美作女子大学・美作女子大学短期大学部紀要
1991, Vol. 36, 1〜10

認知理論における「自己像」の機能的役割

妻藤真彦

「意識」を認知モデルに取り込もうとする試みがいくつかなされている（e.g., Johnson-Laird, 1983; 1988; Yates, 1985)。しかし，そもそも「意識」が何であるかについての，議論自体がまだ錯綜しており(e,g., Dennett, 1988; Eccles, 1976; Marcel, 1988; 妻藤，印刷中; Savage, 1976; Sperry, 1976; Weimer, 1976; White, 1980; Wilkes, 1988)，意識モデルへの評価は立場によって大きく変わってしまう。つまり概念規定が問題なのであるが，この「意識」の概念規定は，「状態」と「観測」の関係を定める理論が開発されない限り，心理学の枠内での解決は難しく、哲学的論争をどこかで含んでしまうことになると思われる(妻藤，印刷中，参照：もちろん「観測問題」自体が哲学上の難点を含むことになるかもしれないが，現時点では議論できない)。

本稿では，「機能主義的」認知理論の枠内での，「意識」の意味づけをもう少し分析するために、その機能的存在理由を与える（かもしれない）議論について検討を加え，その過程でJohnson-Lairdタイプのメンタルモデルについて，「自己像」という「機能的存在」の役割を考察する。

「意識」の前提としての「自己」

Oatley（1988）は，「意識」の機能的存在理由を論証しようとして，4つのタイプ分けを行っている。まずヘルムホルツ型の「意識」とは，無意識的推論による感覚情報処理の結果を指す。バージニアウルフ型では，心像としての「意識」，ヴィゴツキイ型は心的操作の自発的方向づけ，そして，ミード型は「社会的関係」の内在化である。これら4つに共通しているのは，イメージあるいはメタファーの中で，世界のシミュレーションを行う能力である（このあたりは，引用されていないが，Yates,1985の理論が前提とする作業仮説とほぼ同じである)。また後の2つは，明確に「自己」概念を内包している。さらに，ヘルムホルツの知覚やウルフのイメージには，明確に「自己」が登場するわけではないが，それでも暗黙のうちにその存在が認められる。つまり，単に「見ている」だけではなく，「自分がそれを見ている」ということを知ることができ，また，それによって「注意」を方向づけることもできるからである。結局，彼の主眼は次のようになる。「知っている（見ているなど)」を，少なくとも潜在的に持ち得るということが「意識」の条件であり，またそのために，「自己」が前提として必要とされるということである。そして，そのようなシステムは「意図」という概念で捉える方が適切であるような（そのためには社会的関係も問題になる）ことがらと，活動を結び付けることができなくてはならない。(彼は，有望な一つの例として，SussmanによるプログラムHACKERを挙げている。これは「積木の世界」での問題解決を行うとき，障害が発生すると自分自身の方略やゴール設定を「反省」して，自分自身を書替えることができる。)

しかし，このような立論に対して，おそらく異なる「立場」からの批判はまぬがれない。彼が述べているのは，「自己」の機能であり，そこから「意識」の現

象体験を導き出すことは困難である（妻藤，印刷中，参照)。もし仮にこのような機能全てを実現している人工知能が完成したとしても，人間の「心」と同等ではないと主張する論者は必ず現れるであろう。意識体験の「必要条件」を述べているだけであって，結局のところチューリングテストへの反論と同じものを適用することができるからである。しかも，この議論は「意識」を前提として，それが存在するためには「自己」が必要だというものである。Wilkes（1988）のように，「意識（consciousness)」がデカルト以来の哲学的虚構だとする論者は，論法自体を認めないかもしれない。

「意識」と「自己」との関係については，メタ認知の研究に当然関わりがある。この分野では，自己効力感と能力や動機づけとの関連について，大量のデータが蓄積されている（e.g.,Brown,1978)。言い替えると，「自己像」の重要性はすでに確立されているといってよいであろう。しかし，だからといって，「それ」が何であるのかが判明したわけではない。「意識」をまず「自己」の問題から捉えようとしても，また類似の問題に踏み込んでしまう可能性がある。例えば，「意識」ではなく，注意・メタ認知・自己組織化などの説明概念で十分であって，「意識」を（機能的）実在と考える必要はないのかもしれない（妻藤，印刷中，参照)。「自己」についても，「行動主義的」な議論も可能なのである。つまり，人間は実際には全く（刺激に反応する，あるいは社会的文脈に応答するだけの）「機械的存在」なのだが，自分は「自発的に」行動しているという（虚構の）情報が，システムの安定性を支えている，という「可能性」すら考え得る（それでも「自己像」は重要なのである)。

