

公正な社会への抵抗

スピノザと行動経済学

柴田健志

はじめに

「公正」な社会とは何であろうか。すべての人間が共通のルールを遵守して協力するならば、それが「公正」な社会といえるであろう。人間はそのような社会を求めてきた。では、人間はいったいなぜ「公正」な社会を求めるのであろうか。驚くべきことに、スピノザはその根底に「妬み」という感情を見出した。「妬み」とは自分が享受しえないものを享受する者に対する「憎しみ」にはかならない。他者に「罰」を与えたいという欲望はこの感情から発生している。例えば、フリーライダーに対して「罰」を与えるのはなぜか。それが「不正」な行為だからだと人間はいうだろう。しかしこの答えは表面的である。スピノザによれば、罰則の根底にはフリーライダーに対する「妬み」がある。みんなが自分の利己心を抑制して協力的に行動するのに乗じて自分の利己心を満足させる者への「妬み」が人間をして「罰」を与えさせる。人間がけつして他人の「不正」を許さないのは「妬み」という強烈な感情があるからなのだ。結局のところ、「公正」に違反するから「罰」を与えるのではなく、

く、「罰」を与えたいという欲望が「公正」を求める欲望として意識されているにすぎない。「公正」に違反するから「罰」を与えるという正義の論理がじつは転倒した論理であるということがここから明らかにされるだろう。この論理によって人間は他者に対する「妬み」を自分自身に対して覆い隠しているのだ。このような欺瞞に気づかない限り、人間は決して倫理的な生を生きることができないだろう。このような思想をスピノザがそのまま述べているわけではない。しかし、以下に示すように、主として「罰」に関する行動経済学の実証成果を援用してスピノザの論理を再構成していくと、近代以後の市民社会が自明とみなす「公正」に対して、スピノザが批判的な視点をもっていたという点が浮かび上がるのである。

1 罰

フリーライダーによって代表される「不正」な行為を罰すことに対する人間の意欲はきわめて高いという点を示した行動経済学の実証結果がある。通常2人のプレーヤーで行なわれる「独裁者ゲーム (Dictator Game)」に3人目のプレーヤーを加えたゲームがある (Fehr&Fischbacher 2004)。「独裁者ゲーム」においては、Aに対して10万円 (ここでは便宜的に日本円にする) が原資として与えられる。Aはその金額を自分とBで分けるが、それぞれの取り分はAが単独で決定するものとする。どんな分け方をしてAは報復されないというのがルールである。だから「独裁者ゲーム」なのだ。このゲームのもともとのねらいは公平な配分に対する人間の傾向を見ることにあるが、その点は省略しよう。ここに

3人目のプレーヤーであるCが導入される。CとはAを罰することのできるプレーヤーである。このゲームでは「第三者 (The Third Party)」と呼ばれる。Aが自分の取り分を多くするような不公平な配分を行なった場合、Cは一定の金額を支払ってAを罰する(罰金を課す)ことができる。実証結果によれば、5万円ずつの配分でAを罰する「第三者」はもちろん存在しないが、Aの取り分が6万円になった時点で60パーセント以上の「第三者」がAを罰するのである。「第三者」であるC自身はAの行動によって何ら不利益を被っておらず、しかも罰するためのコストはCが自分で負担するという点を考慮すれば、驚くべき数字であるといわなければならないであろう。

では、人間はどうしてこれほどまでに「不正」を罰したいと思うのであろうか。じつはこの問いかけ自体はべつに難しい問いかけではない。人間は「不正」な行為をおこなった者に対して「憎しみ」あるいは「怒り」を感じるがゆえに彼らを罰したいと思うのである。例えば、被験者がフリーライダーに対して感じる「怒り」の程度をレベル1からレベル7までの7段階(レベル1:まったく感じない、レベル7:非常に強く感じる)で評価した場合、被験者の47パーセントがレベル6ないし7を選択し、37パーセントがレベル5を選択したという調査結果がある(Fehr & Gächter 2002)。8割以上の人間が強い「怒り」を感じるのだ。罰したいという欲望が「怒り」から生じていることはまず間違いないであろう。フリーライダーを罰するとき人間が快を感じるということが神経学的に実証されている (Singer & Steinbeis 2009)。ここが何よりの証拠である。自分が「憎しみ」ないし「怒り」を感じている人間が害悪を被ることは人間にとって快以外の何ものでもないからだ。スピノザもこの点を明言

