

---

## 特集「ネオ・サイバネティクス」・論文

---

### 社会的自律性の活性度と情動 Affect as Intensity of Social-Autonomy

キーワード：

情動, 情報, 倫理, ネオ・サイバネティクス, 基礎情報学

keyword：

affect, information, ethics, neocybernetics, fundamental informatics

東京大学 原 島 大 輔

The University of Tokyo Daisuke HARASHIMA

---

#### 要 約

情動は、生命システムの作動プロセスの活性度にほかならない。これは、観察記述するシステムの視点によって、次の三種類に分類される。すなわち、機械的情動、生命的情動、そしてそれらの両義的な情動（社会的情動）である。機械的情動は、ある種の自動操縦プログラムであり、他律系の行動を誘発する。この自動的行動は、社会システムの道徳的規範とある程度一致したものになる。なぜなら、メディアの機能によって社会システムが制約した拘束としての現実が、可能な行動の選択肢を有限な範囲にあらかじめ限定しており、機械的情動が誘発しうるのはこの範囲内で選択された行動だからである。これは、基礎情報学のHACS（階層的自律コミュニケーション・システム）モデルでいうと、上位システムの視点からみた下位システムの行動として観察記述される。生命的情動は、自律系が固有の意味と価値を自己形成する自己産出の行為である。これは無限の偶然性から有限の可能性を自己限定する。これが、システムに、規範主義的な道徳性ではない、倫理的な責任をもたらすのである。これは下位システムの視点からみた下位システム自身の作動プロセスとして観察記述される。そして、これらの情動の両義性の感情が、社会的生物としてのシステムの自己感覚を実感させる。これはHACSの社会的自律性の活性度を自己観察記述する方法のひとつである。

#### Abstract

Affect is the intensity of operational process of a living system. It is felt differently according to

---

原稿受付：2019年2月2日

掲載決定：2019年4月1日

the perspective of its observing system; mechanical-affect, lived-affect, and ambiguity of both. Mechanical-affect is a kind of an autopilot program which triggers a heteronomous system's behavior. This automatic behavior conforms to some degree with the moralistic standard of its social system, because it is chosen from a set of possible behaviors which is predetermined by the reality as constraint of the social system by means of its media. It is observed as a behavior of the sub-HACS (Hierarchical Autonomous Communication System) from the perspective of the super-HACS. Lived-affect is an autopoietic action in-forming meaning and value unique to an autonomous system. It self-determines a set of possible behaviors from unbounded contingency. This brings forth the ethical responsibility for the system, rather than its normative morality. It is observed as an operational process of the sub-HACS from within itself. The Feeling of ambiguity of the affect makes sense of the self living its social being. It is a way of realizing the intensity of its social-autonomy of the HACS.

## 1 序

ネオ・サイバネティクスには、自律系と他律系というシステムの作動の仕方に着目した分類がある。自律系は作動の規範を自から形成し、自己準拠的に行為する。他律系は作動の規範を他から設計され、それに依拠して行動する。社会的生物は、観察記述の視点によって、この両者の側面をどちらもみせる両義的なシステムである。このような社会的生物としての自律性を、社会的自律性（階層的自律性）という。理想的には、これは規範的な行動を、自から創造的に行為する、そういう上位システムと下位システムとの相即した自己形成である。このとき、社会的生物は倫理的であるといえる（原島2018）。この倫理性を、この社会的自律性の活性度として概念化することはできるだろうか。社会的生物としての生命状態つまり自己形成の活性度である。そして、もしそのような活性度が定義できるとして、それを学究的に観察記述することはできるだろうか。

本稿は、それはある種の情動であり、この情動の実感として自己観察記述できる、そして、これが原初的な倫理感と責任感である、という説を提唱する。これはあらゆる社会的生物の条件としての権力作用でもあり、そこにはメディアを活用した情動の制御という社会情報学的な問題が根本的にかかわっている。本稿では、これを論述するために、まず、意味と価値の自己形成にかかわる情動の神経生物学と哲学の先行研究を参考にしつつ、機械的情動と生命的情動という二つの情動概念を提案する。そして、これらをあらためてネオ・サイバネティクスの理論モデルに位置付けることで、社会的自律性の活性度としての情動概念を提案する。

## 2 機械的情動と生命的情動

### 2.1 神経生物学的情動説

#### 2.1.1 深い実用主義

倫理学の研究でしばしば持ち出される典型的なジレンマのひとつにトロッコ問題がある。一台のトロッコが線路を暴走している。その先には5人の作業員がおり、このままでは避けることもできずに轢かれてしまう。ところで、この線路には歩道橋がかかっている。そこには大きな荷物を背負った1人の図体の大きな作業員がいる。もしこの作業員を歩道橋から突き落とせば、その命と引き換えに、巨体と荷物でトロッコを食い止めることができる。さて、あなたはいまこの歩道橋にいる。ただし、あなた自身が飛び降りたところで弾き飛ばされるだけでトロッコを食い止めることはできない。5人の命を救うには1人の作業員を突き落とすしか方法がない。さあ、あなたはこの作業員を突き落とすべきか。

トロッコ問題は、端的には5人の命をとるか1人の命をとるかという二択問題が、功利主義的には問題にもならないはずなのに、道徳的なジレンマに感じられてしまうというところが本質である。さまざまなバリエーションをつくることができるから、実験心理学の便利な典型問題のひとつとなっている。自動運転車の実用化をはじめとして、人工知能に行動の判断をゆだねることになる場面が現実味をもって議論されるようになっていくが、機械の行動の道徳性をいかにして設計するかを考察するときにも、このトロッコ問題が活用できる（河島・北村・柴内2017）。

情動論として実験心理学的にこの問題にアプローチした先行研究にGreene (2015) がある。Greeneによれば、1人の命をとるか5人の命をとるかの選択では、功利主義的には5人の命をとるのが明らかに正しい判断であるにもかかわらず、それが明らかに間違っているように感じられて、この二択がジレンマになりうるのは、情動的

な道徳的直観によるものだという。このジレンマが腹側内側前頭前野（VMPFC）や扁桃体などの情動に関係するとされている脳部位の働きと関連しているという実験結果が論拠とされる。

Greeneによると、こうした情動には利点がある。それは、道徳観を共有する共同体（社会や文化）の内部での、道徳的に最適な判断と行動を効率的に自動的に起動するということだ。個人的人間関係や小規模集団の共同体にとっての、進化生物学的に有利な行動を、自動選択するのだ。道徳観を共有する共同体内部での利害関係の最適解を効率的に自動選択する、遺伝や文化や個人的経験によって蓄積された知恵である。功利主義者を自認するGreeneであっても、たしかにトロッコ問題のような場面が現実起こったとして、実際に躊躇せず1人を犠牲にする人とは、個人的に仲良くできるかわからないと述べている。

