東京大学教養学部哲学・科学史部会　哲学・科学史論叢第二十四号　令和 4 年 2 月（1–22）

# チャーチランドの神経計算論的理論観
## ―その意味論的基盤の問題点―

藤原　諒祐

## 1.　はじめに：自然化された認識論，脳，コネクショニズム

我々の知的実践のあり方を解明する上で，認知についての科学的知見を頼りにするのは 1 つの有効な手である．特に，認知にとって重要な器官である脳のはたらきを明らかにしてきた神経科学の知見は，知識の獲得や取捨選択の現実的なメカニズムの解明にとって有用かもしれない．ポール・M・チャーチランドはまさにこのような方針に従った認識論･科学哲学を展開している．彼は，ニューラルネットワーク（コネクショニストネットワーク）モデルに基づいて脳による知覚や学習，概念把握のあり方を論じ，その上で知識や科学理論が脳においてどのように実現されているかを明らかにしようとする．

ニューラルネットワークは，複数のユニット（ニューロン）からなるいくつかの層と，隣接する層のニューロン同士の結合からなる．ネットワークでは，重み付き結合を通して入力層から出力層へと活性パターンが伝わっていく．

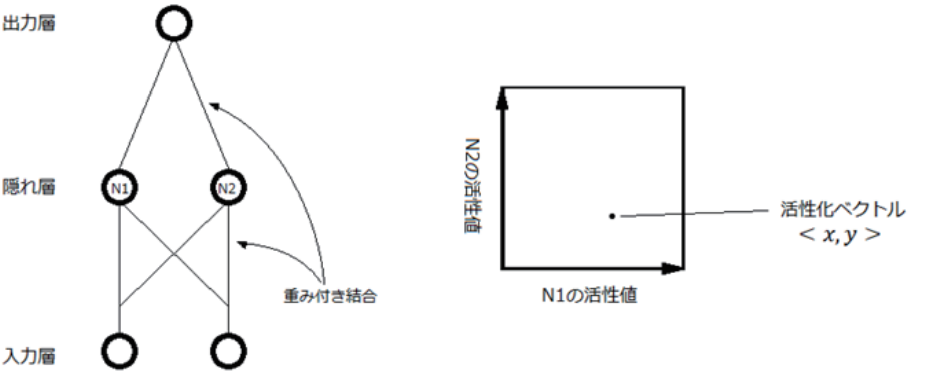


**図 1**　シンプルなネットワーク（左）と隠れ層の活性化空間（右）
（Churchland, 2012, p. 39, Fig. 2.2 を参考に作成）

ある層の活性パターンは「活性化空間」（その層の各ニューロンの活性値が各次元に対応する空間）上の点（「活性化ベクトル」）として表わすことができ，ネットワークの情報処理は各層の活性化ベクトルの非線形変換過程として捉えられる（**図 1**）(Churchland, 2012, pp. 38–50)．

こうしたモデルのもとで，チャーチランドは，広範な認知活動を活性化ベクトルの変換過程として統一的に捉えようとする．例えば，彼は様々なタイプの説明的理解を，ネットワークによる知覚的判断と類比的に，ベクトルの活性化として考える「説明の統一的理論」(Churchland, 1989) を提示する．また，運動技能の獲得もネットワークの学習により可能になることから，彼は命題知／技能知の区分は何ら重要なものではないと指摘する (Churchland, 2012, pp. 45–50)．さらに，非言語的なモデルを用いることで，言語をもたない動物たちと我々のような人間の双方に適用できる認知の説明を得ることができると彼は考える [1](Churchland, 2012, pp. 4–6)．このように，様々な主体による様々な種類の認知活動を統一的に説明できるならば，それは大きな利点である．

しかし，チャーチランドの認識論は本当に我々や動物の認識のあり方を適切に捉えるものになっているのだろうか．本稿の目的は，この疑問に否定的に答えることである．特に，我々が有する理論についての彼の見方には問題がある．このことを示すために，2 節で彼の認識論・意味論的見方を概観し，3 節で，彼の理論観が下敷きにしている概念の意味論に問題があることを論じる．彼の認識論は，その土台となる意味論に問題があるために，問題があるのだ．もちろん，これらの議論はコネクショニズムや神経科学的知見の認識論的応用が不可能ないしは不毛であることを示すものではない．むしろ，チャーチランドのモデルの問題点を指摘することで，コネクショニズムや神経科学と親和的な認識論アプローチの適切な方向性を示唆するものである．

## 2. チャーチランドの認識論と意味論

### 2.1 コネクショニズム的な認識論

科学理論や科学的説明がどのようなものかについては，既に様々な議論がなされている (Winther, 2021; Woodward & Ross, 2021)．世界のあり方を記述し，現実の諸現象に説明を与えるものとして科学理論を捉える点で，チャーチランドの理論観は古典的なものと重なる．しかし，コネクショニズムの枠組みで理論や説明の本性について捉えようとする点に彼の議論の独自性がある．以下では，説明と理論についてのチャーチランドの考えを順に見ていく．

チャーチランドは説明の本性に接近するために，説明の認知的側面，すなわち，問題となる状況の説明が当の状況の理解を促すという点に着目する．チャーチランドによると，説明による理解（「説明的理解」）の本質は，「問題となるケースを，包括的なタイプ，すなわち，それについて詳細で情報の十分含まれた表象を生物が有するようなタイプの一例として捉えること」(Churchland, 1989, p. 210) にある．ある事象のプロトタイプと紐付けることで，目下のケースをよく知っている事象の一例とみなすことができ，それによって，当のケースについての良い理解が得られるということだ．彼の提示する「説明のプロトタイプ活性化 (prototype-activation) モデル」[ 以下 PA モデル ] は，説明的理解についてのこのような見方にもとづいて説明を捉えるものである．このモデルの下では，様々なタイプの説明がプロトタイプの活性化として統一的に理解される．

