間接互惠性の成立

——非寛容な選別主義に基づく利他行動の適応的基盤——

真島理恵 高橋伸幸 北海道大学

The emergence of indirect reciprocity: Evolutionary foundation of altruistic behavior based on “strict discriminator”

Rie Mashima and Nobuyuki Takahashi (Hokkaido University)

Although there have been a number of studies that theoretically and empirically examined altruism based on direct reciprocity, few have been conducted on how altruism based on indirect reciprocity emerges. Recent advances in biological research, however, have suggested possible answers to the question. For instance, Nowak and Sigmund (1998a, b) proposed that what they called image scoring strategy made indirect reciprocity possible. After critically examining their work, Leimar and Hammerstein (2001) pointed out several limitations to the theory, and instead proposed standing strategy as an explanation. Although careful attempts to replicate the findings by them and Panchanathan and Boyd (2003) supported the arguments against image scoring, we reveal that standing strategy was not a satisfactory answer either. Based on a series of evolutionary simulations, we propose a new strategy, which we call strict discriminator, as an alternative. Strict discriminators are discriminating altruists, similar to the altruists with image scoring or standing strategy, but they are different in that its criterion for discrimination is stricter: unconditional altruists are excluded from their reciprocity.

Key words: altruism, indirect reciprocity, evolution, cooperation, discrimination.

The Japanese Journal of Psychology
2005, Vol. 76, No. 5, pp. 436-444

人は、なぜ利他的に振舞うのだろうか。利他行動とは外的報酬を期待せず他者のために自発的に行う、行為自体が目的の行動と定義される (Bar-Tal, Sharabany, & Reviv, 1982)。なぜ人間はそのような非合理的な行動をとるのだろうか。これまでの社会心理学における多くの研究では、この問いに対して、人間には無償の愛他動機が備わっているからである、という回答が提示され、利他的動機の促進・抑制要因の解明が中心的な焦点とされてきた (明田・岡本・奥田・外山・山口, 1994; 大坊・安藤・池田, 1989)。しかしこの説明は、そもそも人間にはなぜそのような非合理的な動機が備わっているのか、という更に一歩踏み込んだ問いに対しては答えをもたない。なぜなら、利他動機は人間に利他行動をとらせるためのメカニズムとして備わっているのだ、という説明は、上述の利他行動の説明 (人間が利他行動をとるのは利他的動機があるからだ) と循環する同義反復に過ぎないか

らである。これに対し本研究では、利他行動を、そのような行動傾向を備えることが行為者自身に利益をもたらすために備わっている適応の産物であると考え、この観点から利他行動の成立する仕組みを説明することを目指す。つまり本研究の目的は、一見非合理的に見える利他行動がなぜ、いかにして成立し得るかを、適応論的観点から明らかにすることにある。

互惠的利他主義

合理的な個人間における利他行動 (協利行動) を可能とする原理として 1990 年代まで焦点があてられていたのが、互惠的利他主義 (Trivers, 1971) である。繰り返しのつきあいのある二者間においては、自分が利他的に振舞うことによって相手から返報を引き出すことができる。この原理を互惠的利他主義と呼ぶ。これは、囚人のジレンマ研究における応報戦略 (Tit For Tat 戦略: 以下 TFT とする; Axelrod, 1984) ——最初は協利し、その後は相手が協利すれば協利し、相手が非協利すれば非協利する、条件つき協利戦略——に集約される。二者間で相手が TFT を採用している限り、相手への協利は自分に対する協利を、裏

切りは自分に対する裏切りを引き出すため、短期的な自己利益を追求して非協力をとるよりも、協力する方が長期的には合理的な選択となる。つまり、このような状況で利他的（協力的）に振舞うことは、自身に利益をもたらす適応的な行動といえることができる。

しかしこの議論は、二者間に繰り返しのつきあいが存在することを前提とする点で、その説明範囲に限界がある。あらかじめ直接の返報が期待できるとわかっている二者間での利他行動は、社会における利他行動のごく限られた一部に過ぎない。直接の返報は期待できないような、二者間を超えた助け合いは社会に決して少なくないだろう。だが、二者間を超えた助け合いの状況においては、利他的な他者にただ乗りすることが可能であるため、利他行動は適応的な行動とはなり得ない。では、そうした状況における利他行動を支える原理は何なのだろうか。

