

意識の統合情報理論

大泉 匡史

はじめに

医療の現場においては、外界とのコミュニケーションが取れない患者の意識状態を判定しなければならないという難しい状況に遭遇する。脳に損傷を負った患者、もしくは全身麻酔下にある患者の意識状態を判定しなければならないといった状況である。このような場合に一体どういった基準で判定すればよいのだろうか。例えば、植物状態の判定には患者が意識的な行動を示すかどうかということで判定するが、振る舞いを基にした判定はしばしば誤診断をもたらすことが知られている¹⁻³⁾。従って、患者の脳活動を測定することで、意識を判定することが重要な課題として求められている。

医療から離れて他にも意識の判定が難しい例をあげると、赤ん坊に意識があるか、あるとすれば一体いつから、どれくらいあるのかという問題もある。そして、より難しい問題として人間以外の他の動物(犬、からす、蛙、ハエ等)には意識があるのかということも、古くから人々の興味を引いている重大な問題である。さらにいえば、この先 iPhone の Siri にみられるような優れた人工知能が発達し、あたかも人間と会話ができるロボットができたと仮定して、そのロボットに意識はあるといえるだろうか。ないとすれば、なぜないといえるのか。このような問題に答える際も、それぞれの個体の振る舞いによって意識を判断するという方法ではなく、脳活動(人工知能であれば頭脳に相当する部位の活動)を観測することによって意識を判定することが求められる。

上記したような様々な意識にまつわる問題を統一的に理解するためには、個々の現象の実験的観測だけではなく理論が必要となる。本稿では、意識の理論として近年注目を集めている意識の統合情報理論(integrated information theory of consciousness; IIT)を解説する。IITはウィスコンシン大学の精神科医、神経科学者である Giulio Tononi によって提唱された理論である。

IITは思考実験からいくつかの公理および公準を仮定する。そして、その公準からあるシステム(例えば脳)が持つ

おおいずみ まさふみ 理化学研究所脳科学総合研究センター

意識の「質」そして「量」を記述する数学的な方法を提案する。ここでいう意識の「質」とは、われわれが赤いものを見た時に感じる赤の赤らしさ、ヴァイオリンの音色など、意識の中身(content of consciousness)のことである。そして、量とは意識レベル(level of consciousness)といわれるもので、睡眠時の脳は意識レベルが覚醒時に比べて低いというような使い方をする。

本稿では IIT を全く知らない方、または論文を読んだことはあるが要点がピンとこなかった方を対象として、IIT の要点のみをできる限り丁寧に解説することを試みる。IIT の詳細に関する正確な理解に関しては文献 4~9 を参考にさせていただければ幸いである。本稿では多少の正確性を犠牲にして、筆者自身の見解も加えながら要点を解説する。前もって注意しておきたいことは、IIT は現在もなお発展中の理論であるということである。理論としてはまだ赤ん坊の段階といつてよい。従って、理論の中でも不完全な点、議論が必要な点が存在する。この理論のキーポイントを理解した上で、さらなる改善、発展をはかるにはどうしたらよいかを考えていただくことが読者に求められる態度と考える。筆者は本稿がその一助となることを期待してこれを執筆する。

思考実験

はじめに述べたように、IIT は思考実験、そして現象論的な考察から生まれた理論である。既に得られている実験事実に整合するように作られた理論ではないということに注意されたい。ある特定の現象に整合するように作られた理論は、いわばオーバーフィットした理論といえ、その特定の現象を説明することはあるだろうが、他の別な現象をも説明するような一般性がないことが多い。歴史上の優れた理論(例えばアインシュタインの相対性理論など)は思考実験により作られたことが知られており、科学に飛躍をもたらす際に思考実験はしばしば有効な道具となる。以下では IIT において二つの鍵となる思考実験を説明する。

1. フォトダイオードの思考実験

今、あなたがまっさらなスクリーンの前に立っているも

のとしよう。このスクリーンは明るくなったり、暗くなったりを繰り返すとする。この時、スクリーンが明るくなったら「明るい」と言い、暗くなったら「暗い」と言うようにあなたが指示されたとする、当然ながらあなたはそれを容易に実行できることであろう。一方、同じスクリーンに一つのフォトダイオードが向き合っているとする。このフォトダイオードはスクリーンが明るくなると光を感知して電流が流れ、「ピー」と音を鳴らすものとする。暗くなると何も起こらない。この時フォトダイオードは、あなたと同じように「明るい」と「暗い」の二状態を弁別する能力を有しているように見える。しかしながら、このフォトダイオードがあなたと同じように「明るい」と「暗い」を主観的な体験として“意識”しているだろうか。多くの人がこの質問に関してはいいえと答えることであろう。それでは、この単純なフォトダイオードとわれわれ人間との決定的な差は一体何か。

