報道資料

令和 2 年 8 月 31 日
株式会社国際電気通信基礎技術研究所（ATR）
カリフォルニア州立大学ロサンゼルス校（UCLA）
科学技術振興機構（JST）

意識下の強化学習能力をメタ認知で開花させる

本研究成果のポイント

- ✔ fMRI[1]ニューロフィードバック[2]手法を利用して、意識下の脳活動パターンが最適な行動を決めるように、試行錯誤の賭けゲームを設定し、ヒトを対象に強化学習[3]実験を行いました。
- ✔ ヒトは自身の脳内の意識下の情報を読み取って利用し、数百試行以内に、試行錯誤で強化学習ができました。
- ✔ 実験参加者が自身の視覚知覚判断についてより確信している試行ほど、ゲームでより最適行動を選択することがわかりました。さらに、学習するにつれて、強化学習を司っている脳部位（大脳基底核）[4]と、メタ認知[5]を司っている脳部位（背外側前頭前野）[6]との間で、学習と確信度の情報が同期しました。
- ✔ 人工知能の難しい問題『学習サンプルが少数個しか無い時に、どのようにして非常に大規模で複雑な問題を学習するか[7]』の解決にメタ認知が必要なことがわかりました。
- ✔ ヒト脳で得られた本研究成果により、次世代 AI 学習アルゴリズムにおいても同様に、複雑な問題の解決にメタ認知が応用できる可能性が示されました。

概　要

ATR 脳情報通信総合研究所の Aurelio Cortese 主任研究員、川人光男所長、カリフォルニア州立大学ロサンゼルス校と香港大学の Hakwan Lau 教授は、ヒトが金銭報酬を得るために自身の意識下の脳内情報を読み取り、成功失敗の情報だけから試行錯誤で（強化）学習[3]できることを発見しました。さらに、確信度（メタ認知[5]能力の重要な要素）がこの学習過程に含まれていることを明らかにしました。これらの結果は、脳がメタ認知を利用して、超多次元の複雑な問題を簡単化し、少数のサンプルだけから高速で効率的な学習をしていること[7]を示しています。本研究成果は次世代の AI 開発に有用な指針を与えます。

背　景

ヒトの脳がどのようにして『次元の呪い[7]』を解決しているかは未だに解けていない謎です。これまで脳が持っている次元の呪いを解く能力は、科学的に実証されていませんでした。その理由は、従来の実験が簡単な条件と小さな次元の問題を扱っていたからです。次元の呪いは大規模な問題の解答が多すぎて実際の経験で試行錯誤できる範囲をはるかに超えている時に顕著となります。いわゆる、AI の最大の難問である『少数サンプルからの学習』とも深い関わりがあります。現在、多次元の問題を小数個のサンプルから上手く学習できる AI アルゴリズムはありません。

メタ認知[5]は、ヒトそれ自身の認知、意志決定、記憶などを監視する機能であると理解されています。そしてメタ認知は前頭前野に 2 次的な神経活動表現を作り出すと示唆されています。多くの著名な研究者が、メタ認知が次元の呪いに対する解決法の一部だと考えてきました。例えば、ディープラーニングの発展に最も大きく貢献した研究者の一人、Yoshua Bengio 教授は、意識に上る心的表現を持つ能力、あるいは少し拡張して、メタ認知の過程を保持することは、次世代汎用 AI の開発の鍵であると提案していました。超多次元の問題を解く時に、強化学習とメタ認知がどのように相互作用するのかを理解すれば、脳の機能を人工的に再現できる道が開けます。しかしながら現在までのところ、AI 開発の視点でのメタ認知の研究はされてきませんでした。そこで本研究では fMRI 非侵襲脳活動計測と機械学習の最新技術を用いて、実験参加者が、自身の意識下の脳活動パターンで決まる、報酬を最大化する行動を、試行錯誤で選択する強化学習課題を用意しました。脳がどの脳部位を使い、どのような神経計算論的アルゴリズムに基づいて次元の呪いを解くかを理解するのは、将来の AI 開発に重要と考えられます。