間接互惠性の原理

この問いに対する回答として Nowak & Sigmund (1998a, b) が紹介し、それ以降生物学を中心に注目を浴び始めた概念が、間接互惠性の原理である。間接互惠性の概念は“情けは人の為ならず”という諺で端的に表現することができる。即ち、“他者に対する利他行動は回り回って別の他者から報われる”という仕組みを間接互惠性と呼ぶ。したがって、間接互惠性が存在すれば、たとえ相手から直接に返報が期待できなくても、他者に対する利他行動は結果的に自分自身に利益をもたらす合理的な行動となる。Nowak & Sigmund (1998a, b) は、これを繰り返しのつきあいが存在しない状況における利他行動の説明原理として挙げ、更に間接互惠性成立の鍵が条件つき利他行動（選別的利他戦略）の存在にあると指摘した。仮に人々が、“利他的な他者に対してだけ利他的に振舞う”という選別的利他戦略に則って行動しているとしよう。そうであれば、誰かに対して利他的に振舞った人は将来別の人から利他的に振舞ってもらえる、という間接互惠性が成立し、利他行動は当人に利益をもたらす適応的な行動となる。この議論を受け、Nowak & Sigmund (1998a, b) 以降、どのような種類の選別戦略が間接互惠性の成立に最も有効かを焦点とする理論研究が行われ続けてきたが (Leimar & Hammerstein, 2001; Panchanathan & Boyd, 2003)、未だ明確な結論は得られていない。本研究ではこの間接互惠性に焦点をあて、進化的シミュレーションにより、間接互惠性の成立を可能とする要件を原理的に明らかにすることを通し、二者間を超えた利他行動の適応的基盤を解明することを目的とする。以下ではまず、間接互惠性の成立を可能とする解決策として先行研究で提示された選別戦略を再検討し、その限界点を指摘する。そして、間接互惠性の成立を可能とする新たな解決策を提

唱する。

ギビング・ゲームパラダイム

先行研究の内容に具体的に触れる前に、一方的な利他行動を検討するためのパラダイムとして先行研究でも用いられ (Leimar & Hammerstein, 2001; Nowak & Sigmund, 1998a, b)、本研究でも用いるギビング・ゲームのパラダイムについて簡単に説明しておこう。ギビング・ゲームは、相手からの直接の返報が期待できない（即ち、直接互惠性が存在しない）状況で一方的に相手に対して資源を提供するゲームである。ゲームの内容とシミュレーションの流れを以下に述べる。ゲームでは、複数のプレイヤーからなる集団が想定され、毎回、その中から一組のペアがランダムに選出され、一方に“渡し手”の役割が、もう一方に“受け手”の役割が割り当てられる。渡し手は、コスト (c) を支払って受け手に資源を提供するか提供しないかを決定する。資源を提供された受け手は b の利益を獲得する ($b > c$)。

渡し手が資源を提供するかどうかを決定する際には、受け手の“スコア”を参照する。スコアは Good もしくは Bad の値をとる評判情報である。全プレイヤーは、スコアが Good の相手に対しては資源を提供し、Bad の相手に対しては提供しないという行動規則に従う。ただし、どのような相手を Good とみなし、どのような相手を Bad とみなすかは戦略によって異なる。つまり、各人の主観的なスコアの付け方を戦略と呼ぶ。極端な例でいえば、あらゆる相手を Good とみなす戦略が無条件に提供する無条件利他主義戦略 (All Cooperator: All-C)、逆にあらゆる相手を Bad とみなす戦略が無条件に非提供をとる利己主義戦略 (All Defector: All-D) である。毎回、全員が渡し手の行動を観察し、その渡し手に対して各人が持つ主観的スコアを更新する。

また、プレイヤーは小さな確率で、行動のエラーと知覚のエラーという2種類のエラーを起こすと想定されている。行動のエラーとは、渡し手が資源を提供するか否かを決定する際に、誤って本来とるはずの行動とは逆の行動をとってしまう（例えば、提供するつもりだったのに非提供をとってしまう）エラーである。これに対し知覚のエラーとは、プレイヤーが、他の渡し手の行動を見誤るエラーである（他者の提供を非提供と勘違いする、もしくは非提供を提供と勘違いする¹⁾）。

渡し手と受け手が選ばれてから全員が渡し手の行動を観察してスコアを更新するまでの一連の流れを試行、試行の繰り返しを世代と呼ぶ。1世代終了後、各

¹ ただし Leimar & Hammerstein (2001) における知覚のエラーは、“Good を Bad と見誤る”という方向でしか生じないと設定されていた。

プレイヤーの獲得利益に応じて“淘汰”が生じる。淘汰では、その世代で相対的に高い利得を得た戦略の割合が増加し、利得の低かった戦略は減少するという規則に従い、次世代の構成員の戦略分布が決定される。世代を繰り返した結果、非利他的な戦略が集団から駆逐され、利他的な戦略同士が資源を提供しあうようになった状態を、ここでは間接互惠性が成立している状態と定義する。本研究の具体的な焦点は、ギビング・ゲームにおいて間接互惠性の成立を可能とする戦略を明らかにすることにある。

なお本研究では説明にあたり、便宜的に“淘汰”、“遺伝子”等の進化生物学の用語を用いているが、これらは一種の比喻であり、本研究のねらいは真に生物学的な意味での遺伝子レベルの進化過程を再現することではないことを断っておきたい。進化的シミュレーションは、集団内でプレイヤーが各々の行動規則（戦略）に従って相互作用を繰り返し、その結果より高い利得を得た戦略が増加して利得の低い戦略が減少していくよう設定することで、特定のゲーム状況においていかなる戦略を採用する個人が高利得を獲得し得るかを原理的に導き出すためのツールである。本研究におけるシミュレーションの目的も、ある戦略が他の戦略との相互作用においてどの程度有利となるかを演算することのみにある²。