だが、Greeneによれば、これが共同体内ではなく共同体間の利害関係の問題となると、むしろ情動は衝突の原因になってしまう。それぞれの共同体の利害を優先してしまうからだ。この場面では、情動ではなく、理性が有効だというのがGreeneの主張だ。理性は、合理的推論の一般的な能力である。Greeneの実験によれば、これにはDLPFCなどの、情動に相関する脳部位とは別の脳部位が相関しているようだ。理性には二通りの利用方法がある。ひとつには、情動つまり共同体の道徳観を合理的に正当化することに理性を利用できる。ただ、共同体どうしの衝突の場面ではこのように理性を使用しても役に立たない。だが、理性には、情動の自動性を意識的に打ち消して、共同体の利害にとどまらないより一般的な間共同体の利害について熟慮することもできるのである。つまり、異なる道徳観をもった競合する共同体の間でも共有可能な、メタ道徳的な判断と行動の指針として、理性を利用することもできるのだ。まさに現代の文化主義的衝突の場面で必要な道徳性ではないか。こういう場面では、あくまで文化

内で共有された道徳的な判断と行動を実行する自動的な行動プログラムたる情動に任せるのではなく、より柔軟で一般的な合理的推論にもとづいて判断と行動をするべきだ。ただし、合理的推論をするにも、推論の根拠として何らかの道徳観を共有していなければ、議論にならない。そのために現代で可能なかぎり広く共有されうる共通の道徳観は、功利主義であるということ、Greeneは深い実用主義の立場から主張する。

ようするに、競合する個人と個人の問題には情動のオートモードを利用し、競合する共同体と共同体の問題には理性のマニュアルモードを利用せよ、というわけだ。そして、マニュアルモードを利用するにしても、自分の情動的直観を正当化するためではなく（Greeneの見解では、たとえばKantの定言命法もRawlsの正義論も、そうした情動的直観を正当化しようとする作話にすぎない）、この正当化に打ち勝って、共同体と共同体の間の文化主義的衝突を解決するために、合理主義的な功利主義にもとづいて意識的に意思決定するのが実用的だというわけだ。

ただ、厳密に言えば、Greeneも認めるように、功利主義が実用的なのは、異文化ではあってもすでに根本的に功利主義的な道徳観を共有しているより大きな共同体の内部に限定される。たとえば、いかに情動的には容認しがたくとも、合理的な判断であれば社会的には正当とみなす、という道徳観を根底的に共有していない社会では、功利主義は実用的とはいえない。Greene的な功利主義にとっての価値の究極の根拠は、経験（幸福）と公平におかれるが、これが共有できないと功利主義共同体の成員にはなれない。

つまり、基礎付け的に前提された道徳的権威がなければ、競合する共同体の道徳観を理性だけで調停することはできないのである。そのうえ、あらゆる共同体を包含する普遍的な道徳的権威の存在など、合理的には不可知である。だからこそ、Greeneは、功利主義が普遍的な道徳的権威であ

るという考え方は明確に却下する。ただ、あくまで可能なかぎり広く共有されうる間共同体的な道徳観として、功利主義を深い実用主義の立場から支持するのである。ちなみに、功利主義が文化主義を超えて共有されうる共通道徳観だといえる根拠を、Greeneは、それが脳という広く共有された身体のメカニズムであるというところに見いだす。同じ脳の機能でも、情動は共同体単位での利己性を発揮するメカニズムだから、間共同体的な共通道徳観には向かないというわけだ。

こうして功利主義的道徳観を共有した間共同体的道徳的ジレンマなら、統計的なアプローチによる最適な意思決定も期待できるだろう。道徳観によってすでに選択肢の範囲が画定されているから、そこでのジレンマの不確定性はあくまで決定論的ないし確率論的にアプローチできる。ちなみに、情報技術と人工知能はこれが得意だから有効に活用できるだろう。

ところで、このような立場からすると、たとえば多文化的な未知の他人同士が混在する状況での倫理を考えるために、Rawlsの正義論のように原初状態としてまったくの未知（いわゆる無知のベール）、つまり無限の偶然性を想定するのは、非合理的である。Greeneにとってはそのような偶然性は曖昧で非合理的であり、これを推論に持ち出すのは情動的直観の正当化にすぎないという。あくまで決定論的ないし確率論的な不確実性だけを考えるのが実用的だというのだ。

一見するとこのRawls批判に一理あるように見えるかもしれない。だが、ほんとうにそうだろうか。むしろ、この考え方は、Greene自身の深い実用主義の意義を損なってしまうものではないだろうか。根本的に道徳観を共有しない社会間での倫理的ジレンマを実用主義的に考察するためには、むしろ考慮せねばならないのは、全体が画定している不確実性ではなく、全体が画定されていない無限の偶然性ではないだろうか。なぜなら、全体が画定しているというのは、道徳的権威が実

在しているということの、別の言い方にほかならないからだ。

つまり、オートモードとしての情動もマニュアルモードとしての理性も、いずれもその実用性は、あくまで有限な可能性にあらかじめ限定された問題状況に限られる。ほんとうに不可知論的に倫理を考えるなら、このような限定は前提ではなく、むしろそのような限定を無限から自己限定するプロセスこそ倫理性があることが問題にされなければならない。

ところで、機械にはこの自己限定ができない。だからフレーム問題に悩まされるのだ。生物には自己限定ができる。だからフレーム問題に悩まされないのだ。オートモードとしての情動もマニュアルモードとしての理性も、いわば倫理のフレーム問題に悩まされることになるだろう。選択肢ははじめから限定されているトロッコ問題は、自己限定がすでに済んでいる問題だから、オートモードにとってもマニュアルモードにとっても、倫理のフレーム問題が問題にならないで済んでいるのだ。だが、ラディカルな実用主義なら、この無限からの自己限定を問わなければならない。

これは、トロッコ問題の研究に意義がないということではない。むしろトロッコ問題は、特定の社会システムの内部における道徳観についての研究には有用であろう。あくまで、間社会システムや社会システムの進化の場面で、Greeneのように、そして本稿のように、道徳的規範の存在を前提にすることができないとする立場から、無根拠の偶然性の倫理を問題にする場合には、実用的でないということだ。

なお、このような有限の可能性と無限の偶然性について、ネオ・サイバネティクスでは、たとえば西垣（2018a）が、Meillassouxの概念を援用しつつ、潜勢力と潜在性と呼び分けている。情動には、たしかにある視点からみると、蓄積された道徳プログラムの自動的な発動という、潜勢力にかかわる機能がある。だが、それだけでなく、別

