それでは「道徳的」な行動はどうして生じるのであろうか？この問題に重要な示唆を与えるのがロバート・アクセルロッドの有名な実験である。彼は、上の図のような利得行列であらわされる囚人のジレンマのもとで、さまざまな戦略をコンピュータのプログラムとして表現して互いに組み合わせるトーナメントを行ない、このシミュレーションを数百回くり返して、長期的にはどんな戦略がもっとも高い点数を上げるかを調べた。その結果、1位となったのは、意外にもしっぺ返し (TIT FOR TAT) と呼ばれる、たった4行のもっとも短いプログラムだった。<sup>5</sup>これは「最初は協力せよ、そして後は相手と同じ行動をとれ」、いいかえれば「善人には親切に、悪人には罰を」というきわめて単純な相互性のルールにもとづくものであるが、もっと合理的なしかけをほどこした複雑なプログラムにも長期的にはつねに勝った。短期的には損に見える利他的な行動も、それをくり返しているうちに相手も協力した方が得だということを学習しはじめ、結局、長期的には得になるのである。これは、「善人は結局、得をする」という民間定理 (Folk Theorem) の印象的な例であり、「利己的な遺伝子」から利他的な行動が生じる「協力の進化」のモデルとして、生物学的にも多くの実証研究によって支持されている。

しかし、このしっぺ返し戦略は機械的に相手の行動を反復するものだから、期待効用最大化という基準から見ると合理的ではない。これを改善して、たとえば「一度、裏切ってみて相手が協力するようなら裏切る」というような合理的なプログラムを考えることによってもっと高い点数を得ることができそうだが、実際にはそういう機会主義的な戦略は相手の裏切りを誘発して、結局しっぺ返しほど高い得点を上げることはできない。しっぺ返しの強さの原因は、そうした短期的な利益よりも相手と同じ行動をするという原則を優先することによってエゴイストではないという評判 (reputation) を形成し、相手を自発的に協力に誘導する点にあるのである。

### 4. チューリング・ゲーム

合理性が必ずしも最適の結果をもたらさないようなパラドックスはゲーム理論において数多く

発見され、理論家たちを当惑させるようになった。例えば100回くり返される四人のジレンマを考えると、これが100回で必ず終わるということを双方が知っているとする、彼らは100回目にはいずれも裏切るであろう。なぜならこれが最後のゲームだから、その裏切りに対して報復される心配はないからである。そうすると、99回目には彼らはいずれも次の回には裏切られ、協力しても報いられることはないことを知っているから、どちらも裏切る。そして98回目には……と論理は続き、結局、最初から協力しないことがもっとも合理的である、という結論が出るが、この解は前にも見たように明らかに最適ではない。これは後ろ向き推論の逆説としてよく知られた問題である。

このようなパラドックスが生じる理由は、Aにとって最適な戦略は、Bがそれに対してどう出るかにかかっているが、Bがどう出るかはAがそれに対してどう出るかにかかっている……という相互依存の無限ループが発生するため、一方の当事者にとって何が最適かという答が必ずしも一義的には出せないためである。相手の行動を読んで行動するプレイヤーの戦略は、最適な解を見つけようとするこのループに落ちこまざるをえず、この無限退行が安定な解に収束するという保証は一般にない。これは「私は嘘つきである」という自己言及の代わりに二人が互いに言及しあうことによるパラドックスだから、「相互言及」とでも呼ぶことができよう。

容易に想像されるように、このように複雑な入れ子型になったゲームの構造は、ゲーデルの不完全性定理の応用問題である。ゲーデルの証明のポイントは、数についての演算そのものに数（ゲーデル数）を割り当て、数についての演算についての演算……という自己言及の無限ループが一般に非決定になることを示す点にあった。後にチューリングは、仮想的な万能計算機を想定し、ゲーデル数にプログラムという具体的な姿を与えることによって汎用計算機の基礎を築くとともに、ゲーデルの定理に「別解」を与え、現在のコンピュータの可能性と不可能性を最初に証明した。

