- title: A deep learning framework for neuroscience
- author: Blake A. Richards, Timothy P. Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, Colleen J. Gillon, Danijar Hafner, Adam Kepecs, Nikolaus Kriegeskorte, Peter Latham, Grace W. Lindsay, Kenneth D. Miller, Richard Naud, Christopher C. Pack, Panayiota Poirazi, Pieter Roelfsema, Joao Sacramento, Andrew Saxe, Benjamin Scellier, Anna C. Schapiro, Walter Senn, Greg Wayne, Daniel Yamins, Friedemann Zenke, Joel Zylberberg, Denis Therien, Konrad P. Kording.
- url: https://www.nature.com/articles/s41593-019-0520-2
- year: 2019
- journal: Nature Neuroscience
- pages: 1761-1770

神経科学のための深層学習の枠組み

A deep learning framework for neuroscience (2019)

Blake A. Richards, Timothy P. Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, Colleen J. Gillon, Danijar Hafner, Adam Kepecs, Nikolaus Kriegeskorte, Peter Latham, Grace W. Lindsay, Kenneth D. Miller, Richard Naud, Christopher C. Pack, Panayiota Poirazi, Pieter Roelfsema, Joao Sacramento, Andrew Saxe, Benjamin Scellier, Anna C. Schapiro, Walter Senn, Greg Wayne, Daniel Yamins, Friedemann Zenke, Joel Zylberberg, Denis Therien, Konrad P. Kording

システム神経科学は，知覚，認知，運動などのさまざまな課題を脳がどのように実行しているかを説明しようとするものである。 一方，人工知能は，解決しなければならない課題に基づいて計算系を設計しようとするものである。 人工ニューラルネットワークでは，目的関数，学習則，アーキテクチャの 3 つの要素が設計で指定される。 脳に触発されたアーキテクチャを利用した深層学習の成功に伴い，これらの 3 つの設計要素は，複雑な人工学習系のモデル化，エンジニアリング，最適化の方法の中心となってきている。 ここでは，これらの構成要素にもっと焦点を当てることが，システム神経科学にも役立つと主張する。 この最適化に基づく枠組みが，神経科学における理論的・実験的な進歩をどのように促すかについて，例を挙げて説明する。 システム神経科学に対するこのような原則的な視点が，より迅速な進歩をもたらすことになると考えている。

大きな技術的進歩により，脳を大規模に観察・操作したり，複雑な行動を定量化したりする能力に革命が起きている (1,2)。では，これらのデータをどのように活用して脳のモデルを構築していけばよいだろうか？ システム神経科学の古典的な枠組みが構築された当時は，小さなニューロンセットからしか記録できなかった。 この枠組みでは，研究者は神経活動を観察し，個々のニューロンが何を計算しているかについての理論を構築し，次にニューロンがどのように操作を組み合わせているかについての回路レベルの理論を組み立てる。 この方法は, 単純な計算には有効である。 例えば，中枢パターン発生器がリズミカルな動きを制御する仕組み (3)，前庭眼反射が視線の安定化を促進する仕組み(4)，網膜が動きを計算する仕組み (5) などがわかっている。 しかし，この古典的な枠組みは，何千ものニューロンの記録や，説明したいと考えているすべての行動に対応できるのだろうか？ おそらく，新皮質や海馬のように，多数の機能を実行する大規模な神経回路では，古典的なアプローチはあまり成功していないのではないだろうか。 このような回路では，簡潔にまとめるのが難しい反応特性を持つニューロンがよく見つかる (6,7)。

古典的な枠組みでは限界があるため，実験の進歩を生かすためには新しいアプローチが必要である。 神経科学と人工知能 (AI) の間の相互作用から，有望な枠組みが生まれつつある (8-10)。 機械学習の代表的な手法として深層学習が登場したこと

で，人工ニューラルネットワーク (ANN) が見直されている。 ANN は 実際のニューロンの統合および活性化の特性を大まかに模倣した単純化されたユニットを用いて，神経計算をモデル化するものである (11)。 ユニットは，非常に単純化された線形演算から，複数のコンパートメントやスパイクなどを含む比較的複雑なモデルまで，さまざまな抽象度で実装される (11-14)。 重要なのは ANN が実行する特定の計算は，設計するのではなく，学習するということである (15)。

すなわち，学習目標は，シナプスの重みの更新として表現される学習規則のセットを最大化または最小化されるべき目的関数 (または損失関数) として表される。 そして，情報を流すための経路や接続として表現されるネットワークアーキテクチャである (図 1)(15)。 この枠組みでは，計算がどのように実行されるかを要約するのではなく，どのような目的関数，学習則，アーキテクチャがその計算の学習を可能にするかを要約している。

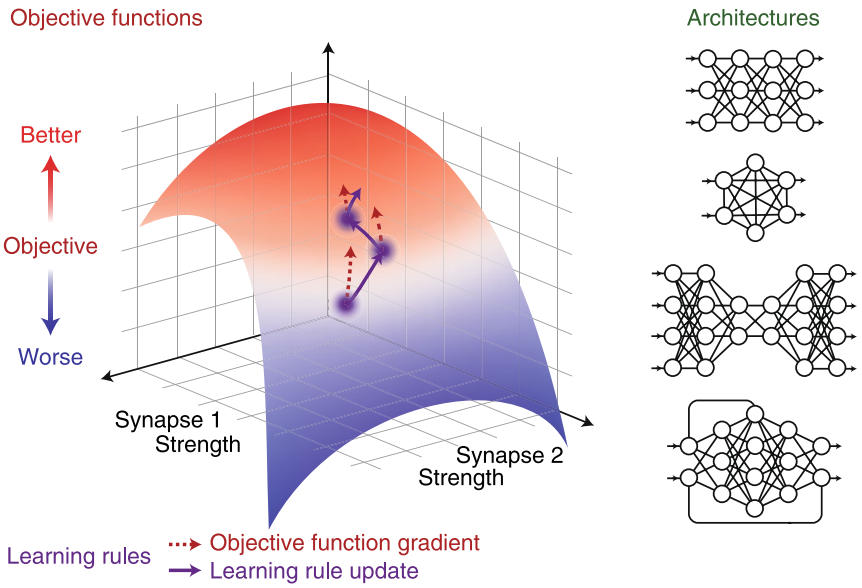

図 1. ANN 設計の 3 つの核となる要素。 ANN を設計するとき ネットワークが実行する特定の計算を作るのではなく 以下の 3 つの要素を指定する。 **目的関数** は課題におけるネットワークの成績を定量化する。 学習は目的関数を最大化または最小化するシナプスの重みを見つける。 **学習則** はシナプスの重みを更新するためのレシピを提供する。 これにより，目的関数の明示的な勾配に従わなくても目的関数を上昇させることができる。 **アーキテクチャ** はネットワーク内のユニットの配置を指定し，情報の流れや，ネットワークが学習可能な計算，不可能な計算を決定する。

