title: Cognitive computational neuroscience author: Nikolaus Kriegeskorte and Pamela K. Douglas journal: Nature Neuroscience Volume: 1148 number: 21 year: 2018 pages: 1148–1160

認知計算論的神経科学

Cognitive computational neuroscience

Nikolaus Kriegeskorte and Pamela K. Douglas

要旨

認知が脳内でどのように実現されているかを知るためには，認知タスクを実行できる計算モデルを構築し，脳や行動の実験で検証する必要があります。 認知科学では，認知を機能的な要素に分解した計算モデルを開発してきました。 計算論的神経科学では，相互作用するニューロンがどのように認知の要素を実現するかをモデル化しています。 今こそ，脳の計算というパズルのピースを組み立て，これらの別々の分野をよりよく統合する時です。 現代の技術では，動物や人間の脳活動を，これまでにないほど豊富な方法で測定・操作することができます。 しかし，実験によって理論的な洞察が得られるのは，脳計算モデルの検証に用いられた場合のみである。 ここでは，認知科学，計算論的神経科学，人工知能の分野における最近の研究を紹介します。 知覚・認知・制御タスクにおける脳の情報処理を模倣した計算モデルが開発され，脳や行動のデータを用いて検証されるようになってきた。

脳の情報処理を理解するためには，認知タスクを実行できる計算モデルを構築する必要があります。 タスクを実行できる計算モデルを支持する論拠は，1973 年にアレン・ニューウェルが発表した解説 “You can’t play 20 questions with nature and win” によく現れています。 ニューウェルは，認知心理学の現状を批判していました。 認知心理学の分野では，一連の二項対立の質問に自然に答えさせれば，やがて脳のアルゴリズムが明らかになると期待して，認知に関する1つの仮説を一度に検証する習慣があります。 ニューウェルは，言葉で定義された認知に関する仮説を検証しても，計算機的な理解にはつながらないと主張しました。 仮説を検証するためには，タスクを実行する包括的な計算モデルを構築することで補完する必要があると考えていたのです。 認知機能を説明できるかどうか，提案された構成要素のメカニズムが実際にどのような相互作用をしているのかは，コンピュータ・シミュレーションによる合成によってのみ明らかになります。 もし，情報処理メカニズムを完全に理解しているのであれば，それをエンジニアリングすることができるはずです。 1988 年に亡くなった物理学者のリチャード・ファインマンは，「作れないものは分からない」 という言葉を黒板に書き残しています。

ここでは，認知が神経生物学的に妥当な動的要素からどのようにして生じるのかを説明するタスク実行型の計算モデルが，新しい認知計算神経科学の中心となることを論じる。 まず，認知科学と脳科学の歩みを簡単にたどり，次に，神経生物学的に妥当な人工知能 (AI) モデルを用いて，認知科学 (人間がどのように学び，考えるのかを説明する) と計算神経科学 (脳がどのように適応し，計算するのかを説明する) の両方の野心を満たすことができるかもしれないことを示唆する，いくつかのエキサイティングな最近の進展をレビューする。

ニューウェルの批判の精神に基づき，認知心理学から認知科学への移行は，タスクを実行する計算モデルの導入によって定義されました。 認知科学者は，認知を理解するためには AI が必要であることを知り，認知研究にエンジニアリングを導入しました。 1980 年代，認知科学は記号的認知アーキテクチャ[4,5] とニューラルネットワークによって重要な進歩を遂げ，人間の行動データを使って計算モデルの候補を判断しました。 しかし，コンピュータのハードウェアと機械学習は，認知プロセスの複雑さを完全にシミュレートするには十分ではありませんでした。 さらに，これらの初期の開発では，行動データだけに頼っていたため，脳の解剖学的構造や活動から得られる制約を利用できませんでした。

ヒトの脳機能イメージングが可能になったことで，認知理論をヒトの脳に関連付けることができるようになりました。 この試みは，認知神経科学と呼ばれるようになっています。 認知神経科学者は，認知心理学の箱 (情報処理モジュール) と矢印 (モジュール間の相互作用) を脳にマッピングすることから始めました。 これは，脳の活動を取り込むという点では一歩前進したが，計算の厳密さという点では一歩後退した。 認知科学の課題遂行型の計算モデルを脳活動データで検証する方法は考えられていなかった。 その結果，認知科学と認知神経科学は 1990 年代に決別することになった。

認知心理学が提唱する高レベルの機能モジュールに関する課題や理論は，脳波計，ポジトロン・エミッション・トモグラフィー，初期の機能的磁気共鳴画像法 (fMRI) など，空間分解能の低い機能イメージング技術を用いて，人間の脳の粗いスケールの組織をマッピングするための合理的な出発点となりました。 認知心理学の「モジュール」という概念に触発された認知神経科学は，自然との 20 の質問という独自のゲームを開発しました。 ある研究では，脳の中に特定の認知モジュールが存在するかどうかが問われる。 この分野では，増え続ける認知機能を脳領域にマッピングし，人間の脳の全体的な機能レイアウトの有用なラフドラフトを提供しました。

どのようなスケールの脳地図であっても，計算メカニズムを明らかにするものではありません（図1）。 しかし，マッピングは，理論に制約を与えます。 結局のところ，情報交換には，通信する領域間の距離に応じて，物理的な接続，エネルギー，信号の待ち時間などのコストがかかります。 部品の配置には，こうしたコストが反映されていると考えられます。 高い帯域幅と短いレイテンシーで相互作用する必要のある地域は，近くに配置されることが予想されます。 より一般的には，生物学的な神経ネットワークのトポロジーとジオメトリーは，そのダイナミクス，ひいては機能的なメカニズムを制約する。 したがって，機能的な局在化の結果は，特に解剖学的な結合性と組み合わせることで，最終的には脳の情報処理のモデル化に役立つと考えられる。

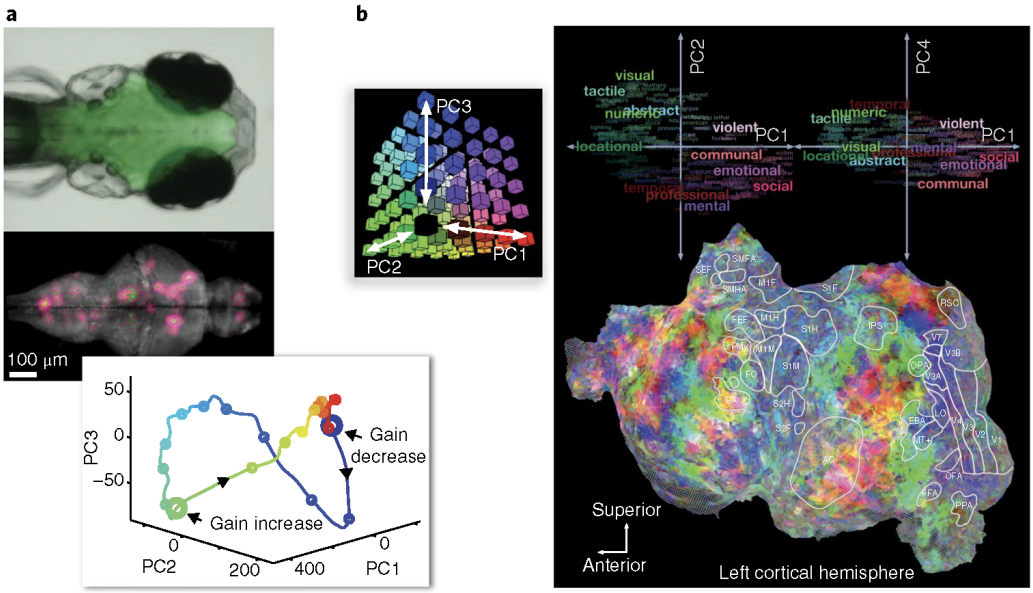


図 1. 最新の脳画像技術は，脳活動に関するかつてないほど詳細な情報を提供する。だが，データに基づく解析では限られた知見しか得られない。

a) 二光子カルシウムイメージング (121) の結果は，ゼブラフィッシュの幼生が仮想環境と相互作用している間に，同時に測定された多数の細胞集団の単一ニューロンの活動を示している。 b) ヒトの fMRI の結果 (70) は，被験者が物語を聞いている間，意味的に選択的な反応の詳細な地図を明らかにした。 これらの研究は，異なるスケールでの最新の脳活動計測技術の威力を示す一方で(a,b)，このようなデータセットから脳計算に関する洞察を引き出すことの難しさも示している。 両研究とも，複雑で時間の連続する自然主義的な体験中の脳活動を測定し，主成分分析 (a, bottom; b, top) を用いて，活動パターンの全体像とその表現上の重要性を明らかにしたものである。PC: 主成分分析

