道徳判断の機能について

――進化的互恵性概念からの一仮説――

矢島 壮平

私たちが道徳的に善いとか悪いとか呼ぶ行為については、古来多くの道徳哲学者たちがそれを機能的観点から説明しようとしてきた。その典型であるのが功利主義者であり、彼らは大雑把に言って、行為の道徳性をその行為によりもたらされる帰結という観点から機能的に定義する。例えば、ある行為が快をもたらすのであれば、その行為は快をもたらすというその機能ゆえに善と判断されるのであり、逆にある行為が不快をもたらすのであれば、その行為はそうした機能ゆえに悪と判断される。義務論者でさえ、道徳的行為に何らかの機能性があることを認めることができる。(もちろん彼らは、その機能性ゆえに道徳的行為を行うということは否定するだろうが。)

こうした道徳的行為が持つ機能性の問題については論じられることが多かった一方で、道徳判断自体が持つ機能性については比較的等閑に付されがちであったと言える。私たちは何らかの行為について時にそれが道徳的に善いものであると判断し、また別のときにはそれが道徳的に悪いものであると判断する。しかし、私たちがどのような判断を下しているのか、あるいは下すべきであるのかについては頻繁に論じられる一方で、そもそも私たちがなぜそのような判断を下しているのかについて問われることは稀である。この問いについて問われることが少ないのは、おそらくは道徳的行為に機能性があることがかなり明確である一方で、道徳判断自体には何の機能性もないように思われるからかもしれないし、あるいは、道徳判断の機能性を問うこと自体に何の意義もないように思われるからかもしれない。しかし、そこに何の機能もないとすれば、私たちはなぜ道徳判断を下しているのだろうか？　あるいは、なぜそのような判断を下すようにできているのだろうか？　道徳判断の機能を(もしそれがあるとして)知ることができれば、例えばはさみが物を切るという機能を持つと知ることで私たちがそれをうまく使うことができるように、私たちは自身の（そして他人の）道徳判断をもっとうまく活用する方法を知ることができるのではないだろうか？

本稿では以上のような問題関心の下、生物学的進化理論における「互恵性

(reciprocity)」概念を手がかりとして、道徳判断の機能についていかなる仮説が導かれるか検討したい。なお、以下本稿で論じるのは、道徳判断の機能が事実としてどのようなものであるかについてであり、道徳判断の機能がいかなるものであるべきであるのかについてではない。近年は本邦でも進化理論から何らかの規範を正当化しようとする野心的な試みがいくつかある[1]が、本稿の目的はもっとはるかに控えめなものであることを最初に明記しておくのは、無用な誤解を招かないためにおそらく有用なことだろう。

1．進化的利他性と説明

互恵性の概念は、進化理論において生物個体の利他性（altruism）を説明する概念として編み出されたものである。進化的利他性について説明するために、最初に簡単に適応度（fitness）の概念について述べておきたい。適応度とは、基本的には集団内の各遺伝子の残りやすさについての相対的指標であり、それが個体について言われるときには、その個体の保持する遺伝子と同型の遺伝子の残りやすさについて言われている。進化理論における適応主義（adaptationism）の立場からすれば、生物個体は自然選択によって自らの適応度を高めるように進化してきたことになる。

しかし、仮に自然界に自個体の適応度を低めて他個体の適応度を高めるような行動的傾向性、すなわち利他性が存在している場合、それは適応主義の立場に対する反証となる。そして、現実の生物においては進化的利他性の例であるように外観上見える行動が多く存在しており、こうした深刻な理論的挑戦に応答することが前世紀後半の進化理論の重要な課題の一つであった。現在では、進化理論において利他性の例であるように見える行動を説明する主な概念として、（1）血縁選択（kin selection）、（2）互恵性、（3）集団選択（group selection）、（4）性選択（sexual selection）が挙げられる。（2）については後で詳しく説明するとして、（1）、（3）、（4）について簡単に説明しよう。

（1）の血縁選択は、血縁者間での利他行動が存在することを説明するものであり、進化生物学者のウィリアム・D・ハミルトンによって提唱された（Hamilton 1964）。彼の理論は次のようなものである。仮にAとBの2個体がここにいて、AにとってBは唯一の血縁個体であるとしよう。ここで、集団内の個体間で繁殖成功度に差をもたらすAの遺伝子が、Aの繁殖成功度に及ぼす効果をa、Aのそれらの遺伝子がBの繁殖成功度に及ぼす効果をbとする。また、共通祖先に由来す

