

第 10 章 社会的交換と互惠性——なぜ人は 1 回限りの囚人のジレンマで協力するのか

山岸 俊男・清成 透子・谷田 林士

10.1 社会的交換ヒューリスティックスと裏切り者検知

なぜ人々は、匿名性が保証された 1 回限りの囚人のジレンマで協力行動を選択するのか。完全に匿名性が保証された 1 回限りの囚人のジレンマには、長期的にも短期的にも、協力行動の誘因は存在していない。それにもかかわらず、これまで無数といってよいほど繰り返し行われた囚人のジレンマ実験では、匿名性が保証された 1 回限りの状況においても、かなりの数の実験参加者が協力行動を選択すること——実験による差は大きいですが、大多数の実験では 3～6 割の参加者が協力行動を選択している——が明らかにされている。本章の目的は、1 回限りの囚人のジレンマ実験での協力行動選択に対する二つの解釈を対比させることで、その背後に存在する「進化的」基盤を明らかにすることにある。

その二つの解釈を、本章では「混乱仮説」と「ヒューリスティックス仮説」と名づけることにする。混乱仮説によれば、1 回限りの匿名状況での囚人のジレンマ実験で協力行動を選択する人間は、囚人のジレンマ実験での利得構造の意味を正確に理解していない人間だとされている。逆にいえば、利得構造の意味、特に非協力の選択が優越戦略であることを正しく理解しさえすれば、全員が非協力行動を選択するはずだとされている。また、混乱仮説と類似の説明として、誘因不十分仮説が存在する。囚人のジレンマ実験で用いられる誘因は現実の囚人のジレンマでの誘因に比べ小さいので、現実の囚人のジレンマでは非協力行動を選択する人間でも、囚人のジレンマ実験では他の動機に基づいて協力行動をとるのだ、という仮説である。この仮説を逆転すれば、1 回限りの囚人のジレンマ実験においても誘因さえ十分に大きくすれば、全員が非協力行動

を選択するだろうと予測されることになる。

実験参加者が実験状況の意味を「誤解」しているという点では、ヒューリスティックス仮説を提唱するわれわれも、混乱仮説に同意している。ヒューリスティックス仮説が混乱仮説と異なっているのは、その「誤解」の背後に、より根源的な人間の情報処理の特性を仮定している点である。すなわちわれわれは、人々は1回限りの囚人のジレンマでの利得構造の意味を「誤解」させるようにはたらく、「社会的交換ヒューリスティックス」を用いた情報処理を行っている——1回限りの囚人のジレンマに直面した実験参加者は、その状況を社会的交換状況だと判断し、社会的交換状況の存在を前提とした状況認知と意思決定を行うため、特定の方向での「混乱」が生じやすい——と考えている。次節では、まず、社会的交換ヒューリスティックスに関するわれわれの考えを述べることにする。

A. 社会的交換と裏切り者検知

Cosmides と Tooby (1989, 1992; Cosmides, 1989) は、人類は進化の過程で社会的交換状況での課題解決に特化した領域特長的認知モジュールを獲得してきたとしている。この彼らの議論の背後には、社会的交換における相互協力の促進が人類の進化史のなかで中心的な適応課題をなしてきたとする理解が存在している。そして彼らは、人類の心の進化を生みだしてきた中心的な舞台である社会的交換状況を、囚人のジレンマに代表される混合動機状況として定義している。このことは、社会科学や生物学のさまざまな分野でこれまで研究されてきた、囚人のジレンマないし社会的ジレンマ (N 人状況における囚人のジレンマ) における相互協力の達成が、人類の心の進化にとっての最も重要な適応課題の一つとなってきたことを意味している。

これまで、利己的な個体の間での相互協力を説明するために、生物学者は二つの説明原理を提案してきた。血縁淘汰の原理 (Hamilton, 1964) と互惠的利他主義の原理 (Trivers, 1971) である。この二つの原理のうち、社会学者は特に後者の互惠的利他主義の原理に着目してきた。その理由は、彼らは非血縁関係にある人々の中の協力関係の成立に関心をもってきたからである。本章におけるわれわれの関心も、多くの社会学者の関心と同様に、非血縁関係に

ある人々の中の相互協力の達成にある。社会学者やゲーム理論家は、生物学者が互恵的利他主義とよぶ行動原理を、TFT 戦略 (tit-for-tat strategy) として取り扱ってきた (e.g., Axelrod, 1984)。たとえば Axelrod による一連の戦略コンピュータ・シミュレーションの結果は、TFT 戦略が、シミュレーションに用いられたすべての戦略のなかで最も優れた成績をあげる適応的戦略であることを示している。さらに彼が行った追加シミュレーションでは、TFT 戦略が進化する——つまり、成績の悪い他の戦略にとって代わって集団内で最も勢力を伸ばしていく——ことが示されている。また、彼のコンピュータ・シミュレーションでは、TFT が進化的に安定した ESS 戦略であることも示されている。

Cosmides と Tooby は、これらの知見や TFT 戦略の有効性に関する理論化に基づき、社会的交換場面に直面した際に人々に TFT 戦略をとらせるような (あるいは、互恵的な利他行動に従事させるような) 領域特殊的認知モジュールを、人類は進化の歴史のなかで獲得してきたとしている。さらに、TFT 戦略を採用したり互恵的利他的に振舞ったりするためには、非協力者のなかから協力者を見抜く能力が必要とされるため、人間は進化の歴史を通して、社会的交換状況で非協力者やフリーライダーを見つけるように、われわれの注意を向けさせる領域特殊的認知モジュールを獲得してきたと議論している。彼らが主張しているように、社会的動物としての人間の適応価を高めるにあたって社会的交換における相互協力の促進が重要であり、かつ TFT 戦略を効果的に用いるためには非協力者を協力者から見抜く必要があるとすれば、われわれ人間が非協力者から協力者を見分けるために最適にデザインされた情報処理のための領域特殊的な認知モジュールを獲得したこと、とりわけ社会的交換領域に特殊の認知モジュールを獲得したという議論は、非常にもっともらしく思われる。

本章では Cosmides と Tooby によって議論された、この「社会的交換の論理」を前提にしながら話を進めていく。ただし、Cosmides と Tooby が「裏切り者検知」に焦点を当てているのに対して、本章では、社会的交換状況での相互協力を推進する別の条件に焦点を当てて考察を進めることにする。すなわち、社会的交換における相互協力の重要性に関する上述の Cosmides と Tooby による基本的な議論を受け入れた上で、本章でのわれわれの議論は、

裏切り者検知能力が相互協力の達成にとって意味があるのは、そもそも協力の達成を望ましいと考えている人の場合のみであるという明確な事実から出発する*。つまり、裏切り者検知能力は、社会的交換状況において相手を搾取したいと思わない人、すなわち協力的な相手から一方的に搾取することを望まない人にとってのみ有益な能力である。このことは、PDにおける相互協力の達成には、協力したいと願うこと（あるいは協力的な相手からの一方的な搾取を望まないこと）と、協力者から非協力者を見きわめる認知的な能力の両方が必要であることを意味している。