研究内容

●　実験方法

実験では１８人の実験参加者が２つの行動（A か B か）のうち１つを選択するための学習をしました。どちらの選択肢がより報酬を得る可能性が高いかは、参加者の意識下の脳活動パターンを機能的磁気共鳴画像法（fMRI）[1]で実時間計測し、機械学習のアルゴリズムで解読（デコーディング[8]）して、試行毎に決めました。参加者は、報酬を得るためには、自分自身の脳状態を読み取って適切な行動を選択しなくてはいけないということは知らされていませんでした。この実験アプローチは研究グループで、よく確立されたデコーディッドニューロフィードバック法[2,9]の新しい拡張になっています。

●　実験準備：デコーダー（脳情報解読器）の構成（セッション 0）

ゲーム（行動選択課題）で各試行の条件を決めるためには、まず機械学習でデコーディング[8]のための脳情報解読器（デコーダー）を構築しなければなりません。主なゲーム学習実験のおよそ１週間前に参加者は簡単な知覚判断とその確信度評定を行いました（図１）。知覚課題はランダムドットパターンの動きを眺めてその全体的な動きが右か左かを判定するものです。そして、その選択に関する確信度を報告します。質問は、『あなたの知覚判断が正しいかどれくらい自信がありますか？』というもので、答えは『すごく自信があります』『当てずっぽうで言っただけです』などになります。

デコーダーは脳活動パターンが右か左の動きのどちらを表しているかを二者択一で分類

するように学習を行いました。スパースロジスティック回帰アルゴリズム[10]によって、fMRI の数千次元のボクセルパターンから右か左の分類が可能になります。デコーダーは一旦構築されると、視覚刺激が提示されていない場合も含め、任意の脳活動パターンについて、それが右または左のどちらの運動情報を表現しているかを判定できます。

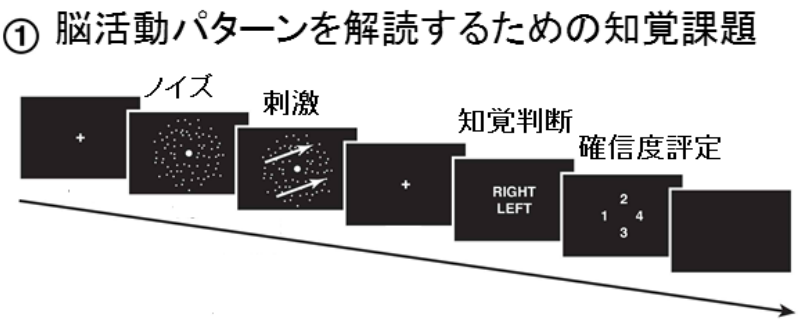


図１　視覚刺激（ランダムドットパターン）の動きの向きの知覚判断とその確信度評定

●　メインのゲーム学習実験：意識下の隠れた脳状態に関しての強化学習

主な行動選択課題では参加者が自身の脳状態に応じて決まる、報酬につながる最適行動を学習できるかを調べました。重要なのは参加者が試行錯誤だけを通して課題を学習することです。参加者はどのような規則で最適行動が決められているかは教えられていません。

実験は３日連続のセッションで行われました（図 2 参照）。最初の 2 日間、参加者がスクリーン上で見るのは、全くランダムで彼らの脳活動の情報を含まない刺激でした。参加者は試行錯誤だけで 2 つの選択肢のうち、どちらが最適かを学ばなければいけません。３日目のセッションのみ、視覚刺激はデコーダーが読み取った脳状態に応じた右か左かの動きの情報を提示します。

学習課題の流れを図 2 下部に示します。最初に参加者は 6 秒間真っ暗なスクリーンを見ます。この時の脳活動をデコーダーが右か左の動きに判別します。その後 8 秒間ランダムドットパターンの動きを観察します。ランダムドットパターンの動きは、第１、2 日のセッションでは、左右の動きを全く含まない完全にランダムな動きで脳の状態を全く反映していません。一方、第３日のセッションではデコーダーの出力で決まる動きで、左右どちらかの方向です。つまり、第１，第 2 日では脳は、自身の意識下の状態を読み取らない限り、正解はできません。一方、第 3 日ではランダムドットパターンの動きに応じて行動を決めれば良いだけなので、学習は簡単で、すぐ正解を見つけられます。

