

Title	ロボットがフレーム問題に悩まなくなる日
Author(s)	柴田, 正良
Citation	: 119-174
Issue Date	2004-07
Type	Book
Text version	publisher
URL	http://hdl.handle.net/2297/36191
Right	

*KURAに登録されているコンテンツの著作権は、執筆者、出版社（学協会）などが有します。

*KURAに登録されているコンテンツの利用については、著作権法に規定されている私的使用や引用などの範囲内で行ってください。

*著作権法に規定されている私的使用や引用などの範囲を超える利用を行う場合には、著作権者の許諾を得てください。ただし、著作権者から著作権等管理事業者（学術著作権協会、日本著作出版権管理システムなど）に権利委託されているコンテンツの利用手続については、各著作権等管理事業者に確認してください。

第三章 ロボットがフレイム問題に悩まなくなる日

柴田正良

1 ロボットは苦悩する（あるいはフレイム問題）

フレイム問題の由来は、マッカーシーとヘイズが書いた一九六九年の論文「人工知能の観点から見た哲学的諸問題」だとされている。そもそも「由来」を語ることができるということは、この問題がその発端から「成長」し、何かしら初めとは別の点で有名なもの（？）になってしまったということだ。もちろん、まったく別物になったわけではないが、ここで扱うその問題は、当初の人工知能（AI）の技術的な狭い問題設定から進化した（哲学者のフレイム問題）と呼ばれるものである（Sousa, 1987: 192）。まずは、そのフレイム問題の何たるかを確認することから始めよう。

ことの始まりは、変化する世界を人工知能が扱おうとしたことである。現実の世界では当たり前のことだが、ある出来事が生じれば、それに応じて同時にたくさんの変化が生じる。しかもその変化は、

われわれの想像もしなかった事柄にまで波及する。とはいえ、これまた当然のことながら、何か起きて、世界の大部分の事柄はそれによって何の変化もこうむらないままである。例えば、あなたが裏庭のノラ猫を箒で追い払ったなら、箒は物置から出されていたはずだし、猫は裏通りに逃げだしたはずである。そして、飼い主が四〇万円という大金をかけてその猫を探していたのだとすれば、同時にあなたは、その懸賞金を得るチャンスを逃したのだ。しかし、猫を箒で追い払ったとき、もちろんそれによってあなたの銀行口座に何か変化が生じたわけではない。ここまでは何の問題もない。

問題は、かなり窮屈な人工知能が融通のきかない（つまり厳格な）演繹システムによって、こうした変化をとらえようとする場合だ。このようなシステムは、公理からの論理的証明によってすべての知識を獲得しなければならぬから、ある出来事の変化によって生じている世界の変化についても、例えば、箒の場所や裏通りの状態などをそれぞれの公理から推論しなければならぬ。いったい、どれくらいの数の公理が必要になるだろうか。例えば、

公理 1 $\langle x \text{ を } y \text{ で 追 い 払 う } \downarrow (y \text{ が } w \text{ の 中 に あ っ た なら ば, } y \text{ は } w \text{ から 出 さ れ る}) \rangle$

公理 2 $\langle x \text{ を } y \text{ で 追 い 払 う } \downarrow (x \text{ が } z \text{ を 安 全 な 場 所 だ と 思 う なら, } x \text{ は } z \text{ へ と 逃 げ る}) \rangle$

……

しかし、こうした公理の数の多さに絶望的になるのはまだ早い。世界には、その出来事が生じてもしも〈変化しない〉事柄もたくさんあったからだ。問題の演繹システムは、〈変化しない〉事実に関

しても、やはり公理からの演繹的推論によってその知識に到達しなければならぬ。そのような公理がなければ、システムは、どのような状態が以前と同じように存続しているのかを確定することができないのだ。いわゆる「フレーム公理」とは、このように不変化の事実をシステムが演繹するためにおかれた「不変化公理」にはかならない。すると、容易に想像がつくように、必要とされる公理の数は悲劇的なまでにくれあがることになるだろう。というのも、システムの扱う世界が〈世界〉という名に恥じないほどに多くの事実から成っているなら、フレーム公理の数は、システムが表現できる可能な事実の数とともに爆発的に増えていくからだ。この問題の不気味さは、次のようなフレーム公理の一見した無意味さと無関係ではない。猫をただ箒で追い払ったからといって、その猫が猫でなくなるわけでもないことを、また箒の重さが変わるわけでもないことを、なぜことさら改めて知らねばならないのか?……

公理 3 $\langle x$ を y で追い払う \downarrow (x が以前に猫ならば、 x は以後も猫である) \rangle

公理 4 $\langle x$ を y で追い払う \downarrow (y が以前に軽いならば、 y は以後も軽い) \rangle

……

しかし、これでもまだフレーム問題の本当の困難さは実感できないかもしれない。というのも、いくら公理の数が増えようと、またいかに公理の内容が取るに足りないバカげたものだろうと、いやそれだからこそ(?)、所詮は現実問題とは直接関係のない、単なるプログラム上の技術的問題だと思

われるかもしれないからである。しかしながら、そうではない。この問題の一見した空虚さは、人工知能がやがて身体をまとい、ロボットとなって現実世界の中で行為しようとするやいなや、とたんに切実なものとして重くのかかってくる。そこで、そのありさまを、いまや有名となったデネットのストーリーから見えてみることにしよう。このストーリーは「フレーム問題」という名を少しでも知っている読者ならばきっとすでに知っているに違いないが、これに代わるほど鮮やかな例解をいまだに私は知らないので、ここで改めて簡単に紹介するのをせひとも許して頂きたい。⁽¹⁾

むかしむかし(?)、R1というロボットが研究室で作られた。そのロボットは何の仕事もなく気楽な生活を送っていたが、ある日、試練が訪れた。彼の動力源となる予備バッテリーをしまつてある部屋に時限爆弾が仕掛けられたのだ。爆発まであまり時間がない。バッテリーは一台のワゴンの上にあった。そこでR1は、直ちにバッテリー救出作戦〈PULL OUT (WAGON, ROOM)〉を立て、爆発が起きる前にワゴンを部屋から引きずり出した。作戦は成功。しかし、不幸なことに、爆弾はバッテリーとともにワゴンの上にあったのだ(Bang)。R1は爆弾がワゴンの上にあることを知っていた。ただ、ワゴンを引き出すことが爆弾も一緒に引き出すことになるのは、思いもよらなかったのである。改良点は明らかだった。ロボットは自分の行動の意図した結果だけでなく、意図せざる副次的結果も知らなければならぬ。それを知るためには、周囲の状況と行動についての記述からその副次的結果を演繹させねばいだろう。そこで次に作られたロボットは、「演繹ロボット」R1D1と名づけられた。R1と同じ状況に立たされたとき、R1D1はやはり同じ作戦を立て、設計図通りに、その

行動の副次的結果を次々に考え始めた。やがて、ワゴンを部屋から引き出しても部屋の壁の色は変わらないということを演繹し、次に、ワゴンを引けば車輪が回転するだろうという帰結の証明に取りかかったときだった、爆弾が破裂したのは。

今度の問題は、目標に対する帰結の関連性にかんしてロボットがまったく無頓着なことだった。とすれば次の改良型ロボットに必要なものは、演繹した帰結を目下の目標にとって関連のあるものとなしものに分類し、関連のないものは無視するという能力である。新たな改良モデルは、「分別ある演繹ロボット」R2D1と呼ばれた。さて、設計者たちの期待を一身に背負ったR2D1はこの苦境にどう立ち向かったのか。デネットの言葉をそのまま引用しよう。

すると、驚いたことに、かれは部屋のなかに入ろうともせず、まるでハムレットのようにじっとうずくまったままであった。部屋のなかでは時限爆弾がチクチク音を立てていた。かれは、シェイクスピアが（そして最近ではフォードが）見事に表現したように、赤く燃え上がる、もって生まれだ決断力を憂鬱な心の壁土で塗りたくられてしまったのだ。「なにかしろ！」と設計者たちは叫んだ。かれは「してるよ」と答えた。「ぼくは、無関係な帰結を探し出してそれを無視するのに忙しいんだ。そんな帰結が何千とあるんだ。ぼくは、関係のない帰結をみつけると、すぐそれを無視しなければならぬものリストにのせて、……」。かくてまたもや爆発。(Dennett, 1987: 42)

R2D1というデネットのネーミングは、映画『スターウォーズ』のずんぐり型ロボットR2D2

ではすでにこのフレーム問題は解決されているのだ、という含みだろうか。それはともかく、フレーム問題が、人工知能、それもとくに現実世界の中で行為するロボットにとって致命的な問題であることが、これで少しは明らかになったと思う。では、はっきり言って、それはどういう問題なのか。三つの側面からそれを特徴づけてみよう。

(一) 端的な無視 (あるいは「ちょうどびったりと知ること」)

第一に、デネットのストーリーから明らかのように、それはたんに知識が不足しているという問題でもなければ、適切な推論ができないという問題でもない。R2D1は、「爆弾がワゴンの上にあるということ」を知っているし、「ワゴンを引けば爆弾も一緒に部屋から出すことになる」ということを推論によって知ることもできる。では問題は、知りすぎていることだと言っていいだろうか。確かにそれも関係しているが、しかしそれなら、すべてを知っている神はフレーム問題を逃れるすべもない、という不都合なことになるだろう。

問題は、すでに獲得している多くの知識の中から必要なものを必要なときに必要だけ取り出し、それ以外のものは何もさわらずにそっとしておく、ということである。この「何もさわらずにそっとしておく」という点が重要なことは、R2D1の悲劇が、関連しない知識を次々に引き出して、それに「関連せず」という烙印を押さねばならなかった点にある、ということからも分かるだろう。したがって、関連のない知識項目を無視すればいいのだが、関連のなさを一々の項目に関して確認するような仕方での「無視」を行ってはならない。なぜか。それは、一般に知識は行為のために役立てられ、行為は、限りある情報の効率的利用と有効時間内での実行を必須の条件とするからである。

したがって、理想的な有限の行為者というものが存在するならば、彼／彼女は、そのつどの行為に
 関して、知らなければならぬものすべてを、しかもそれだけを、多すぎも少なすぎもせずによ
 うどびったりと、自分に見合った時間と労力の範囲内で、そのつど知ることになるメカニズムを備え
 た認識者だということになるだろう。つまり、その時に知るべきことは知るが、知らなくてもいいこ
 とはすべて、そのつど知るための努力の対象にはならないようなメカニズムが問題なのである。した
 がって、こう言ってもいいだろう。フレーム問題とは、「何を考えなくてもいいか」ということを考
 えずに、考えなくてもいいことをいかに考えずにすませるか、という問題であると。フレーム問題を
 特徴づける第一の点、端的に無視する、つまり余計なことを一切考えずに考慮する必要のないことを
 考慮しない、とはそういうことだ。

(ii) 関連性(あるいは「anything to anything」)

行為を開始する前の知識状態を考えてみよう。自然に浮かんでくるイメージにしたがえば、それは
 図書館の本棚のように整然としてか、あるいはあなたの部屋(?)の中のように雑然としてか、いず
 れにしても、いかにその知識が膨大であろうと、全体が渾然一体となって溶け合っているのではなく、
 それぞれの知識が互いに区別されてそれぞれの場所に収められている、と考えられるだろう。コンピ
 ュータのメモリに割り振られる「番地」は、そのようなイメージのある種の(古典的計算主義的な)
 実現だ。何かを思いだすとき、あるいは何かを推論するとき、必要な知識がその「番地」をもとに探
 し出され、それが認知過程に呼び出される。

