環境との相互作用に基づく
自律適応系の構成論的研究

谷口 忠大

2006

目次

図目次

表目次

第Ⅰ部

序論

第 1 章

はじめに

「自律適応的に言葉を獲得し，人間と意思疎通出来るようになる人工物を作ることは可能だろうか？」

本論文の目的は上の単純な問いかけに答えていくことにある．

近年，情報処理や機械技術の発展に伴って，半世紀前には夢でしかなかったような自律ロボットを作ることが徐々に可能になってきている．HONDA の ASIMO は大衆の前で歩き，ボールを蹴って見せた．SONY の AIBO は家庭にやってきて，ひとりでに歩き，主人を見つけると手をふったり尻尾をふったりしてみせる．これはその機械技術的側面においては著しい進歩といえる．

しかし，最近よく目に付くのは，家庭で電池切れのまま放置された AIBO であり，画面の向こうで手を振るだけで町には決して出てこない ASIMO である．これは何を意味しているのだろうか？

数年前，AIBO が出始めたときは AIBO をペット代わりに可愛がる人が多数現れ，社会現象とまで言われた．しかし，そのブームは次第に後退し最近ではそのような話はあまり聞かない．周りを見渡しても，ペットロボットにはまっている人は少ない．私自身もそうであるが，実際にペットロボットと遊んだ経験のある人に尋ねると多くの人がこう言う．

「飽きたんです．」

犬を大事に飼う人の口からはそうそう聞かない言葉である．我々はこの言葉の意味するところを，次のように捉えている．多くのペットロボットのような社会ロボットと呼ばれるロボットは，現在相互作用のための法則を様々に埋め込まれて刺激反応的

に振舞う．故に，ユーザはその内の大半のパターンを経験し終えたらそれ以上の相互作用の喜びが湧かない，つまり飽きるのだ．それに対して，実際の犬を飼う多くの人は，犬の持つ行動の多様性に飽きない．ここで言う多様性とは，決して多くの行動のレパトリーをもっているということに終始しない．それは，人間との相互作用を通すなかでより豊かになっていく適応的な多様性である．そして，その豊かな相互作用が二人，もしくは一人一匹の関係性を徐々に構築していく基盤となるのである．

「うちの○○（犬の名前）は私の言ってる言葉をわかるよ.」

と，多くの犬を大事に飼う人が口にする．犬は人間の言葉が話せるわけではないし，言葉を理解しているかについての科学的な知見は多くない．しかし，犬はパブロフの犬で有名な条件付けを行う能力など高度な関係生成能力を持っていることは知られているし，現在の機械システムが持ち得ない多自由度なモータ系とセンサ系を有している．また，哺乳類という進化論的にはかなり高度な種に属することから，自律適応システムとしては人間に近い高度な適応機能を有していることが考えられる．

ならば，人間と同じレベルの言語は持たなくても，言語の原始的なもの，つまり，広い意味での記号の解釈・設計能力はあるのではないかと考えられる．もし，そのような能力があるならば，上のような飼い主の言葉を導いても不思議ではない．自然言語ほど統語論的にも豊富な特性をもつ高度なものでなくとも，記号を相互作用の中から見出し，その意味を飼い主と共有できるようになれば，適応的な記号的相互作用が可能になり，飽きないペットロボットが生まれる可能性があるだろう．

本論文では以上のような文脈から，自律系としての行動主体が環境や人との相互作用を通して自らの内部に内的表象を組織化し，また，それを外部の聴覚・視覚信号と連合させることにより，他者と相互作用するための記号を獲得するプロセスの構成論的モデルを構築する．

本研究にあたっての一つの動機は上のような適応的な社会ロボットの設計論の必要性に負っている．しかし，このような記号の自律系内部での適応的組織化や，記号の意味の相互作用を通した形成という問題は，ロボットの運動学習や，人間機械系，記号論，人工知能研究，生命研究等における重要な問題と密接に関連している．冒頭に掲げた問題意識は，自律適応系の構成論的研究を謳う上では，あまりに具体的過ぎる問題と思われるかもしれないが，この誰でも思いつく身近なテーマが自律適応系研究についての非常に良い例題となっていることに読者は気付いていくことであろう．

まず，第 I 部の後半となる次章では幾つかの角度からこの問題の研究背景について議論し，第 II 部以降の導入とする．第 II 部では環境や相互作用する対象の概念が自

律ロボット内部で組織化されていく機構について論じる．具体的には運動球の追跡問題をとり上げる．第３章では環境の相転移的な変化に自発的に気づき，シェマとよばれる内的表象を分化させることにより，学習を行いながら概念生成を行う双シェマモデル (Dual-Schemata model) の提案を行う．第４章ではこのような自律適応系内部での表象生成は，あくまでその自律ロボット自身のセンサ・モータ系に閉じた領域，つまり現象学的な領域での出来事であり，客観的な表象の境界は存在し得ないことを示す．また，このとき身体パラメータと表象の粒度との間に非線形な関係が存在することを構成論的に示す．第 III 部では行為概念の獲得に焦点を当てる．環境や対象の概念が “名詞的” なものであるとするならば，これは “動詞的” な意味を持つ内的表象の生成に焦点を当てたものである．特に非明示的で自律的な学習を実現するため，強化学習を導入する．他の系も対象にするが，主には平板上での運動球の制御問題をとり上げる．第５章では行為概念の獲得において，環境ダイナミクスの変化に依存しない汎化行為概念を双シェマモデル上の強化学習で実現することが出来ることを示す．第６章では強化学習シェマモデル (RLSM: Reinforcement Learning Schemata Model) を導入し，第３章で導入した双シェマモデルにおける環境概念の分化による累増的獲得と同様な行為概念の累増的獲得が可能なことを示す．第 IV 部ではこのような記号生成のプロセスが実際の人間の脳における可塑性と同型的であるかについて考える．構成論的な枠組みでは実際に記号生成にどの部位が関係しているというような，科学的な結論を導出することは出来ないが，計算論的に等しい構造を見つけ出すことが出来れば，本研究で定式化する記号現象と実際の脳内現象との連関についての示唆を得ることが出来る．第７章では小脳のモデルとして提案されている MOSAIC に対して，それと交換可能と考えられる計算論的なモデルとして NGSM(Schema Model with a Normalized Gaussian network) を提案する．これは第３章において提案された双シェマモデルにおける分化プロセスを非線形な系へ拡張したものである．実験を通して，MOSAIC が元々持っていた不安定性や累増性についての制限を NGSM が取り払うことを示す．また，第８章において，近年，大脳新皮質や海馬で発見されたスパイクタイミング依存のシナプス可塑性 (STDP: Spike Timing-Dependent Plasticity) とシェマモデルを結合することにより，自律系内部に獲得された表象と外的な聴覚・視覚情報（遠隔情報）を連合させ適応的に行動できるようになることを示す．第 II 部，第 III 部を通して内的に獲得できるようになった表象が外界の信号と連合することにより，感覚運動系を通して自律的に形成された知覚行為概念が記号として縮約されてコミュニケーションに利用されうる記号となり創

