アップストリーム型とダウンストリーム型間接互恵性の統合モデルのダイナミクス分析

An integrated model of upstream and downstream reciprocity

佐々木達矢*1　内田智士*2　岡田勇*3　山本仁志*4
Tatsuya Sasaki　Satoshi Uchida　Isamu Okada　Hitoshi Yamamoto
*1 郡山女子大学短期大学部　Koriyama Women's College, *2 倫理研究所　RINRI Institute
*3 創価大学　Soka University, *4 立正大学　Rissho University

要旨: 人間社会における協力進化の主要なメカニズムの一つである間接互恵性は，アップストリーム型とダウンストリーム型に分けられるが，これまでほぼ個別に研究されてきた．特に，アップストリーム型は協力相手を選ばないため，そのままでは非協力者の搾取で消滅すると理論的に言われている．近年では，アップストリーム型とダウンストリーム型を統合させた実証研究も現れてきているが，その数理的研究は未だ少ない．そこで我々はこれまでの間接互恵性研究のフレームワークを応用して，これら二つのタイプの統合モデルを構築し，進化ゲーム理論的分析を行ってきた．今回は，その分析結果から，互恵主義者と非協力者の二戦略の間に，従来モデルでは通常見られなかった，安定的な共存均衡が存在できることを示す．

キーワード: 間接互恵性，アップストリーム互恵性，ダウンストリーム互恵性

Abstract: Indirect reciprocity is one of the major mechanisms for the evolution of cooperation in human societies. There are two types of indirect reciprocity, upstream reciprocity and downstream reciprocity. Cooperation in downstream reciprocity follows the pattern "I help you, and someone else will help me". The direction of cooperation is reversed in upstream reciprocity, which instead follows the pattern "You help me, and I will help someone else". In reality, these two different types of indirect reciprocity often occur in combination. In theory, however, upstream and downstream reciprocity have been studied mostly in isolation. Here we propose a standard model that integrates both types. We apply a framework for indirect reciprocity based on image scoring in pairwise donation games. We analyze the model by means of evolutionary game theory. We show that this model can explain the stable coexistence of reciprocators and defectors, a result that is new compared with the corresponding previous models.

Keywords: indirect reciprocity, upstream reciprocity, downstream reciprocity

1. はじめに

互恵的な協力関係は，持続可能社会には不可欠なものであり，Trivers (1971)による互恵的利他主義に関する研究から半世紀近く経った現在でも，その適応的な環境に関すること等，進化生物学や社会科学分野で盛んに研究されているテーマである．ただし，協力には通常コストがかかるため，ただ乗りが発生しやすい環境下では，無条件に協力する固体が進化できる可能性は乏しい．よって，従来の研究における標準的なパラダイムは，相手の協力度に応じた条件付き協力であり，互恵主義者が助けるべき相手を区別することは当然に考えられてきた．

互恵的関係は，相互作用が同一の相手との間で繰返される場合，X が Y を助ければ次に Y が X を助ける(図 1A)，直接互恵性と呼ばれる構図になる(e.g., Axelrod, 1984)．そのような同一相手との出会いの繰返しが期待できない場合，つまりより一般的な交換状況では，直接互恵性における固定二者間の関係が外部の第三者へ開かれた，間接的な関係となる(e.g., Alexander, 1987)．従って，間接的な互恵性を上手く働かせるには，次やり取りする可能性のある相手の情報を，直接観察や噂話、評判などを通して知り，自身の行動を決めていく必要がある．