このわれわれの議論の出発点は、Pruitt & Kimmel (1977) の協力の目標/期待理論にある。Pruitt & Kimmel (1977) は、1970年代半ばまでに行われた1000を超える実験ゲーム研究の蓄積を概観し、PDにおける協力の達成には次の二つの要素が必要であることを結論づけた。すなわち、(1) 自分自身の利益の一方的な追求から、相互協力の達成への目標の変換と、(2) 相手も同様に利己的な利益の追求をあきらめ、相互協力を目指しているとする期待、の二つである。このPruittとKimmelの議論との対比でいえば、CosmidesとToobyの「社会的交換の論理」は、「期待」に焦点を当ててはいるが、「目標」の部分に対してはあまり注意を払っていない。

協力的な相手を搾取しようと望む人は、たとえ協力者から非協力者を見分ける能力をもっていたとしても、相互協力を達成できないことは自明である。逆に、非協力的な相手による搾取の可能性に注意を払わずに一方的に協力する人も、相互協力を達成できない。この二つの点を考えれば、社会的交換状況で相互協力を達成させるためには、目標の変換を促進するメカニズムと、協力者と非協力者を見きわめるメカニズムの両方が必要であることは自明である。したがって、社会的交換状況における裏切り者検知に特化された認知的なモジュールをわれわれ人間が進化させてきたのであれば、協力的な相手を一方的に搾取しようとするのをやめさせるための別の認知的モジュールも併せて進化させて

* この議論は、他人も同様に裏切り者検知能力をもつという前提に基づいている。この前提が成立している限り、裏切り者はたとえ自分の餌食を見つけたとしても、その相手からは交換の相手として受け入れられないだろう。したがって、裏切り者検知能力は、他の人から交換の相手として受け入れられることのできる人にとってのみ、すなわち、協力者にとってのみ有益なものである。

きたはずである。本章では後者に焦点を当て、この認知的メカニズムを「社会的交換ヒューリスティックス」とよぶことにする。

B. PD から安心ゲームへの主観的構造変換

社会的交換ヒューリスティックスは、いかにして私たちに交換相手の搾取をあきらめさせ、相互協力を目指すよう仕向けるのだろうか。この中心的な疑問に対するわれわれの答えは、人々が囚人のジレンマ (PD) を安心ゲーム (Assurance Game: AG) へと主観的に構造変換することにより生じるというものである。つまり、社会的交換に直面した人間は、囚人のジレンマに類した状況を安心ゲームとして知覚する認知的バイアスをもっている、とする答えである。囚人のジレンマでは非協力が優越選択である。つまり、相手の選択が何であれ、非協力が個人的にはよい結果を生み出す。これに対して、表 10.1 に示されているように、安心ゲームには優越選択が存在しない。安心ゲームでは、相手が非協力を選択している場合には、自分も同様に非協力を選択することが個人的に望ましい結果を生む。しかしながら、相手が協力を選択している場合には、自分も同様に協力を選択したほうが個人的によりよい結果を生む。主観的な構造変換が起こるということは、人々が囚人のジレンマ的な状況を安心ゲーム状況として知覚する、つまり、相手が自分と同じように協力する限りは、非協力よりも協力したほうが個人的にはよりよい結果を生み出すと直感的に知覚することを意味している。

これまでの実験研究の結果は、一貫して、1 回限りの PD でも参加者はしば

表 10.1 囚人のジレンマゲームと安心ゲームの例
プレイヤー A の利得はセルの右上、プレイヤー B の利得はセルの右下に表している。

		(a) 囚人のジレンマの例		(b) 安心ゲームの例	
		プレイヤー A の選択		プレイヤー A の選択	
		Cooperation	Defection	Cooperation	Defection
B プレイヤーの選択	C	2 / 2	3 / 0	2 / 2	1 / 0
	D	0 / 3	1 / 1	0 / 1	1 / 1

しばゲームの構造を主観的にはAGとして理解していることを示している。このことを示す直接的な証拠としては、まず、プレーヤー自身が、一方的な非協力よりは相互協力の結果を選好するという実験結果があげられる。たとえばKollock (1997) は、場面想定法を用いた実験で、PDの相手が敵対集団の成員でないかぎり、実験に参加したプレーヤーたちは、一方的な非協力よりも相互協力の結果を好むことを示している。同様の結果は、渡部・寺井・林・山岸 (1996) により行われた実験においても報告されている。彼らの実験の参加者は1回限りのPDをプレーし、その後、PDの結果生じ得る四つの結果の望ましさについてそれぞれ評定した(7件法尺度)。その結果、この実験の参加者たちは、相手が協力している場合には、自分が非協力するよりも協力するほうが個人的により満足の高い結果を生むと判断していた(6.22 vs. 4.62)。しかし、相手が非協力のときには、協力しないほうがより満足な結果を生むと判断していた(3.16 vs. 1.82)。このパターンは、アメリカ人参加者や韓国人参加者を用いた同様の実験でも確認されている(Cho & Choi, 2000; Hayashi, Ostrom, Walker, & Yamagishi, 1999)。さらに、表10.2に示された、山岸を中心とする研究グループにより実施された一連の実験でも、同様の結果が一貫して得られている。

表10.2には、これらの研究を含む一連の研究で得られた、PDでの四つの結果(セル)に対する「望ましさ」得点の平均値が示されている。また同表は、選好がAG構造に合致している参加者の比率と、PDのインセンティブ構造に合致している参加者の比率も同時に示している。質問紙に対する回答に基づくと、渡部ら(1996)の実験の参加者のうち41%がPDをAGとして経験しており(すなわち、DCよりもCCを好み、CDよりもDDを好む)、PDをPDとしてそのまま経験していた(すなわち、CCよりもDCを好み、CDよりもDDを好む)プレーヤーはわずか18%しかいないことがわかる*。アメリカ人参加者を用いた実験(Hayashi et al., 1999)では、PDをAGとして変換したプレーヤーは30%であり、PDをPDとしてそのままとらえた参加者は28%であった。さらに、寺井(1995)をはじめとする四つの実験では、参加者は

* PDプレーヤーとAGプレーヤーの比率は、他の選好を示した参加者が存在するため合計100%とはならない。

表 10.2 PD における四つの結果に対する満足度と選好順位の平均値
PDゲームとしてプレーした人とAGとしてプレーした人の比率

実験	n	望ましさの平均値 (7件法)				選好順位の平均値				比率			
		CC	DC	CD	DD	CC	DC	CD	DD	望ましさ	順位	望ましさ	順位
渡部ら(1996)	148	6.22	4.62	1.82	3.16					.18	.41		
Hayashi et al. (1999)	167	6.01	5.05	1.76	3.66					.28	.30		
寺井(1995)	81	6.33	4.71	1.94	3.79	1.38	2.46	3.47	2.70	.22	.46	.04	.40
清成・山岸・中嶋(1998)	79	6.44	4.82	2.13	3.59	1.43	2.43	3.43	2.68	.22	.42	.30	.47
山岸・小杉(1999)	40	6.18	4.78	2.50	4.08	1.23	2.58	3.49	2.67	.18	.43	.20	.58
Yamagishi, Kikuchi, & Kosugi (1999)	90					1.37	2.32	3.60	2.69			.27	.51