る遺伝子である「同祖遺伝子（genes identical by descent）」の２個体間での共有率を「血縁度（relatedness, degree of relationship）」と呼ぶが、ここでＡとＢの血縁度をｒとしよう。このとき、Ａの上記遺伝子がＡとＢの繁殖成功度に及ぼす効果を無視したときのＡの繁殖成功度を１とすると、

１＋ａ＋ｒｂ

という値が「包括適応度（inclusive fitness）」と呼ばれる。何が「包括」であるのかというと、Ａの形質がＡの繁殖成功に及ぼす効果ａのみならず、Ａの形質がＢの繁殖成功に及ぼす効果、すなわち、ＢがＡと共有している同型の遺伝子の複製に及ぼす効果ｒｂについても考慮しているという点で「包括」なのである。そして、もしＡが持つ遺伝子によって発現する形質が

ａ＋ｒｂ＞０・・・(α)

を成立させる形質であれば、Ａの持つ遺伝子の複製率は高まり、その形質は進化する。また逆にその形質が、

ａ＋ｒｂ＜０

とするような形質であれば、Ａの持つ遺伝子の複製率は低まり、その形質は進化しない。血縁選択理論は、仮に上記（α）式（ハミルトン式と呼ばれる）が成立するのであれば、血縁者間で一見利他的であるように見える行動が進化することを示したのであり、集団遺伝学者Ｊ・Ｂ・Ｓ・ホールデンによるあまりに有名な言葉に即して言えば、私たちが２人の兄弟姉妹、あるいは８人のいとこの命を救うために自身の命を投げ打つことは、進化的に理にかなっているのである。

次に（３）の集団選択であるが、かつて、比較的最近の1990年代くらいまで、集団選択の主張は「集団存続のために有益な形質が進化する」というものであると広く認識されていた。この観点からの利他行動の説明は容易である。「自己を犠牲にして全体に奉仕する」というまさに全体主義的な形質が進化するのであり、そのような形質として利他行動形質が進化してきたと言える。しかし、進化生物学者のジョージ・Ｃ・ウィリアムズは、こうした集団を益する一方で個体の適応度を下げる形質、すなわち「生物相適応（biotic adaptation）」の概念が、自然選択が集団内での個体の適応度を上げる方向に働くという適応主義の理論上ありえないことを強く主張し、この集団選択批判は少なくとも進化生物学の内部において広く受け入れられた（Williams 1966）。

しかし一方で、ウィリアムズが批判の対象としたのは生物相適応の概念であり、彼は集団選択そのものを否定していたわけではなかった。実際、仮に利他性をもたらす遺伝子の頻度が集団（group）の内部で下がるとしても、複数の集団によっ

て構成される個体群（population）全体において上昇するのであれば、個体群全体として利他的形質は進化すると考えられるのであり、1990年代以降、進化生物学者のデイヴィッド・S・ウィルソンと哲学者のエリオット・ソーバーによって、こうした方向から進化的利他性を説明する理論としての集団選択の復権が図られている（Wilson & Sober 1994; Sober & Wilson 1998）。

（4）の性選択については、性選択自体はダーウィン以来の古い概念である。有性生物個体にとっては、配偶者選択（mate choice）は子孫を残すことに、そしてそれゆえ遺伝子の存続と適応度とに関わる重要なファクターであり、それゆえ有性生物個体は適応の観点において優秀な配偶者を選び、かつ選ばれるための形質を進化させてきたと考えられる。人間が持つ道徳的徳がそのようにして配偶者として選ばれるための徳として進化してきた可能性は十分にあり（Miller 2008）、利他性もそうした道徳的徳の一つとして説明することができる。