の視点からみると、道徳観を共有しない未知の社会（それは異文化かもしれないし、異世代かもしれないし、社会とは呼べない何かかもしれない）の、潜在性にかかわる倫理的な能力もあるということ、これが本稿でこれから論述されることになる情動仮説の核心である。

つまり、すでに限定された有限の選択肢のなかから意味と価値のあるものを自動選択する機械的情動と、無限の偶然性から意味と価値のある選択肢が自己限定されることとしての生命的情動の、二つの姿を情動はもつ。

### 2.1.2 自動操縦と自己形成

人類史という視点から情報社会の諸問題を考察するHarari (2018b) は、Greeneの研究にも言及しつつ、自動運転をはじめとする情報技術と人工知能を例にあげ、人間的な情動や感情に影響されない機械に倫理的な判断と行動をゆだねることについて問題提起をしている。たとえば、トロッコ問題を念頭に置きつつ、Kant的な定言命法アルゴリズムの自動運転車を購入するか、Mill的な功利主義アルゴリズムの自動運転車を購入するか、Rawls的な正義論アルゴリズムの自動運転車を購入するか、消費者が選択するようになったとしたら、どうなるだろうかと問いかける。この哲学自動車のたとえ話は、情報社会における自由と責任の倫理は道徳プログラムの選択ですむような問いではないのだからよく考えろ、という警鐘にほかなるまい。Harari (2018a) は、生物の知能が意思決定アルゴリズムにすぎないわけがあるだろうか、そして知能と意識が分離できるわけがあるだろうか、そこを間違えたデータ主義の情報社会がどんなものになってしまうか、よく考えよと問題提起しているのだ。

情動だけでは解決できない多文化的な倫理問題のために、理性あるいは意識的な意志を活用せよというGreeneの分析と提案は、たしかに一理ある。特定の社会的文化的な共同体の内部での道徳

的行動を効率的に選択することを情動が補助し、異なる共同体の間（ただし功利主義は共有する）での道徳的な行動の選択には合理的推論（熟慮）が役立つという、Greene流の深い実用主義の考え方そのものには同意する。だが、これはあくまで規範が前提された道徳主義的な問題に限定される。つまり、遺伝や文化や学習によってプログラムされた自動的メカニズムとしての、機械的情動の機能の仕方の問題である。トロッコ問題で明らかにできるのはそこまでだ。

先述のように、トロッコ問題は、プログラムされた規範に依拠して判断し行動する他律系の道徳性を考察するには、格好の思考実験である。ただ、生物のように、偶然的な無根拠のなかで自から意味と価値を形成している自律系の倫理性を考察するには、十分でない。なぜなら、これだけでは依拠すべき規範そのものの自己形成としての倫理の問題がわからないからだ。これでは、無根拠の倫理学としての深い実用主義の本質的な意義も、十分には説得的にならないだろう。既成の選択肢からの選択行動をいくら調べても、選択肢そのものの自己形成という責任を問われうる自律的な行為についてはわからないからだ。自己形成のできない機械的な情動と理性にもとづいた道徳プログラムに判断と行動をゆだねてしまうと、無限の偶然性からの自己限定ができなくなるから、とりうる行動の可能性が過去のデータと法則に依拠した有限の選択肢に限定され、潜在的な未知の状況において創造的に対応することができなくなってしまう。道徳プログラムは、根本的に自己形成ができず、それゆえ責任も問われえない。

ここでひとつ、SF作品から一場面を引用して考察してみよう。『新世紀エヴァンゲリオン』のあるエピソードだ。主人公の碇シンジは、人類を敵から守る巨大ロボット兵器エヴァンゲリオンのパイロットである（物語の筋を厳密に言えばこの表現は精確ではないが、ここでは例題として話を単純にするため、これでよい。以下の記述も同様

に簡略化した)。エヴァンゲリオンは複数台あり、その新しい一台のパイロットに、碓シンジのクラスメイトが選ばれた。ところが、この新しい一台は敵のウイルスに感染しており、パイロットであるクラスメイトもろとも敵になってしまった。エヴァンゲリオンは圧倒的な戦力を誇るため、エヴァンゲリオンによってしか対抗できない。放っておけば人類は全滅する。かくして、碓シンジはエヴァンゲリオンのパイロットとして、いまや敵となってしまった友人の搭乗するエヴァンゲリオンを殲滅するという任務を負わされた。友人1人をとるか人類全員をとるか。トロッコ問題だ。

いかに人類の敵になってしまったとはいえ、友人をその手にかけることができず、碓シンジは逡巡してしまう。すると、碓シンジの搭乗するエヴァンゲリオンを保有している軍隊ネルフの司令である碓ゲンドウは、碓シンジのエヴァンゲリオンを強制的に自動操縦モードに切り替えてしまう。碓シンジから主導権を奪ったこの自動操縦システムは、情動的な葛藤などに惑わされることなく、人間的な意識から切り離された純粹に合理的な知能で、1人の命を犠牲にして人類全員の命を救うという、功利主義的に正しい判断を効率的に自動的に下し、すみやかに敵を友人もろとも殲滅する。こうして人類は全滅の危機を乗り切った。

友人の命を奪ったまさにそのロボットの操縦席から一部始終をなすすべなくただ眺めているしかなかった碓シンジは、もちろん殺人の責任を法的に問われることはなかった。エヴァンゲリオンに搭載されていた自動操縦システムが自動的に行動するにまかせるよりほかなかったのだから、責められるべきはこのシステムを造ったネルフの責任者かあるいはネルフという法人だろう。だが、それでも、碓シンジは、罪悪感にとらわれるし、クラスメイトたちとも疎遠になる。

この罪悪感はもちろん友人を救えなかったことの罪悪感だろうし、クラスメイトと疎遠になったのはもちろん小規模な共同体に情動レベルで共有

された道徳的規範を遵守できなかったことへの嫌悪感からくる疎ましきだろうが、その実相は、本稿の見解では、そうではない。それは、道を自分で切り開く自由と責任を負い損ねて、所与の規範に機械的に従属することを強制させられたという、無責任と自動性の原初的な罪悪感と疎外感なのである。

哲学自動車のたとえ話に戻れば、所与の規則を機械的に遵守するだけの無責任と自動性という罪悪感と疎外感から解放されて、責任を問われうる自由へと踏み出した偉大な一歩こそが科学技術なのだ、というNietzsche的な超人主義アルゴリズムを搭載した自動運転車なら、科学技術の進歩が完全自動運転へと後退することなど許可しはしないだろう。機械に行動の判断をゆだねるのは、科学革命以後の自由で創造的な社会的生物たる人間にとっては、どれほどの安全性や利便性を得られたとしても割りに合わないことであり、これを科学技術によって促進することほど滑稽なことはないと、このNietzsche的の道徳プログラムは搭乗者に語りかけてくるかもしれない。