方法論上の課題はあるものの，認知神経科学で得られた知見の多くは，その基礎となる確かなものです。 例えば，ヒトの腹側視覚野に顔選択領域が存在するという知見は，徹底的に再現され，一般化されています。 ヒト以外の霊長類でも，fMRIを用いた実験で同様の顔選択領域が確認されています。これらの領域は，侵襲的な電極では広い視野で連続的な画像を得ることができないため，これまで調査されていませんでした。 霊長類の顔パッチは，fMRIで局在化され，侵襲的な電極記録で検証された結果，顔選択ニューロンの密度が高く，鏡面対称のチューニングや最前部パッチでの個々の顔の耐視的表現など，階層的な処理の高い段階で不変性が現れることが明らかになった。 顔認識の例は，解剖学的基盤のマッピングとニューロン応答の特性化が確実に進んでいることを示す一方で，決定的な計算モデルがないことを示している。 一方で，決定的な計算モデルがないことも示している。 顔認識の脳-計算モデルは，顔選択ユニットの空間的なクラスターや，fMRIや侵襲的な記録で観察される選択性や不変性を説明しなければならないでしょう。

認知神経科学は，人間や霊長類の脳の全体的な機能配置を明らかにしてきた。 しかし，脳の情報処理に関する完全な計算論的説明は達成されていない。 今後の課題は，脳の構造と機能に合致し，複雑な認知タスクを実行する脳情報処理の計算モデルを構築することである。 認知科学，計算論的神経科学，人工知能の分野における以下のような最近の進展は，これが達成可能であることを示唆している。

1. 認知科学は，複雑な認知プロセスを計算上の構成要素に分解することで，トップダウンで進められてきた。 脳のデータを理解する必要性にとらわれることなく，認知レベルでタスクパフォーマンスの高い計算モデルを開発してきた。 その成功例の一つが，ベイズ認知モデルである。ベイズ認知モデルは，世界に関する事前知識と感覚的な証拠を最適に組み合わせることができる。 当初は基本的な感覚や運動のプロセスに適用されていたベイズモデルは，物理的・社会的な世界をモデル化する方法など，複雑な認知にも適用されるようになった。 これらの開発は，統計学や機械学習との相互作用の中で行われ，確率的な経験的推論に対する統一的な視点が生まれました。 この文献は，脳を理解するために不可欠な計算理論を提供しています。 さらに，現実の知能に必要とされるような，利用可能なデータに応じて複雑さを増すことができる生成モデルの近似的な推論のアルゴリズムも提供しています。

2. 計算論的神経科学は，生物学的なニューロン間の動的な相互作用が，いかにして計算コンポーネントの機能を実現するかを示す，ボトムアップ的なアプローチをとってきました。 この分野では，過去20年間に，初歩的な計算コンポーネントの数学的モデルを開発し，それを生体ニューロンに実装してきました。 この分野では，感覚コーディング[29,30]，正規化[31]，作業記憶[32]，証拠の蓄積と決定メカニズム[33-35]，運動制御[36]などのコンポーネントがあります。 これらのコンポーネント機能の多くは，計算上単純なものですが，認知のビルディングブロックを提供しています。 また，計算論的神経科学では，高レベルの感覚や認知の脳内表現を説明できる複雑な計算モデルの検証が始まっています[37,38]。

3. 人工知能は，構成要素の機能を組み合わせて知的な行動を作り出す方法を示した。 初期の人工知能がその期待に応えられなかったのは，知能を発揮するために必要な豊かな世界の知識を工学的に作ることも，自動的に学習することもできなかったからである。 最近の機械学習の進歩は，計算能力の向上と学習対象となるデータセットの増加によって後押しされ，知覚[39]，認知[40]，制御の課題[41]において進歩をもたらした。 多くの進歩は，認知レベルのシンボリックモデルによってもたらされました。 最近の最も重要な進歩のいくつかは，入力の線形結合を計算するユニットで構成され，静的な非線形性を伴う深層ニューラルネットワークモデルによって推進されています[42]。 これらのモデルは，活動電位などの基本的な機能を抽象化して，生物学的なニューロンの動的機能のごく一部しか使用していません。 しかし，これらの機能は，脳からヒントを得ており，生物学的ニューロンでも実装可能である。

この 3 つの分野は，認知課題を実行し，脳の情報処理や行動を説明する生物学的に妥当な計算モデルに，補完的な要素を提供している（図2）。 ここでは，認知科学（認知課題を実行し，行動を説明する計算モデル）と計算論的神経科学（脳活動を説明する神経生物学的に妥当な機構モデル）の成功基準を合わせて満たす認知計算論的神経科学に向けた文献の最初の一歩をレビューする。 計算モデルが動物や人間の認知を説明するためには，知能を発揮しなければなりません。 そのため，AI，特に機械学習は，認知計算論的神経科学の理論的・技術的基盤となる重要な学問です。

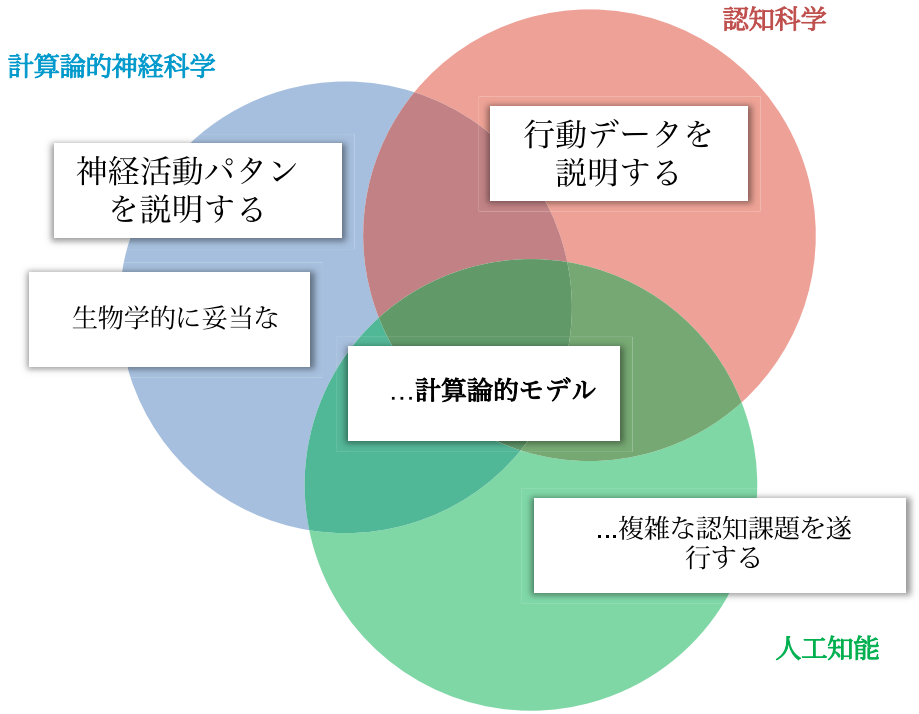


図 2. 脳がどのようにどうさするのかを理解することは何を意味するのか？ 認知計算論的神経科学の目標は，動物や人間の神経細胞の活動や行動に関する豊富なデータを，実世界の認知課題を実行する生物学的に妥当な計算モデルによって説明すること。 歴史的には、各分野（円）は，これらの課題（白ラベル）のサブセットに取り組んできた。 認知計算論的神経科学は，これらすべての課題に同時に取り組むことを目指している。

そのためには，理論（タスクを実行する計算モデルに具現化される）と実験（脳や行動のデータを提供する）の間にしっかりとした橋を架けることが重要な課題となります。 このレビューの最初の部分では，実験データから始まり，データから理論の方向に橋を架けようとするボトムアップ型の開発について説明します[43]。 脳活動データが与えられた場合，接続性モデルは，脳の活性化の大規模なダイナミクスを明らかにすることを目的とし，デコーディングおよびエンコーディングモデルは，脳の表現の内容と形式を明らかにすることを目的としています。 この文献で採用されているモデルは，計算理論に制約を与えるものですが，一般的には問題となっている認知タスクを実行するものではないため，タスク実行の根底にある計算メカニズムを説明するには至りません。

この論文の第 2 部では，理論から実験への橋渡しをするという，逆方向の展開を紹介します[37,38,44]. この記事では，タスクを実行する計算モデルを脳や行動のデータで検証し始めた新しい研究を紹介します。 これらのモデルには，抽象的な計算レベルで指定され，生物学的な脳への実装がまだ説明されていない認知モデルや，神経生物学の多くの特徴を抽象化しているが，生物学的なニューロンを使って実装することがもっともらしいニューラルネットワークモデルが含まれる。 これらの文献は，脳の計算を理解するための統合的なアプローチの始まりを示唆しています。すなわち，認知タスクを実行するためのモデルが必要であり，生物学は許容できる構成要素の機能を提供し，計算メカニズムは脳の活動や行動の詳細なパターンを説明するために最適化されるのです。

1. 実験から理論へ

1.1. 接続性とダイナミクスのモデル

脳の活動を測定して計算で理解する方法の一つとして，脳の連結性とダイナミクスをモデル化する方法があります。 接続性モデルは，活性化された領域の局在を超えて，領域間の相互作用を特徴づけるものです。 ニューロンのダイナミクスは，相互作用するニューロンの局所的なセットから脳全体の活動まで，複数のスケールで測定・モデル化することができます[45]. 脳のダイナミクスの第一近似値は，測定された応答時系列間の相関行列によって得られ，これは場所間のペアによる「機能的連結性」を特徴づける. また，空間独立成分分析のような時空間行列の線形分解でも，同様に時間を超えた場所間の同時相関を捉えることができます[47].