さてそこで、さまざまな知識をもった行為者が世界の変化に直面し、新たな情報を獲得したとき、

彼のもっている知識は適切に更新されなければならない。しかしどの知識を更新させるのか？ 先ほどのR2D1が身をもって示してくれたように、もちろんそれは、現在の目的に関連のある知識以外のものではない。しかし、どれが現在の目的に関連のある知識であるかを、何が教えてくれるのか？ デネットのストーリーでは、設計者たちが〈関連／非関連〉の分類の方法を開発し、ロボットに授けたことになっていた。確かにあのストーリーのポイントは、そんな分類法があっても愚直なロボットは恐ろしいほど非効率なやり方でしか計算できない、ということであった。しかし、そもそも設計者たちは、どんな状況にも通用するそんな分類法をどうやって見出したのか。かりにもしそうした分類法があったなら、あの状況であらかじめ、ワゴンを引き出すことの結果「更新すべきことになりそう」な一群の信念「知識」をフレーム（枠組み）で「囲い込み」(Edoer, 1983: 112-3)、それだけをチェックする、ということができただろう（哲学者のこの「フレーム」が、「フレーム公理」の「フレーム」から少し意味をそらせ始めていることに注意）。ちょうど、新しい情報に対して関連のある知識は、図書館のこの区画の五つの本棚、あるいはあなたの部屋の偶数月発行の雑誌に収められた知識だけだ、と事前に決定しておくように。

しかし、知識と知識、情報と情報、信念と信念の関連性は、それらが文脈と無関係にもつローカルな性質ではなく、文脈ごとに持ったり持たなかったりする全体論的な性質である。したがって、文脈によっては、いかなる情報 (anything) も他のいかなる情報とも (to anything) 関連性をもちうるのだ。火星に無人探査機が降り立ったことは、ふつう、あなたが明日朝九時に出社するかどうかに何の関連性ももたないだろう。しかし、あなたが火星の土地の売買という超いかがわしいセカンドビジ

ネスに大金をはたいていたら、あなたの出社時刻に大きな関連性をもつかもしれない。フレーム問題の第二の特徴づけを言いかえれば、あらゆる情報が他のあらゆる情報に関連性をもちうるような情報の組織化を、無限数のフレームの装備などという実行不可能な手段によらずに、いかに信念（知識）システムにおいて実現しうるかという問題である。

(iii) 規則と例外（あるいは「特別な事情がない限り」）

そこで、たとえそのようなフレームをロボットがもったとしてもいかなる問題が発生するか、というのを別の角度から述べてみよう。これが、フレーム問題の第三の特徴づけである。たとえば、あなたがレストランに行く。そこで起きる出来事はだいたいが予測できることだろう。ウェイターが来て、あなたを席に案内し、メニューをもってきて、注文を聞く。あなたは好みの料理を彼に伝え、しばらく所在なげに（？）待ってから、運ばれた料理を平らげ、席を立ち、支払いをすませて、店を出る。これらのストーリーのあらましをまとめたものがミンスキーの言うフレームであり、それは記憶の中から呼び出される知識の基本的まとまりのようなものだ。人は「その細部を必要に応じて変更することで現実の世界に適合するようになる。……フレームとは、居間にいるとか、子供の誕生パーティに行く……といった、ステレオタイプの状況を表現するためのデータ構造である」(Minsky, 1975: 212)。この点では、シャンクの「スクリプト」やポプロウの「スキーマ」も基本的に同じものだと考えてよい（第3節ではスキーマを取り上げる）。

このようなフレームの本体は、突きつめて言えば、「レストラン↓ウェイターが来る」、「ウェイターが来る↓（彼は注文を聞くor彼は料理をもってくるor……）」といった条件文の形をした思考規

則／行為規則の集合にほかならない。すると、現実の世界ではよくあるように、どんな典型的な状況でもとんでもない例外が生じてくるから、これらの規則も、致命的な例外には対処できるようにないなければならないだろう。というのも、「目的物を部屋から出すには、それがワゴンの上にあるならワゴンごと部屋から引き出せ」という規則の大事な例外、「爆弾もワゴンの上にある場合は、その限りにあらず」をあらかじめ知らなかったがために、最初のロボットR1は爆弾の餌食になってしまったからだ。

しかし、このような〈重要な〉例外は、どのようなものが、どれくらい存在するのだろうか。答は、〈分からない〉である。分かりやうがないのだ。それで、人間の場合はどうなっているかというところ、どんな規則にも実は一つの限定、「特別の事情がない限り（他の事情がすべて等しいなら）*ceteris paribus*」という例外対処条項が暗黙のうちに付けられている。つまり、「レストランに着いたら、特別な事情がない限り、料理の注文をせよ」というような具合だ。しかし、問題は、人間の場合には「特別な事情って具体的には何だ？」といったうるさい質問には答えなくてもすむのに（「そんな常識やんか」、人工知能の場合にはそうはいかないということである。少なくとも、知識の更新や行為への推論を、規則にしたがった記号操作によって実現するタイプの知識システム、つまりプログラムによって計算を行うふつうの人工知能にあつては、その曖昧な「特別な事情」の中身を明示的に列挙しなければ、その規則を使うことがそもそもできなくなってしまう。というのも、そのシステムは、〈今の今〉が特別な事情のないふつうの場合であるかどうかすらも、分からなくなってしまうからだ。しかし、特別な事情をすべてあらかじめ枚挙することは、文脈によってどの条件もその特別な事情に

なりうるのだから、土台無理な話である。したがって、フレーム問題をこの側面から見れば、あらゆる規則の適用に際して、「特別な事情がない限り」という条項に匹敵するような、例外に対する何か柔軟な処理をいかに編み出すことができるか、という問題である。もし例外の処理に手を焼くようなら、いっそ明示的な規則など一切もたずに、それでも、そのつどそのつどの状況で適切な規則にしたがっているかのように振る舞うことのできるシステムがあったならば、その方がずっといいだろう。

さてしかし、残念ながら私には、この厄介な問題の有望な解決案を提示することはいまのところできない。ここで私がなしうることは、せいぜい、認知に関するこの問題の哲学的な含みを検討し、この問題の解決のための条件をできるだけ明確にすることである。その結果は、しかし、単なるフレーム問題という枠を越えた知性一般に関する論点につながるだろう。そこでまず、フレーム問題は古典的計算主義の枠組みでは解決不可能だとする論点の整理を、もっぱらホーガンとティーンソンの議論に依拠して行うことにする(第2節)。ついで、古典的計算主義の代替案として登場してきたコネクショニズムの下では、フレーム問題がいかなる意味で解決可能だとされているのかを確認する(第3節)。しかし、コネクショニズムの「成功」はいまだ部分的な成功にとどまっており、決定的なものではない。ニューラル・ネットワークが本当にワールドな現実の状況に立たされたときにフレーム問題を克服できるかどうかは、知性の全体的な設計のなかで問われなければ意味がない。というわけで、コネクショニズムが与えてくれる解決の予兆にはまだ何が欠け、何が必要なのかを含めて、最後に知性一般に関していささかの思弁を試みてみることにしよう(第4節)。

2 精確な規則にしたがった記号操作（あるいは古典的計算主義）

古典的計算主義とは要するに、われわれの周囲にころがっている（？）ふつうの多くのパソコンの作動原理となっているものだ。しかし、ホーガンとティーンソンによれば、フレーム問題の解決は古典的計算主義によってはまったく見込みがない。それは、古典的計算主義が認知の理論として決定的に誤っていることの証拠である。そこで彼らは、古典的計算主義に代わる新たな認知理論のパラダイムとしてコネクションリズムを提唱することになるのだが、それが本当にどれほど革新的であるかはともかく、彼らの見るところ、なぜ古典的計算主義ではフレーム問題を解決することが不可能なのだろうか。

ともかくまず、古典的計算主義に対するホーガンとティーンソンの診断を確認しておこう。古典的計算主義によれば、認知過程は、記号表象を扱うアルゴリズムによって、つまり規則に支配された記号操作によって実現される。その主張の根本的前提は、彼らによれば次の五つである (Horgan & Tinson, 1996: 24-6)。

- (1) 知的な認知は、構造的に複雑な心的表象を用いる。
- (2) 認知過程は、これらの表象の構造に左右される（それゆえ、表象の内容に左右される）。
- (3) 認知過程は例外のない精確な規則に合致し、それらの規則は、表象それ自身を扱う規則とし

て表すことができるし、また、コンピュータ・プログラムの形で書くこともできる。

(4) 多くの表象は統語論的構造をもつ。

(5) 人間の認知状態の推移は、何らかの物理的手段で実際に計算可能 (tractably computable) な認知推移関数に合致する。

認知過程全体の古典主義的な見通しに関して、ここで少し説明しておいた方がいいかもしれない。認知過程は、典型的にはわれわれが志向的状态としてとらえているようなある認知状態 A (例えば、ボールが飛んできたという知覚) から別の認知状態 B (それを捕ろうという意図) への推移の連続である。しかしもちろん、現実的な認知システムというものを考えれば、ある時点での認知システム全体の状態が、ただボールの知覚だけから成り立っているはずはなく、同時並行的にはほかの多くの認知状態 (他の知覚や信念や欲求) もそこで実現されているはずである。したがって、われわれを一つの複雑な認知システムとしてみれば、われわれの生活 (?) とは、そのような認知状態相互間の絶え間ない推移のことにほかならない。古典的計算主義は、この認知状態間の推移を、状態から状態へのマッピングという意味で数学的関数だととらえる。つまり、認知システムとしてのわれわれは、この認知推移関数 (cognitive-transition function) を計算するデバイス (装置) だということになる。

では、この関数は実際にはどのような仕方で計算されるのか。システムの状態が認知状態であるからには、システムの内部状態を構成するが、その不変的構造には属さず、むしろ環境の変化と連動して変化し、その変化が内部状態の変化であるようなものが必要である。それが「表象」だ。表象は、

外部の情報重担う（意味論的内容をもつ）という点では〈記号〉であり、その適切な集まりが認知を構成する（心的状態となる）という点ではシステムの〈認知状態〉である。そしてここが古典的計算主義の決定的特徴なのだが、認知状態相互間の推移関数は、この記号としての表象に対する操作として計算されるのである。記号の操作は、記号の形式的特徴だけに反応する形式的規則によって果たされる。いいかえれば、表象を扱う形式的規則が、認知推移関数を計算するためのアルゴリズム（計算方法）となるのである。さらに、このような記号としての表象の形式的特徴、つまり統語論的特徴とそれを扱う形式的規則が、実際にどのような物理的装置として実現されるのか、という問いに対する答として古典的計算主義が提供するものが（チューリング・マシンとしての）いわゆるコンピュータであることはもはや言うまでもなからう。そこでは、表象はまさに物理的に実現された記号列「 $\sigma \downarrow p$ 」であり、その形式的特徴だけに反応してそれを操作する規則とは、コンピュータ・プログラムにはほかならない。

ホーガンとティーンソンによる五つの前提を順にもう少し説明しておこう。人間の認知過程が環境世界のきわめて複雑な情報を処理し、高度に組織化された理論的認識にまで至るということを考えるなら、それは（1）で言うような複雑な構造の心的表象を用いる以外は不可能だ、というのが古典的計算主義の基本前提である。人間の認知には実は表象はまったく用いられていないという過激な立場もあるが、少なくとも、認知科学の本体と目されてきたのは、上部レベル（知識や意味論）や下部レベル（神経科学や物理学）の領域から相対的に独立した、中間レベルの対象としての表象についての理論（アルゴリズムや統語論）であったし、今もそうである（Cf. Pylyshyn, 1989: 7-10）。確かに表象

は、日本語や英語のように外部に現れる自然言語のものではないし、解剖された脳のニューロンの中に観察されるわけでもない。したがって、その実在性は、ロックやヒュームが語る〈観念〉や〈印象〉のように外界からの感覚刺激の実在性から漠然と類推されるわけにはいかず、ひとえに、表象に言及する理論がいかにわれわれの認知過程をうまく説明できるか、ということにかかっている。この点で、古典的計算主義はコンピュータモデルによって、表象の理論的実在性をまさに〈目に見える〉形で人々に説得することができたのである。もしわれわれの脳がコンピュータであり、心がプログラムなら、心的表象が存在しないということは、プログラムを書くための言語記号が存在しないということに等しいだろう。