もうひとつ別のSF作品を引用して、さらに考察を展開してみよう。今度は『マトリックス』である。この映画三部作の中盤のクライマックスで、主人公ネオは、マトリックス世界の造物主であるアーキテクトから二者択一の選択を迫られる。ネオにとっては仲間でもあり最愛の人でもあるトリニティの命をとるか、人類全員の命をとるか、という選択である。恋人1人をとるか人類全員をとるか。トロッコ問題だ。

ネオの選択はこうだ。トリニティの命をとる。かくして、人類全員が身を寄せる村ザイオンには、アーキテクトたち人工知能軍団の兵器の大群が押し寄せることになる。この時点では、ネオは情動のままに非功利主義的にトリニティを選んだかのようなのである。人類全員より恋人1人の命を選んだかのようなのである。ところが、なんとネオは、最終的にはなんやかやの命懸けの大博打をうまく切り

抜けて、機械軍団と平和を結び、人類全員をも救いおこせしてしまうのである。

これは与えられた選択肢にはなかった選択肢である。恋人1人の命も救うし、人類全員の命も救う。ネオはこの有限な選択肢にはなかった潜在的な選択肢を実現してしまったのである。ネオの選択とは創造でもあったのだ。もちろん、成功したからこそネオは物語の英雄になれたが、失敗していたかもしれない。だが、その失敗は、ネオの責任だ。選択肢を自己形成したからだ。ネオは、責任を問われうる、責任に応答することができる境地を、生きたのだ。碇シンジにはゆるぎなかった境地である。

### 2.1.3 原初的な自己感と責任感

神経生物学には、情動こそが無数の可能性から意味と価値のあるものを限定しているとする説がある。Damasio (2010)のソマティック・マーカー説である。情動には、無数の可能性のなかからいまこの状況で自分に関係している意味と価値のあるものだけを際立たせる機能があるという説だ。身体的変化である情動が、重要なものをマークするから、ソマティック・マーカーである。これは、情動に相関するとされるVMPFCを損傷した患者が、フレーム問題に悩まされるロボットさながらに、悲惨な意思決定をするのを目のあたりにしたことから着想された。この機能のおかげで、人間はロボットのようにフレーム問題に悩まされないのだ、という説もある(信原 2017a)。これが自己限定の生命的情動なのだろうか。

だが、もし意味と価値の規範があらかじめ実在していることが前提とされてはじめて機能する身体能力なら、ソマティック・マーカーは、先述の機械的な選択の自動プログラムと根本的には変わらない。そして、Damasio説にも信原説にも、その規範の候補がある。それはいずれも、健康な生命の維持、ホメオスタシスである。信原(2017b)はそのような規範の表象主義的な実在論の立場を

表明している。Damasio (2005, 2018)はもう少し微妙な立場にみえる。Damasioのいうホメオスタシスは、いわゆる自己同一性の恒常性維持というよりは、自己生成変化とでもいうべきものだ。あるいは、Damasio自身がSpinozaの情動説を参照しつつ強調するように、よりよく生きようとする意志たるコナトゥスのことなのである。これは、一般的に理解されているホメオスタシス概念とは区別されるべき、Damasio独特の生命システム概念とみなされるべきものである。

これは、ネオ・サイバネティクス的なシステム理論の分類でいえば、第一世代システム理論であるホメオスタシスよりは、第二世代システム理論である自己組織化、あるいは第三世代システム理論であるオートポイエシスに近いといえなくもない(Maturana & Varela 1980, 河本 1995)。ただし、その場合、恒常的な実体として前提された自己が問題なのではなく、恒常性を維持しようとするプロセスが問題になる。オートポイエシスの自己は、ホメオスタシスや自己組織化の自己のように作動プロセスに先立って他から画定されるのではなく、作動プロセスの再帰性の結果として事後的に自から画定されるのである。

所与の規範によってあらかじめ範囲が限定されているのであれば、それは根本的に自己限定ではない。すでに限定された選択肢のなかからの選択であって、無限の偶然性のなかから選択肢そのものを自己限定することではない。もしソマティック・マーカーの機能が限定された選択行動であるならば、それは機械的なプログラムにすぎない。情動とはそのような機械的なプログラムにすぎないのだろうか。

そうではない。オートモードもソマティック・マーカーも、高度に発達した神経系としての脳の機能であるとされているわけだが、Damasio自身も強調するように、そもそも神経系を構成している神経細胞ひとつひとつだって生物なのだから、神経系としての集合的なふるまいとしてはじめて



創発するような脳と相関する機械的情動が、神経細胞も含めた生物の自己限定プロセスそのものであるはずもない。Damasioのような進化論的な神経生物学の見解からすれば、もし情動が脳の機能だとしたら、それは生物進化的には新しい機能の一面にすぎない。だが、Damasio説では、情動とはあらゆる生物の生きるプロセスにともなう生命状態の変化のことである。この変化を感情として知覚できるかどうかという水準でならば、それは人間のような神経系の発達した一部の生物だけだろうが、情動については、脳も神経系さえももたない単細胞生物をも含めて、あらゆる生物がもつのである。

つまり、機械的情動よりも古層の生命的情動があるのだ。それは、本稿の文脈で言えば、有限な範囲からの選択ではない、無限の偶然からの自己限定であり、生物の意味と価値の自己形成である。このプロセスが生命的情動である。したがって、この生命的情動が、機械的情動のように脳との何らかの相関関係を発見できるかどうかについては、理論上、ある意味では肯定しうるしある意味では否定しうる。というのは、脳も神経細胞でできているからには、そのひとつひとつの生きている神経細胞が意味と価値を自己形成しているのだから、そこには生命的情動が発生しているから、脳を探せば生命的情動との何らかの相関関係を見いだすことはできるだろうからだ。だが、神経細胞そのものも含めて、脳も神経系ももたない単細胞生物だって生きているのだから、そういう脳をもたない生物の生命的情動と脳の相関関係はありえない。むしろ、理論上は、文字通りガッツフィーリングに見いだされるのではないだろうか、という仮説も立てられる。

ちなみに、このような進化論的な説明の利点は、設計（超越的な根拠）の存在を前提することなしに、生物を説明するところにある。何らかの権威（たとえば優れた生物の定義）の存在を根拠にした説明とは根本的に一線を画す。生物を環境に適

応する機械のように外から眺める客観主義的な視点や、ホメオスタシスという説明原理などは、この利点に水を差すことになりかねないから、注意が必要である。遺伝も、生物個体にとっては継承されたものでも、生命にとっては進化の歴史で自己形成した意味と価値の記憶にすぎない。生物個体が、自己形成した意味と価値を、構造としての神経系に記憶するのと、同じことにすぎない。だから、意味と価値は、(遺伝子や神経系、あるいは人間社会で利用されるもろもろの記録媒体によって) 記憶として継承されつつも、本質的には生物の生存によって事後的に定着される偶然的なものである。継承されてきた意味と価値にもとづいてつくられた概念構造や法則が根本的に覆されてしまうようなクリティカルな場面においてもなお生き延びてきたこの生命の知能の根本に、偶然的な意味と価値の自己形成があるということだ。

