コネクショニズムと消去主義についてのノート：
ラムジー・スティッチ・ガロンの議論への応答

美濃　正（大阪市立大学）

『コネクショニズムの哲学的意義の研究』pp.22-30
平成 12 ～ 14 年度科学研究費成果報告書(2003)
研究課題番号 12410003
研究代表者：南山大学人文学部教授・服部裕幸

本科研の研究チームメンバーによって編まれた書物（『心の科学と哲学：コネクショニズムの可能性』昭和堂、現在印刷中）に寄稿した拙論（「新しい認知の理論としてのコネクショニズムの可能性」）において、私は認知の理論としてのコネクショニズムの可能性を擁護する議論を展開した。それは実質的には、コネクショニズムは心的なものに関する消去主義にはつながらない、という主張の展開であった。これに対して、数多くの哲学者（および認知科学者）が正反対の主張、つまり、コネクショニズムが正しい認知の理論であるとしたら必然的に心的なものに関する消去主義が導き出される、という主張に与する議論を提出してきている。そのなかでも、最も有名でかつ論議の的となることも最も多かった議論の一つは、ラムジーとスティッチとガロンがその共著論文（Ramsey-Stich-Garon (1990)）において示した議論であろう。（じっさい、上に言及した書物に寄せられた論文のなかにも彼らの議論を取り扱い対象としているものがある。そのなかでも代表的なものは柏端論文（柏端(2003)）であり、これはほとんど全篇、彼らの議論に対する直接的な応答とみなしてよい内容となっている。）

しかしながら、上記拙論においては、ラムジーらの議論に対するコメントを行う余裕がなかった。主たる理由は、消去主義が拙論の中心テーマではなかったからである。そこで、この報告において、彼らの有名な議論についての批判的検討を行うことにしたい。何と言っても、仮に彼らの議論が正しいとすれば、拙論で展開された主張は崩されることになるのだから、彼らの議論は拙論にとっても無視できるものではないのである。

さて、Ramsey-Stich-Garon(1990)におけるラムジーらの主張は次のようなものである。まず、いわゆるフォーク・サイコロジー（心的なものの概念はそのうちにいわば体現されている）は、命題的態度と呼ばれる心的状態（信念、欲求など）の「命題的モジュール性」を前提している(cf. 315f.*)。彼らによれば、「命題的モジュール性」とは、命題的態度の次の三つの性質の連言を意味する。すなわち、

(1) 各命題的態度は「機能的に分離可能 functionally distinct」、つまり、単独で獲得ないし喪失されることが可能である。

(2) 各命題的態度は「意味論的に解釈可能 semantically interpretable」、つまり、「　．．だと信じる」とか「．．したいと欲する」といった形で記述されることができ、しかも命題的態度のそのような記述は法則論的説明において有効である。

そして、

(3) 各命題的態度は他の命題的態度や行動の産出にさいして「因果的役割を

果たす」ことができる（場合により、果たさないこともできる）。

これに対して次に、認知のモデルとしてのコネクショニスト・モデルは、彼らによれば、典型的に次の二つの性質を有する(cf. 320f.)。

(1) これらのモデルにおいては情報のコード化はユニット間の結合の重みづけと各ユニットのバイアス値のうちに実現されるが、このようなコード化は局所的コード化ではなく、広く分散的なコード化である（いわゆる「分散表象」）。

(2) 隠れ層の個々のユニットはいかなる記号的解釈も許さない。それらはサブ・シンボリックである。

そして最後に、ラムジーらは、このような性質を有するコネクショニスト・モデルは、命題的態度の「モジュール性」を満たすことが不可能である、と主張する(cf. 324ff.)。したがって、もしコネクショニズムが認知の理論として正しいのなら、フォーク・サイコロジーは根本的に誤っており、少なくとも命題的態度という心的状態はじつは存在しえない、と主張するのである。

ラムジーらの議論の検討に入る前に少し先回りして言っておくと、彼らの議論を批判するさいの私の基本的ストラテジーは、コネクショニスト・モデルと命題的モジュール性が両立可能であることを示す、というものである。この点では、上記柏端(2003)の方針と変わるところはない。ただ、ラムジーらの議論により密着した仕方で、今述べた論点を示していきたい。

