

「囚人のジレンマ」のパラドックス*

水戸康夫 進本眞文 内藤徹 八島雄士

〈要約〉

繰り返しのない「囚人のジレンマ」型のゲームにおいては、ジレンマが生じるはずであるが、「実験」の結果、必ずしもジレンマは発生しないことを確認した。発生すべきジレンマが生じないという「囚人のジレンマ」のパラドックスが生じる理由として、利得にかかわらず「協力」を選択するグループと、「協力」を選ばないことによる評判悪化の可能性を気にするグループの存在に注目した。

キーワード：評判、利得水準、囚人のジレンマ、パラドックス

1 はじめに

「囚人のジレンマ」に関する実験は多く行なわれてきた¹。初期（1950年代後半）には、実験のための実験を行なっているのではないのか、という疑問を持たれるほど多くの実験が行なわれた。しかし、そのような実験データの蓄積からは理論を生み出せないという反省から、1960年代後半以降には被験者の行動の意味が検討されるようになってきた。その結果として、利益の和の向上を目指す「協力的」動機、自己の利益の向上を目指す「個人主義的」動機、相手と自己の利益の差の拡大を目指す「競争的」動機、相手の利益最小化を目指す「攻撃的」動機、相手の利益最大化を目指す「愛他的」動機が提唱され、実験されるようになった。

「囚人のジレンマ」において、1回限りのゲーム（one-shot game）における、ゲーム理論に基づく均衡戦略は、「非協力行動」、「非協力行動」のみであり、繰り返しゲーム（repeated game）の均衡戦略は複数存在する。1回限りのゲームにおけるプレイヤーの最適行動は単純な

*本稿の執筆に関わった4名は、九州共立大学経済学部ゲーム理論研究会に属し、本学経済学会から研究助成をいただいた。記して感謝の意を表したい。

ので、経済学部の学生であれば直観的に理解できるはずであり、「囚人のジレンマ」に関する実験（以下、「実験」と呼ぶ）において100%かほぼ100%のプレイヤーは、「非協力行動」を選択することが予想された。

ところが山岸（2000）によれば、山岸の行なった1回限りのゲームにおいて、56%が「協力行動」を選択しており、世界各国での実験においても3～6割が「協力行動」を選択すると報告している²⁾。九州共立大学経済学部ゲーム理論研究会（以下、「研究会」と呼ぶ）において、理論的予想と山岸報告のどちらが、九州共立大学経済学部学生の行動をより適切に説明できるのかについて議論が行なわれ、「研究会」で「実験」を行なうこととなった。

結論を先取りするならば、「研究会」での「実験」結果は、山岸報告に近いものであった。山岸（2000）は、「協力行動」が生じるのは、自分の利益以外の動機が原因であり、「例えば友情にもとづいて相手の利益のみを考える純粋に愛他的な動機とか、こういう場合には協力するのが正しい行動なのだ」という内面化された社会規範ないし良心にもとづく動機などが考えられ³⁾」としている。上述の山岸（2000）の分析は、山岸俊男氏が社会心理学者であり、社会的ジレンマに関する分析を行なうという山岸（2000）の目的から考えれば、妥当なものである。「研究会」としては、「実験」における条件を微妙に変化させて、複数の「実験」を行ない、さまざまなデータを入手することで、山岸分析をより深めることを目指した。「実験」を通していくつかのことが明らかとなったが、明らかにされるべき多くの課題も残った。本稿では、入手したデータを基にして行なった「研究会」における議論を示すこととし、残された課題を今後の「実験」によって明らかにしていきたい。

2 「実験」

2.1 「囚人のジレンマ」

本節では、ゲーム理論でもっとも有名であるものの1つである「囚人のジレンマ」と呼ばれる現象を説明する。いま、2人のプレイヤー（学生）がそれぞれ交渉が出来ない状態にある。プレイヤーはそれぞれ質問を受け、自分の選択と相手の選択の組み合わせによって得点が確定するものとする。ゲームは1回のみとし、各プレイヤーが選択することが出来る戦略は、「協力」か「裏切る」である。双方の学生が協力するを選択した場合、2人のプレイヤーに6点ずつが与えられる。