している。「自分の憎むものが悲しみに刺激されることを表象するひとは喜びを感じるであろう」(Eth. III23P)。

では人間はいつたいどうしてフリーライダーのような「不正」な行為をおこなう者に対して「憎しみ」を感じるのであろうか。この点を考えてみなければならない。

2 憎しみ

人間がフリーライダーに対して「憎しみ」を感じるということはある意味で自明のことである。フリーライダーは相互的な義務を命じる「公正」の原理に反する。だから彼らに対して「憎しみ」の感情をもつことは当然なのだ。しかし、この説明は明らかに皮相である。相互的な義務を遵守しない人間を眼にしたらおのずと「憎しみ」がわき上がってくるという点を見えていないからである。フリーライダーへの対応が、相互的な義務を命じる「公正」の原理に反するという認識を出発点にしているというのであれば、フリーライダーに対して冷静に対処することもできるはずである。いや、そうでなければならぬ。しかし、現実にはそうっていない。人間はフリーライダーに対して本能的に強い「憎しみ」を感じるのだ。では、彼らに対するこの「憎しみ」はいつたいどこから発生するるのであろうか。「公正」の原理を前提せずにその発生が説明できなければならない。

フリーライダーとはみんなが協力的であることを利用して自分だけ得をしようとする人間を指す。みんなが同等に負担しているのに自分だけは負担を逃れ、にもかかわらずみんなと同等の恩恵を受ける人間だ。負

担なしに同等の恩恵を受けるのだから、結局はみんなよりも多くの恩恵を受けていることになる。以下に見るように、行動経済学が「公共財ゲーム (Common Goods Game)」によってモデル化したのはこのような状況である (Fehr & Gächter 2002)。

金額を日本円にしてこのモデルを具体的に示すと次のようになる。4人のプレーヤーに対して1人10万円の原資が与えられる。プレーヤーがそれを出資すると出資額の2倍の金額が得られるというルールになっている。また、その金額はプレーヤー全員に等分に配当されることになっている。したがって、4人全員が10万円出資すれば40万円が2倍になり80万円が得られる。それを4人で分ければ配当金は1人20万円になるだろう。協力的に行動することで元手が倍に増えたわけである。ところが、3人のプレーヤーが10万円ずつ出資し1人がまったく出資しなかった場合、出資金30万円の2倍の60万円が得られるが、4人で分けるというルールはかわらないので、1人12万円の配当になる。ということは、出資した3人が12万円の配当を受け取るだけなのに対して、出資しなかったプレーヤーは手元に残した10万円プラス12万円の配当を受け取ることになる。これがフリーライダーのモデルである。このように、他人よりも多くの恩恵を受けているということは、見方を変えれば、他人の所有できないものを所有しているということだ。実際、フリーライダーは他の3人が出資してしまった10万円を所有しているのだ。そこで、きわめて形式的に、フリーライダーに対する「憎しみ」は自分が所有できないものを所有している人間への「憎しみ」であると考えてみるができる。フリーライダーに対する「憎しみ」がどうして発生するかを説明するために「公正」を前提できないとすればこ

う考えてみるほかないだろう。この観点から、フリーライダーの問題というよりも、むしろより一般的な問題として次のように問いかけてみなければならぬ。——はたして人間は自分の所有しないものを所有している人間に対して「憎しみ」の感情をもつのであろうか。もつとすれば、いったいどうしてなのであろうか。

3 妬み

スピノザは上記のような「憎しみ」を「妬み」と呼び、その発生を証明している。スピノザによれば、誰かが何らかの事物を所有しそれを「享受」していることを表象すると、人間はおのずと自分もその事物を「享受」したいと感じる (Eh. III32Dem.)。人間は他者の感情を表象するとほとんど自動的に同じ感情をもつからだ。スピノザはこれを「感情の模倣」(Eh. III27Sch.)と呼んでいる。