解決策 1：イメージ・スコアリング戦略

間接互惠性の成立を可能とする戦略として最初に提唱されたのが、イメージ・スコアリング戦略（image scoring：以下 IS とする）である（Nowak & Sigmund, 1998a, b）。これは端的に言えば、提供者を Good、非提供者を Bad とみなす、単純な TFT 的戦略である。具体的には、第 1 試行、つまり過去の行動履歴が存在しない初期時点では、IS は全員のスコアを Good とみなす。そしてその後は、渡し手の役割になった際に資源を提供したプレイヤーは Good、提供しなかったプレイヤーは Bad としてスコアを割り振っていく。集団内に、無条件利他主義者 All-C と、利己主義者 All-D、そして IS の、3 戦略が存在する状況を想定してみよう。渡し手にとっては本来、提供よりも非提供が有利な選択である。しかし集団内にある程度 IS が存在する場合、非提供をとると、IS から資源を提供してもらえなくなるという、将来的な損失を被ってしまう。そしてこの損失は、資源提供により一時的に支払うコストよりも大きい（ $b > c$ ）ため、All-D は All-C や IS に比べ低い利得しか得ることができないことになる。したがって、最終的には集団から All-

D が駆逐され、利他的戦略だけが残る共栄状態が達成される。Nowak & Sigmund は、数理解析（Nowak & Sigmund, 1998b）とシミュレーション（Nowak & Sigmund, 1998a）の結果から、IS が間接互惠性を成り立たせる戦略であると結論づけている。

イメージ・スコアリングの問題点 しかし後続の研究は、Nowak & Sigmund (1998a, b) の結論がごく限られた条件下でしか成り立たないことを明らかにしている。まず Leimar & Hammerstein (2001) は、IS が勝利するという結果は、資源の提供コストが非常に小さい、もしくはシミュレーションモデルにおける遺伝的浮動の影響が非常に大きい、という特殊な条件が満たされていない限り成り立たないことを指摘した。更に Panchanathan & Boyd (2003) は数理解析の結果、IS と All-C により構成される均衡は最終的には All-D により侵入され崩壊することを数理的に証明している。彼らの指摘する IS の弱点とは、IS の、互いが互いを罰しあう点である。IS は非提供者に対して非提供をとるが、非提供をとった IS 自身は、他の IS から Bad とみなされ、将来資源を提供してもらえる機会を失ってしまう。これに対し、常に資源を提供する All-C は、将来 IS から提供を受ける機会を失うことは決してない。ゆえに IS の期待利得は All-C のそれを下回り、集団内では IS より All-C が高い比率を占める状態となる。そして All-C が大半を占める集団は All-D の侵入に対して無力であるため、最終的には All-D が集団に侵入、制覇し、共貧状態に陥ることになるのである。

解決策 2：スタンディング戦略

Leimar & Hammerstein (2001) 及び Panchanathan & Boyd (2003) は、IS のこのような問題点を指摘した上で、間接互惠性の成立を可能とする新たな解決策としてスタンディング戦略（standing：以下 ST とする；Sugden, 1986）を提唱している。

ST のスコアの付け方は、基本的な部分は IS と類似している。ST も IS と同様、最初は全員を Good とみなし、その後、提供者には Good のスコアを割り振っていく。ただし ST は、非提供者にスコアを割り振る際に、その非提供者が Good な受け手に対して非提供をとったのか、それとも Bad な受け手に対して非提供をとったのかに応じてスコアの割り振り方を変える点で、IS とは異なっている。ST は、Good な受け手に対して非提供をとった相手は Bad とみなすが、Bad な受け手に対して非提供をとった相手は、Good とみなす。つまり、Bad に対する罰として“正当化できる”非提供をとった相手と、Good に対する“正当化できない”非提供をとった相手を区別し、後者を罰する一方、前者を許容する点が、ST の特徴である³。この特徴ゆえに ST では、All-D を罰することで他の選

² この意味で本研究におけるシミュレーションは、限定的な合理性をもつ個人を仮定した場合に、利他性を備える戦略が有利となり、集団内に一般化することが原理的に可能か否かを確認するための思考実験である。

Table 1
Score assignment by standing strategy

Previous behavior		Score of recipient's previous recipient	
		Good	Bad
		(a) Good (c) Bad	(b) Good (d) Good
	gave		
	didn't give		

別戦略から罰されてしまう、という IS の弱点が克服されている。ST のスコアの割り振り方を、Table 1 に整理して示す⁴。前回その相手が資源を提供したかどうかが行に、相手の前回の受け手が Good だったか Bad だったかの区別が列に表されている。そして当てはまる相手、即ち“(a)前回 Good な受け手に提供した相手”、“(b)前回 Bad な受け手に提供した相手”、“(c)前回 Good な受け手に非提供した相手”、“(d)前回 Bad な受け手に非提供した相手”のそれぞれを ST がどのようにみなすかが示されている。Leimar & Hammerstein (2001) と Panchanathan & Boyd (2003) はそれぞれシミュレーションと数理解析の結果から、ST が、非提供者を駆逐し間接互惠性を維持する戦略であると提唱した。