ここでの論点は，このような設問の仕方だけでよいのかどうかということである。つまり，問題設定の仕方によっては，その設問自体が無意味になってしまう危険が大きい。例えば，パソコン上でエディタープログラムを走らせて原稿を書いているとき，その人が使っている機械がコンピュータなのかワープロなのかを考えてみよう。コンピュータとワープロは，しばしば別の機械だと扱われる。しかしワープロというのは実は，機械としてはコンピュータと同じものだということを聞きかじった人が店頭で，「原稿を書くのに使えるコンピュータをください」といえば，ワープロソフト（あるいはエディータソフト）をコンピュータと一緒に売りつけられるであろう。そして，システムソフトその他のインストールなどという言葉を説明書に見いだしてお手上げとなり，自分が買ったのが「ワープロ」ではないと判断することになる。実用上この区別は，ワープロ専用機を一般的なコンピュータと区別するときに用いられており，その限りでは有用な用語である。しかし，それを「本質的」相違ととらえると，「結論のない問題」を提出してしまうことになる。つまり，原稿を書くのに使っているパソコンはワープロであるのかないのか，と問うても，観点によるとしか答えようがない。一見馬鹿げた例のようであるにもかかわらず，この種の問題はあらゆるところで発生してきた。素粒子は波動であるのか，粒子であるのかという設問では，この選択肢の中には答えがない，ということを知るために大量の実験と理論的研究が必要であった。またクダクラゲのようなある種の群体について，それが独立した生物の集まりのなか，全体として一つの個体なのか，という大論争も，結局どちらでもなく群体としかいいようがないと結論された（Gould, 1985)。これも設問が無意味であったという有名なケースである。これらの「結論」を導くのに必要だった研究は「議論」ではなく，データの収集と「具体的な」個別理論だったということが重要である。今問題にしているのは，無意味な設問をしないことではなく（それは避けられない)，それが無意味であるということを知るための手段があるかどうかなのである（無意味であるという結論は貴重なものであり，またそこにたどり着くまでの研究こそ，たいていの場合，新しい理論のパラダイムを作り上げる)。

おそらく，Oatley（1988）のような方法での「議論」は，何を探すべきかという，かなりメタレベルの方針を探る手段として価値があるとしても，具体的な個別理論を導くには抽象的すぎるように思われる。も

しこれが，正しい議論だが意味がない，という場合でも，このような方法ではそれを判定することができないかもしれない（つまり議論だけに終り，反対者はそのままで残ってしまうかもしれない）。また「意識」の前提として「自己」の「機能的」必然性を論じている場合，意識自体についての錯綜した議論が問題をさらに複雑にしてしまうかもしれない。そこで，この議論をもう少し具体的に扱うために，すでに提案されている個別理論の応用的拡張において，「自己」ないしそれに関連する構成概念が理論的に要請されることを論証するべきであろう。データの解釈では，少なくとも「自己像」が重要であることがわかっているにも拘らず，（それを受け入れた上での解釈理論はあっても）それがメカニズムを正常に働かせるために果たしている役割の分析や意味づけは行われていない。以下では試論として，これらの問題を異なる角度から考察し，さらに（「自己像」を持つことができると想定されている）メンタルモデル論が人格理論や社会的関係を制約する基礎理論になり得る可能性があることを論ずる。

メンタルモデルにおける「自己像」

メンタルモデルが提案されて以来（Johnson-Laird, 1983），数多くのバリエーションが発表され，もともと「古典論」（この用語については，妻藤，1990参照）としての役割をもっていたのに並列分散処理による基礎づけの試みまで現れている（Johnson-Laird,1988a)。ここでは創始者であるJohnson-Lairdのものをとりあげる。このモデルは，人間が論理的推論を行うときに，いわゆる論理規則の集合を用いるのではなく，前提命題を（抽象的表現と具体的表現の中間にあたるような）

S＝h
S＝h
S＝h

図1　メンタルモデルの例：Sは心理学者，hは変人，イコールは「である」を示す。

モデルの形で表現し，これによって結論を導くというものである。例えば，「全ての心理学者は変人である」という命題は図1のように表される。ここで項の数は絶対的なものではなく，最初は3個程度が設定され，必要に応じて増減される（ここで推論エラーが生ずるかも知れない)。そして次に，「ある変人は哲学者である」が与えられると，対応するトークンの関係がこのモデルの中に挿入される。この時，論理関係を完全に表すには，モデルが一つでは足りなくなる。つまり図2のようになる。厳密にはこれでも不足であるが，こ