IITによれば、その決定的な差とは生み出される“情報”の差ということになる。フォトダイオードは「明るい」と「暗い」の二つの状態しか取り得ない。もし仮にスクリーンが青く光ったとしても赤く光ったとしても、フォトダイオードにとっては同じ「明るい」状態であり、「ピー」と言うだけである。一方、人間は青い光を見れば「青い」という別の体験をするし、赤い光を見れば「赤い」というまた別の体験がある。もちろんわれわれは単なる赤や青だけではなく、非常にバラエティに富んだ視覚的体験を持つことができるのは承知の通りである。われわれの、ある時刻における視覚的体験(例えば、今この瞬間にあなたが見ているもの)は非常に多くの別な可能性のうちの一つであって、たった二状態しか取り得ないフォトダイオードとは生み出されている情報の量が大きく異なる(ここでいう「情報」という言葉が何を意味するかに関しては後に詳しく述べる)。

このフォトダイオードの思考実験から、意識がないと思われるフォトダイオードと意識があるわれわれ人間との決定的な差が、生み出されている情報の差にあるということが考察できる。

2. カメラの思考実験

前節のフォトダイオードの思考実験によると、生み出される情報の大きさが意識にとって決定的に重要であるということが述べられた。そうだとすると、フォトダイオードよりはるかに多くの情報を処理しているように“見える”デジタルカメラに意識はあるのだろうかという疑問が浮かぶ。話を単純化するために、デジタルカメラが100万個のフォトダイオードの集合体だとしよう。このデジタルカメ

ラがある一つの状態になった時は、2の100万乗の状態のうちの一つであるため情報量は膨大である。しかしながら、このカメラがわれわれのように意識して景色を見ていると考える人はほとんどいないだろう。それではこのカメラとわれわれ人間との決定的な差は一体何であろうか。

それは、カメラの中ではそれぞれのフォトダイオードによって独立に処理された情報が全く統合されていないのに対して、われわれ人間の脳内では神経細胞が密につながり情報が強く統合されているという点にある。カメラの中の一つのフォトダイオードが生み出した情報は別のフォトダイオードと共有されることはなく、その中にのみ留まっている。つまり、カメラというシステムの中には非常に小さな情報を生み出す素子がばらばらに存在しているだけであり、それを統合する存在がいらない以上、大きな情報を生み出しているとはいえないのである。

このカメラが映し出した写真を外からの観測者(例えばあなた)が見れば、そこから多くの情報を得ることができよう。しかしながら、その情報とはあなたの脳の中で100万個のフォトダイオードが生み出した情報が統合された結果、あなたの脳が作り出した情報であって、カメラというシステムの内的な観点から存在する情報ではない。このことは、次節で詳しく述べる内的な情報(intrinsic information)と外的な情報(extrinsic information)の区別とも関連した重要な話で、IITの肝となる部分である。

以上をまとめると、カメラには情報が統合されていない以上、内的に存在する情報は極めて微小であり、個々の素子によって生み出された情報が強く統合されている脳のようなシステムとは決定的な違いがあるということになる。

IITの公理と公準、そして数学的な定量化

IITは上記した思考実験から端を発し、現象論的にいくつかの公理(axioms)を置く。そして、公理に付随していくつかの公準(postulates)を提案する^{7,8)}。IITにおける公理と公準の使い分けの意味は、公理はいわば誰もが認めるような自明の理のこと、そして公準は公理より自明性はないがIITが前提、要請するものという意味で使っている。

1. 情報(information)

IITにおける第一の公理は“意識には情報がある(Consciousness is informative)”である。ある一つの意識体験は、他に可能であった様々な別の体験と異なるという意味で情報がある。この公理からさらに踏み込んで、IITは次の公準を提案する。“あるシステムが意識を持ち得るのは、そのシステムが情報を生み出した時だけである”。

システムの中にあるものが、神経細胞であろうが半導体素子であろうが、意識に関係するのはそこで生み出される「情報」だけであり、システムが何でできているかなどの物理的性質とは無関係であるというのが、この公準の意図することである。

注意しなければならないのは、ここでいう「情報」とは一体何かという問題である。IITにおける情報とは、“観測者によらない”システムそのものにとっての内的な情報(intrinsic information)のことを指している。内的な情報という概念はIITにおける最も重要な概念であり、ある特定の観測者にとっての外的な情報(extrinsic information)とは区別しなければならない。

われわれが普段「情報」という言葉を使う時は外的な情報のことを指している。これは観測者によって依存する量である。例えば温度計を観測することを考えよう。外の気温に関心がある観測者が温度計を見たすると、温度計は明らかに情報を持っている。Iを情報量として、

$$I(\text{温度計の状態；気温}) > 0$$

となる。この式は温度計の状態と気温との間の情報量が0より大きいということを意味している。情報量とはこの場合、温度計の状態を知ること、外の気温をどれだけ正確に当てることができるかを定量化したものである。

