

「シンギュラリティの十分条件」

山口優

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うたらどうなってしまうのだろうか。実際、重度なてんかんの治療として、脳梁が部分的に切除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。(中略) ここで一つ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。(中略) 脳梁の結合を弱くしていくと、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、二つに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ことになる。

——大泉匡史「意識の統合情報理論」

*Clinical Neuroscience* vol. 32, no. 8 (2014-8)

「サーボモーターと意識」「サブリミナルへの福音」に続き、本稿でも意識の問題について考察を進めたいと思います。

SF及びかつての科学においては、人と人が分かり合うことのできる手段として、思考や感情を共有するテレパシーという概念がよく用いられてきました。

現実には、特殊能力という形ではなく、TMS等の脳磁図を媒介とした技術的手段が最近によく研究されるようになってきています。また、海馬にアクセスすることでマウスに対して記憶操作をすることも可能になる等、人の脳内と外部との間で情報を入出力する科学技術的な方法が徐々に確立されつつあります。

私は「アルヴ・レズル」や「アンノウン・アルヴ」等の一連の小説において、人体を透過する帯域の電磁波のアンテナを備え、神経細胞と類似の構造をしたナノマシンを仮想することで、より精密かつ確実な脳と外部の情報のやりとりの思考実験を行ったことがあります。

こうした思考実験を行うにあたり、特に気にしたのが、「人は外部とどこまでもつながっても、自身の人格を保てるのだろうか」ということです。人は通常、皮膚や感覚器を使って外部の情報を得、運動器を使って外部に情報を出力しています。その情報量は毎秒一〇〇万ビット程度とされていますが、もしこれが人間の思考の全情報量（一兆

ビット程度)と同等になれば、人の人格、或いは意識そのものはどうなるのでしょうか？ 私には外部とのつながりの帯域があまりにも拡がると、意識は外部と一体化してしまい、「私」というものもなくなってしまうだろう、と考えました。

そのような帯域の広い脳情報の入出力技術は未だ存在せず、人の人格が外部と混じってしまふようなことも未だ起こっていませんが、逆に、一つの人格を二つに分割する事例ならば既に存在します。

脳梁切断手術、というのがそれです。

重度のてんかんの症状を呈する人に対して、その症状を抑えるため、右脳と左脳を接続させる脳梁を切断してしまう手術だそうです。素人には極めて乱暴に聞こえる手術ですが、かつてはよく行われていたとのことでした。

冒頭に紹介したのは、その切断手術を紹介した一節です。この論文では、「意識とは何か」という本質的な問いへの答えも徐々に確立しつつある様が描かれています。この論文で紹介されている理論の提唱者、ジュリオ・トノーニ博士らの唱えた「意識の統合情報理論」によれば、内部情報の統合度が最も高い範囲のみが意識を持つだろう、とされています。人間で言えば大脳の皮質―視床系にあたります。そこが人間の「意識の座」というわけです。松果体のように狭い領域ではありませんが、かといって小脳等も含め、

「脳全体にある」というわけでもなさそうです。

内部情報の統合度、というとなにやらわかりにくい概念ですが、まず内部情報というのは、内部の情報ノード（人にとっては神経細胞）のつながり方だけで、どれだけ多様な情報を生み出せるか、ということですよ。ある人に対して、ある絵を見せたときの感想を予想してみましよう。どんな感想をその人が言うか、あなたには予想できるでしょうか？ おそらく実際に見せてみないと分からないでしょう。「この構図がいい」「色合いがいい」「昔見た風景に似ている」「はやくこの絵のような場所に休暇に行きたいものだ」などなど、その人の口にする感想には無数の可能性があるからです。では、絵を見せたら必ず「きれいですね」というロボットだったらどうでしょう。そのロボットは「きれいですね」と言うに違いないとあなたは予想しますし、確実にそうなります。それはそのプログラムされたロボットには一つの可能性しかないからです。「絵を見せる」という外部からの刺激に対して、どのぐらい多様な可能性があるか、その点において大脳の視床―皮質系は優れているのです。そして、その多様性を生み出している脳領域は、適切なプロセスを辿れば特定でき、それが大脳の視床―皮質系だということです。

ちなみに小脳にはそのような多様性はありません。小脳は運動をスムーズに行うことに特化していて、「歩く」「投げる」等の動作命令に対しては（スピードなどのパラメ

ータはありますが)毎回異なる反応を示すことは有り得ません。そんなことになったら、あなた日々の単純な動作をどのようにするのか、いちいち調整しなければならなくなるでしょう。

人間の脳には無数の神経細胞があり、その神経細胞同士の結びつきの強さの組み合わせは天文学的な数字になります。しかし、「最も高い内部情報統合度」つまり、ノードの数あたり最も多様な可能性を生み出す組み合わせというのは、既存の系に対して一つに決定することが出来、その中で最大のものが意識を担う脳領域になるわけです。