視覚刺激の後で、参加者はランダムドットパターンの全体的な動きが右か左かの知覚判断を報告し、その選択に関する確信度を報告します。第１日、第 2 日ではランダムドットパターンは動きを含んでいないので、自信があるはずは無いのですが、試行によっては高い確信度が報告されます。最後に参加者は図 2 に示すＡボタンかＢボタンかの１つの行動を選択します。最も大事な点は、この行動選択が強化学習のポイントであることです。行動に応じて３０円の報酬がもらえるか、０円になるかが決まります。行動が選択された後で、報酬の結果がスクリーン上に示されます。

この学習課題の難しさは次の点にあります。各試行毎に、デコーダーは真っ暗なスクリーンを見ている時の脳活動パターンを解析し、右か左のラベルを出力します。このラベルが、その試行で行動ＡかＢのどちらがより適切かを決めます。実験実施者だけが脳活動－デコーダー－最適な行動の関係を知っており、実験参加者はこのことを知らされていません。さらに、もちろん実験参加者は脳のどの部位のどのような情報で最適行動が決まっているかは知りません。とりわけ第１，第 2 日にはランダムドットパターンは動きの情報を含んでいません。従って、脳は強化学習を規定している意識下の状態を、1 千億ものニューロンによって非常に大きな次元で表現される大量の情報の中から探索しない限り、適切な行動は選べません。このような困難な状況の下、脳は一体どうやって、試行錯誤だけで、最適行動を決めている脳活動パターン（特定の fMRI ボクセル群）を見つけることができるのでしょうか。