次に、1人が裏切る、もう1人が協力するという戦略を採った時、裏切るを選択した学生は8点を獲得し、協力するを選択した学生は1点を得る。また、双方が裏切るを選択した場合、双方共4点ずつを獲得する。したがって、利得表は以下の通りになる（表1）。

この時、各プレイヤーが選択しうる戦略と均衡利得は、どのようなものになるであろうか。いま各プレイヤーが利得にのみ興味があると仮定する。プレイヤー2の戦略を所与にした時の

表1 「囚人のジレンマ」のゲーム

		プレイヤー2の行動	
		「協力」	「裏切る」
プレイヤー1 の行動	「協力」	(6, 6)	(1, 8)
	「裏切る」	(8, 1)	(4, 4)

利得：(プレイヤー1、プレイヤー2)

出所)「研究会」作成。

プレイヤー1の行動を考える。プレイヤー2が協力するという戦略をとったと仮定しよう。この時プレイヤー1は「協力」を選択すれば6点を得るが、「裏切る」を選択すると8点を得られるために、「裏切る」を選択するであろう。次にプレイヤー2が「裏切る」という戦略をとったと仮定しよう。この時、プレイヤー1が「協力」を選択すると1点しか得ることができないが、「裏切る」を選択することで4点を得ることができる。したがって、プレイヤー1にとって「裏切る」という戦略は支配戦略となる。さらに、プレイヤー1の行動を所与にした時のプレイヤー2の行動も同様に推論できる。結果、各プレイヤーは相手の戦略に関わらず「裏切る」という戦略をとるため、均衡戦略はともに「裏切る」を選択するものになる。すなわち、2人とも4点ずつを得ることになる。

2. 2 「実験」結果

「研究会」では、平成13年1月に「囚人のジレンマ」ゲームを行なった。第1番目の「実験」は、資料1を「交通論」の講義時に配り、3回生42人、4回生19人の計61人を対象とした。学生同士の相談を禁じ、**「実験」**を行なうに前に、以下の説明を行なった。集めた**「実験」**用紙は離れた列の学生の用紙と無作為に組み合わせ、各学生の解答は他の人にはわからないように回収し、その結果は公表しないことを明言した。これは、以下のことを考慮するからである。知人・友人に対しては、「協力行動」が選択されやすいことが予想される。「囚人のジレンマ」の設定を踏襲するためにはこのことを避ける必要があるため、同じ講義を受けただけの無関係な人とゲームを行なうことを明確にした。加えて、無関係な人とのゲームであっても、「非協力行動」という戦略を選択することが、友人・知人に観察されることを避けたいと考える学生が存在していれば、「協力行動」が選択されやすいことが予想される。このことを避けるため、誰がどのような選択を行なったかがわからないように工夫した。

「研究会」での**「実験」**においては、「協力行動」として「協力」と、「非協力行動」として「裏切る」という戦略を設定した。表2によれば、「協力」という戦略を選択した学生は11人(18.0%)、「裏切る」という戦略を選択した学生は50人(82.0%)という結果が得られた。つまり、(「裏切る」、「裏切る」)という均衡戦略から逸脱して、「協力」を選択した学生が、18.0%に達した。

したがって、「実験」結果は、山岸報告に近いものであった。少なくない比率(約5分の1)の学生が理論的予想から逸脱したが、それは「交通論」の受講学生にだけ見られる現象であり、例外的な学生が多く存在したからかもしれない。理論的予想から逸脱する学生が、恒常的に存在しているか否かを検証するため、「管理会計論」の受講生に対しても、同じ「実験」方法で、同じ利得行列での「実験」を行なった。資料1を2回生42人、3回生4人、4回生1人の計47人に配付した。その結果、「協力」という戦略を選択した学生は9人(19.1%)、「裏切る」という戦略を選択した学生は38人(80.9%)という結果が得られた。やはり、「裏切る」、「裏切る」という均衡行動から逸脱して、「協力」を選択する学生は、約5分の1(19.1%)に達した。

表2 資料1に基づく「実験」結果 (単位：%、人)