以上が互恵性以外の進化的利他性の説明として主立ったものである。注意してほしいのは、これらの説明がすべて「実際には進化的利他性など存在しない」とする説明であるという点である。どの説明においても、生物個体の一見利他的であるように思える行動は、その個体の（あるいは、その個体が持つ遺伝子の）適応度上の利益にかなっているのであり、それゆえこうした行動はどれも実際には進化的に「損して得を取る」行動であると言える。[2] しかし、次に見る互恵性により進化する利他的行動については、少なくとも一時的に適応度上の損失を被るのであり、その意味で真の利他的行動であると言える。

2．互恵性とは何か

歴史的に言えば、1964年のハミルトン論文によって血縁者間の利他的（に見える）行動については進化的説明が可能となったわけであるが、それは人間社会において広範かつ大規模に見られる非血縁者間の利他的行動を説明するものではなかった。そこに風穴を空けたのが進化生物学者のロバート・トリヴァースであり、彼が非血縁者間の利他性を説明する原理として提唱したのが互恵性である（Trivers 1971）。

２個体間で互いに時間差をおいて利他的行動を繰り返すことで、両者の適応度が上がる、というのが互恵的利他性の原理である。具体例を用いて説明しよう。例えば、血縁のないＡとＢの２個体がいたとしよう。まず、Ａが繁殖成功度上の損失１を被る代わりにＢが繁殖成功度上の利益２を得るような利他的行動をＡが

行うとする。次に、しばらく時間をおいて、今度はBの繁殖成功度上の損失が1、Aの繁殖成功度上の利益が2であるような利他的行動をBが行ったとする。こうして互いに利他的行動を行った結果、AとBはそれぞれ損失1に対して利益2を得ているので、最終的には共に繁殖成功度上の利益1を得ることになる。それゆえ、こうした利他的行動の交換を行う傾向性である互恵的利他性は、両者の繁殖成功度を上昇させることで両者の適応度を上昇させ、そのことにより進化すると言える。この意味で互恵的利他性は、長期的に見れば適応的利益をもたらすとはいえ、少なくとも一時的には適応度上の損失となる利他的な行動をもたらすのであり、真の意味での進化的利他性であると言える。

以上のような互恵的利他性が進化するには、以下の3つの条件が必要であるとされる。

1、利他的行動の場合は、行動を受ける側の繁殖成功度上の利益が行動を行う側の損失を上回ること。

2、特定の個体間である程度の長期にわたる相互作用が存在すること。

3、個体が他個体を識別し、さらに他個体との間で授受した行動を記憶することができるだけの認知能力が前適応として進化していること。

第1の条件が必要であるのは、もしこの条件が満たされないならば、互恵的利他性によって結果として得られるはずの利益がもたらされないからである。例えば、AとBが互いに行動を行う側の損失が2、行動を受ける側の利益が1の利他的行動を交換するとする。このとき行動を受ける側の利益は行動を行う側の損失より低いため、結果として両者は繁殖成功度上の損失1を被る。したがって、こうした利他的行動の交換は適応的ではない。

次に第2の条件だが、この条件が必要であるのは、互恵性においては同時に行動がやりとりされるのではなく、時間をおいて交互に利他的行動が繰り返されるからである。例えば、AがBに対して何らかの利他行動を行ったとして、AとBとの相互交渉がそれっきりになってしまえばAはBにお返しとして利他的行動をしてもらう機会を持たないのであり、Aは一方的な損失を被ることとなる。それゆえ互恵的利他性が進化するには、特定の個体間で利他的行動を交換し合えるだけの長い期間にわたって相互作用があることが条件となる。

第3の条件のうち、他個体識別能力については第2の条件が満たされるための条件であると考えてよい。特定の個体間で相互作用が続くには、まず個体同士が互いのことを識別していなければならないと考えられるからである。記憶能力については、どの個体にどのような行動を行い、どの個体からどのような行動を受

けたかが何らかの形で記憶されていなければ、そもそも特定の個体間で利他的行動の交換が行われ得ないと考えられるからである。

こうした互恵性は、2個体間で直接的に利他的行動を交換するので、「直接的互恵性（direct reciprocity）」と呼ばれる。直接的互恵性は、1対1の関係で時間を置いて互いに利他的行動をやりとりすることによって、結果的に両者が繁殖成功度上の利益を得るというものでり、それゆえ、互恵的利他性が進化するためには、特定の個体間の相互作用がある程度長く続く必要があった。つまり、ある人が相手に利他的行動をしたならば、利他的行動を施したまさにその相手から「直接的」に利他的行動を返してもらう必要があったこれに対し、2個体間のやり取りを観察する第三者を加えた3者間の相互作用を考える「間接的互恵性（indirect reciprocity）」の概念を提唱したのが、進化生物学者のリチャード・D・アレグザンダーである(Alexander 1987)。彼はそれを次のように説明している。