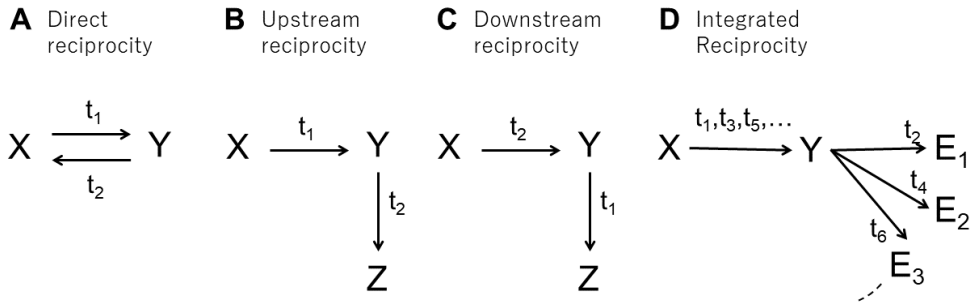

図 1. 互恵性メカニズムの構図. 各矢印のラベル t_i の添え字 i はその方向の C 行動が発生する順序を表す.

そして大事なポイントは, 間接互恵性は, アップストリーム型とダウンストリーム型の 2 つに大きく分けられるということである(Boyd and Richerson, 1989). まずダウンストリーム互恵性の構図は, Y が Z を助ければ, 次に X が Y を助ける, である(図 1C). つまり, Y の行為に対する応答は, Y が助けた Z から直接帰るのではなく, Y を評価する第三者 X によってなされる. これが所謂, 評判報酬(rewarding reputation)と呼ばれるものである(Baker and Bulkley, 2014).

一方, アップストリーム互恵性の場合は, ダウンストリーム互恵性の場合の構図が反転し, X が Y を助ければ, 次に Y が Z を助ける, となる(図 1B). これは, 条件付き協力のパラダイムに沿ったダウンストリームとは異なり, 直接的なお返しや評判による間接的な報いを期待して行動するのではなく, 感謝や恩義などの感情によって起きる, 利他的行動の連鎖であると説明され(Simpson et al., 2018), しばしば恩送り(paying it forward)とも呼ばれている.

この 2 種類の互恵性は, どちらも実験環境や実際のフィールドでよく観察できるものだが(e.g., Stanca, 2009; Yoshikawa et al., 2020), 進化ゲーム理論的には, アップストリーム互恵性は自然選択で淘汰されると分析されている. 従来研究によると, 単独では無理だが, 直接互恵性, 空間構造的(Nowak and Roch, 2007)等との組合せがあれば生き残り可能とされる.

1.1. 取り組む課題

一方現実では, 互恵性の様々異なるメカニズムが併せて適用されることは珍しくない(Melamed, et al., 2020). Baker and Bulkley (2014)は, 評判報酬と恩送りが互いに補完的なメカニズムであるとの仮説を立てている. 最近の実証研究では, 報酬を与える観察者がいる環境では, 恩送り行動がより寛大になることが示された(Simpson et al., 2018). しかし理論的には, アップストリーム互恵性とダウンストリーム互恵性は別個に研究されることがほとんどであり, これらの相互作用が協力行動の進化に与える影響は未だよく知られていない.

明らかなことは, これまでの研究には, 様々なタイプの互恵性, 特に, 異なるタイプの間接互恵性を総合的に理解するための理論的枠組みが欠けている点である. 今回のモデルでは, 直接互恵性や空間構造等を織り込むことなく, 間接互恵性のみに拠る手法で, 互恵的協力とフリーライダーとの間に安定した共存均衡を実現できることを示す.

具体的には, Baker and Bulkley (2014)で示唆された, 次のようなアップストリーム Y とダウンストリーム X のサイクル(図 1D)をモデル化し分析する. まず初めに X が Y を助ければ(t_1), これにより気持ちを動かされた Y は次に他の誰か Z_1 を助ける(t_2); 更にこの Y の行動を評価した X が Y に報いようと Y を助ける(t_3); するとまた同じく Y は他の誰か Z_2 へ恩送りする(t_4); 以降も同様に報酬と恩送りのサイクルが続く(t_5,t_6,…). この様に外に無条件に開かれたサイクルを, 同種の互恵性をペアにして起こさせようとしても結果は閉じたループとなり, 直接互恵性と変わりがない. 図 1D は, そのような開放性を備えた互恵性のサイクルとして最小であり, 総合的研究の基点としてふさわしいと考える.