説明に用いられるプロトタイプは，ネットワークの出力層と入力層の中間にある「隠れ層」の活性化空間によって提供される．例えば，ソナーによって海中の機雷と岩とを見分けるネットワークの隠れ層の活性化空間は，次のような構造になっている (Churchland, 1989, pp. 202–204)．ネットワークの訓練後，活性化空間は機雷に対応する領域と岩に対応する領域とに分割される．そして，それら領域内部には，それぞれプロトタイプ的な機雷と岩に対応するベクトルの領域が存在する（**図 2**）[2]．このように，活性化空間上には，学習対象とな

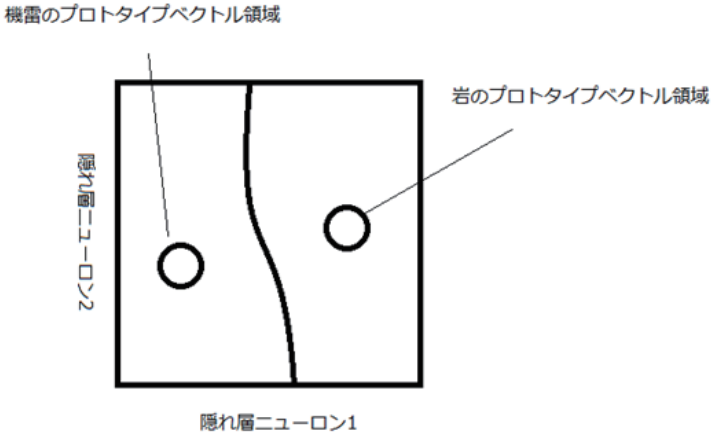


**図 2**　機雷識別ネットワーク隠れ層の活性化空間
（Churchland, 1989, p. 203, Fig. 10.2 を参考に作成）

るカテゴリーのプロトタイプとしてはたらくベクトルが存在する．これら「プロトタイプベクトル」を活性化させることで，生物は状況を識別・理解するのである (Churchland, 1989, p. 208)．これは説明的理解の場合も同じである．例えば，茂みから出ている紐に対して “ ネズミ ” プロトタイプベクトルが活性化することで，それがネズミの尻尾だとわかったり，木星に対して “ 回転する塑性体 ” プロトタイプベクトルが活性化することで，木星表面の縞模様についての説明的理解が得られたりする (Churchland, 1989, p. 211, Fig. 10.3)．したがって，「説明的理解とは，良く訓練されたネットワークの特定のプロトタイプベクトルを活性化させること」(Churchland, 1989, p. 210) なのだ [3]．

チャーチランドの理論観は，以上のような説明モデルに対応するものになっている．いくつものプロトタイプベクトルが配置された活性化空間は，あるカテゴリー群についての概念枠組みとして捉えられる．例えば，顔を識別するネットワークの活性化空間は，人間の顔についての概念枠組みである (Churchland, 2012, p. 74)．このような活性化空間はわれわれの脳内に無数に存在する．そして，それぞれの活性化空間が対象とするカテゴリー群は多岐にわたり，人間の顔についての空間だけではなく，“ 人間の声 ” 空間や “ 楽器 ” 空間もあれば，

リンネの分類体系に沿った“動物”空間や，天動説的な，あるいは地動説的な“天体”空間もありうる (Churchland, 2012, p. 77)．学習によって形作られた活性化空間は，日常的事物についての概念枠組みだけでなく科学理論としても捉えられるということだ．チャーチランドの言葉を使えば，理論とは「塑造された活性化空間 (sculpted activation-space)」（Churchland, 2012, p. 218[ 強調原文 ]）[ 以下 SAS] なのである．

ただし，次節で見るように，次元構成の異なる活性化空間同士も理論や概念枠組みとしては同一でありうる．その意味で，理論や概念枠組みと活性化空間は同一のものではない．活性化空間は理論や概念枠組みを「具体化している (embody)」(Churchland, 2012, p. 45)，というような関係がより正確な記述となるだろう．

## 2.2 コネクショニズム的な概念の意味論

前節では活性化空間は，概念枠組みないし理論として，対象カテゴリー群にかかわる世界のあり方を捉えていることを見た．活性化空間の内部では，プロトタイプベクトルを中心とする特定領域が概念としての役割を果たしている (Churchland, 2007, p. 144–145)．こうした領域は，概念の学習の中で形成されていく．ある概念をもつということは，特定の領域内部に位置づけられるものとして，関連性をもつ個々の事例を表象する能力をもつことなのである．そして，このようにして捉えられた概念の内容ないし意味内容は，「世界の何らかの側面の非常に特異的な描写 (portrayal)」（Churchland, 2007, p. 135[ 強調原文 ]）であるとチャーチランドは考える [4]．それでは，活性化空間（そしてその内部領域）はいかにして外界を表象する描写となるのか．そして，個人がもつ概念の意味内容の同一性や類似性についての基準はどのようなものなのか．彼が提唱する概念の意味論，「ドメイン描写意味論 (Domain Portrayal Semantics)」[ 以下 DPS] は，このような疑問に答えるものである [5]．