S＝h＝T		S＝h
S＝h＝T	OR	S＝h
S＝h		S＝h
		h＝T
		T

図2　複数個必要なメンタルモデルの例：Sは心理学者，Tは哲学者，hは変人，イコールは「である」を示す。

の場合一応この2つがあれば妥当な推論を導くことができる（心理学者の中に哲学者を兼ねている者がいるとも，いないともいえない)。しばしば（訓練を受けた人でも），2個以上のモデルがなければ妥当な結論が出せないような場合に，（作業記憶容量の制約や訓練不足のため）すべてのモデルを展開しないで結論を出してしまう。推論エラーの多くはこのために起こる。実際には，この理論はもっと複雑であり，（関係演算を含む）一階述語論理について，人間の推論エラーと反応時間を予測でき（Johnson-Laird, Byrne, & Tabosi, 1989），またこのモデル操作のオペレーティングシステムと言語解釈についての仮定を加えて，かなり幅広いデータを説明できる。このオペレーティングシステムは，並列処理に伴うコントロールの問題を解決するために導入された。つまり，各サブシステムがそれぞれの作業を並列で進めていくと，どこかで衝突が起こったり必要なときにデータが手に入らないサブシステムができてしまう。そのために，調整をとる

メタルレベルのシステム（オペレーティングシステム）が必要になる（これが「意識」を司るとされる）。ただし，すべてのサブシステムの詳細を監視するのは無駄であると同時に不可能であり，目的や方針のレベルでの監視や調整を図るほうがよい。このため，「意識」できる内容は，階層構造化されている並列システムの上層部のみである（このように「意識」をとらえることの問題点については妻藤，印刷中，参照）。

Johnson-Laird（1988b）は，意識論との関連で，このメンタルモデルが「自己意識」を持ち得ると述べている。ただしその仮説は，認知系が自分自身の状態を解析するために再帰的に全体を埋め込むというものである。比喩的に述べると，自分自身の縮少コピーをメンタルモデルの中に加えるということになる。このようにすると「自分自身の状態の縮少コピーを含んだメンタルモデル」を持っている自分が出来上ることになる。すると，そのような「自分」の縮小コピーをつくることもできるために，『「自分自身の状態の縮小コピーを含んだメンタルモデル」を持っている自分の縮小コピー』を含んだメンタルモデルも出来ることになる。このようにして，無限の埋め込みが起こる可能性がある（彼自身これを認めており，むしろ積極的に，このような一種のフラクタル様の全体が「自由意志」に関係するとみている）。もう一度言い替えると，≪『「悩んでいるわたし」をわたしは嫌っている』そういうわたしが情けない≫と思っているわたしがつまらなく思え……といった構図になる。

しかし，この仮説の論拠として，これだけは，強いものとは思われない。すでにオペレーティングシステムがサブシステムの状態を認識できるとされており，むしろオペレーティングシステムが「自己像」をメンタルモデルの中に埋め込むのであるから，これだけでは「何のために」という疑問が残ってしまう。さらに，（実際には有限であるとしても）何度でも実効可能な再帰的自己言及を行ったとき何が起きるのか明確でない。現時点では（メンタルモデル理論の基本部分とは異なり）このようなシステムのシミュレーションも困難である。

「人間関係」の性質から要請される自分自身のシミュレーション

Johnson-Laird（1988）とOatley（1988）の着想を合成することによって最初の疑問に答えることができるであろう。後者はミードの考察に基づいて社会的関係の内在化が「自己意識」の成立に関係があるとみている。そこで社会的行動に必要な情報処理のタイプを考えてみる。