さて、他者が「享受」するものを自分も「享受」したいという欲望は、まさに他人がそれを「享受」していることによつて否定される。それゆえ人間は「彼にそれを所有させないように努めるであらう」(Eh. III32P.)。こゝして、自分が「享受」したいものを「享受」している人間に対する「憎しみ」の感情が生じる。それが「妬み」である。スピノザによれば、「妬み」とは「人間が他人の不幸を喜び、また反対に他人の幸福を悲しむようにさせると考えられるかぎりでの憎しみにほかならぬ」(Eh. III24Sch.)。事実、自分が「享受」できないものを「享受」している人がそれを失うことを考えると、人間は「喜び」を感じるのだ。ところで、自分が「享受」できないものを他者が「享受」すること自

体はなら「不正」ではない。人間は「不正」の認識とはまったく関係なくただたんに「妬み」の感情をもつのだ。ところが、「不正」という概念はまさにこの「妬み」から発生すると考えられるのである。この点を理解するには、次の定理の意味を鮮明に認識しなければならぬ。

「善および悪の認識は我々が意識するかぎりにおける喜びおよび悲しみの感情にはかならない」(Eh. IV8P)。

この定理を援用すれば、人間が他者に対する「妬み」という自分の「悲しみ」を意識することで「不正」という「悪」の概念が発生するのである。人間は他者が「不正」をおこなっていると判断するがゆえにその人に対して「憎しみ」を感じるのではない。他者に対して「憎しみ」を感じるがゆえにその人を「不正」と判断するのである。彼は私のもっていないものをもっている。だから「不正」だというわけなのだ。

行動経済学における「最後通牒ゲーム」(Ultimatum Game)の実証結果によつてこの論理が支持されていると考えることができる。「最後通牒ゲーム」とは、2人のプレーヤーのうちの1人に与えられた金額をそのプレーヤーが分配するというルールでおこなわれる。日本円でこのゲームを具体的に示すと次のようになる。10万円がAに与えられる。AはBに対して自分の裁量で一定の金額を配分し残りを自分の手元に残すことができる。Bが金額の配分に関するAの提案を受諾した場合、2人のプレーヤーはそれぞれAの提案した金額を受け取ることができる。しかしBがAの提案を拒否した場合、2人のプレーヤーともに何も受け取ることができない。これが「最後通牒」という名称の由来である。問

題はBがAの提案を拒否するという行動をとることが現実にかかるかどうかである。実証結果はある意味で驚くべきものである。たとえば、Aに対して9万円、Bに対して1万円という配分をAが提案した場合、その提案はBによつて拒否される確率がきわめて高いことが実証されている。Bは1万円よりも0円を選択するわけである。

どうしてそんな行動をとるのであるのか。BはAに9万円を受け取らせたくないからである。そのために自分の取り分までなくなってしまうことも厭わないのである。行動経済学によれば、この行動はAに対する怒りがもたらした行動として理解することができる(Sanft et al. 2003)。しかし問題はいつたい何がBをそこまで怒らせたかという点にある。Aの提案は「不正」であるという判断がBを怒らせたのであろうか。言い換えれば、Bの怒りは「公正」な配分という規範の認識から生じているのであろうか。この点を考えてみなければならない。そのために、以下の実証結果を参照してみよう。Bの脳状態をfMRIでモニターした結果、「怒り」「苦痛」「嫌悪」等の感情をもつときに活性化される領域(anterior insula)に大きな活動がみられるが、効率性の認知に関する領域(dorsolateral prefrontal cortex)での活動は比較的小さい(Sanft et al. 2003)。すなわち、他者に対する「怒り」が効率性(損得の計算)を上回った結果としてBはこのような行動していると考えられる。その証拠に、これらの活性化の割合が逆転した場合には、「怒り」を押さえ1万円を受け取るという行動が見られる(Sanft et al. 2003)。もしかりにBの行動が規範の認識にしたがつてなされたのであればこういふことはありえない。もつとも、あとからきかれれば、BはAの配分が「不正」だと答えるだろう。しかし、その認識にもとづいて行動したわけではない。