深層学習は，古くからある ANN の発想を再ランディングしたものと考えられる (11)。 深層 ANN は，フィードフォワード，または時間経過による複数のリカレント層を持つ。 「層」は 生物の脳の特定の層というよりも，脳の領域に類似していると考えるのが最適である (16,17)。 深層学習とは, 特に, 階層化された ANN をエンド・ツー・エンドで学習することである。 階層の各層の可塑性が学習目標に寄与するようにする(15)。 これには，「信用割り当て問題」(囲み記事 1) の解決が必要である (18,19)。 近年，深層学習の進歩は，必要な計算を効率的に処理できる GPU (Graphics Processing Unit) を使用して，より大きなデータセットで学習された，より大きな ANN を使用することによってもたらされた。 このような開発により，画像(20)，音声(21) の分類と生成， 言語処理と翻訳(22)，触覚と把持(23)，ナビゲーション(24)，感覚の予測(25)，ゲームプレイ(26)，推論(27) など，多くの新しい問題に対する解決策が生み出されている。

最近の多くの知見は，深層学習が脳の理論に役立つことを示唆している。 まず，深層 ANN は，霊長類の知覚系における表現変換を，場合によっては忠実に模倣することができ (17,28)，それによって神経活動を操作することができることが示されている (29)。 第二に，グリッドセル(24)，形状同調(30)，時間的受容野(31)，錯視(32)，モデルベース推論(33) など，多くのよく知られた行動および神経生理学的現象が，動物が解決したものと同様の課題で訓練された深層 ANN で現れることが示されている。 第三に，多くのモデリング研究により，典型的な誤差逆伝播アルゴリズム (backpropagation-of-error algorithm) の力を模倣できる学習アルゴリズムなどのエンド・ツー・エンドの学習則の見かけ上の生物学的非現実性が実証されている。 of-errorアルゴリズム（backprop）を模倣した学習アルゴリズムなど，エンド・ツー・エンドの学習則は，一見すると生物学的にはあり得ないことが明らかになった (図 2 および 囲み記事 1)。 細胞や細胞内の電気生理，抑制性微小回路，スパイクタイミングのパターン，短期的な可塑性，フィードバック接続などについて比較的簡単な仮定を置くことで，生物学的系が深層 ANN における逆伝播のような学習を近似的に行うことができる (12,14,34-39)。 したがって，ANN に基

づく脳のモデルは，これまで考えられていたほど非現実的なものではないかもしれないし，同時に，多くの神経生物学的データを説明することができると思われる。

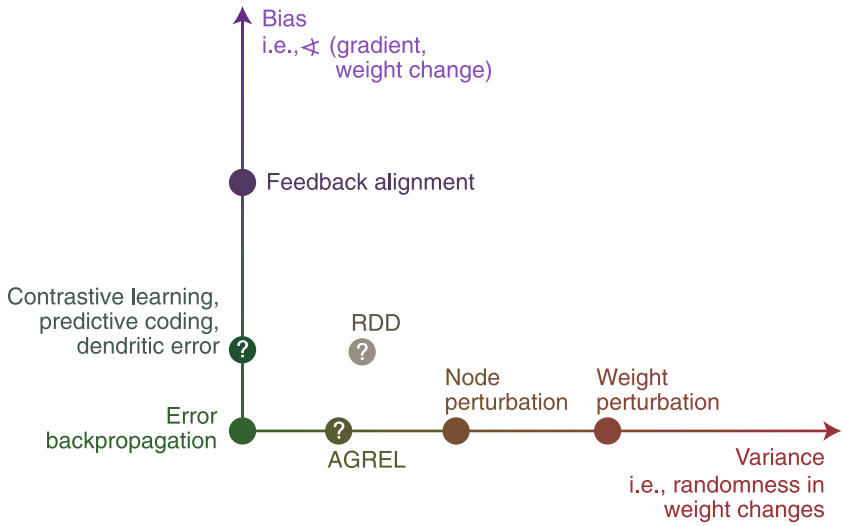

図 2. 学習則のバイアスとバリアンス。 多くの学習則は, 明示的に勾配ベースでなくても, 目的関数の勾配の推定値を提供する。 しかし, 他の推定器と同様に, これらの学習規則は, 勾配の推定値に様々な程度の分散や偏りを示す可能性がある。 ここでは, 提案されている生物学的に妥当な学習規則のいくつかが, 誤差逆伝播法 と比較して どの程度のバイアスとバリアンスを持つかを大まかに説明する。 ここで重要なのは, 多くの学習規則の正確なバイアスとバリアンス の特性は不明であり, これは単なるスケッチに過ぎない。 例えば, 対照的なヘブ学習則, 予測コーディング(35), 樹状突起エラー学習(14), 回帰不連続計画(RDD)(91), 注意喚起強化学習 (AGREL) (37) については, その場所をクエスチョンマークで示している。 その他, バックプロパゲーション, フィードバック・アライメント(36), ノード/ウェイト摂動 (92) については, それらの既知の相対的位置を示している。

囲み記事 1. 学習と信用割当問題。 学習の自然な定義は 「性能を向上させる系への変更」である。 目的関数 $F(W)$ があると仮定する。 これは，現在のシナプスの重みの N 次元ベクトル W が与えられたときに, 系が現在どれだけうまく機能しているかを測定する。 もしシナプスの重みが W から $W + \Delta W$ に変化した場合，成績の変化は $\Delta F = F(W + \Delta W) - F(W)$ となる。 W に小さな変化を与え，F が局所的に滑らかであれば ΔF は重みの変化と勾配の内積でほぼ与えられる(41)。

$$\Delta F \approx \Delta W^{\top} \cdot \Delta_w F(W)$$

ここで, $\Delta W F(W)$ は W に関する F の勾配であり, $\top$ は転置を示す。 改善された性能を保証したい, つまり, $\Delta F \geq 0$ を保証したいとする。 このとき, W の局所的な変化がすべて同じ改善につながる $(N-1)$ 次元の多様性があることがわかっている。 どれを選べば良いのか？ 勾配ベースのアルゴリズムは，「そのステップサイズで最大の改善が得られる方向に小さなステップを踏みたい」という直観に基づいている。 勾配は目的の最も急な方向を指しているので, 小さなステップサイズ η, 勾配の倍数 $\Delta W F$ を選択すれば, そのステップサイズで可能な限りの改善を行うことができる。 したがって 次のようになる:

$$\Delta F \approx \eta \Delta_W F(W)^{\top} \cdot \Delta_W F(W) \geq 0$$

つまり 目的関数値は, ステップごとに (η が小さい場合), 勾配ベクトルの長さに応じて増加する。

信用割当の概念は，与えられたニューロンやシナプスが与えられた結果に対してどれだけの 「信用」や「不平」 を得るべきかを決定する問題を意味する。 より具体的には $\Delta F \geq 0$ を確実にするために, システムの各パラメータ (例えば 各シナプスの重み) をどのように変化させるべきかを決定する方法である。 最も単純な形では信用割当問題は，複雑なネットワークにおいて信用を割り当てることの難しさを意味する。 目的関数の勾配 $\Delta F(W)$ を使って重みを更新することは ANN の信用割当問題を解決する優れた手段であることが証明されている。 システム神経科学が直面している問題は，脳も勾配法のようなもので近似しているのかということである。