まず個人の社会的場面での振舞いが，その時の社会的「状況（ここでは特定の社会的場面）」と自分自身の性格傾向との働きあいで決るのだとしてみる。この仮定は個人が有限オートマトン型のシステムであるものだとすることに等しい。ただし，これは認知システムがオートマトンだということではなく，行動を決定する部分が認知システム（おそらく万能チューリングマシン）の出力に応じてオートマトン的に振舞うということである。このような場合，「意味的」に一貫性のある社会的行動をおこなうことは，相当困難であると思われる。ある社会的「状況」を一対一で表す情報があるものとして（実際にはそんなことは不可能であるが，それが可能だとしても）それを入力として内部状態が変化して出力を決定するのであれば，内部状態から出力への遷移確率は行動連鎖の相当先の方までの連を考慮した条件つき確率の形になる。しかし，問題は連続性をもつ社会的場面での適切な（少なくとも意味的に見当はずれでない）行動は，その時々の他人の行動との関係で決る。また，後で述べる並列分散処理とは異なり，相互作用が興奮と抑制という単純な信号で決るのではなく，「状況」を記述するには（少なくとも）2次元以上の空間が必要になる。これをm次元とすると，n人の相互作用はm×n次元の仮想空間上の点の運動で表すことができる。もしここで仮定したように，「状況」の入力によって，各人の行動が（学習的に）変化していくものとすると，ここでの「適切」な行動は，一時点だけで決ってしまう可能性は少ない。相手が変化していくのに応じてお互いが，同時並列的に変化していくため（含まれる変化単位つまり各個人の性質の集まり方によっては），相互作用を記述する

点が,(m×n次元空間上で)振動的変化を繰り返すかもしれない。あるいは,場合によっては解が定まらずランダム様のパターンになることもあり得る。適当な不動点に落ち込んで安定することもかなりあるかもしれないが,何らかのリミットサイクルになるかもしれない。

もちろん,このような「社会」システムに適当な制約条件を加えることで,ある程度定まった相互作用パターンを作り出すようにできるかも知れない(必ず不動点やリミットサイクルを持つように,すべての個人が定まった性質を持っているなど)。しかし,ここで問題になるのは,たいていの場合,ある社会的相互作用を観察している人がある程度の正確さで成行きを了解的に予測できることである(了解が何を意味するのかは,また別の問題ではあるが)。しかも,それは必ずしもいつも同じ結果になるのではなく,相当程度の可塑性を持っている(必ず有限の領域内での不動点などに落ち着くというものでもなく,喧嘩別れという形でシステムが崩壊することもあり,それでも「観測者」は適当な時点からであれば,ある程度予想することができる)。環境あるいは自分自身以外の状況が固定されているか,そうでない場合,何らかの(初期条件だけから予測可能であるような紋切り型の行動などの)固い法測性を持っている場合には,上記のようなシステムでも意味的に一貫性のある相互作用を示すように設計できる。しかし,そうでない場合について,このシステムが我々の社会的相互作用に関する直感と(ほとんど常に)一致するような振舞いを示すためには,個人間の相互作用の仕方そのものについても相当の制約条件を持たせるような設計が必要になるであろう。

例えば,並列分散処理の認知理論では,認知システムの中の各処理単位が複数あって,上記のようなやり方でパターンを生成するのであるが,この場合,必ず興奮型と抑制型の相互作用が,ある目的との関連でパターン化されている。つまり,パーセプトロンやボルツマンマシンなどでは,教師信号のベクトルに基づいて,出力パターンから後向きに相互結合の強さを変更していく(バックプロパゲーション)ことによって,入力と出力の関係を目的に適合させている。つまり,(システム)全体の外側に(それ自身はそのシステムから影響を受けない)絶対者があって,これがシステム全体の構造に「意味」を与えている。この点が最大の違いである。「競合による学習」では教師信号はないけれども(Rumelhart & Zipser, 1986),学習結果はどれかのユニットが一つ勝ち残るというかたちで意味を持つものである。もしこの各ユニットが社会心理学の個人に対応するものとして,社会的相互作用のモデルとして使えるとしても,ごく狭い状況設定に限られるであろう(いずれにしても,このユニットでは「個人」のモデルとしては単純すぎる)。ライフゲームも,相互作用によって全体的パターンを形成するものであるが,隣り合う要素同士の関係で決る生成と消滅に基づくので,最終的全体が最初の社会的状況の「意味」に照らして適切といえるパターンに落ち着くかどうかは判らない(パターンといえるようなものにならないこともある)。