> 間接的互恵性においては、恩恵の受け手以外の誰かから返報を受けることが期待される。こうした返報は本質的に、集団内のどの個体もしくは個体の集合からでも与えられるだろう。間接的互恵性は評判と地位を伴い、社会集団内の全ての人が他人との相互作用に基づいて、過去と将来の相互作用の相手から絶えず評価（assess）され再評価されるという結果をもたらす。(ibid., p. 85)

> 私は間接的互恵性を、利害関心を持った観衆のいる前で生じる直接的互恵性の結果だと考える—その観衆というのは、自らの社会のメンバーを将来の潜在的な交渉相手として絶えず評価（evaluate）し続ける個体の集団であり、彼らはそれらの交渉相手から失うよりも多くのものを得る（こうした結果は、もちろん相互的でありうる）ことを望むのである。(ibid., pp. 93-94)

間接的互恵性においては、利他的行動を直接交換することによってのみならず、そうした交換をしている他個体を観察して評価することによって、また逆に、他個体によって自身の利他的行動を観察され評価されることによって、適応度上の利益を得ることが可能となる。それでは、そのような利益はどのようにして得られるのだろうか？　アレグザンダーは次のように整理している。

> 間接的互恵性による返報は、少なくとも主には次の三つの形を取るだろう。
> （1）利他的（beneficent）な個体は、直接的な互恵的相互作用における彼の

行動を観察し、彼を潜在的に報酬をもたらしてくれる相手だと判断した個体によって、後になって有益な互恵的相互作用へと引き込まれることになるだろう。(彼の「評判（reputation)」や「地位（status)」が高められ、彼の究極的利益につながる。)（2）利他的な個体は、集団全体もしくはその一部から直接的な補償（金銭や勲章や英雄としての社会的称揚など）によって報いられ、そのことで、彼（と彼の血縁者）が付加的な特典を得る可能性が高まるだろう。(3）利他的な個体は単純に、彼がそれに属して利他的に振る舞っているところの集団の成功が、彼自身の子孫と傍系親族の成功に貢献する、ということによって報われるだろう。(ibid, p. 94)

（3）については血縁選択や集団選択について言われていると言ってよく、また（2）については１種の直接的互恵性の結果であると言えるだろう。特に間接的互恵性に特有の適応度上の利益として着目すべきは（1）である。直接的な互恵的関係における自身の利他的行動を他個体に観察されることが、自身の「評判」や「地位」を高める効果を持ち、それが今後直接的互恵性の相手として選ばれ、適応度上の利益を得る機会へとつながっていくのである。

このように、互恵性には直接的互恵性と間接的互恵性があるとされるが、間接的互恵性の核心にある他個体の行動傾向の評価は、直接的互恵性の成立においても鍵となっているのであり、実は両者の違いはそれほど大きなものではない。このことを見るため、次に互恵性に関するゲーム理論的研究について見ていきたい。

３．互恵性とゲーム理論

進化生物学者のジョン・メイナード・スミスは、表現型（phenotype）としての行動のセットを「戦略（strategy)」と定義し、集団内の様々な戦略を持つ個体同士がある決まったルールの下で対戦して得点（適応度上の利益）を得て、その得点に応じて個々の戦略を持つ個体の頻度（つまりは個々の戦略の頻度であり、その戦略を表現型として持つ遺伝子の頻度）が次世代において変化する進化ゲームの理論を提唱した。そして、進化ゲームにおける対戦を何世代も繰り返した結果、もし集団内にある特定の戦略のみが存在するようになり他のどのような戦略も侵入できなくなった場合、その特定の戦略のことを「進化的に安定な戦略（evolutionarily stable strategy, ESS)」と呼んだ（Maynard Smith 1982)。つまりＥＳＳとは、いかなる突然変異戦略が集団内に侵入しようとも、常にそれらの戦略よ