一方、湿度には全く関心がなく、例えば湿度に関心がある観測者にとってこの温度計は情報があるといえるだろうか。温度計の状態を知っても湿度が分かることはないので、当然ながら情報量は0である。

$$I(\text{温度計の状態；湿度}) = 0$$

非常に当たり前のことを長々と記述して恐縮であるが、温度計が情報を持つのか否かを議論する際にはあくまで、“何”に対しての情報か(例えば温度なのか湿度なのか)という対象を決めない限りは全く意味をなさないということに今一度着目してほしい。情報を測る“対象”を決めるのは観測者であって、そこには観測者依存の恣意性があるのである。このように外部の観測者に依存して決まる情報を外的な情報と呼んでいる。

一方、意識とはあるシステムにとっての主観的な体験であって、外部の観測者に依存せず一意に決まる量であるべきである。内的な情報量とは数学的にどのように定義されるべきであろうか。これは非常に深い問題であるが、IITはこの内的な情報量を測る一つの方法を提案している。それを次節で簡単に紹介する。

2. 内的な情報量(intrinsic information)の測り方

先に述べたように、意識と関係があるのは内的な情報量であり、外的な情報量とは無関係である。外的な情報量とは、システムの内側の状態(例えば温度計の状態)とシステムの外の世界(例えば気温)とにどれくらい関係性があるかを定量化したものであった。一方、内的な情報量はシステムの外側の世界とは全く無関係に、システムの内側の状態だけを使って定義されるべき量である。IITにおける内的な情報量とは「システムの現在の状態が、システムの過去、未来の状態に対して持つ情報」を定量化したものである。これはいわば、システムの現在の状態を知った時に、それを引き起こした原因(過去)は何であり、結果(未来)が何であるかをどれだけ予測できるか、すなわち現在と過去・未来との因果関係を定量化することに他ならない。システムが多くの内部状態をもち(様々な状態を取り得る可能性があり)、かつ、現在から過去、および未来への因果性が強ければ強いほど内的な情報量は大きくなる。

逆に、システムの現在の状態が、過去と未来に対して何の因果関係も持たなければ内的な情報量は0である。例えば、先にも出た単純な温度計などは外部の気温によってシステムの状態が決まるので、現在の状態が過去と未来に対して因果関係を持たない。すなわち内的な情報量は0である。温度計は気温に対して外的な情報量はあるので、通常情報はありと勘違いしてしまうが内的な情報量は0なのである。フォトダイオードの思考実験において述べたフォトダイオードも同様に、そもそも内的な情報量が0なので意識は全く存在し得ないということもいえる(意識がある最小限のフォトダイオードの例は文献9を参照)。

では具体的に、どのように内的な情報量を定量化するかをみていこう。今、簡単な例として、二つの素子からなるシステムを考える。この素子はそれぞれ0か1(ONかOFF)かの状態を取り、自分自身およびもう片方の素子からインプットを受け取っているものとする。両方のインプットが1であれば、次の時刻で8割の確率で1となる。それ以外の時は8割の確率で0となる。この二つの素子は自分と他方の素子のインプットのANDゲートの機能を持っていて、8割の確率で正常に機能し、2割の確率で失敗する素子だと思ってもらえればよい。このシステムの状態遷移をまとめた遷移確率行列(transition probability matrix; TPM)は図1のようになる。TPMはシステムがある状態にあった時、次の時刻である状態へ移る確率を示した行列で、システムに内在するメカニズムによって定まる行列である。

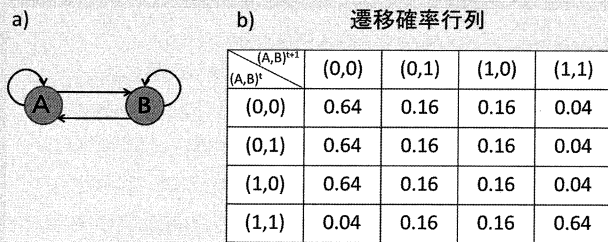


図 1

a) 二つの要素 A,B からなるシステム。それぞれ自分自身と他方の要素からインプットを受け取っている。

b) システムの遷移確率行列。時刻 t におけるシステムの状態, $(A,B)^t$ から次の時刻 $t+1$ におけるシステムの状態, $(A,B)^{t+1}$ へと移る遷移確率を示した行列である。

IIT では TPM に基づいて内的な情報量が計算されることになる。IIT において情報量は現在の状態に依存して決まる量である。例として、現在のシステムの状態が $(A,B) = (1,1)$ である時の情報量を計算することを考えよう。このためには、現在(時刻 t)の状態が得られた時に、過去(時刻 $t-1$)のシステムの状態(原因)が何であったかの確率分布、

$$p_{cause}^{whole} = p((A,B)^{t-1} | (A,B)^t = (1,1))$$

を求める。この分布は現在の状態と、システムのメカニズム(遷移確率行列)から制約を受けた原因の確率分布という意味で原因のレパートリー(cause repertoire)と呼ばれる。このシステムの場合原因のレパートリーは図 2 b1 のようになる。現在の状態が $(1,1)$ であった時、過去の状態は $(1,1)$ であった可能性が非常に高いことを示している。