スタンディング戦略の問題点 それでは ST が、間接互惠性の成立を可能とする解決策なのだろうか。しかしそう結論づける前に、ST の弱点（知覚のエラー）について言及しておく必要がある。先に紹介した、ギビング・ゲームで想定される 2 種類のエラーに関し、特に IS は行動のエラーに弱いこと、そして ST は知覚のエラーに弱いことが先行研究において指摘されている (Boerlijst, Nowak, & Sigmund, 1997; Nowak, Sigmund, & El-Sedy, 1995)。この指摘の前半部に関し、Panchanathan & Boyd (2003) は、IS による共栄状態は行動のエラーがある場合には崩壊するのに対し、ST はそれに耐えて共栄を維持するという数理解析の結果を得ている。しかし、ST は知覚のエラ

ーに対しても頑健な戦略なのだろうか。例として、ST を採用するプレイヤー P1, P2, P3 を想定してほしい。いずれのプレイヤーも、他の ST から Good とみなされているとする。ここでは、P3 が知覚のエラーを起こした場合にいかなる帰結が生じるかを例示したい。ある試行で P2 が渡し手、P1 が受け手に割り当てられ、P2 が P1 に資源を提供したとする。そしてその後の試行で P3 が渡し手、P2 が受け手に割り当てられたとする。前回の P2 の提供行動を観察していた P3 は、P2 を Good とみなし、資源を提供する。ゆえに提供した P3 は他の ST から Good とみなされる。しかしここで P3 が、P2 の行動を非提供と見誤るというエラーを犯した場合を考えてみよう。その場合、P3 から見ると P2 は“Good に対して非提供をとった”Bad な相手であるため、P3 は P2 に提供しない。しかし、実際は P2 は資源を提供したため Good である。よって、P3 が罰のつもりでとった非提供行動は他の ST から見ると、“Good に対する非提供行動”という罰されるべき行動になる。したがって ST は、知覚のエラーを起こした場合、他の ST から Bad とみなされ、将来資源を提供してもらう機会を失う可能性がある。これに対して、相手の行動にかかわらず常に提供する All-C は、知覚のエラーが生じて Good なスコアを失うことはない。したがって、知覚のエラーが存在する場合、ST の利得が All-C のそれを下回り、最終的には All-D が侵入する帰結がもたらされる可能性がある。このことは、Panchanathan & Boyd (2003) の結論に疑問を提示するものである。なぜなら、彼らのモデルは知覚のエラーを投入しないものであったため、彼らの主張する ST の優位性は知覚のエラーが存在する場合には成立しない可能性があるからである。

一方、Leimar & Hammerstein (2001) は行動のエラーと知覚のエラーの双方を投入したシミュレーションを実施し、ST が間接互惠性を成立せしめる結果を示している。しかし、彼らの結果はすべて、遺伝的浮動の影響を抑制するために採用した、island model (Wright, 1943) という特殊なモデルを用いたものであった。したがって、彼らの結論が、標準的な淘汰モデルを用いた場合にも当てはまるものであるかどうかは明らかではない。そこで本研究では、標準的な淘汰モデルを用いて、知覚のエラーが存在する場合の ST

³ ただしこの ST の定義は先行研究のものを簡略化したものである。Leimar & Hammerstein (2001) における ST は、自分のスコアが Bad である場合には、渡し手になった際に自らのスコアを Good に変えるために必ず資源を提供する、というより複雑な特徴を備えていた。しかし、後述のシミュレーションにおいて、この特徴を備える ST を用いた場合と、これを備えない ST を用いた場合で、質的に等しい結果が得られることが確認できたため、議論の単純化のために本研究では簡略化した定義を用いることとした。

⁴ ただし、Table 1 の (d) におけるスコアの割り振り方は、先行研究 (Leimar & Hammerstein, 2001; Panchanathan & Boyd, 2003) における ST とはやや異なる。先行研究では、誰かが Bad に対して非提供をとった場合、ST はその人のスコアを変化させず元のままにするということになっていた。しかし、ST の定義を、Bad に対する非提供者を常に Good とみなすとした場合にも質的に等しい結果が得られたため、本研究ではこの簡略化した定義を用いることとした。

の頑健性について再検討を行う。

第1シミュレーション

本研究では、これまで間接互惠性を成立させるとして提唱されてきた戦略の頑健性を再検討することを目的とし、まず、標準的な淘汰モデルで2種類のエラーを投入した進化的シミュレーションを行い、先行研究の結果が再現されるか否かを検討する。第1シミュレーションでは、Nowak & Sigmund (1998a, b) により提唱されたISの頑健性を再検討する。具体的にはAll-CとAll-DとISの3戦略を投入した進化的シミュレーションを実施する。

第1シミュレーション方法

シミュレーションでは、300人からなる集団を仮定する⁵。毎試行1組の渡し手と受け手がランダムに選ばれ、ギビング・ゲームを行う。渡し手以外のすべてのプレイヤーは渡し手の行動を観察し、それぞれが渡し手に関するスコア情報（GoodもしくはBad）を更新する。1500試行を繰り返した時点で、1世代が終了する。

淘汰 世代終了後、各プレイヤーが残す“子孫”の数が決定される。この際、その世代においてより高い利得を獲得したプレイヤーほど多くの子孫を残す確率が高く、逆に獲得利得の低いプレイヤーほど子孫を残せる確率が低くなる（ただし集団サイズは世代にかかわらず一定となるよう調整されている）。子孫は親から戦略のみを継承し、親の利得やスコアは継承しない。ただしごく小さな確率で、親とは異なる戦略を備える子孫が生産される（突然変異）。淘汰後、親自身