深層 ANN の勾配を計算する最も一般的な方法は 誤差逆伝播法 (15) である。 誤差逆伝播法は，合成関数の微分則 (連鎖法則) を用いて, 出力から逆方向に勾配を再帰的に計算するものである (11)。 しかし，誤差逆伝播法は，フィードバックの重みが対称であるとか，情報の前後の受け渡しが明確であるなど，生物学的にはありえない仮定に基づいている(14)。 誤差逆伝播法に限らず，多くの学習アルゴリズムが勾配の推定値を提供しており，それらの中には誤差逆伝播法の生物学的にありえない仮定の影響を受けないものもある (12,14,34-38,91,93,94)。

しかし，アルゴリズムはそのバイアスとバリアンスの特性に違いがある (図 2)(36,92)。 報酬によってシナプスの重みのランダムな変化を強化する weight/node 摂動法 のようなアルゴリズムは，勾配に沿った経路の分散が大きい(92)。 勾配情報を伝えるためにランダムなフィードバック重みを用いるアルゴリズムは，高いバイアスを持つ (36,95)。 生物学的な現実性を維持しつつ，アルゴリズムのバイアスやバリアンスを最小化するための様々な提案がなされている (37,38)。

こうした動きを受けて，システム神経科学のための深層学習に触発された枠組みを検討するのに適した時期に来ている (8,19,40)。 ANN の基盤となる主要な原理については理解が進んでおり，これらの洞察が一般的に適用されると考える理論的な理由もある (41,42)。 同時に，大規模な神経集団を監視・操作できるようになったことで，深層学習に関する文献から得られた仮説を検証する新しい方法が生まれている。 ここでは， 現代のシステム神経科学のための深層学習の枠組みの骨子を説明する。

1. 「課題セット」を用いた人工ニューラルネットワークおよび脳の学習の制約

「ノーフリーランチ定理」は，どんな学習アルゴリズムも，起こりうるすべての問題でうまく機能することはできないということを大まかに示した (43)。 そこで 21 世紀最初の 10 年間の ANN 研究者たちは，AI は 「…ほとんどの動物が難なくこなすことのできる，知覚や制御，さらには… 長期的な予測，推論，計画，(通信) などの一連の課題に主眼を置くべきだ」と主張した (44)。 この一連の課題 「AI セット」と呼ばれており，人間や動物と同様の能力を持つコンピュータを構築することに焦点を当てていることが，AI 課題をコンピュータサイエンスの他の課題と区別している点である (44) (なお，ここで言う「課題」とは，教師なしのものも含めて，あらゆる計算を広く指している)。

深層学習が成功した理由の多くは AI セットでの学習を考慮したことにある (15, 44)。 特定の課題を学習するのに適した ANN を設計することは「帰納バイアス」を取り入れることの一例である (囲み記事 2)。 これは，ある最適化問題の解の性質についての仮定を意味する。 深層学習がうまく機能しているのは，AI セット(15,45)，特に階層型アーキテクチャに適切な帰納バイアスを使用していることが一因である。 例えば，画像は，エッジから，エッジの単純な組み合わせ，物体を形成するより大きな構成まで，複雑さを増す特徴の階層的な集合に構成することで，うまく説明できる。 言語もまた，音素が単語に，単語が文章に，文章が物語にと，階層的に構成されていると考えることができる。 しかし，深層学習では，人手によるエンジニアリングを排除し，系が計算する機能が学習中に現れるようにしている (15)。 このように，深層学習は，計算能力の向上だけに依存しているとか，「白紙状態」で知能を発揮するという通説があるが，深層学習の成功の多くは，有用な帰納バイアスと創発的な計算のバランスから生まれたものであり，大人の脳を支える自然と育成の融合に呼応するものである。

囲み記事 2. 機能バイアスとは何か？ 解決しなければならない問題の種類についての事前知識があれば，学習はより簡単になる (43)。 帰納バイアスは，そのような事前知識を最適化系に埋め込むための手段である。 このような帰納バイアスは，一般的なもの (階層など) もあれば，特殊なもの (畳み込みなど) もある。 重要なのは，脳内に存在する帰納バイアスは，地球上の生活という広い文脈 (例えば，食料，水，シェルターなどを得る必要のある 3 次元世界での生活)と，特定の生態系ニッチにおける動物の適応度を高めるために，進化によって形成されたものである。

- **説明の簡便性**: 世界を理解しようとするとき オッカムのカミソリ (96) で示されているように, 単純な説明が好まれることがある。 ベイズ的枠組みや，疎性表現 (59) のような他の機構を用いて ANN に組み込むことができる。

- **物体永続性 object permanence**: 世界は時空間的に一定の物体に整理されている。 感覚空間での一貫した動きを仮定した表現を学習することで, ANN にこれを組み込むことができる(97)。
- **視覚変換不変性 visual translation invariance**: 視覚特徴は その場所に関わらず同じ意味を持つ傾向がある。 これは，畳み込み演算を用いて ANN に組み込むことができる(98)。
- **焦点化注意 focused attention**: 系に入ってくる情報の中には，他の情報よりも重要なものがある。 これを ANN に注意機構で組み込むことができる (99)。

同様に，神経科学者は，ある種の生物が実行するために進化した行動や課題に注目する。 この課題集合は，完全ではないにしても，AI セットと重なる部分がある。 研究者は特定の種の「脳の集合」 (その種の生存と繁殖にとって重要な課題) を考慮することで，学習の鍵となる可能性が最も高い機能に焦点を当てることができる。 例えば AI セットに有用な帰納バイアスを持つ ANN デザインに注目するなど，純粋な白紙状態から出発することが，現代の ANN の成功の鍵であったように，深層学習の開発にもそれが重要になると考えられる。 システム神経科学のための深層学習枠組みの開発には，与えられた動物が適切な脳の集合でどのように課題を解決するかに焦点を当てることが重要であると考えている。