りも高い点（つまりは適応度上の利益）を得ることができるような戦略である。進化ゲーム理論が目的とするところは、いかなる行動戦略が自然選択によって進化するのかを数理モデルによって導くことであり、例えば利他行動戦略がＥＳＳとなるのであれば、少なくともモデル上では利他行動が進化しうることが示されるわけである。

互恵性の進化に関するゲーム理論的研究は、進化ゲーム理論が最も活躍したと言える研究の一つであり、そこで用いられたのは、有名な「囚人のジレンマ（prisoner's dilemma）」ゲームである。囚人のジレンマゲームは次ページにあるような行列で表現される。プレイヤーＡとＢはお互いの間での情報交換や合意が全く無い状態で、それぞれ協力（cooperation）と裏切り（defection）の二つの選択肢のうちどちらかを同時に選ぶ。結果として得られるＡの得点が各欄の左下、Ｂの得点が各欄の右上の数字である。例えば、Ａが裏切り、Ｂが協力を選択した場合には、Ａが５点、Ｂが０点を得ることになる。欄中のＲは互いに協力し合ったことによって得られる報酬（reward）、Ｐは互いに裏切りあったことによって与えられる罰（punishment）、Ｔは裏切りによってより大きな得点を得られるという誘惑（temptation）、そしてＳは安易に協力して痛い目を見るお人よし（sucker）をそれぞれ表している。このゲームは次の二式を満たすものとして一般化でき、行列はその一例を示している。

Ｔ＞Ｒ＞Ｐ＞Ｓ・・・（１）

Ｒ＞（Ｔ＋Ｓ）／２・・・（２）[3]

		Player B	
		協力	裏切り
Player A	協力	R=3 R=3	T=5 S=0
	裏切り	S=0 T=5	P=1 P=1

ここで、もし両者が各自の得点を最大化しようとするとＴ＞ＲかつＰ＞Ｓであるので、相手の選択にかかわらず裏切りを選ぶことが自己の得点の最大化という観点から合理的であるということになる。上の行列で言えば、例えばＢが協力を選択した場合、Ａの得点はＡが協力すると３点、裏切ると５点となるので、Ａは裏切ったほうが高い得点を得られる。同様にＢが裏切りを選択した場合も、Ａの得点はＡが協力すると０点、裏切ると１点となるので、やはりＡは裏切ったほうが

高い得点を得られる。つまり、AはBの選択にかかわらず常に裏切りを選択した方がよいのであって、それはBについても同様である。そこで、AとBがもし自己の得点を最大化しようとするならば両者は共に裏切る事となり、それぞれ得点Pが得られることとなる。そしてR＞Pであるから、AとBは各自の得点を最大化しようとして選択した結果として、互いに協力し合っていたなら得ることができたはずの点数より低い点しか得ることができない、というジレンマに陥るのである。

こうした囚人のジレンマゲームがなぜ互恵性の研究に用いられるのかと言えば、このゲームにおいてプレイヤー同士が互いに協力することが、互恵性における利他的行動の交換を模していると考えられるからである。つまり、「協力」が「利他的行動を行う」という選択、「裏切り」が「利己的行動を行う」という選択であると考え、(1)互いに協力を選択するということは互恵性が成立しているということ、(2)一方が協力して他方が裏切っているということは一方の利他的行動に対して他方が利己的行動を行い相手から搾取したということ、そして(3)両者が裏切っているということは両者が互いに対して利己的行動を行ったということをそれぞれ意味していると考えるのである。

だが、この囚人のジレンマゲームは行われるのが一回限りの場合、合理解、すなわち各プレイヤーが自己の得点を最大化しようとする際に選ぶ選択肢は「裏切り」である。それゆえ、もしある個体が一生の間にどの特定の他個体とも一度きりしか交渉しないのであれば、その個体が自己の適応度上の利益を最大化するためにとるべき戦略は、どの個体に対しても「利己的行動を行う」という戦略であると言える。だが、現実には社会性生物は特定の個体同士で繰り返し相互作用するのであり、この条件が組み込まれたのが反復型囚人のジレンマゲームである。これは、囚人のジレンマゲームを特定のプレイヤー間で複数回繰り返し、その結果得られた総得点をそのプレイヤーの得点とするものである。一回限りの囚人のジレンマゲームでは「協力」か「裏切り」かの二者択一の戦略しかありえなかったのに対して、反復型ゲームではゲーム開始後の経過に関する情報を利用してプレイヤーが選択を決めることが許されるので、きわめて多様な戦略が可能となる。そうした多様な戦略の中でどの戦略が他の戦略を圧倒しうるのかの検証には膨大な計算が必要とされるので、通常コンピューターシミュレーションが用いられる。