発する．最後に第 V 部において，まとめと展望を記す．

第 2 章

研究背景

第 1 章の冒頭にあげた，問いの答えに近づくためには，その背景と問題設定を整理して，接近のための戦略を練る必要がある．本章はそれらに割くことにする．重要なキーワードは自律適応系，記号，分化 (差異化) の三つである．これらのキーワードで本研究の作業仮説を表現するならば「自律適応系内部における概念分化を通して記号的相互作用の基盤が形成される.」となる．この意味と理由を本章で確認していく．

前章で述べた，人間機械コミュニケーションの適応的構築は既存の人工知能的手法を組み合わせることで解決する問題ではない．その根本的問題は，現在様々な領域で問題視されている記号の適応的構成の問題そのものになっている．例を挙げると人工知能における記号接地問題，哲学における他我問題，情報学における情報の基礎付けなどである．このような問題を扱う上で，機械か生体か人間かといった区別は不要であり，自律適応系としての記憶の組織化を通した適応のダイナミクスが問題となってくる．まず，自律適応系設計がどの様な流れで必要となってきているかについて論じる．

2.1 自律適応系の設計論という要望

近年，社会における自律ロボットの認知度が上がってきている．ペットロボットが家庭で飼われ [2]，テレビ CM や歌手のプロモーションビデオで二足歩行ロボットが現れ [109]，名古屋万博では大きな話題をさらった [168]．この流れは，なんの原因も無い一時的なブームなのだろうか．それは違う．最近のこの傾向は自律ロボットの研究における，「工場環境から人間環境へ」という研究の文脈を背景にしているとい

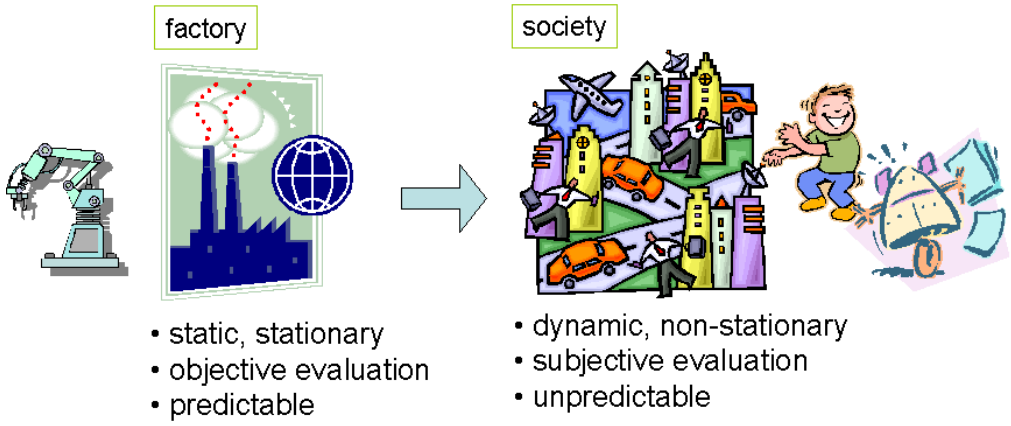


Fig.2.1: From factory environment to social environment

える．

産業用ロボットの技術は 20 世紀後半に飛躍的な進歩を遂げ，日常生活に物質的繁栄をもたらしてきた．その技術は工場の外，つまり人間の生活空間へと活躍の場を求めている．特に，日本ではアニメーションやＳＦの浸透度が高いこともあり，その機械技術を日常生活上の生活支援へと拡大できないかという機運が盛り上がるのも不思議ではない．しかし，その新たなフロンティアへの拡大は自然な拡大として成立するのだろうか．

実際には，そのフロンティアで自律ロボット達は全く新たな問題に遭遇することになる．常時，工場内のロボットの稼動状況を監視する労働者の存在によって常に一定の環境に保たれている工場環境に比べると，人間環境は捉えどころのない複雑さに満ち溢れている．故に，人間社会に進出するロボットのための設計論研究は，工学的に考えて非常にチャレンジングな問題を数多く提供することになる[*1]．工場環境における産業ロボットに対し，人間社会で活動するロボットは広く社会ロボット (social robot) と呼ばれる．前章で例に挙げたペットロボットはその一例に過ぎない．

人間環境へ社会ロボットが進出するとき，その環境の複雑さ・動的変化は工場でのそれと比べ物にならない．人間環境は制御工学的に時変・非線形な系であるのみならず，人間特有の言語的・社会的ルールといった未だ数理的に捉ええないものをも含んでいる．このような人間要素の持つ不確かさは，数理的に扱いがたいのみならず，往々にして客観的情報ではなく，各人内部で固有の主観的なものであり，首尾一貫し

[*1] 問題は工学だけに収まらないことも明らかになるが．

ていることすら望めない.

上で述べた複雑さ，動的さは大きく分けて物理的な複雑さ (時変・予測不可能性・多様性) と，人間由来の社会的な複雑さ (主観性・価値の局所性・多様性) に分けられるかもしれないが，どちらにしろこれらはまず工場環境下で活動していたロボットにとっては未知のものである[*2]．これらをすべて設計者が事前に想定し，計測，解析，モデル化し，制御器・行動アルゴリズムを設計することは，もし出来たとしても膨大な時間も手間もかかる[*3].