②隠れた脳状態の強化学習課題:
機械学習（スパースロジスティック回帰）を用いた脳活動のリアルタイムデコーディング

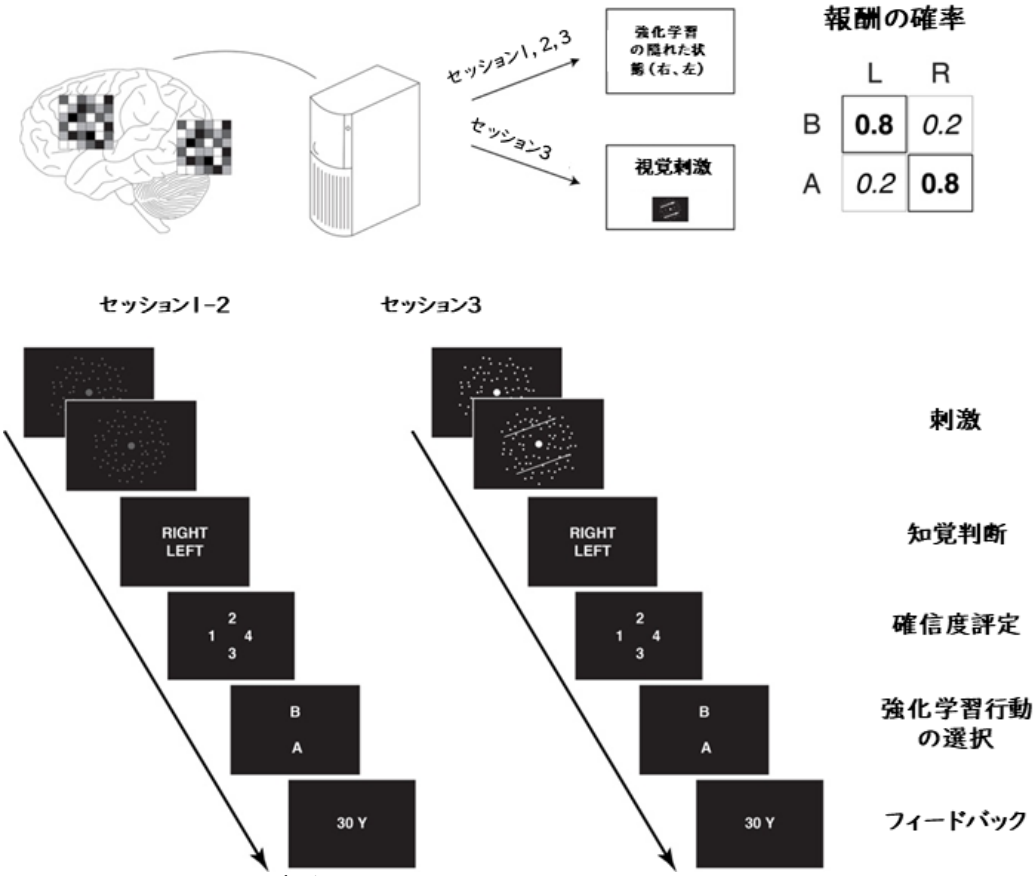


図２　主な実験のデザイン：意識下の脳状態で決まる最適行動を強化学習する

●　確信度の助けを借りて、報酬につながる最適行動の選択を学習できる

図３の左側に示すように、参加者は２日目のセッションからすでに偶然のレベル 50%以上に最適な行動を選択できています。つまり、視覚刺激が脳状態を反映せず全くランダムな 2 日目に、強化学習がすでに進んでいます。これは、脳が意識下の状態を発見し、行動選択に利用できたことを示しています。さらに、図３の真ん中からわかる通り、主実験の約１週間前にデコーダー構成セッション０で検査された参加者のメタ認知能力（自身の知覚判断成績について、確信度としてどれ程正確に評価できているか）から、本実験（セッション１，２）の強化学習の成績が予測できます。図３の右側には知覚課題での確信度と行動選択の成績が強く相関することを示しています。この結果は、メタ認知の１つ、確信度が意識下の強化学習と関係していることを初めて示したという意味で重要です。

③被験者は報酬行動を選択できるようになり、メタ認知は被験者の学習能力を予測している

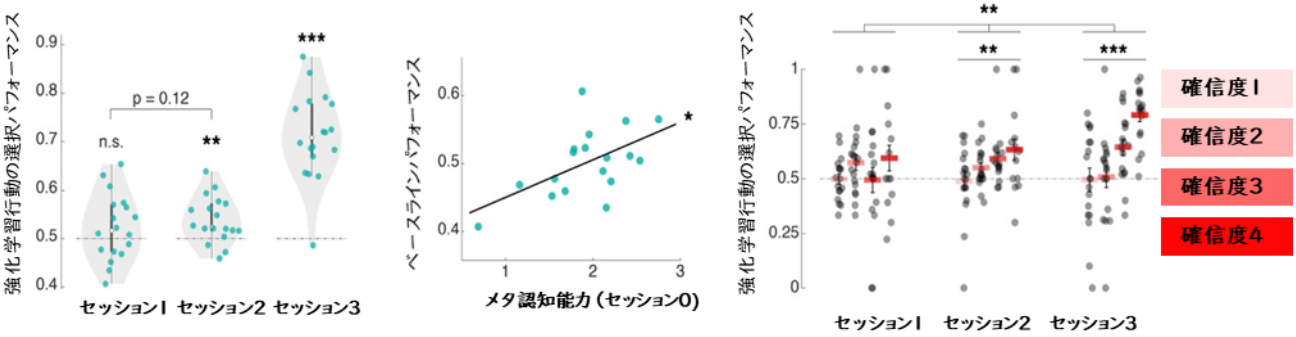


図３　強化学習課題の実験結果：(左)2 日目に最適行動がすでに選択できています。(中)メタ

認知能力が強化学習の成績を予測。セッション 0：デコーダーの構成のセッションです。(右)確信度が高いと最適行動をより選択。

次に、強化学習のアルゴリズムから参加者の行動を説明しました。簡単なアルゴリズムを通じて、実際の行動選択とその結果（報酬）に基づいて、状態と行動の組み合わせの報酬予測モデルが学習を通して更新されます[3]。下の図 4 では、神経科学の強化学習計算モデルを用いた解析で、確信度と意識下の強化学習の関係が表れることを示しています。確信度が大きい試行では、強化学習アルゴリズムの驚き信号（報酬予測誤差）が小さくなります。