政治学者のロバート・アクセルロッドは、世界中のゲーム理論の専門家に呼びかけてこの反復型囚人のジレンマゲームの戦略プログラムを募集し、それらのプログラムに完全に等しい確率で協力と裏切りを選択する戦略を加え、総当たりの

リーグ戦によるトーナメントを２度にわたって行った (Axelrod 1984)。第１回の参加者は１４名、第２回の参加者は６２名であったが、いずれの大会でも優勝したプログラムは「しっぺ返し（tit for tat, TFT)」と呼ばれる戦略であった。ＴＦＴはこの大会に応募されたプログラムの中でも極めて単純な部類に入る戦略であり、（１）必ず初回の対戦では協力を選択、（２）２回目以降は前回の相手の選択をまねる、というものである。このＴＦＴは、相手が「裏切り」を選択する傾向にあるときには「裏切り」を選択し、逆に相手が「協力」を選択してくる傾向にあるときには「協力」を選択する。つまりそれは、相手が利己的な個体のときには、こちらも利己的行動で返して搾取され続けることを避ける一方で、相手が利他的な個体のときには、こちらも利他的行動を行って互恵性の利益を得るという、相手の行動に応じて行動を変化させる条件的戦略（conditional strategy）である。ＴＦＴは厳密に言うとＥＳＳではない[4]が、それにきわめて近い強力な戦略であり、アクセルロッドの研究は互恵的利他性がＴＦＴのような条件的戦略として自然選択によって進化しうることを示していると言える。

囚人のジレンマは２者間で行われるゲームであるので、直接的互恵性を模したものであると言うことができるが、ここで注目されるのは、ＴＦＴが条件的戦略であるという点である。条件に応じて行動を変化させるには、その条件に対する「評価」というのが必然的に伴うのであり、そうした評価がなければそもそも条件的に行動することができない。ＴＦＴは対戦相手の行動に応じて自身の行動を変化させるという意味で条件的なのであり、そこに必然的に伴うのは、対戦相手の行動を評価するということである。対戦相手の前回の選択が「協力」と「裏切り」のいずれであったのか、すなわち、対戦相手の行動が「利他的」であったのか「利己的」であったのかを評価することができなければ、そもそもそうした条件に基づいて行動することが不可能である。このように直接的互恵性においても、間接的互恵性と同様に他個体の行動傾向の評価が不可欠の要素として組み込まれていると言える。したがって、間接的互恵性の場合は観察や伝聞によって他個体の行動傾向の評価の参考となる情報を得られるという意味で、情報源が直接交渉に限られる直接的互恵性よりはるかに情報量が多いという違いはあるものの、少なくとも原理的には、直接的互恵性と間接的互恵性との間に違いはないと言える。どちらの場合にも、他個体に対する評価が行動を決定する際の重要な要因となるのである。

結論——道徳判断の機能

数理生物学者のマーティン・ノワックとカール・ジグムントは互恵性に関する数々のシミュレーション研究を行っており、間接的互恵性に関する単純化された分析モデルについて検討する際、彼らは次のように述べている。

> 0（悪を意味する）と1（善を意味する）の二つの印象レベルのみが存在すると仮定しよう。(Nowak & Sigmund 1998, p. 575)

ここで彼らは（おそらく無意識的に）互恵性における評価を善悪の判断、すなわち道徳判断であるとみなしている。仮に互恵性における他個体の行動傾向の評価が、私たちが道徳判断と呼ぶものに相当するのであるとするなら、それが持つ適応的機能はかなり明確である。