活性化空間における領域が概念としてもつ意味内容（外界を表象するあり

方）は個々独立に定まるものではなく，同じ空間上の他の要素との距離関係によって定まる．チャーチランドはこのことを活性化空間と地図との類比を用いて説明する (Churchland, 2012, pp. 74–77)．地図上の点や線が特定の地域の建造物や道を表わすことは何によってわかるのか．その手がかりとなるのは，他の点や線との位置関係である．つまり，地図における他の点や線と当のアイコンとの位置関係が，地図の対象地域にある建造物や道路と特定の建物との位置関係を写し取ったものとなっていることから，そのアイコンがその特定の建物を表わすことがわかるのである．したがって，地図上の点や線は，地図全体から独立に何かを表わすわけではなく，対象地域全体を描いた地図上で他の点や線と特定の位置関係をとることで特定の建造物や道を表わすのだ．活性化空間と空間上のベクトルについても同様に考えられる．地図が対象地域に含まれる対象の位置関係を地図上の位置関係に写し取ることで対象地域全体を描写するように，活性化空間は対象「ドメイン」に含まれる対象（例えば特定の顔）の類似関係を空間上の位置関係（距離）に写し取ることで対象ドメイン全体を描写する（**図 3**）．そして，地図上の点や線と同じように，活性化空間上の個々の点は他の点と特定の位置関係をとることで特定のカテゴリーに対応づけられるのである[6]．

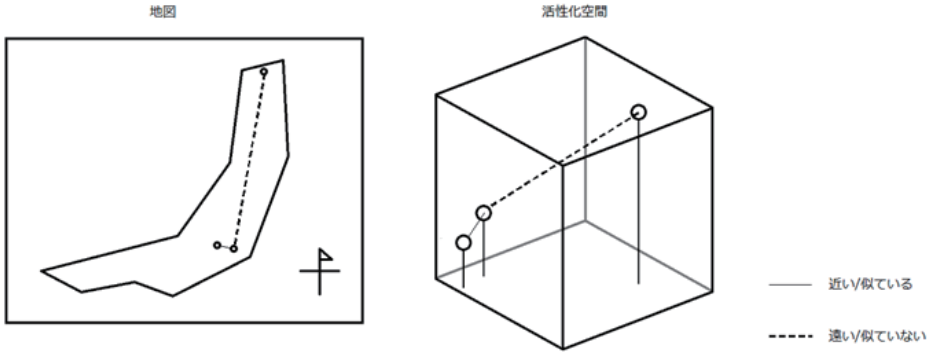


**図 3** 地図と活性化空間のアナロジー
（Churchland, 2012, pp. 74–77 の記述をもとに作成）

ここで，活性化空間の次元，すなわち，隠れ層の各ニューロンが反応する特徴（「選好刺激」）がそれぞれの概念を構成しているわけではないことに注意したい (Churchland, 2012, pp. 85–87)．例えば，顔識別ネットワーク隠れ層の各ニューロンの選好刺激は，入力の顔画像と同じぐらい複雑なパターンであり，特定カテゴリーの構成要素とみなせるようなものではない．また，ネットワークの訓練ごとに，活性化空間の分割のされ方が同じでも，ニューロンの選好刺激は全く異なっていることがある．2 つの活性化空間の内的構造は同じでも，それぞれが鏡映しになっていたり，向きが異なっていたりするのである．しかし，こうした差異は最終的な顔識別のあり方には無関係なのだ[7]．

構成する次元のあり方が異なりうる 2 つの活性化空間同士の（概念枠組みとしての）同一性や類似性は次のように決まるとチャーチランドは考える (Churchland, 2012, pp. 104–112)．2 つの地図がある場所を同じように写した地図であるかどうかは，回転や縮小・拡大によってそれらを重ね合わせられるか試せばわかる．同じように，2 つの活性化空間があるドメインのあり方を同じように写し取っているかどうかは，それらを重ね合わせられるか試せばわかるのである．もちろん，実際に求めたいのは完全に同一ではない 2 つの空間の類似性かもしれない．この場合は，完全な重なり合いではなく，2 つの空間上の対応する要素同士の距離の和が最小になるような対応関係を見つけることになる，そして，空間内の 2 点（プロトタイプベクトル）を結ぶ辺について，空間同士で対応する辺の長さ（それぞれ $L_1$，$L_2$）の和と差の割合の平均をとることで類似性を測ることができる．つまり，次の式で類似性は表わされる．

$$\mathrm{Sim} = 1 - \mathrm{avg.}[(|L_1 - L_2|) \div (|L_1 + L_2|)] \qquad \text{(Churchland, 2012, p. 113)}$$

さらに，ある空間の一部と別の空間との類似性を測ることで，空間同士の部分的な重なり合いや詳細さの差異を測ることができる[8](Churchland, 2012, p.113)．このように，個人間での概念枠組みの同一性・類似性や重なり合いは，空間を構成する次元のあり方ではなく，活性化空間の同型性を基準に測られるのである[9]．

以上の議論は，活性化空間としての理論にも同様に適用される．例えば，理論間還元は，活性化空間同士の包摂関係として捉えられる (Churchland, 2012, pp. 210–212)．チャーチランドによれば，理論 T が包括的な理論 G へと還元されることの必要十分条件は，T に対応する活性化空間のプロトタイプ配置と，G に対応する活性化空間のプロトタイプ配置の一部とがおおよそ「準同型である (homomorphic)」(Churchland, 2012, p. 211[強調原文])ことである．このように，理論間の比較も，活性化空間の同型性や部分的重なり合いを測ることで可能になる．

以上のことから明らかに，理論の SAS 的見方は，DPS を基盤にした考えだといえる．第 1 に，DPS は活性化空間をなぜ理論と呼べるのかを説明する．活性化空間を理論と呼べるのは，DPS が記述するような仕方で活性化空間が世界を表象しているからである．第 2 に，より重要なことだが，DPS は，理論間の同一性・類似性の基準を提供する．DPS が提示する概念枠組みの同一性・類似性の基準，すなわち空間的同型性という基準は，そのまま理論の同一性・類似性の基準ともなるのだ．つまるところ，DPS が概念の意味内容や概念枠組みの比較について語ることは，理論的概念の意味内容や理論間の比較の場合にも全く同様に適用される．この点で概念枠組みと理論に差異はないのだ．
DPS と SAS 的見方の依存関係からは，DPS に問題があるならば，理論の SAS 的見方にも問題があるということがいえる．したがって，DPS の問題点を検討することで，本稿の目的である SAS 的見方の妥当性の検討が可能になる．