	「協力」比率	「裏切る」比率	被験者数
「実験」1 (「交通論」)	18.0	82.0	61
「実験」2 (「管理会計論」)	19.1	80.9	47

出所)「研究会」作成。

「実験」の第1の目的は、九州共立大学経済学部学生の行動は、理論的予想と山岸報告のどちらに近いものであるかを検証することであった。1番目と2番目の「実験」から、逸脱が少なくない比率で存在することが明らかとなった。つまり、九州共立大学経済学部学生の行動は、山岸報告の方に近いものであった。そこで、理論的には存在しないはずの逸脱が起こる理由の解明が、「実験」の新たな目的となった。そこで3番目の「実験」では、利得構造は変更しないが、利得の数値を変更してみた。学生がどの程度合理的に行動しているかを確認するためである。学生が経済的な観点で合理的に行動していれば、利得構造は変化していないので、行動は変化しないことが予想された。

3番目の「実験」では、「多国籍企業論」を受講している3回生43人、4回生6人、科目履修生1人の計50人に資料2を配付した。その結果、「協力」という戦略を選択した学生は13人(26.0%)、「裏切る」という戦略を選択した学生は37人(74.0%)という結果が得られた。「裏切る」という均衡行動から逸脱する学生は、約4分の1(26.0%)に達した。3番目の「実験」結果が、1、2番目の結果と相違するか否かに関して、この結果だけでははっきりしたことは言えない⁵⁾。

1～3番目の「実験」によって、理論的予想から逸脱する学生が少なからず存在することを確認できたが、その理由については不明なままである。そのため、「研究会」は「実験」4～5において、どちらの戦略を選択するのかと同時に、その選択理由についても述べさせることにした。

4番目の「実験」は、「ことばと文化特講B」を受講している3回生52人、4回生8人の計60人に資料2を配付した。その結果、「協力」という戦略を選択した学生は30人(50.0%)、「裏切

る」という戦略を選択した学生は30人(50.0%)という結果が得られた。「裏切る」という均衡行動から逸脱して、「協力」を選択する学生は、2分の1(50.0%)に達した。この結果は、1～3番目の「実験」結果と明らかに相違する。もし、選択理由を書かせたことが原因で、結果が大きく相違したとすれば、「実験」方法はさらに注意深く検討しなければならないし、4人が行なう「実験」方法の統一についても細心の注意を払う必要がある。

「研究会」では、「実験」4において、「協力」を選択した被験者が挙げた理由を整理し、以下の2つのグループの存在に注目した。第1のグループは、利得にかかわらず協力を選択するグループであり、第2のグループは、協力を選ばないことによる評判悪化の可能性を気にするグループである。そこで「研究会」では、評判に影響を与える方法で「実験」を行なうことで、評判の影響を探ることにした。「実験」1～4では、集めた「実験」用紙は離れた列の学生の用紙と無作為に組み合わせ、各学生の解答は他の人にはわからないように回収し、その結果は公表しないことを明言するなど、評判に影響を与えないように「実験」を行なってきた。「実験」5では逆に、学生の解答が他の人に見られる可能性を示唆し、相手プレイヤーが特定できるようにした。

5番目の「実験」方法は、次のように変化させた。相手プレイヤーが誰であるか特定できるようにし、「裏切る」を選択すれば、そのことが一部の人間には伝わるようにした。具体的には、列の前の人と後ろの人がペア(例えば、列の1番前の人と2番目の人、3番目の人と4番目の人、5番目の人と6番目の人がペア)となってゲームを行ない、後ろの人は自分の選択と選択理由を書いたレポート用紙と、前の人レポート用紙を回収し、提出するようにさせた。つまり、ペアになった前の人選択は、後ろの人には観察可能であった(前の人には後ろの人の選択は観察できない)。後ろの人には、ペアになった前の人選択を観察可能とはいえ、回収して提出するまでの間にその内容を読み取ることは、実際上困難であった。しかし、前の人には自分の選択が見られることを前提として、選択をすることが予想された。したがって、協力を選ばないことによる評判悪化の可能性を気にする学生が多ければ、「協力」は大幅に増えるだろうことが予想された。5番目の「実験」は、北九州大学経済学部の「国際貿易論」を受講している3回生32人、4回生19人の計51人(奇数の列がいくつかあった。奇数の列の最後の1人には自分自身とゲームすることを伝えたが、誰が最後の1人になるかはあえて伝えなかった。)に対して、資料2に基づいた「実験」を行なった。その結果、「協力」という戦略を選択した学生は26人(51.0%)、「裏切る」という戦略を選択した学生は25人(49.0%)という結果が得られた。「協力」を選択する学生は、約2分の1(51.0%)に達し、北九州大学経済学部学生の選択においても、本学学生の選択との類似性が見られた。