さて、ラムジーらの議論は、彼ら自身が制作した一つないし二つのシンプルなコネクショニスト・モデルに即して展開されている(cf. 324f.)。これらのモデルについてごく手 短に説明しておこう。図１に示されているように、これらのモデルはいずれも 16 個のユニットから成る入力層、４個のユニットから成る隠れ層、そして１個のユニットだけをもつ出力層という三つの層から構成されている。その唯一の機能は、入力層に与えられた命題の真偽を判定することである。二つのモデルの違いは次の点にある。第一のモデル（ネットワークＡ）は、以下の 1 から 16 の命題について、その真偽を適切に判定できるように訓 練されたものである。それに対して第二のモデル（ネットワークＢ）は、それに加えて以下の 17 番目の命題についても適切に真偽判定する能力を与えられたものである。

	命題	入力	出力
1	犬には毛がある	11000011 00001111	1 真
2	犬には爪がある	11000011 00110011	1 真
	.....		
5	猫には毛がある	11001100 00001111	1 真
	.....		
15	犬にはヒレがある	11000011 00001100	0 偽
16	猫にはヒレがある	11001100 00001100	0 偽
17	魚は卵を産む	11110000 11001000	1 真

図１

では、コネクショニスト・モデルと命題的モジュール性の両立不可能性を根拠づけるラムジーらの議論はどのようなものであろうか。主要な議論は次の一節に含まれている。

．．コネクショニスト・ネットワークのうちには、ある特定の命題を表象する役目をつ　とめるネットワークの分離可能な状態あるいは部分、というようなものはまったく存在しない。［というのも］ネットワークＡにおいてコード化されている情報は、全体論的　に蓄積され、ネットワーク全体に分散している。情報がネットワークＡから取り出される－－それは、入力コード列をネットワークに与え、ネットワークが出力ユニットへ高　い値を算出するかそれとも低い値を算出するかを見ることによってなされるのだが－－　ときはいつでも、多数の結合の強さ、多数のバイアス、多数の隠れユニットが、その計　算において役割を果たす。また［逆に］、個々のどの重みづけ、バイアス、ユニットも　多くの異なる命題についての情報のコード化に関与している。それゆえ、ある特定の命　題の表象がネットワークの計算において因果的役割を果たしているのかどうか、と問う　ことは端的に意味をなさないのである。記憶に関するわれわれのコネクショニスト・モ　デルが、常識心理学における命題的モジュール性と根本的に食い違うと思われるのは、まさにこの点においてである。(326-327)*

この主要議論は次のように整理できるだろう。
(1) ネットワークＡは、先に述べた 16 個の命題の真偽に関する情報を蓄積（記憶）してい　ると言える。
(2) しかし、この情報蓄積はネットワーク全体に分散された仕方での蓄積である。つまり　個々のどの命題に関する情報の蓄積にも、ネットワークＡのすべての部分ないし部分状　態が関わっている。言い換えれば、ネットワークＡのどの個々の部分ないし部分状態も 16 個すべての命題に関する情報蓄積に関与している。
(3) したがって、ネットワークＡのどの個々の部分ないし部分状態をとっても、それがある特定の命題だけに関する情報を蓄積している、ということは言えない。
(4) したがって、ネットワークＡのうちには「機能的に分離可能で、意味論的に解釈可能　な」部分ないし部分状態は存在しない。
(5) したがってまた、ある特定の命題の表象がネットワークＡによる［ある特定の］情報　処理において因果的役割を果たしているのか否かと問うのは無意味である。なぜなら、「ある特定の命題の表象」と呼ぶに値するもの、つまり「機能的に分離可能で、意味論　的に解釈可能な」部分ないし部分状態はそこには存在しないからである。
(6) ゆえに、ネットワークＡにおいては（したがってまたコネクショニスト・ネットワーク一般においても）、命題的モジュール性は成立しえない。

ラムジーらの主要議論に関しては、まずもちろん前提(1)を問題にすることができるだろう（なぜなら、ネットワークＡはあまりにも単純なシステムだから、そもそもそれに対して何らかの命題的態度を帰属させることに意味があるのかどうか、問題であろうから）。しかし、より大きな問題は、(3)から(4)への推移にあるだろう。ラムジーらがネットワークの「［部分］状態」と言うとき彼らの念頭にあるのは、上の引用から明らかなように、「あるユニットの活性化値はかくかくである」とか、「あるユニット同士の結合の強さはしかじかである」とか、「あるユニットのバイアス値はこれこれである」というような状態である。これらはすべてネットワークの非傾向的状態である。そして、これらの非傾向的状態に関するかぎり、ラムジーらによる(3)から(4)への推論は正しいと考えられる。しかし、ネットワークは傾向的状態 dispositional states ももっている。だから厳密に言えば、(3)から(4)は帰結しない。ラムジーらの議論はこの点で明らかに