## 3. ドメイン描写意味論の問題

### 3.1 DPS への批判

3 節では，DPS が適切な意味論として機能するのか考察することで，理論の SAS 的見方の妥当性を検討する．そのためにまず，プリンツによる DPS 批判 (Prinz, 2006) をとりあげる．この批判は科学理論の捉え方の検討という文脈でなされたものではないが，SAS 的見方の基盤としての DPS に潜む問題を明ら

かにする上で，検討する価値がある．

2.2 で見たように，DPS において，活性化空間を構成する次元のあり方（対応するニューロンの選好刺激）は，概念枠組みの類似性に無関係とされる．こうした想定の問題について，プリンツは次のように論じる (Prinz, 2006, pp. 97–100)．心理学実験において被験者に課せられた，概念と結びつく特徴のリストアップ課題の結果は，概念使用のあり方（典型性評価やカテゴリー判断）と相関する [10]．これは，こうした特徴が「心理学的に実在し，かつ因果的に効能を有する」(Prinz, 2006, p. 97) こと，また，それら特徴は概念から取り出せることを示唆する．また，特定の特徴を有する／欠く対象（例えば，耳のない犬）を我々は想像できるので, こうした特徴は「独立に操作可能」(Prinz, 2006, p. 98) だといえる．つまり，我々が有する概念は「操作可能」な特徴から構成されているのだ．それゆえ，概念の類似性もこうした特徴を基準に測られるべきである．そのために必要なのが，概念を構成する特徴として空間の次元に対応する特徴を同定する方法，すなわち「意味論的空間の次元にラベルをつける方法」(Prinz, 2006, p. 99) なのである．

プリンツの批判に対しては次のような 2 段階の応答が考えられる．第 1 に，DPS は，ネットワークの各次元が具体的特徴に対応する可能性を排除しない．例えば，顔識別ネットワーク隠れ層の活性化空間のある次元は瞳孔と眉の距離に反応するかもしれない (Churchland, 2012, p. 89)．分類や識別に役立つ具体的特徴に対応する次元が存在するならば，我々の概念使用においてそうした特徴が重要性をもつことは，DPS と矛盾しないかたちで説明できる．

ある次元が具体的特徴に対応することを認めてしまうならば，結局プリンツの議論に与することになるのではと思われるかもしれない．しかし，ある次元が具体的特徴に対応することは，これら特徴が概念の意味内容の同一性や類似性の基準になることとは別である．それゆえ，次元と具体的特徴の対応関係を認めることと，概念の意味がこれら特徴と独立であるとすることは両立可能なのだ．これが第 2 の応答である．

この応答は，認識論と意味論の区別についてのチャーチランドの議論にもと

づいている．チャーチランドは，実際のベクトル活性化がどのようになされるかについて，すなわち，彼が言う「ミクロな認識論」への関心と，ベクトルがいかにして意味内容をもつかについて，すなわち，彼が言う「意味論」への関心とを区別し，次元が対応する特徴の同定は認識論にとって有意義かもしれないが，意味論にとってはそうではないと論じる (Churchland, 2012, p. 87)．次元と特徴の対応関係は個々のネットワークのベクトル活性化のあり方を左右するが，外界の描写としての概念枠組みや概念の意味内容の同一性や類似性には無関係ということだ．例えば，“ 瞳孔と眉の距離 ” 次元を含む空間と同型であるが，“ 瞳孔と眉の距離 ” に対応する次元をもたない空間があるかもしれない (Churchland, 2012, p. 89)．前者を実現するネットワークと後者を実現するネットワークは異なる特徴に反応するため，そのベクトル活性化メカニズムは異なる．それでも，2 つの空間が同じ概念枠組みであり，それぞれが活性化させるベクトルは同じ意味内容をもつということがありうるのだ．

このことは，感覚器官のはたらきが異なっていても同じ概念枠組みの共有が可能だと考えるならば，より理解しやすいだろう．例えば，色覚をもつ人の “ 鳥 ” ドメインについての活性化空間を構成する次元の中には，色覚的特徴に対応するものがあるかもしれない．しかし，色覚をもたない人や動物，そしてそもそも視覚に頼らないコウモリも，色覚をもつ人と同じように，他の鳥とは違うものとしてフラミンゴを同定できるかもしれない (Churchland, 2012, p. 89)．つまり，次元と特徴（感覚的刺激）との対応関係が異なっていようと，“ 鳥 ” についての同じ概念枠組みをもつことができるのだ．概念枠組みの同一性にとって本質的なものは，次元のあり方とは無関係に測定可能な空間の同型性（プロトタイプベクトルの位置関係）なのである．それゆえ，主体がどのような特徴を概念と結びけるかということは，概念の意味内容の同一性基準を与えるという課題にとっては無関係ということになる．

もちろん，こうした応答の妥当性には議論の余地があるだろう．特に，DPS が説明的理解や科学理論についての考えを下支えしていることを考えると，認識論と意味論の分離戦略には問題があるように思われる．DPS 単体で考えると，

主体がある概念にどのような特徴を結びつけるかということが概念の意味内容の類似性・同一性に無関係となることに問題はないかもしれない．しかし，次節で見るように，概念からいかに明示的特徴を取り出すかということと無関係に，我々の説明実践を説明することはできないのである．

### 3.2 SAS 的見方の基盤としての DPS の問題点

前節の議論では，DPS の妥当性は，説明の PA モデルや理論の SAS 的見方とは独立に論じられている．しかし，2.2 で見たように，DPS はチャーチランドの認識論的モデルの基盤となっている．ところが，以下で見るように，DPS はこうしたモデルの基盤として適切に機能していない．そしてこのことは同時に，チャーチランドの認識論がその意味論的基盤に問題を抱えていることを示している．