2. 手法

2.1. 贈与ゲームを元にした間接互恵性モデル

今回のモデル構築にあたっては, 従来と同じく, 贈与ゲームでの評判に基づく間接互恵のフレームワークを使う(Ohtsuki and Iwasa, 2004,2006). このフレームワークでは, 互恵的戦略は行動ルールと評価ルールの組合せによって表現される. 贈与ゲームの進め方としては, まずランダムマッチングで母集団から二人のプレイヤーを選び, 次に贈与側と受取側の役割をランダムに割り振り, 1 ラウンド限りの贈与ゲームを行う. その時, 贈与側となったプレイヤーには, もう一方の受取側プレイヤーを助ける(C)/助けない(D)の選択肢が与えられる. そこで C をしたは, 受取側には$b > 0$の利益がもたらされ, 贈与側には$c > 0$のコストがかかる(ただし$b > c$). 逆に, D をした場合には, 贈与側と受取側どちらの利得にも影響はないとする.

ここで互恵的戦略を採用している場合の, 誰に C をするかの決定は, 前提とする行動ルールによって定められる. 今回この行動ルールは, (a)受取側プレイヤーの最新の評判値に拠るとし, 加えてアップストリーム互恵性を考慮するために, (b)贈与側プレイヤーが以前最後に受取側の時に何をされたかにも依存するとする.

評価値については, good (G)または bad (B)の 2 値で評価を行う. ゲームの各回後に, 贈与側プレイヤーの G/B は, (i)その回における贈与側プレイヤーの行動内容(C/D), および, (ii)贈与側プレイヤーが以前最後に受取側だった時に相手から何をされたか(C/D)で決まるとする(評価ルール).

(a)と(i)のペアは, よく知られているように, ダウンストリーム互恵性で最も研究されている一次情報モデルに対応する. これを(b)と(ii)に拡張した点は本研究の独自性の一つである. この拡張により, 恩送り行動に対する C を, 単なる C 行動に対する C から区別することが可能となる.

2.2. アップストリームとダウンストリームの統合

次に, 今回のメインである 2 つの互恵性の統合を説明する. ポイントは, アップストリーム互恵性とダウンストリーム互恵性の間に, 図 1D で示したような, 持ちつ持たれつの好循環を確立することである. 我々の考える統合モデルの行動ルールは： 前回 C を受けた場合は, 恩送り行動をする(相手の評判が何であれ C をする); また, 前回 D を受けた場合は, 今回の相手の評判が G なら C を, B なら D をする(つまり条件付きで協力する)(表 1). この行動ルールは, 特に断りのない限り, 以降全体で前提とする.

また, 今回の分析では, 統合型互恵戦略に加えて, 協力戦略と非協力戦略も考慮する. ゲーム内でのそれぞれの行動は, 無条件に相手を助ける(C)か, または無条件に助けない(D)かである.

表 1. 贈与側の行動ルール

	受取側が G	受取側が B
直近に C をされた	C	C
直近に D をされた	C	D

表 2. 贈与側の評価ルール I

	C を実行	D を実行
(他の条件なし)	G	B

表 3. 贈与側の評価ルール II

	C を実行	D を実行
直近に C をされた	G	B

この研究では, 統合型互恵戦略用に, 次の評価ルール I と II を考慮する. 評価ルール I では, C をする贈与側は G と評価され, D をする場合は B と見なされる(表 2). つまり Scoring ルール(Nowak and Sigmund, 2005)である. 評価ルール II では, もし以前最後に受取側であった際に C を受けたとき, 今回自分が贈与側で C を与えた場合は G と評価され, 一方, その前回 C を受けたにもかかわらず, 次に自分が贈与側で D をした場合は B と評価される(表 3). また, 以前最後に受取側であった際に D を受けた時は, そのプレイヤーの評価は贈与側としての行動によらず変化しないとする. これは所謂, Staying ルール(Sasaki et al., 2017)の一種である.