不十分である。

それでも(4)は結局、正しいのだろうか。少なくともラムジーらがその正しさを示して はいないことはこれから明らかにしていくが、このように傾向的状態あるいはむしろ傾向性 disoposition こそがネットワークにおける「命題的モジュール性」の担い手であり、それゆえにコネクショニスト・モデルは「命題的モジュール性」と両立可能であると考える点においても、私は柏端(2003)と軌を一にしている。柏端(2003)の特に第２節の論述はこの点に関する非常に有益な議論を含んでいるが、本稿では、先ほども述べたように、柏端論文では明示的に取り扱われてはいないラムジーらの議論の諸側面を検討していくことに努めたい。

さて、ラムジーらは先ほど検討した主要議論を補う二つの補助的議論を提示している。これらはともにネットワークＡとネットワークＢの比較対象にもとづく議論である。しかし、これらの議論の本質的ポイントは主要議論と変わるものではなく、したがってまた主要議論と同様の問題点を抱えている。そのことを次に示したい。

補助的議論の第一は次のようなものである(cf. Ramsey-Stich-Garon(1990), 327-328)。ネットワークＢは、最初の 16 個の命題に加えて、17 番目の命題「魚は卵を産む」に関する情報をも蓄積しているとみなされるネットワークである。つまり、それはネットワークＡよりもこの命題一つ分だけより多くの情報を蓄積している。しかし、ネットワークＢのユニット間結合の重みづけ配置や隠れユニットのバイアス値は（したがってまた、同じ入力が与えられたさいの隠れ層の活性化パターンも）、ネットワークＡのそれとは驚くほどに異なっており、しかもその相違はネットワーク全体に及んでいる（図２と図３を参照）。このことが意味するのは、17 番目の命題（だけ）に関する情報の蓄積に関与しているとみなすことのできる、機能的に分離可能でかつ意味論的に解釈可能な「部分構造」(328)は、ネットワークＢのうちには存在しない、ということである。コネクショニスト・ネットワークにおける情報蓄積のあり方が命題的モジュール性と両立するものではないことは、このような仕方でも確認できるであろう。

図２、図３

以上のような第一の補助的議論の問題点が主要議論の場合と同様であることは、もはや明らかだろう。17 番目の命題に関する情報蓄積に関与する、機能的に分離可能で意味論的に解釈可能な「部分構造」は、たしかにネットワークＢのうちにはないだろう。しかし、このような「部分構造」は明らかにネットワークの非傾向的な状態である。ネットワークの傾向的状態もしくは傾向性への考慮が、ここでもまた脱け落ちている。したがって、コネクショニスト・モデルと命題的モジュール性との非両立という結論は、やはり導かれない。

さらに、第一の補助的議論は、「機能的分離可能性」概念についてラムジーら自身が一種の誤解を犯していることを鮮やかに示している。命題的態度の機能的分離可能性とは、彼ら自身の定義によって、各々の命題的態度が単独で獲得ないし喪失されうる、ということであった。ところで、（一定の前提のもとにではあるが）ネットワークＡは明らかに一つの命題的態度、つまり「魚は卵を産む」という信念を単独で獲得することができるのである。どのようにしてか？ネットワークＢになることによってである。逆に、ネットワークＢは明らかに一つの命題的態度、つまり「魚は卵を産む」という信念を単独で喪失することができる。もちろんそれは、ユニット間結合の重みづけ配置と各ユニットのバイアス値をネッ

トワークＡと同じものに戻すことによってである。したがって、ラムジーら自身の元々の定義に従うかぎり、コネクショニスト・ネットワークは機能的に分離可能で意味論的に解釈可能な何かをもっていることになるのである。

次に、第二の補助的議論の検討に移ろう。この議論は次に示すように、命題的態度を表す意味論的述語の投射可能性を直接の攻撃対象とするものである。

さて、すでに確認したように、ネットワークＡとネットワークＢはともに同じ16個の信念（そのなかには「犬には毛がある」という信念も当然含まれる）を共有するとみなしうるかもしれないモデルでありながら、その重みづけ配置やバイアス値に関しては驚くほど異なりあっていた。しかし、