CC, DC, CD, DD: 最初の文字は参加者自身の選択を示し, 2番目の文字はパートナーの選択を示す。

* CCよりもDC(相互協力よりも搾取すること)を好み, CDよりもDD(搾取されるよりも共貧)を好む「PDプレーヤー」の比率

* DCよりもCC(搾取するよりも相互協力)を好み, CDよりもDD(搾取されるよりも共貧)を好む「AGプレーヤー」の比率

四つの結果の「望ましさ」をそれぞれ独立に評定するのとは別に, 四つの結果を望ましい順番に順位づけるやり方でも結果の望ましさを評定している。これらの選好ランキングの平均値は, 四つの結果に対する「望ましさ」得点の平均値のパターンと一貫しており(数値が低いほど順位が高い), これらの実験に参加したプレーヤーの多くがPDをAGとして経験していることを示している。PDをAGとして変換している参加者(以後, 「AGプレーヤー」とよぶ)の比率は40~58%の範囲にわたり, PDそのものとしてとらえた参加者(以後, 「PDプレーヤー」とよぶ)は, 4~30%の範囲にわたっている。このように, 表10.2で示したすべての研究で, 大多数の参加者の選好はPDの利得構造ではなく, AGの利得構造と対応していることが示されており, 多くの参加者が実際にPDをプレーする際に, ゲームの内容をAGへと変換してプレーしていた可能性が示唆されている。

ただし, 質問紙に対する回答のみでは, 相互協力の望ましさに対する参加者の選好の表出が単なる「リップサービス」に過ぎず, 実際の選好を表していない可能性が残されている。しかしながら, 日本人参加者を用いた渡部ら(1996), アメリカ人参加者を用いたHayashi et al. (1999), 韓国人参加者を用いたCho & Choi (1999)の実験の結果は, このリップサービスの可能性を

否定している。これらの実験では、参加者は決定前に相手の決定を知らされ、その後自身が協力するか非協力するかを決定するという、順次 PD ゲームを用いる条件が設定されていた。この順次 PD で第1プレーヤーが非協力を選択していることを知らされた上で、協力・非協力の選択を行った第2プレーヤーの行動が、非協力ではなく協力を選擇するという結果が得られれば、上述の質問紙研究で表明された選好どおりに実際に行動していることになり、質問紙研究で表明された選好が単なるリップサービスである可能性が否定されることになる。

この順次 PD において、まず第1プレーヤーが非協力を選擇した場合には、誰もが予測するとおり、ほとんどすべてのプレーヤーが非協力を選擇していた（日本人参加者23名中20名、アメリカ人参加者13名中全員、韓国人参加者10名中全員）。それでは、第1プレーヤーがすでに協力を選擇していることが知らされている条件の結果はどのようなものだろうか。もし参加者の「真」の選好が PD の利得行列に示された金銭的価値と一貫しているとすれば、相手が協力した場合にも相手が非協力を選擇した場合と同様、非協力を選擇するはずである。しかし、実際の結果はこの予測とはまったく逆であり、相手が先に協力した条件では、第2プレーヤーの過半数が協力を返すことが示されている（日本人参加者20名中15名（75%）、アメリカ人参加者18名中11名（61%）、韓国人参加者11名中8名（73%）。また、相手が先に協力している条件における協力率は、同時に決定する通常の1回限りの PD 条件での協力率（日本人参加者56%、アメリカ人参加者36%、韓国人参加者46%）よりも高いものであった。さらに、この通常の1回限りの PD 条件においては、AG の選好パターンを示した AG プレーヤーのほぼ全員が協力していた（日本人参加者10名中9名、アメリカ人参加者6名中全員）のに対して、PD プレーヤーの大多数は非協力を選擇していた（日本人参加者の3名中2名、アメリカ人参加者の3名中全員）。これらの結果は、(1) 多くの参加者は PD を AG へと主観的に変換しており、(2) 彼らの実際の行動は、事後質問への回答に表出された PD ないし AG の構造を示す選好パターンと一貫していることを示している。

C. 混乱かヒューリスティックスか

上述の実験結果は、人々が1回限りの囚人のジレンマ状況をAG状況として主観的に経験していることを明確に示している。このこと、つまりPD状況をAG状況として経験するということは、相互協力が望ましい状況だと直感的に理解することと同じである。われわれはPD状況をAG状況として理解させるように人々を導く心理的メカニズムを社会的交換ヒューリスティックスとよぶが、上述の議論は、社会的交換ヒューリスティックスが活性化すると人々は相互協力を目指そうと動機づけられることを意味するものである。本章では、次に、この社会的交換ヒューリスティックスが、1回限りの囚人のジレンマに直面したプレーヤーに作用していることを示す、Kiyonari, Tanida, & Yamagishi (2000) による実験を紹介する。

社会的交換の分析を合理的選択アプローチからとらえる人たちは、社会的交換ヒューリスティックスのはたらきを示唆する上述の実験的証拠に直面しても、それらの結果は、参加者の混乱によって生じたものだと主張することが多い（たとえば Andreoni, 1995; Shafir & Tversky, 1992）。合理的選択理論の立場からのこの「混乱仮説」（Andreoni, 1995）によれば、われわれが社会的交換ヒューリスティックスとよんでいるものは、実際にはプレーヤーが混乱しており、利得構造を十分に理解していないことを示しているに過ぎないとされる。つまり、実験参加者はゲームの利得構造、特に、非協力が優越選択であるという1回限りのゲームの性質を理解していない実験参加者が、1回限りのゲームをあたかも繰り返し相互作用があるかのように「混乱して」理解している可能性である。論理的に考えれば、われわれが社会的交換ヒューリスティックスとよんでいるPDのAGへの主観構造変換は誤りである。われわれはこの「誤り」を、社会的交換状況において相互協力をつくりだす人間の認知機能の重要な構成要素と考えているのに対して、合理的選択理論家は、この「誤り」に基づく相互協力を偶発的な混乱の結果と見なしている。