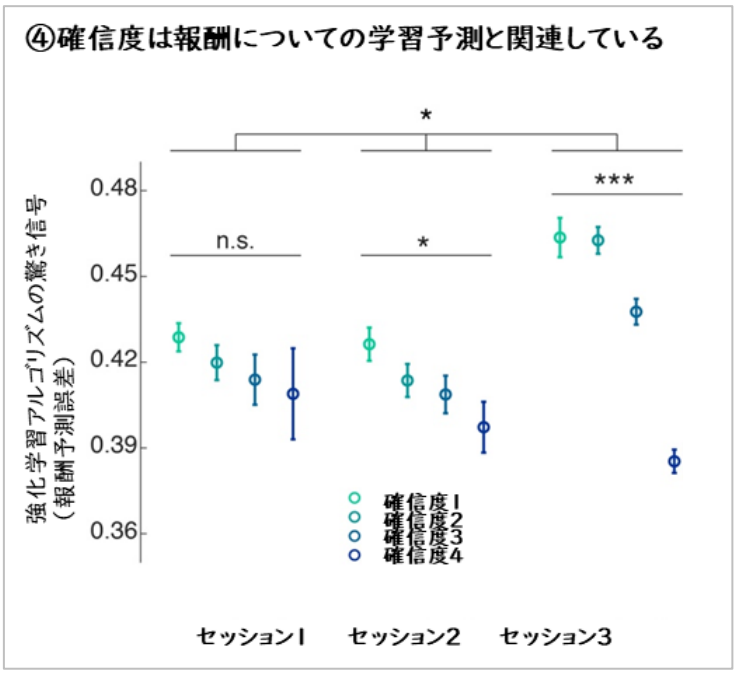


図 4　確信度と強化学習の進み具合が関連している

最後に、これらの行動レベル・計算論レベルでの確信度と強化学習の関連の発見は、神経活動レベルでも見ることができました（図 5）。確信度と強化学習を表現している 2 つの脳部位、DLPFC（背外側前頭前野）[6] と BG（大脳基底核）[4] は、学習が進むにつれ、活動の強さでも、表している情報の上でもより同期するようになります。つまり、この研究によって、脳は天文学的に大きな次元の難しい問題を解く能力を持っていて、メタ認知（ここでは参加者の知覚課題に関する確信度）が重要なことがわかりました。