また，深層学習における帰納バイアスの重要性を認識することは，既存の誤解を解くことにもつながる。 深層ネットワークは，大量のデータに依存するため，脳とは異なると考えられがちである。 しかし (i) 多くの生物種，特に人間は，大量の経験的データを用いてゆっくりと発達すること, (ii) 深層ネットワークは，優れた帰納バイアスを備えていれば，データ量の少ない体制でもうまく機能すること，は注目に値する(46)。 例えば，深層ネットワークは，学習方法を素早く学ぶことができる (47)。 脳の場合，このような帰納バイアスを獲得する手段の一つとして進化が考えられる (48,49)。

2. 脳の深層学習の枠組みの 3 つの主要要素

深層学習は, 人間の設計と自動学習を組み合わせて課題を解決するものである。 設計とは, 計算内容 (ANN の具体的な入出力関数) ではなく, 3 つの要素で構成される。 設計は 計算 (ANN の具体的な入出力関数) ではなく, (i) 目的関数, (ii) 学習則, (iii) アーキテクチャの 3 つの要素で構成される (図1)。 「目的関数」は, 学習システムの目標を記述するものである。 目的関数は, ニューラルネットワークのシナプスの重みと受け取るデータの関数だが, 特定のタスクやデータセットを参照せずに定義することができる。 例えば, 機械学習でよく使われる 交差エントロピー目的関数は, ImageNet データセット における犬種の区別から, ツイートの感情分類まで, あらゆる分類課題の成績を計算する手段を規定している。 脳に対して提案されている具体的な目的関数については 以下で説明する (50-53)。 「学習則」は, モデルのパラメータがどのように更新されるかを記述するものである。 ANN では一般的に, これらの学習則は目的関数を改善するために使用される。 学習則は, 教師あり学習 (エージェントが模倣すべきターゲットを明示的に受け取る) だけでなく, 教師なし学習 (エージェントが何の指示もなしに学習しなければならない) や強化学習システム (エージェントが報酬や罰だけを使って学習しなければならない) にも当てはまります。 最後に 「アーキテクチャ」とは ANN のユニットがどのように配置され, どのような演算を行うことができるかを示すものである。 例えば, 畳み込みネットワークでは, 同じ受容野を入力の空間的範囲に渡って繰り返し適用するという接続パターンが採用されている。

なぜ多くの AI 研究者は，特定の計算機を設計するのではなく，目的関数や学習則，アーキテクチャに注目しているのだろうか。 それは，現実世界の問題を解決するためには，この方法が最も扱いやすいと考えられているからである。 もともと AI の実務家は，初歩的な計算をつなぎ合わせることで知的系を手で設計できると考えていた (54)。 AI セットでの結果は圧倒的なものであった (11)。 現在では，複雑な問題をあらかじめ設計された計算 (例えば，手作りの機能) で解決することは，通常，難しすぎて実際には実行できないことが明らかになっている。 これに対して，目的関数，アーキテクチャ，学習則を指定することはうまくいく。

しかし，高次元データで学習した大規模な ANN では，計算結果の解釈が難しいという難点がある。 ニューラルネットワークは数行のコードで構築でき，ANN の各ユニットに対して，刺激に対する反応や行動との関係を決める方程式を指定することができる。 しかし，学習後のネットワークは，ネットワークが学習した内容を集約した数百万個の重みによって特徴づけられており，このような系を少数のパラメータだけで記述する方法は，言葉では想像できない(55)。

このような複雑性を考慮することは，神経科学にとって有益である。 数十個のニューロンからなる小さな回路であれば，個々の神経の反応や計算のコンパクトなモデルを構築することができるかもしれない (つまり，少数の自由なパラメータや単語を使って通信できるモデルを開発することができる) (3-5)。 しかし, 動物が多くの AI セット問題を解いていることを考えると, 脳は ANN が使う解法と同じくらい複雑な解法を使っていると考えられる。 このことから，神経の反応がなぜそうな

るのかを説明する規範的な枠組みは，神経の反応を，目的関数，学習則，アーキテクチャの相互作用の結果として現れるものと見なすことで得られるのではないだろうか。 このような枠組みがあれば，実際に神経反応をうまく予測する ANN モデルを訓練することができる (29)。 もちろん，そのような ANN モデルは，何百万，何十億，あるいは何兆もの自由パラメータを含む非コンパクトなものになる可能性が高く，言葉ではほとんど説明できない。 したがって，我々が主張したいのは，コンパクトなモデルで神経反応を予測できるかどうかではなく，コンパクトな枠組みで神経反応の出現を説明できるかどうかである。

そこで，動物が遭遇する環境，すなわちデータを神経科学の第 4 の要素に据えるべきかどうかという疑問が自然に湧いてくる。 ある動物の脳セットを決定するためには，その動物の進化的，発生的環境を考慮する必要がある。 自然主義的な刺激を効率的に記述し，倫理的に適切な行動を特定する努力は，神経科学にとって極めて重要であり，神経系の多くの側面を形成してきた。 しかし，我々は，複雑で階層的な脳回路のモデルをどのように開発するかということが重要な課題であり，環境を構成要素の一つとしてではなく，構成要素を固定化するための重要な検討事項としてとらえている。

適切な脳セットが特定されると，最初の疑問は「回路の構造はどうなっているか」である。 これには，細胞の種類とその結合性 (ミクロ，メゾ，マクロ) を記述することが必要である。 このように，脳の回路レベルの記述は，システム神経科学者にとって重要なテーマであることは論を待たない。 回路トレースや遺伝的系統決定のための最新の技術のおかげで，急速な進歩が見られる (56,57)。 しかし，繰り返しになるが，アーキテクチャの理解は回路の理解に十分ではなく，むしろ学習則や目的関数に関する知識によって補完されるべきであると主張したい。

多くの神経科学者は，学習則やアーキテクチャの重要性を認識している。 しかし，学習や進化の過程で脳を形作った目的機能を特定することは，あまり一般的ではない。 アーキテクチャや学習則とは異なり，目的関数は脳内で直接観察できないかもしれない (図 3)。 しかし，特定の環境や課題に依存することなく，数学的に定義することができる。 例えば，予測符号化モデルでは，神経表現を用いて感覚データを符号化するために必要な情報量を表す記述長と呼ばれる目的関数を最小化することができる。 このほかにも，脳を対象とした目的関数がいくつか提案されている (囲み記事 3)。 本展示では，モデルではなく，フレームワークとして提案しているため，これらの目的関数を推奨しているわけではない。 我々の重要な主張の一つは，たとえ推測しなければならないとしても，目的関数は，アーキテクチャや学習規則がどのように計算目標を達成するのに役立つかについての完全な理論の達成可能な一部である，ということである。