2.3. 情報観察と情報共有, そしてエラーについて

今回のモデルでは, 代表的な観察者によって各ゲームの情報は提供され, 同一の評価ルールに従う全てのプレイヤー間で平等に共有されるという状況を想定している. また簡易化のため, どの場合も贈与側プレイヤーが受取側のイメージを知っている確率は 100%と仮定する(完全情報). さらには, 双方向の行動実行エラーも考慮する.

2.4. 進化ダイナミクスと評判ダイナミクス

今回のモデルの分析は, 進化ゲーム理論の手法にレプリケータダイナミクス(Sigmund, 2010)で行う. よって, 無限母集団を仮定し, 解析を簡単にするため, 各個人はその生涯で, 毎回異なる相手との 1 ラウンド対戦のゲームを無限回プレイすると考える.レプリケータ方程式は一般的に, $ds/dt = s(P_S - P)$ と表され, そこでsは戦略Sを持つ個体集団の相対頻度, P_Sは戦略Sのラウンド当たりの期待利得である(P_Sは無限回ラウンドを繰返した末に与えられるものとする). そして P は集団全体の平均期待利得を表し, $\sum_S sP_S$で与えられる. 今回の 3 戦略分析の場合、協力戦略 X, 非協力戦略 Y, 統合型互恵戦略 Z の相対頻度をx, y, zとすると, $x + y + z = 1$で, $P = xP_X + yP_Y + zP_Z$である.

また, 各戦略集団内で G 評価を持つ個体の割合をg_Sとすると, 全集団内で G 評価を持つ個体の割合gは$g = xg_X + yg_Y + zg_Z$である. 更に, あるラウンドで戦略S(= X/Y/Z)を採用する贈与側プレイヤーが, 直近に行動 *Act* (= C/D)を受けていて, かつ現在の受取側が評価 *Img* (= G/B)の場合に, 行動した結果, 評価ルールによって G と評価される確率を, $g_{S,Act,Img}$とする. よって, ゲームを無限回繰返した末では, 各戦略で G 評価を持つ個体の割合g_Sは次式を満たす:

$$g_S = u_S\left[g\, g_{S,C,G} + (1-g)\, g_{S,C,B}\right] + (1-u_S)\left[g\, g_{S,D,G} + (1-g)\, g_{S,D,B}\right]. \quad (1)$$

この式で，u_Sは戦略Sを持つ個体が受取側の時に C を受ける確率を表し，戦略毎に以下の式で与えられる:

$$\begin{aligned} u_X &= xe + y\bar{e} + z[u_Z e + (1-u_Z)(g_X e + (1-g_X)\bar{e})] \\ u_Y &= xe + y\bar{e} + z[u_Z e + (1-u_Z)(g_Y e + (1-g_Y)\bar{e})] \\ u_Z &= xe + y\bar{e} + z[u_Z e + (1-u_Z)(g_Z e + (1-g_Z)\bar{e})] \end{aligned} \quad (2)$$

そこでeは行動実行エラーが発生しない確率を，$\bar{e}(=1-e)$はそのエラーが発生する確率を表す．基本的に，式(1,2)を解くことで，各状態(x,y,z)に応じて戦略毎の$g_S(x,y,z)$と$u_S(x,y,z)$を求められる．

毎ラウンド後に更新される評価ダイナミクスは十分速いと仮定する．これによりレプリケータダイナミクスは，式(1, 2)から求められる，評判が平衡状態にあるときのg_Sとu_Sを元にした期待利得によって決定される(注: 評価ダイナミクスの初期状態では全個体が G であると仮定)．各戦略の期待利得は

$$P_S = bu_S - cv_S \quad (3)$$

で与えられる．そこでv_Sは，戦略Sを持つ個体がゲームの贈与側である時に C を実行する確率を表し，戦略毎には以下の通りである:

$$\begin{aligned} v_X &= e \\ v_Y &= \bar{e} \\ v_Z &= u_Z e + (1-u_Z)[ge + (1-g)\bar{e}] \end{aligned} \quad (4)$$

