

## 反実仮想的な情報生成による人工意識の構築

金井良太

Ryota Kanai

株式会社アラヤ・ブレイン・イメージング  
ARAYA

意識という主観的な現象も自然現象の一部であるはずであり、そのため意識も何らかの普遍性を持つ自然法則に従っているはずである。つまり、どのような物理的条件または情報的条件において意識が生じるのかを正確に同定することのできる意識の科学的理論というものが存在するはずである。このような意識を自然現象と位置づける見方においては、生物学的な脳が唯一の意識を生じさせる唯一の基盤ではなく、脳以外の物理的基盤であっても、生成の必要十分条件を満たす物理システムであれば、意識が生み出されるはずである。つまり人工意識の構築は原理的には可能であると考えられる。しかしながら、その本質的な条件については、現時点ではわかっていない。そのような理論を見つけるためには意識の機能の本質を見つけ出す必要があるが、意識の主観的な性質により機能の同定が困難であった。本研究では、心理学や神経科学の研究を俯瞰することで、意識を必要とする認知課題とその神経学的機構を同定し、意識の機能的本質を見極めることを試みた。そこから現れてきた意識と関連に深い機能として代表的なものとして、非反射的行動、トレース条件付け、短期記憶、行動のプランニングがある。このような機能の共通項を分析することで、意識の本質的な機能は「反実仮想的な状況の感覚表現を内的なモデルに基いて生成する能力」であると結論するに至った。すなわち、現在目の前で起きていることではなく、数秒程度の過去や未来の出来事を、視聴覚等の感覚情報のフォーマットによって内的に生成することが意識の機能であると考えられる。そのような機能を実現する必要条件としては、感覚運動ループを通じた環境との相互作用により「自己」を含んだ生成モデルの獲得が必要である。この「意識の情報生成理論」によって、意図・注意・思考といった主観的な心理状態に対してメカニズムとしての解釈を与えることができる。さらに、我々は反実仮想の生成過程を現代的なニューラルネットワークによって実装し、反実仮想的に未来を生成することで行動の方策を選定するエージェントを構築した。「意識の情報生成理論」では、このような方針で人工意識のプロトタイプが構築できる可能性を含意している。本研究を通じて、情報生成するエージェントのプロトタイプでは、新しい目標が設定されてもフレキシブルに対応ができ、新しい環境を効率的に探索する「好奇心」を自然な形で実装できるという利点があるということが明らかになった。また、このような反実仮想を扱うために必要となる「情報生成」という観点が、これまでの脳の中やコンピュータ上での「情報処理」というものと、どのように違うのかについて検討する。特に両者はそれぞれ、機械学習的な文脈では、デコーディングとエンコーディングに対応し、予測符号化理論の観点では、予測と予測誤差に対応し、意識の統合情報理論の観点では、結果情報と原因情報に対応付けられる可能性がある。これまでは、脳の中での「情報処理」がなぜ意識的経験を生み出すのかとハードプロブレムが問われてきたが、今回の情報生成理論では、新たな「情報生成」という観点から意識の問題を捉え直す。特に、情報生成理論では、「情報処理」をエンコーディングや情報の圧縮と捉えると、これは意識を生み出さず、一方で、デコーディングとしての「情報生成」は意識を生み出すと考えている。