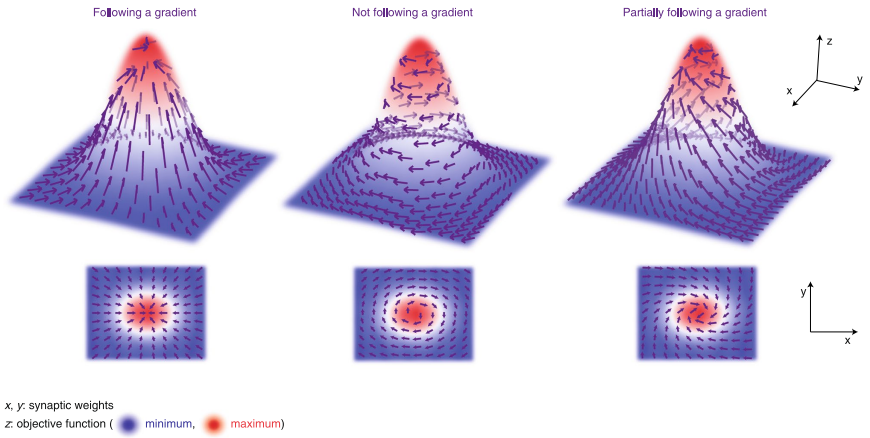

図 3. 勾配に従わない学習規則。 学習は最終的に目的関数で測れるような何らかの改善をもたらすはずである。 しかし，すべてのシナプス可塑性規則が勾配に従う必要はない。 ここでは，シナプス加重空間におけるベクトル場として特徴づけられた 3 つの異なる仮想的な学習則を示すことで，この考えを説明する。 x と y の次元はシナプスの重みに対応し，z の次元は目的関数に対応する。 どのようなベクトル場も勾配とそれに直交する方向に分解することができる。 左は目的関数の勾配に忠実な可塑性則で，系を直接的に最大化させる。 中央は勾配に直交する可塑性則で，そのため系を最大値に近づけることはない。 右側は学習則で，勾配に部分的にしか従わず，間接的にではあるが系を最大値に近づける。 理論的には，これらの状況はいずれも脳内で成立しうるが，学習目標が達成されるのは勾配に完全に従う場合と部分的に従う場合(左と右) だけであろう。

囲み記事 3. 脳に客観的機能はあるのか？

動物には，明らかに基本的な目的関数がある。 例えば，ホメオスタシスは，血中酸素濃度などの生理的変数とその変数の設定値との差に対応する目的関数を最小化する。 このように，ホメオスタシスが生理学的に重要であることを考えると，目的関数は脳が関与しなければならないものであることは間違いない。

しかし，読者の中には，機械学習で使われるような目的関数が脳に関係するのかどうか，疑問に思われる方もおられるかもしれない。 例えば，カテゴリー分類を学習する ANN で用いられる交差エントロピーの目的関数は，脳で使われる可能性は低い。 というのも，この関数は感覚入力ごとに正しいカテゴリを指定する必要があるからである。 しかし，他の目的関数はより生態学的に妥当である。 例えば，予測符号化モデルで用いられる記述長の目的関数 (50)，行動系列の対数確率をそれらが生み出した報酬でスケーリングしたもの (報酬を最大化する強化学習で用いられる)(51)，環境との相互情報の増加 (100)，エージェントが環境を制御できる度合いを測るエンパワーメント (52, 53) などである。 これらの目的関数は，特定のデータセット，課題，環境を気にすることなく，すべて脳に対して数学的に指定することができる。

しかし，神経科学における経験的・理論的モデルに目的関数を結びつけるには，現実的な課題がある。 多くの潜在的な可塑性則は，どの目的関数の勾配にもまったく従わないか，部分的にしか従わないかもしれない (図 3)。 このことは，明らかに問題を複雑にしており，神経可塑性に目的関数が常に関与していることを保証することは不可能である。 また，脳は複数の目的関数を最適化している可能性が高く(40)，そのうちのいくつかは実際に学習する可能性があり (すなわち「学習する」可能性がある。例えば，人間は新しいボードゲームの学習方法を学ぶ)，いくつかは個々の動物ではなく，進化の過程で最適化されてきたかもしれない (すなわち反射や生殖行動など)。

このように複雑ではあるが，システム神経科学においては，目的関数を考慮することが重要であると考えている。 例えば，ドーパミン放出などの生物学的変数は，強化学習による目的関数と有意義に関連することが分かっている(64)。 また，多くの潜在的な学習則は，目的関数の勾配に直接従わないかもしれないが，それでもその目的関数の改善につながるだろう。 ここで，目的関数を特定することで，神経回路の表現型の変化を学習とみなすべきかどうかを確定することができる。 もし，何らかの指標に従って「良くなる」のでなければ，表現型の可塑性を，単なる「変化」ではなく「学習」と呼ぶことができるのだろうか？

この最適化の枠組みには，ANN と同様に，脳のアーキテクチャ，学習則，目的関数は，少なくとも個々のニューロンが行う計算のリストと比較すると，比較的単純でコンパクトであると思われる，という利点がある (58)。 なぜなら，これら 3 つの構成要素は，おそらく限られた情報ボトルネック，すなわちゲノム (ゲノムには大型脊椎動物の脳の配線を完全に規定するほどの容量はないかもしれない(48)) を介して子孫に伝えられなければならないからである。 これに対して，我々が生活する環境は，ゲノムの能力を凌駕するほどの複雑で変化する膨大な量の情報を伝達することができる。

個々の神経細胞の反応は環境によって形作られるため，その計算にはこの巨大な情報源が反映されるはずである。 その証拠に，脳内には，その活動において高いエントロピーを持ち，実験家が今日まで探求してきた多数の刺激や行動と容易に説明できる相関を示さないニューロンが偏在している(6,7)。 我々の主張を明確にするために，3 つの要素を用いて規範的な説明を特定することは，特定のニューロンを活性化するための最適な刺激を決定するために課題最適化深層 ANN を用いる最近の研究 (29) が示すように，回路内のニューロンの応答特性に関するより良い，コンパクトではないモデルの開発に進むための有益な方法であるかもしれないと示唆している。 自然淘汰による進化論は，なぜ種がそのように出現するのかについて，比較的少ない単語で表現できるコンパクトな説明を提供する。 この種の出現に関するコンパクトな説明は，特定の種の系統樹に関するより複雑で非コンパクトなモデルを開発するために利用することができる。 我々が提案するのは，3 つの要素に基づく規範的な説明によって，神経反応の低レベルのモデルを生成するための同様の高レベルの理論を提供することができ，それによって，多くの科学者が求める「理解」の形に一歩近づくことができる，ということである。