3. 結果

3.1. 善と悪の共存

結果として今回のモデルでは，評価ルール I または II のいずれの場合でも，統合型互恵戦略と非協力戦略が共存する均衡を実現し，そこでは 100%では無いものの，実質的にハイレベルの協力を安定化させることが可能である(図 2,3)．共存状態では，統合型互恵戦略のアップストリーム互恵行動による無条件の協力は非協力戦略によって搾取されるものの，これは同じ戦略のダウンストリーム互恵行動の条件付き協力によって埋め合わすことができる．ここから非協力戦略と共存できる余地が生まれる．この結果は，そもそもアップストリームとダウンストリーム互恵性を今回のように統合することによるメリットであり，従来のようにダウンストリーム単独の間接互恵性モデルでは見られなかった現象である．以下では評価ルール毎にダイナミクスを詳しく確認する．

3.2. 評価ルール I

まず評価ルール I (表 2)では，統合型互恵戦略と非協力戦略との間に，孤立したアトラクターを持つことには成功するが，そのアトラクターは協力戦略の侵入に対して漸近的に安定しない．

図 2 では，その 3 戦略レプリケータダイナミクスを，エラー無し(図2A)と有り(図2B)の場合に分けて描いている．ルール I の場合，一般に，戦略毎の G の確率は

$$\begin{aligned} g_X &= e \\ g_Y &= \bar{e} \\ g_Z &= u_Z e + (1-u_Z)[ge + (1-g)\bar{e}] \end{aligned} \quad (5)$$

となる．式(2,5)より，

$$\begin{aligned} P_Z - P_Y &= g_Z[P_X - P_Y] \\ P_Z - P_X &= -(1-g_Z)[P_X - P_Y] \end{aligned} \quad (6)$$

が成り立ち、そこで

$$P_X - P_Y = bz(1-u_Z) - c \quad (7)$$

である．従って，この式(7)の零点集合が状態空間$\{(x,y,z): x+y+z=1\}$内部に均衡点の連続体を形成する．図 2A,B では内部曲線 PQ がそれにあたる．

式(6,7)から，<u>エラーなしの場合</u>，境界 ZY 上$(x=0)$ではg_Zの値は，$(3-\sqrt{5})/2 < z \le 1$の区間 ZR で，

$$g_Z = -\frac{z^2 - 3z + 1}{z} \quad (8)$$

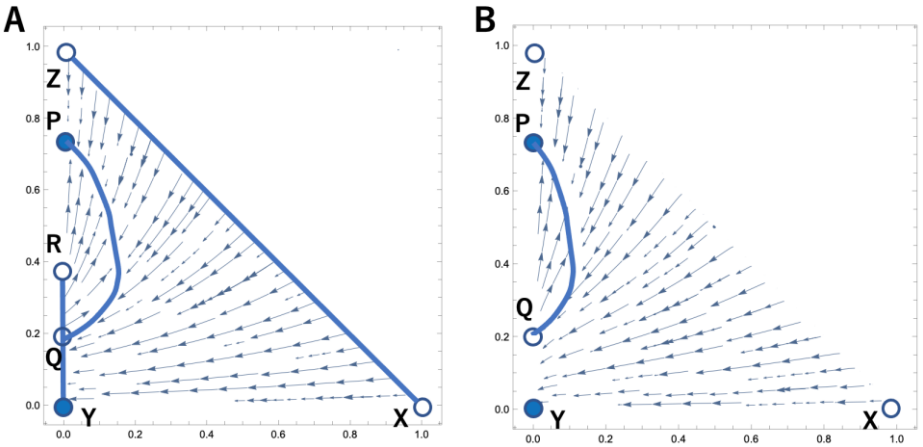

図 2. 評価ルール I の場合の 3 戦略ダイナミクス．$\boldsymbol{b=5}$，$\boldsymbol{c=1}$, A エラー無し$(\boldsymbol{e=1})$, B エラー有り$(\boldsymbol{e=0.95})$.