もしありとあらゆるユーザの個別性，行動する環境の在りえる可能性を想定して社会ロボットを設計したとしても，それは膨大な人的コストを必要とするであろうし，その設計に関する労働力はペイされないと考えられる．それは個々の製品に多大な研究費を費やす，極端な多種少量生産のようなものになるだろう．このような意味でも，現実問題として，設計者がユーザと社会ロボットとの相互作用を全て想定することは不可能であろう．実際，現在の社会ロボットは，設計者の想定した，いくらかの対人間相互作用のためのロジックを自律ロボットに搭載し売り出すことしか出来ていない．このような社会ロボットに人間環境でタスクを行わせるためには，そのユーザは，まさに工場で産業用ロボットの動作を監視する労働者のように，そのロボットにとって想定外の状況を排除することに努めなければならない．例えば，お掃除ロボットを使うときには，前もってユーザがある程度掃除して障害物が無いようにしておかないとまともな仕事をしないので，今まで余り掃除をしなかった亭主が掃除をするようになったという笑い話がある．つまり，労働を肩代わりしてくれるロボットに労働してもらうための，新たな労働が生まれるのだ．また，前章で扱ったペットロボットなどの場合は，不可能，可能の議論ではなく，その適応性の無さはコミュニケーション能力というエンターテインメントの核部分を失うことになる．後にも論じるが，西垣の基礎情報学によれば，刺激ないし環境変化に応じ，あくまで自分自身の構成に基づいて自己変容を続ける，その変容作用こそ意味作用である [178]．つまり，相互作用のロジックのみを起動し，自己変容を遂げない社会ロボットと人間の間には真のコミュニケーションは成立しない．変容なしに意味のあるコミュニケーションは存在しないのだ.

[*2] この静的環境から，動的物理環境，社会的複雑さへの 3 段階については後にもとりあげる.

[*3] もちろん我々は「多大な労力を払えばそれらが可能になる」ということについてさえ悲観的，否定的であるが.

2.1.1　「自律知」と「道具知」の違い

では，上のような時変で複雑な世界を克服するためにはどのようなロボット設計論を考えていけばよいのだろうか．ここで注目すべきは我々人間の「知」がその世界の問題を少なくともある程度は克服しているという事実である．

従来の知的な，もしくは機能的な自律機械を作ろうという機械設計の観点は，知的な機能を人工物に付与することによってより便利な機械，つまり道具を作ろうといういわば「道具知」の観点である．そこでの “賢さ” とは，あくまで道具としての賢さであり，それを見つめる解釈者の視点はその機械の外側，つまりユーザ・設計者の視点にある．すなわち，「道具知」の知的さを支えるのはあくまで，解釈者としての設計者とユーザの視点である．

しかし，機能を**直接設計**するという方法論では，設計された機能は工場環境のような厳しく限定された環境，つまり設計時に設計者がモデル化し，その条件下で設計が成された環境においてのみ有効であり，環境の多様性，状況の変化，さらにはユーザの個別性といった設計者の想定からの逸脱は忌むべきノイズ，もしくはモデル化誤差として悲観的に捉えられてしまう．機能とは元来，解釈者が自らにとって有用な物理的作用を呼称しているに過ぎない存在であり，主観的なのである．そこには必ず機能を見出す解釈者の存在がある．機能自体は客観的な実在ではなく，解釈されない機能は作用でしかない [177]．このデザインされた機械がユーザの手に渡ったあと，ユーザと言う解釈者によって，デザインが補完されていくプロセスはポストデザインと呼ばれる [205]．しかし，設計者はその機械の用いられる状況，その自律機械が活動する状況を全て把握できるわけではないので，実際にその機械が活動する環境に即したものになっているとは限らない．この問題は，その機械の利用方法やその機械の対面する環境が一般的，普遍的であった場合には誤差程度になるが，個別的で，局所的であった場合にはデザインされた機能にとっては致命的な問題となる．また，ポストデザインのプロセスは機械側ではなく，ユーザ側の機械に対する適応，熟練，そして機械の活動する状況の維持という絶え間ない努力によって支えられている点も考慮されるべきである．先に見たお掃除ロボットの例がその一つであろう．

これに対し，生体のような自律適応系は，本来環境との相互作用を通して機能を自律的に構成していくことができる能力を有している．この構成的な機能を見つめる視点は生体内部にあり，その機能とは自律主体それ自体によって機能と解釈される．こ

のような自律的な知的さは上で述べた道具としての知的さとは別の次元に位置する．「道具知」に対して「自律知」と呼ぶのが良いかもしれない．このように，主体としての知的さと客体としての知的さは別次元の存在である．自律ロボットを生活環境で用いる際に，我々が期待しているのは，我々の直接的な命令を確実に実行する受動的な「道具知」ではなく，自ら状況を解釈し，行うことをも学ぶ能動的な「自律知」ではないだろうか．

「知能とは何か？」[106] という問いは人工知能を始めとする知能研究では常に立てられる問いであるが，その評価は外部観察者の視点からなされることが多い．R. Pfeifer は参照フレーム問題という言葉で，「外見上知的で複雑な振る舞いをするものが必ずしも，内部に複雑な機構を持っているとは限らない.」という指摘をし，知能を考える上で，観察者の視点が重要な意味を持つことを指摘した [63]．なお，この知能研究における問題については後に再掲する．この外部観察者の評価により，知能を評価するという考え方は「科学的」であり，知能を客観的に測ろうとするときには重要である．この知能を刺激応答の入出力関係のように外部から捉える姿勢はアメリカを中心に発展した行動主義心理学において顕著である．しかし，果たして，知能とは外部から評価されるものが全てであろうか．我々は外部から測られる「道具知」ではなく内部から構成される「自律知」こそ知の本質であると考える．生体において「道具知」は外部から測られる副次的なものに過ぎない．我々自身，科学絶対主義的な時代にあっても，自らの知性はその大半について，内観・自省によってのみでしか感知しえないことには頷かざるをえないだろう．

さて，「道具知」を作ろうという視点から「自律知」に照準を切り替えて，視点を系内部に移すことにより，より進んだ人工物 (機械) 設計へのアプローチが見えてくる．生体は先にも述べたように自ら環境に適応し，機能を発現していくという，非常に高度な適応性を持っている．近年，生体の持つ高度な機能を真似て，機械を作ろうと言う生物模倣 (biomimetics) の研究が盛んであるが，それらの研究の多くは外面から生物の機能を捉え設計する直接設計的な態度であり，生命の本質である長期的な適応性・システムの構成能力に払われる注意は少ない．機械は適応性が無いものという一般的なイメージは，日常会話で「機械的」という言葉が，融通の利かないことや，単調で変化の無い繰り返し作業を意味するところにも現れているが，生命の適応性は，現在，我々の作る機械のレベルから考えると，非常に高度な適応性である．たとえば，人間の認知発達を例にとってみれば，生まれた時はバタつくのみで何も出来ないのが，次第に環境の中を歩き回ることを覚え，言葉を覚え，社会生活を送るに到