一方、現在のシステムの状態が与えられていない時の過去の状態の確率分布, $p((A,B)^{t-1})$, は図 2 a1 のように一様な分布となる。すなわち、どの状態も等確率で起こり得るという分布である。

システムが現在ある特定の状態, $(A,B) = (1,1)$ にいること、そしてシステムに内在するメカニズム(AND ゲート)によって過去の分布が図 2 a1 の一様分布から図 2 b1 の非一様な分布へと変化する。この変化こそがシステムが生み出した内的な情報であると IIT は考えている。そして、図 2 a1 から図 2 b1 へとどれだけ変化したか、図 2 a1 と図 2 b1 の確率分布間の距離 D が内的な情報量の大きさである。

$$ci = D(p((A,B)^{t-1} | (A,B)^t = (1,1)) \| p(A,B)^{t-1})$$

この情報量は原因の情報量(cause information)と呼ばれる

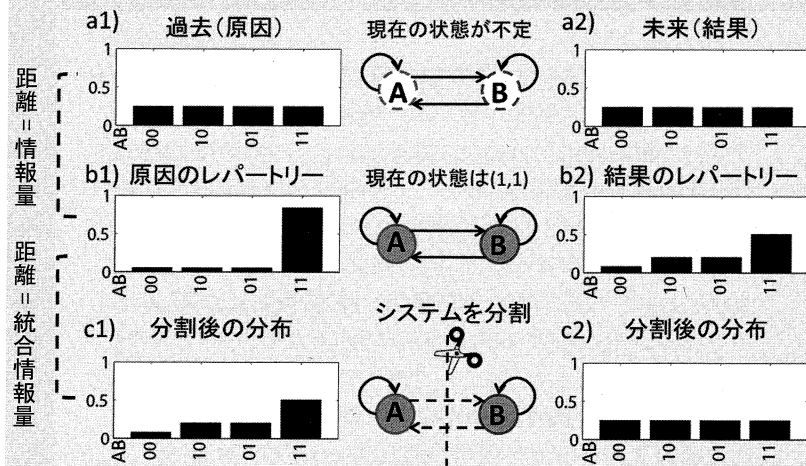


図 2 システムが生み出す内的な情報量と統合情報量の測り方

a1, a2) システムの現在の状態が定まっていない時の、過去・未来の分布。b1, b2) システムの現在の状態とシステムに内在するメカニズムによって、制約を受けた過去と未来の分布(原因・結果のレパートリー)。a1 と b1, a2 と b2 の差が内的に生まれる情報量である。c1, c2) システムを分割した際に定まる、過去と未来の分布。b1 と c1, b2 と c2 の差が統合情報量である。

(確率分布間の距離 D をどう測るかについては文献 9 を参照されたい)。

現在から未来(結果)に対する情報も同様に求められる。すなわち、まず、現在(時刻 t)のシステムの状態が得られた時に未来(時刻 $t+1$)のシステムの状態が何かの確率分布

$$p_{effect}^{whole} = p((A,B)^{t+1} | (A,B)^t = (1,1))$$

を求める。この確率分布は結果のレパートリー(effect repertoire)と呼ばれ、この場合、図 2 b2 のようになる。現在の状態が $(1,1)$ であるので、未来の状態も $(1,1)$ になる可能性が一番高く、両方の状態が 0 になる可能性が一番低くなる。現在の状態が与えられていない時の未来の状態の確率分布 $p((A,B)^{t+1})$ (図 2 a2)と結果のレパートリー(図 2 b2)の差が、現在から未来に対する結果の情報量(effect information)となる。

$$ei = D(p((A,B)^{t+1} | (A,B)^t = (1,1)) \| p(A,B)^{t+1})$$

以上まとめると、IIT では原因・結果の情報量(cause-effect information)が意識の量的な側面と関係する量と考えている。そして、現在の状態からシステムのメカニズムに制約される過去と未来の状態(取り得る原因と結果)の確率分布、原因・結果のレパートリー(cause-effect repertoire)が意識の質に関係すると考えている。

3. 統合(Integration)

情報の公理に続く次の公理は“意識は統合されている(Consciousness is integrated)”である。例えば、われわれが赤い三角形を見た時に、われわれの意識にのぼるのは「赤い三角形」という統合体であって、「赤」だけ、もしくは「三角形」だけを独立に意識することはできない。別な例として、左視野に見えているもの、右視野に見えているものを独立に意識することもできない。常に、左視野と右視野とが統合されたもののみがわれわれの意識体験となる。つまり、われわれの意識体験というのは常に統合されたものである。

この公理からさらに一步踏み込んで、IIT は次の公準を提案する。“あるシステムが意識を持ちうるのは、そのシステムが生み出した情報が統合されている時だけである”。

情報の公準では、内的に情報を生み出すことが意識に不可欠だということを述べた。統合の公準はさらに、内的に生み出された情報が統合されることが意識に不可欠なものであるということを主張している。