次のようなケースを考えよう．水族館にやってきた S1 と S2 は水槽の中にいる生物がなぜ背泳ぎをしているのか説明しようしている．ここで，両者は“水生生物”ドメインについて同型の活性化空間をもつとする．また，S1 と S2 はその生物を見たとき，DPS において同じ内容とされるベクトル，すなわち“カブトガニ”プロトタイプベクトルを活性化させる．つまり，両者はともにその生物をカブトガニとして見ている．

説明の PA モデルに則れば，「a が F である」ことの説明がなされるためには，a は特徴 F をもつようなカテゴリーのプロトタイプないし概念と紐付けられる必要がある．このケースにおいて S1 と S2 はともにこの条件を満たしている（実は，カブトガニは背泳ぎをするのである）．しかし，この条件は説明の成立条件としては不十分である．“カブトガニ”概念から，「背泳ぎをする」という特徴を取り出すことができなければ，その生物が背泳ぎをすることを説明できないからだ．

ここで，DPS の枠組みでは，S1 と S2 は同じ対象に同じ意味内容の概念を紐付けているにもかかわらず，S1 は「背泳ぎをする」という特徴を取り出すこ

とができて，S2 はできない，ということが起こりうる．なぜならば，前節でみたように，DPS によれば，ある概念とどのような特徴が結び付けられるかということは，概念の意味内容の同一性の基準からは切り離されているからだ．今考えているケースにおいて，S1 と S2 は同型の活性化空間をもっている（それゆえ同じ意味内容の概念を有している）が，S1 の活性化空間には泳ぎ方の特徴に対応する次元が含まれており，S2 の活性化空間には含まれていない，ということがありうる（**図 4**）．このとき，S1 と S2 では概念から取り出せる特徴が異なることになる．

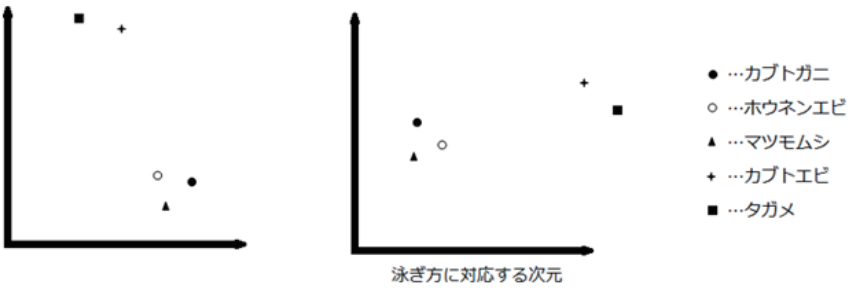


**図 4**　同型だが，次元のあり方は異なる “ 水生生物 ” 空間

しかし，そうだとすると，DPS に則れば S1 と S2 は同じ意味内容の “ カブトガニ ” 概念を有しているにもかかわらず，S1 は「背泳ぎをする」という特徴を取り出せるが, S2 はその特徴を取り出せないということになる．そして,「背泳ぎをする」という特徴を取り出せなければ,「この生き物が背泳ぎをする」ということを説明できない．つまり，S1 と S2 は説明の成否について異なっているということになる．

さて，2.2 で見たように，DPS は SAS 的見方の基盤となっている．理論の同一性や還元は，DPS のもとでの概念枠組みの同一性や重なりの基準によって

捉えられるのである．したがって，同型の活性化空間をもつ S1 と S2 は，あるドメインについての同じ理論をもっていることになる．しかし，上の想定に従えば，同じ理論をもっているにもかかわらず，S1 ができる説明を S2 はできないということになってしまう．もちろん，説明の成否についての不一致も多少は認められるだろう．しかし，S1 と S2 はあるドメインの事例に同一の意味内容の概念を紐付けるが，S1 がその概念から取り出す特徴のほとんどを S2 が取り出せない，という状況も考えられる．このとき，S1 に可能な説明のほとんどが S2 には不可能な説明ということになるが，DPS を下敷きにした SAS 的見方に従えば，両者は同じ理論をもっていることになるのである．このような状況を許容してしまう見方には明らかに問題がある．なぜならば，理論の同一性にとって説明の成否が無関係だとすると，ある理論が対象ドメインについての諸現象に説明を与えるということを説明できないからだ．したがって, DPS は, SAS 的見方の基盤として役割を果たさないのである．

ここで，次のように応答できるかもしれない．たしかに，DPS に則れば，概念とどのような特徴が結びつくかは概念の意味内容の同一性に無関係である．しかし，同型の活性化空間をもつ人ならばその概念に結びつけることができるような特徴はあるかもしれない．そして，説明的理解はそうした特徴を取り出すことで得られるのだ．したがって，DPS のもとで同一とされる理論をもつ S1 と S2 は，説明の成否についても一致することになる．

ここで求められる特徴は，同型の活性化空間をもつならば取り出せるようなものでなければならないので，各ニューロンの選好刺激などはその候補にならない．この条件に明らかに合致するものは，ある概念と他の概念との相対的類似関係である．こうした相対的類似関係は，プロトタイプベクトル間の相対的距離関係に写し取られている．そして，活性化空間が同型ならば，ベクトル間の相対的距離関係についても同一である．それゆえ，同型の活性化空間からは，同一の相対的類似関係を概念の特徴として取り出せる．そこで，我々がプロトタイプ活性化によって何かを説明するときにはこうした相対的類似関係に訴えているのだ，と論じることができるかもしれない．