4番目の「実験」では、選択理由を書かせたことにより、協力を選ばないことによる評判悪化の可能性に強く反応する学生が「協力」を選択したと考えられる。つまり、評判低下を強く恐れる学生ならば、選択理由を書かせるとい程度の評判悪化の小さな可能性によって、選

択は変化する。そこで、5番目の「実験」では、4番目の「実験」よりも「協力」が多く選択されるように、「実験」方法の変更（4番目の「実験」よりも評判悪化の可能性を高めること）を行なった。しかし、5番目の「実験」結果は4番目とほとんど変わりがなかった。被験者の母集団が相違（北九州大学経済学部学生と九州共立大学経済学部学生）することの影響を考慮しなければならないことと、今回の「実験」では、評判低下を恐れる程度に関する調査は行なっていないため、ここでの結果に関する分析は行なえない。次回以降の「実験」では、評判低下を恐れる程度に関する調査を同時に行ない、評判低下を恐れる程度が中程度である学生比率の低さが、4番目と5番目の「実験」結果を類似させたか否かの検証を行ないたい。

表3 資料2に基づく「実験」結果 (単位：%、人)

	「協力」比率	「裏切る」比率	被験者数
「実験」3（「多国籍企業論」）	26.0	74.0	50
「実験」4（「ことばと文化特講B」）	50.0	50.0	60
「実験」5（「国際貿易論」）	51.0	49.0	51

出所)「研究会」作成。

3 まとめ

「囚人のジレンマ」の利得行列のもとで、両プレイヤーにとっては（「協力」、「協力」）が実現する方が望ましいが、均衡戦略は（「裏切る」、「裏切る」）であり、理論で予想するような均衡戦略が実現するのなら、ジレンマが生じる。そして、「囚人のジレンマ」における均衡戦略が実現している、と思い込んでいるゲーム研究者は多い。しかし先行研究によれば、必ずしも（「裏切る」、「裏切る」）という均衡戦略は実現していない。ゲーム研究者の多くは、アクセルロッドによる、コンピューター・プログラムによる「囚人のジレンマ」の繰り返しゲームの実験については知っていても、社会心理学の分野で多くの実験が行なわれていることを知らないであろう。多くの実験が行なわれていることを知っている研究者でも、1回限りのゲームにおいて、「協力」を選択する被験者が3～6割も存在していることを知っている研究者は少数であろう。「研究会」では、「実験」を行なって、理論的な予想が実現することを確認しようとした。

「実験」の結果、先行研究で示されたように、（「協力」、「協力」）が多く見られた。つまり、理論ではジレンマが生じるはずであるが、ジレンマの生じないケースが多く見られた。発生すべきジレンマが生じないということに対して、「研究会」では、「囚人のジレンマ」のパラドックスと呼ぶことにする。社会心理学の分野では、「協力」が選択される理由についての分析が行なわれているが、経済学の分野では「囚人のジレンマ」のパラドックスが存在しているという事実がほとんど知られていないため、分析されてこなかった。

「実験」1～3によって、「協力」を選択する被験者が実際に少なからず存在すること（約4分の1から5分の1）を確認した。そこで、「協力」を選択する理由や選択基準を明らかにしようとして、「実験」4～5においては、選択理由も書かせた。書かれた選択理由に基づき、「研究会」では2つのグループに注目することにした。第1のグループは、利得にかかわらず「協力」を選択するグループであり、第2のグループは、「協力」を選ばないことによる評判悪化の可能性を気にするグループである。本稿では第2のグループに関する「実験」を行なったが、明快な結論を得ることはできなかった。

