

# 学生フォーラム

## 第98回 渡邊正峰先生インタビュー 「意識をアップロードする」

今回は、東京大学大学院工学系研究科准教授で脳神経科学が専門の渡邊正峰先生にお話を伺った。渡邊先生は、大学で意識を機械にアップロードするための研究を行う一方、近年の意識研究の成果に基づき、人間の意識をデバイスに移植することを目指すスタートアップ「MinD in a Device」社の技術顧問としても活躍されている。インタビューを通じて、意識のアップロード構想の全貌から、それがもたらす価値観・死生観の変化まで議論することができた。

——まず、渡邊先生の研究内容について教えてください。

20年後の意識のアップロードを目指して研究を進めています。現在、二本立てとなっていて、ドイツのマックス・プランク研究所では、意識のアップロードのプロトタイプ実験として「マウスと機械の視覚的意識の一体化」、MinD in a Device社では、意識をアップロードする対象としての「生成モデル」の研究開発に取り組んでいます。

——「意識のアップロード」というとなかなか想像がつかないのですが、脳活動が再現できれば良いのでしょうか？ そもそも意識とは何でしょうか？

まずは意識を定義しないと話が始まらないですね。哲学者などは、意識の定義についてあれやこれやと議論を繰り返すのが仕事ですが、私達のように何年もかけて一つの実験に取り組む立場からすれば、その土台となる定義がぐらついてはたまったものではありません。そんな私達にとっての意識の定義とは、ずばり、「感



図1 渡邊先生の研究室にて、資料を用いながら「意識のアップロード」構想に関する説明をいただいた

覚意識体験（クオリア）」です。

脳にしても、人工神経回路網にしても、第三者的な視点から眺めれば、ネットワーク上を信号が行き交っているに過ぎません。しかし、我々の脳が「赤いリング」の視覚入力を受け、それを処理したときには、私達自身に「赤いリングが見える」との感覚が生まれます。ここでの私とは、まさに脳の神経回路網であり、つまり神経回路網が第一人称を有していることになります。私達にモノが見えている時点で、そのこと自体に疑いの余地はありませんが、よく考えれば非常に不思議なことです。神経回路網とはいえ、それは単なるモノに過ぎず、そこに第一人称が発生しているわけですから。

普通に考えれば、コーヒーカップのようなモノには第一人称的な感覚は存在しない。一方で、脳の神経回路網にはどうしてもなくそれが存在している。視覚処理に

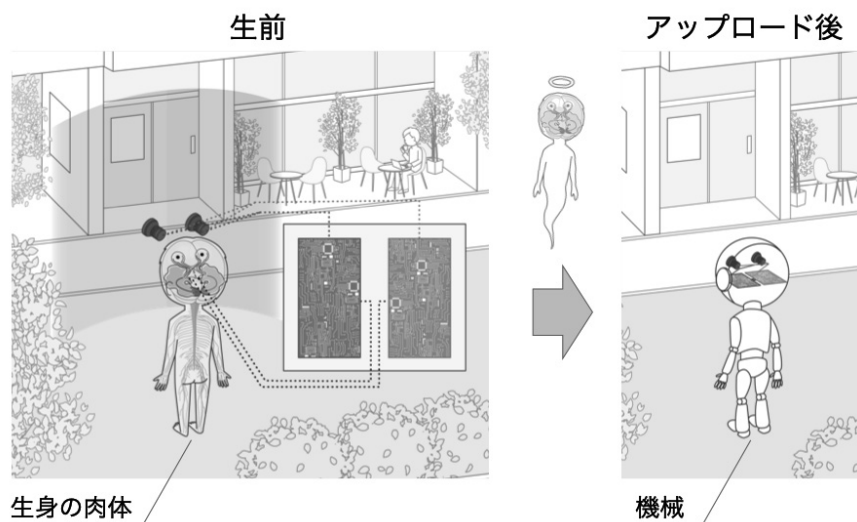


図1 「意識のアップロード」の概念図。生体脳半球と機械半球の意識を一体化する

は視覚体験が伴い、聴覚処理には聴覚体験が伴い、意思決定には意思決定の体験が伴う。この神経回路網の第一人称的な感覚、すなわち、感覚意識体験こそが私達の意識の定義となります。