D. 仮 説

以下に、Kiyonariら（2000）が行った実験を紹介するが、この実験の目的は、1回限りのPDにおける協力行動に対するこれらの二つの説明、すなわち

混乱仮説とヒューリスティックス仮説のうち、いずれが妥当であるかを検討することにある。この実験では、1回限りのPDでの協力が「混乱」によって引きだされているという説明と、本章で提唱する社会的交換ヒューリスティックス仮説による説明のいずれが正しいのかを、実験参加者がいかに真剣にPDで決定を下すかを操作することによって検討する。さきに紹介したように、われわれは、社会的交換ヒューリスティックスはPDをAGへと主観的に変換させると考えている。したがって、社会的交換ヒューリスティックスが活性化されやすい状況では、そうでない状況よりも協力率が高くなると予測される。以下の実験では、社会的交換ヒューリスティックスを活性化しやすいと考えられる状況と、しにくいと考えられる状況をつくりだすために、実験で用いた社会的交換場面の現実性を操作している。

(1) 現実性

われわれは、プレーヤーが実験状況を2人の参加者が相互依存関係にある状況として設定する場合には、交換される資源が現実性を欠いていたり、交換が成立しないゼロサム状況としてフレームされている場合に比べ、社会的交換ヒューリスティックスが活性化されやすいだろうと考えている。より具体的には、参加者が実験課題を、自分や相手の決定によって自分や相手の利益に重大な結果をもたらすと知覚するときのほうが、そうでないときよりも協力率が高くなるであろうと予測している。すなわち、次のように仮定する。

仮説1： 交換の現実性（すなわち、参加者や相手の決定が自分自身や相手の利益に対して重要な影響をもたらす程度）は、プレーヤーの協力率を上昇させる。

他方で、混乱仮説（Andreoni, 1995）では、結果の重要性の効果を逆に予測することになる。結果が重大な場合には、プレーヤーは利得構造の性質とゲームが1回限りだという点に、より注意を払うだろう。そしてその結果、混乱は減少すると考えられる。そして混乱仮説によれば、混乱が減少したときにはプレーヤーはより合理的な決定を下す、すなわち非協力的行動を選択するようになるはずである。したがって、混乱仮説からは、以下の代替仮説が導かれる。

代替仮説 1 : 交換の現実性は参加者の協力率を減少させる。

(2) 同時 PD 対順次 PD

実験の第2の目的は、同時 PD と順次 PD の比較をとおして社会的交換ヒューリスティックスのはたらきを例証することにある。われわれはさきに、順次 PD での第2プレイヤーの協力率が高いのは、社会的交換ヒューリスティックスがはたらいているためだと議論した。ゲームが順次にプレーされる場合には、2人のプレイヤー間のコンティンジェンシーの感覚、つまり第2プレイヤーの行動は第1プレイヤーの行動に依存するという感覚が強まるだろう。そのため、第1プレイヤーと第2プレイヤーの間に交換関係が存在しているという直感的理解が促進され、そのために社会的交換ヒューリスティックスが活性化されやすくなるだろうと予測される。すなわち、社会的交換ヒューリスティックスは、同時 PD でよりも順次 PD でより引きだされやすいと予測される。

仮説 2 : 順次 PD において、第1プレイヤーがすでに協力したことを知っている第2プレイヤーの協力率は、同時 PD のプレイヤーの協力率よりも高いだろう。

この比較に関しても、混乱仮説からは逆の予測が導きだされる。すなわち、第2プレイヤーは同時 PD のプレイヤーよりも複雑さの少ない課題に直面するからである (Shafir & Tversky, 1992 参照)。第2プレイヤーは囚人のジレンマの四つの結果をすべて考える必要がなく、第1プレイヤーが選択した手に対応する二つの結果のみを比較すればよい。また、第1プレイヤーの思惑を考える必要もない。さらに、すでに相手は決定を下しているので、自分の決定が相手の決定に影響するといった「コントロール幻想」、つまり、自分自身の行動が相手の選択に影響するだろうという信念が入り込む余地がない。したがって、混乱仮説からは、以下の代替仮説が導かれる。

代替仮説 2 : 順次 PD において、第1プレイヤーがすでに協力したことを知っている第2プレイヤーの協力率は、同時 PD のプレイヤーよりも低いだろう。

この代替仮説2は、Shafir & Tversky (1992)の研究では支持されている。彼らは、1回限りのPDでも相手の行動をコントロールできると考えるコントロール幻想が、同時ゲームでよりも順次ゲームで減少するだろうと予測した。とりわけ、第1プレーヤーがすでに自分の決定を終えており、そのため第1プレーヤーの決定に影響できないことを知っている第2プレーヤーのコントロール幻想はきわめて低くなるだろうと考えられる。したがって、第1プレーヤーがすでに協力していることを知らされている第2プレーヤーの協力率は、2人のプレーヤーが同時に決定する同時PDでの協力率よりも低くなるだろうと彼らは予測した。そして彼らの実験結果は、この予測を支持するものであった。すなわち、彼らの実験の参加者は、PDが同時に行われたときよりも、相手がすでに協力の決定をしていると知らされた場合に、より低い協力率を示した。

Shafir & Tversky (1992)により示されたこの結果は、上述の仮説2や、Cho & Choi (1999), Hayashi et al. (1999), 渡部ら (1996)によってさきに示された実験結果と矛盾している。また、第1プレーヤーがすでに協力したことを知っている順次ゲームの第2プレーヤーが、同時ゲームの参加者よりも高い協力率を示すという、渡部ら (1996)と同様の結果を報告している、Morris, Sim, & Giretto (1998)の結果とも矛盾している。さらに、若干文脈が異なるが、McCabe, Smith & LePore (2000)もまた、ゲームが通常形(すなわち利得行列の形)で提示される場合よりも、展開形で提示される場合——ゲームが展開型で提示される場合にはプレーヤーは相手と順次に対戦し、通常形態の場合には同時に対戦する——に、プレーヤーはより協力の選択をとりやすいという実験結果を示している。

Shafir & Tversky (1992)の結果とそれ以外の結果との間のこのような対照的な違いは、いかにして説明可能だろうか。われわれはこの違いは、交換の現実性の大小により生みだされたと考えている。すなわち、Shafir & Tversky (1992)の実験では交換の現実感が乏しかったために、社会的交換ヒューリスティックスが活性化されにくかったのだとわれわれは考えている。彼らの実験参加者は、実験ゲームを40試行繰り返し経験しており、その40試行全体で得た得点を集積したものが報酬とされた。合計得点は実験の終わりに実際のお金に交換されたが、ゲームの最中には得点の換金率は知らされていなかった