神経回路の機能を説明するために，研究者は長い間，目的関数や可塑性規則を仮定してきたことを認識する価値がある (59-62)。 しかし，その多くは，深層学習の鍵である階層的な信用割当問題を避けてきた(15)。 また，予測符号化(31,63)，強化学習 (64,65)，階層的な感覚処理 (17,28) などの実験的成功事例もある。 このように，ここで述べる最適化ベースの枠組みは，個々のニューロン応答特性の研究と並行して行うことができ，また行ってきた。 しかし，この 3 つの中核的な要素に焦点を当てた枠組みがより広く採用されれば，さらに大きな成果が得られると信じている。

3. ウェットラボにおけるアーキテクチャ，学習則，目的関数

ここで紹介した枠組みは，どのように実験と連携させることができるのだろうか？ 進歩のための一つの方法は，3 つの主要要素を用いて作業モデルを構築し，そのモデルを脳と比較することである。 このようなモデルは，理想的にはすべてのレベルでチェックされるべきである。 (i) 検討中の脳セットに含まれる複雑な課題を解くこと， (ii) 解剖学と可塑性に関する我々の知識から情報を得ていること， (iii) 脳で観察される表現と表現の変化を再現することである (図 4)。 もちろん，これらの

基準をそれぞれ確認することは，自明なことではない。 多くの新しい実験パラダイムが必要になるかもしれない。 モデルが与えられた課題を解決できるかどうかの確認は比較的簡単であるが，表象と解剖学的な一致の確認は簡単ではなく，これは活発な研究の領域である (66,67)。 幸いなことに，最適化フレームワークのモジュール性により，研究者は 3 つの構成要素のそれぞれを単独で研究することを試みることができる。

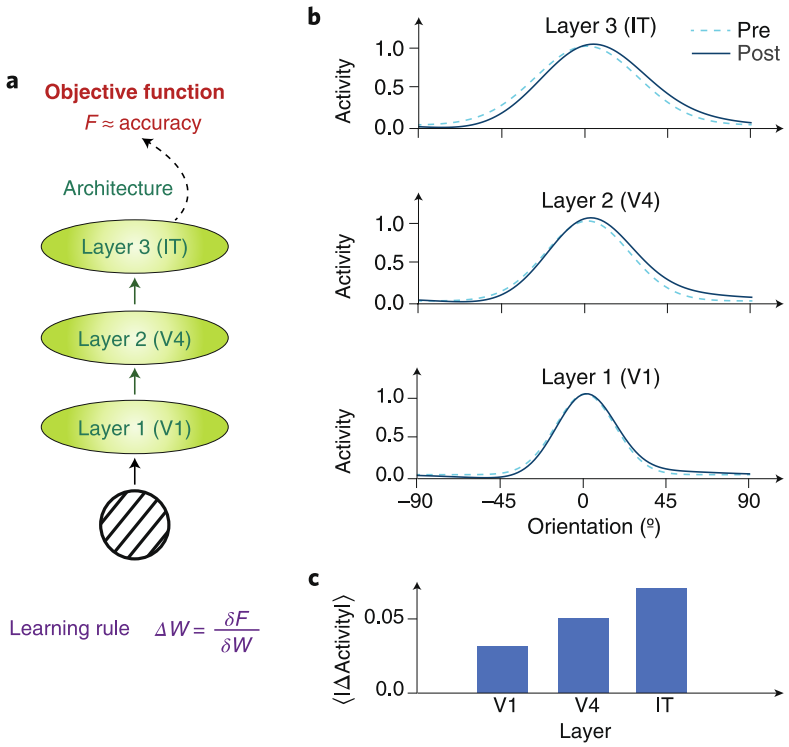

図 4. 深層 ANN モデルと脳を比較する。

3 つの要素を同時に評価する方法の一つとして，3 つの要素を取り入れた深層 ANN の表現の変化と実験データを比較する方法がある。

- 例えば，階層的なアーキテクチャを持つ深層 ANN を用い，グレーティングの向きをうまく識別できたときに送られる報酬を最大化するための目的関数で学習させることができる。 勾配に基づくエンド・ツー・エンドの学習則で学習させる。
- b, 階層の異なる層における集団の方位調整を調べるとき、このようなモデルは予測を行うことができる。 例えば，このモデルは，同調の最大の変化が皮質階層の高い位置 (上) で起こり，その中間の例えば V4 (中) でより小さな変化が起こり，階層の低い例えば V1 (下) で最も小さな変化が起こるはずだと予測することができる。
- c, これは，動物が学習しているときに実験的に観察されるべき神経活動の変化の平均的な大きさについて，実験的に検証可能な予測につながる。

4. 脳内アーキテクチャの実証的研究

脳の帰納的バイアスを定義するアーキテクチャを特定できるようにするためには，回路レベルで神経解剖学を探求する実験を継続する必要がある。 また，神経解剖学を最適化の枠組みの中に組み入れるためには，行動の結果に関する信号がどこから来るのかなど，回路が利用できる情報を特定する必要がある。 最終的には，解剖学のこうした側面を，学習を担う発達過程を導く具体的な生物学的マーカーと関連づけられるようにしたいと考えている。

神経系の解剖学的構造を説明するために，すでにかなりの実験的取り組みが行われている。 我々は，回路の解剖学的構造と発達を定量化するために，様々なイメージング技術を使用している (57,68)。 また，細胞型特異性を持つ神経回路の投射を写像するための広範な作業も行われている (56)。 脳の階層構造を写像しようとする研究は古くから存在するが(69)，現在い

くつかのグループが，深層 ANN 階層のどの部分がどの脳領域を最もよく反映するのかを探っている (17,70)。 例えば，線条体皮質の表現 (例えば非類似度行列で測定) は，深層 ANN の初期層によく一致し，側頭下皮質の表現は後期層によく一致する (8,71)。 また，この種の研究では，例えば，異なる再帰的接続モチーフを探索することによって，脳内の表現ダイナミクスに近い形で深層 ANN のアーキテクチャを最適化することも行われている(66)。 これまでになされた，あるいはこれからなされるであろう解剖学的観察の惑わしさに直面したとき，解剖学を目的関数や学習規則と一緒に置く理論と枠組みとは，最も説明力のあるこれらの特徴に照準を合わせる方法を提供するだろう。