よって、意識をアップロードするということは、自身の神経回路網の第一人称を人工の神経回路網に移すことに相当します。

——機械と人間のハードウェアの違いが問題にはならないのでしょうか？

私はハードウェアには依存しないと考えています。二つの材料から説明させてください。

はじめに、哲学者のチャーメーズが考案した「フェーディングクオリア」と呼ばれる思考実験を紹介しましょう。ここでは、意識が発生している状態で、脳のニューロンを一つ一つシリコン製の人工ニューロンに置き換えていくことを考えます。その際、もとのニューロンの結合関係を完全に保持し、スパイク列の入出力関係を完全に再現可能な人工ニューロンを想定します。すると、残りの脳がその「置換えに気付かない」、正確には、残りの脳に全く影響が及ばない状態が生まれます。

では、このような操作を繰り返す、すべてのニューロンがシリコン製のものに置き換わった後には、意識はどうなっているのでしょうか？チャーメーズは、出来上がったシリコン製デバイスがもとの脳と全く同じ働きをしていることから意識は維持される、すなわち、機械にも意識が宿り得ると論考しています。

ただし、フェーディングクオリアの果てに出来上がるデバイスは、個々のニューロンが抽象化されてはいるものの（必ずしも、神経伝達物質やイオンチャネルなどの機構をもたなくてもよい）、その三次元配線構造はもとの脳のまま非常に複雑で、工学的に実現しようとするとは非常に厄介な代物です。

そこで二つ目の材料として登場するのが、著書<sup>\*1</sup>の中で提案した「デジタルフェーディングクオリア」です。20年後の意識のアップロードに向け、ノイマン型コンピュータにも意識が宿るかを問うものになっています。

ここでは、脳のニューロンをシリコン製のものに置き換える代わりに、ニューロンをコンピュータの中に一つずつ取り込んでいきます。ブレインマシンインタフェースが完璧で、数値シミュレーションされた仮想ニューロンが、スパイク列の入出力を完全に再現したなら、1個目の取込みに関しては「フェーディングクオリア」と同様、脳側には全く影響が及ばないはずで、少々事情が異なるのは2個目以降で、脳の中でもともとつながっていたニューロンを複数取り込むことになるかもしれませ

ん。ただ、それにしても、それらのニューロンの相互作用も含めてシミュレーションすることにより、脳が取込みにいっさい気付かない状況をつくることができます。同様にして、すべてのニューロンがコンピュータの中に取り込まれた後にも、その動作は元の脳を完全に再現したのになります。よって、コンピュータの中に仮想的に存在する神経回路網にも意識が宿り得ると論考することができます。

これらの思考実験により、シリコン製の超並列デバイスにも、ノイマン型のコンピュータにも意識が宿ることとなり、意識はハードウェアに依存しないと私は考えています。

——特定の人間の神経回路網のコネクションを解明して、それをそのまま完全に再現する必要はないのでしょうか？

それはとても興味深いポイントで、私は脳の「完全コピー」は、少なくとも当分の間は不可能であるとの立場をとっています。というのは、脳の中で三次元的に配線されている軸索のコネクション自体は、死後に脳を取り出して、冷凍して（豆腐のように柔らかい脳を固めるために）、薄くスライスしたものを走査型電子顕微鏡で観察し、軸索を1本1本追っていくことで、現在の技術でも読み解くことが可能です。事実、この方法で*C.elegans*（線虫の一種）の中樞神経系などは完全に解明されています。しかし、ニューロン間の結合の有無の特定と、シナプス結合の重みの特定とは全く別の話です。結合の有無だけであれば数 $\mu\text{m}$ の厚さのスライスで十分読み取ることが可能ですが（それにしても上記のプロセスが必要）、さらにシナプス結合値を決めるとなると、シナプス小胞やイオンチャネルの数の検出が可能となるようなオングストローム単位の精度が必要となり、少なくとも現時点では道筋が立っていません。

また、仮に私の脳の完コピーができたとしても、そこで蘇った意識は果たして私のものなのかとの疑問が大いに残ります。例えば、「どこでもドア」の仕掛けとして、ドアに入るのび太くんを分子スキャンした後に殺し、反対側のドアで再構成していたとしたらどうでしょう？ そんな「どこでもドア」にあなたなら入りますか？ 死後に脳を取り出して完コピーする手法は、この「殺人どこでもドア」と等価です。