脳の神経ネットワークの活動パターンが(先に述べたような仮想空間で)カオス軌道をとる場合,そのストレンジアトラクターをうまく利用した自己組織化が可能だとする理論的研究がある(e.g.,津田, 1990)。これとの類推からいえば,オートマトンとしての個人が,脳のカオス制御と類似の相互作用を示すことで,「社会的」関係を作れないとはいえない。しかし,このような脳内での「制御」と個人間の相互作用を同列に見るのは危険である。ここまでの論点は社会的なレベルで人間を見るとき,そこには「了解可能」な構造があり,しかもその時間的発展を仮想空間で表すとすれば,意味的に了解可能な形で軌道を追跡できるということである。この「了解可能性」と理論の関係について,物理学にも類似のものがある。例えば量子力学などのようにわれわれの直感とは遠く隔たった内容を持っている理論がある。しかし,そのような理論から予測される量を意味づけるのは,我々の感覚レベルとほぼ同じであるような古典理論である〔例えばある演算子に関係する平均値が我々の(知覚可能な)世界での(平均値ではない)ある量に対応しているなど〕。同様に,心理学の理論においても,内的メカニズムが

並列分散処理であったり，またその特殊な場合としてのストレンジアトラクターによる制御であったとしても，それらの働きあいから出来上がる心理世界（日常言語が対応するレベル）では我々の「了解」を越えるようなものであるとは考え難い。

また，コンピュータでストレンジアトラクタをプロットしていくときの時系列にそった振舞いは一見ランダムであり，ある程度の時間が経過したときに「パターン」が「見える」ようになる。脳がこれを利用しているとしても，その時間経過にそって出力されるのではなく「結果」が利用されているのである。一方、社会的相互作用の場合，一見ランダムに見えるような関わり合いが，しだいに形をなくしてくるということはまずない。「古典」レベルでの理論を考える場合，むしろ理論の構成要素が意味的に了解可能であるようなものになるはずである（ただし，物理学でもマクロレベルの現象に量子効果が重要な要因になることはあるように，心理学の場合にも，例えば互いに見知らぬ者が集まった群衆などについては，単に統計的振舞いではなく，カオス的振舞いが次第にストレンジアトラクターに焦まるということは考えられる）。そこで以下では「古典論」レベルでの理論を考察する。

オートマトンを要素とする「社会」システムの問題点は，複数の個人が互いに相手の行動を予測できないことから発生している。つまり，（社会的）環境自体が不確定であるために，上記のような局所的学習を行うと，自分だけではなく環境自体が変化してしまうために，最終的にどのようなパターンに落ち着くのかあらかじめ予想できない。しかもそのことが，自分自身の行動の予測すら困難にしている。各個人は，自分自身の内部状態（性格と社会的応答手続き）に，当面の「状況」が入力されて出力が決定される。このとき，応答手続きのどれを実行するかは，他者の行動に応じて変化していくため，他者との相互作用がある程度進んだとき，自分自身がどのような応答パターンでその（特定の）他者に応ずるのかあらかじめ知ることができないのである。

そこで相互作用空間の領域を限定して自由度を減らすのに，おそらく最も簡単なのは，各個人が「関係」に加わっている人の全てをトークンとして含むメンタルモデルを持つことである（単に万能チューリングマシンであればよいというのではなく，特定の表象とその操作プログラムが必要である）。ただし，各トークンは単なる記号ではなく構造を持っていなければならない。つまり，各々の行動傾向を，場面毎に予測できるような情報を持ったものである。この予測は完全なものである必要はなく（もともと不可能である），領域を限定するだけでよい。つまり，相手の出方の可能性をある程度限定するだけで，相互作用全体としては自由度を大きく失うからである。もし，メンタルモデルの操作が，（不十分ではあっても）正確な情報に基づいて行われ，かつ（「目的」にとって）「合理的」であるという理想化を行うと，メンタルモデルによる行動のコントロールは，以下のような結果をもつことになる。まず，相手の出方についてある程度の予測ができれば，それによって自分の（不利にならない，あるいは目的にかなう）行動の幅も限定されることになる。すると，自分のとるであろう行動の集合が限定されると，それに応じて相手の行うであろう行動も可能性の幅が狭くなる。これを繰り返すことによって（対人関係のシミュレーション），自分のとるべき行動あるいは実効可能な行動のレパートリーは，相当限定されるであろう。ただし，実際にはそう簡単にはいかない。相手についての認知がゆがんでいることは大いにあり得る。しかし重要なのは，それが誤っていても，それに基づいて行動を決めていくとすれば，それがその人の（特定の人に対する）行動傾向になるため，相手はそのことを自分のメンタルモデルに組み込むことができる。その他に，「目的」が自分にとっても明確に認識されていないとか，「目的」自体が複合的であるために，シミュレーションを単純化しないと処理能力を越えてしまうかもしれないなどの要因はある。しかし原理的にはメンタルモデルを導入することで，先に挙げた問題を解決することができる。