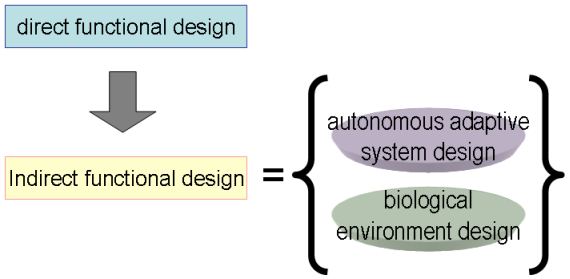


Fig.2.2: From direct functional design to indirect functional design

る．この間に，主体は乗り越えられない段差や，歩きにくい足場，意味のわからない言葉，理不尽な他人など様々な状況に直面するが，それらに適応しながら，段々に自らの認知システムを発達的に構築していくのである．この幼児から成人までの，あまりに大きな発達的飛躍は議論の焦点を吹き飛ばしてしまうように見えるが，これが示している発達を思い切って一文で表現すると以下のようなものとして捉えられる．

“**自律的なシステムが環境との相互作用を通して，新たな機能を適応的・累増的に獲得していき，それらを自らの内部で構造化していくプロセス．**”

これは本当に生物だけの特権だろうか？この表現を満たす人工物を工学的に作ることは本当に無理だろうか？上のような一文として生体の適応能力を端的に捉えることによって，目指すべきものが明確化され，生体の発達的機能生成についての議論と新たな機械設計論のパラダイムを交差させることが出来ると考えられる．

2.1.2　直接的機能設計から自律適応系設計と生体環境設計へ

これまでに，機能を直接設計すること，つまり**直接的機能設計** (direct functional design) が人間環境のような複雑で動的な環境下では上手く働かない可能性が高いことを論じてきた．また，それを乗り越えるためには，生体や人間の認知システムのような自律適応系にヒントを求めるべきだと主張してきた．しかし，自律適応系を作ること，それ自体が直接的機能設計に代わるものになるのだろうか．

学習機械の研究等において忘れられがちになるが，学習能力とは決してそれ単独で機能するものではない．自律適応系が機能を発現するためには，その双対概念が必要

になる．この自律適応系設計の双対概念が，生体環境設計である [177]．生体環境設計は富田らによって提唱されている再生医療において組織を上手く「育てる」ための概念である．これは決して新しい概念ではなく，医学や農学において広く認識されてきた知恵的なものである．自律適応系が環境と相互作用する中で自らの機能を組織化していくというシナリオを考えると，その組織化される機能，構造は環境との相互作用の歴史に強く依存することになる．生体において万能細胞と呼ばれる ES 細胞は相互作用する環境によって様々に分化していく．山本らは ES 細胞の化学的条件は変えずに，力学的な環境だけを変化させることで，ES 細胞の分化を軟骨様組織へと導けることを示した [136]．近年，分子生物学の台頭により生命の発現は DNA のプログラムを実行していくのみというような機械論的見方が蔓延してきたが，このように細胞レベルの複雑さの系ですら，環境との相互作用を通じた機能形成を見せる．また，人間が学習環境の与えられ方に強く依存して成長することも経験的によく知られた事実である．このような人間の発達における学習環境の重要性は J. Piaget などの流れにある構成主義的な教育理論で重要視されており，学習環境設計とよばれる．このように自律適応系設計と生体環境設計が一体となることにより，設計者の手を離れた後の適応的な機能設計が可能になる．ポストデザインの過程における，このシステム自身による機能設計を**間接的機能設計** (indirect fuctional design) と呼ぶ．デザインを設計者の手で完成させず，ユーザと機械の作動環境に委ねるのだ．これは生命が進化の過程で獲得してきた方策と同じである．系統発生としての種のデザインと個体発生としてのポストデザインである [107].

これらのことは主体が自律適応系であるからこそ起こる高度な適応性の産物である．直接的機能設計により設計された機械は決して与えられた環境との相互作用のパターンに依存して機能を変化させることはない．間接的機能設計では機械が与えられた環境に対して機能をその場に応じて自ら設計していくために，与えられた環境が獲得される機能に決定的な影響を与えることになる．故にその環境の設計，つまり**生体環境設計**がもうひとつの機能設計の自由度として残されることになる．ただし，場合によってはこれがシステム自体を不安定にしたり，機能自体を破壊することもありえる．これは先の ES 細胞の例でも確認されている．また，生体環境設計は生体システムの自律適応性を陽に見た考え方であるが，認知システムや社会システムに対しては，学習環境設計や生活環境設計 [123] といった言葉がそれと同型的な意味を持つ．つまり，環境設計は自律適応系において普遍的に存在する機能設計の自由度である．ここではそれらを代表して生体環境設計という言葉を用いることにする．

現在の制御論や設計論ではこの自律適応系の「どうなってしまうかわからなさ」は安定性保証の困難さに結びつき，悲観的に捉えられてしまう場合もある．吉川の一般設計学は設計を機能から設計解への写像として捉えている．これは明らかに直接機能設計を，設計論の対象と見做しており，本研究で議論する自律適応系設計という立場とは対象を異にする [129]．しかし，自律適応系が有する生成・崩壊という構成的プロセスは不安定性が基本であり，安定性保証を過剰に求める負のフィードバックをその基本にすえるサイバネティクス，もしくは制御工学的思考が必ずしも正しいとは，自律適応系設計論の立場からすれば言えない．

2.1.3　自律適応系の構成論的研究

生体の適応についてや，人間の教育についての研究の場合，すでに「そこに自律適応系が有る」という場に研究者は遭遇する．しかし，その適応のダイナミクスは多くの場合謎である[*4]．これに対し，人工物 (機械) の設計者はそこに非適応的なものしか存在しないという場に遭遇する．そこから自律適応系に必要な作動原理を考え出し，ロボット等に埋め込み自律適応系を設計することが研究者の仕事である．しかし，ここで重要なのは，結局両者にとって自律適応系の適応・発達のダイナミクスは未だ多くが謎に包まれているという点である．このため，お互いがお互いに示唆を求め合う状況がある．つまり，生体の適応のモデルを考えることは，それ自体が自律適応系の設計論を考えることに他ならない．逆もまた然りであり，設計し自律適応系を作ろうとすることが，ひいては生体や認知システムの発達のダイナミクスを探り，理解するということに結びつく．このような工学的視点からの自然科学，認知科学へのアプローチは**構成論的アプローチ**と呼ばれており，複雑系や知能研究，社会学研究においても注目されている [106, 63, 110]．ここで，「我々の目的は設計論であり，構成論的理解はおまけである.」ということが述べたいのかというと，そうではない．実は現在，自律適応系を設計することではなく，モデル化し，構成論的に理解するということも，我々が目指している人間と調和する機械設計において求められているのだ．その一例がユーザのモデル化であり，適応・熟達のモデル化である．