「研究会」において、2つのグループに注目することを議論した時点では、ほとんどの講義が終了していた。そのため、第1のグループに関する「実験」を行なうことができず、また、第2のグループに関する「実験」は行なったが、更なる「実験」を必要とする。また、「実験」に際して、「協力」と「裏切る」という戦略を提示したが、「協力」という言葉にプラスのイメージ、「裏切る」という言葉にマイナスのイメージがあることにより、「実験」結果に何らかの影響を与えた可能性がある。今後の「実験」においては、選択肢A、選択肢Bとするべきであろう。選択理由が述べられている「実験」用紙の中には、利得水準によっては、別の選択を行なうかもしれないと書かれたものが何枚かあった。ゲームの利得構造に変化はなくても、利得水準が変化すれば、選択が変化するの否かは、1番目と2番目の「実験」に対する、3番目の「実験」によって確かめようとしたが、明快な結論は得られていない。更に「実験」を行ないたい。今回は報告しなかったが、「研究会」では利得行列を少し変化させ、弱い支配戦略のもとでのゲームも行なった。被験者数が少なく、「研究会」が分析するための時間を十分持てなかったため、報告しなかった。弱い支配戦略のもとでのジレンマゲーム、「チキンゲーム」、「両性の闘い」等の実験も今後の課題としたい。

資料 1

学年 年 組 学籍番号
名前

説明

いま、あなたとあなたにはわからない学生がペアになっています。ここで皆さんにある選択をしてもらいます。内容はあなたが誰だかわからない相手を裏切るか裏切らないかです。あなたが選んだ選択と相手が選んだ選択によってあなたのボーナスの出席点が決定します。

		相手の行動	
		協力	裏切る
あなたの行動	協力	(6 , 6)	(1 , 8)
	裏切る	(8 , 1)	(4 , 4)

かっこないの見方：(あなたの点数、相手の点数)

資料 2

学年 年 組 学籍番号
名前

説明

いま、あなたとあなたにはわからない学生がペアになっています。ここで皆さんにある選択をしてもらいます。内容はあなたが誰だかわからない相手を裏切るか裏切らないかです。あなたが選んだ選択と相手が選んだ選択によってあなたのボーナスの出席点が決定します。

		相手の行動	
		協力	裏切る
あなたの行動	協力	(6 , 6)	(1 , 8)
	裏切る	(8 , 1)	(2 , 2)

かっこないの見方：（あなたの点数、相手の点数）

注

- 1) 山岸 (2000) p.52によれば、1000以上の実験がある。
- 2) 山岸 (2000)、p.56。
- 3) 山岸 (2000)、p.53。
- 4) ゲーム理論で分析される「囚人のジレンマ」の重要な状況設定に、「各プレイヤーは個々に交渉ができないようにされている」というものがある。「囚人のジレンマ」は非協力ゲームなので、上述のように想定している。理論的には、相談を行なっても、拘束力のある約束を行ないえないので、「協力行動」という戦略は選択されえない。バカラック (1981) 4章で詳述されているように、理論的には相談を行なうことの影響は存在しない。しかし、実際に影響が存在しないと断定することは困難である。そこで、「実験」では相談を禁止した。
- 5) 平均の差に関する検定を行なえば、相違しているか否かに関する結論は出る。しかし、同じ「実験」方法をとるように努めたが、「実験」は手分けして行なったため、完全に同じであるとは言い切れない。1人で「実験」を行なうか、「実験」を手分けして行なう場合には、同様な実験を多く行なうことが必要である。微細な「実験」方法の相違が、「実験」結果に大きな影響を与えるため、本稿では平均の差に関する検定を行なわない。

参考文献

- [1] Axelrod, R., *The Evolution of Cooperation*, New York : Basic Books, 1984 (松田裕之訳『つきあい方の科学 — バクテリアから国際関係まで —』HBJ出版局、1987年。).
- [2] Bacharach, M., *Economics and the Theory of Games*, Curtis Brown Ltd., London, 1976 (鈴木光男・是枝正啓訳『経済学のためのゲーム理論』東洋経済新報社、1981年。).
- [3] 林直保子『協調行動の社会関係的基盤』多賀出版、1999年。
- [4] 中山幹夫『初めてのゲーム理論』有斐閣、1997年。
- [5] Nash, John, "Non-Cooperative Games," *Annals of Mathematics* 54, pp.286-295, 1951.
- [6] 利島保・生和秀敏編『心理学のための実験マニュアル — 入門から基礎・発展へ』北大路書房、1993年。
- [7] 山岸俊男『社会的ジレンマ「環境破壊」から「いじめ」まで』PHP研究所、2000年。