た。すなわち、それぞれのゲームでの意思決定と、その結果である報酬を受け取るまでの間に長い隔たりがあり、参加者にとっては自分の決定が報酬に直接結びついていていたことを実感しにくい状況であったといえる。そのような単調で退屈な作業や、毎回の決定は結果にわずかしか影響しない（参加者は自分の決定の結果いくらかを稼いだかあるいは失ったかをゲームの途中では知らなかったが、実際には1回の決定につき2～3セント程度の影響しかもたなかった）ような作業に直面した場合には、参加者は毎回の決定にあまり真剣には取り組まないだろう。さらに、彼らの実験では、試行ごとに新しい相手がランダムにマッチされるというゲームの性質によって、社会的交換状況であるという知覚が妨げられたと考えられる。まとめると、実験ゲームの試行数が多く、試行ごとに新しいパートナーがランダムに選択されるという Shafir & Tversky の実験デザインには、社会的交換ヒューリスティックスの活性化を促進する要素が欠如していたといえる。逆に Cho & Choi (1999), Hayashi et al. (1999), 渡部ら (1996) の実験の参加者は、PD ゲームを一度だけプレーし、報酬は1回のPDでの自分と相手の意思決定の結果に直接依存していた。当然自分の課題に真剣に取り組む、少なくとも Shafir & Tversky の実験に参加した参加者よりも自分たちの決定の重要性を理解していたと考えられる。

Shafir & Tversky (1992) の結果に対するわれわれのこの解釈は、以下に紹介する実験で検討する第3の仮説を導きだしている。われわれは、第1プレーヤーがすでに協力していることを知っている第2プレーヤーの協力率よりも同時PDでの協力率のほうが高いという Shafir & Tversky (1992) の結果が、社会的交換ヒューリスティックスが活性化されにくい状況でのみ得られるだろうと予測する。つまり、結果が些細であり、そのためプレーヤーが交換の現実性を感じない状況でのみ得られる結果だと予測している。逆に、仮説2で予測したパターンは、ゲームの結果が些細ではなく、プレーヤーが交換を現実的に感じる時に生じると予測される。

仮説3： 仮説2で予測された第2プレーヤーの協力率の増加は、ゲームの結果が些細でないときにのみ生じるだろう。

この点に関しても、混乱仮説は異なるパターンを予測するだろう。混乱は、

第2プレーヤーよりも同時PDプレーヤーの間でより深刻だと考えられる（代替仮説2）が、同時プレーヤーの混乱は、特にゲームの結果が些細なときにより大きくなると予測されるからである。ゲームの結果が重大なものであれば、プレーヤーはゲームの性質を注意深く検討し、その結果混乱は減少するようになるだろう。したがって、代替仮説2で予測した同時PDプレーヤーと第2プレーヤーとの協力率の差は、ゲームの結果が重要なときよりも些細な場合により顕著になると考えられる。

代替仮説3： 代替仮説2で予測された、第2プレーヤーと同時プレーヤーとの間の協力率の差は、ゲームの結果が些細な場合にのみ生じるだろう。

（3） 順次PDでの第1プレーヤー

以下に示す実験では、「交換の現実性」の水準、つまり、プレーヤーと相手の決定の結果が自分自身と相手の利益に関係する程度を操作した。具体的には、二つのタイプの実験を比較した。実験参加者の報酬が自分と相手の決定に実際に依存している完全実験と、実験参加者の決定が自分や相手の報酬に対して無関係な場面想定法実験の二つのタイプの実験である。

またこの実験では、同時PD条件と順次PDにおける第2プレーヤー条件の二つの条件に加えて、第3の条件として、第1プレーヤー条件が加えられている。この条件の参加者は、第2プレーヤーが決定を行うまえに、協力か非協力かの決定を行うように教示されていた。また第1プレーヤーには、第1プレーヤーの決定を知らされた上で第2プレーヤーが決定することが教示されていた。この条件は、コントロール幻想がもはや幻想ではない条件である。Cho & Choi (1999), Hayashi et al. (1999), 渡部ら (1996) の結果が示しているとおり、第2プレーヤーの選択は、第1プレーヤーの選択によって大きく影響されるからである。混乱仮説からは、第1プレーヤーの協力率が、同時PDプレーヤーや順次PDの第2プレーヤーよりも高いか低いかについては予測できない。これに対して社会的交換ヒューリスティックス仮説からは、順次PDの第1プレーヤーの協力率は、第2プレーヤー条件の参加者とほとんど変わらないだろうと予測される。社会的交換ヒューリスティックスが活性化されると、第1プレーヤーは第2プレーヤーが互恵的に行動すると期待すると考えられるた

めである。社会的交換ヒューリスティックスが活性化された状況では、第1プレイヤーの協力に対する第2プレイヤーの互惠性と、第2プレイヤーからの互惠性に対する第1プレイヤーの期待の両方が高まるだろう。第2プレイヤーの互惠性に対する期待が存在する場合には、第1プレイヤーにとっては非協力するよりも協力したほうが主観的にはよりよい選択となる。したがって以下の二つの付加的な仮説が導きだされる。

仮説 2f (第1プレイヤーに対する仮説)： 順次 PD ゲームにおける第1プレイヤーの協力率は、同時 PD ゲームのプレイヤーの協力率よりも高いだろう。

仮説 3f (第1プレイヤーに対する仮説)： 仮説 2f で予測された同時 PD プレイヤーと第1プレイヤーの協力率の差は、ゲームの結果が些細ではない場合にのみ生じるだろう。

10.2 実験の説明

以下に紹介する実験では、交換の現実性を操作するために、すなわち、プレイヤーの決定が自分自身や相手に対して重大な結果を生み出すという理解を操作するために、二つの異なる形態の実験を実施した。まず第1に、実験参加者は PD を1回だけプレーし、利得表に明記されたとおりの金額がそのままプレイヤーに支払われる「完全実験」が実施された*。この完全実験では、実験参加者は、同時 PD 条件、第2プレイヤー条件、第1プレイヤー条件の三つのプレイヤー条件にランダムに配置された。つまり、プレイヤー条件は被験者間要因として操作されている。各参加者は3条件のうちの一つの条件だけで、1回のみ PD をプレーした。参加者は実験室で教示を与えられ、自分と相手の報酬が自分と相手の決定によって決まることを完全に理解するために十分な時間が

* 実験参加者に実際に支払われた報酬は、PD の実際の結果とは完全には対応していなかった。というのは、表 10.3 に示した利得表では一方的に搾取されたプレイヤーの利得はゼロであるが、実験参加者の報酬がゼロになる事態を避けるために、一方的に搾取された参加者には実際には 1200 円を支払ったからである。この場合、参加者には、相手が協力したというフィードバックが与えられた。

与えられた。これに対して、第2のタイプの実験は場面想定法を用いて行われた。実験参加者はシナリオに記述された実験に参加している場面を想定するように教示され、もしシナリオに記述された実験に自分が参加したとすれば、自分なら協力を選択すると思うか、非協力を選択すると思うかを尋ねられた。この場面想定法実験では、三つのプレーヤー条件は被験者内要因として操作されている。以下に、それぞれの実験についてさらに詳しく説明する。

A. 完全実験

149名（男性108名、女性41名）の参加者が、1500名以上登録している北海道大学社会心理学研究室の実験参加希望者プールのなかからスケジュールされた。実験希望者プールは、大学の教養部のさまざまな学部から参加可能な学生（主として1年生）を募って作成された。希望者プールへの登録は完全に自由意思でなされており、登録の勧誘に際しては金銭的報酬が強調されていた。