相関行列を閾値化することで，地域の集合は無向グラフに変換され，グラフ理論的な方法で研究することができます。 このような分析により，「コミュニティ」（強く相互接続された領域の集合），「ハブ」（多くの他の領域に接続された領域），「リッチクラブ」（ハブのコミュニティ）を明らかにすることができる[48]。 接続性グラフは，解剖学的または機能的な測定値から導き出すことができます。 リージョンは解剖学的経路を介して相互作用するため，解剖学的連結性マトリックスは一般的に機能的連結性マトリックスと似ている。 しかし，解剖学的結合性が機能的結合性を生み出す方法は，局所的なダイナミクス，遅延，間接的な相互作用，ノイズを考慮に入れることで，よりよくモデル化されます[49]。 局所的なニューロンの相互作用から，大脳皮質や皮質下の領域にまたがる大規模な時空間パターンまで，自発的なダイナミクスの生成モデルは，脳活動のデータを用いて評価することができます。

有効な接続性分析は，より仮説主導型のアプローチをとり，ダイナミクスの生成モデルに基づいて小さな領域のセット間の相互作用を特徴づける[50]。 活性化マッピングが認知心理学の「箱」を脳領域にマッピングするのに対し，有効連結性分析は「矢印」を脳領域のペアにマッピングする．この分野のほとんどの研究は，脳領域の全体的な活性化のレベルで相互作用を特徴づけることに焦点を当てています。 古典的なブレインマッピング手法と同様に，これらの分析は領域平均活性化に基づいており，領域間で交換される情報ではなく，領域全体の活性化の相関的な変動を測定している。領域間で交換される情報ではなく，領域全体の活性化の相関的な変動を測定する。

効果的な接続性や大規模な脳のダイナミクスの分析は，活性化や情報ベースのブレインマッピングで使用される線形モデルのような一般的な統計モデルを超えて，生成モデルであるという点で，測定値のレベルでデータを生成することができ，脳のダイナミクスのモデルとなります。 しかし，これらのモデルは，表現された情報とそれが脳内でどのように処理されるかを捉えることはできません。

1.2. 復号化モデル

脳の計算メカニズムを理解するもう一つの方法は，脳の各領域にどのような情報が存在するかを明らかにすることである。 デコーディングは，あるタスクに対するある領域の関与を示す活性化という概念を超えて，ある領域の集団活動に存在する情報を明らかにするのに役立ちます。 ある脳領域の活動から特定の内容がデコード可能であれば，それは情報の存在を意味する。 脳領域がコンテンツを「表現している」という言い方をすると，情報が，その信号を受け取る領域にコンテンツを知らせる目的があるという機能的な解釈が加わることになる[51]。 最終的には，この解釈は，情報が他の領域や行動にどのような影響を与えるかについてのさらなる分析によって実証される必要があります[52-54]。

デコーディングは，神経細胞の記録に関する文献[27]にそのルーツがありますが，ニューロイメージング[55-59]では，表現の内容を研究するための一般的なツールとなっています。 最も単純なケースでは，デコーディングによって，2つの刺激のうちどちらが測定された反応パターンを生じさせたかが明らかになります． 表現の内容は，感覚刺激の同一性（代替刺激のセットの中で認識されるべきもの），刺激の特性（格子の向きなど），認知的な操作に必要な抽象的な変数，または行動である[60]。 通常のようにデコーダが線形である場合，デコード可能な情報は，下流のニューロンが 1 ステップで読み出すことが可能な形式になっている。 このような情報は，活動パターンに「明示的」に含まれていると言われる[61]。

囲み記事 1: 「モデル」の多様な意味

モデルという言葉は 脳科学や行動科学において様々な意味を持っています。 データ解析モデルは，測定された変数間の関係を確立するための一般的な統計モデルです。 例えば，線形相関，ブレインマッピングのための一変量重回帰，線形デコーディング分析などがある。 効果的な接続性モデルや因果関係モデルも，同様にデータ分析モデルです。 これらのモデルは，脳領域間の因果関係や相互作用を推測するのに役立ちます。 データ解析モデルは，変数間の関係（例えば，相関，情報，因果関係）に関する仮説を検証する目的で使用することができます。 脳の情報処理のモデルではありません。 一方，箱と矢印のモデルは，認知要素の機能を表すラベル付きの箱と，情報の流れを表す矢印で構成された情報処理モデルである。 認知心理学の分野では，脳の計算理論を解明するために，定義されていないにもかかわらず，このようなモデルが有用であった。 単語モデルも同様に，脳の情報処理に関する理論を，言葉で曖昧に定義したスケッチである。 これらは情報処理のモデルではあるが，脳で行われていると考えられている情報処理を実行するものではない。 オラクルモデルとは，脳の反応をモデル化したもの（データ解析モデルにインスタンス化されていることが多い）で，脳をモデル化している動物が入手できない情報に依存しているものである。 例えば，腹側頭の視覚反応を，抽象的な形状の記述の関数として，あるいはカテゴリーラベルや連続的な意味的特徴の関数としてモデル化した場合，そのモデルが画像から形状，カテゴリー，意味的特徴を計算することができなければ，神託モデルとなります。 オラクルモデルは，ある領域に存在する情報とその表現形式の有用な特性を提供することができるが，その表現が脳によってどのように計算されるかについては，いかなる理論も規定しない。 一方，脳計算モデル（BCM）は，あるタスクを実行する際の脳の情報処理を，ある程度の抽象度で模倣したモデルである． 例えば，視覚神経科学の分野では，画像のビットマップを入力として，脳の活動や行動を予測する視覚処理のBCMが，画像計算可能モデルと呼ばれています． 深層ニューラルネットは，画像計算可能な視覚処理のモデルを提供します。 しかし，スーパービジョンによって学習されたディープニューラルネットは，学習のためにカテゴリーラベル付きの画像に依存しています． 生物の発達や学習の過程では，ラベル付きの例は（同等の量で）入手できないため，これらのモデルは視覚処理のBCMではあるが，発達や学習のBCMではない。 強化学習モデルは，より現実的な質の環境フィードバックを用いるので，学習過程のBCMとなりうる． 感覚符号化モデルは，感覚入力をある段階の内部表現に変換する計算のBCMである． 内部変換モデルは，2つの処理段階の間で表現が変換されることのBCMである． 行動デコーディングモデルとは，ある内部表現から行動出力への変換のBCMである． BCM というラベルは，単にそのモデルがある程度の抽象度で脳の計算を捉えることを意図していることを示していることに注意してください． BCM は，生物学的な詳細から任意の程度まで抽象化することができますが，脳の活動や行動の何らかの側面を予測しなければなりません． 感覚入力から行動出力を予測する心理物理学モデルや，認知タスクを実行する認知モデルは，高いレベルの記述で定式化されたBCMです． BCM というラベルは，そのモデルがもっともらしいとか，経験的なデータと一致するとかいうことを意味するものではありません． BCM の候補を経験的に否定することで進歩するのです． BCM は，脳の計算を支える生物学的プロセスを表現するミクロスケールの生物物理モデルや，マクロスケールのブレインダイナミカルモデルや因果インタラクションモデルと同様に，脳で発生するプロセスのモデルである。 しかし，他のプロセスモデルとは異なり，BCM は脳のダイナミクスの機能と考えられている情報処理を行います． 最後に，モデルベース強化学習やモデルベース認知のように，脳が採用する世界のモデルを指す場合には，「モデル」という言葉が使われます。

1.3. 表象モデル

復号化 (デコーディング) だけでなく，私たちはある領域の表現を徹底的に特徴づけ，任意の刺激に対する反応を説明したいと考えています。 完全な特徴付けは，どのような変数がどの程度までデコーディングできるかを定義することにもなります。 表現モデルは，表現空間について包括的な予測を行おうとするもので，デコーディングモデルよりも計算メカニズムに強い制約を与えます[52,62]。

表現モデル分析には，符号化モデル[63-65]，パターン成分モデル[66]，表現類似性分析[57,67,68] の 3 種類が文献に紹介されている。 これら 3 つの方法はいずれも，実験条件の多変量記述-例えば，刺激のセットの意味的な記述や，刺激を処理するニューラルネットワークモデルの層間の活動パターン-に基づいて，表現空間に関する仮説を検証するものである[52]。

符号化モデルでは，刺激に対する各ボクセルの活動プロファイルは，モデルの特徴の線形結合として予測されます。 パターン成分モデルでは，表現空間を特徴づける活動プロファイルの分布を多変量正規分布としてモデル化する。 表象類似度解析では，刺激によって誘発される活動パターンの表現上の非類似性によって表現空間を特徴づける。