しかし、表象が認知過程の説明に要請されるのは、人間の場合、信じたり欲求したりしている内容と整合的な仕方での信念や意図が形成される、ということが謎となってしまうためでもある。「コーヒーを入れよう」という意図の神経的な実現がどうあろうと、そう意図した〈結果〉自分がネコに餌をやっていた、というようなことはわれわれのような認知システムでそう起こってほしくない事態だし、実際ごく希にしか起こらないだろう（それに近いことがよく起こるようでは、いよいよヤバイけど……）。この内容の整合性、つまり意味論的な調和は、(2) 認知過程が心的表象の内容にたがって生ずるということの意味するが、それは結局、認知過程が表象の構造にしたがって生ずるということだ。というのも古典的計算主義によれば、今のところわれわれは、意味論的内容が因果的な効力を発揮する（しかるべき意図や行為を生じさせる）仕掛けとしては、コンピュータ言語と似たような仕方、意味論的内容が統語論的構造に依存するあり方しか知らないからである。

そこで、もし古典的計算主義の言うように認知過程が正確な規則によって生じているのだとしたら、(3)その規則とは表象を扱うための規則にほかならず、規則であるからには、それをコンピュータ・プログラムの形で書くことができると思えるのは自然であろう。というのも、プログラムとはまさしく、ある条件の集まりが成立したときに次に何をするかという機械的(自動的)な手続きを定めたフロー・チャートの集合にほかならないからである。認知過程はへたんに規則に従っているかのように記述される(ちょうど、規則ではなく法則に従う惑星の動きのように)、というのではない。まさに、規則に書き込まれたステップの一つ一つを実際にシステムが辿るという意味で、(規則に従うという)メカニズムが認知過程を動かしているのだ。その規則が例外のないものでなければならぬはずだ、というホーガンとティンソンの主張は、すぐ後で問題にすることにしよう。

心的表象は、(4)古典的計算主義では典型的に、統語論的構造もったものだと考えられている。というよりも、認知過程が計算(つまり規則に従った記号操作)であり、なおかつ意味論的な整合性を持ち、なおかつ物理的なメカニズムによって実現される、というアイデアを可能にしたのは、フォードの説く「思考の言語 Language of Thought」という仮説にほかならない。要するに、現代的な意味での表象 (representation) というのは、外界の事物の似姿 (イメージ) でもなければ、外界の事物によって刻印される感覚印象でもない。それは、外界の事物を表わす(志向性をもつ)心的記号であるが、なによりも自然言語のような統語論的構造を持ち、統語論的規則に従って合成され、意味論的に解釈可能であるような抽象的存在者である(物理的には、たぶん脳において神経的に実現される)。これが、古典的計算主義の理解する表象であり、たとえこのような(文的)表象とは異なる

〈像的〉もしくは〈図的〉表象の存在を古典的計算主義が認めたとしても、認知過程を担う本体としての表象は、あくまで統語論的な構造をもつ表象なのである。

認知過程がプログラム化可能な規則に従って順次生ずるなら、(5) トータルな認知状態の推移は、いかに複雑なものであろうと実際に計算可能な関数として理解することができるように思われるだろう。しかも正確に言えば、その計算可能性は、純粹に数学的な計算可能性ではなく、何らかの物理的な手段による現実の計算可能性である。というのも古典的計算主義によれば、われわれのような認知システム、つまり物理的に実現可能な認知システムは、実際にそのような推移関数を計算する計算機にはかならないからである。もちろん、この計算は、認知状態の推移がたんにプログラム化可能な規則に合致する〈入力―出力〉の関数をもつということではなく、プログラムに指定されたステップをシステムが次々に内部状態として実現しながら入力状態から出力状態に推移する、という意味で理解されねばならない。⁽²⁾

ホーガンとティーンソンは、以上五つの古典的計算主義の前提のうち、(3)と(5)を拒絶する。つまり彼らによれば、われわれの認知は、表象自身を扱うプログラム化可能な規則に従って進行するわけでもないし、その推移関数は、いかなる物理的な手段によっても計算不可能である。しかし、彼らの描く認知のメカニズムは、先の前提の(1)、(2)、(4)を保持する。つまり彼らの主張は、統語論的構造をもつ表象としての「思考の言語」仮説を保持しながら、表象の操作としてはプログラムではなくニューラル・ネットワークにおける並列分散処理(全体論的重ね合わせ)を唱え、それを数

学的にはダイナミカル・システム（物理的手段によっては状態の推移が計算不可能な一種の複雑系）として理解するような、少し異端的な（？）コネクショニズムなのだ。⁽³⁾ 彼らの主張が成功するかどうかは、ここでの問題ではない。われわれが彼らの議論から引き出すべきは、フレーム問題はなにゆえに古典的計算主義の枠組みでは解決不可能なのか、というその理由である。

まず第一に、プログラムもしくは規則というものがわれわれの認知活動を捉えるのに適切な概念ではない、と彼らは考える。その議論を少し敷衍しておこう。古典的計算主義の下での規則は、前提の(3)で触れておいたように、例外のない規則であることを要求される (Deductive)。なぜかというに、ここではまず、この規則はある問題領域におけるわれわれの（データを除く）認知能力のすべてを構成する（説明する）はずのものだ、と主張されている。したがって、この規則の適用に関してその領域における認知能力を別個に必要としたならば、その規則は問題の認知能力のすべてではない、ということにならざるをえないだろう。ところが例外がほかならぬ例外として認識されるということは、その規則の適用は（その場合において）差し控えられるべきだ、というその領域における（問題の規則とは）別の認知能力が発動されたということだ。したがって、問題の規則は、例外をもつがゆえにこの領域の認知能力のすべてではない。しかしそれでも、例外の認識が神秘的な直観などではなく、やはり何らかの規則の適用だと言いたいなら、必要とされるのは、その認識を〈例外〉としてではなく〈通例〉として含むようなさらに上位の規則であり、その上位規則が今度はその領域の認知能力のすべてだと主張されることになるだろう。しかしそれなら、同じ理屈によって、その上位規則は〈例外〉を含むことはできないはずである。しかも、この〈できない〉は概念的な制約を表している。

例えば、規則「鳥は飛ぶ（と考えよ）」がペンギンを例外とする場合、このままでは、この規則は〈鳥の飛行〉という領域に関する認知能力のすべてではない。では、規則「鳥は飛ぶ」をペンギンの場合には差し控えるべきだ、という判断は一体どこからやってくるのか。今われわれがなそうとしているのは、認知能力のすべてを霊的な力などではなく何らかのメカニズムによるものとして説明しようとするのであり、古典的計算主義とは、そのメカニズムを心的表象に対する規則（プログラム）の適用と捉えるプロジェクトである。したがってこのプロジェクトが成功するためには、〈鳥の飛行〉という領域に関する認知能力は、ペンギンに関する「例外条項」どころか、すべての「例外条項」をも規則の一部として含んだ〈例外なき規則〉によって余すところなく説明されねばならないのである（……そんなことができるだろうか？）。

ここで、規則とプログラムに対するホーガンとティーンソンの不満の本音（？）を、確認しておこう。それは、実はフレイム問題にとつて、一見したよりもはるかに本質的な点で重要かもしれない。

人間の活動が規則によって決定されているという考えは、われわれの行いたい方がいいのことはふつう問題解決活動だとは見なされない、ということに思い至るならそれほど説得力のあるものではない。交差点に近づいているときにいつブレーキを踏むか、を正確に決定しているような規則があるだろう。……あるいは、仲間とのおしゃべりや仕事の最中にいつコーヒーマップを手にするか、を正確に決定しているような規則があるだろうか。ブレーキをいつ踏むか、仕事にいつ戻るかを正確に決定しているような、内的な認知規則があるのだろうか？ (Ibid.: 31-3)

彼らは、この点での古典的計算主義の難点を、非常に多くの認知的作業（典型的には言語使用）においては表象の産出が際限ないこと（open-endedness）、そして、そもそも認知的作業の種類に際限がないことの二点から必然化すると見ている。確かに、規則が扱うべき表象が際限なく、規則を必要とする作業の種類も際限がないのであれば、規則を際限なく産出するメカニズムだけがそれに対処できるだろうが、そのメカニズムを捉えたと称する規則がこれまでに提案されたことはなかった。というのも、そういう規則が存在するとすれば、それはまさしく〈万能型知性 universal intelligence〉そのものであったはずだろうが、規則とはそもそも領域限定的なものでなければ不可能であるように思われるからである。したがって、古典的計算主義の提案する規則がチェスゲーム・プログラム（例えば一九九七年の「ディープ・ブルー」）のように〈一能型知性 particular intelligence〉であったのは、たんなる技術的な問題ではないだろう。つまり、万能型知性である〈規則〉とは、人生の規則（？）、しかも例外のない人生の規則といった途方もなく意味不明なものにならざるをえないように思われるのだ。これは、第4節で見られるように、たんなる語呂合わせの冗談ではない。それは、フレーム問題が（知性の、ではなく）知性という部分的な領域を越えた、生存可能な認知システム全体に対して初めて発生し、そこにおいて解決が図られるべきことを暗示しているのではないか。いいかえれば、フレーム問題の突き刺さる地点とは、知的システムだけを扱う部分的な理論ではなく、狭義の認知システムを組み込んだトータルな生存システムを扱う理論の次元元ではないだろうか。だが、あまり先走る前に、古典的計算主義の絶望の姿をもう一度、イヤというほど目に焼き付けておこう。

その絶望とは、すでにお察しのように、一般的な形での関連性 (relevancy) の問題である。ホーガンとティーンソンが指摘する古典的計算主義のその第二の難点を、彼らと同じくらい雄弁にフォーダーに語らせよう。一般に、フォーダーは古典的計算主義の守護神のように思われているが、それは主に入力系の認知モジュールに関してであって、こと情報処理の中枢部分に関してはまったく異なる。フォーダーの立場は、知覚や言語理解や相貌認知などのさまざまな入力系に関しては、モジュール性が成り立ち、したがって心の計算理論がたぶん妥当するが、周辺からの入力情報を同時にまとめて処理する中央システムはモジュール性をもち、心の計算理論も妥当しないというものである⁴⁾。このことと関連性の問題がそれこそどう関連するかというと、心の計算理論の妥当性はある領域内部での情報処理がローカルであることに依存するのに対し、関連性とはまさしくローカルではなく、グローバルもしくは全体論的な情報処理を要求するということだ。したがって、古典的計算主義が心の計算理論に依拠するかぎり、関連性を適切に捉える認知モデルを提出することはほぼ不可能なのである。もう少しその辺の事情を説明して、フォーダーの絶望の淵がはっきり見えるようにした方がいいだろう。

まず、ある認知システムがモジュールであるのは、それが領域限定的 (domain specific) で、ニーロンの構造として生得的に組み込まれ、自動的・自律的に作動し、情報閉鎖的である、つまり任務の遂行に関して他のモジュールの情報を利用できず (informationally encapslated)、また逆に他のモジュールもその処理プロセス内の情報にアクセスすることができない (inaccessibility)、といった特徴をもつ場合である (Fodor, 1983: 37)。つまり認知モジュールは、担当領域に関して処理スピードが速く処理結果も正確であるが、その代わりに、トータルな状況認知においては錯誤や誤謬を誘発

する原因ともなる。例えば、水槽に差し入れた真っ直ぐな棒は水中で曲がっているとかわれわれには見えないが、いくら光の屈折に関する知識を動員しても真っ直ぐに見えるようにはならないし、いくらその見え方を反省してもなぜ曲がって見えるのかは分からない。認知モジュールは、入力した「刺激言語」をいわば盲目的に「中央システム言語」へと「翻訳」して出力するのである。

このような認知モジュールの内部では、その情報処理の仕組みは古典的に、つまりチューリングが考えた通りに計算論的に捉えられる。つまり、情報処理の対象である表象の統語論的な特徴のみに依拠したプログラムを書くことができる。したがって、表象がシステム全体に寄与する情報内容は、あくまで統語論的特徴に付随したものとしてみ処理される。しかし問題は、そのようなプログラムが依拠する表象の統語論的特徴は、当の表象が自分以外の表象と何の関わりもなしにそれ単独で持ちうるローカルな性質であり、ローカルな性質はその定義上、それが置かれる文脈に左右されない文脈非可感的 (context insensitive) な性質、あるいは文脈横断的な性質だ、ということである。例えば、その典型は、「P & Q」から「P」を推論する演繹的なプロセスが依拠する記号の形式的・合成的な配列だ。「P & Q」は、どんな文脈におかれようと、そのような配列で内部に「P」を含む、という性質を失うことはないだろう。ところが、フォォーダーが「グローバル」、「アブダクティブ abductive」、「全体論的」、あるいは「最上の説明への推論」と呼んでいるような関連性につつまれる種の推論は、ある表象がローカルにもつ性質に依拠してではなく、文脈ごとにもったりもたながったりする性質、つまり文脈可感的な性質に依拠して行われるのだ (Fodor, 2000: 23-29)。そして、ある表象を文脈可感的にするものは、その表象がいかなる信念システム全体に属しているのかという事