いということは脳の活動に示されているのである。では、「不正」という認識はいったいどこから発生したのであるうか。スピノザのいうように、それは「妬み」が意識されたものだと考えられるのである。

4 共感

「最後通牒ゲーム」によってモデル化されているのは2人の人間のあいだに生じうる「憎しみ」である。したがって、この一般モデルはフリーライダーの分析には使用できない。フリーライダーとはただたんに自分にとって害になる人間のことでなく、むしろみんなにとって害になるとみなされる人間だからである。そういう人間に対する「憎しみ」がどのように発生するかという問題にアプローチするには、はじめに出てきた「独裁者ゲーム」の「第三者」をモデルにしたほうが適切だろう。そのため、上記の結論を踏まえた上で、他者に対する「共感」の作用に注目しなければならない。

他者に対する「憎しみ」が感じられる典型的な事例は、自分が他者から危害を加えられた場合に見られる。しかし、「独裁者ゲーム」を観察している「第三者」のように、自分以外の人間が危害を加えられるのを見たときにも、その加害者に対する「憎しみ」が生じるであろう。自分は何も被害を被っていないのに、どうしてこういうことになるのであるうか。被害者に「共感」するからである。このような「憎しみ」をスピノザは「義憤 (Indignatio)」 (*Eth. III22Sch.*) と呼んでいる。スピノザはこのような感情の発生を以下のように説明している。

スピノザの用語法では「共感」は「憐憫 (Commisericordia)」と呼ばれ

る。スピノザは「憐憫」とは「他人の被害から生じる悲しみ」 (*Eth. III22Sch.*) であると説明した上で次のように述べる。

「他人に善をなした人に対する愛をわれわれは好意と呼び、これに対し他人に悪をなした人に対する憎しみを義憤と呼ぶであろう」 (*Eth. III22Sch.*)。

このテキストでは「好意 (Favor)」および「義憤」が言及されている。ここではもっぱら「義憤」に着目しなければならない。他人が何らかの被害を被ったことに対して人間は「憐憫」を感じる。すなわち他人への「共感」によって「悲しみ」が発生する。ところで、他人の「悲しみ」の原因となっているのは加害者である。そこで、人間は他人の「悲しみ」だけでなく、その「悲しみ」を引き起こしたと考えられる原因つまり加害者をも認知した場合には、被害者とその加害者に対して「憎しみ」を感じるのと同じように、加害者に対して「憎しみ」を感じるだろう。「共感」のはたらしきによって、加害者に対する第三者の「憎しみ」すなわち「義憤」はこうして発生するのだ。

人間は個体発生のごく初期の時期にこのような「憎しみ」を感じることで発達心理学的に明らかにされている。驚くべきことに、3ヶ月の幼児にすでにこのような感情が観察されることが実証されているのだ。それは「共感による怒り (Emphatic Anger)」と呼ばれる (*Hoffman 2000:96ff.*)。幼児は誰かが目的を達成することを援助する人物を好むが、それを妨害しようとする人間を嫌うのだ (*Hanlin et al. 2007, 2010*)。さらに、8ヶ月の幼児は、非協力的な人間に対して積極的に振舞う人間

よりも、むしろ否定的な振舞いを見せる人間の方を好む (Hanlin et al. 2011)。非協力的な人間に対して「憎しみ」を感じているからこそ、その人間を否定的に取扱う人間は好まれるのだ。重要な点は、まだ文化的な条件づけのない8ヶ月の幼児にこのような感情のはたらきが観察できるといふ点である。

「共感による怒り」の延長線上に分配の公平さに対する関心の発達がみられる (Geraci & Surian 2011)。12ヶ月〜18ヶ月の幼児が分配の公平さに関心を示すことが実証されているのだ。ある子どもが他の子どもより少なくものを分けられることに對して、2歳にも満たない幼児が反対するのだ。しかし、子どもは自分が損をしていないのにいったいなぜ反対するのだろうか。そこには「共感」のはたらきが認められなければならない。「共感による怒り」に関する上記の研究を前提すれば、当然のように考えられる。つまり、他の子どもより多く分配された子どもに對する「憎しみ」を感じて不公平な分配に反対するのだ。この「憎しみ」はスピノザのいう「妬み」であろう。というのも、自分より多くのものを分配された子どもは、自分のもっていないものをもっていることになるからだ。そうであるとすれば、「人間は本性からして憎しみおよび妬みに傾いていることは明らかである」(Eth. III 555sch.) というスピノザの主張はまさしく正解である。なぜなら、18ヶ月までの幼児を對象にした以上の実証結果を踏まえれば、分配の公平さに関する関心が文化的に身につけられたものであるとは考えられないからだ。それはまさしく「本性」と呼ばれるべきものなのだ。