に, それ以外の区間 RY: $0 \le z < (3-\sqrt{5})/2$で, $g_Z = 0$に収束する. 区間 ZR で式(8)から式(7)の零点集合を求めると, $z = z_0$で,

$$z_0 = \frac{b-2c}{b-c} \tag{9}$$

という点が得られる. 区間 ZR では, このz座標を持つ点がアトラクターP であり, P は$z_0 > (3-\sqrt{5})/2$の場合に現れる. 一方, $g_Z = 0$となる区間 RY は, 式(6)より, 全ての点が均衡点となる.

また, 統合型互恵戦略 Z と協力戦略 X の間の境界ダイナミクスは中立で, 境界 ZX は全て均衡点の連続体となる. そして, 協力戦略 X と非協力戦略 Y の間の境界 XY 上では; Y が X を支配する.

次にエラー有り(図 2B)の場合, 数値計算から, 統合型互恵戦略 Z と非協力戦略 Y の間には, パラメータに応じて, 均衡点が 2 点現れ, 頂点 Z に近い方から順に, 境界アトラクターP と境界リペラーQ になる. エラー無しの場合の境界上の均衡点の連続体 RZ は消える一方, 内部の均衡点連続体 PQ は依然存在する. 中立ドリフトや突然変異を考慮した場合, 長期的には, 集団は点 P に止まることはできず, 最終的には頂点 Y の 100%非協力状態へ収束するとみられる(c.f. Scoring (Nowak and Sigmund, 2005)の場合).

協力戦略の侵入に対する点 P の脆弱性については次のように理解できる: 評価ルール I における G の定義は, 単に C をするかどうかであり, 従ってこれは, “正当化されない裏切り”という問題に繋がる(Okada, 2020; Yamamoto et al., 2020). 統合型互恵戦略の贈与側プレイヤーが B 評価の受取側プレイヤーに C をすることを拒否した場合, その贈与側プレイヤーの評判は B になり, 同じ戦略を持つ個体から C を受ける機会が減少することになる. このような B の連鎖が戦略集団内で発生すると, 最終的に, 条件付き協力により非協力戦略からの搾取を防ぐことで得た, 統合型互恵戦略の協力戦略に対する利得上の優位が失われることになる.

3.3. 評価ルール II

この問題に対処するために, 我々は新しく評価ルール II (表 3)を提案する. このルールは, ルール I からの変化を最小限に抑えながら, ダウンストリーム互恵行動で C する対象をアップストリーム互恵行動(をした個体)に集中させることを意図している. その結果, ルール I とは異なり, 統合型互恵戦略と非協力戦略の共存状態は協力戦略の侵入を許さないようになる.

ルール II の場合, G の確率は, Staying を考慮すると,

$$\begin{aligned} g_X &= u_X e + (1-u_X) g_X \\ g_Y &= u_Y \bar{e} + (1-u_Y) g_Y \\ g_Z &= u_Z e + (1-u_Z) g_Z \end{aligned} \tag{10}$$

と表され, 統合型互恵戦略の G 確率も定数となる: $g_X = e$, $g_Y = \bar{e}$, $g_Z = e$. そして, ルール I とは著しく対照的に, ルール II では内部平衡点は存在せず, 全ての内部軌道は状態空間の境界へ収束する. 実際,

$$P_Z - P_X = c(1-u_Z)(1-g_Z)(2e-1) \tag{11}$$

から, 内部空間では$(1-u_Z)(1-g_Z) \neq 0$のため, 非エラー率eが十分大きければ$P_Z - P_X > 0$が成り立つ.