ここで重要なのは，相互作用は登場人物全ての働きあいであるから，このようなメンタルモデルの中に，

自分自身もトークンとして含まれていなければならないということである。そして先の理想化を多少緩めて「合理的」に行動の方略を決定するとは限らないということにすれば，シミュレーションを行うために，自分自身の感情的反応も考慮に入れなければならない。その方が自分の利益になるということが（仮に）推測できていても，そのとき自分が強い自己嫌悪を持つ可能性があるとすれば，さらに他の可能性も検討するであろう〔この方が「生存的に合理的」だという議論もあるが（水島，1990参照），ここではとにかく「目的」を達成するために最小の努力で実行できる方略を選ぶこと，という意味で用いている〕。すでに述べたように，「論理的推論」を行うとき，人間は必ずしも論理的に必要な全てのモデルを展開してみるとは限らないということが確認されている。人間関係についての推論では，はるかに多くの可能性（つまり異なるモデル）があるために，すべてのモデルを展開するようなシミュレーションは不可能であろう。特に，一階述語論理についての実験で，（本当は）2つのモデルが必要な問題でも，一個しか展開しないためにエラーをおかす被験者が少なからぬ割合でいることが見いだされている（Johnson-Laird, Byrne, & Tobossi, 1989）。もちろん，このような実験では厳密な解があることを知っているため，個々のモデルを厳密なものにしようとするため，各モデルに必要とされる処理容量が大きく，そのために2個でもオーバーフローになるということかもしれない。しかし，それほどの厳密さを必要としない（あるいはもともと不可能である）社会的場面のモデルでも，一度に展開できるモデルの数はさほど多くないと考えてもよいであろう。そうだとするなら，この感情的要因に基づいて展開するモデルの数を極端に減らしてしまう可能性が強いといってよい。もし感情的に問題のあることが予想されると，そのモデルの展開を打ち切り，他のモデルを探すかもしれない（もちろんこれは個人と場面によって異なる）。

したがって，このようなメンタルモデルは自分自身の，社会的行動のレパートリーだけでなく感情的な面も含んだ「自己像」を持っていなければならない。これで，（少なくとも人間が）社会的行動を行うためにはメンタルモデルの中に「自己像」を組み込まねばならないということをて示せたと思われる。（ここではとくに取り上げないが，動物の「社会的行動」については，その「程度」を考慮する必要があるだろう。前述のように，「環境」の各要因が，それ自体強い法則性を持っているときは，つまりある「状況」への応答が紋切り型に近ければ，オートマトン的振舞いでもかまわない。社会的相互作用のレベルに応じて，「自己」のメンタルモデルが決ることになるだろう。）

再帰的埋め込み

次に発生する問題は，再帰的埋め込みの程度である。Johnson-Lairdの問題設定から導かれるのは，「自己モデル」の再帰的埋め込みのみであったが，ここでの問題設定から明らかなように，それだけではなく，「他者」のモデルについても，それを考慮しなければならない。「わたし」の（考え方，認知の癖，そして感情傾向も含めた意味での）行動傾向を相手が（ある程度）知っているということも当然，人間関係シミュレーションの重要な要因だからである。すると，『そのことを「わたし」が知っている』ということを相手が知っている，ということがまた行動の傾向を変化させるかもしれない。そして，同様に自分自身についても，ある場面での行動に関連して発生する感情がどのようなものになるか，ということを自分で予測することにより，それも考慮に入れた（自分自身についての）シミュレーションの結果が異なってくるかもしれない。

しかし，ここではJohnson-Lairdとは異なり，そのような再帰的埋め込みはさほど重要ではないという立場をとりたい。まず他者についてのモデルの場合，人間に可能な他者の理解には決定的に限界があると仮定すれば（あるいは少なくとも通常さほどの理解には達しないなら），「わたし」が知っているということを相手が知っている，というレベルを越えて埋め込みを行っても，その時の他者の行動の変化をどれだけ予測できるか疑問である。そのような意味で，おそらく一段程度の埋め込みしか起こらないであろう（彼は私のこと

を良く知っているから，このようなときに，ひどい言葉を投げつけたりはしないだろう，など)。ただし，共感的理解（それが何であるかはさしあたり別として）が優れている人が，さらに一段深い埋め込を行うことはあるかもしれない（彼は私がそう考えるということも判っているだろうから，わざとイヤミを言うかもしれない)。いずれにしても，無限の埋め込みが原理的には可能であるとしても，処理容量と能力レベルとの関連で有限にしかなりえないであろう。