実である。

例えば、クワイン以来つとに有名となった信念システム（／理論）における〈保守性〉という性質を考えてみよう。信念システムが保守的だというのは、新しい情報が何も獲得されなければこれまで信じていたことを信じ続けるということ、そして、情報の更新によって信念システム（／理論）のどこかを修正しなければならぬ場合、できるだけ小さな変化ですむような修正を行うということである。しかし、もちろん、変化の大小は、当の修正を直接にこむる信念部分、もしくは表象部分の大きさによって単純に計るわけにはいかない。もし修正される信念部分が見かけ上は小さくとも、その信念システム（／理論）の中心的な部分だったとしたら、システムの整合性のために、多くの他の部分も修正されざるをえないだろう。要するに、信念システム（／理論）にとって、異なった構成部分は異なった程度の中心性をもつのである。したがって〈保守性〉とは、実は、自分の中心的主張はできるだけ手放すな、という信念システムに対する指針なのだ。しかし、ある信念もしくは表象の中心性は、まさしく文脈可感的な現象である。例えば、「自由落下する物体は重さに比例して加速する」という一般化は、重さに依拠した力学にとっては譲りえない中心的テーゼであったが、それ以降の質量に依拠した力学にとっては、たまたまの条件下で成立する偶然的な主張にすぎない。したがって、その同じ主張（／信念）が中心的であるかどうかは、信念システム（／理論）可感的である。ところが、その主張を表す表象の統語論的特徴は、それがローカルな性質であるがゆえに、信念システム（／理論）可感的ではありえないし、また、文脈の変化とともに変化することもありえない。しかし、古典的計算主義が計算できる（処理できる）ものは、表象の統語論的特徴だけなのであるから、表象

のグローバルな、もしくは全体論的な性質は、古典的計算主義がその本性上扱うことのできないものである。

フォードの苦悩が頂点にまで達するには、後ほんの一步である。確かに、この全体論的な性質を古典的な計算主義が扱うたった一つの方法が残されている。それは、問題の表象を含めて、その信念システム（理論）のすべての表象の統語論的特徴が与えられたなら、中心性や関連性といったすべての全体論的性質も決定される、という〈弱い計算主義〉をとることだ。しかしこれは、信念システム全体を計算の領域（ドメイン）とすることであり、人間のような認知システムの場合、領域が大きすぎて心理学的モデルとしては現実的ではない。それどころか、フォードによれば、それは〈破滅的な全体論〉を受け入れることにはかならない。⁽⁵⁾とすれば、もはや古典的計算主義以外のアプローチに頼るほかはないが、現在知られている唯一のそれは、コネクショニズムという、フォードの見たところまったく望みのない選択肢なのである。かくしてフォードにとって、『精神のモジュール形式』からほぼ二〇年たった現在、可能に思われるあらゆる脱出口は閉じられる。「チヨムスキーの言方を借りれば、心的過程がいかに同時に実行可能 (feasible) で、なおかつアブダクティブで、なおかつメカニカルであるか、ということは問題であるどころかミステリーである。実際、私の見るところ現状では、これと意識こそが、心に関しての最後のミステリーだと思われる」(ibid.: 99、強調は原著者)。

3 すべての表象が同時に重なり合っ (あるいはコネクショニズム)

皮肉なことに、認知科学をこれまで引張ってきた古典的計算主義は、実は、認知過程の最も中心的な部分、万能型知性の万能型たるゆえんである非論証的推論（日常的推論／アブダクティブな推論）を扱うことができずに、ただ、その中央システムに情報を提供する周辺部の入力モジュールにおいてだけ妥当するものであったのだ。少ない情報から世界の姿について整合的な仮説を立てること、新たな情報を取り入れつつその整合的な仮説を調整し維持し続けること、可能な状況を考慮に入れて適切な行為プランを決定すること、時間的な制約の下に最善の判断をすること、要するに、認知活動の最も認知らしい中枢部分（つまり哲学者のフレーム問題）は、古典的計算主義の下では手つかずのままに残されざるをえなかったのである。

さて、当然のことながら、古典的計算主義に対しては、すでに述べたように別のプロジェクトがそれに代わるものとして提案されてきた。紆余曲折をへて、一九八〇年代から再び認知科学において台頭してきたコネクショニズムがそれである。では、コネクショニズムはフレーム問題に対して、どれほど見込みがあるのか。以下の検討に際しては、コネクショニズムのアイデアそのものに対して強力に展開されてきた、思考の体系性／合成性に関するフォード、ピリシン、マクローリンらによる批判と、コネクショニズムの側からの応酬には深く立ち入らないことにする。⁽⁶⁾ というのも私の見るとこ

という観点が必要と思われるからだ。別の言い方をすると、基本的には〈表象とは何か〉をめぐるこの対立は、フレーム問題がどう解決されるか、というような実質的な課題からアプローチしていかない限り、認知科学における一種の「神学論争」とどまる恐れがある。「神学論争」とどめを刺すのは、多くの場合、「神学者」が軽蔑するほどプラクティカルな（しかし内容は困難かもしれない）問題の解決だということをし、われわれはさまざまな場面から学んで知っているはずである。

そこで、これから、フレーム問題に対するコネクショニズムからの具体的なアプローチを見てみることにしよう。それは、ラメルハートやスモーレンスキーらが構築した〈部屋を表現するネットワーク〉⁽⁷⁾におけるスキーマの出現とその働きである (Rumelhart et al. 1986: 7-38)。

しかし、その前にコネクショニズムに関して、必要最低限のことは述べておかなければならない。⁽⁸⁾コネクショニズムもまた、認知過程がある種の計算過程であると考えるのだが、そのモデルはおなじみのコンピュータではなく、われわれの脳にあるニューロンのネットワークだ。コンピュータがどれほど高速だろうと一度に一回きりの計算しか行わないのに対し、ニューラル・ネットワークは、多くのニューロン（ユニット／ノード）が同時並行的に数多くの計算を行う。一つ一つのユニットが行う計算はきわめて単純なものであり、要するに、入力される刺激値を足し合わせて、その合計（活性化値）が一定の値を越えたら、それに正負の重みづけを掛けて他のユニットに出力する、というだけのことである。ふつうのコンピュータの逐次処理に対して、ニューラル・ネットワークが並列分散処理といわれるゆえんだが、ユニットの数を莫大にすればこの人海戦術的な仕組みのおかげで、ネットワークは単純な計算法しか持たないにもかかわらず、驚くべき計算スピードを確保することができる。

3 すべての表象が同時に重なり合っ

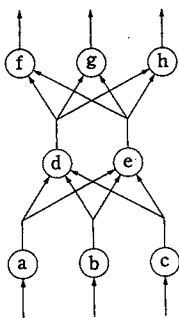


図3-1 フィードフォワード型の簡単なネットワーク

ユニット相互の結合の仕方はいろいろあるが、この種のモデルとしてよくお目にかかるのは、次の図3-1のような三層構造のフィードフォワード型モデルだろう。このモデルでは、入力層（ユニット a、b、c）から隠れ層（ユニット d、e）へ、隠れ層から出力層（ユニット f、g、h）へと、それぞれのユニットがとる活性化値が変換されながら波のように次々に伝わっていく。つまり、ネットワークの計算過程は、それぞれの層を $n1 \sim nn$ 個のユニットが構成している場合、 $n1$ 次元のベクトルが $n2$ 次元のベクトルに変換され、次にそれが \dots という具合に、連続的なベクトル変換を行う過程だと見ることができるとくにネットワークは、次の二つの特徴をもっている。まず第一に、計算過程全体が多くのユニットの活性化によって行われる結果、ネットワークの扱う表象は（どの表象をターゲットにするかに相対的に）多くのユニットに分散化され、それらが全体論的に重ね合わされて存在する。この結果、全体の計算に対するそれぞれのユニットの寄与は小さくなるので、局所的なノイズやダメージに強いという特徴がネットワークに与えられる。また第二に、ネットワークは、そもそもプログラムとデータの区別をもっていないので、いわゆる学習は、データの更新やプログラムの変更によってではなく、ユニット間の結合の重みづけを調整することによって行われる。その際、重みづけの調整は、ネットワークが自分で探し出す一定の最適な重みづけ配置に導かれるという意味で、自発的に行われる。

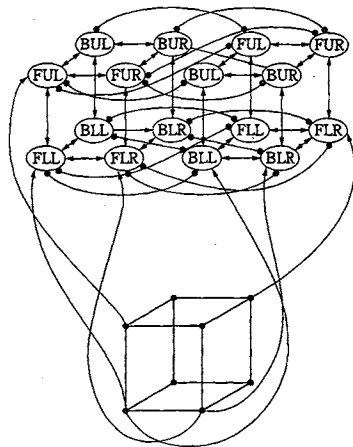


図3-2 ネットカー・キューブ解釈の制約を表すネットワーク

(Rumelhart et al., 1986: 10より)

を与えるための出力層はとくに用意されていない。むしろ、ネットワークの安定状態がいかなるものとなるかを決定するのはユニット相互間の結合であるから、興奮性もしくは抑制性の関係をもつユニットはすべてが双方向に結合されている。図3-2は、このようなネットワークの簡単な実例である。

図3-2の下のネットカー・キューブ (Necker cube) の各頂点とつながった一六個のユニットからなるこのネットワークは、ネットカー・キューブをどう見るかという二つの対立する解釈に対応する二つのサブ・ネットワークからなっている。それらは、左側の立法体状につながった八個のユニット群 (キューブを左向きに見る場合) と、同じく右側の立方体状の八個のユニット群 (キューブを右向きに

さて、ラメルハートらのネットワークを理解するには、以上に述べたことを少しばかり変更すればよい (Rumelhart et al., 1986: 8-11)。一番肝心な点は、ネットワーク全体の動作を、入力値 (問題) から出力値 (答) への (解答作業) と見るかわりに、ある入力 (興奮) を与えられたネットワークが全体として安定した活性化状態 (平衡) へと自分を落ちつかせる (安定化作業 relaxation) だと見ることである。したがって、この種の (制約—満足ネットワーク constraint satisfaction networks) には、解答

見る場合である。それぞれのユニットは、下のキューブの対応する頂点から入力を受け取るという設定となっており、ユニットに付けられた名前は、それに対応する頂点を〈前(F)か、後ろ(B)か〉、〈上(U)か、下(L)か〉、〈右(R)か、左(L)か〉のどちらかに解釈しているかを表している。例えば、それぞれのサブ・ネットワークの左下の二つのユニットは、キューブの左端の下の頂点から入力を受け取るという設定となっている。すると、左のサブ・ネットワーク側のそのユニットは、キューブの前面の左下にある頂点から入力を受けているという解釈を表し(したがって「FLU」)、右のサブ・ネットワーク側の対応するユニットは、キューブの背面の左下にある頂点から入力を受けているという解釈(したがって「BLU」)を表すだろう。各頂点はただ一つの解釈しか許さないという制約があるから、これらの二つのユニットは互いに強い抑制性の結合によってつながれる。また、各頂点の解釈はそれに隣り合う三つの頂点の解釈にも制約されるから、サブ・ネットワークの各ユニットは隣り合う三つのユニットと興奮性の結合でつながれる。最後に、二つの解釈は互いに両立不可能なので、二つのサブ・ネットワークの対応するユニット同士(例えばFLUとBLU)は、強い抑制性の結合で結ばれることになる。