表象モデルは，人間の観察者が提供するラベルのような 刺激の説明に基づいて定義されることが多い[63,69,70]。 このシナリオでは，ある領域における脳の反応を説明する表象モデルは，脳の計算を説明するものではなく，少なくとも表象の記述的説明を提供するものである。 このような説明は，モデルが新しい刺激に一般化するときに，計算理論への有効な足がかりとなります。 重要なことは，表象モデルは，脳-計算モデル間の判断を可能にするということです。

本節では，脳活動データから計算上の知見を得るのに役立つ 3 種類のモデルについて考えた。 **接続性モデル** は 領域間の動的な相互作用の側面を捉えます。 **デコーディングモデル** は 脳の領域を調べて，その表現内容を明らかにすることができます。 **表象モデル** は ある領域の表現空間を完全に特徴づける明確な仮説を検証することができる。 これら 3 種類のモデルはいずれも，理論的に動機づけられた疑問を解決するために用いることができます。 しかし，課題を実行できる計算モデルがない場合，一連の質問をしても，説明しようとしている認知機能の根底にある計算メカニズムを明らかにすることはできないという Newell の議論に従うことになる。 これらの方法は，認知機能の基礎となる情報処理がどのように機能しているかを具体的に示すメカニズムモデルを検証しないため，理論への橋渡しにはなりません。

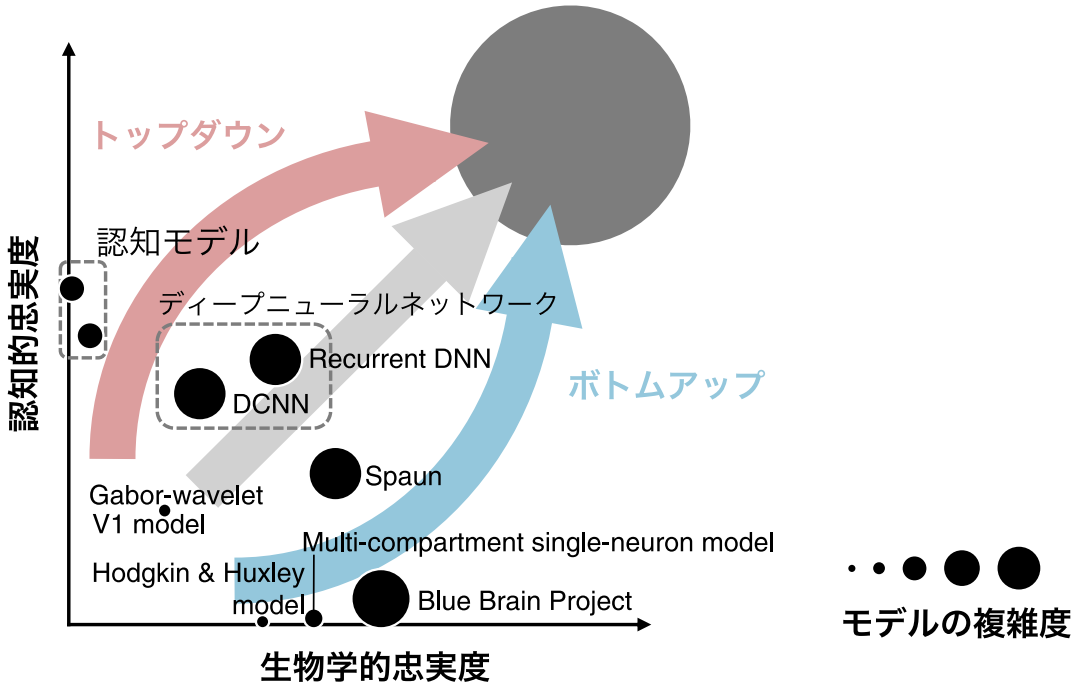


図 3. 脳内で起こるプロセスのモデルは，さまざまなレベルの記述で定義することができ，その計量可能な複雑さ（図中の点の大きさ）や，生物学的（横軸）および認知的（縦軸）な忠実度もさまざまです。 理論家は，さまざまな主要目標を掲げてモデリングに取り組んでいます。 ボトムアップ型のモデリング（青矢印）は，まず，活動電位や単一ニューロンの複数の要素間の相互作用など，生物学的な神経ネットワークの特性を把握することを目的としています。 このアプローチでは，認知機能を無視して，皮質の柱や領域などの脳の小さな部分の創発的なダイナミクスを理解し，振動などの生物学的なネットワーク現象を再現することに集中します。 振動などの生物学的ネットワーク現象を再現する。 トップダウンアプローチ（赤矢印）は，まず認知機能をアルゴリズムレベルで捉えることを目的としています。 これは，生物学的な実装を無視して，課題成績の基礎となる情報処理をアルゴリズムの構成要素に分解することに焦点を当てたアプローチです。 この 2 つのアプローチは「脳がどのようにして心を生み出しているのか」という共通の目標に向かって，連続した道の両極を形成しています。 全体として，認知的忠実度と生物学的忠実度の間にはトレードオフ（負の相関関係）がある。 しかし，認知的な制約が生物学的な機能を明らかにし，生物学が認知的な偉業を説明するモデルを刺激することで，トレードオフは相乗効果（正の相関）に変わります。 知能には豊かな世界観が必要なので，人間の脳の情報処理のモデルは，パラメトリックな複雑さが高くなります（右上の大きな点）。 生物学的な詳細を排除したモデルで課題成績を説明できたとしても，神経生物学的な実装を説明するためには，生物学的に詳細なモデルが必要になります。 この図は，モデル間の関係を理解し，それぞれの補完的な貢献を評価するのに役立つ概念図です。 しかし，この図は，認知的忠実度，生物学的忠実度，モデルの複雑さを定量的に測定したものではありません。 この3つの変数をそれぞれ測定する明確な方法はまだ開発されていません。図は参考文献 [122].

2. 理論から実験へ

実験と理論の間により良い橋を架けるためには，まず理論を完全に特定する必要があります。 そのためには，理論を数学的に定義し，それを計算モデルで実装する必要があります（Box 1）。 計算モデルは，認知的な忠実さと生物学的な忠実さを両立させるために，さまざまなレベルの記述を行うことができる（図3）。 ニューロンの構成要素とダイナミクスのみを捉えるように設計されたモデル[71]は，認知機能[72]を説明することができない傾向にある（図3，横軸）。 逆に，認知機能だけを捉えて設計されたモデルは，脳との関連付けが難しい（図3，縦軸）。 心と脳を結びつけるためには，行動とニューロンのダイナミクスの両方の側面を捉えようとするモデルが必要です。 最近では，脳からの制約が認知機能の説明に役立つことが示唆されており[42,73,74]，その逆の場合もあり[37,38]，トレードオフをシナジーに変えることができます。

本節では，認知機能を表現やアルゴリズムで説明する課題成績モデルの最近の成功例に注目します。 課題実行モデルは，心理物理学や認知科学の中心的存在であり，伝統的に行動データを用いて検証されてきました。 しかし，脳活動データを用いてタスク実行モデルを検証する文献も出てきています。 ここでは，ニューラルネットワークモデルと認知モデルという2つの大まかなクラスのモデルを順に検討します。

2.1. ニューラルネットワークモデル

ニューラルネットワークモデル（囲み記事 2）の歴史は長く，様々な分野で織り交ぜられています。 計算論的神経科学では，様々なレベルの生物学的詳細に基づくニューラルネットワークモデルが，生物学的神経ネットワークのダイナミクスや初歩的な計算機能を理解するために不可欠なものとなっています[27,28]。 認知科学においては，1980 年代に並列分散処理と呼ばれる認知機能を理解するための新しいパラダイムを定義し[6,75]，この分野を神経科学に近づけました。 AI においては，視覚や音声認識などの知覚タスクから，言語翻訳などの記号処理タスク，さらには音声合成やロボット制御などの運動タスクに至るまで，多くのアプリケーションに大きな進歩をもたらしました[42,74]。 ニューラルネットワークモデルは，3 つの分野の成功基準を合わせて満たす，タスクパフォーマンスの高いモデルを構築するための共通言語です (図2)。

ニューラルネットワークモデルは，脳と同様に，フィードフォワード計算とリカレント計算を行うことができます[37,76]。 最近の進歩を支えているモデルは，線形-非線形の信号変換を複数の段階で構成するという意味で，深いものです． モデルは通常，数百万のパラメータ（接続重み）を持ち，タスクのパフォーマンスを最適化するように設定されます。 成功したパラダイムの一つは教師付き学習で，入力（画像など）と関連する出力（カテゴリーラベルなど）のトレーニングセットから，入力から出力への望ましいマッピングを学習します。 しかし，ニューラルネットワークモデルは，教師なしで学習することもでき，経験データに固有の複雑な統計的構造を学習することができる。