このチューリング・マシンと呼ばれる万能の計算機の内部構造はきわめて単純で、紙テープを前後に送ってそこに書かれたデータを順に読みこみ、それを内部の記憶装置に記憶し、記憶した内容を命令に応じて紙テープに書きこむ、という3種類の作業をするだけである。ここで重要なのは、この読み書きの手続きそのものが記憶されたデータによって決まる、というしくみになっている点である。この機械が読むことのできるデータは、「1, 2, 3……」といった文字列だけであるが、「最初に1番目のデータを読み、次にそれに3番目のデータを足せ」というような命令（プログラム）もゲーデル数を割り当てて文字列に帰着できるからデータとして読みこむことができ、それを記憶することによって（記憶容量に限界がなければ）理論的にはどんな複雑なプログラムも実行できる。こうして、データを処理するプログラムそのものをデータと考える自己言及的な構造によって初めて無限の汎用性の実現し、プログラム駆動型の逐次計算機という現在のコンピュータの基本的な構造（不当にもフォン・ノイマン型と呼ばれる）が決められたのである。

ここで2台のチューリング・マシンX, Yによってゲームの解を計算する場合を考えよう。記憶容量が無限大であり、互いに相手の戦略について完全な情報を持つとすると、彼らはつねに解を求めることができるだろうか？残念ながら、その答は否である。これはチューリングの論法を応用することによって証明できるが、その大筋をごく直観的という次のようなものである：

今、プレイヤーXはその予測 $p$ にもとづいて行動 $q$ を決め、Yは予測 $\pi$ にもとづいて行動 $\rho$ を決めるとすると、XはYの行動を完全に予見できるから、 $p = \rho$ 、同様に $\pi = q$ となる。とこ

ろが、ここでXのプログラムが $q = \pi + 1$ となるように作られているとすると、どういうことが起きるだろうか？まず初期状態において、Xは自分に与えられたデータとYの戦略についての情報をもとに $p$ を計算して行動 $q$ を決め、同時にYはXの行動を $\pi$ と予測して行動 $\rho$ を決める。しかし、Yの予測はつねにXによって裏切られる宿命にある。なぜなら、XはYの予測 $\pi$ を完全に予測してつねにそれに1を加えた解を行動とするように作られているからである。これに対してYもXの戦略を完全に知っているから $\pi' = \pi + 1$ となるように予測を修正するが、Xはこれを完全に予測できるから、 $q' = \pi' + 1$ となるような行動をとり、これをYが予測して……というようにYの予測とXの行動は相互言及の無限ループに入ってしまう、2台のマシンは答を出すことなく永遠に計算を続ける。このように相手の予測をつねに「ずらす」ようなパラドキシカルなプログラムXを構成することはどのようなチューリング・マシンに対しても容易だから、つねに意味のある解を出し、かつ有限時間内に計算が終わるような合理的なプレイヤーは存在しない。<sup>6</sup>

いかに計算能力が高くても、計算を終えることのできないコンピュータというのは、あまり役に立つとは思われないから、問題はこの計算をどこかで終わらせるための一般的な停止ルールがあるのかどうかということである。そのもっとも簡単な候補は、一定の時間がたったら計算を打ち切る、というものだが、こういう乱暴なルールによっては、マシンが停止した時点でなんらかの答が出ているという保証さえない。もうひとつの考え方は、ただ相手の裏をかくだけのような無意味なプログラムを除外し、合理的（最適化）行動だけを対象とすることによって停止を保証しようというものだが、この場合でも合理的ではあるが無限に循環するようなケースを必ず作ることができ、問題は解決しない。これは、論理的に一貫しているという無矛盾性かつつねに正しい答を出せるという完全性は両立しないというゲーデルの定理からみちびかれる系である。