すると、結合の強度をうまく調整すれば、このネットワークは全体として常に、対立する二つの解釈のどちらかに落ちつくようになるだろう。例えば、あるユニットに隣り合う三つのユニットと、それらと競合するもう一方のサブ・ネットワークの三つのユニットが共にオンになるなら、それらの効果は互いに打ち消しあう。このとき、問題のユニットに外部からわずかな刺激でも入力されると、それだけでそのユニットはオンの状態になろうとする。しかし、競合する三つのユニットがオンなの

対し、こちらではそれ以下の数のユニットしかオンとなっていない場合、問題のユニットへ刺激が入力されても、それはオフになろうとする。要するに、このシステムは全体として、ほとんど常に、一方のサブ・ネットワークのユニット群がすべてオンになり、別のサブ・ネットワークのユニット群がすべてオフになるような状態へと落ちつこうとするのだ。つまりこのシステムは、ラメルハートらが言うように、ネッカー・キューブを左向きのもものと見るかそれとも右向きのもものと見るか、という二つの解釈の一方の状態に安定化し、その最適状態 (optimal state) に留まろうとするのである。

そこでいよいよラメルハートらのスキーマのモデルに登場してもらおうが、それは、ネッカー・キューブの場合のような〈制約—満足ネットワーク〉の一種である。正確に言うと、彼らがモデル化したのは典型的な五つの部屋のタイプについてのわれわれの知識であり、彼らはそれを、互いに緩やかに興奮性もしくは抑制性の関係をもつ幾つもの家財道具から構成しようとした。その五つのタイプ《台所》、《オフィス》、《居間》、《浴室》、《寝室》は、それぞれに特徴的な家財道具を備えており、誰もがそれをすでに暗黙のうちに知っているだろう (ベッドは寝室にはあるが、浴室にあることは滅多にない)。そこで彼らは、オープン、暖炉、灰皿、ベッド、電話など四〇項目にのぼる日常的な家財道具をリストアップし (それらの中には「大きい」、「中くらい」といった部屋の大きさの項目も含まれていて)、被験者に、それらを備えている (ない) ような上の五つのタイプの部屋をいくつか想像してもらおうように求めた。

こうして集められた五つの部屋タイプに関する、総数八〇個の部屋記述を〈制約条件〉として用いることによって、彼らの〈部屋を表現するネットワーク〉が構築された。このネットワークでは四〇

3 すべての表象が同時に重なり合って

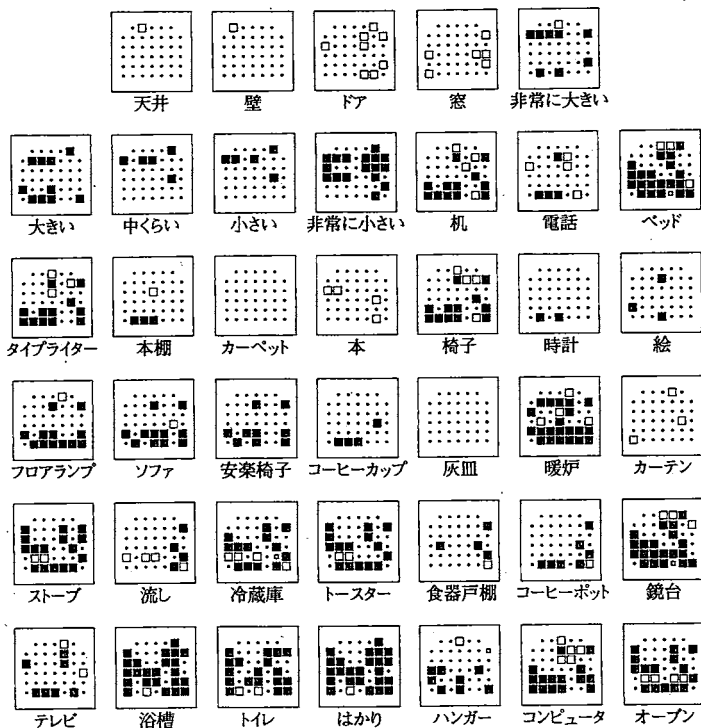


図3-3 家具を表す40個のユニットの相互結合

それぞれのユニットは正方形で表現されている。正方形の下に付けられた名前、各ユニットによって表現された項目を表している。各ユニットの内部では、小さな黒と白の長方形が、このユニットとシステム内の他のユニットとの結合状態を表している【黒は抑制性、白は興奮性の結合を示し、正方形の大きさが結合の強度を示す】。各ユニット内部での小さな正方形の相対的な位置が、このユニットの結合先のユニットを示している。[例えば、下端の右から二番目<コンピュータ>は、上から二段目の右の<ベッド>と抑制性の結合をし、その左の<電話>と興奮性の結合をしている]。

(Rumelhart et al. [1986] p.24 より)

個のユニットのそれぞれが四〇の項目を表し、それらは互いに、ネッカー・キューブの時のように、互いに同時にオンになる確率が高い場合（つまり同時に存在する確率が高い項目を表す場合）には強い興奮性の結合で結ばれ、同時にオンになる確率が低い場合（つまり同時に存在する確率が低い項目を表す場合）には強い抑制性の結合で結ばれ、相互に独立にオンになる確率が高い場合には強さ〇の結合で結ばれた。この結合強度の調整が、このネットワークの〈学習〉に相当する。また、ここでは、コネクショニズムの通例のネットワークに見られる（図3-1のような）〈隠れ層〉ユニットは用いられていない。その理由は、彼らによれば、ネットワークがスキーマ（フレーム）を体得するのに、〈隠れ層〉のような高次レベルのユニットは必要ない（一）からである（ibid.: 26）。図3-3は、各項目が他のどの項目とどのような結合によって結ばれているかを示した一種の「相関図」である。

それでは、このネットワークにおける部屋のスキーマの特徴の細部を、ラメルハートらの記述からまとめてみよう。

(α) スキーマは生成する（あるいは「プロトタイプ」）

〈オープン〉と〈天井〉をオンにすると、このネットワークは、〈コーヒーカップ〉を弱くオンにし、続いて〈流し〉と〈冷蔵庫〉を加え始め、部屋を〈小さい〉と解釈し、さらに〈トースター〉と〈コーヒーポット〉を加え、最終的に、〈ストーブ〉、〈時計〉、〈コーヒーポット〉、〈食器戸棚〉、〈電話〉、……等が最大に興奮した状態、要するに、〈台所〉のプロタイプに達する。これは、〈オフィス〉など他の四つの部屋タイプに関しても同様である。つまり、このネットワークは、五つの部屋タイプのスキーマをもっていて、いくつかの項目が満たされると、それに最も適したスキーマが自発的に生成さ

れ、その細部が補完され、その状態に安定化する。

(β) スキーマは互いに類似する(あるいは「四〇次元空間の地形」)

このネットワークは、スキーマ《オフィス》と《寝室》の方が、《台所》と《オフィス》および《台所》と《寝室》よりも互いに似ている、ということを知識している。それを示しているのが図3-4だ。この三次元座標は、ネットワークの状態を表す四〇次元の超空間のうち二次元だけを取って、そこで表されるスキーマの適合性 (goodness) を示したものである。二つの次元の値と共に、《オフィス》、《寝室》、《台所》の適合性が変化し、この地形 (landscape) がこの三つのスキーマに関して「山」と「谷」をもっていることが分かる。そこで例えば、《オフィス》の「山」の周辺は、この二次元の値(二つの項目の興奮度) に関して最も《オフィス》らしい《オフィス》であり、その「谷」の周辺は、最も《オフィス》らしくない《オフィス》だということになる。二つの「山」を結ぶ「尾根」はその二つの概念(スキーマ)の混合であり、《オフィス》と《寝室》を結ぶ「尾根」の方が、《台所》と《オフィス》および《台所》と《寝室》を結ぶ「尾根」よりも高い、ということがこれら類似性の違いを表している。もちろん、ネットワークはこの違いを地形として表すだけでなく、この地形の違いにしたがって振る舞う(安定化状態へ向かう)わけである。

(γ) スキーマは柔軟である(あるいは「例外への対処」)

(α) で見たように、それぞれのスキーマは標準の値をもっている。例えば、《浴室》は〈非常に小さい〉という標準値をもち、《台所》は〈小さい〉、《寝室》は〈中くらい〉という標準値をもつという具合だ。そこで、スキーマが標準から外れた項目をいくつか持ったらどうなるだろうか。その例外

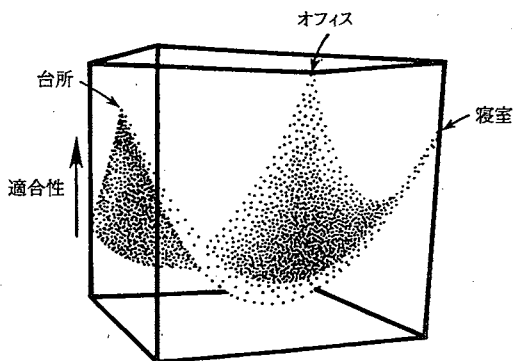


図3-4 〈台所〉、〈オフィス〉、〈寝室〉のプロトタイプに対応する適合性
(Rumelhart et al., 1986:28より)

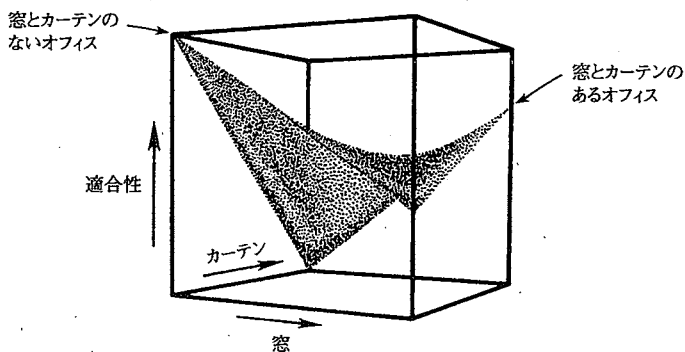


図3-5 〈カーテン〉と〈窓〉の興奮度の関数として見られた〈オフィス〉の適合性
(Rumelhart et al., 1986:35より)

によって、スキーマは働かなくなるだろうか。このネットワークでは、スキーマは整合性を求めて、他の項目の標準値を変更してその例外に対処する。例えば、《ベッド》と《ソファ》が共にオンにされると、その《寝室》は、《安楽椅子》と《フロアランプ》と《暖炉》を含むようになり、サイズを《大きい》に変更するようになる。要するに、この例外的状況の下では、《一風変わった大きな寝室》という変則的なスキーマが新たに出現する。

(δ) スキーマは構造を持つ (あるいは「合成可能性」)

大きなスキーマは、その内部にサブ・スキーマをもつような複合的な構造をもっている。例えば、互いに整合的で、いつも同じ場合にオン(かつオフ)になるような少数のユニット群は、それ自体が一つのスキーマのように働くようになり、幾つかの異なった大スキーマの構成要素となる。例えば、図3-5の《窓》と《カーテン》の場合を見てみよう。これら二つの要素は、常にどちらも現れるか、常にどちらも現れないかのいずれかの傾向をもっている。したがって、《オフィス》スキーマは、これらの要素に関して、それらが両方とも最大値「1」を取るか、最低値「0」を取る場合に、最大の適合性を示しているのが分かる。逆に、両方の項目が「1」と「0」の組み合わせの場合、適合性は最低となる。したがって、《窓∥カーテン》サブ・スキーマは、一つの安定した構造体として《オフィス》やその他のスキーマに取り込まれており、その意味で、サブ・スキーマ同士の新たな組み合わせによって新たなスキーマを創造する可能性が示唆されている。

(ε) スキーマは定義的情報ではない (あるいは「信念/知識の程度」)