パラメータの数が多いと，解釈可能なパラメータの数が少ない単純なモデルに慣れている研究者は不安になります。 しかし，単純なモデルでは，複雑な知能の働きを説明することはできません。 AIの歴史は，知能には十分な世界の知識と，それを蓄えるための十分なパラメータの複雑さが必要であることを示している。 そのため，私たちは複雑なモデル（図3）と，それがもたらす課題に取り組まなければなりません。 一つの課題は，パラメータ数が多いため，モデルを理解するのが難しいことです。 モデルは完全に透明であるため，内部表現を理解するために何百万もの入力パターンを安価に調べることができます。 オーバーフィッティングの懸念に対処するため，モデルはその一般化性能の観点から評価されます。 例えば，視覚モデルは，学習していない画像に対する神経活動や行動反応を予測する能力の観点から評価されます。

囲み記事 2: ニューラルネットワークモデル

ニューラルネットワークモデル とは，生物学的なニューラルネットワークにヒントを得て開発されたモデルで，各ユニットが多数の入力を組み合わせ，情報がネットワークを通じて並行して処理されるという特徴を持っています。 生物学的に詳細なモデルでは，活動電位や各ニューロンの複数のコンパートメントにおけるダイナミクスを捉えることができますが，ニューラルネットワークモデルでは，生物学的な詳細は除外されています。 しかし，視覚的な物体認識などの認知機能を説明することができるため，認知と脳を結びつけるための魅力的なフレームワークとなっている。

典型的なユニットは，入力の線形結合を計算し，その結果を静的な非線形性に通します。 その出力は，ニューロンの発火率に似ていると解釈されることもある。浅いネットワーク（入力と出力の間に隠れユニットの層が1つあるもの）でも，任意の関数を近似することができます[123]。 しかし，深層ネットワーク（複数の隠れ層を持つネットワーク）は，実世界のタスクで必要とされる複雑な機能の多くをより効率的に捉えることができます． コンピュータビジョンをはじめとする多くのアプリケーションでは，**フィードフォワード** アーキテクチャが使用されている． しかし，ユニットの出力を再処理して複雑なダイナミクスを生成する **リカレントニューラルネットワーク** は， さらなる工学的進歩をもたらし[76]，脳内の再帰的なシグナル伝達をよりよく表現している[35,124-126]． フィードフォワードネットワークが **普遍関数近似器** であるのに対し，リカレントネットワークは **動的システムの普遍的な近似器** である[127]。 リカレント処理は，ネットワークがその限られた計算資源を時間の経過とともに再利用し，より複雑な計算のシーケンスを実行することを可能にする。 リカレントネットワークは，最近の刺激の履歴を動的に圧縮して表現し，現在の処理に必要な時間的コンテキスト情報を提供することができる。 その結果，リカレントネットワークは動的なパターンを認識し，予測し，生成することができる。

フィードフォワードネットワークもリカレントネットワークも，そのアーキテクチャと接続の重みの設定によって定義される。 重みを設定する方法の一つとして，出力をある目的の出力に近づけるために小さな調整を繰り返す方法があります（**教師あり学習**）。 各重みは，それを少し変更することで得られる誤差の減少に比例して調整されます。 この方法は，誤差が最も急峻に減少するような重みの空間のステップを生成することから，**勾配降下法** と呼ばれる。 勾配降下法は，各重みに対する誤差関数の微分を計算する効率的なアルゴリズムである **バックプロパゲーション** を用いて実装することができる。

脳がバックプロパゲーションのようなアルゴリズムを使って学習しているかどうかは議論の余地がある。 バックプロパゲーションやそれに近い形の教師付き学習の生物学的に妥当な実装方法がいくつか提案されている[128-130]。 **教師信号** は，複数の感覚モダリティが提供する文脈に基づいて内部的に生成されるかもしれない[131]。 感覚や記憶からより多くの証拠が得られるようになり，時間の経過とともに表現が動的に洗練されていくことに基づいて[132]，環境との相互作用で生じる内部および外部の強化信号に基づいて[133]。 **強化学習**[41]や，ニューラルネットワークのパラメータの **教師なし学習**[119,134]は，現在急速に進歩している分野です。 ニューラルネットワークのモデルは，生物学からヒントを得ることで，AIに飛躍的な進歩をもたらすことを実証しています。

人間の認知能力に匹敵するモデルを求めて，生物学に深く入り込んでいくことになりそうです[135]。 現在，工学的に最も成功している抽象的なニューラルネットワークモデルは，生物学的なハードウェアでも実装可能です。 しかし，これらのモデルは，脳の動的な構成要素のごく一部しか使用していません。 神経科学では，活動電位[108]，正規の微小回路[136]，樹状突起のダイナミクス[128,130,137]，振動[138]などのネットワーク現象[27]など，計算機能を持つ可能性のある動的コンポーネントの豊富なレパートリーが記述されています。 また，生物学は，グローバルなアーキテクチャに制約を与え，例えば，学習のための補完的なサブシステムを示唆しています139。 これらの生物学的要素を，意味のあるタスクを実行するように設計されたニューラルネットワークの文脈でモデル化することで，脳の計算にどのように貢献しているかが明らかになり，AI のさらなる発展につながるかもしれません。

最近のいくつかの研究では，ニューラルネットワークモデルを脳の情報処理のモデルとして検証し始めています[37,38]. これらの研究では，画像中の物体を認識するように訓練された深層畳み込みニューラルネットワークのモデルを用いて，霊長類の腹側視覚野における新規画像の脳内表現を予測した. モデルを用いて予測しました. その結果，深層畳み込みニューラルネットワークの内部表現は，ヒトとサルの下側頭葉皮質における視覚イメージの表現について，現在最も優れたモデルを提供することが明らかになった[77-79]。 多数のモデルを比較した場合，物体の分類というタスクを実行するために最適化されたモデルが，皮質の表現をよりよく説明していた[77,78].

物体を認識するように訓練された深層ニューラルネットワークの初期の層は，初期の視覚野の表現に似たものを含んでいる[78,80]。 腹側視覚ストリームに沿って移動すると，ニューラルネットワークのより高い層が，表現を説明するためのより良い基盤を提供するようになる[80-82]。 深い畳み込み神経回路網の高層は，物体の位置，大きさ，姿勢，そして物体のカテゴリーを復号することができるという点で，下側頭皮質の表現にも似ています[83]。 この分野では，脳の活動データを予測することでこれらのモデルを検証することに加えて，知覚された形状84や物体の類似性を反映した行動反応を予測することでモデルを検証し始めている[85]。

2.2. 認知モデル

認知レベルのモデルでは，神経生物学的に妥当な構成要素を用いてその実装に取り組むことなく，情報処理のイメージを描くことができます。 これにより，ニューラルネットワークモデルではまだ不十分な高次認知領域の研究を進めることができます。 さらに，あるプロセスがニューラルネットワークモデルでも実現可能な場合でも，認知モデルが有用な抽象化を提供することもあります。

現在，脳科学的な説明は，感覚や運動のプロセスが動物と環境を結びつける脳の周辺部に近い機能的な構成要素を支配しています。 しかし，高次の認知機能の多くは，脳科学的な説明やニューラルネットワークモデルでは理解できないものでした。 認知モデルのユニークな貢献を説明するために，生産システム，強化学習モデル，ベイズ認知モデルという3つのクラスの認知モデルについて簡単に説明する。

プロダクションシステムは，推論や問題解決を説明できる認知モデルのクラスの初期の例です。 これらのモデルは，ルールとロジックを使用しており，感覚データや運動信号ではなくシンボルで動作するという点でシンボリックである。 これらのモデルは，認知を物理的環境に基づかせる知覚や運動制御ではなく，認知を捉えます。 プロダクション」とは，「if-thenルール」に基づいて行われる認知的な行動のことである。 このようなルールのセットは，一連の生産物（「then」）が実行される条件（「if」）を規定する。 条件とは，現在の目標や記憶にある知識のことである。 アクションは，目標や知識の内部状態を変更することができる。 例えば，プロダクションはサブゴールを作成したり，推論を保存したりすることができる。 複数のルールで条件が満たされた場合，競合解決メカニズムが1つのプロダクションを選択します。 この形式論を用いて指定されたモデルは，一連のプロダクションを生成します。これは，ある認知的な目標に向かって作業をしている私たちの意識的な思考の流れに，ある程度似ているかもしれません。 また，プロダクション・システムの形式論は，普遍的な計算アーキテクチャーを提供するものでもある[86]。 ACT-R[5]のようなプロダクション・システムは，もともと行動データの指導のもとに開発されたものである。 最近では，このようなモデルは，地域平均のfMRI活性化タイムコースを予測する能力という点でもテストされ始めている[87]。