——脳の「完コピー」が不可能だとして、それではどのように意識をアップロードしようとしているのでしょうか？

端的に言えば、まずはニュートラルな意識を機械に宿し、その機械を脳と長時間接続することで、機械の意識を徐々に「わたし色」に染めていくことにより、意識をアップロードできると考えています。ここで鍵を握るのは、生成モデルこそが意識の源である可能性です（そ

\*1 渡邊正峰：脳の意識 機械の意識—脳神経科学の挑戦, 336 pp., 中公新書 (2017)

の意味合いについてはぜひ著書\*1を読んでみてください)。この場合、ニュートラルな意識とは、視覚を例にあげるなら、生まれてこの方の個々人の視覚体験を完全に再現しなくとも、一般的な視覚刺激を生成モデルに与えることで実現される汎用的な視覚的意識を指します。色については哲学的な問題があることは承知していますが(私の「赤」とあなたの「赤」は同じか?)、例えば形や奥行きなどは、万人に共通であると考えています。

次に、このニュートラルな機械の意識と自身の脳の意識を一体化させるプロセスについて説明させていただきます。はじめに、左の生体脳半球と右の機械半球、右の生体脳半球と左の機械半球といった形で、たすき状に接続します。そのような状況での機械半球と生体脳半球の意識の一体化は、意識の源が生成モデルであったなら決して難しくはないはずで、両者の間で高次シンボリック表象さえ共有できれば達成されることとなります。なぜなら、例えば左視野の対象物を生成する際に、右視野の光源なども情報として必要ではあるものの、シンボリックなレベルで視野をまたいで共有されれば、生成過程自体(CGレンダリングをニューラルに実装したもの)は片視野内(よって、片半球内)に近似的には収まるからです。

以上のように、機械半球と生体脳半球の意識を一体化することで、先ほどの「殺人どこでもドア」とは異なり、「シームレス」に意識をアップロードすることが可能になります。目指しているのは、私の右脳半球が梗塞などによって壊死した後にも、私自身の意識が左脳半球で生き続けるような状態です。機械半球と生体脳半球の意識を一体化させることができたなら、肉体が減び、生体脳が終わりのときを迎えたとしても、私は機械半球の中で生き続けることになるでしょう。さらに、たすき状にないでいたそれぞれの機械半球どうしを結合したなら、一つの意識としての完全形で成立することになります。

—— 私自身は意識のアップロードに対してワクワクする気持ちをもつ一方で、不安を抱く人がいると思うのですが、どのように社会に受容されると考えていますか？

著書の刊行以来、一般聴衆を対象に講演させてもらう機会が増え、その際に簡単なアンケートをとるようにしています。面白いことに、機械に意識をアップロードしたいと考える人の割合は極端に少なく、逆に私が驚かされています。私自身は、中学生の時分から死にたくないと思っていたものですから(笑)。

しかし、実際に技術として確立され、隣人の意識がアップロードされるのを目の当たりにしたなら、それを願う人は格段に増えるのではないのでしょうか(そうでないと会社として成立しないですよ!)。また、そのときには「死」の概念自体が変わってしまうでしょう。現代の感覚からすれば、江戸時代の大飢饉で人口の半分が

餓死したなどという過去はとても残酷に思えますが、同様にして、自身の意識が完全に消失し、金輪際蘇らないということが、すごく野蛮に思える日が来るのではないのでしょうか。

—— 「メメント・モリ\*2」という言葉がありますが、従来の死生観も変化していくかもしれませんね。

ドイツのケルン大聖堂を訪れたとき、光り輝くステンドグラスに感動すると同時に、建立当時の市民の大半が文盲で、ステンドグラスはまさにそんな彼らでも理解できるように、聖書の物語を絵画的に表現したものであることを知り愕然としました。支配層が、荘厳な教会建築という大掛かりな仕掛けを通して、「死んで天国に行けば皆が幸せになれる。その代わり現世では我慢しろ」と都合の良いように市民を従わせようとしていたともいえます。同様に、「命ははかないからこそ美しい」などといった現代の考え方が普遍的に正しいとは限りません。

—— 先生の学生時代のお話をお伺いします。先生は工学系のバックグラウンドをおもちですが、一見、工学としては取付きようのない「意識」をなぜ研究しようと思われたのですか？