は姿を消し、子孫により次の世代が始められる。10000世代を繰り返し、これを1レプリケーションとする。なお本研究では遺伝的浮動が淘汰の結果に与える影響を抑制するため、Baker (1987) が開発しGrefenstette (2000) により遺伝的浮動の影響を低減させる方法として提唱された、確率論的サンプリング法を組み込んだ淘汰モデルを用いた。集団サイズと試行数の比率（1:5）、突然変異率（ $\mu=0.0001$ ）、行動のエラーの確率（ $\varepsilon=0.025$ ）、知覚のエラーの確率（ $\delta=0.025$ ）のパラメータ設定は、先行研究（Leimar & Hammerstein, 2001）に準じる。また提供コストを $c=1$ 、受け手の受け取る利益を $b=2, 4, 6, 8, 10$ の5段階で設定し、資源が何倍になって受け手に渡るか（ b/c ）を5条件で設定した。

第1シミュレーション結果

各条件においてそれぞれ20レプリケーション中、間接互惠性が成立した回数をFigure 1に示す（10000世代終了時の平均提供率が0.7を超えるレプリケーションを、間接互惠性が成立しているとみなした）。Figure 1の左半分にシミュレーション開始時の戦略の分布をランダムとした場合の結果を、右半分に集団の全成員が当該戦略（第1シミュレーションの場合はIS）を採用している状態から開始したシミュレーションの結果を示す。第1シミュレーションの結果を白い棒グラフで示す。Figure 1の左半分の結果は、全条件で一度も間接互惠性が成立しなかったことを示している。どの条件でも共通して、シミュレーション開始後数世代でまずAll-Dが集団から駆逐されるものの、その後ISよりもAll-Cが集団のほぼすべてを占める状態となり、最終的にはAll-Dが集団を制覇し共貧状態に陥る、という推移が観察された。またFigure 1右半分に示される結果は、初期状態として全員がISを採用している状態から始めた場合でさえ、

Figure 1. The number of replications where indirect reciprocity emerged.

⁵ シミュレーションにおける乱数の発生は、すべて一様乱数を用いて行った。また本シミュレーションプログラムの開発及び実行は、Boaland C++ Builderを用いて行われた。第2・第3シミュレーションも同様である。

Table 2
Score assignment by four genes

Previous behavior		Score of recipient's previous recipient	
		Good	Bad
	gave	(a) Good or Bad	(b) Good or Bad
	didn't give	(c) Good or Bad	(d) Good or Bad

ほとんどの場合間接互惠性は維持されなかったことを示している ($b/c=2$ の条件では 20 レプリケーション中 7 回で共栄が維持されたものの、残り 13 回、及び他の条件の全レプリケーションでは、すべて All-D が集団を制覇する結果となった)。

第 1 シミュレーション考察

第 1 シミュレーションの結果は Leimar & Hammerstein (2001) と Panchanathan & Boyd (2003) の議論を支持し、IS は All-C を増加させるという弱点のために、間接互惠性成立の解決策たり得ないことが確認された。

第 2 シミュレーション

第 2 シミュレーションでは、IS の問題点を克服しているとされる、ST の頑健性を再検討する。第 1 シミュレーションと同様の方法で、All-C, All-D, ST を投入したシミュレーションを実施した。

第 2 シミュレーション結果

第 2 シミュレーションの結果を、Figure 1 に灰色の棒グラフで示す。開始時の戦略の初期分布をランダムとした場合には (Figure 1 左側)、 $b/c=4, 6, 8$ の条件ではそれぞれ 20 回中、17 回、14 回、4 回のレプリケーションで間接互惠性が成立したが、残りのレプリケーションでは All-D が集団を制覇する共貧状態に陥った。この結果は、知覚のエラーが存在する場合、特殊な淘汰モデルを用いずに標準的な淘汰モデルを用いると、ST が間接互惠性成立の解決策だとする先行研究の主張は必ずしも成り立たないことを示している⁶。共貧に陥ったレプリケーションには、シミュレーション開始後すぐに All-D が増加するパターンのほか、一旦は All-D が集団から駆逐され、共栄状

態が達成されるが、その後 All-C の比率が増加し、最終的には All-D が侵入して共貧状態に陥るというパターンがかなり多く見られた。このことは、知覚のエラーがある場合、ST も All-C の侵入を許すという IS と同様の弱点のために、必ずしも間接互惠性を維持できないことを示す結果である。また、全成員が開始時に ST を採用する状態から始めた場合でさえ、間接互惠性が一貫して成立する結果は得られなかった (Figure 1 の右側に、 $b/c=6, 8, 10$ ではそれぞれ、4 回、15 回、20 回で共貧に陥ったことが示されている)。

第 2 シミュレーション考察

ST が間接互惠性を成立させるという先行研究の主張は、island model という特殊なモデルを用いた場合にしか成り立たず、知覚のエラーがある場合、ST は All-D の侵入を完全には防げないことが示された。