5. 脳内学習則の実証研究

神経科学の分野では，シナプス可塑性の規則を研究する長い伝統がある。 しかし，これらの研究では，信用割当がどのように行われるかをほとんど研究していない。 しかし，前述したように (囲み記事 1)，信用割当は ANN における学習の鍵であり，脳においても同様であろう。 ありがたいことに，最近のシナプス可塑性の研究では，トップダウンフィードバックと神経調節系が焦点となっている(72-76)。 これにより，例えば，頂端樹状突起がどのように信用割り当てに関与しているか (12,14)，あるいは，神経調節因子と組み合わせたトップダウン型注意機構が信用割り当て問題をどのように解決するか (37,38) など，いくつかの具体案が出てきた(図 5)。 また，表象の変化を観察し，その観察結果から可塑性の規則を推測することもできるかもしれない(77)。 学習中に表現がどのように進化していくかを捉えるために，実験者は動物が安定した性能に達している最中と後の両方で神経応答を測定することが重要である。 信用割当を視野に入れた学習則の研究は，可塑性に影響を与える無数の要因について，よりきめ細かな理解をもたらしている(78)。

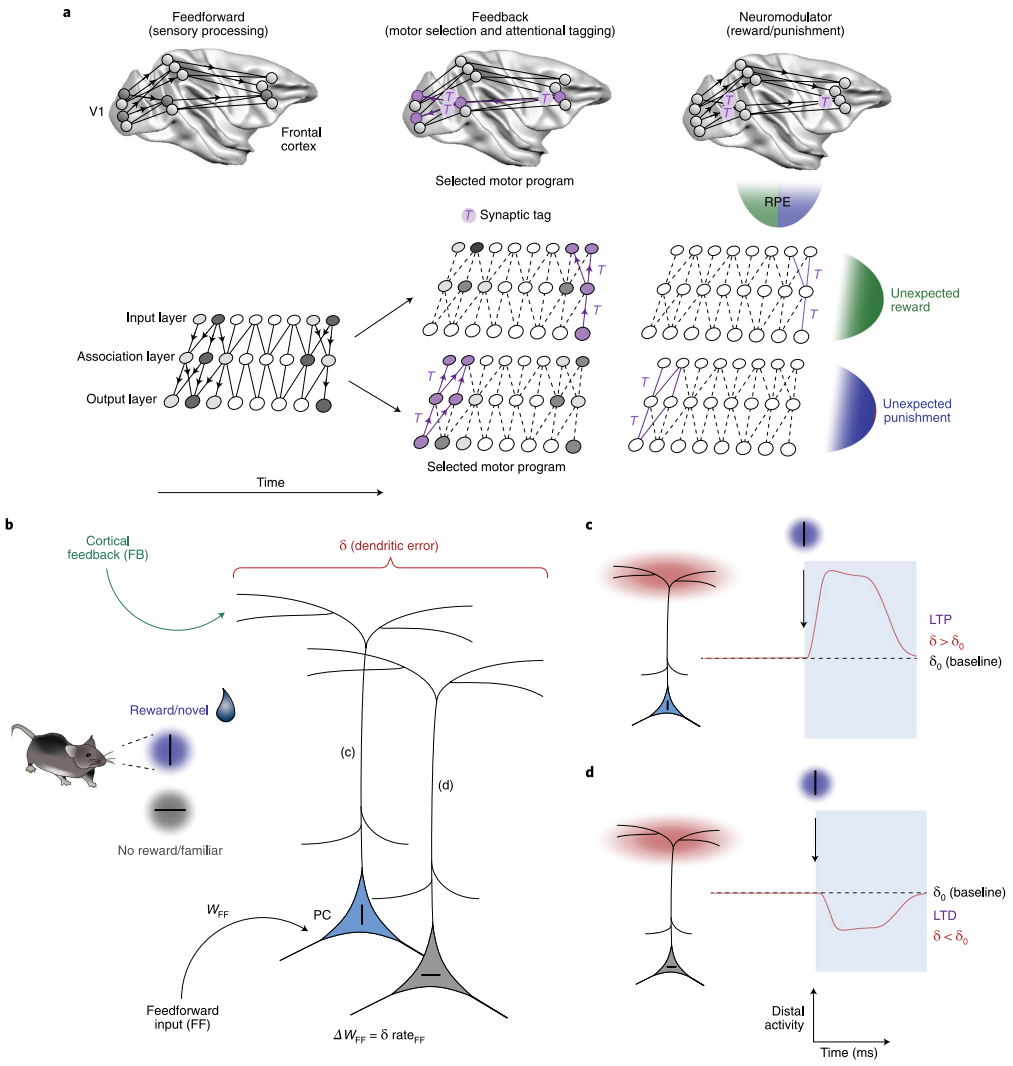

図 5. 信用割当の生物学的モデル

- (a) 注意に基づく信用割当モデル(37,38) は，信用割当問題が，注意と神経調節信号を用いて脳によって解決されると提唱している。 これらのモデルによると，感覚処理は初期段

階では大部分がフィードフォワードであり，次にフィードバックがニューロンやシナプスに単位を「タグ付け」し，報酬予測誤差 (RPE: reward prediction errors) が可塑的変化の方向を決定する。 下図では，円形がニューロンを示し，灰色の色調はその活動レベルを示している。 これらのモデルは，特定の出力単位の活性化を担当するニューロンが，注意のフィードバックによってタグ付け (T) されることを予測する。 そして，正の RPE を受けた場合，シナプスが増強されるはずである。 一方，負の RPE を受け取った場合，シナプスは抑制されるはずである。 b-d, 信用割当の樹状突起モデル (12,14) では，錐体細胞の頂端樹状突起の「樹状突起誤差」信号によって勾配信号を伝達することを提案している。

- (b) これらのモデルによれば，フィードフォワードの重み更新はフィードフォワード入力と δ の組み合わせによって決定される。 2 種類の刺激を提示し，片方だけを強化する実験では，具体的な予測につながる。
- (c) 強化された刺激に対してニューロンが同調する場合，強化によって頂端の活動が増加するはずである。
- (d) 反対に，強化されない刺激に同調している神経細胞は，強化されると頂端の活動が減少するはずである。

6. 脳における客観的機能の実証的研究

場合によっては，脳が最適化する目的関数が神経信号として直接表現され，それをモニターし記録することができるかもしれない。 また，シナプスの更新を支配する可塑性規則に関してのみ，暗黙のうちに目的関数が存在する場合もある。 最適制御のような規範的な概念が適用可能であり (80)，進化的な考え方が我々の思考に影響を与えることができる。 より具体的には，動物がどの機能を最適化することが有用であるかという指針を倫理学から得ることができ (81)，目的関数について考えるための有意義な直観的空間を提供することができる。