実験に際しては、1度に数名の参加者が電話でスケジュールされた。スケジュールされた個々の参加者は、文学部ロビーに設置された受付に到着すると、1人1人IDカードを渡され、匿名性を保持するためにID番号によって識別されることが教示された。参加者はその後1人ずつ実験室に案内され、実験室のなかの個室に通された。したがって、各参加者は受付に同時に到着した場合を除いて、他の参加者に会う機会をもたなかった。その後、各参加者は各個室でインストラクション用紙を渡され、全員がインストラクションを読み終わるまで待つように求められた。インストラクションでは、実験で用いるPDゲームの利得が表10.3に示された形で提示され、その利得表を用いながら実験の説明がなされた。各参加者はゲームの利得構造に関する説明を読んだ後で、利

表10.3 実験で用いたPDの利得表

あなたの選択	相手の選択			
	K		P	
L	あなたに 1200円	相手に 1200円	あなたに 0円	相手に 1800円
S	あなたに 1800円	相手に 0円	あなたに 600円	相手に 600円

得構造を理解したかどうかを確認するための簡単な確認問題に解答した。実験者が各解答をチェックし、必要に応じて説明が口頭で加えられた。各参加者には、インストラクションのなかで、いくつかの異なるやり方でゲームがプレーされる可能性のあることが予告されていた。以下に紹介される3条件の操作は、参加者の確認問題に対する解答が確認された後で与えられた「追加インストラクション用紙」のなかでなされた。

同時PD条件 ($n=48$) では、最初に教示されたのと同じ利得表が「追加インストラクション用紙」で再び与えられ、L (協力) かS (非協力) かのいずれかを選択するように求められた。同時に、相手も参加者と同じようにK (協力) かP (非協力) のいずれかを選択するように求められていると教示された。第1プレーヤー条件 ($n=51$) では、同じ利得表が提示された後、参加者は最初に決定するように求められた。さらにこの条件では、第2プレーヤーには、決定前に第1プレーヤー (参加者) の決定が知らされることも教示された。最後に、第2プレーヤー条件 ($n=50$) では、利得表が提示され、他の参加者 (第1プレーヤー) が意思決定をしている間、その決定を待つように求められた。その後、他の参加者がK (協力) を選択したという内容の情報が実験者によってもたらされ、その後で、参加者は自分自身の決定を行うように求められた。

参加者は協力・非協力の選択を、「追加インストラクション用紙」と同時に配布された「意思決定用紙」上のLかSの文字に○をつけることで行った。その後、参加者の決定がわからないようにするために、参加者は封筒の中に決定用紙を入れ実験者に手渡した。実験者に手渡された封筒は、別の部屋にいる集計者に渡され処理されることが参加者に教示されていた。集計者は封筒に入れている被験者の決定に基づいてそれぞれのプレーヤーの結果を決定し、それぞれのプレーヤーが受けとるべき報酬額を参加者のIDが表面に記入された封筒に入れ、実験者がその封筒を参加者に渡した。この手続きをとることで、参加者と対面する実験者は参加者の決定内容はもちろん、それぞれの参加者がいくら稼いだかもわからないように配慮されていた。つまり、集計者にはそれぞれの参加者がどのような選択を行ったかはわかるが、参加者をID番号でしか識別できないし、参加者と対面する実験者には参加者の行動がわからないよ

うにされていた。この手続きを用いることで参加者の決定の匿名性が他の参加者に対してだけでなく、参加者と対面する実験者に対しても完全に保持されていることは、インストラクションのなかであらかじめ参加者に対して説明されていた。

B. 場面想定法実験

場面想定法実験には112名が参加した。これらの参加者は、二つの異なるやり方で実験に参加した。まず、半数の56名の参加者（男39名、女17名）は、上述の完全実験と同様に、同じ北海道大学の実験参加希望者プールからスケジュールされ、完全実験に用いられたのと同じ実験室で、1人1人個室に入って実験に参加した。これらの参加者には、実験参加の謝礼として一律1000円が支払われた。参加者は各個室で完全実験用のインストラクションを受けとり、もしこの実験に自分が参加したとすれば、L（協力）とS（非協力）のどちらを選択すると思うかを尋ねられた。その後、完全実験と同様の手順で「意思決定用紙」のLかSに○印をつけ、その用紙を実験者が回収した。

残り半数の参加者は、別の大学（千葉大学）の心理学入門の授業に出席している学生であった*。参加者は授業を担当している教官（実験者）から、授業の15分を実験に費やしてほしいと要請され実験に参加した。この実験では報酬は支払われなかった。この方法で実験に参加した参加者も、完全実験の参加者と同じインストラクションを与えられ、「意思決定用紙」上のLかSに○印を付けるように求められた。ただし、これらの参加者に対しては、インストラクションの配布は授業の教室で行われ、15分後に実験者が意思決定用紙を回収するという手続きが用いられた。このように、場面想定法実験には北海道大学の学生と千葉大学の学生が異なった方法で参加していたが、予備的分析の結果、これら二つのグループの被験者の間に主要な変数に関して有意な差が見られなかったため、以下に示す分析はこの違いを無視するかたちで行われている。

上述の二つのいずれの方法で参加した場合にも、参加者は実際に実験で用い

* この千葉大学での授業を利用した実験では、参加者の性別は尋ねられなかった。

られたインストラクションを読み、その実験に参加したとすれば自分はどのように行動するかを想起するように教示された。その後、二つの選択肢L（協力）とS（非協力）のうち、その実験に参加したとすればどちらを選択するかを考えて、意思決定用紙上に○印を付けるように求められた。参加者には、三つのプレーヤー条件（同時PD条件、第1プレーヤー条件、第2プレーヤー条件）で用いられた3セットのインストラクションが与えられており、三つの条件のそれぞれで、自分がどのように行動するかを想像して決定するように教示されていた。なお、第2プレーヤー条件では、相手はすでにK（協力）を選択していることが教示されていた。これらの3条件が提示される順序は、参加者ごとにランダム化されていた。

場面想定法実験では、完全実験で用いられた条件に加え、金銭条件と得点条件の二つの被験者間条件が設定されていた。金銭条件に割り当てられた48名の参加者は、オリジナルの完全実験と同様に、金額（つまり円）で表現された利得表が提示された。残りの54名の参加者には、同じ利得表が単なる得点の形で提示された。この得点条件の参加者に対しては、オリジナルの実験の参加者は自分自身の得点を最大化するように求められていたとの説明が与えられた。金銭条件と得点条件の操作は、北海道大学の参加者に対しても千葉大学の参加者に対しても与えられた。