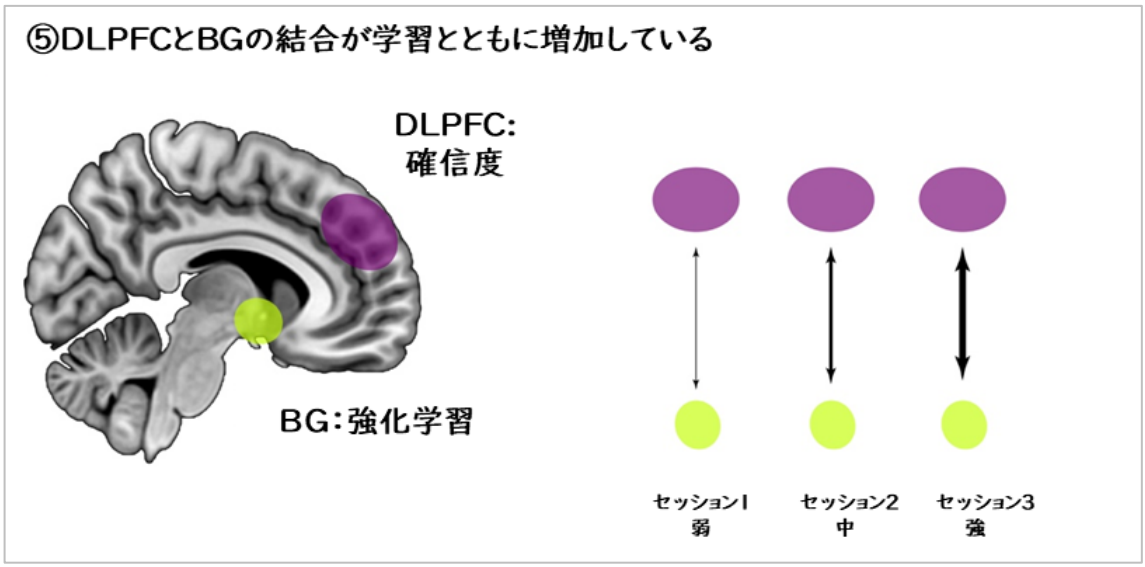


図 5　DLPFC（背外側前頭前野）と BG（大脳基底核）との結合が学習とともに強まる

本研究の意義と今後の展望

● 科学的意義

本研究ではデコーディッドニューロフィードバック[2,9]に基づく革新的な実験パラダイムを用いて、脳が超多次元の問題を少数サンプルから学習する能力があり、確信度がその計算論的な解法で重要な役割を果たしていることを示しました。

● 技術的な新規性

研究者が属するグループはデコーディッドニューロフィードバックを用いてこれまで、実験参加者の確信度を変化させたり、色の感覚を作り出したり、恐怖記憶反応を消去したりすることに成功してきました。今回の研究は従来研究の発展ですが、脳の状態を実時間で計測・解析して強化学習課題の最適行動を決めるという意味で全く新しいものです。学習課題の条件を単純な視覚刺激などで実験者が外的に決めるのではなく、脳の状態そのもので学習課題を規定できるからです。この実験パラダイムは神経科学や AI の新しいタイプの疑問に答えるための革新的な研究開発に有用であると期待されます。

● 今後の展望

この研究によってヒトはわずか数百回の試行錯誤のうちに、意識下の脳状態で規定されている恐ろしく複雑な問題を強化学習することができ、しかも確信度がそのような学習能力を裏打ちしていることが示されました。この結果は、将来、脳科学と AI を融合させることにつながり、より高い性能の AI を開発するために有益であると考えています。たとえば、メタ認知あるいは自己モニターのモジュールが強化学習ユニットと相互作用して行動し学習する AI は、よりヒトに近い問題解決が可能になると期待されます。また純粋な神経科学の観点からは、意識下の状態の強化学習の成功が確信度と関連していることから、確信度と意識的な体験が乖離できることを示しました。これはメタ認知に関する新しい考え方につながり、学習や意識の新しい研究パラダイムを切り拓くと期待できます。

● 倫理面での懸念に対する対応

脳内のメカニズムを用いた、新しいテクノロジーや実験デザインを利用・開発することは、新しい知識と応用によって全く新しい事態が起こりうるため、特に神経科学と AI の境界領域で特別な倫理的な配慮が必要です。たとえば、ヒトがどのようにして非常に複雑な問題を少数のサンプルから学習するかを理解すること自体は無害ですが、それが新しい AI に翻訳され、ヒトに匹敵するレベルになるとさまざまな問題が生じ得ます。また、ヒトがどのように意識下で学習するかを理解することも重要なポイントでしたが、一方でこの実験プロトコルを洗脳に近い、つまり本人が意図していないのに、ある行動を取るように訓練されている実験結果だと考える人もいるかもしれません。研究グループはこれらの倫理的問題を慎重に検討するとともに、倫理あるいは生命倫理の専門家も含めて、倫理安全委員会で研究を評価しています。

論文著者名とタイトル

Nature Communications 誌（英国時間　2020 年 08 月 31 日公開）
Aurelio Cortese, Hakwan Lau, Mitsuo Kawato: Unconscious reinforcement learning of hidden brain states supported by confidence. *Nature Communications.*
https://doi.org/10.1038/s41467-020-17828-8