新たな解決策の検討 第 1, 第 2 シミュレーションの結果、先行研究で提唱された IS も ST も、間接互惠性を成立させる解決策としては不十分であることが明らかとなった。それではこれらの戦略のほかにも、間接互惠性の成立を可能とする戦略は存在するのだろうか。次に、論理的に想定可能な全戦略を検討し、間接互惠性の成立を可能とする戦略を探索的に検討する。

第 3 シミュレーション

ギビング・ゲームにおいて、プレイヤーが、他者が資源を提供したかどうかという行動履歴と、その時の受け手のスコアという 2 種類の情報を使用できることを前提とすれば、IS と ST のほかにも様々な選別戦略のパターンが想定可能なはずである。第 3 シミュレーションの目的は、この前提から論理的に想定可能な全戦略の有効性をそれぞれ検討し、間接互惠性の成立を可能とする、新たな戦略が存在するかどうかを探ることにある。まず、想定可能な戦略の種類について説明する。

各プレイヤーは、Table 2 の各セルに対応する 4 種類の相手のそれぞれを Good とみなすか、Bad とみなすかを決定する四つの遺伝子をもつ。この四つの遺伝子の組み合わせが“どのような相手のスコアをどうみなすか”を決定する戦略であり、計 $2^4=16$ 種類の戦略が想定されることになる。以下それぞれの戦略を、

⁶ 第 2 シミュレーションと同様のシミュレーションを island model を用いて実施したところ、Leimar & Hammerstein (2001) が用いたパラメータである $b/c=4$ の条件では、全 20 レプリケーションで集団を ST が制覇し、間接互惠性が成立するという先行研究の知見と一貫する結果が得られた。 $b/c=6, 8, 10$ でも全レプリケーションで共栄が達成されたが、これらの条件では必ずしも ST が集団において優勢であったわけではなく、 b/c が高くなるにつれ、集団内に ST よりも All-C が高い比率を占める頻度が増加する傾向にあった。

遺伝子の組み合わせの形で表現する⁷ (e.g., GGGG = All-C, BBBB = All-D, GGBB = IS, GGBG = ST)。

第3シミュレーション方法

第1, 第2シミュレーションと同様に, All-C, All-D, IS, ST を除く 12 戦略の有効性を検討するシミュレーションをそれぞれ行った。具体的には, All-C, All-D に加え, もう 1 戦略を投入する, 12 種類のシミュレーションを実施した。方法及びパラメータ設定は第1, 第2シミュレーションと同様である。

第3シミュレーション結果

まず, 開始時の戦略の初期分布をランダムとしたシミュレーションを 12 戦略それぞれについて行った結果, 唯一, 戦略“GBBB”を投入したシミュレーションにおいてのみ, b/c が 4 以上の全条件の全レプリケーションで, All-D が集団から駆逐され, 共栄状態が達成された (Figure 1 の左側の黒い棒グラフ)⁸。GBBB とは, “前回 Good に提供した相手に対してのみ提供し, 他の相手には一切提供しない”という非常に厳しい基準をもつ選別戦略である。このことから本研究ではこれを strict discriminator (以下 SDISC) と呼ぶこととする。更に SDISC による均衡の頑健性について調べるために, 集団内の全員が SDISC を採用しているという初期分布から開始したシミュレーションを実施した。その結果, ($b/c=2$ の条件を含む) 全条件の全レプリケーションを通じて, SDISC は All-D を駆逐し, 一貫して間接互惠性を維持せしめた (Figure 1 の右側の黒い棒グラフ)。しかし $b/c=10$ の条件では 10 000 世代終了時には SDISC だけではなく All-C が集団内である程度の比率を占めるようになっていた ($b/c=10$ における, 10 000 世代終了時の All-C の平均比率 0.42, SDISC の平均比率 0.58)。

第3シミュレーション考察

第3シミュレーションの結果, 間接互惠性を成立させる新たな戦略として, SDISC が示された。それでは, なぜこの戦略のみが, 間接互惠性の成立を可能とし得たのだろうか。その鍵は, SDISC が, 非提供者を排除するのみならず, “非提供者への提供を罰する”という特徴を備えている点にあると考えられる。

IS と ST の弱点は, 共栄を達成した後に All-C の増加を許し, 最終的に All-D の侵入を招いてしまうと

いう点であった。SDISC がこれらの戦略と異なるのは, 非提供者に資源を提供する All-C の増加を許容せず, むしろ罰する行動をとるという点である。IS や ST は, それがどのようなスコアの持ち主に対する提供かにかかわらず提供者を常に Good とみなすが, SDISC は “Bad への提供者” を Bad とみなし, 提供しない。つまり, IS や ST は All-C に常に提供するのに対し, SDISC は Bad に提供する All-C には提供せず, 排除するのである。更に, SDISC はこの “Bad への提供者を排除する” という特徴に加え, IS や ST が備えている “Good への非提供者を排除する” という要素, 即ち Good にさえ提供しない All-D を排除する行動傾向も備えている。これらのことから, 第3シミュレーションで SDISC のみが間接互惠性を成立させたのは, この戦略が, 非提供者を罰し, かつ All-C の存在を許容せず All-D が侵入する土壌を作らないという特徴を備えていたためであると考えられる。