これは、カメラの例のように、独立に情報処理をしている要素が組み合わさったシステムを考えれば理解が容易である。いくら、多数の独立な要素が情報を生み出していたとしても、それらの情報が統合されない限りはシステム全体に情報が共有されず、システム全体の内的な観点から大きな情報が生み出されているとはいえない。このようなシステムの中に意識があるとすれば、それぞれの要素の中の極めて微小な意識だけであり、システム全体に意識は存在し得ない。

さて、ここで当然問題になってくるのは、一体情報の統合というのは数学的にどのように定量化するべきかということである。次節では IIT が提案する情報の統合の測り方を紹介する。

4. 統合情報量(integrated information)の測り方

IIT では情報の統合の度合いを統合情報量(integrated information)と呼び、ギリシャ文字 ϕ (ファイ) で表す。IIT において統合情報量とは、システムを分割した際にどれだけ情報処理の質が変化するかを定量化したものと定義される。カメラの例でいえば、フォトダイオードの集団をばらばらに分割したとしても、そもそも独立な集団なので、その中で行われている情報処理に何らの質的な変化を起こさない。一方、われわれの脳の神経細胞集団をばらばらに分割してしまったとしたら、われわれの脳は全く正常に機能しないだろう。

簡単な例として、再び図 2 のシステムを考えよう。この

システムの統合情報量を測るためには、まずシステムを二つに分割する(図 2 c)。そして、素子 A (または B) だけの現在の状態だけを使って(すなわち、もう片方の状態は全く知らない状態で)、A (または B) の過去・未来の状態が何であったかの確率分布を計算する。そして、それぞれ独立に計算された分布を掛け算したものが、システムを分割した際の原因・結果のレポートリーである。

$$p_{cause}^{part} = p(A^{t-1} | A^t = 1) \times p(B^{t-1} | B^t = 1)$$

$$p_{effect}^{part} = p(A^{t+1} | A^t = 1) \times p(B^{t+1} | B^t = 1)$$

これらのシステムが分割された際の確率分布は図 2 c1, c2 に示す。システムが分割されたことによりシステムの情報処理の質が変わり、元のシステムにおける原因・結果のレポートリー(図 2 b1, b2)と異なっていることが分かる。この二つの分布がどれくらい違うか、すなわち確率分布間の距離が、統合情報量である。

$$\phi_{cause} = D(p_{cause}^{whole} || p_{cause}^{part})$$

$$\phi_{effect} = D(p_{effect}^{whole} || p_{effect}^{part})$$

前節では内的な情報量の大きさが意識の量に関係すると述べたが、いくら大きな内的な情報量を生み出しているも、その情報が統合されていなければ意識とは無関係というのが統合の公準が示すところである。従って、より正確にいうと統合情報量の大きさが意識の量に関係する。意識の質と関係するのは原因・結果のレポートリーである。

5. 排他(exclusion)

ここまでのことを踏まえて、IIT が要請する、あるシステムが意識を生み出すために必要な条件をまとめてみよう。

- 1) 内的に情報を生み出していないといけない。
- 2) 内的に生み出された情報が統合されていなければならない。

これらの条件だけを考えると、情報が統合されてさえいれば意識が発生するというのであれば、世界には互いにオーバーラップした無数の意識が存在してしまうことになり、われわれの直観に反するのではないかという懸念が生じる。例えば、二人の人間(二つの脳)が話していることを考えよう(図 3 a)。この時、われわれの直観的には二人の人間の間に境界があつて、二つの独立な意識が存在していると思える。しかし、会話によって二人の人間の間に情報の統合が多少なりとも生まれているとすると、二人の人間を全体のシステムと考えた時の意識というものも存在してしまうのではないかというような発想に至る。さらに考えていけ