そして，同じことが自分自身の再帰的重層モデルについてもいえる。「自己像」の構造が多層化するほど，その時の自分の感情的状態などを予測することは難しくなるであろう。というより，実際，意識的にそれを試みれば直感的には納得できる：私がそんなことを言ったら，自己嫌悪に陥るだろう，ということを私は知っているが，それを知っていることによって，意識的に感情を押えようとすれば，私の自己嫌悪の程度がある程度減るかもしれない，ということを私は知っているが，それを知っていることによって，私の自己嫌悪の程度が変化するだろうか？このような，ある程度の深さの多層化を意識的に行った後で（それを記憶にとどめ)，実際にその（自己嫌悪を起こすであろう）言動を行ってみたときにはたしてどうなるであろうか。そんなことをやっても別に変化はないという人が，おそらく多いと思われる。また，変化があると考える人でも（そのようなことを考えてからであれば，相当程度感情を自己制御できると考える人でも)，４重５重以上の再帰的予測となると，意味がないと答えるであろう。そして，実際にその場面になったとき，いずれにしても，そのような「確信」は，無意識のダイナミクスによって裏切られるかもしれない（意外なほど感情がわかなかった。あるいは隠しておけないほど取り乱してしまったなど)。

このように，少くとも，意識レベルでは多数回の埋め込みは無理であろうと思われる。メンタルモデル自体はオペレーティングシステムによって意識化されない限り無意識に留まるのだとしても，もし多数回の埋め込みが意味をもつほどの自己洞察が（無意識的に）可能であるのなら，（少なくとも）その洞察結果を意識的に認識できないのはなぜなのか説明を要する。現時点では（特に抑圧が起こりそうな感情的問題に関係していない場合には)，これを説明する理論はみあたらない。むしろ，そのような多重再帰表現は，さしあたり問題とする必要はないと思われる。またこのような考察から，「自由意志」の問題は，無限の再帰的自己言及によって生ずるような一種の状況独立性とは別に考える必要があるということになる（このテーマはこれ以上ここでは扱わない)。

結論と要約

「意識」理論の可能性を探るために，「自己」の問題を無視できないという議論について論評がなされ，そのようにして認知理論に「自己」を持ち込むのは，かえって混乱を起こす可能性があることが指摘された。そして「自己」を導入するためには，むしろ逆に「自己」（少なくとも「自己像」）について，システムが適性に働くにはそれが必要であることを論じるべきであると結論された。

さらに，複雑な人間の社会的行動についての考察から，人間が「社会的状況」に応じて柔軟に対応しながら，かつ他者との相互作用の仕方を局所的に学習していくという仮説では無理があるということが論じられた。「社会的人間」がそのようなものであるとすると，個人間の相互作用パターンが（普通想像されているのとは異なり)，関連する人の集団の構成がわずかに変わっただけで（つまり新しい人がその集団に入ってきたり，あるいは誰かがいなくなったりすると）全く違うものになってしまう可能性が強い。それどころか，安定不動点からリミットサイクルのようなものに突然変化したり，ランダムに近い（というよりカオスのような）パターンになったりするかもしれない。並列分散処理のように，その集団から影響を全く受けないコントロール信号（教師信号）があったり，またその集団の個々のメンバーの相互作用の仕方をコントロール信号との関連でバランスをくずさないで調整するバックプロパゲーションのような学習規則があれば話は別

である。しかしそのような仮定は，少くとも一次集団では考えられない。

このため社会行動の理論においても，（論理的推論などと同様に）メンタルモデルを導入する必要があり，しかも，そのメンタルモデルは「自己像」を再帰的に，ただし無限の埋め込みではなく一重か二重程度に含んでいる必要があると結論された（哲学や現象学的議論ではなく，システム上の必要性が論じられた）。おそらく自己像の崩壊が危機的状態を引き起こすことがあるのは，この仮説から導くことが出来る。ノイローゼの治療が終結したその日に自殺する危険がある。というのは臨床心理学の基礎的知識である。結局，問題を抱えていた自分が変化したとき，新しい自分についてのメンタルモデルがまだ完全には成立していないために，自分の（感情も含んで）行動の予測が困難になること，またそれ以上に，自分の内的世界（メンタルモデル）が違うものに変っていくために，（もし「意識」がこれに関係するとしたなら），意識的に認識できる拠り所を失うことに関係があるのかもしれない。もしこのように考えてよいとするなら，「自己」は「自己像」を介して認識される「虚像」であるかもしれない。ただし，それがないと後者も意味を持つことができないという点で，「機能的」実在であろう。

文献表

Brown,A.L.（1978）Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser（Ed.), *Advances in instructional psychology Vol. 1.* 和訳「メタ認知－認知についての知識」，湯川良三 & 石田祐久 共訳 サイエンス社 1984.