道徳判断はしっぺ返し戦略がそうであったように、条件的適応行動を動機付けるのであり、その行動とは、利他的行動に対し利他的行動を、利己的行動に対して利己的行動を返すという、まさに互恵的（あるいは後者を考慮すれば、互酬的と言った方が適切であろうが）な行動である。これらの行為が適応的機能を持つことは、心理学者のエドワード・L・ソーンダイクが提唱した「効果の法則（law of effect)」と呼ばれる動物の学習傾向性と兼ね合わせて考えれば、かなり明確になる（Thorndike 1911, p. 244)。効果の法則とは平たく言えば、動物は一定の状況下で快をもたらす行動を行い、不快をもたらす行動を行わないように学習する、というものである。効果の法則を学習規則として既に進化的に獲得している動物は、利他的行動を行った際に利他的行動によって報いられ、そのことによって快を感じるのであれば、利他的行動を行うように学習し、逆に、利己的行動を行った際に利己的行動によって報いられ、そのことによって不快を感じるのであれば、利己的行動を行うように学習する。このことが意味するのは、私たちが他人の利他的行動を善であると判断し、そのことにより利他的行動を返すことを動機付けられるのであれば、その他人は利他的行動を行うよう学習するのであり、つまり私たちは善なる行為に善なる行為を返すことで、他人の善行を促進することになるのである。利己的行動についても同様に、悪行に対して悪行を返すことで、私たちは他人の悪行を抑止することができる。そしてこのことは、他人の利他的行動を受ける確率を高め、利己的行動を受ける確率を低めるいう意味で私たちの適

応的利害にかなうことになる。このように道徳判断により動機付けられる行動とは、典型的には私たちが賞罰と呼ぶ行動だろう。

この仮説が示唆することはいくつかある。例えば、こうした機能を持つ道徳判断が生得的であり、かつ、効果の法則のような学習規則が生得的でありさえすれば、道徳的行動傾向が必ずしも生得的傾向性でなくとも、私たちは周囲から与えられる賞罰などによって道徳的に行為することを学習しうるだろう。また、仮に道徳判断が上記機能を持つものとして生得的であるとするならば、私たちが善や悪と呼ぶ道徳的性質の（全体ではなく）少なくとも一部[5]が、進化的利他性／利己性という性質とかなり密接に結び付いているということになる。

以上のことは仮説以外の何物でもなく、ただ事態を整合的に説明するに過ぎないのかもしれない。しかし、私たちによって善いとか悪いとか呼ばれる行為自体のみならず、私たちの善いとか悪いという判断自体が何らかの機能を持つということを認識し、その事実を正確に把握しようと努めることは、私たちが何らかの規範を（それを究極的に正当化できるかどうかは別として）定めようとする際にもおそらくは役立つことであり、その意味でこうした仮説を提示することにも多少なりとも意義はあると言えるのではなかろうか。

註

[1] 例えば、内井（1996, 1998-2000）、内藤（2007, 2009）参照。本稿筆者は、進化的観点が道徳的事実の説明に多大な貢献を成しうると考えているが、一方で、進化理論が（そしておそらくは倫理学、道徳哲学が）何らかの規範の正当化を究極的に成しうるかどうかに関しては懐疑的である。この点において本稿筆者の態度はフィリップ・キッチャー（Kitcher 1993）に近いものであるが、キッチャーと本稿筆者との最大の違いは、キッチャーが進化理論による規範の正当化可能性のみを疑問視する一方で、本稿筆者は進化理論、道徳哲学理論を問わず、規範の究極的正当化可能性自体に対して懐疑的である点である。

[2] 集団選択が個体や遺伝子の適応度上の利益にかなっているかどうかは、議論の余地があるところかもしれない。例えばソーバーとウィルソンの主張であるのは、利他的な個体は、自身の集団内における相対適応度を下げつつも、自身が属する集団の個体群内における相対適応度を上げているがゆえに、その遺伝子の頻度を個体群内で上昇させることができるというものである。つまり彼らの主張は、利他的個体は集団内においては少なくとも相対適応度を下げているので、実際に進化的利他性は存在するというものである。しかし、ここでも集団レベルではなく個体群全体のレベルで見れば、利他的行動をもたらす遺伝子は（そしてそれゆえその遺伝子を持つ個体は）相対適応度を上げており、個体群全体として見れば進化的利他性は存在しないと言える。ソーバーとウィルソンはこれに対し、そのように個体群全体で遺伝子頻度を見ることを「平均化の誤謬（averaging fallacy）」と呼んで批判している（Sober & Wilson 1998, pp. 31ff）が、本稿では深く立ち入らない。