しかし，このアイデアはうまくいかない．カブトガニのケースを再度考えよう．このとき，概念から取り出されるべき特徴「背泳ぎをする」は，“ カブトガニ ” の相対的類似関係「“ ホウネンエビ ” に似ている，“ マツモムシ ” に似ている，“ カブトエビ ” に似ていない，“ タガメ ” に似ていない，…」に置き換えられるようなものではない．なぜなら，このような相対的類似関係を理解しつつも，“ ホウネンエビ ” や “ マツモムシ ” は背泳ぎをし，“ カブトエビ ” や “ タガメ ” はそうではないことに気づかない人もいるかもしれないからだ．したがって，概念の相対的類似関係のみから説明的理解を得られるわけではない．

次に考えられるのは，活性化空間の次元のあり方とは無関係に空間上に描ける直線ないし部分空間に訴えるという戦略である．これは 2 次元の散布図に引かれる近似直線のようなものである．つまり，空間の内的構造に相対的に取り出せるものであり，かつ，空間上の各点の分布を分析する尺度を与えるかもしれないものだ（**図 5**）．

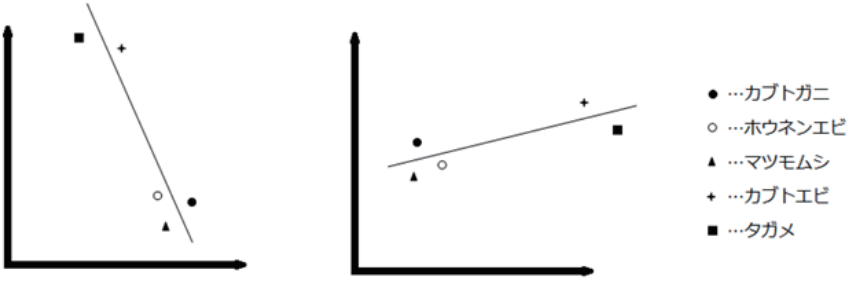


**図 5**　同型の空間上に引かれる直線

例えば，“ 顔 ” 空間の場合，活性化空間の次元がどんな特徴に対応していても，同型の活性化空間であるならば，“ 瞳孔と眉の距離 ” に対応する直線を空間上に引けるかもしれない．このとき，この直線上の端に位置するプロトタイプベクトルが表わす顔は，極端に瞳孔と眉が近い顔であるだろう．同じように，

S1 と S2 がもつ同型の“水生生物”空間上には“泳ぐときの体の向き”に対応する部分空間を描くことができ，かつ“カブトガニ”プロトタイプベクトルがこの部分空間上で一定の位置をとるのであれば，両者はこのベクトルから「背泳ぎをする」という特徴を取り出せることになるだろう[11].

こうした方策の問題は，部分空間に相当する尺度の解釈に多義性があることだ．部分空間が外界の何らかの特徴に対応しているとしても，認知主体によってその対応関係は様々に解釈されうる．例えば，泳ぐときの体勢と体の構造のあり方（重心の位置）には相関があるかもしれない．このとき，ある部分空間上に並ぶプロトタイプベクトルの順序は，“泳ぐときの体の向き”に応じたものとも，“重心の位置”に応じたものとも捉えられる．この解釈について S1 と S2 が異なっているのであれば，両者は説明の成否についても異なることになる[12].

このように，同型の活性化空間から取り出せる特徴に訴える応答には困難が伴う．DPS の枠内で説明の成否の差異を捉えることは難しいのだ．しかし，DPS にそのような分解能を求めることが間違いなのかもしれない．ある概念をもつことで説明が可能になることと，ある概念が特定の意味内容をもつことは別である．そして，前者は認識論によって，後者は意味論によって説明されるものである．それゆえ，概念の説明上の貢献を説明できないことは，概念の意味内容のあり方についての理論の瑕疵とはならない．よって，DPS が説明の成否の差異を捉えられないことは，概念の意味論としての DPS の問題ではないのだ．これは，3.1 で見たものと同じような，認識論と意味論の分離に訴える応答である．しかし，この応答は，SAS 的見方の基盤としての DPS の問題を解決するものではなく，その解決の放棄である．それゆえ，理論の SAS 的見方の意味論的基盤については依然として問題が残ることになる．

この応答に則りつつ SAS 的見方の問題を回避するためには，空間的同型性に加えて別の条件を理論の同一性・類似性の基準に含める必要がある．もちろん，このような追加条件は，説明の成否について異なる活性化空間を異なる理論とするようなものでなくてはならない．例えば，空間的同型性と，空間を構

成する次元が対応する特徴の一致という基準が考えられる．この基準からは S1 と S2 が異なる理論をもつことが帰結する．しかし，この場合，部分空間が対応する尺度の解釈について上で述べた問題と同じ問題が浮上する．ある次元 D は，“ 泳ぐときの体の向き ” に対応しているとも，“ 重心の位置 ” に対応しているとも解釈できる．この解釈の差異は説明の成否の差異につながることは上で見たとおりである．次元の多義性の問題を解決するためには，次元に対する精緻な同一性基準が求められることになるが，これでは状況が好転したようには思えない [13]．

あるいは，理論の同一性の基準に，他の活性化空間との関係性も含めるということも考えられる．つまり，その理論が知覚から運動に至るまでの一連の認知過程においてどのような役割を果たすかに目を向けるということである．例えば，S1 の場合，“ カブトガニ ” プロトタイプベクトルの活性化によって，別の活性化空間の “ 背泳ぎ ” プロトタイプベクトルも活性化することになるが，S2 の場合はそうではない，ということがあるかもしれない．このとき，S1 と S2 は説明の成否について異なる．そして，両者の “ 水生生物 ” 空間は他の活性化空間との関係性について異なっているので，両者は異なる “ 水生生物 ” 理論をもっているといえるのだ．こうした考えは先ほどの応答よりはうまくいきそうである．しかし，どのような関係性が理論の類似性・同一性にとって重要なのだろうか．例えば，“ 水生生物 ” 理論と “ 食べ物 ” 空間の関係性は，S1 と S2 で異なっている（S1 にとってカブトガニは食べ物だが S2 にとってはそうではない）かもしれない．あるいは，“ 水生生物 ” 理論と情動反応にかんする空間や行動生成に寄与する空間との関係性が異なっている（S1 はカブトガニを怖がり，水槽から逃げだそうとする）かもしれない．このとき，両者は異なる “ 水生生物 ” 理論をもっていることになるのか．異なっている，あるいは，同じだとするならば，その理由は何なのか．そして，異なっている場合には，（2.2 で見たような）類似性の度合の測定方法を考えることができるのか [14]．このような問題に解答を与えなければ，活性化空間同士の関係性は実質的な基準として役立たないだろう [15]．