研究グループ

株式会社国際電気通信基礎技術研究所（ATR）
Aurelio Cortese、川人 光男、
カリフォルニア大学ロサンゼルス校
Hakwan Lau

研究支援

本研究は、科学技術振興機構（JST） ERATO「池谷脳-AI ハイブリッドプロジェクト」（JPMJER1801）（研究総括 池谷裕二）の一環として行われたものです。一部は、日本医療研究開発機構（AMED）戦略的国際脳科学研究推進プログラムの「脳科学と AI 技術に基づく精神神経疾患の診断と治療技術開発とその応用」課題 JP18dm0307008（代表 川人光男）の支援を受けています。また Hakwan Lau 教授は、米国 National Institute of Health（NIH）R01NS088628 からの支援も部分的に受けています。

お問い合わせ先

＜研究内容に関すること＞
（株）国際電気通信基礎技術研究所（ATR）経営統括部 企画・広報チーム
〒619-0288 京都府相楽郡精華町光台 2-2-2
Tel: 0774-95-1176, Fax: 0774-95-1178, E-mail: pr[at]atr.jp
https://www.atr.jp/

＜JST 事業に関すること＞
科学技術振興機構 研究プロジェクト推進部 グリーンイノベーショングループ
内田 信裕
〒102-0076 東京都千代田区五番町 7 K’s 五番町
Tel: 03-3512-3528, Fax: 03-3222-2068, E-mail: eratowww[at]jst.go.jp

補足説明

[1] 機能的磁気共鳴画像法（functional Magnetic Resonance Imaging, fMRI）

酸化型と還元型ヘモグロビンの磁化率の違いを利用して、粗く言えば、脳全体の血流量の変化を画像化する技術です。酸化型と還元型ヘモグロビンの量の違いは脳活動の度合いを反映しているため、この画像を解析することで、各脳部位の活動度合いを推定することができます。

[2]（実時間）ニューロフィードバック

ニューロフィードバックは、脳の状態をモニタリングしながら、特定の脳の状態を誘導する方法です。一般的なニューロフィードバックでは、対象とする脳領域の活動度を上げたり下げたりします。実時間ニューロフィードバックでは、計測した脳活動をリアルタイムで解析し、実験参加者に解析結果を即座にフィードバックとして知らせます。

[3] 強化学習(RL, Reinforcement Learning)

強化学習は、AI・ロボティクス・神経科学に大きな影響を与えている、経験から学習する強力なアルゴリズムの 1 つです。学習者である「エージェント」がある設定された環境の中で、環境から与えられる報酬が一番多くなるような行動を獲得するための機械学習法です。強化学習ではエージェントが環境と相互作用して行動を学習する様相を特徴付けます。エージェントは、自分の取った行動が良かったか悪かったかを、その行動を取った際に得られた報酬に基づいて「価値」として計算し、価値が高い行動を高い頻度で選ぶように学習を行います。経験を通して、エージェントは世界のある状態に関する信念（報酬に関わる）を獲得し、最も適切な行動を選べるように学習します。強化学習は試行錯誤によって学習を行う方法で、ヒトや動物の学習方法と類似すると考えられています。

[4] 大脳基底核 BG(basal ganglia)

大脳基底核と呼ばれる部分で、大脳皮質と視床・脳幹を結びつけている神経核の集まりです。線条体・淡蒼球・黒質・視床下核からなります。役割として、運動調節・認知機能・感情・動機づけや学習などさまざまな機能を司っています。

[5] メタ認知(Metacognition)

メタ認知は自分自身の心的な能力や認知過程を監視する能力です。メタ認知はまた、私たちの認知、さまざまな行動決定、記憶、そして思考でさえも、その現実性あるいは信頼度を監視する機能でもあります。自己モニタリングも含まれます。視覚認知の確信度判定も、メタ認知の一種です。さらにメタ認知は意識と深く関係しています。

[6] 背外側前頭前野 DLPFC(Dorsolateral prefrontal cortex)

背外側前頭前野と呼ばれる部分で、意思決定を行うための脳部位です。合理的に物事を考え、情動を抑制するなどの機能を司っています。

[7] 次元の呪い(Curse of dimensionality)

機械学習、特に強化学習では、次元の呪いは状態空間の次元、あるいは特徴の次元が非常に大きい時に生じる問題と定義される。そのような状況では、問題は天文学的な複雑さを示し、学習に必要とされるサンプル数（学習回数、データなど）は次元の指数関数的に増大する。強化学習では、このような状況では、エージェントは最適行動を合理的な時間内に探索