[3] この二つ目の式は、反復型囚人のジレンマゲームにおいてプレイヤーが互いに搾取し合っても、「互いに協力し合っていたなら得ることができたはずの点数より低い点しか得ることができない」というジレンマから抜け出すことができない、という条件を付けるための仮定である。

この仮定は、「互いに搾取し合うことで得られる両者の利益の総和よりも、協力し合うことで得られる両者の利益の総和の方が大きい」ということが、逆のケースよりも現実に近いという前提の下で、モデルを現実に近づけるために与えられるものである。

[4] ＴＦＴのみの集団は、全面協力戦略（ALLC）の進入を許し、遺伝的浮動によって集団内にＡＬＬＣが増えると、それを搾取する利己的な戦略の進入を許すこととなる。こうしたＴＦＴの欠点を補う戦略として「パブロフ（Pavlov）」という戦略が考案されている（Nowak and Sigmund 1993）。

[5] 私たちは道徳的性質を定義する際、どうしてもそれを全体として定義しうるシンプルな方法を無理に模索しがちであり、それがムーアの未決問題論法（open question argument）を有効たらしめた一つの理由であるように思える。しかし、道徳的善悪と呼ぶ性質が１種類しかないと考えるべき理由は特にないし、ある１つの対象が複数の道徳的性質を同時に具有することは当然ありうるのであり、私たちが道徳的ジレンマと呼ぶ事象はその典型である。５人の命を救うために１人の命を犠牲にすることは、善であると同時に悪であるからこそジレンマなのである。

参考文献

Alexander, R. D. (1987). *The Biology of Moral Systems.* Hawthorne, New York: Aldine De Gruyter.

Axelrod, R. (1984). *The Evolution of Cooperation.* New York: Basic Books. 邦訳『つきあい方の科学』（松田裕之訳）ミネルヴァ書房 1998 年

Hamilton, W. D. (1964). "The Genetical Evolution of Social Behavior." *Journal of Theoretical Biology* 7: 1-52.

Kitcher, P. (1993). "Four Ways of "Biologicizing" Ethics." Reprinted in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology 3rd Ed.*, Cambridge, Massachusetts: The MIT Press, 2006: 575-586.

Maynard Smith, J. (1982). *Evolution and the Theory of Games.* New York: Cambridge University Press. 邦訳『進化とゲーム理論：闘争の論理』（寺本英・ 梯正之訳） 産業図書 1985 年

Miller, G. (2008). "Kindness, Fidelity, and Other Sexually Selected Virtues." In W. Sinnott-Armstrong (ed.), *Moral Psychology Vol.1: The Evolution of Morality: Adaptations and Innateness*, Cambridge, Massachusetts: The MIT Press, 2008: 209-243.

Nowak, M. A. and K. Sigmund. (1993). "A Strategy of Win-stay, Lose-shift That Outperforms Tit-for-Tat in the Prisoner's Dilemma Game." *Nature* 364: 56-58.

Nowak, M. A. and K. Sigmund (1998) "Evolution of Indirect Reciprocity by Image Scoring." *Nature* 393: 573-576.

Sober, E and D. S. Wilson. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, Massachusetts: Harvard University Press.

Thorndike, E. L. (1911). *Animal intelligence.* New York: Macmillan.

Trivers, R. L. (1971). "The Evolution of Reciprocal Altruism." *The Quarterly Review of Biology 46*: 35-57.

Wilson, D. S. & Sober, E. (1994) "Reintroducing group selection to the human behavioral sciences." *Behav. Brain Sci. 17*: 585-654.

内藤淳 (2007). 『自然主義の人権論：人間の本性に基づく規範』. 勁草書房.

内藤淳 (2009). 『進化倫理学入門：利己的なのが結局、正しい』. 光文社.

内井惣七 (1996). 『進化論と倫理』. 世界思想社.

内井惣七 (1998-2000). 「道徳起源論から進化倫理学へ」. 『哲学研究』565 号、567 号、569 号. 京都哲学会.