序論からここまで，人間機械系協調を議論する際に，我々が自律適応系と見做して

[*4] 近年，細胞内の分子ダイナミクスは徐々に明らかになっているが，その具体的データが我々に生命理解を与えてくれるかというと，悲観的な意見が広がりだしている [190]．故に一段粗視化した自律適応系としてのダイナミクス理解が求められている．

きたのは人間とロボットの関係においてロボットの側であった．しかし，この設定は現在の問題というよりも，近い将来の夢と問題といえる．これに対し，現在，直近の人間機械系の不協和の問題として，ユーザインタフェースの問題がある．

例えば，航空機を例にとっても現在は自動化システムが航空機の全体にいきわたり，パイロットが行うのはその要所要所でのチェック作業といったものが大半を占めている．しかし，自動化の範疇を超えたような非常時や緊急時には人間が再介入せざるをえない．一度制御のループから外に出ていた状態から復帰時に人間は自動化の内部状態，いうなれば「意図」といったものを取り違えることが起こり，このオートメーションサプライズと呼ばれる現象は，近年特に問題視されている航空機事故を引き起こすヒューマンエラーの一因となっている [209]．このような安全・安心の問題を乗り越えるための人間要素の理解が急務とされている [161]．

また，ユーザインタフェースの研究では，インタフェースの改善案を認知科学，心理学，情報学等様々な示唆をうけて提案するが，その評価の結果が熟練者と素人で分かれるということがよく起きる．また，素人に対して補助的機能をつけたインタフェースが逆に素人の熟達を阻害したり，緊急時のオペレーションを阻害することがある．最近では，素人にとっての使いにくさを熟練へ導くための不便益として再評価する流れすらある [132]．ユーザが変化する自律適応系であることを考えれば，刺激応答系としてユーザモデルを捉えるだけでは不十分であり，その熟達のプロセスも考慮に入れた人工物設計が重要である．

道具としての人工物を人間に与えるということはそれ自体が，人間が適応していくための環境を変化させることになる．ある機能的外在物を与えることで自律適応系は「それありき」での機能形成へ進みだす．これが，道具の熟練であり，ユーザが熟練した道具は身体の一部，身体の拡張として認識されることすらある[*5]．

このような項を含めた人工物設計の問題は自律適応系としての人間要素を理解することなしに解決しない．しかも．これらの例は全て人間の認知システムとしての適応性，状態認識の機構や，熟練のプロセスに関わっている．そのような動的なプロセスを理解するためには，定性的な議論だけでは難しく，自律適応系としての人間の構成論的モデルの存在が望まれるのだ．つまり，適応的な認知システムの設計論を考えた

[*5] 生体でも，骨折した骨に補助具をつけたところ，それに荷重が逃げ，骨自身が「それありき」で環境適応を進めるために十分な骨再生が進まないという現象が例として示される [177]．学習環境設計では過保護に育てた子供がスポイルされて無能になるという例が身近である．このような異なったシステム階層間での自律適応系性の同型性は注目に値する．

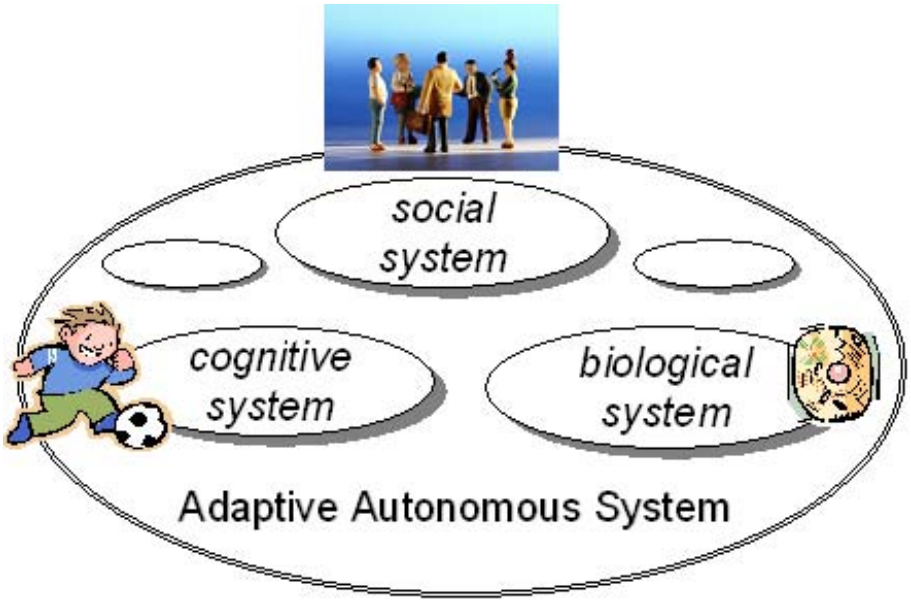


Fig.2.3: Adaptive Autonomous Systems

場合は，それがそのまま認知システム理解の枠組みとして利用されうる場合が多い．実際，脳科学においてはロボットの設計と脳の計算論的理解が並列に進行しており，この表裏一体性が体現されている [166].

2.1.4　自律適応系とオートポイエーシス

さて，ここまでは本論文の題目にも含まれる自律適応系という言葉が何を指しているのかを明らかにせず，導入としての背景を語るため，その言葉を使ってきた．次に本研究において理解しようとしている自律適応系とは何なのかということについて議論を進めたい.

自律適応系研究では一般的なシステム論の系譜と同じく，本質的にはシステムの階層によらずシステム階層間の変換を通して不変な自律システムの構造を探ることを目的としている．例えば，自律適応系には人間の認知システムや，社会システム，そして生体システムなどが含まれる．このような次元の違うシステムの作動原理が同型性を持っているという考え方は，N. Wiener のサイバネティクスに始まるといわれる [170]．サイバネティクスに始まるシステム論の歴史の中で，システム論の興味は当初のサイバネティクスに見られる負のフィードバックによる生命の安定的な (stable)

性質，つまりホメオスタシスのような生命理解から，H. Haken のシナジェティクスや Y. Prigogine の散逸構造のような動的な (dynamic) パターン形成，そして H. Maturana や F. Varela らのオートポイエーシスのような「自らを生成する」動的かつ構成的 (constructive) なシステム観へと移ってきた [115].