最後に, SDISC の, $b/c=10$ の条件における結果の解釈を補足する。第3シミュレーションの結果は, 基本的には SDISC が All-C と All-D を排除するというパターンを示している。しかし $b/c=10$ の条件においては, ある程度の比率の All-C が SDISC と共存する結果となった。これは, 集団から All-D が完全に淘汰された状況においては, (エラーを考えなければ) SDISC と All-C の振る舞いがほぼ等しくなるために見られた結果であると考えられる。ただしここで注目すべき点は, そのような状況下でさえ, SDISC が All-C の比率を一定以下に抑え, All-D が侵入可能となるまでの増加を許していないことである。つまり, エラーや突然変異などにより集団内にわずかでも非提供行動が存在するならば, SDISC は All-C の過度の増加を防ぎ, 間接互惠性を維持することが可能な頑健な戦略であることが示唆されている。

総合考察

本研究では, 二者間を超えた利他行動の成立基盤を明らかにすることを目指し, 間接互惠性の成立を可能とする選別的利他戦略の検討を行った。第1, 第2シミュレーションの結果, これまでに提唱された IS と ST のいずれもが間接互惠性の解決策としては不十分であることが明らかにされ, 第3シミュレーションの結果, “Good に提供した相手にのみ資源を提供する”という選別基準をもつ SDISC のみが間接互惠性の成立を可能とすることが示された。この結果は, IS や ST は利己主義者 (Bad) に対して寛容な All-C を許容するために脆い戦略であること, そしてそれ故に間接互惠性の成立を可能とするには, 寛容な All-C をも排除する, 厳しい戦略が必要であることを示すものである。

この結果が示唆するのは, “直接互惠性が成り立た

⁷ なお, 過去の行動履歴が存在しない第1試行においては, いずれの戦略も全員を “(a) 前回, Good に対して提供した” とみなしてスコアを割り振る。このように設定したのは, 先行研究において IS や ST は “情報がない場合には相手を Good とみなす” と設定されていたことを踏襲するためである。

⁸ なお 12 戦略を検討した中で, $b/c=2$ の条件で一度でも間接互惠性が成立する結果は一つも見られなかった。

ない状況においては、利己主義者を許容する寛容な利他主義者の存在が、利他主義を破壊する”という命題である。直接の見返りが期待できない状況における利他行動は一見、無条件に寛容な利他性の発露としてとらえられるかもしれない。しかし本研究の結果は、無条件の利他行動を許容することは、それを食べ物にする利己主義者の増殖を招き、最終的には社会における利他性の崩壊を導くことを明らかにした。本研究の結果は、二者間を超えた利他行動を支える基盤は寛容な利他主義ではなく、利己主義者を排除するのみならず、利己主義者を助ける者も排除するという、いわば村八分的な厳しい選別的利他主義にあることを示唆した点において、利他行動研究に新たな視点を提供するものである。

展望 それでは、社会における利他行動は、本当に人々が実際にこのような形での選別的利他主義を備えることにより成立しているのだろうか。この点には未だ疑問が残る。なぜなら、SDISCを採用するには、偶然 Bad な相手と出会ってしまった人には決して資源を提供しないという極端な厳しさを身につけなければならないからだ。不運にも Bad な人と出会ったために、やむを得ず資源を提供しなかった相手に対しても非提供をとるという選別基準は、直感的にはあまりに不合理的な厳しさであるように思われる。これまでも、人々の行動を検討する実験研究が行われてきたが、一貫した結果は得られていない (Bolton, Katok, & Ockenfels, 2003; Milinski, Semmann, Bakker, & Krambeck, 2001; 高橋・山岸・林, 1999)。現実には人々が SDISC のような戦略をとっているのか、あるいは全く異なる行動傾向を備えているのかを、今後実験やフィールドワークにより実証的に検討する必要がある。

最後に、本研究を含む間接互惠性研究の成果と、人類学や社会学、及び社会心理学を中心に進められてきた一般交換研究を連結することの有用性について考察する。本研究では主に、Nowak & Sigmund (1998a, b) 以降生物学者を中心に展開されてきた間接互惠性研究に沿った形で議論を展開してきたが、ここで彼らの議論が、社会心理学を中心に行われてきた一般交換研究と非常によく似ていることを指摘しておきたい。両者の扱う間接互惠性の概念と一般交換の概念はほぼ同義であり、間接互惠性の成立要件を探る研究成果のかなりの部分は一般交換のそれと重なっているにもかかわらず、これまで両者を統合する研究はほとんど行

われてこなかった。間接互惠性（一般交換）の成立要件を探る研究の成果は共通して、（選別基準は戦略により異なるものの）条件付きの利他行動をとる、選別的利他戦略が必要であると示している⁹。これは、非選別的な利他主義は成り立ち得ないとする Hardin (1982) の議論とも一貫する結論である。

ただしこれら二つの研究の流れは、それぞれ仮定する状況の種類が異なる。直接互惠性が存在しない状況を扱う上で、Nowak & Sigmund (1998a, b) に代表される生物学者は資源の提供者と被提供者がランダムにマッチングされる状況を想定してきたのに対し、Takahashi (2000) に代表される一般交換研究では提供者自身が資源の提供相手を選択する状況が想定されてきた。囚人のジレンマ研究 (Yamagishi & Hayashi, 1996) での用語を用いると、前者は強制的プレイ (forced play)、後者は選択的プレイ (selective play) の状況に相当する。これは理論的に重要な違いであることが知られており、間接互惠性（一般交換）の文脈においても、この状況の相違がそこの適応的行動の種類に質的な違いを生み出す可能性は十分に考えられる。今後、強制的プレイ状況を想定する間接互惠性研究の結論が選択的プレイ状況においても同様に導かれるか否かを検討する研究が望まれる。