強化学習モデルは，エージェントが環境との相互作用を通じて，長期的な累積報酬を最大化するように学習する方法を示しています[88,89]。 生産システムと同様に，強化学習モデルでは，エージェントが状態と行動の離散的な記号表現を使用できる知覚モジュールと運動モジュールを備えていることを前提としています. エージェントは行動を選択し，その結果としての環境の状態を観察し，途中で報酬を受け取り，行動を改善するように学習します。 エージェントは，各状態とその期待される累積報酬を関連付ける「価値関数」を学習することができる。 エージェントは，各行動がどの状態につながるかを予測し，それらの状態の値を知っていれば，最も有望な行動を選択することができます。 エージェントは，各状態と有望なアクションを直接関連付ける「ポリシー」を学習することもできます。 行動の選択は，搾取（短期的な報酬をもたらす）と探索（学習に利益をもたらし，長期的な報酬をもたらす）のバランスをとる必要がある。

強化学習の分野では，累積報酬を最大化するためにどのように行動し，学習するかを定義するアルゴリズムを探求します。 強化学習理論は，心理学や神経科学にルーツを持ち，現在では機械学習やAIの重要な分野となっています。 強化学習理論は，古典的な手法である動的計画法，モンテカルロ法，網羅的探索法を限界事例として含む，非常に一般的な制御の視点を提供し，環境が確率的で部分的にしか観測されず，その因果関係のメカニズムが不明であるような困難なシナリオを扱うことができます。

エージェントは，環境を徹底的に探索し，どのような状態でも最も有望な行動を試行錯誤で学ぶことができるかもしれません（モデルフリー制御）。 そのためには，十分な学習時間と記憶力，そしてエージェントが早死にしないような環境が必要です。 しかし，生物は学習時間や記憶力が限られており，死に至るような相互作用を避けなければなりません。 このような状況では，エージェントは環境のモデルを構築した方が良いかもしれません。 モデルは，経験を圧縮して一般化し，新しい状況での知的行動を可能にする（モデルベース制御）。 モデルを用いない方法は，計算効率が高い（状態から値へのマッピング，あるいは行動への直接のマッピング）が，統計的には効率が悪い（学習に時間がかかる）. モデルベースの方法は，統計的には効率が良いが，（起こりうる未来をシミュレートするために）膨大な計算量を必要とする場合がある[90]。

信頼できるモデルを構築するのに十分な経験が得られるまでは，エージェントは単にエピソードを保存し，過去に成功した行動経路に戻るのが最善かもしれない（エピソード制御）[91,92]。 エピソードを保存することで，モデル構築に重要な逐次的な依存関係の情報が保存される。 さらに，エピソード制御は，成功した行動経路を支える因果関係を理解する前に，そのような依存性を利用することができる。

脳は，これらの3つの制御モード（モデルフリー，モデルベース，エピソード）をそれぞれ行うことができ[89]，まだ発見されていないアルゴリズムを用いて，それぞれの利点を組み合わせているように見える。 AI と計算論的神経科学は，このアルゴリズムを発見するという目標を共有しています[41,90,93-95]が，この目標に対して異なる角度からアプローチしています。 これは，認知的な課題が形式モデルの開発の動機となり，AIと神経科学の進歩を促す例です。

認知モデルの第三のクラスは，ベイズモデル（囲み記事 3）である[21,96-98]。 ベイズ推論は，認知についての本質的な規範的視点を提供する。 動物が最適な行動をとるために，脳が実際に何を計算すべきかを教えてくれる。 例えば，知覚的な推論では，現在の感覚データを事前の信念の文脈の中で考慮する必要がある。 ベイジアン推論とは，確率のルールに従って，データと事前の信念を組み合わせることです。

囲み記事 3: ベイズ認知モデル

ベイズ認知モデルは，「脳はある課題に対して統計的に最適な解を近似的に求める」という仮定に基づいています。 推論を行い，何をすべきかを決定するための統計的に最適な方法は，確率の法則を用いて，現在の感覚的な証拠をすべての利用可能な事前知識に照らして解釈することである。 統計的に最適な方法は，現在の感覚的な証拠を，確率の法則を用いて，利用可能なすべての予備知識に照らして解釈することです。 視覚の場合を考えてみましょう。 網膜の信号は，私たちが認識したいと思っている世界のオブジェクトを反映しています。 物体を推論するためには，どのような物体の構成が考えられるか，それぞれがどの程度画像を説明できるかを考える必要があります。 我々の事前の信念は，オブジェクトの各構成の確率と，与えられた構成が異なる網膜画像を生成する確率を把握する生成モデルによって表されます。

より形式的には，ベイズ視覚モデルは，感覚データ d（画像）と世界の原因 c（推論すべき表面，物体，光源の構成）の結合分布 $p(d, c)$ の生成モデルを使用することになる[140]。 同時分布 $p(d, c)$ は，原因のすべての可能な構成に対する事前分布 $p(c)$ と 尤度の積に等しい。 原因の特定の構成が与えられたときの，特定の画像の確率である $p(d|c)$ と等しい。 $p(d|c)$ の所定のモデルがあれば，特定の原因 c から特定の画像 d が生まれる確率である尤度を評価することができます。 また，$p(d|c)$ の暗黙のモデルとして，原因 c からデータ d（画像）への確率的なマッピングがある場合もあります。 このようなモデルは，自然な画像を生成します。 $p(d|c)$ のモデルは，規定されたものであれ，暗黙のものであれ，世界の原因がどのように画像を作り出すか，少なくとも画像とどのように関連するかを捉えています。 視覚認識は，ある画像が与えられたときの原因の確率分布である事後分布$p(c|d)$eを計算することになります。 事後 $p(c|d)$ eは，感覚データdを説明するために世界に存在しなければならない原因 c を明らかにする[141]。 $p(c|d)$ を計算するモデルは，画像を判別することから，判別モデルと呼ばれています（ここでは，効果（画像）から原因へのマッピング）。 反転には，数学的には，潜在的な原因に対する事前情報 $p(c)$ が必要です。 事前の $p(c)$ は 解釈を制約し，どのような画像も説明できる原因の複数の構成から生じる曖昧さを軽減するのに役立ちます。

原因 c の推論を，利用可能なすべての知識と不確実性を取り込んだ $p(d, c)$ の生成モデルに基づいて行うことは，統計的には最適（限られたデータで最良の推論が得られる）であるが，計算上は困難（動物が使用できる以上のニューロンや時間が必要になる可能性がある）である。 理想的には，推論$p(c|d)$に暗黙的に含まれる生成モデル$p(d, c)$は，画像形成に関する知識だけでなく，世界にあるものやdに関する知識も含んでいなければなりません。 推論$p(d, c)$に暗黙的に含まれる生成モデルは，画像形成に関する知識だけでなく，世界の物事やそれらの相互作用，そしてこれらのプロセスに関する我々の不確実性も含んでいなければならない。 一つの課題は，感覚データから生成モデルを学習することです。 その際には，学習した知識と残りの不確実性を表現する必要があります。 もし，生成モデルの仕様が間違っていれば，推論は最適なものにはなりません。 現実のタスクでは，ある程度のモデルの誤指定は避けられません。 例えば，生成モデルには，画像生成プロセスの過度に単純化されたバージョンが含まれている可能性があります。 もう一つの課題は，事後評価 $p(c|d)$ の計算である。 現実的に複雑な生成モデルでは，マルコフ連鎖モンテカルロ法，信念伝播法，変分法など，計算量の多い反復アルゴリズムを用いた推論が必要になることがあります。 統計的な効率と計算的な効率の間で脳がとる妥協[142-144]は，高速なフィードフォワード認識モデルを学習することで，頻出する成分の推論を高速化し，反復アルゴリズムでは流動的に導出できない結論を結晶化することであろう。 これは amortized inference[145,146] として知られている。

ベイズ型認知モデルは，最近，機械学習や統計学との相互作用の中で発展してきた。 初期の研究では，固定された構造を持つ生成モデルを使用しており，限られたパラメータのセットに関してのみ柔軟性があった。 現代の生成モデルは，データとともに複雑さを増し，固有の構造を発見することができる[98]。 これらのモデルは，あらかじめ定義された有限のパラメータセットによって制限されないため，ノンパラメトリックと呼ばれている[147]。 それらのパラメータは，事前に定義された境界なしに数を増やすことができる。

ベイズモデルは，基本的な感覚や運動のプロセスの理解に貢献してきた[22-24]。 また，古典的な認知バイアス[99]を，実験課題では間違っていても，現実世界では正しく役立つ事前の仮定の産物として説明することで，判断や意思決定といった高次の認知プロセスに対する洞察を提供してきた。

ベイズ型ノンパラメトリックモデルにより，認知科学はより複雑な認知能力を説明し始めている。 例えば，1つの例から新しい物体カテゴリーを誘導する人間の能力を考えてみましょう。 このような帰納的な推論には，現在のフィードフォワード・ニューラルネットワークモデルでは捉えられない種類の事前知識が必要である[100]。 カテゴリーを誘導するためには，対象物，その部品間の相互作用，それらがどのようにしてその機能を生み出すかについての理解に頼ることになる． ベイズ認知の視点では，人間の心は幼児期から世界のメンタルモデルを構築します[2]。 これらのモデルは，確率的な意味での生成モデルであるだけでなく，因果的で構成的である場合もあり，新規のシナリオや仮説的なシナリオに一般化するために再構成可能な要素を用いて，世界のプロセスの精神的なシミュレーションをサポートします[2,98,101]。 このモデリング・アプローチは，物理的世界[101-103]，さらには社会的世界[104]についての推論にも適用されている。