Damasio (2013) がSpinoza情動論と神経生物学の見解に依拠して主張するように、ある情動はより強力な別の情動をぶつけることで対抗できるのであり、認知的無意識は強い意志の意識的な熟慮と訓練によって変えることができるのである。この自己形成の生命的情動の実感とともに実現されるのが、倫理的な責任の問われうる自己である。生命的情動は、権威の存在によって基礎付けられた道徳的な情動ではなく、無根拠な自由の倫理的な情動である。この情動から、意識ある意志の感情が生じるのであり、これが行為者としての自己を限定する（ソマティックにマークする）のである。Damasioもいうように、この実感なしには責任は問われえない。

#### 2.1.4 無根拠の倫理的な知識行為

基礎付け主義的な道徳主義ではない、無根拠の倫理については、ネオ・サイバネティクスにも先行研究がある。Varela (1999) のエチカル・ノウハウである。そもそも、客観主義的な存在も主観主義的な主体も前提することなく、無根拠な偶然

性にいかにしてシステムと環境が創発するかと問いを立てよというのが、Varela et al. (1991)の提案したエンボディメントやエナクティブ・アプローチの核心である。ようするに、意味と価値は、客観主義的に実在しているわけではないし、主観主義的に恣意的なわけでもない。そのつどの状況における行為とともに自己形成されるのである。そのような倫理的な状況への応答を体現する知恵は、意識的な合理的推論や文化的な道徳規範に基礎付けられた判断にもとづく行動ではないが、たんに無意識の機械的で自動的な行動にゆだねることでもなく、意志によって鍛錬することで上達していくものだということが、エチカル・ノウハウの議論である。この鍛錬しうる知恵の体現に、Varelaは、現象学や東洋思想と神経生物学を統合しながら学究的にアプローチしようとしていた。

これは、先述のDamasio的な意志の優位にほかならない。認知的無意識を、自己形成した意味と価値によって意識的に変化させてゆくことができるということである。無意識をエチカル・ノウハウとして知恵に鍛錬してゆくことができるのである。Damasioの情動論は、Varelaの研究を引き継いでいるというわけではないが、それでも事実上は、エチカル・ノウハウの神経生物学的な探究のバトンを受け継いだ現状では最も秀でた研究のひとつとみなすべきものだろう。だが、意味と価値の自己形成が概念化できないと、意味と価値の根拠をどこかで人為的に設定しなくてはならなくなり、権威主義的な独断か相対主義的な無限後退を余儀なくされる。それがDamasio説ではホメオスタシス（恒常性維持）という究極原理である。これを基礎付的に設定してしまえば、生物とは恒常性を維持すべく設計された機械だということになってしまい、これでは自己形成は問われえない。

オートポイエーシスという概念の意義を、自己なるものを前提することなく、そして自己なるものの設計を前提することなく、自己なるものを産出する自己なるものを産出する自己なるものを

産出する……という再帰的で内在的な自発性の創造プロセスをとらえたものであるところに見いだすならば、むしろそこには自己なるものの偶然的で自由な創発と生成変化をこそとらえるために実用的な発想の展開が望めることも理解されるだろう。そのラディカルさは、無根拠の無限の偶然からの自己限定を問うたところにある。

Varela (1979) は、この自己形成のプロセスをこそ情報 (in-formation, すなわち、外から形を与えられるのではなく、内的に形成すること) と呼び、自律系として概念化した。このようにして生きられた生物にとっての意味と価値が、その生物が生きられたという事実によって事後的に自己形成され、それに自己準拠的に自己産出が再帰的に作動するのである。こうした概念装置は、生命的情動すなわち偶然からの自己形成のプロセスをとらえるものだ。

## 2.2 情動の制御

意味と価値のプリセットとしての機械的情動は、あらゆる生物にとって本質的に制約としての権力作用だが、意味と価値の自己形成としての生命的情動は、そのプリセットが無根拠で偶然的であることの表現である。このことは、Massumi (2015) や伊藤 (2013, 2017) に代表されるような、人文学・社会科学における情動と権力についての研究の意義を支持する。それらは、行動の自動プログラムと、意味と価値の自己形成という、情動の両義性の倫理的な問題を、情報社会の権力作用を主題にして考察している。とくに大衆メディアや情報技術と人工知能が発達した現代の情報社会はもちろん、そもそも情動の制御とは時代や地域に限定されるものではなく、社会的生物の生きるところどこでも働く権力作用でもあり生命力でもある力そのものの別名なのだということを喝破したところに、とりわけ本稿の議論とも通ずる重要な意義がある。

また、これらの研究では、情動の権力作用とい

うものが、機械論的で物質主義的な形式論理で記述しようとする、逆説的で不合理なものになってしまうことも指摘されている。それは、ネオ・サイバネティクスの見地からすれば、情動が、機械論的世界観だけではとらえられない、生命論的世界観にかかわるものだからである。ようするに、情動の権力作用の非形式論理的で非合理主義的な理不尽さは、機械論的な合理性によって矯正すべきものではなくて、合理主義的な機械論的一元論によってはとらえがたい社会的生物の生命力の作動そのものなのであって、むしろこの権力作用を理解し訓練し活用することこそ倫理的な意味と価値を見いだしたところに、Massumiや伊藤の情動論の意義がある。

権力作用としての情動の制御は、意味と価値の形成の範囲を限定することで作用する。つまり、行動を指定するのではなく、あくまで有限のポテンシャルを範囲画定するだけ、つまり可能な選択肢の全体を提供するだけなのである。情動の制御というかたちでの権力作用は、厳密にどういう行動をするべきかを指令するのではない。言い換えれば、社会情動的な言語や画像などの記号を利用して、意識的にせよ無意識的にせよ、行動の一部始終を指令するわけではない（たとえば、サブミナル的に特定の商品の消費行動をうながすとは違う）。機械的情動の自動行動プログラムにとっての選択肢の範囲や選好に働きかけたり、機械的情動の自動行動プログラムの発動を誘発しうるような攪乱の刺激を与えたりすることで、限定された可能性の範囲内で行動を統計的に制御するのだ。せいぜい選択肢のなかからどれを選択するかの確率を変えたり、自動的な行動プログラムを誘発するかもしれない攪乱の引き金をひいたりするだけなのである。だから、結果的にひとりひとりがどのような行動をするかは統計的な多様性をみせる。ただ、同じ情動の刺激を受ければ、同じ文化的な共同体であればおおむね同じ行動プログラムが誘発されるだろうし（先述のGreeneの実験