さて、冒頭から述べている、「人の脳が外部とのつながりを密にしていけばどうなるか」という問いは、この分離脳の逆を考えれば良いのだと言うことは、既に読者の皆様もお気づきの事と思います。

例えば、NLNでも進化したTMS技術でも良いですが、人と人の脳の間を、現在では考えられないような高密度さで情報連結することができたとしましょう。徐々に情報連結の帯域幅を広げていけば、いつか(分離脳が元の統合された脳に戻るように)人と人の脳の意識を担う領野は、個々の統合情報量よりも、それらを合わせた統合情報量の方が大きくなる「相転移点」に至るでしょう。そうなったとき、個々の脳からは意識はなくなり、新たな統合された人格が生まれることになるのだらうと思います。

統合する相手は人間とは限りません。より進化していく機械的な知能である可能性もあるでしょう。現在の機械的な知能との結びつきは、ディスプレイ、マウス・キーボードなど、極めて限定されたもので、とてもではありませんがそれらと結びついている方がウェットな（脳内に存在する）意識よりも統合情報量が大きいというものではありません。しかし、機械と人との結びつきが様々な技術により、より密になっていくとすれば、いつか、こうした、人の大脳と機械を含んだノード群の方が、大脳の皮質―視床系単独よりも統合情報量の多い情報コアとなっていくことも予想できます。そうなる、「私たちそのもの」である意識は機械的な部分も含んだものとなり、意識の変容は避けられない事態となります。

しかし、「人と機械の情報連結がより密になっていく」という前提の是非を、この先、話を進める前に問わねばならないでしょう。

今後人工知能がより発展していく中で、確実に問題になってくのが、その人工知能の判断の妥当性の検証です。現在、自動運転車において議論になっているように、たとえ、機械の方が判断を誤る可能性が人よりも少ないのだとしても、人は確実に、機械の判断を検証し、場合によっては正すというプロセスを確保したがるでしょう。重要な判断を機械に任せようになればなるほど、そのニーズは強まるはずで、株取引などは純粹

に利益が多少増えたり減ったりするだけで、自動運転では人命が関わります。介護や医療でも同じ問いに直面するでしょう。

人工知能の判断の高度化は、人に判断の意味を素早く理解させ、的確に検証させる技術のニーズをも生むのだと私は思っています。その際、高速化していく機械に対して、より密にコミュニケーションするため、そして機械の判断についていくために、人間の判断力をもエンハンスしなければなりません。

無論、これは一つの予想であり、そうならない可能性もあります。機械を信頼し、全てを任せてしまう、ということですが、しかし、同じ人間同士であっても他者を信頼できない人間です。機械ならば従順だから人間よりも信頼できる、ということにはならないでしょう。その機械を造っているのもまた、他者である人間なので、今はまだ、人工知能によるカタストロフ——或いはもしそれが自分に起こっていたらと人を震えさせるような事故は起こっていません。しかしそれが起こったが最後、こうした考えは主流ではなくなっていくでしょう。

そして、人間自身をエンハンスするという考えも、機械に全てを任せるという考えも、いずれも肯定しない人々は存在すると思います。未だ地球上には、文明を受け入れず、原始のままの生活を送っている人がいるように。しかし、最先端の技術力を持たないそ

うした人々の相対的な影響力は少なくなっていくでしょう。

三つの考え方がずれが「正しい」のか、私には分かりません。しかし、一つの選択肢として、——おそらく最も急進的な人々は——人間自身をエンハンスすることになるのだろうかと思います。

従来、我々はAIやコンピュータの能力が向上することにはばかり目を奪われてきました。技術的特異点、テクノロジカルシンギュラリティというのも最近はそのような事象として語られることの多い概念です。特異点というのは、数学上の概念で、複素平面上において、その点を囲むなめらかな関数の閉経路の積分がゼロとならない、という「コーシーの定理」に基づく不連続な点という意味ですが、その概念を技術史へ援用したものとと言えるでしょう。技術進歩の不連続な飛躍。それより先は、現在の発展状況からどのようななめらかな関数によって外挿しようとする合理的な予測ができない点です。それを為す主体はAIに限りません。エンハンスされた我々人類自身かもしれません。もしエンハンスされた人類が主体となるならば、AIという言葉は過去の古い技術を描写するだけのものになるでしょう。我々の一部でしかないものに特別の名前をつける必要はありません。ただ、「人間」あるいは「私」の定義が少し変わるといっただけです。

いずれにせよ、その必要条件として、コンピュータの能力とAIの能力の急速な進歩



が挙げられるのは妥当なところだ。しかし、もし人類が状況に流されず、主体的に自らの未来を選ぶ意志を持つべきだという考えを採るならば、そうした機械の進歩に対して、我々人類がどのような選択肢を採るべきなのか、選ぶことができるようになって初めて、我々は技術的特異点の十分条件を備えたと言えるのではないだろうか。その条件として、私は意識のメカニズムの解明を挙げたいと思います。