Dennett,D.C.（1988）Quinting qualia. In A.J.Marcel & E. Bisiach(Eds), *Consciousness in contemporary Science*（*pp. 42-77*）. Oxford: Oxford University Press.

Eccles, J.C.(1976) Brain and free will. In G. G. Globus, G. Maxwell, & I. Savodnik (Eds), *Consciousnesrs and the brain* (*pp.101-122*). New York; Plenum Press.

Gould, S. J.（1985）*The flamingo's smile.* W. W. Norton & Company, Inc. 和訳 「フラミンゴの微笑－進化論の現在」 新妻昭夫訳 早川書房 1989.

Johnson-Laird,P.N.(1983) *Mental models.* Cambridge: Cambridge University Press. 和訳「メンタルモデル：言語・推論・意識の認知科学」海保博之監修 AIUEO訳 産業図書 1988,

Johnson-Laird,P.N.（1988a）*The computer and the mind:An introduction to cognitive science.* London: William Collins Cons & Co., Ltd. 和訳「心のシミュレーション」 海保博之，中溝幸夫，横山詔一，& 守一雄 訳 新曜社 1989.

Johnson-Laird,P.N.（1988b）A computational analysis of consciousness. In A.J. Marcel & E. Bisiach (Eds), *Consciousness in contemporary science*（*pp. 357-368*）. Oxford: Oxford University Press.

Johnson-Laird,P.N., Byrne, R.M.J., & Tabbosi, P.（1989） Reasoning by model: The case of multiple quantification. *Psychological Review,* **96**, 658-673.

Marcel,A.J.（1988）Phenomenal exprience and functionalim. In A.J. Marcel & E. Bisiach（Eds）, *Consciousness in contemporary science*（*pp. 121-158*）. Oxford: Oxford University Press.

水島賢太郎（1990）生存的合理性と論理的合理性の概念－合理性の認知科学 神戸女子短期大学学会「論攷」,**35**, 101-137.

Oatley,K.（1988）On changing one's mind: a possible function of consciousness. In A.J.Marcel & E.Bisiach (Eds), *Consciousness in contemporary science*（*pp. 369-390*）. Oxford: Oxford University Press.

Rumelhart,D.E., & Zipser,D.（1986） Features discovery by competitive learning. In D.E. Rumelhart, J.L.McClelland,& the PDP research group（Eds）, *Parallel distributed processing: Explorations in the microstructures of cognition, Vol. 1 : Foundations* (*pp.151-193*). Cambridge: The MIT Press.

妻藤真彦（1990）認知心理学における３つの理論的立場 美作女子大学・同短大部紀要,**35**,9-20.

妻藤真彦（印刷中）認知心理学における「意識」と「測定」. 大阪市立大学文学部心理学教室40周年記念出版（仮題）.

Savage,C.W.（1976） An old ghost in a new body. In G.G. Globus, G. Maxwell, & I. Savodnik （Eds）, *Consciousness and the brain* (*pp. 125-154*) . New York: Plenum Press.

Sperry, R.W.（1976） Mental phenomena as causal determinants in brain function. In G.G. Globus, G. Maxwell, & I. Savodnik （Eds）, *Consciousness and the brain* (*pp. 163-178*). New York Plenum Press.

津田一郎（1990）カオス的脳観－脳の新しいモデルをめざして. サイエンス社

Weimer, W.B.（1976） Manifestations of mind: Some conceptual and empirical issues. In G.G. Globus, G. Maxwell, & I.Savodnik （Eds）, *Consciousness and the brain* (*pp.5-32*) . New York: Plenum Press.

White,P.A.（1980） Limitations on verbal reports of internal events:A refutation of Nisbett and Wilson and of Bem. *Psychological Review,* **87**, 105-112.

Wilkes, K.（1988） --, yishi, duh, um, and consciousness. In A.T.Marcel & Bisiach(Eds), *Consciousness in contemporary science* (*pp.16-41*). Oxford: Oxford University Press.

Yates, J.（1985） The content of awareness is a model of the world. *Psychological Review,* **92**, 249-284.

（1990年12月1日受理）