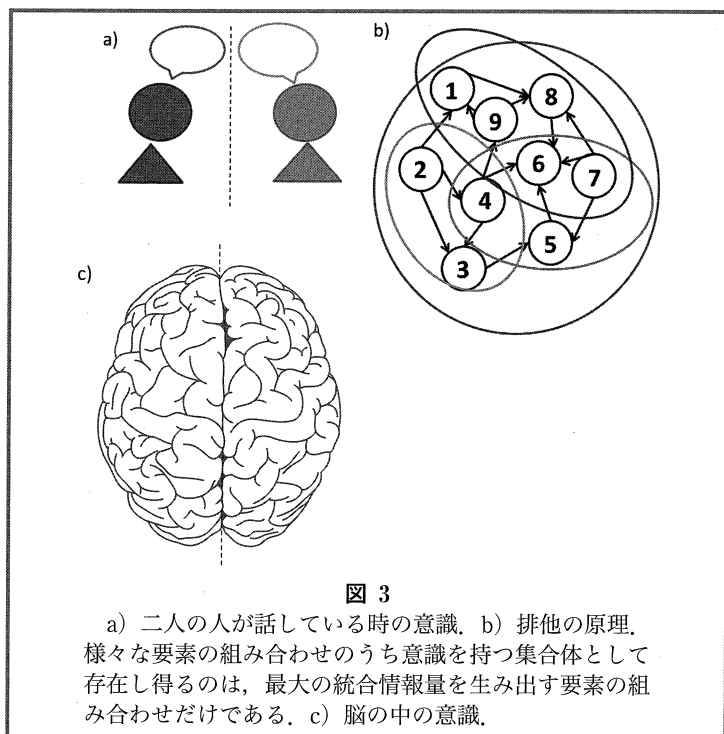


図 3

a) 二人の人が話している時の意識。b) 排他原理。様々な要素の組み合わせのうち意識を持つ集合体として存在し得るのは、最大の統合情報量を生み出す要素の組み合わせだけである。c) 脳の中の意識。

ば、国家という人間の集合体にも意識があることになってしまふのではないかなど、際限なく意識を持ち得るものの範囲が広がっていつてしまふ。

IIT はオーバーラップした無数の意識の存在を次の排他の公理、そしてそれに伴う排他の公準によって否定している。“意識は排他的である (Consciousness is exclusive)”。

例えば先ほどの A と B という二人の人間が会話をしているという状況の場合、世界に存在するのは A の意識と B の意識であり、A と B を合わせた集合体 AB に意識はない。すなわち、これは A の意識と B の意識が AB の意識を排他している状況である。それでは、どの意識が残り、どの意識が排他されるのか。これを決める原理として、IIT は次の公準を要請する。“ありとあらゆる要素の組み合わせのうち、意識を持つのは最大の統合情報量を生み出すものだけである。”

最大の統合情報量を生み出す要素の組み合わせをコンプレックス (complex) と呼ぶ。IIT によればコンプレックスが意識の座となる。排他の公準は排他の原理 (exclusion principle) とも呼ばれる。

排他の原理の意図するところを理解するために図 3b を見てほしい。この図のシステムには 1 から 9 までの要素が存在する。コンプレックスを同定するためには、要素のセットの全ての組み合わせを考える。図では [1 6 7 8 9], [2 3 4], [4 5 6 7], [1 2 3 4 5 6 7 8 9] などが組み合わせの例として示されている。さて、今仮に [1 2 3 4 5 6 7 8 9], すな

わちシステム全体がどの部分システムよりも大きな統合情報量を生み出しているとしよう。この時はシステム全体がコンプレックスであり、システム全体に一つの意識があることになる。排他の原理により、それに含まれる部分系、[1 6 7 8 9], [2 3 4], [4 5 6 7] の組み合わせなどはコンプレックスではなく、意識が全くない。

今度は、逆に部分系の中の一つ、例えば [1 6 7 8 9] の組み合わせが、他のどの組み合わせが生み出す統合情報量よりも大きかったとしよう。この時は [1 6 7 8 9] がコンプレックスであり、意識を持つ集合体である。システム全体の意識というのは存在しない。それでは、[1 6 7 8 9] のコンプレックスに含まれない残りの部分 [2 3 4 5] に意識はあるのだろうか。この部分系の中でも同様に排他の原理を適用し、やはり最大の統合情報量を生み出すものがコンプレックスとなる。例えば、[2 3 4 5] の全体が最大であれば [2 3 4 5] は全体として意識を持つことになるし、部分系 [2 3 4] が最大であればこの部分だけが意識を持つ。このように、あるシステムには複数のコンプレックスが存在することもあるが、コンプレックス同士は互いにオーバーラップしない。排他の原理を適用することによって意識の境界線が決まっていくことになる。

この排他の公理と公準を認めると興味深い予測がいくつか成り立つ。その一例がわれわれの脳に関してである (図 3c)。われわれの脳は左脳と右脳という二つの脳が脳梁でつながっているという構造を持っている。われわれは通常、自分の脳の中に意識は一つと思っていると思うが、仮に脳梁がなくなってしまったらどうなってしまうだろう。実際、重症なてんかんの治療として、脳梁が部分的に切断されてしまった患者が存在する。この場合、脳梁の切断の程度にもよるが、左脳と右脳にあたかも独立な意識が生まれたかのような振る舞いを示すことが知られている¹⁰⁾。

この例からも分かる通り、左脳や右脳はそれぞれ一つで意識を持つ能力を有している。しかし、脳梁によって強く結合することによって、すなわち左脳と右脳の間で情報が強く統合されることによって、左脳と右脳の全体のシステムに意識があるということになっている。排他の原理に従えば、われわれの脳が通常の状態であれば、左脳と右脳の中だけで生み出されている統合情報量よりも脳全体で生み出されている統合情報量の方が大きいため、脳は全体として意識があるという説明になる。これとは逆に、脳梁が部分的に切断されてしまった患者では、脳全体として生み出されている統合情報量が左脳と右脳の部分で生み出されている統合情報量よりも小さくなり、意識が二つに分離する