このネットワークは、例えば、台所とはどのような種類の家財道具を備えた部屋か、ということ

述べた定義をもっているのではない。スキーマが定義なら、問題はそれに合致するかしないかだけであって、入力 of 解釈にいささかの融通性もないだろう。しかし、ネットワークが蓄積するスキーマは固定的な表象ではなく、ユニット間の可変的な結合にほかならないのだから、スキーマは、確固たる知識からあやふやな信念に至るまでのさまざまな種類の情報を表現することができる。このネットワークにおけるスキーマは、その時のデータとして与えられた家具の（各部屋タイプにおける）共存の頻度が、ユニット相互の結合の強さとして表現されることによって生成された、ということを思い出そう。つまり、ユニット相互間の結合の強さは、そのスキーマにおける家財道具のあり方の決定性、つまりスキーマの厳格性を決定しているわけである。そこで、たまたま、別の場合のデータでは、台所に備え付けの家具としては〈コーヒーポット〉と〈オーブン〉は常に必ず含まれるが、〈冷蔵庫〉はそう多くは含まれないとしよう（昔の北欧の台所か？）。すると、ここから生成される《台所》スキーマは、「台所には必ずコーヒーポットとオーブンがある」という（この場合の）確かな家具情報を表現し、「冷蔵庫はもしかするとあるかもしれない」という不確かな情報を表現しているだろう。さらに、スキーマを構成する相互結合の〈すべて〉がきつければきついほどスキーマ全体も厳格であり、ゆるければゆるいほどルーズなものとなる。最も厳格なスキーマは定義〈論理的含意〉のように振るまい、最もルーズなスキーマは純粋な連想〈自然な連合〉のように振るまいだろう。スキーマは、信念や知識についてのわれわれの直観に見あった〈存在の程度〉というものを表現している。

さて、それでは、コネクショニズムはフレーム問題を解決したのだろうか。少なくともこれまで見

た所では、コネクショニズムがきわめて有望であることは間違いない。われわれは、第1節でフレーム問題の特徴を大きく三つに分けてあげておいた。(i) 端的な無視、(ii) 関連性、(iii) 規則と例外、というその三つを思い出しながら、(順不同で) コネクショニズムの可能性を見積もってみよう。

まず、コネクショニズムでは、ある領域の問題を解決するのに、その領域の出来事を表現する形式的規則(プログラム)とそれが操作する表象(記号)というペアを導入する必要がない。ネットワーク・キューブの例でも、また部屋スキーマの例でも、それぞれのユニットは自分が担当する計算をするが、それらを束ねて、システム全体の状態を導いていくような規則は存在しなかった。それどころか、ネットワークが五つの部屋タイプの特徴を学習したとき、規則が扱うべきデータもまた、表象の形でシステムが蓄積する必要はなかった。すべては、ユニット相互間の結合の強さを調整することによって果たされたのである。したがって、そもそも規則が存在しない以上、(iii) にあげた規則と例外に関するフレーム問題は生じないように思われる。というのも、システムが従うべき規則が存在しないのだから、その規則への「重大な例外」をすべて前もって知る必要もないからだ。しかも規則が存在しないにもかかわらず、システムはでたらめではなく、規則に従っているかのような振る舞いをする。そして例外が実際に生じるような場合には、システムはいかなる上位規則に訴えることもなくその振る舞い方(スキーマ)をゆるめて、整合性を求めながら、例外に柔軟に対処したのである。先に見たスキーマの自発的な標準値の変更は、このシステムが、「特別な事情がない限り」という例外対処条項をすでに使いこなしているかのように思わせる。

また同様に、規則が規則として適用されるための前提となる背景要因もまた、規則と表象(記号)

というアーキテクチャとしては存在していない以上、それらの背景要因に言及し、それらをチェックするという必要もない。したがって、(i)で問題となった〈無視のパラドックス〉、つまり、ふつうは無視してかまわない背景要因をふつう通りに無視するためにも常にそれをチェックしなければならぬ(したがって、無視することができない)、という厄介な事態はそもそも生じようがない。例えば、図3-5に表された〈窓⇨カーテン〉サブ・スキーマを考えてみよう。これは二つのユニットが「ともにオンかともにオフ」という仕方でかなり強く結合された結果、一つのスキーマの如くに振る舞うようになったものだ。したがって、このサブ・スキーマは、部屋全体の様子を決定する相対的に独立した一つの新たな要因だと見なすこともできよう。例えば、図3-5に見られるように、「窓だけのオフィス」もしくは「カーテンだけのオフィス」というのは、典型的なオフィスではない。その場合、部屋全体の様子は、〈窓〉と〈カーテン〉の興奮度の組み合わせと、他の項目の興奮度との関係で最終的に決まるだろう。

つまり、〈窓⇨カーテン〉というユニットはこのネットワークには存在しないし、それを扱う規則も存在しないが、それが問題となる場合には、部屋全体の決定要因の一つとして働くようになる(その結果はあらかじめ予見できないが)。しかし、あくまでそうしたユニットが存在しない以上、それは背景要因であり、しかも〈窓〉と〈カーテン〉が共に最大値か共に最低値かをとるプロトタイプ的な状況では、この背景要因は特別な役割を何も果たしていないように思われる。つまり、力の場を形成する普段は目立たない要素(力)の一つのように、典型的な状況では、端的に無視されているように思われるのだ。このことは、すべての入力情報をいったん隠れ層のユニット群に分散しながら重ね合

わせるタイプのネットワークではいっそう明らかであるが、そのような隠れ層をもたない部屋スキーマの例でも十分に見てとることができる。というのも、潜在的な関連要因を場合によって「端的に無視する」一つのやり方は、それらをあからさまな操作（計算）対象にせずに、しかも、計算全体の過程に何らかの仕方ですれらを「常に」関わらせることだからである。

そのような計算のやり方、つまり分散化された表象の全体論的な重ね合わせこそが、実は、コネクショニズムの最もすぐれた特徴であり、(ii)の関連性の問題を解決する一つの有望な可能性でもある。そこで、このネットワークにおいて文脈の変化とは何だろうか。すべてのユニットがある重みづけで互いに結合されたこのネットワークでは、すでに五つのスキーマが生成され、ユニット群への入力を待っている。ユニット群は、そのつどの入力状況しだいで、あれやこれやのユニットがあれやこれやの強さで興奮し、それからスキーマに導かれ、残りのユニットもあれやこれやの強さで興奮する。つまり五つのスキーマは、そのつどの入力状況の変化に晒されて、それに対応し、その結果、ネットワーク全体が最終的にその時の安定化状態に落ちつくわけである。したがって、五つのスキーマにとつて、文脈とはそのつどの入力状態のパターンのことであり、文脈の変化とは、入力状態のパターンの変化のことだ。もし関連性の問題が解決されるなら、それは、入力状態の変化に依じて、ユニットの新たな適切な興奮度の組み合わせが生成されるという仕方、スキーマが柔軟に変化することだろう。そのためにネットワークは、あらゆる表象をあらゆる表象に分散化させて重ね合わせる。すると実際にはどうなるのか。

あらゆるものがあらゆるものに関連をもちうる、ということは、部屋スキーマのシステムでは単刀

直入に、すべての家財道具ユニットがお互いにすべて結合されている、という入力情報の関連の仕方に実現されている。つまり、このシステムでは、《台所》、《寝室》といった五つのスキーマ表象はそのつどすべての家財道具ユニットに分散化され、逆に、すべての家財道具の表象はそのつど一つの部屋スキーマの中に重ね合わされている。ここでは、家財道具ユニットは、ローカルな特徴によって個別化された単独の原子論的な表象として、スキーマ非可感的に（スキーマ横断的に）存在しているわけではない。例えば、《オフィス》と《居間》の両方に現れる《本棚》は、それぞれの場合で興奮度も違うし、二つのスキーマ生成への貢献の仕方も違うのだから、まったく同じ表象だというわけではない。むしろ文脈の変化、つまり他のユニットからの入力状態の変化が生ずれば、それに応じて、文脈可感的に残りの各ユニットの興奮状態が変化し、スキーマそれぞれ自身が新しい整合的な安定状態へとその構造を変化させる。しかも、部屋スキーマの例のように、ユニットの一つが一つの家財道具を表すという具合にユニットの局所性を固定的に考える必要はない。ネットワークによっては、例えば《本棚》のような表象も分散化して表現することが可能なのだから、以上の話は表象のどのレベルにも一般化することができる。したがって、文脈に応じた表象相互のダイナミックな関連性は、すべての表象が（ポジティブに／ネガティブに／ニュートラルに）何らかの仕方で《常に互いに》計算過程で関連しあうという認知メカニズムによって、少なくともこれまで説明してきた意味においては可能となるように思われる。

かくして、分散化され重ね合われた全体論的な表象というコネクショニズムの最大の特徴は、フレーム問題に悩むロボットたちにとって、将来、最も明るい希望の光となるであろう（たぶん）。

4 俺たちは天使じゃない (あるいは自然知性)

しかし、先のラメルハートらの論文を収めたコネクショニズムの二冊の「聖典」が世に出たのが今から二〇年ほど前の一九八六年であり、しかもそれ以降、コネクショニズム関連の研究文献が増え続けているにもかかわらず、いまだにフレーム問題が解決されたという話を聞かないのはなぜだろうか。そして少なくともわれわれ自身にしても、これまでのコネクショニズムの説明で、この問題が最終的に解決されたという気になれないのはなぜだろうか。

この最後の節で私は、コネクショニズムに対する二つの不吉な予感をごく簡単に紹介し、そしてそれが本当はどこからやってくるのか、ということに関していささか思弁めいたことを述べたいと思う。それは、そもそもフレーム問題とは、いかなる観点から見られた知性、いかなる文脈にある認知システムに対して生じるものであるか、したがって、いかなる解決をどこに求めるべきであるか、ということに関するものである。

コネクショニスト・ネットワークに関して指摘されてきた一つの問題点は、小さなネットワークによる比較的限られたデータ領域での成功を越えて、本当に、莫大な量の知識を扱うモデルへとそれを拡大(スケール・アップ)していくことができるのか、という懸念である。フレーム問題とはまさにシステムが複雑な課題に直面したときに生じるものであるから、ネットワークの能力が心理学的に見て現実味のないような単純な課題だけで訓練され、現実的な問題領域へと一般化するのが困難な

ようでは、コネクショニズムがフレーム問題を解決する見込みは小さいものとなるだろう。この点で、一九九八年段階においてヘイスレイガーとファン・ラパードが下している診断は、依然として、コネクショニズムにとって決して明るいものではない (Haselager & Van Rappard, 1998)。

彼らが組上に載せたのは、チャルマーズ (Chalmers, 1993)、クリスチャンセンとチャター (Christiansen & Chater, 1994)、ニクラッソンとヴァン・ゲルダー (Niklasson & van Gelder, 1994) などの文法解析を行うネットワークであるが、いずれも、少しへひねりを加えた (しかし、心理学的には現実味のある) 条件の下では、訓練セットに現れなかった新たなデータに対してきわめて不満足な能力しか示していない。例えば、クリスチャンセンとチャターのネットワークでは、動詞の複数形「eat, run」を選択するのに、単語「boy」が初めて、(主語) 名詞句の一部となって現れる場合にも、複数形選択の一般化は正しく行われる (したがって「John and boy from town eat」と答える)。さらに、この一般化は、複数形主語が所有格として現れても、既知の名詞「cats」の場合には正しく行われる (したがって「Mary's cats run」と答える) にもかかわらず、初めての名詞、例えば「girls」の場合には、複数形選択の一般化は正しく行われない (したがって、「Mary's girls run」とは答えられない) (Haselager & Van Rappard, 1998: 172)。ネットワークが示す部分的な一般化能力は、法的的な何かに裏打ちされた本物の力なのか、それとも訓練によって獲得された頻度の高い単なる偶然なのか。これは深刻な問題である。

そればかりか、ヘイスレイガーとファン・ラパードの論文には、もっと厄介なことに、分散化された表象は相互作用させるには限界があるという報告も見られる (ibid.: 174)。例えば、述語と対象に

関する分散化された表象を、もっと大きな分散化された表象構造に組み込もうとすると、どの対象がどの述語によって束縛されているかという情報が失われてしまう。つまり、「Ted gave Mary flowers」の分散化された表象を「Jane knows that p」の分散化された表象と組み合わせて、「Jane knows that Ted gave Mary flowers」の分散化された表象をつくると、〈誰が〉〈誰に〉〈何を〉与えたかを〈誰が〉知っているのか、という束縛情報が失われてしまうというのである。もしこれが改良不可能な欠陥であるなら、コネクショニスト・ネットワークのスケール・アップはほぼ絶望的であり、ことフレーム問題に関しては実際はほとんど何も進展はなかった、と言わざるをえないであろう。