実験データを目的関数に関連付けようとする文献が長年に渡って存在する。 これは既知の可塑性規則と潜在的な目的関数とを関連付ける理論的研究から始まる。 例えば，実験的に観測された神経活動と自然な情景で学習した ANN の神経活動を比較することで目的関数を推定しようとする研究がある(59,82)。 また，逆強化学習を用いて系が最適化するものを特定するアプローチもある(83)。 さらに，ある目的に対して最適化された表現幾何学と実際の神経表現幾何学との相関を調べることで，目的関数を把握することができると主張することもできる(28,84)。 また，ブレイン・コンピュータ・インターフェイス装置を制御する際に，動物の回路が何を最適化できるかを問うアプローチも新たに登場した(85)。 このように，過去の研究成果(80) を基にした文献が増えつつあり，脳における目的関数の探索に役立っている。

7. 注意点・懸念点 Caveats and concerns

アーキテクチャ，学習則，目的関数に焦点を当てると，ニューロンの符号化特性の研究から遠ざかり，方位選択性，周波数同調，空間同調 (場所細胞，グリッド細胞) など，これまで学んできたことの多くが失われる，という反論があるかもしれない。 しかし，我々が提案する枠組みは，このような知見に大いに助けられている。 畳み込み ANN は，視覚系の複雑な細胞の観察から直接生まれた (86)。 さらに，同調曲線はしばしば学習実験の文脈で測定され，同調の変化は学習規則や目的関数について我々に情報を与えてくれる。

同様に，計算論的神経科学の多くは，神経活動のダイナミクスのモデルを重視しているが (87)，これは我々の議論では主要なテーマになっていない。 そのため，我々の枠組みはこの過去の文献と接続できていないのではないかと心配されるかもしれない。 しかし，我々がここで明確にした枠組みは，ダイナミクスを考慮することを排除しているわけではない。 ダイナミクスに注目することは，アーキテクチャ，学習則，目的関数に関する推論に再利用することも同様に可能であり，これらは長い間，神経ダイナミクスのモデルの特徴であった(49,88)。

神経科学と深層学習の関連性についてのもう一つのよくある反論は，動物が行う多くの行動は比較的学習を必要としないように見えるというものである (48)。 しかし，そのような生得的な行動は，進化のタイムスケールでのみ，実際に「学習」されたものである。 なぜなら，あらかじめ組み込まれた行動でさえ，学習によって修正されることがあるからだ (例えば，馬は生まれた後も走るのが上手になる)。 したがって，神経回路が適度な量の学習しか行わない場合でも，最適化の枠組みを用いることで，その動作をモデル化することができる (48)。

ここで示した枠組みは，目的関数の最適化を脳のモデルの中心に据えたものである。 しかし，どのような脳でも，包括的な理論を構築するためには，目的関数の最適化とは無関係な他の制約に注意を払う必要がありそうだ。 例えば，生理学の多くの側面は，進化上の祖先から受け継いだ系統的な制約によって決定される可能性がある。 これらの制約が神経科学のモデルにとって重要であることは間違いないが，我々は，これらの制約の中で目的関数を最適化することが，脳で観察される神経回路と行動の豊かな多様性を生み出すと信じている。

ボトムアップ的なアプローチで脳を理解しようとする私たちの中には，脳に客観的な機能や学習則を求めるのは時期尚早ではないか，現在の我々が持っているよりもはるかに詳細な脳の働きが必要ではないか，と心配する人もいる。 しかし，科学的な問いは，必ず何らかの思考の枠組みの中で提起されるものである。 重要なことは，ボトムアップの説明を放棄せよということではない。 むしろ，ANN が示唆する枠組み (図 5) から，新しい重要な実験的疑問が生まれることを期待している。

最後に，研究者の中には，深層 ANN のパラメータの多さを，オッカムの剃刀の違反であり，単なるデータへの過学習であると見て，懸念している者もいる。 興味深いことに，AI における最近の研究は，膨大な数のパラメータを持つ学習系の振る舞いが直感に反することがあることを示している。 つまり，パラメータを過剰に持つ学習系には，優れた汎化を可能にする数学的特性が内在しているようだ (42, 89)。 脳そのものが膨大な数の潜在的適応パラメータ (例えばシナプス結合，樹状突起のイオンチャネル密度など) を持っているらしいので，深層 ANN のパラメータ数が多いほど，脳のモデルとしてより適切であると言えるのかもしれない。

8. 結論

システム神経科学の多くは，脳内の個々の神経細胞の機能について簡潔な記述を試みてきた。 このアプローチは，いくつかの (比較的小さな) 回路と特定のハードワイヤーされた行動を説明するのに成功している。 しかし，数千，数百万，数十億のニューロンを持つ可塑的な回路の優れたモデルを開発するためには，このアプローチを他の知見で補完する必要があると考える理由がある。 残念ながら，中枢神経系にある個々のニューロンの機能を，人間が解釈しやすく，言葉で表現できるような形に圧縮できる保証はない。 現在のところ，深層 ANN の個々のユニットの機能を言葉にする良い方法はなく，また，実際の脳はより複雑である可能性が高いことを考えると，システム神経科学は，ANN 研究プログラムで成功している種類のモデル，すなわち，3 つの基本要素に基づくモデルに焦点を当てることが有益であることを提案する。

システム神経科学の現在の理論は美しく，洞察に富んでいるが，最適化に基づいた首尾一貫した枠組みの恩恵を受けることができると我々は考えている。 例えば，ヘブのような局所可塑性則は，多くの生物学的データを説明することができる。 しかし，複雑な課題で優れた成績を発揮するためには，ヘブ則は目的関数とアーキテクチャを考慮して設計されなければならない (34,90)。 同様に，他の研究者は，正当な理由によって，脳が用いる帰納バイアスの利点を指摘している(48)。 しかし，AI セットや様々な脳セットのような複雑な課題を解くには，帰納バイアスだけでは十分ではない。 これらの難題を解決するためには，帰納バイアスは学習や信用割当と対になる必要がある。 これまで述べてきたように，動物が解くことのできる課題の集合が神経科学にとって不可欠な考慮事項であるならば，これらの課題を実際に解くことのできるモデルを構築することが極めて重要である。

システム神経科学の進歩のためには，ボトムアップの記述的研究とトップダウンの理論的研究の両方が必然的に必要とされる。 しかし，正しいトップダウンの理論的枠組みから始めることが重要である。 現代の機械学習が AI セットや数多くの脳セットの問題を解決できることを考えると，機械学習の知見でシステム神経科学研究のトップダウンの枠組みを導くことは実りあることであろう。 このような考え方が提供する枠組みの中で研究データを考察し，ここで明らかにした 3 つの本質的な要素に注意を向けるならば，現在の神経科学の技術革新の恩恵を最大限に享受できる脳の理論を開発できると信じている。