## 5. 計算複雑性

このような抽象的な不完全性が具体的にどんな意味があるのか、といふかる向きもあるが、期待効用の最大化計算にはもっと実的な困難もある。以上の議論は無限の記憶容量を想定しているが、実際にはこれは有限だから、たとえ論理的に解ける問題であっても、じゅうぶん短い時間内で解けるかどうかはわからないのである。たとえば将棋のような確率的要因を含まない1対1のゼロ・サム・ゲームは決定可能であることがツェルメロによって証明されているが、実際には可能な手の組み合わせはべき乗で爆発的に増えるから、最善手を計算によって求めることは、論理的には可能であっても有限の時間内では不可能である。とことん最適の手を追求して長考していたら時間切で負けてしまうから、どこかで思考を打ち切って行動に移らなければならないが、どこで思考を打ち切るかという停止ルールは、思考そのものからは出てこないのである。

計算量の問題がゲームのみならず意思決定にかかわる理論においてきわめて重要な問題であることが、1980年代になってようやくわかってきた。例えば、人間の意思決定を機械的なアルゴリズムで模倣しようとした人工知能は、人間の情報処理の含んでいる桁違いの複雑性に直面してゆきづまってしまったが、その結果としてわかったのは、こうした意思決定にかかわる計算量の大きさは本質的なものであって、アルゴリズム的な工夫によって克服することはできない、ということである。計算量の理論においては、将棋のように組み合わせがべき乗で増えていく問題は、計算量が多項式のオーダーを超える（Non-Polynomial）という意味でNPで完全問題とよばれ、このような問題は単純な問題に分解して解くことができない。人間の情報処理にかかわる言語や

視覚処理などの多くの問題はNP完全であり、しらみつぶしに組み合わせを調べあげていく人工知能的な方法では、原理的に解けないのである。この例としてよく知られているのが、人工知能の挫折の大きな原因となった「フレーム問題」である。

こうした論理的な決定不可能性と計算複雑性 (algorithmic complexity) の問題は、コンピュータによって人間の知能をシミュレートできるという80年代初期の意思決定理論の夢を打ち砕き、ゲーム理論においても、永遠の未来までにわたって相手の行動を予測して期待効用の最大化を行う超合理主義的な解法は、意味のある解を出すという点ではほとんど失敗に終わってしまった。多人数のくり返しゲームにおいては、戦略の数と人数とくり返しの回数はすべてべき乗になるから、たとえば価格が一定でないような寡占企業のモデルでは、知らなければならぬ情報は莫大なものになり、きわめて単純なケースでも実際に最適な解を求めるのは不可能である。市場とは、ハイエクのいうように、価格という指標に情報を集約することによってこのような相互作用による組み合わせの爆発を防ぎ、計算複雑性を減らすためのしくみなのである。

囚人のジレンマの重要な意味は、協力の成立する条件が自己の利益の最大化という意味での合理性ではないという点である。もしもプレイヤーが両方とも合理的であり、かつ互いに合理的であることが共通知識 (common knowledge) になっている場合には、囚人のジレンマは解くことができない。相手が合理的だということを知っている以上、自分は裏切ることが合理的な時には裏切りし、それを相手は知っている、ということを知っている……という疑心暗鬼のループが生じてしまうからである。合理的な主体はしつぱ返し戦略をまねることによって単純なナッシュ均衡の解 (相互裏切り) よりも高い利益を得ることができるが、そのためには少なくとも相手が裏切らない (合理的ではない) ことを信頼しなければならない。つまり、協力が成立するには互いが合理的であることが共通知識になっていないことが必要なのである。<sup>7</sup>しかし、意思決定を合理性によってみちびくことが不可能だとすれば、それはいったい何によってみちびかれるのか？