以上のように，空間的同型性以外の条件によって理論の同一性の基準を考える方針にも困難がある．結局，説明の成否の差異に対応するように理論の差異化をするという課題の遂行は，DPS だけを下敷きにしていても，そして DPS の枠組を超えたところに訴えても難しいということだ．

## 4. 結論

以上で見たように，チャーチランドの認識論・意味論モデルに従えば，説明の成否について異なる個人同士でも，同じ理論を有することになってしまう．これがチャーチランドのモデルの大きな問題なのである．

もちろん，これはコネクショニズムに則って認識論を組み立てるというアプローチが不可能であるということを意味するわけではない．コネクショニズムの考えを維持しつつ，上の問題に対処する方針はいくつか考えられる．第 1 に，3.2 で検討したような応答によって問題を解決するのは困難ではあるが，不可能ではないかもしれない．もちろん，そのためにはこれらの応答をより精緻に検討する必要がある．

第 2 に，より大きな修正を許容するのであれば，我々の概念枠組みや理論は脳内の活性化空間ではなく，むしろ，脳外表象，特に言語によって表現されるものなのだという考えもあるだろう．こうした考えはベクテルやギャリによって提示されている (Bechtel, 1996; Giere, 2002)．これは外的表象の認知的重要性を認める考えであるが，ニューロンの活性パターンが脳内表象を担うという考えと両立可能である．以上のような方策が考えられるため，コネクショニズム的な認識論を構築するという試みにはまだ成功の見込みが残っている．

また，本稿で提示した問題によって，チャーチランドが提示するコネクショニズム的な認識論が全面的に問題のある理論である，ということが示されるわけではない．説明的理解というタイプの認知的達成のあり方をチャーチランドのモデルで上手く捉えられないというのが前節でみた問題なのであり，これは知覚や運動についても彼のモデルが不適格であることを示すようなものでは

ない．例えば，理論的説明とは異なり，ある対象から特定の特徴を取り出せなくても，知覚的判断における対象の同定は可能だといえるかもしれない．このような認知プロセスにかんしてはチャーチランドのモデルをうまく適用できるかもしれない．また，本稿では検討しなかったが，因果過程を表象するプロトタイプ軌道によって，ある種の因果的説明は捉えられるかもしれない．知覚や運動，因果的説明の一部にチャーチランドのモデルを適用できるのであれば，彼の理論の真の問題は．モデルの適切な射程を見誤り，あまりにも広範囲の事象を統一的に理解しようとしてしまったことにあるといえよう[16]．

## 註

1　この点にかんして，チャーチランドの議論は，（特に科学における）言語の重要性を軽視していると批判されている（Bechtel, 1996; 戸田山 , 1999）．ただし，チャーチランドも言語の認知的重要性を認識していないわけではない．彼は，科学の発展において言語が重要な貢献を果たしていることを認めている (Churchland, 2012, Ch. 5).

2　図 2 に表わされているのは 2 次元の平面であるが，実際の活性化空間は隠れ層ニューロンの個数と同じ次元数の高次元空間となる．

3　ただし，プロトタイプベクトルそのものが活性化する必要があるわけではなく，その周辺領域に含まれるベクトルが活性化すればよい．例えば，非典型的な事例に対応するベクトルは領域の周縁に存在する (Churchland, 2007, p. 144)．また，プロトタイプベクトルに対応するような典型例に一度も遭遇しない，ということもある (Churchland, 2007, p. 145)．説明的理解におけるはたらきのような概念的な作用を現実に担っているのは特定のプロトタイプベクトルというよりも，その周辺の領域（ないしはその領域に含まれるベクトル）である．しかし，本稿の議論においてはあまり重要でないため，以降では両者を厳格に区別することはしない．

4　チャーチランドは，DPS が捉えるのは概念の「狭い内容 (narrow content)」ないしはフレーゲ的な「意義 (sense)」であり，それ単体で概念の指示対象を定めるものでは

ないと考えている (Churchland, 2007, pp. 132–135)．概念や概念枠組みの指示にかんするチャーチランドによる議論については，本稿の議論にとって副次的であるため詳しく論じることはしない．

5　この考え方は「状態空間意味論 (State-Space Semantics)」とも呼ばれる (Churchland, 2012, p. 77)．

6　ベクトル間の距離関係からは，カテゴリーの階層構造（例えば，“ ○○家 ” というカテゴリーの内部にその家族の構成員のカテゴリーが含まれる）を読み取ることもできる (Churchland, 1998, pp. 14–18)．また，活性化空間上でプロトタイプベクトルが描く軌道は，物体の運動や衝突などの「プロトタイプ因果過程」(Churchland 2012, p. 149[ 強調原文 ]) を表象する．

7　こうした議論の背景には，フォーダーとルポアによる批判があると思われる．彼らは概念枠組みとしての活性化空間の同一性や類似性を測るためには，空間を構成する次元に対応する特徴を同定しなければならないと論じる (Fodor & Lepore, 1996a)．これに対して，チャーチランドは上で見たものと同様の議論によって応答している (Churchland, 1996a; 1996b)．フォーダーらとチャーチランドの間の論争については，Prinz (2006) に詳しい．