本研究で問題にする自律適応系はオートポイエーシスほど過激な自己構成の問題は論じないが，そのシステム観に最も近い.

自律適応系としての生命システムは H. Maturana や F. Varela のオートポイエーシスで論じられてきた．オートポイエーシスとは，構成素が構成素を産出するという産出 (変形及び破壊) 過程のネットワークとして，有機的に構成 (単位体として規定) されたシステムである．このとき構成素は次のような特徴をもつ.

- 変換と相互作用を通じて，自己を産出するプロセス (関係) のネットワークを絶えず再生産し実現する.
- ネットワーク (システム) を空間に具体的な単位体として構成し，また空間内において構成素は，ネットワークが実現する位相的領域を特定することによってみずからが存在する.

この表現では，理解しがたいが，要点は以下の 4 つを持つことである.

1. 自律性　システムは自分に起こるどのような変化に対しても，自分自身によって対処できる能力をもつ
2. 個体性　システム自身でみずからの構成素を産出することによって自己同一性を維持する
3. 境界の自己決定　自己の境界を，産出のネットワークの中から自分自身で決定している
4. 入力も出力もない

これらは彼らが生命を公理的に捉えようとして編み出した定義である．彼らはこの公理が，細胞という一次のオートポイエーシス，多細胞生物という二次のオートポイエーシス，多細胞生物が作り出す社会としての三次のオートポイエーシスに普遍的に成立するという考えを示した．この点については批判もあるが，実際，社会学においては N. Luhmann がオートポイエーシスの考え方を社会システム理論に適用している [178].

また，オートポイエーシスではないが，認知システムについては J. Piaget が発生

的認識論において人間の認知システムを閉じた自律適応システムとして見做し構成主義的な理論を展開した．彼の発生学的認識論の要点は，いかに自律適応系が環境との相互作用を通して，その認識，行動を組織化していくかという議論にある．本研究で自律適応系についての構成論的研究を実際に行うのは認知システムについてであり，J. Piaget の哲学がその基礎にある．H. Maturana， F. Varela， N. Luhmann， J. Piaget らは構成主義者としてまとめられることがあり [145]，その共通性は自律適応系の議論の上で注目すべきである．

自律的知的さを持ち，自ら発達的に機能を獲得していくシステムこそ自律適応系である．今まで，十分な定義をせずにこの自律適応系と言う言葉を用いてきたが，次からその意味に踏み込んで行きたい．

2.1.5　自律適応系の定義

自律適応系とは何かを問うには，まず，自律について触れねばならないだろう．自律とは何かということは，歴史的には反対語である他律との対比によって論じられてきた．

まず教育学，倫理学においては，古くは I. Kant によると，神の意思や快楽を求める自然的衝動にしたがって行動することが他律であり，そこでは意志を規定する法則は外部から，あるいは意欲の対象の側から与えられていたが，自律では意志が他なるものや意志の対象に依存することなく，自分で自分に課す法則にのみ従う．つまり，I. Kant は自律の概念の中核を理性的存在者の自由な自己立法として語る．一方，発生的認識論の J. Piaget は，他律は子供が親の権威に従属して行動決定をしている時期，自律は親の権威を離れて，自己の欲求に基づいて，あるいは自己で理由を意識して行為を決定する時期と述べている [114]．ここでは I. Kant に比べて，自律の定義における理性の色合いは薄れ，システムの行動を決定するものが他のシステムか自分自身かという点に焦点があてられる．これらは，全て行為主体が「人間」であることを前提とした自律の定義である．I. Kant に比べ J. Piaget の定義の方が本研究における自律性についての感覚には近いが，それでも意志決定基準が明記されている以上，「人間以外が行う意思決定とは何か？」という問題がおきてしまう．

また，ロボットの遠隔操作の世界では，オペレータの動きをそのままロボットの動きに反映する他律的な制御方式を master-slave 方式，オペレータの操作を受け入れず完全に独立の機構として動くロボットを完全自律方式，この中間として両者の意思

決定がロボットの操作に乗る方式を共有自律 (shared autonomy) 方式と呼び人間機械協調のための方式として研究されている [35]．この区分は J. Piaget のそれに近いが，他律を無くすことが自律と言っているようで，自律系の積極的な意味合いが汲み取り難い．もちろん当研究は機械設計を研究のルーツに持っており，この区別が我々の自律の定義の出発点になっている．しかし，そのような単純な自律性の定義は人間と相互作用する自律機械を議論する基礎としては不十分である．

これに対して，本研究では H. Maturana，F. Varela らがオートポイエーシスとして定義した，生命システムの持つ自律性こそ問題とする自律系のイメージに最も近いと考える[*6]．生命システムでは適応性が無く自律的に振舞うのみでは生存は難しく，いかなる生命とて適応的でなければならないので，生命は自律適応系について問うとき一番に参照しなければならない項である．なお，生命システムの議論では自律システムと言うだけで，そこに当然適応性のニュアンスも含まれる．しかし，我々は研究領域として機械システムを自律適応系に近づけるという作業をしており，機械システムでは「自律」と銘打つだけでは適応性は当然のものとは考えられないので，自律適応系という言葉が意味を持つ．ところで，オートポイエーシスにおける自律性の概念が我々の自律性の概念に一番近いと言ったが，オートポイエーシス自体は

1. 生命の個としての認識論的な閉鎖性とそれに基づいた自律適応性
2. 再生産のネットワークとしての自己複製能力

という大きく分けて二つのことを同時に論じているように思われる．もちろん H. Maturana にとってはこの不可分性こそオートポイエーシスの力点だったのかもしれないが，この二つを一体としたところにオートポイエーシスの難解さ，不明瞭さの一つがあるように思える．後に，F. Varela はこれを明瞭にするため，前者に重点を置き自己組織化現象と関連して「作動上の閉じ (operational closure)」こそ重要であるという認識を示している．本研究での自律適応系設計における目標は “**自律的なシステムが環境との相互作用を通して，新たな機能を適応的・累増的に獲得していき，それらを自らの内部で構造化していくプロセス．**” を実現することであるので，系自体が構成されていくプロセスについての議論は外す．よって，このオートポイエーシスの議論から前者の自律適応性だけを我々の議論に組み入れることにする．ここでの自

[*6] もっとも彼らは自律性は生物にとって当たり前であり，考えるまでもない前提条件であると考えていたようだが [51].