引用文献

- 明田芳久・岡本浩一・奥田秀宇・外山みどり・山口勸 (1994). パーシク現代心理学 7 社会心理学 有斐閣
(Aketa, Y., Okamoto, K., Okuda, H., & Toyama, M., & Yamaguchi, S. (1994). *Basics of contemporary psychology*. Vol. 7. *Social psychology*. Tokyo: Yuhikaku.)
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- Baker, J. (1987). Reducing bias and inefficiency in the selection algorithm. In J. Grefenstette (Ed.), *Proceedings of the 2nd International Conference on Genetic Algorithms*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 14-21.
- Bar-Tal, D., Sharabany, R., & Raviv, A. (1982). Cognitive basis for the development of altruistic behavior. In V. J. Derlega & J. Grzelak (Eds.), *Cooperation and helping behavior: Theory and research*. New York: Academic Press. pp. 377-396.
- Boerlijst, M., Nowak, M., & Sigmund, K. (1997). The logic of contrition. *Journal of Theoretical Biology*, **185**, 281-293.
- Bolton, G., Katok, E., & Ockenfels, A. (2003). Cooperation among strangers with limited information about reputation. *Working paper*. PHL PA: Laboratory for economic management and auctions, The Pennsylvania State University.

⁹ 間接互惠性研究が IS や ST を提出してきたのに対し、一般交換研究においては、その成立を可能とする戦略として、選別的提供戦略 (Takahashi, 2000)、下方 TFT (Boyd & Richerson, 1989; Yamagishi & Cook, 1993)、その他いくつかの条件付き提供戦略 (Takagi, 1996) が提唱されてきた。ただしここでは各戦略の詳細は割愛し、その名前を紹介するにとどめる。

- Boyd, R., & Richerson, P. (1989). The evolution of indirect reciprocity. *Social Networks*, **11**, 213-236.
- 大坊郁夫・安藤清志・池田謙一 (1989). 社会心理学 パースペクティブ 1 誠信書房
- (Daibo, I., Ando, K., & Ikeda, K. (1989). *Social psychology: Perspectives*. Vol. 1. Tokyo: Seishin Syobo.)
- Grefenstette, J. (2000). Proportional selection and sampling algorithms. In T. Back, D. Fogel & T. Michalewicz (Eds.), *Evolutionary computation 1: Basic algorithms and operators*. Bristol, UK: Institute of Physics Publishing. pp 172-180.
- Hardin, G. (1982). Discriminating altruism. *Zygon*, **17**, 163-186.
- Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, **268**, 745-753.
- Milinski, M., Semmann, D., Bakker, T., & Krambeck, H. (2001). Cooperation through indirect reciprocity: Image scoring or standing strategy? *Proceedings of the Royal Society of London. Series B. Biological Sciences*, **268**, 2495-2501.
- Nowak, M., & Sigmund, K. (1998a). Evolution of indirect reciprocity by image scoring. *Nature*, **393**, 573-577.
- Nowak, M., & Sigmund, K. (1998b). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, **194**, 561-574.
- Nowak, M., Sigmund, K., & El-Sedy, E. (1995). Automata, repeated games, and noise. *Journal of Mathematical Biology*, **33**, 703-732.
- Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: The importance of standing in the evolution of indirect reciprocity. *Journal of Theoretical Biology*, **224**, 115-126.
- Sugden, R. (1986). *The economics of rights, cooperation and welfare*. Oxford, UK; Basil Blackwell.
- Takagi, E. (1996). The generalized exchange perspective on the evolution of altruism. In W. Liebrand & D. Messick (Eds.), *Frontiers in social dilemmas research*. Berlin: Springer-Verlag. pp. 311-336.
- Takahashi, N. (2000). The emergence of generalized exchange. *American Journal of Sociology*, **105**, 1105-1134.
- 高橋伸幸・山岸俊男・林 直保子 (1999). 一般交換の自発的形成——選別的利他行動に関する実験研究—— 心理学研究, **70**, 9-16.
- (Takahashi, N., Yamagishi, T., & Hayashi, N. (1999). Spontaneous formation of generalized exchanges system: An experimental study of discriminating altruism. *Japanese Journal of Psychology*, **70**, 9-16.)
- Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, **46**, 35-57.
- Wright, S. (1943). Isolation by distance. *Genetics*, **28**, 114-138.
- Yamagishi, T., & Cook, K. (1993). Generalized exchange and social dilemmas. *Social Psychology Quarterly*, **56**, 235-248.
- Yamagishi, T., & Hayashi, N. (1996). Selective play: Social embeddedness of social dilemmas. In W. Liebrand & D. Messick (Eds.), *Frontiers in social dilemmas research*. Berlin: Springer-Verlag. pp. 363-384.

——2004. 4. 8 受稿, 2005. 5. 14 受理——