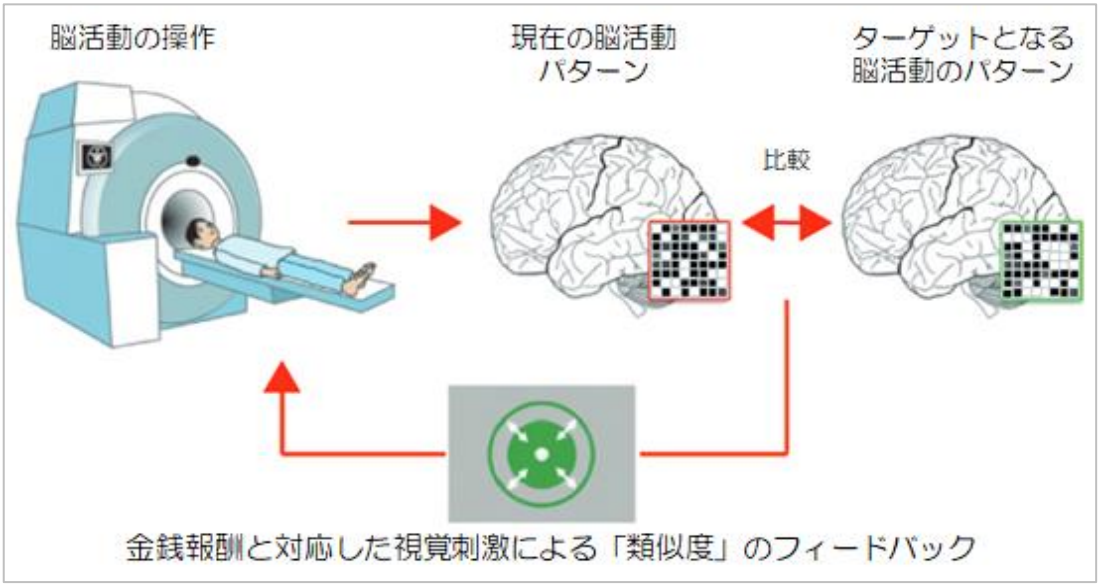
することは実質的に不可能となる。

[8] デコーディング

空間的な脳活動パターンを解読（デコード）し、人の認知や知覚の状態を推定することを一般的にデコーディングと呼びます。計測した fMRI データは、ボクセルと呼ばれるたくさんのデータ点を含みます。それぞれのデータ点が、数ミリメートルという脳内のごく小さなエリアの活動量を指し示しています。空間的脳活動パターンとは、多数ある fMRI データ点のうち、どのデータ点（エリア）が大きな値（脳活動を示す信号）を持ち、どのデータ点が小さな値を持つのか、という空間的な情報を指します。空間的脳活動パターンを解読するためにはさまざまな AI アルゴリズムが用いられます。

[9] デコーディッドニューロフィードバック （DecNef）

fMRI（[5]を参照）と AI 技術を組み合わせ、対象とする脳領域に特定の活動パターンを誘導する方法です。従来のニューロフィードバックと異なる点として、脳領域全体の活動度ではなく、脳領域を細かく分解して「この部分は活動が上がっているけれどもここは下がっている」、というように、脳領域内の活動パターンを誘導する点です。著者らによる ATR で実施された先行研究（Shibata et al., Science 2011; , Cortese et al. Nature Communications 2016）において、世界に先駆けて開発されました。



[10] スパースロジスティック回帰アルゴリズム

ATR で開発された AI 技術の 1 つ（Yamashita et al., NeuroImage, 2008）。計測した fMRI データは、ボクセルと呼ばれる非常にたくさんのデータ点を含みます。しかし、すべてのボクセルが参加者の認知状態についての情報を持っているわけではありません。fMRI データを用いて参加者の認知状態を精度よく推定するためには、この推定にかかわるボクセルのみをうまく選別する必要があります。スパースアルゴリズムを用いることによって、自動的かつ効率的にボクセルを選別することが可能になります。