次に, 統合型互恵戦略 Z と非協力戦略 Y の間のダイナミクスを確認する. 式(2-4,10)から, $x = 0$の場合,

$$\begin{aligned} &P_Z - P_Y \\ &= bz(1-u_Z)(2e-1) - c[u_Z + (1-u_Z)g_Z]. \end{aligned} \tag{12}$$

更に, エラー無し($e = 1$)の場合は

$$P_Z - P_Y = -z^2(b-c) + z(b-2c), \tag{13}$$

となる. よって$b > 2c$ならば, $z = z_0$が ZY 上の内部アトラクターとなる. その座標はルール I と同じく式(9)で与えられる. 他に, 境界 ZX, XY 上のダイナミクスはルール I の場合と変わらないことから, 以上を踏まえると, エラー無しの場合では, 内部軌道は全て点

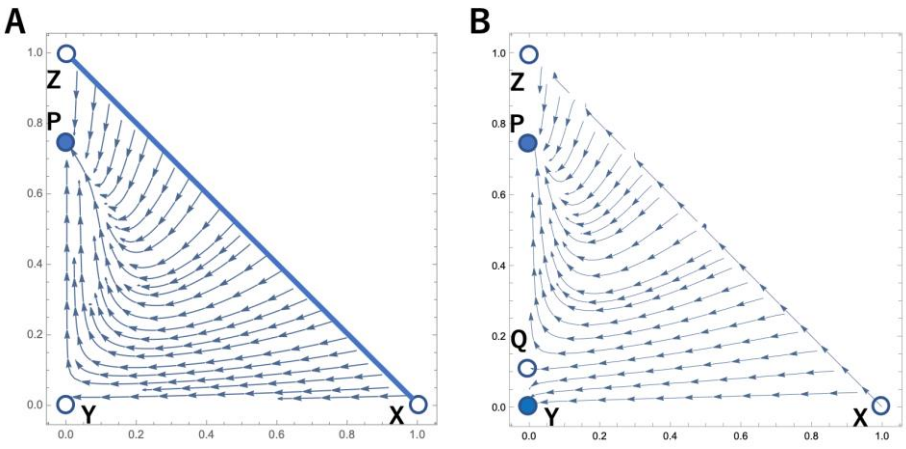

図 3. 評価ルール II の場合の 3 戦略ダイナミクス. $\boldsymbol{b = 5}$, $\boldsymbol{c = 1}$, A エラー無し($\boldsymbol{e = 1}$), B エラー有り($\boldsymbol{e = 0.9}$).

P: $(x, y, z) = (0, 1 - z_0, z_0)$に収束することが分かる. よって点Pは大域安定である(図3A). そして式(4)から, P における統合型互恵戦略の協力率は$z_0(2 - z_0)$, よって全体の協力率は$z_0^2(2 - z_0)$である.

またエラー有りの場合(図3B)では, ZY の間に均衡点がある場合は, アトラクターP だけでなくリペラーQ も加わりダイナミクスは双安定状態となる: 内部軌道は初期状態に応じて, P か頂点 Y へ収束する.協力戦略の侵入に対する点 P の頑健性については次の様に説明できる. 統合型互恵戦略のプレイヤーが, 仮に前回 D を受けたことから, 次に出会った評判 B の相手に対して D を行ったとしても, ルール II の Staying 要素のため評判の変化は起こらず, "正当化されない裏切り"の発生は防止される. よって, 統合型互恵戦略のプレイヤーは G を維持できるため, 引き続きダウンストリーム互恵行動から C を受けることができる.

4. 考察

アップストリーム互恵性の進化を説明する理論については, これまで直接互恵性や空間構造等と組み合わせたモデルが使われてきたが, 今回初めて, アップストリームとダウンストリーム互恵性が相互作用する数理モデルを我々は提案した. その結果, 今回の統合型モデルでは, エラーを前提にしなくとも, 適切な行動ルールと評価ルールを置くことで, 十分にハイレベルな協力をもたらす大域安定な均衡点が可能であることを示した. そこでは, 無条件協力戦略の侵入を抑止しながら, 互恵戦略と非協力戦略の共存がアトラクターとなっている.