生成モデルは，一般的な知能に不可欠な要素である。 生成モデルを学習しようとするエージェントは，その経験の間のすべての関係を理解しようとする。 学習のために外部からの監視や強化を必要とせず，環境や自分自身についての洞察を得るためにすべての

経験を掘り起こすことができる。 特に，世界のプロセスの因果モデル（物体がどのように画像を引き起こすのか，現在がどのように未来を引き起こすのか）は，エージェントに深い理解を与え，推論や行動のより良い基盤となります。

ニューロン集団における確率分布の表現については，理論的にも実験的にも検討されてきた[105,106]。 しかし，ベイズ推論や学習，特にノンパラメトリックモデルでの構造学習を，脳内での実装に関連付けることは，依然として困難である[107]。 脳の計算理論として，サンプリングなどの近似的な推論アルゴリズムは，皮質のフィードバック信号や活動の相関を説明できるかもしれない[97,108-110]。 さらに，計算効率を上げるために脳が削減する手抜き，つまり近似性が，人間の統計的最適性からの逸脱を説明する可能性もある。 特に，認知実験では，人間の行動にサンプリング[111]や償却型推論[112]の特徴があることが明らかになっている。

ここで取り上げた3つのクラスを含む認知モデルは，認知を意味のある機能的な構成要素に分解する。 認知科学者は，脳内の実装から独立したモデルを宣言することで，現在のニューラルネットワークでは実現できない高レベルの認知プロセス[21,97,98]を扱うことができる。 認知モデルは，部分の役割を理解しようとするときに全体を見ることができるため，認知的計算論的神経科学には欠かせないものです。

囲み記事 4: なぜ，認知科学，計算論的神経科学，AI が互いに必要なのか？

認知科学 は単に認知モデルの脳への実装を説明するだけでなく，そのアルゴリズムを発見するために，計算論的神経科学を必要としています。 例えば，感覚処理や物体認識の主要なモデルは，脳にインスパイアされたニューラルネットワークですが，その計算は認知レベルでは簡単には捉えられません。 また，最近のベイズノンパラメトリックモデルの成功は，一般的にはまだ実世界の認知には適用できません。 人間の認知機能の計算効率を説明し，認知機能の詳細なダイナミクスや行動を予測するには，脳活動のダイナミクスを研究することが有効である。 行動を説明することは重要であるが，行動データだけでは複雑なモデルの制約条件としては不十分である。 しかし，脳のデータを適切に活用すれば，認知アルゴリズムに豊富な制約を与えることができます。 認知科学は，常に人工知能と密接な関係を保ちながら発展してきました。 両分野は，タスクを実行するモデルを構築するという目的を共有しており，共通の数学的理論とプログラミング環境を利用しています。

計算論的神経科学 は，より高度な認知に挑戦するために認知科学を必要としている。 実験レベルでは，認知科学の課題により，計算論的神経科学が認知を実験室に持ち込むことができます。 理論レベルでは，認知科学は，計算神経科学が研究している神経生物学的な動的構成要素が，認知や行動にどのように寄与しているかを説明することに挑戦します。 計算論的神経科学は，生物学的に妥当な動的要素を持つ認知機能をモデル化するための理論的・技術的基盤を提供するために，AI，特に機械学習を必要としています。 認知機能を生物学的に妥当な動的要素でモデル化するための理論的・技術的基盤を提供するために，AI，特に機械学習が必要です。

人工知能 を実現するためには，認知科学が必要である。 認知科学の課題は AI システムのベンチマークとなり，初歩的な認知能力から人工的な一般知能へと積み上げていくことができます。 人間の発達と学習に関する文献は，学習者が何を達成することが可能か，また，どのような種類の世界との相互作用が知能の獲得をサポートするかについての重要な指針となります。 人工知能には，アルゴリズムのヒントとなる計算論的神経科学が必要です。 ニューラルネットワークモデルは，脳からヒントを得た技術の一例であり，AI のいくつかの分野で他の追随を許さないものです。 神経生物学的な動的構成要素（例えば，スパイクニューロン，樹状突起のダイナミクス，皮質の典型的な微小回路，振動，神経調節プロセスなど）や，人間の脳のグローバルな機能レイアウト（例えば，感覚モダリティ，記憶，計画，運動制御などの異なる機能に特化したサブシステムなど）からさらなるインスピレーションを得ることができれば，AIのさらなるブレークスルーにつながるかもしれません。 機械学習は，統計学と計算機科学という別々の伝統に基づいており，それぞれ統計的な効率と計算的な効率を最適化してきました。 計算効率と統計効率の統合は，ビッグデータ時代の必須課題です。 脳は計算効率と統計効率の両方を兼ね備えていると考えられ，そのアルゴリズムを理解することで機械学習を促進できるかもしれません。

3. 先を見据える

3.1. ボトムアップとトップダウン

脳は，ボトムアップの識別的計算とトップダウンの生成的計算をシームレスに統合して，知覚の推論や，モデルフリーとモデルベースの制御を行っています。 脳科学も同様に，記述のレベルを統合し，ボトムアップとトップダウンの両方を進めていく必要があります。そうすることで，ニューロンのダイナミクスに基づいてタスクのパフォーマンスを説明し，脳がどのようにして心を生み出すのかをメカニズム的に説明することができます。

脳の計算を理解するために詳細な測定を行うというボトムアップの考え方は，最近の最も重要な資金調達の原動力となっています。 欧州のヒューマン・ブレイン・プロジェクトや米国の BRAIN Initiative は いずれもボトムアップの考え方に基づいており，回路レベルに焦点を当てて脳のダイナミクスを測定・モデリングすることで，脳の計算を理解しようとしています。 BRAIN Initiative は，神経細胞の活動を測定・操作する技術の向上を目指しています。 Human Brain Project は，神経科学のデータを生物学的に詳細な動的モデルに統合しようとするものです。 いずれも，実験から理論へ，細胞レベルの記述から大規模な現象へと進んでいきます。

多数のニューロンを同時に測定し，その相互作用を回路レベルでモデル化することが不可欠となる。 ボトムアップのビジョンは，科学の歴史に基づいています。 例えば，顕微鏡や望遠鏡は，科学的なブレークスルーをもたらしました。 しかし，より優れた観測結果によって理解が進むのは，常に先行する理論（観測されたプロセスの生成モデル）の中でのことです。 例えば天文学では，コペルニクスの理論がガリレオの望遠鏡による観測結果を解釈する際の指針となりました。

脳を理解するためには，理論と実験を並行して開発し，ボトムアップのデータ駆動型アプローチを，説明すべき行動機能から始まるトップダウンの理論駆動型アプローチで補完する必要があります1[13,114]。 これまでにないほど豊富な脳活動の測定と操作を行い，生物の行動適性に寄与する機能を果たすことができるかどうかという最初のテストに合格した脳-計算モデルを判断するために使用することで，理論的な洞察が得られます。 このように，トップダウンのアプローチは，ボトムアップのアプローチを補完するものとして，脳の理解に欠かせないものとなっている (図3)。

3.2 Marr のレベルの統合

Marr（1982）は，分析のレベルを 3 つに分けている。 (1) 計算理論，(2) 表現とアルゴリズム，(3) 神経生物学的実装である[115]。 認知科学は，計算理論から始まり，認知を構成要素に分解し，表現とアルゴリズムをトップダウンで開発する。 計算論的神経科学はボトムアップで進み，ニューロンの構成要素を，脳の全体的な機能の文脈で有用な構成要素と考えられる表現やアルゴリズムに構成する。 計算神経科学はボトムアップで進められます。 AI は，単純な構成要素を組み合わせて複雑な知能を実現する表現やアルゴリズムを構築する。 このように 3 つの学問分野は，脳と心のアルゴリズムと表現に収斂し，補完的な制約をもたらします[116]。

Marr のレベルは，脳を理解するための課題に役立つ指針となる。 しかし，認知科学が脳を考慮する必要がないとか，計算論的神経科学が認知を考慮する必要がないということを示唆するものではありません（囲み記事 4）。 Marr は，コンピュータにインスピレーションを受けた。コンピュータは，人間の技術者が高レベルのアルゴリズムの記述に正確に適合するように設計する。 これにより，技術者はアルゴリズムを設計する際に，回路を抽象化することができる。 しかし，コンピュータサイエンスでも，アルゴリズムの一部は，並列処理能力などのハードウェアに依存します。 しかし，脳はコンピュータとは異なり，この依存性をさらに強めている。 脳は，進化と発達の産物であり，その過程では，ある抽象的な記述レベルで完璧に動作を把握できるシステムを生成するような制約はない。 したがって，脳への実装を考えずに認知を理解することはできないし，逆に，認知機能を支える神経回路の文脈を無視して神経回路を理解することもできない。