実は、大学進学時は理論物理学を目指していたのですが、70人そこそこのロシア語クラスのなかに自分より数学ができるのが何人もいて、「これはとてもかなわないな」と思い諦めました(笑)。で、その夢が叶わないなら、将来NASAにでも行こうと思い、進路振分けの第一希望は宇宙工学に出しました。それがわずかながら点が足りず、夏の暑い日の立て看板に、無情にも「原子力」の三文字がおどりました。第四希望として、「火星に行くなら核融合ロケットしかないだろう」と、ほんの冗談のつもりで記したに過ぎなかったのですが(笑)。それが、チェルノブイリ原発事故の直後ということもあり、正直、目の前が真っ暗になりました。ところが、授業に出たらこれが面白かった。調子に乗せられて、本気で核融合ロケットの開発を目指そうと思い修士に進学したところ、今度は、当時准教授であった古田一雄先生に「そんな研究は核不拡散条約の締結以来、世界のどこにも存在しないよ」と言われ、またまた目の前が真っ暗になりました。そんななか、古田先生が輪読の材料としてカオスの分厚い英語の本を指定したのです。そして運命の瞬間が訪れました。英語で理解するのが億劫で、ちょっとズルしようと日本語の解説書を書籍部に探しにいった私の目に、合原一幸先生の「カオス」という本が飛び込んできたのです。これまた結構ズルい本で(笑)、最初の数章は決定論カオスの一般的な説明に当てられているものの、後半は、合原先生の提唱するカオスニューラルネッ

\*2 memento mori. 「いつか死ぬことを忘れるな」というラテン語の警句。

トワークのみを扱っていました。それが、高校生の時分から研究者を目指していたものの、期せずしてテーマ難民と化していた私の心に深く響き、あれよあれよという間に脳の理論研究を始めることになりました。もしあのとき、別のカオスの本が目に入ったなら、全く異なる人生を歩んでいたことでしょう。

私の放浪の旅は、脳科学の門を叩いてからも続けました。修士に進学した1993年は、ちょうど第二次ニューラルネットワークブームの火が消えかかっていた時期に当たります。しだいにSVMなど脳の制約お構いなしで学習性能を追求する手法が台頭し、ニューラルネットワークは隅に追いやられていきました。押し寄せる波をもろに受け、私自身も含め、当時の「研究仲間」は、クモの子を散らすように理論研究から離れていったのです。

そんななか、私は「生モノの脳」への関心を高めていきました。はじめは自身の理論研究の延長で、修士後半から取り組んでいたある仮説をもとに電気生理実験データの解析を試みました。ところが、何を試しても仮説に反する結果ばかりで途方に暮れました。そこで食い下がる手もあったのですが、それまでの経験から、無理に流れに逆らわない習性のようなものが身についていたのか、4～5年かけて積み重ねてきた構想をあっさりと打ち捨ててしまいました。

そこで湧きおこったのは、自身でゼロから実験を組みたいとの思いです。先の試みの失敗は実験系のせいではけっしてなかったのですが、一方で、本来の実験目的からずれた二番煎じの解析には限界があると感じたのです。そこから心理物理実験を自身の研究室で始めて、さらに1年間のサバティカル期間を利用して、カリフォルニア工科大学の下條信輔先生のラボにお邪魔しました。そこで、金井良太さんや土谷尚嗣さんと出会い「意識の洗礼」を受けたのです。

長らく放浪の旅を続けた私も、「意識の不思議」に出逢い、それに取り憑かれてからはぶれていません。そして、今後もぶれないつもりです(四十にして惑わず!)

——これから先の時代を見据えて、今の学生・若手はどういったことを学び・考えるべきでしょうか？

私自身の経験から言えることは、無理に流れに逆らわず、身を任せるべきだということでしょうか。起きてしまったことは起きてしまったと諦め(Shit happens!), 流された先で人生を目一杯楽しむことが肝要ではないかと考えています。

私自身、多くの偶然の積重ねで意識研究にたどり着きましたが、他のパラレル宇宙には(そんなものがあつたとしたら!), 違った自分がたくさんいるのではと、割と本気で考えています。そういう意味で私は、たまたま意識に興味をもち、たまたま意識のアップロードにも応用できる手法(著書\*1の「人工意識の半球接続主観テスト」を参照してください)を思いつき、中学時代に死の恐怖を克服できなかった私は、これ幸いと自身が生きているうちに挑戦しようと思いついたのに過ぎません。

——貴重なお話、ありがとうございました。

松嶋 達也(東京大学), 藤堂 健世(東京工業大学)