ということが起こっているのではないかと考えられるのである。

ここで一つの興味深い実験として、脳梁の結合を徐々に弱くしていく実験が考えられる。IITの予想が正しければ、脳梁の結合を徐々に弱くしていくと、意識が二つに分離する相転移点が存在することになる。そして逆に、二つの意識が分離した状態から徐々に結合を強くしていくと、また同じ相転移点で意識が一つに統合されるということが起こるということになる。この時、分離していた二つの意識の存在が“消えて”しまい、別の意識を持った存在が突如として現れるという、興味深い現象が起きていることになる。

IITの実験的検証

IITは検証可能な実験的予測をいくつかしている^{4,5,8,9}。例えば、IITに従えば、夢を見ていない深い睡眠時、麻酔下、てんかん、植物状態などの意識がなくなる状態では統合情報量は下がるはずである。これを示唆する実験結果はいくつかあるが^{11~13}、直接的に神経活動の記録から統合情報量を計算してこれを示した例はほとんどない。本稿では統合情報量の数学的な詳細は記述を省いたが、統合情報量は計算が非常に困難な量である。数十の要素からなる小さなシステムであれば正確な計算が可能であるが、人間の脳のような数十億個の神経細胞からなる巨大で複雑なシステムにおける計算は事実上不可能といってよい。さらにいえば、計算に必要な遷移確率行列(TPM)を実験から推定することが困難という問題もある。

しかしながら、理論にのっとった“正確な”計算が困難だとしても、何らかの近似的な計算方法によって統合情報量を計算することは重要な意義があると考えられる。そもそもIITにおける数学的な定式化自体、現段階でも発展途上のものであり、理論で提案されている通りに計算することに現段階でこだわる必要はないと筆者は考えている。より重要なのは、理論の本質を理解して、実際に測定可能である“近似的な”統合情報量の指標を導出したり、新たな発想を持って実験を行うことであろう。

IITの提唱者であるTononi自身も、脳波(EEG)から近似的に情報の統合の度合いを測定することでIITの検証を行なっている。情報の統合の度合いは、経頭蓋磁気刺激(transcranial magnetic stimulation)に対してどれくらい脳活動が脳全体に広がっていくかを定量化することで測っている。実際、意識レベルが下がっている時では、磁気刺激によって生じた脳活動が刺激を加えた場所に局所的に留まり、全体に広がらないということが知られている^{11~13}。こ

れは、意識レベルが下がった状態では領野間の機能的な結合が切れてしまっていて、情報の統合の度合いが下がっていることを示唆している。Casaliらは、磁気刺激による脳活動の広がりをPCI(perturbational complexity index)という指標をもって定量化し、その指標をもって、覚醒状態と意識レベルが下がった状態とを弁別することに成功している¹⁴。この研究の重要なポイントは、睡眠、麻酔、植物状態という異なる意識状態を全く同じ指標をもって覚醒状態から弁別することに成功している点である。ある特定の意識状態の変化に関して、特有の脳活動の変化を抽出することは比較的容易であるが、異なる意識状態の変化に共通の脳活動の変化を抽出することは非常に困難なのである。例えば、麻酔の深度を測る指標として臨床の現場でも用いられているBIS(bispectral index)という指標は、脳が損傷して無意識状態にある患者と通常の意識のある人とを、信頼度高く弁別しないことが知られている¹⁵。このような状況では、データからボトムアップで共通する脳活動の変化を抽出するというやり方ではなく、IITのような理論をベースとした指標をもって、トップダウンの方向でデータを説明するという研究の方向性が極めて重要となってくるのである。

今後の方向性とまとめ

今後の方向性として筆者が特に重要と考えるのは、IITを実験的に検証し、理論と実験の両側面から意識研究を発展させていくことである。現状、実験的な検証は十分なされていない状態といってよいが、前節で少し紹介したような検証^{11~14}は徐々にこなされてきている。本稿ではIITの理論そのものの要点を丁寧に説明することに紙面を割いたため、今後、どのようにして実験的な検証を進めるべきか、何が重要な課題かという点に関しては十分記述しなかった。これについては、また別の機会に筆者自身が行なっている研究と併せて紹介したいと思う。以前筆者の書いたレビュー³)でも、本稿ではふれなかった点に関して記述しているので参照していただければ幸いである。

以下、本稿で記述したIITの最も重要な事柄をまとめておこう。

1) IITは思考実験、現象論的な考察から情報、統合、排他の公理そして公準を仮定する。

2) 意識と関係するのはシステムのメカニズムによって内部的に生み出された情報量である。外部の観測者に依存する外的な情報量とは明確に区別されなければならない。

3) 内部的に生み出された情報量がどれだけ統合されてい

るか、すなわち統合情報量が意識の量(意識レベル)を決める。

4) システムのメカニズムによって現在の状態から過去と未来の確率分布がどのように制約されるか(原因・結果のレポートリー)が、意識の質を決める。

5) 排他の原理によると、意識を持つ存在(コンプレックス)とは局所的に最大の統合情報量を生み出す要素の組み合わせである。コンプレックス同士は互いにオーバーラップせず、明確な境界線がある。