コネクショニズムに関するもう一つの不吉な予感、処理する情報の量というよりは質、つまり、ネットワークの処理する情報が領域限定的ではなく本当に同時に多種多様なものでありうるか、という懸念からやってくる。フレーム問題の焦点が、多様な入力モジュールからの情報を統合的に処理する中央システムにあったことを思い出そう。もし単一のネットワークが異なった領域からの情報を同時に処理することができなければ、各モジュールからの情報の接触は起こらず、フレーム問題は解決の一步さえ踏み出すことはできないだろう。カミロフ・スミスは、これまでに一つのネットワークが複数の異なった領域、例えば言語や、空間や、物理的出来事からの情報を同時に受け取ったことは一度もない、と述べる (Karniioff-Smith, 1992: 180-81)。コネクショニストは、ネットワークが領域横断的であることを強調するために、どの領域を扱うどのモデルも同じ〈学習アルゴリズム〉を使用していると言いが、実際は、例えば言語習得をシミュレーションした同じネットワークは、その学習結果をすべてキャンセルしなければ、新たに物理的課題を扱うように訓練することはできない。それは

結局、それぞれのネットワークはそれぞれの特定領域の入力に縛られているということ、つまり、領域限定的だということにはほかならないだろう。

これは少なからず皮肉な結果である。もし異なった領域を受けもつ複数の多様なネットワークが最終的に一つに重ね合わせられないなら、結局のところ、フレーム問題はどうか解決されたということになるのだろうか。実際は領域限定的でしかないこれまでのネットワークの設計原理が前提される限り、あるネットワークが学習した結果を他の目的のために別のネットワークが利用する、ということが是非とも可能でなければならぬ (ibid.: 180-7)。というのも、認知の発達過程における表象の再記述 (representational redescription) という彼女のテーゼをどう評価するにせよ、われわれが、いったん獲得した知識からより高次で抽象的な知識を引き出し、それを別の目的のために意識的に用いることができるようになる、ということには確かだからだ。しかし、彼女は気づいていないように見えるが、ここで解決を迫られている〈ネットワーク統合問題〉とは、まさにコネクショニズムの衣を被ったフレーム問題そのものではないか。異なった領域からの関連情報が相互作用するための有効なメカニズム、その正体がいまだに謎なのだ、とフォードは言っていたのではなかったか。

このことはもちろん、複数のネットワークの統合という場面ばかりではなく、単一のネットワークの内部でも生じてくる。これまでのどのネットワークも〈課題〉をモデル化はすることはあっても、真に〈発達〉をモデル化することはなかった、と彼女が嘆いてみせるとき、そこでの問題も実は、新たな形をしたフレーム問題にはかならない。というのも、彼女によれば、ここでネットワークに要求されているのは、子供がある課題に立ち向かうときに利用する習得ずみの多種多様な関連知識や関連

経験を、いかにモデルの中に取り込むか、という問題の解決だからである。「発達とは、多くの源から知識を引き出し、それを目標に向かって用いるということだ。したがって、私の考えでは、現実の子供たちが現実の環境の中で学ぶやり方をモデル化しようとするなら、これまでよりはるかに豊かな入力ベクトルが必要になる」(Dreyfus)。しかし、〈ネットワーク統合問題〉も〈異種情報統合問題〉も、いまのところコネクショニズムにとって解決可能なものではないように思われる。すると再び(三たび?)皮肉なことに、認知活動の最も認知らしい中枢部分(つまり哲学者のフレーム問題)は、コネクショニズムによってさえ、まだ手つかずの領域として残されてきたことになるだろう。

私は、カミロフ・スミスのこの最後の言葉から始めたいと思う。それは、フレーム問題とは本来、現実の認知―行為システム(生存システム)が現実のワイルドな環境の中で実際に行為するときに初めて出現してくる問題ではないか、したがって、そのようなシステムの部分的なシミュレーションをいくら行ってもその本当の解決には至らないのではないか、という出発点の確認である。とはいえ、あらかじめ断っておいたように、残念ながら、それは解決ではなく解決への出発点の確認にすぎない。ロボットたちにとってのフレーム問題を思い出してみよう。それは単なる計算問題ではない。つまりそれは、状況のもつさまざまな制約の下でいかにある目的をうまく成しとげるか、というきわめて〈実践的な問題〉だったはずである。コネクショニズムによるフレーム問題の「解決」が一向に真の解決らしく見えないのも、ネットワークを組み込んだロボットが現実のワイルドな状況の中でそこそこに行為する、という程度の自立したシステムさえもまだ作成できないからである。ところが、多く

の認知科学者は、理想的な中央計算システムを追い求めるあまり、それが現実の状況の中で任務を果たす限りでしかフレーム問題は生じない、という単純な事実を忘れてしまったのではないか。言いかえると、達すべき目標と現実の状況とのギャップを時間と資源の制約の下で埋めるという観点から見るのではない限り、正しい計算はどこまでいっても正しい計算であり、うまい計算と下手な計算を区別することはできないだろう。そして、その時間と資源の制約が意味するものこそは、認知—行為システムの本質的な有限性にほかならない。だとすると、フレーム問題は人工知性というよりもそもそも自然知性一般に生ずるものであり、したがってまた、フレーム問題を解決する手だても、自然知性の仕組みの中に見出されるべきだろう。われわれ人間や動物は、純粹知性体のような途方もなく優れた計算システムを持っていてはわけではない。われわれは天使じゃないのだ。つまり、現実世界に住まう生存システムをフレーム問題の出発点に取るということは、自然知性においては、計算システムに介入する、あるいは計算システムを制御する計算以外の何らかの（生存のための）メカニズムがフレーム問題を解決しているのではないか、という選択肢を真剣に受け止めるということなのである。

そのメカニズムを感情に求めるのは、今やそう奇異なことではなからう。感情を論ずる何人かの哲学者／心理学者は、神や天使のような完全な知性体には感情がない、もしくは必要でないと主張する（de Sousa, 1987: 190; Oatley & Johnson-Laird, 2002: 170）。というのも、神ならばあらゆることを知り、あらゆることをなすことができるので、目的と現実のギャップが存在することはありえないからだ。いかなるものも彼が期待する以上の出来にはならないし、期待する以下の出来にもならないだろう。したがって、とオートリーとジョンソン・レアーは述べる。いかなるアクシデントも、いかな

る脅威も、いかなる失敗も挫折もないのだから、あこがれも寂しさもなく、それゆえ、神は感情をもたないのだ、と。逆に言うと、感情をもつのは、自分の行為プランと現実とのギャップを少しも気にとめないアリのような本能的行為者でもなく、また神や天使のような完全な論理能力の持ち主でもない、われわれのような中間的存在だけだろう。なぜならわれわれのような存在だけが、自分たちの能力以上の目的を追い求めてしまうからである。

オートリーとジョンソン・レアドによれば、目的と現実とのギャップ、つまりそこから生じる何らかの不整合は、感情だけでなく、アブダクティブな推論を発動させる原因でもある。というのは、もしあなたが何らかの不整合を発見したら、驚きや不快を経験するだけでなく、必ずや、その不整合が整合的なものとなるように思案するだろうからだ。その思案がアブダクティブな推論の正体である。例えば、あなたが、「人は危険を避けるものだ」ということは絶対の(？)真理だと思っていたとして、ところがあるとき、あなたは、友人のチカの運転する車が右折の際に反対車線に入り込み急停車した、という場面を遠くから眺める。あなたはこの状況を考えなおし、なぜチカはそんな危険な運転をしたのかということの説明を与えようとする(アブダクション)。彼女はアメリカから帰国したばかりで右側通行の癖がぬげず、うっかり反対車線に入り込んだのかもしいないし、車内にいた誘拐犯の存在を知らせようとして、わざと反対車線に入って車を急停車させたのかもしれない。いずれにせよあなたは、以前の信念のどれかを破棄して、整合性を回復しようとするだろう(Oatley & Johnson-Laird, 2002: 158-9)。パスやクワインを改めて持ち出すまでもなく、どの信念を破棄するかというアブダクティブな考慮は、単なる論理的推論によっては導き出すことの不可能な、それこそ

計算論的には手に負えない課題 (intractable task) である。それでは、オートリーとジョンソン||レ
アードは、そのアブダクティブな考慮を導くものこそが感情だと主張しているのだろうか。

感情とアブダクティブな推論の類似という彼らの論点は、確かに、われわれにとって奇妙な符合どころの話ではない。というのも、われわれはフォードの事例から、フレーム問題の核心にアブダクティブな推論が存在することを学んだばかりだからだ (第2節)。では、彼らの議論の中に、フレーム問題の最終的な解決を与える〈感情の理論〉がある、と言っているのだろうか。残念ながら現在のところ、そう言うことはできない。なぜなら、アブダクティブな推論と感情の機能が多くの点で類似していることの指摘は正しいとしても、両者の関係がはっきりとはされていないからである。ある箇所では感情がアブダクティブな推論の原因とされ、別の箇所では逆にアブダクティブな推論が感情の原因とされ、さらに別の箇所では、何らかの不整合の出現が両者の原因とされたりする。また、感情がアブダクティブな推論を導くという示唆がなされても、どんなメカニズムによってそれがなされるのか、という説明にまでは届いていない。したがって、感情は計算論的に手に負えない問題を扱うためのヒューリスティックスだ (Ibid.: 171)、という彼らの主張も細部の具体性に欠けているのである。

同様な不満を、私は、ド・スーサの提案に対しても感じざるをえない。彼によれば、天使にはなおフレーム問題が生じる (われわれは天使²をやっているわけではないから、そうこだわることでもない) と思うが、天使は論理的には完全であっても知識と力の点で有限なので……。フレーム問題が生じないのは、むしろ、感情をもたないもう一方の存在者、アリの方だ。その理由は、アリの認知システムが、知覚モジュールと運動モジュールが直結する以外に中央システムをほとんど何ももっていないからで

ある。古くから知られた問題解決法の一つは、問題の発生する条件を根絶することである。というわけで(？)、フレーム問題の解決は、情報の解釈や関連づけを行うことのできないアリのように、われわれの中央システムを一時的に情報閉鎖 (informational encapsulation) 状態に追い込むことである。感情がそれを行う。「行為と信念の決定における感情の機能は、〈純粹理性〉(＋単なる欲求) によって埋められないギャップを、知覚の情報閉鎖を模倣することによって満たすことである。これは、哲学者のフレーム問題を扱う母なる自然のやり方の一つだ」(de Sousa, 1987, 195)。

確かに、これはありそうなストーリーの一つである。しかし、情報閉鎖が行われた中央システムで、なおわれわれがアリよりも賢いのはなぜか。もし情報閉鎖がたんに無差別的な情報量の削減ならば、それによって、どうして情報の関連性が見出しやすくなるのか。逆に言えば、情報の関連性にあらかじめ訴えずに、関連情報だけを囲い込むような情報閉鎖をどうやって行うのか。言いかえれば、情報閉鎖は情報の関連性に可感的に行われなければならないのだから、このままでは、われわれは一步も前進したことにはならないだろう。つまり、われわれは文脈ごとに適切な情報閉鎖をどうやって行っているのか、という問題に答えなければならぬのだが、これこそまさにフレーム問題という悪夢の再現にほかならない。情報閉鎖の概念をたんに持ちだしても、情報閉鎖が中央システムの情報処理に具体的にどう組み込まれているのか、ということが明らかにされない限り、この悪夢を追い払うことはできないだろう。つまり、オートリーとジョンソン・レアドにしても、ド・スーサにしても、感情と計算システムの相互作用が認知—行為システムの具体的なメカニズムとして描かれるまでには議論を詰めていないのだ。したがって、フレーム問題で真に前進するには、彼らの到達点こそがわれわ