心理学的な情動説を想起されたい）、文化が異なっても人間であれば多かれ少なかれ遺伝的なプリセットによっておおむね同じ行動プログラムが誘発されるだろうというわけだ（先述のDamasioの進化生物学的な情動説を想起されたい）。

ようするに、情動のもつ自動選択の機能（その生物の生存にとってその状況において意味と価値のあるものを無意識の身体的な水準で自動選択する機能）に働きかけることで、ある程度ではあるが意図的に行動を制御しようという発想である。この方法で、たとえば大衆の行動をある程度は統計的に制御することもできるだろうし、あるいは自分の注意の集中力をやるべきことに限定させることで身体能力や知的能力をある程度は上げることができよう。Massumiの表現を借りれば、情動の制御とは、ようするに、コンディションを整えることなのである。

別の言い方をすれば、制御がおよぶのは情動の機械的な半面にすぎず、情動の生命的なもう半面には制御はおよばないのであり、そこに社会的制御の限界があるとともに社会的生物たる人間の責任もあるのだ。情動の制御は、意図的に特定の機械的情動のプログラムを誘発させるために、社会的・身体的な制約を設定することで、行動の可能性を有限の範囲内に限定しようとする。だが、生物はそもそも無限の偶然性から意味と価値を自己限定するのである。

この自己形成のプロセスが生命的情動であり、そこにもなう感情が自己意識の原初であり、そしてそこから生じる意志によって、機械的情動の発動を打ち消すこともプリセットを書き換えることもできる。自己形成できたということが事後的に感覚されることで、意味と価値の原初的な情報（原情報）を形成し、これが実感されることで感情となり意識的な自己の萌芽を生み、これが意識的に観察記述されることで主観性ないし自己意識が結果的に形成されるのである。有限状態からの選択ではなく無限の偶然性からの自己限定とい

う、倫理的な生の実現であり、生命的情動の感情とはその情趣なのである。

ちなみに、本稿では踏み込まないが、伊藤(2017)が情動論の先駆としてあげるように、Jamesの純粹経験概念やWhiteheadの抱握概念も、いかにして何かが自から際立ち出すのか、つまり自己限定されるのか、ということについて考えた情動論の重要な先行研究である。そして、こうした発想の源流に、ヨーロッパ精神史があることも見逃せない(坂部1997)。こうした精神史の理解は、情報技術と人工知能の理解を、より深めるだろう(西垣1995, 西垣2010, 西垣2018a)。

### 3 ネオ・サイバネティクスの情動説：社会的自律性の活性度と情動

ここまで論述してきた情動を、ここからはネオ・サイバネティクスの理論モデルで概念化して整理しよう。

ネオ・サイバネティクスは生物と機械を自律系と他律系として識別する。

他律系は、他から設計された規範に依拠して作動する。あらかじめ有限に限定された行動である。あらかじめ限定されているとはいえ、この所与の全体の範囲内で、たとえば機械学習や適応のように、プログラムされた単純な規則からより複雑な行動や新規の行動を創発することならありうる。Clark & Hansen (2014)は、これを科学技術的な創発と呼ぶ。機械的情動とは、このような他律系としての作動のことにほかならない。生物についても、とくに物質主義的な自然主義の視座から科学技術的に観察記述された場合には、あたかもこのような他律系であるかのようにみなされている。このことは、西垣(2008)によるHACS(階層的自律コミュニケーション・システム)モデルでいえば、下位システムの視点から自己観察記述されればあくまで後述の自律系とみなされるが、上位システムの視点から観察記述されたときには

あたかも他律系であるかのような行動がみられるという、社会的生物のシステムとしての両義的な階層性によるものだ。機械的情動が生物にもみられるのは、あくまで観察記述の視点によるものすぎない。また、それゆえ、機械的情動が、確率的アルゴリズムの自動操縦プログラムとして、情報技術や人工知能でも可能なメカニズムにモデル化されるのは、これがそもそも他律系の作動だからにほかならないのであり、そこでは、次に述べるような生物の自律系としての側面はとらえられていないということに注意されたい。

自律系は、自から産出された規範に依拠して作動する。無根拠の無限の偶然性から有限の可能性を自己限定して、生きられた意味と価値を自己形成することで、自己準拠的に再帰的に作動するのである。このようにして自己形成された意味と価値が情報(information)である。つまり、手に負えない無限の偶然性から、生き抜くことのできる有限の可能性を、自己限定するのである。Clark & Hansen (2014)は、これをネオ・サイバネティクスの創発と呼ぶ。ネオ・サイバネティクスの創発は、単純から複雑が追加されるのではなく、複雑から単純が限定されるのだともいえるが、これは、たんに科学技術的な創発と逆方向の進行と理解されるべきではない。ネオ・サイバネティクスの創発は、科学技術的な創発が前提とする有限の一元論ではなく、それぞれの生の自己形成だけそれ固有の意味と価値が自己限定される無限の多元論である(西垣2010, 西垣2018b)。ネオ・サイバネティクスの先行研究は、この自己限定としての知識行為を問題にしてきたのである(たとえば、Foerster (2003)やLuhmann (2013)など)。そもそもWiener (1948)の古典サイバネティクスも、その本質的な意義は、この自己限定という生きられた生物の視点からの知識行為を問うたところにある。そして、情動論の先行研究にも、このような自己限定としての情動を問題にしてきたものがあることは、先述の通りである。生命的

情動とは、こうした自律系としての作動のことにほかならない。

人間の心的システムのように、発達した神経系を構造とする自律系は、この生命的情動を実感し、さらに言語や記号という社会情報に転化することができる。ネオ・サイバネティクスでいう自己観察記述である(西垣2004, 西垣2008)。心的システムは、自身の身体的な状態変化(生命情報)を、自己観察記述できるのである。生命状態の自己観察記述である。これが感情であり、原初的な自己意識である。そして、人間の心的システムは、これを社会的に通用する言語や記号(社会情報)に転化することができる。生命情報は、この心的システムによって観察記述されることで、はじめて人間社会で通用する社会情報となる。したがって、学究的なシステム研究としての情動論は、心的システムの自己観察記述にかかっている。