このことはこれら二つに限られるというわけではない。もし、われわれが［上に示した］17個の命題に加えてさらにいくつかの命題（もしくは、そこからいくつかを減じた命　題）についてネットワークの訓練を行ったとすれば、ネットワークＡとＢとが互いに相違し合うのと同じ程度にこれら両者のいずれとも相異なるような、さらに別のネットワークが得られるだろう。このことから得られる教訓はつぎのことである。つまり、犬に　は毛があるという情報をネットワークＡと同じように表象する無際限に多くのコネクショニスト・ネットワークが存在するけれども、これらのネットワークはコネクショニスト理論の言語で記述可能であるようないかなる投射可能な特徴をも共有してはいない、ということである。(329)

要するに、同じように（たとえば）「犬には毛があると信じている」と記述可能とみなされる多数のコネクショニスト・ネットワークは互いにあまりにも異なり合っており、「混沌とした選言的集合」を成すにすぎない。したがって、この意味論的記述（述語）は法則論的説明に有効なものではない。つまり、上述の命題的モジュール性の条件(2)を満たさ ない。それゆえ、もしコネクショニズムが認知のモデルとして正しいのならば、やはり命題的モジュール性は成立しえない、というのである。

しかしながら、ここにもまた繰り返し指摘してきた同じ問題点があることが容易に見てとれるだろう。たとえば「犬には毛があると信じている」と記述可能かもしれないこれら多数のコネクショニスト・ネットワークのすべてが共有し、しかもコネクショニストの言語で記述できる共通特徴は存在しないのだろうか。ネットワークの非傾向的特徴だけに注目するならば、おそらくそうであろう。しかし、これも繰り返し述べてきたように、ネットワークは多数の傾向的特徴ももっている。そしてじっさい、この点に注目するならば、問題の共通特徴は容易に見出されるのである。つまり、それは＜「犬には毛がある」に対応する同じ入力コード（11000011 00001111）が与えられたならば、出力1（に近い値）を産出するだろう＞という傾向性である（図４を参照）。このように、ネットワークの傾向性ないし傾向的状態に注目することによって、コネクショニズムと命題的モジュール性の両立をはかる有力な方途を得る見込みが生じるのである。

図４

しかし、ラムジーらは、本稿で示唆してきた命題的態度をネットワークの傾向性ないし傾向的状態と同一視しようとする解決策をわざわざ取り上げ、それに対する批判を行っている(Ramsey-Stich-Garon(1990)，333)。そこで最後に、彼らによるこの批判について考 察しておこう。それは次のようなものである。

与えられた入力、たとえばｐという入力に反応してネットワークはある特定の活性化値をとる。それはまた、他の入力、たとえばｑやｒという入力に対しては別の活性化値をとる傾向性をももっているかもしれない。しかし、これらの別の傾向性がｐに対するネットワークの反応において因果的役割を果たしている、あるいはさらに言えば、果たして　いないという主張を明確に解釈する術はないのである。(ibid.)

少なくともネットワークＡのような単純なシステムに関するかぎり、ラムジーらのこの議論は混乱しているとしか考えられない。命題ｐ（「犬には毛がある」であるとせよ）が 入力された場合のネットワークＡの反応において因果的役割を果たす傾向性として考えうるのは、＜「犬には毛がある」に対応する入力コードが与えられたなら、出力ユニットの活性化値は１（に近い値）になるだろう＞という傾向性以外にはない（図４を参照）。ネ ットワークＡは入力層にｐに対応するコードが打ち込まれ、かつ、この傾向性を有するが ゆえに、出力ほぼ１を産出したのである。これが、ｐが入力されたときにネットワークＡにおいて生じる認知的エピソードのすべてである。そこに、別の命題ｑやｒに関する傾向性（たとえば＜「猫にはヒレがある」に対応する入力コードが与えられたなら、出力ユニットの活性化値は０（に近い値）になるだろう＞）もまた因果的に関与すると考える理由は何 もない。それはちょうど、割れ易さと磁性という二つの傾向性をもつ物体が床に落とされたために割れた場合に、そこに関与した傾向性は割れ易さだけであって磁性ではない、としか考えられないのと同様である（柏端(2003)の第２節を参照）。