ヒュームやスミスを悩ませたのも、行為に普遍的な基準があるのかどうかという問題であった。「もし私的な利害や評判に対する関心がすべての正直な行為の正当な動機であるといわれるのなら、その関心がなくなれば、正直はもはやありえない」(『人性論』) ことになってしまうからである。彼らは結局、共感といういささか曖昧な概念によって問題をごまかしてしまうが、これを論理的につきつめたカントは、行為の絶対的な正当化のためには認識論から追放した神を呼び戻さなければならないことに気づく。形而上学的な実体を否定してしまうと個別の状況を越えた絶対の善は存在しなくなるから、「……せよ」という絶対的な命題 (定言命題) をみちびくには、永遠に不滅の魂や至高の神という思考の停止ルールを置かざるをえず、そしてこれはまさに論理ではなく、絶対的な価値にたいする信仰 (共通知識) によって支えるしかないものである。<sup>8</sup>

## 6. 限定された合理性

新古典派経済学が追究した超合理主義的なモデルによる人間行動の記述は、ある意味では自然科学において大きな成功をとげた古典力学的な合理性の概念を人間そのものにも適用しようという英雄的な試みであった。それは、1970年代から80年代にかけて行動科学のさまざまな分野で同時に進められた人工知能や生成文法、ベイズ的な統計的決定理論などによる人間行動のメカニカルな理論化とも共通する、理性が勝利を取めた時代にふさわしい壮大な実験であった。しかし、これに先立って行われた社会そのものの理性化としての社会主義が最終的に崩壊したのとはほぼ時



期を同じくして、こうした試みはほぼ全面的に挫折し、人間の本质をとらえる上で力学的なモデルが基本的に無効であることを決定的に証明した。

ところが同様の事態は、実は彼らがお手本とした自然科学そのものにおいても進行していた。ニュートンを出発点とした近代の合理主義的な科学は、今世紀の相対論や量子力学によって「拡大された合理性」を獲得したが、それはニュートン以来、古典力学につきまとうパラドックスをいまだに解くことができない、とイリヤ・プリゴジンのはのべる：「古典物理学には時間のパラドックスがある。巨視的な過程は不可逆性に結びついているが、基本的な力学は時間について可逆的な決定論的な法則である。」<sup>9</sup> 20世紀の科学においてもっとも輝かしい成功を収めた量子力学はシュレーディンガーの波動方程式という単純な線形の微分方程式にあらゆる物理現象を帰着させることに成功し、その反例は今なお宇宙全体にひとつとして見つかっていない。しかし、この万能の波動方程式は、ニュートン力学や相対論と同様、時間が後ろ向きに流れないという自明の事実を説明することができないのである。

1970年代から発見されるようになった「決定論的カオス」と呼ばれる奇妙な現象は、こうした拡大合理性の有効性に疑問を投げかけはじめた。そこでは、決定論的な微分方程式から出発して非決定論的なカオスが生じ、しかもそれは時間について非対称になる。こうした微分方程式の特徴は、それが非線形だということである。非線形系においては、カオス的なふるまいは個々の要素の動きに分解することはできず、要素どうしの相互作用という還元不可能な要因が決定的な役割を果たす。ここにも古典力学の還元主義的な方法によってとらえきれない本質的な複雑性があるわけである。このような複雑性にかかわる現象は、80年代に入って自然科学のあらゆる分野において発見され、「複雑性革命」とでもいうべき大きな流れになろうとしている。

あらためて考えると、古典力学は基本的に相互作用を捨象した孤立系の科学であり、三体問題を見てもわかるように、相互作用が入ったとたんに見かけ上の整合性を失ってしまう。あらゆるプレイヤーの無限の将来にわたる戦略の確率分布についての完全な情報にもとづいて期待効用の最大化計算をして決定を下すベイズ的な決定理論は、その数学的なソフィスティケーションとは裏腹に、実際の経営的意思決定の支援という本来の用途にはほとんど役立たなかった。なぜなら、そこには「相手の出方を見て戦略を決める」というインタラクティブな行動や「いったんやったことはやりなおせない」という時間の不可逆性がまったく考慮されていないからである。このようなプレイヤーどうしの相互作用を考えたとたん、超合理主義を支える論理の段階は消え去り、エッシャー的な無限循環の階段に姿を変えてしまうのである。