ところで、子どもは自分の目の前でおこなわれる不公平な分配に反対するだけではない。就学年齢に達した児童においては、不公平な分配の

物語を言語によって伝達するだけで同じことが起こる (Nichols 2010)。つまり、実際に不公平な分配を知覚のレベルで経験しなくても、ただだんに誰かが人より多くものを所有していると表象するだけで「妬み」が発生するのである。実際、知覚経験のレベルで発生する「ボトムアップ」型の「共感」だけでなく知覚経験に依存しない「トップダウン」型の「共感」が神経科学においても注目されている (Singer & Lamm 2009, Decety & Lamm 2006)。このように表象のレベルで「共感」が作用するのでとすれば、「共感」すべき人間の数は現実に知覚される人間の数を容易に越えることができる(大平 2015)。スピノザも「人々(Homines)」(Eth. III 29p.) という用語でそのような不特定多数の人間の表象に言及している。

さて、スピノザのいう「人々」の犠牲の上でものを所有している人間を表象する場合、この「妬み」はきわめて強いものになるであろう。多くの人間の「妬み」に「共感」がはたらくからである。とすれば、フリーライダーという反社会的な行為をはたらく人間に對する「憎しみ」は非常に強いものになるはずである。人間は自分が大きなコストを負担することになったとしても「不正」な人間を罰するという行動をとることが行動経済学において実証されている。このことはすでに言及した。いわゆる「第三者」による罰にかんする実証研究である。ちなみに、この「第三者」による罰は別名「利他的罰(Altruistic Punishment)」と呼ばれている (Fehr & Gächter 2002)。「不正」を罰することは自分には何の得にもならないが、自分以外の人たちにとって有益である。いちど罰せられた者は次もまた罰せられると予想してフリーライダー行為をしなくなるからである。だから「利他的罰」なのである。ある意味で不可解な行動

だが、多くの人間の「憎しみ」に対する「共感」がこの行動の動機づけとなっていると考えればそれも納得がいく。「憎しみ」が大きい分だけ罰することへの欲望も強いのだ。

おわりに

では、以上からいったい何が結論できるのであろうか。すでに述べたように、「不正」の概念とは「妬み」の感情から派生したものである。「妬み」の感情が存在しなければ「不正」の概念は存在しない。したがってまたそれと対になるべき「公正」の概念も存在しない。つまり、人間が「公正」な社会を求めるということは、じつのところフリーライダーを罰したいということの意味しているにすぎない。「妬み」とはフリーライダーを罰したいという欲望そのものだからである。

ところが、いったん「不正」の概念が成立すると、人間はそれがはじめから認知されていたものだと考えるだろう。そうすると、フリーライダーに対する罰則の要求はもともと「妬み」から発しているはずなのに、もはやそのようには認識されない。それはむしろ「不正」な人間は罰せられなければならないという理性の判断に転化するだろう。人間は、自分たちが「公正」な社会を求めがゆえにフリーライダーに対する罰則を求めるのだと考えるようになるわけである。こうして「公正」の概念は市民社会を支持する正義の基準となるのだ。ここに欺瞞が隠されていることは明らかだろう。たしかに、「公正」の概念によって罰は合理的に説明することができる。しかし、だからといって「罰」という概念が「公正」の概念から出てきたということにはならないはずなのだ。