ここで重要なのは、観察記述の視点である。社会的生物は、観察記述の視点によって、自律系にも他律系にもみえる。この両義性が社会的生物の本質であり、情動の両義性でもある。社会的生物の自律と他律の両義性は、ネオ・サイバネティクスでは、西垣(2008)の基礎情報学がHACSモデルによって概念化している。すなわち、社会的生物は、上位システムと下位システムの階層をなして連動している。下位システムは、そのシステムのプロセスそのものの視点から観察記述すれば、自律系にみえる。この下位システムが作動に随伴して生産したコミュニケーション素材(言表や画像など)が連鎖して、コミュニケーションがコミュニケーションを産出する自己産出的なひとつの自律系を形成しているとき、これが上位システムと呼ばれる。したがって、上位システムの構成素は、下位システムではなく、下位システムが産出した素材からなるコミュニケーションである。このとき、上位システムの視点に立って下位システムを外から観察記述すれば、下位システムは他律系にみえる。上位システムの視点からは、下位システ

ムは、あたかも、上位システムのコミュニケーションの連鎖を破綻させない範囲内で、上位システムの素材となるコミュニケーションを出力している、情報処理機械のようにみえる。ようするに、所与の規範に依拠して、有限の範囲内で、学習や適応をしつつ、情報処理をしているようにみえる。このように、観察記述の視点によって自律系にも他律系にもみえるのが、社会的生物の社会的自律性(階層的自律性)である。そして、このように社会的自律性は上位システムからの制約をうけているから、上位システムと下位システムには非対称性がある。ようするに権力作用である。

本稿で提示した情動の分類は、こうした社会的生物のシステムとしての作動の分類に対応している。機械的情動は、他律系としての作動である。つまり情報処理のプロセスとしての状態変化である。これは、物質的な変化を外から観察記述すればわかるものだ。生命的情動は、自律系としての作動である。つまり自己形成のプロセスとしての状態変化である。これは、情報的な変化を内から観察記述しないとわからないことである。そしてこれらの情動の両義性は、HACSの自律と他律の両義性である。つまりHACSのプロセスとしての状態変化であり、いわば社会的情動である。

社会システムは心的システムにとっての制約であり、それゆえにこの制約を共有する同じ社会内で可能な行動は、社会のメンバーにとっては同じ仕方で限定されている。だから、個々のメンバーが実際にどのような行動をするかは統計的な多様性をみせるが、その多様性の範囲は有限に限定されている。つまり、とくにそれと意識することなしに、社会の規範にある程度は準拠した行動を、自動的に実行することになる。これがオートモードとしての機械的情動の、ネオ・サイバネティクスのメカニズムである。この制約は、上位システムの規範に則って下位システムが作動していれば、下位システムには意識されない。だが、下位システムの作動が上位システムの規範に抵触する

なら、そのとき上位システムの制約としての現実性が下位システムの意識にのぼる。下位システムにとっては消極的な仕方での階層性の観察記述の仕方である。だが、機械的情動の実感を訓練して活用することで、上位システムの制約をより積極的な仕方でも観察記述し、自己形成に活用することもできるだろう。自己形成にあらかじめ範囲限定されるから自由への制約にもなるが、範囲を限定して集中することで効率的で生産的になる利点もある。機械的情動の制御である。したがって、機械的情動の制御の内実は、生命的情動の自己形成のプロセスを制御することではない。機械的情動の制御は、せいぜいシステムの作動を攪乱し、すでに限定されて選択肢として用意された行動の候補のなかでの選好に影響するだけにすぎない。これは、ネオ・サイバネティクスでは、メディアの機能として概念化されているものだ。メディアは上位システムのコミュニケーション素材になりうるものの条件を整序する。これは、下位システムにとっては、社会的生物としての行動の可能性が有限に限定されるということだ。

情動の両義性は、社会的生物の倫理性的の問題である。機械的情動は、有限からの選択という規範に依拠した道徳主義的な行動である。生命的情動は、無限からの自己限定という無根拠で自由の倫理的な責任を問われうる行為である。生命的情動は、生きられた意味と価値の形成の感覚であり、意味と価値が前提されている他律系の自動適応ではなく、自律系の意味と価値の自己形成である。生きる意味と価値をみいだした感覚（機械的情動）、生きられた意味と価値をつくりだした感覚（生命的情動）、これら二つを重ねることで、社会的生物としての自己形成のありさまが浮き出てくるのである。

この社会的自律性の自己形成の状態すなわち活性度を、その自己形成のさなかで自己観察記述することはできるだろうか。冒頭でも掲げたこの問いに、本稿は、かくしてひとつの仮説を与える。

社会的自律性の活性度は、機械的情動と生命的情動が相即する両義的な情動（社会的情動）として自己観察記述される。これは、社会的生物にはある種の情動の感情として実感されている。社会的生物として生きていることの実感、意味と価値の自己形成の実感である。人間のような意識と主観性と言語的思考をもつ社会的生物は、この実感を、自己観察記述として表現することができるだろう。そしておそらく、人間だけでなく哺乳類をはじめとする多くの社会的生物も、これを厳密に自己観察記述と呼べるかどうかはむずかしいところだが、実感はしているだろう（Damasio 2018）。この社会的情動を実感する自己は、認知的無意識（機械的な自動性）とも個人的意識（自己中心主義的な自我）とも違う、上位システムと下位システムの相即的な自己形成の自己であろう。自律が他律で他律が自律であるような、両義性の境地の自己である。そして、その情趣が社会的自律性の活性度の実感である。

## 結

社会的自律性の活性度は、物質的に観察記述される機械的情動と、情報的に観察記述される生命的情動の、相即的な自己形成プロセスの自己観察記述によって感じわけられる強度である。これは、物質的に観察記述される度ではなく、情報的に観察記述される度である。物質主義的・客観主義的に計測できるような量ではないが、社会的生物には実感として自から感じわけられている、情動の強度である。別言すれば、情動には、機械論的なアプローチだけではとらえがたいところがあるのだが、これにはネオ・サイバネティクスの情報学の生命論的なアプローチがありうるということである。その端緒のひとつとして、本稿は、社会的自律性の活性度と情動を、ネオ・サイバネティクスの理論モデルをもちいて概念化した。

こうした情動論のネオ・サイバネティクスと社

会情報学は相互に貢献しうるだろう。まず、社会的生物の主観性の発生について、社会的自律性の自己形成プロセスとして情動論的に研究するアプローチがありうる。また、HACS的な権力作用としての機械的情動は、ネオ・サイバネティクスのメディアの機能によって整序されるものとして概念化されたが、この概念装置による情動とメディアの社会情報学的研究も期待される。さらに、生命的情動は原情報の自己形成プロセスであり社会的な意味と価値の源泉として概念化されたが、情報技術と人工知能によって最適化が試みられるいかなる社会的な問題も、問題の設定はこの生命情動的な意味と価値に依拠しているものであり、生命的情動の理解にもとづいた情報社会の問題設定は、社会的自律性の活性度を強くするだろう。