統合情報理論という名前の通り、IIT における最重要概念は情報の統合、統合情報量である。情報の統合が意識にとって重要だというメッセージは伝わりやすいものであるが、しばしば見逃されがちなポイントは、まず IIT における情報量とは外部の観測者に依存しない内的な情報量であるという点、そして、排他の原理によって、無数のオーバーラップした意識が存在することを否定している点である。IIT に対する批判はこれらの点を理解せずになされることが多い。本稿では、これらの重要性を強調するために丁寧な説明を試みたつもりである。

何度か繰り返し書いていることであるが、IIT はまだ発展途上の理論であり、特に理論で提案されているアイデアを数学的にどのように定式化するかという部分に関してはまだまだ議論の余地が多く残されている。現段階で重要なのは、理論の本質を理解することであって、数学を理解することは二の次である。本稿では数学的な定式化の説明を最小限度にとどめ、IIT における重要な考え方を紙面の許す限り説明した。筆者が本稿で説明したのは、IIT の理論全体で説明すべきことの半分以下の内容であり、この他にも重要なトピックは多く存在する。本稿を読んで IIT のさらなる詳細に興味を持った方は文献 4, 5, 8, 9 を参照していただければと思う。

IIT は今後、理論的發展と実験的検証の双方向から発展させていく必要がある理論である。もちろん、IIT は意識の理論といっても、現状、物理学の理論のような詳細な予測精度はないため、他にも実験結果と矛盾のない理論は存在し得るかもしれない。しかしながら、IIT は幅広い現象

を説明し得る一般性を持った有力な意識の理論であり、この理論の考え方を基にして実験的検証を行なっていくことは意識研究を大きく進展させ得ると筆者は考えている。

謝 辞： 本稿は筆者が、IIT の提唱者である Giulio Tononi 教授と共同研究することによって得られた IIT に対する理解をまとめたものである。Tononi 教授との長きにわたる議論(丸々一日が議論だけで終わることもしばしばであった)がなければ、IIT を十分理解することはなかったであろう。Tononi 教授には深い感謝の意を示したい。共同研究者である土谷尚嗣教授との議論によっても IIT の理解を深めることができた。併せて感謝したい。また、原稿にコメントをいただいた笹井俊太郎氏に感謝する。

文 献

- 1) Owen AM, Coleman MR. Functional neuroimaging of the vegetative state. *Nat Rev Neurosci*. 2008 ; 9 : 235-43.
- 2) Boly M. Measuring the fading consciousness in the human brain. *Curr Opin Neurol*. 2011 ; 24 : 394-400.
- 3) 大泉匡史, 土谷尚嗣. 温度計に意識はあるか?—意識レベルの定量化へ向けた理論と実践. *LiSA*. 2012 ; 19 : 352-9.
- 4) Tononi G. An information integration theory of consciousness. *BMC Neurosci*. 2004 ; 5 : 42.
- 5) Tononi G. Consciousness as integrated information : a provisional manifesto. *Biol Bull*. 2008 ; 215 : 216-42.
- 6) Balduzzi D, Tononi G. Integrated information in discrete dynamical systems : motivation and theoretical framework. *PLoS Comput Biol*. 2008 ; 4 : e1000091.
- 7) Balduzzi D, Tononi G. Qualia : the geometry of integrated information. *PLoS Comput Biol*. 2009 ; 5 : e1000462.
- 8) Tononi G. Integrated information theory of consciousness : an updated account. *Arch Ital Biol*. 2012 ; 150 : 56-90.
- 9) Oizumi M, Albantakis L, Tononi G. *PLoS Comput Biol*. 2014 ; 10 : e1003588.
- 10) Gazzaniga MS. Forty-five years of split-brain research and still going strong. *Nat Rev Neurosci*. 2005 ; 6 : 6539.
- 11) Massimini M, Ferrarelli F, Huber R, et al. Breakdown of cortical effective connectivity during sleep. *Science*. 2005 ; 309 : 2228-32.
- 12) Ferrarelli F, Massimini M, Sarasso S, et al. Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proc Natl Acad Sci USA*. 2010 ; 107 : 2681-6.
- 13) Rosanova M, Gosseries O, Casarotto S, et al. Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain*. 2012 ; 135 : 1308-20.
- 14) Casali AG, Gosseries O, Rosanova M, et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Transl Med*. 2013 ; 5 : 198ra105.
- 15) Schnakers C, Majerus S, Laureys S. Bispectral analysis of electroencephalogram signals during recovery from coma : preliminary findings. *Neuropsychol Rehabil*. 2005 ; 15 : 381-8.