スピノザの考えによれば、道徳的判断は感情をもとに自動的に行なわれ、その判断に関する合理的な説明はつねに後からつけ加えられる。スピノザのいう「感情 (Affectus)」とは「身体の変容の観念」(Eth. III.3D.)であり、これが道徳的判断のベースとなっている。ところが、いかなる観念についてもその「観念」がある。それが「観念の観念」(Eth. II.2Den.)である。人間がその道徳的判断を合理的に説明するのは「観念の観念」の次元においてである。しかし、「観念」すなわち「感情」が与えられなければ「観念の観念」など存在しない。「妬み」という「観念」がなければ「不正」という「観念の観念」は存在しないのだ。だから道徳的判断が合理的に説明できたからといって、合理的な説明が道徳的判断を生み出したということには全然ならないのだ¹⁾。

「公正」に違反するから「罰」を与えるという正義の論理が転倒した論理であるというばかりではない。この論理によって、人間は他者に対する「妬み」をそうとは知らずに自分自身に対して覆い隠すことになっているからだ。このような欺瞞に気づかない限り、人間は決して倫理的な生を生きることができないであろう。以上の論理に倫理的な意義があるとすれば、それはこの点であるといわねばならない。というのも、スピノザが倫理的な生を生きるための「理性の命令」を締めくくるのは「残りの人間たちのために欲しないようなものでなければ自分のために求めない」(Eth. IV.18Sch.)という規則だが、これを実行するには他者に対する「妬み」が克服されなければならないと考えられるからである。

『エチカ』の参照箇所は以下の要領で本文中に挿入す。

【略号】定義：D. 定理：P. 証明：Dem. 注解：Sch.

【例】『エチカ』第一部定理5証明：Eth. 125 Dem.

注

(1) この点でスピノザは現代の道徳心理学 (Haidt 2001, Greene 2008) の主張を先取りしている点に留意しよう。

文献

- Decety, Jean, Lamm, Claus 2006, "Human Empathy Through the Lens of Social Neuroscience," *The Scientific World Journal* 6, 1146-1163
- Fehr, Ernst, Gächter, Simon 2002, "Altruistic Punishment in Humans," *Nature* vol.415, 137-140
- Fehr, Ernst, Fischbacher, Urs 2004, "Third-Party Punishment and Social Norms," *Evolution and Human Behavior* 25, 63-87
- Geraci, Alessandra, Surian, Luca 2011, "The Developmental Roots of Fairness: Infants' Reactions to Equal and Unequal Distributions of Resources," *Developmental Science* 14-5, 1012-1020
- Greene, Joshua D. 2008, "The Secret Joke of Kant's Soul," Walter Sinnott-Armstrong (ed.) *Moral Psychology Vol.3 The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, 35-79
- Haidt, Jonathan 2001, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* vol.108 no.4, 814-834
- Hamlin, J. Kiley, Wynn, Karen, Bloom 2007, "Social Evaluation by Preverbal Infants," *Nature* vol.450, 557-559
- Hamlin, J. Kiley, Wynn, Karen, Bloom 2010, "Three-Month-Olds Show a Negativity Bias in Their Social Evaluations," *Developmental Science* 13:6, 923-929
- Hamlin, J. Kiley, Wynn, Karen, Bloom, Paul, Mahajan, Neha 2011, "How Infants and Toddlers React to Antisocial Others," *PNAS* vol.108 no.50, 19931-19936
- Hoffman, Martin 2000, *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge UP
- Nichols, Shaun 2010, "Emotions, Norms, and the Genealogy of Fairness," *Politics, Philosophy & Economics* 9-1, SAGE Publications, 1-22
- 大平英樹 2015 「共感を創発する原理」『エコーン・スタディーズ』第一巻第一号56-62
- Sanfey, Alan, Rilling, James K., Aronson, Jessica A., Nystrom, Leigh E., Cohen, Jonathan D. 2003, "The Neural Basis of Economic Decision-Making in the Ultimatum Game," *Science* vol.300, 1755-1758
- Singer, Tania, Steinbeis, Nikolaus 2009, "Differential Roles of Fairness- and Compassion-Based Motivations for Cooperation, Defection, and Punishment," *Annals of New York Academy of Science*, 1167, 41-50
- Singer, Tania, Lamm, Claus 2009, "The Social Neuroscience of Empathy," *Annals of the New York Academy of Sciences* vol.1156, 81-96
- Spinoza 1972, *OPERA II*, Heideberg