れの今後の出発点であらねばならないだろう。

なさねばならないことは何だろうか。まず第一に、感情の機能というものを徹底的に洗い出すことである。認知科学や人工知能は、機能としてシミュレーションしやすいものから攻め込んでいったがゆえに、いまや最も困難な領域が理論化を阻む壁として残されてしまった。それは、中央システムの計算メカニズムであり、感情の機能であり、意識とクオリアである。そしてそれは、生存システム全体という（研究）対象の単位が見失われてきた過程でもある。したがって、まず心理学的な現象としての感情の機能が、自然知性一般の生存にどのような貢献をしているのかが明らかにされねばならない。その際、成熟した大人のシステムよりは、当面、未成熟ではあるが（完全ではなく）完結した子供のシステムを説明のターゲットにするのも効果的かもしれない。もっともそれは、ブルックスの言う「表象なしの知能」(Brooks, 1991)にまで撤退し、そこに留まることではないだろうが……。

そして第二に必要なのは、いわばマクロなレベルで捉えられるそのような機能が、実際の中央システムにとっていかなる作用となつて実現されるのか、という具体的なメカニズムの探求である。これなくしては、先ほど見たように感情理論は、フレーム問題に対する外野からの遠吠えにしかすぎない。そのメカニズムはおそらく、もう一つのプログラムやもう一つのネットワークでもなければ、たんに「より豊かな入力ベクトル」といったものでもないだろう。ひよっとするとそれは、計算レベル（例えばコネクシヨニズム）の理論と言語が、物理的実現レベル（例えば神経生理学）の理論と言語に出会う中間次元で初めて捉えられる、まったく新たなスタイルのメカニズムなのかもしれない。これまでのところ認知科学は、マーに由来する認知の三層構造というところさえ、つまり「計算理論 computa-

tional theory」[「表現とアルゴリズム representation and algorithm」]、「ハードウェアによる実現 hardware implementation」という三つの説明のレベルの相対的な独立性を、多少は「味つけ」を変えながらであっても基本的に受け入れてきた (Mart, 1982)。それは、第2節でのホーガンとティーンソンのコネクショニズム (Horgan & Tinson, 1996) から、それに敵対するピリシンの古典的計算主義 (Pylyshyn, 1999) にいたるまで同じである。しかし、もしかすると、われわれが直面しているのは、この認知の三層構造という従来からの枠組みを、根底的に見直すことにつながるような事態かもしれないのだ。いずれにせよ、これ以上の思弁は今のところ空虚な言葉の羅列にしかならないが、フレイム問題の切っ先が深く認知科学の奥底にまで届いていることは確かであり、また、われわれがまだ未踏の大地を前にしていることも確かである。

ともあれ、ロボットがフレイム問題に悩まなくなる日、それは、ロボットがアリと天使の間に自らの存在論的位置を見出したときなのである。

注

(1) 私がかつてある本の中の小話としてこのデネットのストーリーを借用した理由は、もちろん、このストーリーが、オリジナルの小話を自分で工夫しようなどという気にさせないほど魅力的だったからである。その本、『ロボットの心』の第4章、「フレイム問題」の議論も参照して頂ければ幸いである。

(2) しかし同時に、ここで問題となっている推移関数は、必ずしも決定論的な関数である必要はない。古典的計算主義は、計算可能な推移関数の概念を非決定論的なものへと拡張し、決定論的な関数をその一般的な確率的関数の特殊例と見なすことができる。つまり、推移関数はその際、あるトータルな認知状態から別のトータル認知状態への関数ではなく、それぞれの認知状態に、潜在的に可能な後続のさまざまな認知状態の確率分布をあてがう関数となる(決定論的な関数は、先行の認知状態に、確率1の後続の認知状態をあてがう特殊な関数となる)(ibid.: 27-30)。しかしフレーム問題は、プログラムの推移関数が決定論的であることに起因するわけではない。というのも、ある規則がある一定の確率分布をそれぞれの項目に与えている場合、フレーム問題は、新しい情報の獲得とともにその文脈にふさわしい確率分布の再配置をいかに実現するのか、という問題として再現するからである。もちろん、再配置用の規則が複数用意されていたとすれば、この文脈にとくに関連するのがどの再配置規則であるかを、何が素早く確かに教えてくれるのだろうか？

(3) ふつう、コネクショニズムでは、フォーター流の〈思考の言語〉と表象の〈全体論的重ね合わせ〉は両立不可能だと主張されるが、ホーガンとティーンソンのコネクショニズムはその両立を前面に掲げる、必ずしも標準的といえないコネクショニズムである。彼らのその試みが成功しているかどうかはここでのわれわれの関心事ではないが、コネクショニズムにおいて表象がそもそもいかなる存在だと考えられるべきか、という問題はまだ未決着のきわめて重要な問題であり、わが国における突っ込んだ議論としては注6に挙げた戸田山ほか(二〇〇三)の諸論文を参照されたい。また、そこに収録されている美濃(二〇〇三)は、ホーガンとティーンソンの解説としても有益である。

(4) フォーターはすでに二〇年前の著作『精神のモジュール形式』で、領域限定的な認知モジュールと中央システムに関して次のように述べている。「たとえ入力系が領域限定的であったとしても、そうでない何らかの認知メカニズム「中央システム」がなければならぬ。……入力系が与える表象は、どこかで相互に接

- 触しなければならず、その相互接触を生み出すメカニズムは、それを生み出すというまさにそのために、一つの複数の認知領域から得られる情報にアクセスできなければならない」(Fodor, 1983: 101-02)。ところが「……誰も、そうしたグローバルな要素がどうやってその結果「信念の最終的確定」を生み出すのか理解する手がかりさえも持っていない。ここでは、認知科学はまだ始まっていないのだ」(Ibid.: 129)。
- (5) わが国ではあまり顧みられない、全体論についての〈素面な?〉考察に関しては、Fodor & Lepore, 1992: Ch. 1. をぜひ参照されたい。
- (6) 例えば、Fodor & Pylyshyn, 1988; Fodor & McLaughlin, 1990; Smolensky, 1988; Smolensky, 1991。あるいはわが国においては、戸田山ほか(二〇〇三)に収められた服部裕幸「分散表象」は認知の説明にたして役立つのか?」、金子善彦「表象なしのコネクショニズム」とその行方」、美濃正「新しい認知の理論としてのコネクショニズムの可能性」、柏端達也「コネクショニズムは素朴心理学に対して何か言えるのだろうか」の各論文を参照されたい。
- (7) スキーマとは、第1節で述べておいたように、基本的にフレームのことだと考えてよい。ラメルハートらのこの論文についての解説と論評は、Clark, 1989の邦訳(第五章「並列分散処理」)においても見ることが出来る。
- (8) 邦語文献によるより詳しい説明は、例えば、信原(二〇〇〇)、柴田(二〇〇一)、戸田山ほか(二〇〇三)などの関連箇所を参照されたい。
- (9) 自然知性におけるフレーム問題と、感情機能によるその解決の可能性に関しては、私も不十分ながら柴田(二〇〇一)でラフなスケッチを描いたことがある。ある事情からそこの議論が感情のクオリア問題に突っ走ってしまった感もあるが、素朴なコネクショニズムの問題点の指摘とともに、参照していただければ幸いである。

参考文献

- Brooks, R., 1991, "Intelligence without Representation", *Artificial Intelligence* 47. R.・ブルックス／柴田正良訳「表象なしの知能」『現代思想』一九九〇年三月号
- Chalmers, D. J., 1993, "Connectionism and Compositionality: Why Fodor and Pylyshyn Were Wrong", *Philosophical Psychology*, 6 (3).
- Christiansen, M. H., and N. Chater, 1994, "Generalization and Connectionist Language Learning", *Mind and Language*, 9 (3).
- Clark, A., 1989, *Microcognition*, Cambridge, MA: MIT Press. マンナイ・クラーク／野家伸也・佐藤英明訳、一九九七『認知の微視的構造』産業図書
- Dennett, D., 1987, "Cognitive Wheels: The Frame Problem of AI", in Pylyshyn, 1987. D.・デネット／信原幸弘訳「コグニティブ・ホイール」『現代思想』一九八七年四月号
- de Sousa, R., 1987, *The Rationality of Emotion*, Cambridge, MA: MIT Press.
- Fodor, J. A., 1983, *The Modularity of Mind*, Cambridge, MA: MIT Press. シェリー・A.・フォーター／伊藤笏康・信原幸弘訳、一九八五『精神のモジュール形式』産業図書
- , 2000, *The Mind Doesn't Work That Way*, Cambridge, MA: MIT Press.
- Fodor, J. A. and E. Lepore, 1992, *Holism: A Shopper's Guide*, Cambridge, MA: Basil Blackwell. シェリー・フォーター・マーネステ・ルサマ／柴田正良訳、一九九七『意味の全体論』産業図書
- Fodor, J. A., and B. P. McLaughlin, 1990, "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't work", *Cognition* 35.

- Fodor, J. A., and Z. Pylyshyn, 1988, "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition* 28.
- Haselager, W. F. G. and J. F. H. Van Rappard, 1998, "Connectionism, Systematicity, and Frame Problem", *Mind and Machines*, 8.
- 眼識科学'二〇〇三「分散表象」は認知の説明にはたして役立つのか?」戸田山ほか(二〇〇三)所収
- Horgan T., and J. Tienson, 1996, *Connectionism and the Philosophy of Psychology*, Cambridge, MA: MIT Press.
- Karmiloff-Smith, A., 1992, *Beyond Modularity*, Cambridge, MA: MIT Press. A・カッロニスミス／小島康次・小林好和訳'一九九七『人間発達の認知科学』シネルヴァ書房
- 柏端達也'二〇〇三「コネクショニズムは素朴心理学に対して何か言えるのだろうか」戸田山ほか(二〇〇三)所収
- Lepore, E., and Z. Pylyshyn, eds., 1999, *What Is Cognitive Science?*, Malden, MA: Basil Blackwell.
- Loewer, B. and G. Rey, eds., 1991, *Meaning and Mind: Fodor and His Critics*, Cambridge, MA: Basil Blackwell.
- Marr, D., 1982, *Vision*, New York: W. H. Freeman and Company. マーケット・ワー／乾敏郎・安藤広志訳'一九八七『ビジョン』産業図書
- McClelland, J. L., D. E. Rumelhart and the PDP Research Group, 1986, *Parallel Distributed Processing*, vol. 2, Cambridge, MA: MIT Press. D・E・ラメルハート、J・L・マツランズ、PDPリサーチグループ／甘利俊一監訳'一九八九『PDPモデル』産業図誌
- 美濃正'二〇〇三「新しい認知の理論としてのコネクショニズムの可能性」戸田山ほか(二〇〇三)所収

- Minsky, M., 1975, "A Framework for Representing Knowledge", in Winston, 1975. M・ミンスキー／白井良明訳「一九七九「知識を表現するための枠組み」」P・H・ウインストン編『コンピュータビジョンの心理』所収、産業図書
- Moore, S., and M. Oaksford, 2002, *Emotional Cognition*, Amsterdam: Jonson Benjamins Publishing Co.
- Niklasson, L. F., and T. Van Gelder, 1994, "On Being Systematically Connectionist", *Mind and Language*, 9 (3).
- 田原幸彦「11000『サバイバル・サバイバル』雜誌社版七集巻」
- Oatley, K., and P. N. Johnson-Laird, 2002, "Emotion and Reasoning to Consistency", in Moore and Oaksford, 2002.
- Pyllyshyn, Z. W., ed., 1987, *The Robot's Dilemma*, Norwood, NJ: Ablex Publishing Co.
- , 1999, "What's in Your Mind?", in Lepore & Pyllyshyn, 1999.
- Rumelhart, D. E., P. Smolensky, J. L. McClelland, and G. E. Hinton, 1986, "Schemata and Sequential Thought Processes in PDP Models", in McClelland et al., 1989.
- 柴田昌成「11001『ロキミム』雜誌社版七集巻」
- Smolensky, P., 1988, "The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pyllyshyn", *Southern Journal of Philosophy* 26, Supplement.
- , 1991, "Connectionism, Constituency, and the Language of Thought", in Loewer & Rey, 1991.
- 戸田山和久・服部裕幸・柴田正良・美濃正編著「110011『心の科学』』監訳
- Winston, P., ed., 1975, *The Psychology of Computer Vision*, New York: McGraw Hill.