結局、ラムジーらの議論は、「ネットワークＡのようなシステムにおいては、どの特定の命題に関する情報処理にもネットワークのすべての（部分的な）非傾向的状態が等しく関与するのだから、同じことがネットワークのすべての傾向的状態についても成り立つ」という独断的な前提にもとづく議論である、と判断せざるをえない。しかし、この前提は、柏端(2003)の第２節において説得的に論じられているとおり、誤っている可能性が高いのである。

最後に、明確な議論の形に仕上げられてはいないが、Ramsey-Stich-Garon(1990)において事例として何度か挙げられているケースについて考察しておきたい。それは柏端(2003)によって「等効力性 equipotency のケース」と呼ばれ、同論文の第３節において詳細に検討されている。本稿ではごく手短な考察に限定するが、その事例の一つは次のようなものである。

クルーゾーは執事が嘘をついていると推理したが、そのとき彼の推理に（因果的に）関与したのはどの信念だったのだろうか。クルーゾーは、列車は回送列車だったとも信じていたし、ホテルは閉館だったとも信じていた。可能性は次の三つである。1)クルーゾーは前者の信念だけによって執事の嘘という結論に到達した。2)彼は後者の信念だけを用いて同じ結論を導いた。そして、3)彼は二つの信念を両方とも使って同じ結論に導く推理を実行した。常識心理学的に言えば、もちろんこれら三つの可能性の区別は有意味な区別であり、じっさいクルーゾーの推理は三つの可能な道筋のいずれかを辿ったはずだとしか考えられない。しかし、コネクショニスト・ネットワークにおいても、これらの可能性の、同様に有意味な区別をつけることがはたして可能なのだろうか？

答えは、もし各命題的態度をネットワークの何らかの傾向性ないし傾向的状態と同一視できるならば、もちろん可能だ、というものである。容易に考えつくストーリーの粗筋は次のようなものである。クルーゾーと同等の推理を行うような

ネットワーク（それはもちろんネットワークＡなどとは比較にならないほど複雑なシステムでなければならないだろうが）があるとすれば、それが推理の道筋1)を辿るときと、2)を辿るとき、そして 3)を辿るときとでは、推理に臨んでのネットワーク全体の活性化パターンが互いに大きく（しかし、詳しくは説明できないが、あるシステマティックな仕方で）異なることだろう。たいへん粗っぽく言えば、因果的に関与する傾向性ないし傾向的状態に応じて、たとえば道筋 1)の場合には執事の発言内容に関する情報と「列車は回送列車だった」という命題に関する情報とを「重ね合わせ」た仕方で実現するような活性化パターンが出現することだろう（道筋 2)の場合、同じく 3)の場合にも同様である）。推理において active な傾向性は単なる潜在的傾向性のままにとどまることはできないはずだから、このように考えるのはきわめて自然だと思われる。そして、推理に臨んでのネットワークの活性化パターンがこのように取られる道筋に応じて異なるのだから、その後、出力に到るまでネットワークが辿るであろう状態の推移過程もそれに応じて相異なることだろう（もちろん到達される出力、つまり結論はどの道筋の場合にも同じであるが）。ともかく、コネクショニスト・ネットワークが「等効力的」な推理を行うさいにこのようにして様々の道筋を辿ることが可能だろう、というストーリーのアプリオリな不可能性を説得的に示すような議論は、Ramsey-Stich-Garon(1990)のどこにも見当たらないのである。

以上の検討が当を得たものであるならば、「コネクショニズムが正しければ、命題的態度に関する消去主義が導き出される」ことを示そうとしたRamsey-Stich-Garon(1990)の企ては不首尾に終わっている、と判定すべきであろう。結局やはり「ひょうたんからコマ」は出てきそうもないのである。

＊ Ramsey-Stich-Garon(1990)からの引用、およびそれに対する参照は、本稿ではすべて Macdonald-Macdonald(1995)に所収の版にもとづいて行う。

文献

柏端(2003): 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」　戸田山和久他編『心の科学と哲学：コネクショニズムの可能性』昭和堂（印刷中）　所収.

Macdonald-Macdonald(1995): Cynthia Macdonald and Graham Macdonald (eds.), Connectionism: Debates on Psychological Explanation, Blackwell.

Ramsey-Stich-Garon(1990): William Ramsey, Stephen Stich, and Joseph Garon, "Connectionism, Eliminativism and the Future of Folk Psychology", Philosophical Perspectives 4; Reprinted in Macdonald-Macdonald(1995), pp.311-338.