挫折した拡大合理性の夢にかわるものとしてプリゴジンが提案するのは、ハーバート・サイモンの「限定された合理性」(bounded rationality)である。これは彼がベイズ的な客観的合理性をしりぞけて行動科学の基礎に置こうとした「合理性を求めるが、その限界内に制約されている」<sup>10</sup>という行動様式である。ベイズ的な超合理主義理論は情報の収集や意思決定そのものにとりなうコストをゼロと仮定しているが、例えば買い物をする時、人は少しでも安いものをさがして町中をさがしまわったりはしない。そんなことをすること自体の情報コストの方が大きくなってしまふからである。こうした計算の手續きに要するコストを制約条件に加えて、計算複雑性を最小化するための実用的なルールが日常つかわれている「どんぶり勘定」(rule of thumb)である。

しかし合理性が限定されているということが、実はわれわれが実時間で決定を下し、また他人と協力することを可能にしている。これまでに見たように、囚人のジレンマのもとでは合理的な

個人（であることを互いに知っている個人）どうしても協力が成立しないが、たとえば「次の回以降のゲームのことは考えない」という制約をプレイヤーの決定に課すと協力は容易に達成される。先のことはわからないが、とりあえず裏切ると裏切られることだけは間違いないからである。

ベイズ的合理性がゲームを始める前に最善の戦略を求めてあらゆる可能性を考えつくすのに対して、限定合理性はある意味では行き当たりばったりとその時点で得られる情報の範囲で決めて実際に行動し、間違えたら修正する、というやり方をとる。したがって前者の決定が膨大なパラメータを含んだ線形の微分方程式の解として決まるのに対して、後者ではパラメータの数は最小化されている代わりに相手との相互作用に影響を受ける非線形性を持ち、一般に解析的に解くことはできない。しかし、最近ではアクセルロッドのトーナメントのようにコンピュータを使った数値シミュレーションによってこうした相互作用を分析することが可能になり、「進化ゲーム」の理論が急速に発展している。ゲーデルが予告したように、合理性は宿命的に不完全であり、すべてをあらかじめ合理的に決めようというベイズ的合理性は挫折せざるをえないが、それはまた複雑系のロジックとなる新しい合理性への出発点でもある。

## 7. 結 論

マンデヴィルに始まり、スミスによって基礎が築かれ、そして20世紀に新古典派によって社会科学の中で唯一、自然科学に似た装いをもった「科学」として確立された経済学は、まさに近代社会を支える個人主義のイデオロギーの代表であった。そこでは人間は純粋に合理的＝利己的な個人として理想化され、他の人間は利益最大化のための手段であり、自然環境と同様の「与件」に過ぎない。個人はこのような制約条件のもとで期待効用を最大にする条件つき最大化問題を解いて無限の未来にわたるまでの行動を決め、社会はそうした原子論的個人の欲望の均衡状態として表現される。他ならぬデジタル・コンピュータの父であるジョン・フォン・ノイマンによって作られたゲーム理論は、こうしたメカニカルな合理性の概念をもっとも徹底した形でモデル化したものであった。

しかし、このように極限まで拡大された合理性の追求の果てに今、それが発見したのは、皮肉なことに「自律的な合理的個人」という神話の自己否定であり、個人間の相互作用を考えずには決定そのものが不可能である、というパラドックスである。そして、このように相互作用がモデルに入ったとたんに他人の行動は与件ではなくなるから、新古典派理論のエレガンスを保証していた線形性は維持できなくなり、人間の行動は流体における分子のように、ある主体の動きが別の主体に影響し、そしてそれがまたもとの主体にフィードバックされ……という非線形の複雑なループを形成して、均衡状態は一意的に定まらなくなってしまう。