金銭条件と得点条件の二つの条件を比較すると、得点条件の決定課題は、金銭条件の課題と比べ現実性が乏しい、極度に些細な課題だといえる。たとえ参加者がインストラクションをまじめに読んで実際にその実験に参加した場面を想像していたとしても、得点条件では意思決定の結果が単なる抽象的な得点であるため、想像上の実験参加者にとってさえ重大な結果をもたらすとは考えられない。したがって、交換の現実性が最も高いのは完全実験であり、次いで場面想定法実験の金銭条件で、最も低いのは場面想定法実験の得点条件だと考えられる。

10.3 実験結果の説明

A. 仮説1対代替仮説1

表10.4は、完全実験と場面想定法実験における三つのプレーヤー条件での協力率が、交換の現実性と密接に関係していることを示している。三つのプレーヤー条件すべてにおいて、協力率は完全実験で最も高く、次いで場面想定法金銭条件であり、最後に場面想定法得点条件であった。実験タイプの主効果（完全実験、場面想定法金銭条件、場面想定法得点条件）は χ^2 検定の結果有意であった（ $\chi^2(2)=15.23, p<.001$ ）。この結果は明らかに仮説1を支持するものであり、代替仮説1を棄却するものである。この実験では、プレーヤー条件の実験操作は、完全実験では被験者間要因として、また場面想定法実験では被験者内要因として行われていた。そのため、有意差の検定に際しては、実験タイプの差とプレーヤー条件を同時に分析する際に生まれる分析上の問題を避けるために、場面想定法実験における1回目の回答のみを用いた被験者間要因のみの2要因分散分析が行われた*。1回目の決定における協力率は、同表の「場面想定法実験（1回目の反応）」と記述してある行に報告されている。そこに示されている1回目の協力率は、「場面想定法実験（全反応）」の列で報告されている。被験者内要因であった三つの条件に対する反応すべてを含んだ結果

表10.4 実験での協力選択の比率

	利得の表示	プレーヤー条件		
		同時PD	第2プレーヤー	第1プレーヤー
完全実験	金銭(n=149)	37.5%(n=48)	62.0%(n=50)	58.8%(n=51)
場面想定法実験 (全反応)	全体(n=112)	26.8%	35.5%(n=110)*	31.3%
	金銭(n=58)	29.3%	53.6%(n=56)*	41.4%
	得点(n=54)	24.1%	16.7%	20.4%
場面想定法実験 (1回目の反応)	全体(n=112)	26.5%(n=34)	32.4%(n=34)*	35.7%(n=42)
	金銭(n=56)	23.5%(n=17)	52.6%(n=19)*	55.0%(n=20)
	得点(n=54)	29.4%(n=17)	6.7%(n=15)	18.2%(n=22)

* 第2プレーヤー条件で欠損値が二つある（第2のグループの被験者）

* 場面想定法実験では、被験者内要因であるプレーヤー条件の順番はランダムに参加者に割り当てられていたために、最初に提示された条件での反応のみを被験者間要因として分析することが可能である。

とほとんど一貫している。

B. 仮説 2 と 3 対代替仮説 2 と 3

第 2 プレーヤー（第 1 プレーヤーがすでに協力したことを知らされた場合）の協力率は、完全実験においては同時プレーヤーよりも高く（62.0 % vs. 37.5 %），その差は有意であった（ $\chi^2(1)=5.88, p<.05$ ）。この差は、場面想定法金銭条件でも同様の方向で認められた（全反応に関しては 53.6 % vs. 29.3 %，1 回目の反応のみでは 52.6 % vs. 23.5 %（ $\chi^2(1)=3.20, p<.05$ ）。これに対して、場面想定法得点条件では、有意ではないが（ $\chi^2(1)=2.71, ns.$ ），同時プレーヤーは第 2 プレーヤーよりも高い協力率を示していた（全反応：同時 PD プレーヤー 24.1 % vs. 第 2 プレーヤー 16.7 %，1 回目の反応のみ：29.4 % vs. 6.7 %）。また、実験条件とプレーヤー条件の交互作用は傾向差であった（ $\chi^2(2)=4.98, p<.10$ ）。興味深いことに、同時 PD プレーヤーが第 2 プレーヤーよりもより高い協力率を示すという Shafir & Tversky (1992) の予測が実現しているのは、PD ゲームの結果が極端に些細である場面想定法得点条件のみであった。つまり、被験者が合理的に振舞ったのは、些細な意思決定課題に直面した場合のみであった。これらの結果は明らかに仮説 2 と仮説 3 を支持するものであり、代替仮説 2 と 3 に反している。

C. 仮説 2f と 3f

仮説 2f と 3f の予測どおり、第 1 プレーヤーと同時 PD プレーヤーの対比は、第 2 プレーヤーと同時 PD プレーヤーの対比と一貫していた。すなわち、第 1 プレーヤーは完全実験においては同時 PD プレーヤーよりも高い協力率を示しており（第 1 プレーヤー 58.8 % vs. 同時 PD プレーヤー 37.5 %： $\chi^2(1)=4.50, p<.05$ ），場面想定法金銭条件でも同様の結果が得られた（全反応：第 1 プレーヤー 41.4 % vs. 同時 PD プレーヤー 29.3 %，1 回目の反応のみ：第 1 プレーヤー 55.0 % vs. 同時 PD プレーヤー 23.5 %， $\chi^2(1)=3.78, p<.05$ ）。これに対して場面想定法得点条件ではこのパターンは逆転し、第 1 プレーヤーは同時 PD プレーヤーよりも低い協力率を示している（全反応：第 1 プレーヤー 20.4 % vs. 同時 PD プレーヤー 24.1 %，1 回目の反応のみ：第 1 プレーヤー

— 18.2% vs. 同時 PD プレーヤー 29.4%) が、その差は有意ではなかった ($\chi^2(1)=0.68, ns.$)。実験条件×プレーヤー条件の交互作用は有意水準には達しなかった ($\chi^2(2)=4.04, ns.$)。

10.4 考 察

上に紹介された実験結果のメッセージは明確である。実験参加者の多くは、重大な結果をもたらす PD 状況に直面したときには非合理的に、つまり互恵的に行動しており、逆に、極端に些細な課題（すなわち利得がお金ではなく得点で示されている場面想定法実験）に直面したときにのみ「合理的」に振舞っていた。混乱という考え方では、第1プレーヤーが協力していることを知らされた第2プレーヤーの協力選択は説明できない。さらに、混乱アプローチでは、なぜ第2プレーヤーが些細な意思決定課題に直面したときよりも、重大な意思決定課題に直面したときにより協力的な反応を示すのかを説明できない。したがって、「コントロール幻想」、すなわち自分の行動がパートナーの行動に影響するという直感的な状況理解は、単なる偶発的な混乱ではなく、人間の認知メカニズムに内在するものであると結論づけることができるだろう。われわれはそのメカニズムが、社会的交換ヒューリスティックスであると考えている*。