8　部分的な重なり合いは，次元数が異なる活性化空間同士についても測ることができる．それゆえ，個人の経験や知識の差異（付帯的情報の差異）によって次元のあり方が異なっても概念枠組みの類似性を捉えることができる (Churchland, 2012, p. 115n)．付帯的情報の問題はフォーダーとルポア (Foador & Lepore, 1996a; 1996b; 1999) やカルボ = ガルソン (Calvo Garzón, 2003) が提示した問題であり，この議論はそれに対する応答となっている．

9　ここでみたものが空間同士を比較する唯一の手法というわけではない．その他の手法については Churchland (1998)，Laakso & Cottrel (2000) を参照．

10　Prinz (2006) では具体的に述べられていないが，いわゆる「典型性効果」についての実験が想定されていると思われる．例えば，次のような実験がある (Rosch & Mervis, 1975; Prinz, 2002, p. 55)．2 つにわけたグループの一方には，あるカテゴリーに含まれるアイテムがもつ特徴をリストさせ，もう一方には，それらアイテムがその

カテゴリーに典型的なものかどうか判断させる．すると，あるアイテムの典型性判断の度合と，そのアイテムとカテゴリー内の他のアイテムとの特徴の重なり合いの度合は相関する．

11　何らかの特徴に対応する部分空間の存在は，概念枠組み獲得の後に（対象分類の便利な尺度として）学習されるものかもしれない (cf. Churchland, 2012, p. 88)．

12　より正確には，空間的構造のみでは尺度の解釈が定まらないので，個人の尺度解釈は，（それが定まるのであれば）空間のあり方とは別の基準によって定まることになる．そのような基準についての差異が，尺度解釈の差異，ひいては説明の成否の差異へとつながる．

13　脚注 7 で見たように，次元の同一性基準の必要性は，フォーダーとルポアが DPS に対して提示した問題である (Fodor & Lepore, 1996a)．

14　例えば，“ 食べ物 ” 空間との関係性の差異と，行動生成空間との関係性の差異では，理論の類似性に与える影響に差はあるのか，あるとしたらそれはなぜか，といった問題に答えなければならない．

15　実のところ，この段落で見たものに類似した見方をある時期のチャーチランドは取っている．例えば，Churchland (1996a; 1996b) では，活性化空間内の相対的位置の他に，概念が認知－運動ネットワークの中で果たす因果的－計算論的役割も概念の意味内容の類似性にかかわるものとされている．しかし，チャーチランドのその後の議論では相対的位置に重点が置かれるようになっている (e.g., Churchland, 1998; 2012)．Churchland (1996a; 1996b) の問題点についてはフォーダーらとプリンツが指摘している (Fodor & Lepore, 1996b; Prinz, 2006)．

16　本研究は JSPS 科研費 JP21J20803 の助成を受けたものである．

## 文献

Bechtel, W. (1996). What should a connectionist philosophy of science look like? In R. N. McCauley (Ed.), *The Churchlands and their critics* (pp. 121–143). Blackwell.

Calvo Garzón, F. (2003). Connectionist semantics and the collateral information chal-

lenge. *Mind and Language*, 18(1), 77–94.

Churchland, P. M. (1989). On the nature of explanation: A PDP approach. In P. M. Churchland (Ed.), *A neurocomputational perspective: The nature of mind and the structure of science* (pp.197–230). MIT Press.

Churchland, P. M. (1996a). Fodor and Lepore: State-space semantics and meaning holism. In R. N. McCauley (Ed.), *The Churchlands and their critics* (pp. 272–277). Blackwell.

Churchland, P. M. (1996b). Second reply to Fodor and Lepore. In R. N. McCauley (Ed.), *The Churchlands and their critics* (pp. 278–283). Blackwell.

Churchland, P. M. (1998). Conceptual similarity across sensory and neural diversity: The Fodor/Lepore challenge answered. *The Journal of Philosophy*, 95(1), 5–32.

Churchland, P. M. (2007). Neurosemantics: On the mapping of minds and the portrayal of worlds. In P. M. Churchland (Ed.), *Neurophilosophy at work* (pp. 126–160). Cambridge University Press.

Churchland, P. M. (2012). *Plato's camera: How the physical brain captures a landscape of abstract universals*. MIT Press.

Fodor, J., & Lepore, E. (1996a). Paul Churchland and state space semantics. In R. N. McCauley (Ed.), *The Churchlands and their critics* (pp. 145–159). Blackwell.

Fodor, J., & Lepore, E. (1996b). Reply to Churchland. In R. N. McCauley (Ed.), *The Churchlands and their critics* (pp. 159–162). Blackwell.

Fodor, J., & Lepore, E. (1999). All at sea in semantic space: Churchland on meaning similarity. *The Journal of Philosophy*, 96(8), 381–403.

Giere, R. N. (2002). Scientific cognition as distributed cognition. In P. Carruthers, S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 285–299). Cambridge University Press.

Laakso, A., & Cottrell, G. (2000). Content and cluster analysis: Assessing representational similarity in neural systems. *Philosophical Psychology*, 13(1), 47–76.

Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. MIT Press.

Prinz, J. (2006). Empiricism and state space semantics. In B. L. Keeley (Ed.), *Paul Church-*

*land: Contemporary philosophy in focus* (pp. 88–112). Cambridge University Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605.

Winther, R. G. (2021). The structure of scientific theories. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 Edition). https://plato.stanford.edu/archives/spr2021/entries/structure-scientific-theories/

Woodward, J., & Ross, L. (2021). Scientific explanation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 Edition). https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/

戸田山和久（1999）.「科学哲学のラディカルな自然化」『科学哲学』32 巻 1 号 29–43 頁.