ただ、このような複雑系においても、まれに「非平衡の平衡」が実現して維持されることがある。それは、生態系や生命のように系に対して定常的にエネルギーが供給され、しかもその流れが系の中で一定に制御される場合である。秩序や制度は、ハイエクのいうように、このような散逸構造として形成されるものであるが、それが本当のカオスになって発散してしまわないための条件はきわめて微妙なものであり、またかりに自生的な秩序が維持されたとして、それが各人にとって望ましいものになるという先験的な保証はない。市場による自己補強的な秩序こそがいかなる人為的な強制力による秩序よりも好ましい、というハイエクのドグマを彼自身はついに証明できなかったし、それはゲーム理論の明らかにしたさまざまなパラドックスによって、少なくとも論理的には反証されてしまった。

社会科学が価値や倫理を語らなくなって久しい。20世紀の主流であった、あらゆる存在や認識を記号に帰着させてしまう相対主義にとっては、それはもはや死語であり、新古典派経済学に代表される操作主義的な「部分的社会工学」はこのような倫理的シニシズムによってその論理整合性を守ってきた。しかし今ははっきりしてきたのは、こうした記号一元論が現実からの逃避を正当化し、結果的には既存の価値を追認してきたという事実である。意志や感情などの問題を回避して知識と論理だけによって人間の知性を模倣しようとした人工知能や認知科学の挫折の後、今あらためて意思決定と調整の問題が社会科学のもっとも重要なテーマとなりつつある。20世紀の思想的な対立軸をなしていたイデオロギーが根底から崩れようとしている今、求められているのは、社会を透明な記号の戯れではなく複雑な力の相互作用としてとらえる新しい社会科学であり、そのためにまず必要なのは、合理的な個人というフィクションを捨て、価値や制度の問題に正面から向き合うことではあるまいか。

#### 注

- (1) K. J. Arrow, "Economic Theory and the Hypothesis of Rationality" in J. Eatwell et al (eds.), *New Palgrave*, Macmillan, 1987.
- (2) F. A. Hayek, "The Use of Knowledge in Society" (1945) in *Individualism and Economic Order*, University of Chicago Press.
- (3) ランゲやラーナーが提案した分権的社会主義のモデルによれば、計画当局がせり人をシミュレートして各工場が報告する価格と供給量を調整することによって理論的には市場によるのと同じ資源の最適配分が達成され、それは実際に市場で試行錯誤するよりも無駄がない分だけすぐれている。cf. O. Lange, *On the Economic Theory of Socialism*, University of Minnesota Press, 1938.
- (4) J. M. ケインズ「自由放任の終焉」『世界の名著』69 (中央公論社) 151ページ。強調は原文。
- (5) R. Axelrod, *The Evolution of Cooperation*, Basic Books, 1984. ただし、一般的な  $n$  人ゲームにおいてはしっぺ返しは必ずしも最強の戦略ではないから、アクセルロッド自身もいうように、この結果を社会一般に拡大解釈すべきではない。
- (6) K. Binmore, "Modeling Rational Players I" in *Essays on the Foundation of Game Theory*, Blackwell, 1990.
- (7) D. Kreps, P. Milgrom, J. Roberts and R. Wilson, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma", *Journal of Economic Theory*, 1982 (27) : pp. 245-252.
- (8) 個人の価値を集計することによって社会的な意思決定を「民主的に」行うことが論理的に不可能であることはアロウの「一般不可能性定理」によって証明された。cf. K. J. アロウ『社会的選択と個人的評価』(日本経済新聞社)
- (9) I. Prigogine, "Bounded Rationality : From Dynamical System to Socio-economic Models" in R. H. Day et al (eds.), *Nonlinear Dynamics and Evolutionary Economics*, Oxford University Press, 1993.
- (10) H. A. サイモン『経営行動』(ダイヤモンド社) 104ページ。

(文教大学情報学部非常勤講師)