1回限りの PD において自分自身の行動が相手の行動に影響すると期待するのは、論理的には明らかに誤りである。問題は、なぜ多くの実験参加者がそのような間違いを犯すのか、そして、実質的な報酬がまったくかかっていない些細な PD に直面したときよりも実際に金銭的な結果をもたらされる重要な PD に直面したときのほうが、そのような間違いをより頻繁に犯すのはなぜか、である。社会的交換ヒューリスティックスの考え方は、社会的交換状況においては、そのような誤りを犯すほうが、論理的に正しい意思決定を行うよりもより適応的な行動である（あるいは、前者が後者よりも適応的な社会環境が存在する、ないし、した）、という前提に基づいている。1回の PD 実験でそのような誤りを犯す人は、報酬を余分に得る機会を失っているというのは正しい。し

* 経済学者の Coricelli, McCabe, & Smith (2000) は、われわれが社会的交換ヒューリスティックスとよぶものを「好意会計 (goodwill accounting)」とよんでいる。

かしながら、現実生活における社会交換場面で相互協力を達成するための機会を失うという結果は、1回限りのPD実験で余分に数百円を失うよりももっと深刻なことのほうが多いだろう。社会的交換ヒューリスティックスは、社会的交換状況でわれわれに相互協力を追求させる「バイアス」であり、交換相手を搾取しようとする事からわれわれの注意をそらせるものである。この「バイアス」がわれわれの適応価を高めるか下げるかは、社会的交換状況で一方的な搾取と比較して相互協力の達成が相対的に重要かどうか依存している。社会的交換の領域では、論理的に正しい推論は、われわれに相互協力を達成させようとする誤りを犯すことほどはよい結果を生みださないかもしれない。人間の現実の行動を重要だと考える経済学者やゲーム理論家は、単純な混乱と、適応的な機能をもったヒューリスティックスの違いについてもっと注意を向けるべきであろう。

われわれの実験結果は、経済学やゲーム理論に対してだけでなく、進化心理学に対しても重要なインプリケーションをもっている。もしわれわれ人間が社会的交換状況で相互協力の達成を促進する認知モジュールをもっているとしたら、そこには、裏切り者検知モジュールのみではなく、社会的交換ヒューリスティックスも同時に備わっているはずである。同様の点は Cosmides & Tooby (1989, 1992) の初期の議論でも示唆されてはいるが、その後の議論および実証研究が裏切り者検知に集中するなかで、相互協力を目指させるように促進するメカニズムの重要性が軽視される傾向にあったように思われる。

本章の最初で議論したように、裏切り者検知モジュールと社会的交換ヒューリスティックスの一方だけをもっていたとしても、もう一方をもたない限りは何の役にも立たない。この単純な洞察からは、相互協力の達成に強く動機づけられた人々（あるいは社会的交換ヒューリスティックスが強力にはたらく人々）は、相互協力を目指していない人々に比べ、非協力者と協力者を見きわめるためにより注意をはらっている、あるいは協力者から非協力者を実際により適切に見分けている、という予測が生みだされる。この予測は、より協力的で他者を信頼する人々は、非協力的で他者を信頼しない人々に比べ、社会的交換状況で他者の行動により注意深く、他者にだまされにくいららうという、一般的な常識に反する予測を含んでいる。しかしながら、山岸とその共同研究者

(菊地・渡辺・山岸, 1997; Yamagishi, Kikuchi & Kosugi, 1999; 山岸・小杉, 1999) は, この常識に反する予測の妥当性を実証する証拠を積み重ねつつある。彼らの研究では, 1回限りのPDにおいて, 高信頼者は低信頼者よりも協力者と非協力者をより正確に見きわめることを示している。これらの実験的な知見は, 社会的交換状況において, 人々に相互協力の達成を目指すよう仕向ける認知的メカニズムと, 協力者と非協力者を区別させる認知的メカニズムの間の相互関係が, 今後の研究にとって実り多いテーマとなることを示唆するものである。

参考文献

- Andreoni, J. (1995). Cooperation in public-goods experiments: Kindness or confusion? *The American Economic Review*, 85, 891-904.
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- Cho, K. & Choi, B. (1999). A cross-society study of trust and reciprocity: Korea, Japan and the U. S. A paper presented at the WOW II, Workshop for the Political Theory and Policy Analysis, Indiana University, June 16-19.
- Coricelli, G., McCabe, K., & Smith, V. (2000). Theory of mind mechanism in personal exchange. *Affective Minds*. In G. Hatano, N. Okada, & H. Tanabe (Eds.), Elsevier.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187-276.
- Cosmides, L. & Tooby, J. (1989). Evolutionary psychology and the generation of culture, Part II: A computational theory of social exchange. *Ethology and Sociobiology*, 10, 51-97.
- Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.). *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York: Oxford University Press.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. Parts I, II. *Journal of Theoretical Biology*, 7, 1-52.
- Hayashi, N., Ostrom, E., Walker, J., & Yamagishi, T. (1999). Reciprocity, trust, and the sense of control: A cross-societal study. *Rationality and Society*, 11, 27-46.
- 菊地雅子・渡邊席子・山岸俊男 (1997). 他者の信頼性判断の正確さと一般的信頼——実験研究。『実験社会心理学研究』, 37, 23-36.
- Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior*, 21, 411-427.
- 清成透子・山岸俊男・中嶋 健 (1998). 安心の提供と信頼の生産。『日本グループ・ダイナ

ミックス学会 第46回大会発表論文集], 292-293.

- Kollock, P. (1997). Transforming social dilemmas: Group identity and cooperation. In P. Danielson (Ed.). *Modeling rational and moral agents* (pp. 186-210). Oxford: Oxford University Press.
- 小杉素子・山岸俊男 (1998). 一般的信頼と信頼性判断. 『心理学研究』, **69**, 349-357.
- McCabe, K., Smith, V., & LePore, M. (2000). Intentionality detection and “mindreading”: Why does game form matter? *Proceedings of the National Academy of Science of the United States of America*, **97**, 4404-4409.
- Morris, M. W., Sim, W. M., & Giretto, V. (1998). Distinguishing sources of cooperation in the one-round prisoner's dilemma: Evidence for cooperative decisions based on the illusion of control. *Journal of Experimental Social Psychology*, **34**, 464-512.
- Pruitt, D. G. & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, **28**, 363-392.
- Shafir, E. & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, **24**, 449-474.
- 寺井 滋 (1995). 囚人のジレンマでの協力と相互依存性の知覚. 北海道大学文学研究科修士論文.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, **46**, 35-57.
- 渡部 幹・寺井 滋・林 直保子・山岸俊男 (1996). 互酬性の期待にもとづく1回限りの囚人のジレンマにおける協力行動. 『実験社会心理学研究』, **36**, 183-196.
- Yamagishi, T., Kikuchi, M., & Kosugi, M. (1999). Trust, gullibility and social intelligence. *Asian Journal of Social Psychology*, **2**, 145-161.
- 山岸俊男・小杉素子 (1999). 社会的交換における裏切り者検知. 『認知科学』, **6**, 179-190.