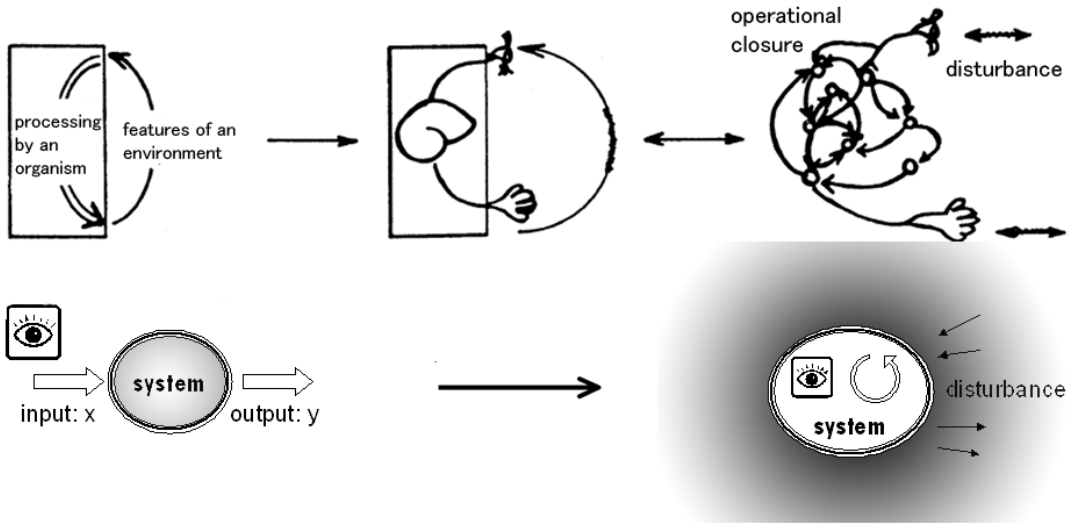


Fig.2.4: Top: Equivalent transformation from input-type description to closure-type description [95]; bottom: input-type description and closure-type description from a viewpoint of an observer of the system

律性の概念は F. Varela の**「作動上の閉じ」**という概念と密接に繋がっている．自己組織化現象は生命にとって普遍的な現象であるが，F. Varela は，作動上閉じたシステムが固有行動をとりながらナチュラルドリフト (natural drift) によって変化することこそ自己組織化の原理であるとしている．F. Varela によると作動上の閉じとは，システムの組織 (organization) の特性を記述する際の記述様式の一つである．記述様式として**入力型の記述**を取るか，**閉鎖型の記述**を取るかということが重要である．F. Varela は以下のように主張する [95, pp.35].

『入力型の記述とは，定義された入力集合と代入される伝達関数を通してシステムと環境が相互作用するというやり方でシステムの組織を定義するやり方である．この方式は現在もサイバネティクスに端を発する制御理論の標準的記述様式であり，これはシステムの「メカニズムが何か？」という点に注目している．だが，生命システムの研究をみると，入力型の記述とは相補的な記述様式を考えざるをえない．直感的な表現をすれば内部決定や自己主張が (つまり自律性が) システムに存在するという事実に基づいている．そのような自律的システムにとって，特性記述の主要なガイドラインは入力の集合からではなく，構成要素の相互連結から生じるシステムの内的コヒーレンス (internal coherence) という性質である．だから，作動上の閉じという用

語を使うのである．システムの組織を定義するためのスタンスを入力型から閉鎖型に変えた主要な帰結として，内的コヒーレンスに注目が集まり，そのため従来は環境からの特定の入力であったものが，不特定の攪乱ないし単なるノイズとして括弧に括られる．』

ここで本質的なのは，視点の移動である．Fig. 2.4 下図にこれら二つの記述様式のイメージ図を示したが，ここでの目のアイコンがその視点を表している． 先に挙げた H. Maturana のオートポイエーシスにおける有名な言葉に「入力も出力もない」という言葉があるが，これもこの作動的閉鎖性のことをさしている．筆者の言葉で言い換えてみる．我々人間というシステムにとっては何が入力で何が出力かと考えてみると，外部の存在が必要になる．しかし，我々の見ている世界は客観的，実在的世界そのものではなく，網膜に映った像を手がかりとして起こっている脳内現象に過ぎず，実際に外部に同様のことが起こっている保障は何もなく，外からの刺激は内部の出来事に対する攪乱でしかない．しかし，このように自らの**現象学的領域**に現れるもの以外我々は知る術が無い．自律系は視点を系内部に持てば，そこに起こるものしか観察できないと言うことを言っているのだ．このようにオートポイエーシスは生命システムの議論であると同時に認識論の学問である．そして我々がオートポイエーシスで考えられる自律性に共鳴する点もここにある．

もちろん視点を系外部に持つことが許されれば (Fig. 2.4 下図左)，単位体は環境との絶え間ない相互作用に晒されていることが観察され，そこには「入力も出力もある」．F. Varela が入力型の表現と閉鎖型の表現が相補的だと言ったのはこのことである．

記号論の文脈で後に引く C.S. Peirce の哲学 (プラグマティズム) にとっては，この立場は基本的な前提である．それは，C.S. Peirce が彼の学問の基本となる現象という言葉を以下のように定義している点からも読み取れる [62].

『私の言う現象とは，それが実在の事物に対応するか否かに全く関係なく，どんな仕方においてであれまたはどんな意味においてであれ，心に現れるいっさいのものの総合的全体を意味する．』

上で現象学的領域と呼んだのはこの心に現れる領域のことである．

これと同じことを，生物記号論の立場から述べたのが J. Uexküll である．J.