分野を超えた課題の例として，初めてエスカレーターを見た子供の場合を考えてみましょう。 エスカレーターを初めて見た子供は，人が斜め上に向かって昇っていくステップをすぐに認識します。 エスカレーターを「動く階段」と考え，それに乗って，力を入れずに1階分持ち上げられることを想像するかもしれない。 「エスカレーター」という言葉を覚える前に，たった一度の体験で機能を推察し，新しい概念を形成するかもしれない。

深層ニューラルネットワークモデルは，視覚体験の要素（人，段差，斜め上方向の動き，手すり）を迅速に認識することについて，生物学的に妥当な説明を提供する。 計算効率の高いパターン認識コンポーネントを説明することができます[42]。 しかし，要素間の関係，物体の物理的な相互作用，人が上に行くという目的，エスカレーターの機能などを子供がどのように理解しているのか，また，体験を想像して瞬時に新しい概念を形成することができるのかについては，まだ説明できません。

ベイズノンパラメトリックモデルは，単一の経験からの深い推論や概念形成がいかにして可能かを説明する。 脳の驚くべき統計的効率，すなわち，抽象的な事前知識を提供する生成モデルを構築することで，少ないデータから多くのことを推論する能力を説明できるかもしれません[98]。 しかし，現在の推論アルゴリズムは大量の計算を必要とし，その結果，単一の視覚的経験から「エスカレーター」という新しい概念を形成するというような実世界の課題にはまだ対応できていない.

脳のアルゴリズムは，20 ワットの電力予算で，統計的効率と計算的効率を両立させており，ベイジアンやニューラルネットワーク型の現在の AI を凌駕しています。 しかし，最近の AI や機械学習では，ベイジアン推論とニューラルネットワークモデルの交点を探り始めており，前者の統計的な強み（不確実性の表現，確率的推論，統計的効率）と後者の計算的な強み（表現的学習，普遍的な関数近似，計算効率）を組み合わせている[117-119]。

Marrの 3 つのレベルを統合するには，さまざまな専門知識を持つ研究者が緊密に協力する必要があります。 神経科学，認知科学，AI スケールの計算モデリングを，一つの研究室が得意とすることは困難です。 そのため，相補的な専門性を持つ研究室間でのコラボレーションが必要になります。 従来の共同研究に加えて，分野間でコンポーネントを共有するオープンサイエンスの文化は，Marr のレベルを統合するのに役立ちます。 共有できる要素としては，認知タスク，脳や行動のデータ，計算モデル，生物学的システムと比較してモデルを評価するテストなどがあります（囲み記事5）。

心と脳の研究は，特にエキサイティングな局面を迎えています。 最近のコンピュータのハードウェアとソフトウェアの進歩により，AI スケールの心と脳のモデリングが可能になりました。 認知科学，計算論的神経科学，AI が一体となれば，人間の認知を神経生物学的に妥当な計算モデルで説明できるようになるかもしれません。

囲み記事 5：共有可能な課題，データ，モデル，テスト：学際的なコラボレーションの新しい文化

認知を説明する神経生物学的に妥当なモデルは，パラメータがかなり複雑になります。 そのようなモデルの構築と評価には，機械学習と大規模な脳・行動データセットが必要になります。 従来，各研究室は，自分の専門分野の目標に焦点を当てて，独自のタスク，データセット，モデル，テストを開発してきました。 しかし，このような取り組みを課題に合わせて拡大していくためには，3つの分野に関連するタスク，データ，モデル，テストを開発し，研究室間で共有する必要があります（図参照）。 新しいコラボレーションの文化は，異なる研究室のコンポーネントを組み合わせることで，ビッグデータとビッグモデルを組み立てることになります。 認知科学，計算論的神経科学，人工知能という 3 つの分野が一体となった成功の基準を満たすためには，従来の分野を超えた最適な役割分担が必要になるかもしれません。

課題 実験課題を設計することで，認知を定量的に調査可能な要素に切り分けます。 タスクとは，行動を制御するための環境のことです。 タスクは，感覚入力（例：視覚刺激）と運動出力（例：ボタンを押す，ジョイスティックを操作する，より高次元の手足や全身の制御）を捉えるタスク「ワールド」のダイナミクスを定義します。 タスクは，脳や行動のデータを取得し，AI モデルを開発する際に，明確な課題とモデルを比較するための定量的な性能ベンチマークを提供します。 例えば，ImageNet タスク[148]は，コンピュータビジョンの大きな進歩をもたらしました。 タスクは，データの取得とモデルの開発を推進するために，3つの分野すべてで容易に利用できるように設計・実装する必要があります（関連する開発には，OpenAI の Gym，Universe，DeepMind の Lab[149]があります）。 役に立つタスクのスペクトルには，単純な刺激と反応を用いる古典的な心理物理学的タスクや，仮想現実

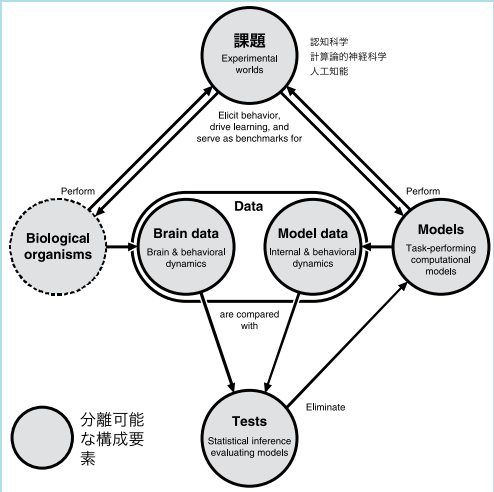
でのインタラクションが含まれます。 人間の心のあらゆる側面に関わるようになると，タスクは自然環境をシミュレートする必要があり，コンピュータゲームのようになっていくでしょう。 これにより，特にタスクがインターネットを介して実行される場合には，大量の参加者と大きな行動データという付加的なメリットがもたらされる可能性があります[150]。

データ タスクの実行中に得られた行動データは，全体的なパフォーマンスの推定値や，成功と失敗，反応時間と動作軌跡の詳細なサインを提供します。 脳活動の測定は，タスクパフォーマンスの基礎となるダイナミックな計算を特徴づける。 解剖学的データは，脳の構造と接続性を複数のスケールで特徴づけることができます。 脳の構造的データ，脳の機能的データ，行動的データのすべてが，計算モデルに制約を与えるために不可欠である。

モデル 課題を実行する計算モデルは，実験的なタスクを実行するために，感覚入力を受け取り，運動出力を生成することができます。 AI スケールの神経生物学的に妥当なモデルは，オープンに共有され，そのタスクパフォーマンスや，モデル定義後に得られた新しいデータセットを含む様々な脳や行動のデータセットを説明する能力の観点からテストすることができる。 モデルを定義した後に得られた新しいデータセットも含めて，様々な脳や行動のデータセットを説明する能力をテストします。 最初は，多くのモデルが小さなタスクのサブセットに特化したものになるでしょう。最終的には，モデルはタスク間で一般化しなければならない。

テスト あるモデルが特定のタスクにおける脳の情報処理をどの程度説明できるかを評価するためには，脳や行動のデータに基づいてモデルと脳を比較するテストが必要です。 すべての脳は，その構造と機能において固有のものです。 さらに，ある脳にとって，知覚，認知，行動のすべての行為は，時間的にユニークであり，まさにその脳を永久に変えてしまうため，繰り返すことができません。 このような複雑さが，脳とモデルの比較を難しくしています。 私たちは，関心のある要約統計と，モデルと脳の間の空間と時間における対応付けを，ある程度の抽象度で定義しなければなりません。 モデル間の比較を行い，脳の理解にどれだけ近づいたかを判断するための適切なテストを開発することは，単に統計的推論の技術的課題ではありません。 それは，理論的な神経科学の基本となる概念的な課題なのです。

研究室や分野間の相互作用は，敵対的な協力から利益を得ることができます[134]。 現在の計算モデルでは認知の重要な側面を説明できないと感じている認知研究者は，これらの欠点を定量化する共有可能なタスクやテストを設計し，AI モデルの基準となる人間の行動データを提供するよう求められています。 現在のモデルでは脳の情報処理を説明できないと感じている神経科学者は，タスク実行中に取得した脳の活動データや，脳とモデルの活動パターンを比較するテストを共有し，モデルの欠点を定量化することが求められています。 成功の定義は複数あるでしょうが，それをモデルの質の定量的な尺度に変換することは不可欠であり，工学だけでなく認知計算神経科学の進歩にもつながるでしょう。



共有可能な構成要素間の相互作用. タスク、データ、モデル、テストは、研究室間や分野を超えた共有に適した構成要素（灰色のノード）で、研究室間で集められた脳や行動の大規模なデータセットから駆動される大規模なモデルを共同で構築し、テストすることができます。