これを踏まえ, ここでエラーの役割について注意したい. よく知られている通り, 互恵性や条件付き協力に関する従来の進化ゲーム理論的研究では, 全体で完全協力状態が確立されたときの, 無条件協力戦略により侵入されるリスクが大きな課題であった(無条件協力戦略がある程度広まれば, 次は非協力戦略の侵入に繋がる). よって殆どの先行研究では, 条件付き協力戦略が無条件協力戦略に対して自然選択での優位を得るにあたり, 協力エラーの発生が非常に重要な役割を果たしている(Ohtsuki and Iwasa, 2004; Brandt and Sigmund, 2006). この問題に関して, 我々のモデルは, 協力エラー要素を後からモデルに外挿するという従来のやり方ではない. 今回は, 非協力個体との共存が進化のプロセスを通じて内生的に構成された結果, この問題にも対処できている点に意義があると考える.

一方, 今回のモデルの限界の一つとしては, 得られた共存アトラクターが, 純粋なアップストリーム互恵戦略の侵入に対して不安定になる点が挙げられる. その戦略はダウンストリーム互恵行動をサボる分だけ, 今回の統合型互恵戦略に対する 2 次フリーライダーとなる. これついては, その様な戦略をダウンストリーム互恵で C を受ける対象から除外するように評価ルールを更新する方法が考えられ, 現在検証中である. また, 純粋ダウンストリーム互恵戦略についても, 影響の確認を進めている.

文　　献

Trivers, R., 1971. The evolution of reciprocal altruism. Q. Rev. Biol. 46, 35–57.

Axelrod, R., 1984. The Evolution of Cooperation. New York, NY: Basic Books.

Alexander, R.D., 1987. The Biology of Moral Systems. New York, NY: Aldine de Gruyter.

Boyd, R., Richerson, P.J., 1989. The evolution of indirect reciprocity. Soc. Netw. 11, 213–236.

Baker, W.E., Bulkley, N., 2014. Paying it forward vs. rewarding reputation: Mechanisms of generalized reciprocity. Organ. Sci. 25(5), 1493–1510.

Simpson, B., Harrell, A., Melamed, D., Heiserman, N., Negraia, D.V., 2018. The roots of reciprocity: Gratitude and reputation in generalized exchange systems. Am. Sociol. Rev. 83, 88–110.

Stanca, L., 2009. Measuring indirect reciprocity: Whose back do we scratch? J. Econ. Psychol. 30, 190-202.

Yoshikawa, K., Wu, C.H., Lee, H.J., 2020. Generalized exchange orientation: Conceptualization and scale development. J. Appl. Psychol. 105, 294.

Nowak, M.A., Roch, S., 2007. Upstream reciprocity and the evolution of gratitude. Proc. R. Soc. B 274, 605–610.

Melamed, D., Simpson, B., Abernathy, J., 2020. The robustness of reciprocity: Experimental evidence that each form of reciprocity is robust to the presence of other forms of reciprocity. Sci. Adv. 6, eaba0504.

Ohtsuki, H., Iwasa, Y., 2004. How should we define goodness?—reputation dynamics in indirect reciprocity. J. Theor. Biol. 231, 107–120.

Ohtsuki, H., Iwasa, Y., 2006. The leading eight: social norms that can maintain cooperation by indirect reciprocity. J. Theor. Biol. 239, 435–444.

Nowak, M.A., Sigmund, K., 2005. Evolution of indirect reciprocity. Nature 437, 1291–1298.

Sasaki, T., Okada, I., Nakai, Y., 2017. The evolution of conditional moral assessment in indirect reciprocity. Sci. Rep. 7, 1–8.

Sigmund, K., 2010. The Calculus of Selfishness. Princeton, NJ: Princeton University Press.

Okada, I., 2020. A review of theoretical studies on indirect reciprocity. Games 11, 27.

Yamamoto, H., Suzuki, T., Umetani, R., 2020. Justified defection is neither justified nor unjustified in indirect reciprocity. PLoS One 15, e0235137.

Brandt, H., Sigmund, K., 2006. The good, the bad and the discriminator—errors in direct and indirect reciprocity. J. Theor. Biol. 239, 183–194.