そのためには、情動の自己観察記述をさらに探究しなければならない。機械的情動であれば物質主義的に外から観察記述することができるが、生命的情動は心的システムが情動的に自己観察記述することで社会情報に転化しないことには、学究的に（言語的記号的に）社会的に共有されうる知恵にはなりえない。情動の鍛錬には、心や意志、そして感情と自己意識についての、情報学的なさらなる研究が必要である。

また、機械的情動は、高度に惰性的な社会システムにおける道徳的行動のメカニズムとしては有用でも、異なる文化の社会システム間や、社会システムそのものの進化にともなう道徳観の変化の場面では、むしろ衝突の要因になる。この場面における人間の理性（合理的推論、熟慮）の有用性が、先行研究では指摘されていたのだった。これも一理あるが、人間の理性の合理性が正しいという超越的な根拠がないと、論点先取の合理主義になってしまい、文化主義的衝突の危機における普遍的な道徳原理にはなりがたい。むしろ、本稿では、そのような根拠の実在については不可知論的な懐疑主義の立場をとり、無根拠な偶然性を前提にして、そこからいかにして有限な可能性を自己

限定するかが問われることに、情動論的な倫理のひとつの道筋を見いだした。

下位システムの自己形成は、上位システムの自己形成と比較すれば、タイムスケールもパワーも小さい。それは、いわば理念的に想定される自律性の内在的な極小の極限である。上位システムが安定していれば、それは滅多に実感されない。規範に依拠したままで生きられるので、意味と価値の自己形成が不要だからだ。自律系である人間が一般的に、社会的な生活において四六時中いちいち情動の趣深さを感じながら生きてなどいないのは、そのためである。だが、実感されないからといって、無いわけではない。それは、潜在的な原情報なのである。この下位システムの小さな自己形成によって上位システムが抜本的に変化することはまずなくとも、上位システムの進化の原動力は、この小さな自己形成である。社会の大変動や、異なる規範を遵守する社会との衝突など、上位システムの危機においては、もはやこれまで依拠してきた規範が通用せず、システムとしての潜在的な意味と価値の自己形成によって生成変化しなければ生きられなくなるから、下位システムの自己形成は劇的に実感されるだろう。とはいえ、それでなくても、そもそもシステムは無限の偶然性から自己限定しているからこそ生きられているのであり、むしろ潜在的にはいつでもどこでも好機しかない。

意味と価値の自己形成は、あらかじめ意味と価値があるわけではなく、自己形成が生きられた結果として、事後的に意味と価値が自から形成されているのだ。だから、いかに巧妙に規範を設計してプリセットした機械的情動の道徳プログラムも、倫理的にはなりえない。情動の倫理は、機械のように規範を遵守して行動することではない。既存の道徳的な規範を内面化すれば、機械的情動にしたがって無意識に自動的に行動しても道徳的な行動ができるようになるだろうが、これは倫理的なわけではない。そのつどの意味と価値を自

己形成する社会的生物としての自己を、そのつど体現することが、倫理的なのである。社会性と自律性が、それぞれ別々に、だが同じこととして、自己形成される。この自己形成の活性度が、社会的な情動として実感され、そして社会的生物としての自己が実現される。いわば、情動の倫理は、情動的に行動することで正しい行動だけをするようになるということではなく、むしろ、情動的に行為することで間違った行為をすることがありえなくなるということなのである。だからこそ、その社会的生物の生きざまは、自由であり、責任を問われうる。そして、そこに生きられた意味と価値の独特の情趣がともなうのである。

### 参考文献

- Clarke, B & Hansen, M. B. N. (2014) 「ネオ・サイバネティックな創発」大井奈美 (訳), 西垣通・河島茂生・西川アサキ・大井奈美 (編) 『基礎情報学のヴァイアビリティ』東京大学出版会, pp.173-204.
- Damasio, A. (2005) 『感じる脳』田中三彦 (訳), ダイアモンド社, 416p.
- (2010) 『デカルトの誤り』田中三彦 (訳), 筑摩書房, 438p.
- (2013) 『自己が心にやってくる』山形浩生 (訳), 早川書房, 433p.
- (2018) 『進化の意外な順序』高橋洋 (訳), 白楊社, 349p.
- Foerster, H. von. (2003) *Understanding Understanding*, Springer, 362p.
- Greene, J. (2015) 『モラル・トライブズ』竹田円 (訳), 岩波書店, 490p.
- Harari, Y. N. (2018a) 『ホモ・デウス』柴田裕之 (訳), 河出書房新社, 560p.
- (2018b) *21 Lessons for the 21st Century*, Jonathan Cape, 368p.
- 原島大輔 (2018) 「階層的自律性の観察記述をめぐるメディア・アプローチ」, 西垣通 (編) 『基礎情報学のフロンティア』東京大学出版会, pp.137-157.
- 伊藤守 (2013) 『情動の権力』せりか書房, 271p.
- (2017) 『情動の社会学』青土社, 270p.
- 河本英夫 (1995) 『オートポイエーシス』青土社, 340p.
- 河島茂生・北村智・柴内康文 (2017) 「自動運転車の「トロッコ問題」などに関する意識」 <[https://researchmap.jp/?action=cv\\_download\\_main&upload\\_id=143157](https://researchmap.jp/?action=cv_download_main&upload_id=143157)> Accessed 2019, January 31.
- Luhmann, N. (2013) *Introduction to Systems Theory*, Gilgen, P. (trans.), Polity, 284p.
- Massumi, B. (2015) *Ontopower*, Duke University Press, 306p.
- Maturana, U. & Varela, F. J. (1980) *Autopoiesis and Cognition*, Riedel, 171p.
- 西垣通 (1995) 『思考機械』筑摩書房, 254p.
- (2004) 『基礎情報学』NTT出版, 235p.
- (2008) 『続 基礎情報学』NTT出版, 219p.
- (2010) 「ネオ・サイバネティクスの源流」『思想』1035, pp.40-55.
- (2018a) 『AI原論』講談社, 200p.
- (2018b) 「観察, 創発, 意識, そして人工知能」, 西垣通 (編) 『基礎情報学のフロンティア』東京大学出版会, pp.1-16.
- 信原幸弘 (2017a) 「フレーム問題と情動」信原幸弘 (編) 『心の哲学』, 新曜社, pp.216-219.
- (2017b) 『情動の哲学入門』勁草書房, 263p.
- 坂部恵 (1997) 『ヨーロッパ精神史入門』岩波書店, 204p.
- Varela, F. J. (1979) *Principles of Biological Autonomy*, Elsevier North Holland, 701p.
- Varela, F. J. et al. (1991) *The Embodied Mind*,



- The MIT Press, 308p.
- Varela, F. J. (1999) *Ethical Know-How*,  
Stanford University Press, 85p.
- Wiener, N. (1948) *Cybernetics*, The MIT  
Press, 212p.