Uexküll は生物学的な研究において，様々な生物が自らの感覚器，行動器に閉じた世界を持っていることに注目した．そして，身体が異なれば，それぞれの生物独自の主観的世界つまり現象学的領域を持っていると考え，これを環境世界 (Umwelt) と呼んだ．この視点は自律適応系の発達を考える上では非常に重要である．なぜなら，自律適応系の発達に利用できるのはこの Umwelt での出来事だけだからである [94]．また，他者の言葉はこの Umwelt の中でしか接地できないのだ．この立場は自律ロボットにとっての認識を考える上で示唆に富み，近年再評価されている．

本研究ではこの「作動上の閉じ」を自律適応系の前提条件とする．これを前提とすれば明らかに，外部からの直接的な意思決定は不可能になる．なぜなら，外部からの情報は自律適応系に解釈されない限り，全て攪乱でしかないからだ．自律性を特徴付ける際に歴史的には意思決定の場所に重きがおかれてきた．我々が重要視したいのは系にとって扱い得る情報についてである．自律適応系においては自らが得る情報以外は意思決定にも使えなければ，適応・学習にも使えない．また環境認識自体がそれらの情報によって生成された内部構造によってなされる再帰的な構造が存在する．

2.1.6　自律適応系はコミュニケーション可能か？

「作動上の閉じ」とはまさに我々自身が持っている認知システムとしての特徴でもある．しかし，この自律的な閉鎖系は，如何にして他のシステムと協調することが出来るのか．オートポイエーシスの中で H. Maturana と F. Varela は構造的カップリングという言葉を用いて，自律システム同士の協調を説明した．その定義は以下である．

『二つ以上の単位体の行為において，ある単位体の行為が相互に他の単位体の行為の関数であるような領域がある場合，単位体はその領域で連結 (カップリング) していると言ってよい．カップリングは，相互作用する単位体が，同一性を失うことなく，相互作用の過程でこうむる相互の変容の結果として生じる.』

しかし，その具体的作動原理については明確ではない．数理的に情報伝達を取り扱ったのは C.E. Shannon が初めてであるが [75]，C.E. Shannon の情報概念はあくまで，送信者がメッセージを信号にエンコードして送信し，受信者がそれを受け取り，決められたコードに従いデコードし解読するといったものであった．これは，入力型のシステム記述でのみ成立しうる形であり，意味作用は捨てられる．そのために，自律適応系を前提にすると，システム間のコミュニケーションも C.E. Shannon

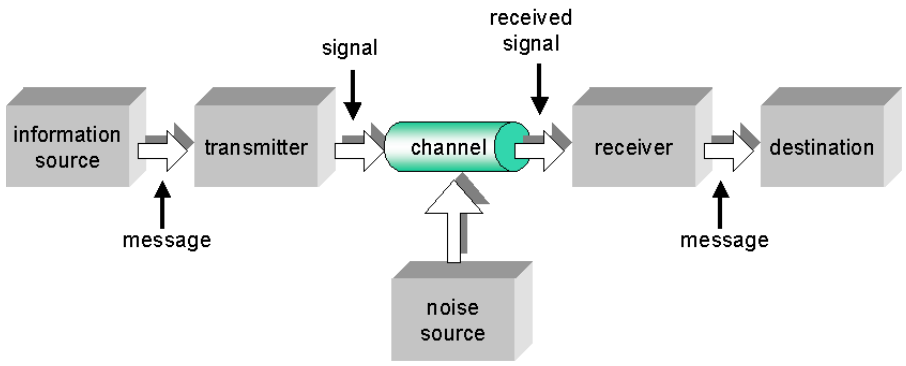


Fig.2.5: The Shannon-Weaver communication model

の情報観のような解釈の自由度を陽に見ないモデルから，個人の解釈の自由度が存在する C.S. Peirce の記号論的なモデルへと移行する必要が現れると，我々は考える．なぜならばコミュニケーションは自律適応系同士の構造的カップリングにより生じると言っても，構成論的理解，設計論展開の両面で作動上閉じた自律適応系同士のカップリングは如何にして生じるのかという問題が残るからである．次節では再び第 1 章で掲げた問いに戻り人工物とのコミュニケーションの可能性を考える．

2.2　人工物とのコミュニケーションとは何か？

本書の冒頭で挙げた問いを再確認しよう．

「自律適応的に言語を獲得し，人間と意思疎通出来るようになる人工物を作ることは可能だろうか？」

本節では，まず，意思疎通，つまりコミュニケーションとはいかなるものかについて議論することにする．イメージをわかりやすくするために人工物として自律ロボットを挙げる．たとえそれがプラントの操作盤であっても問題は変わらない．

近年，人間とのインタラクションを実時間で行うまでに情報処理技術が進歩したこともあり，人間と自律ロボットの間のインタラクションを題材にした研究がここ 10 年ほどで多く見受けられるようになった [17]．そのような研究の中には，「人間とロボットとのコミュニケーションに関する研究」といったように「コミュニケーション」という言葉を明示的に表題に含めた研究も多くある．また，商品としてもペット

ロボットを始めとする多数のロボットが現れてきた．しかし，人間同士で行うような言語的，もしくはジェスチャーなどを含めた非言語的コミュニケーションを通じて感じえる意図の伝達といったようなものを，それらのロボットとの「コミュニケーション」が我々に感じさせてくれるかという問いには，否定的かつ悲観的にならざるをえない．むしろ，我々の行っているコミュニケーションと自律ロボットとの間で行う「コミュニケーション」の間には定性的な差を感じる．もちろん，この議論はコミュニケーションと言う単語をどう定義するかの問題であり，それがない限りはこの漠然とした「定性的な差」を解決するための議論に進むことはできない．

2.2.1　コミュニケーションと記号

まず，コミュニケーションと言うが，それは数理的，もしくは図式的にはどういうことなのか．何を実現すば人間機械のコミュニケーションが可能なのだろうか．この，枠組みの理解が間違っていれば，そこで議論は破綻する．そして，現在の人間機械コミュニケーションの問題はここで往々にして破綻している．

それは, 今まで唯一コミュニケーションを数理的に取り扱いえた, Shannon-Weaver のコミュニケーションモデルに思考によるバイアスを強く受けているためである．コミュニケーションとは辞書的には「言葉による意志・思想などの伝達 (新明解国語辞典)」であるが，この定義自体が Shannon-Weaver のコミュニケーションモデルを受けているように思える．C.E. Shannon 自身は「意味論自体は工学的には取り扱いえない問題だ.」[75] として，信号を劣化無く送信者から受信者に届けるという部分のみをコミュニケーションモデルとして示した．しかし，これが拡大解釈され社会的コミュニケーションの基本的なモデルとして理解されてきている．Shannon-Weaver モデルを基にしたバーロモデルでは岸辺に立った二人の間で小包を送る比喩で情報伝達を説明する [178]．繰り返すが，Shannon-Weaver モデルは「意味」を取り扱っていないのである．コミュニケーションの定義は「言葉による意思・思想などの伝達」と言うが，実際には言葉には意思も思想も詰め込めない．

記号論の祖 C.S. Peirce は『記号は対象を表意し対象について語ることが出来るだけである．記号はその対象の直接知や認知を供給することができない.』と断じている [61]．実際の人間同士のコミュニケーションを思い浮かべてもらいたい．相手の話す言葉は，何も直接的に運んでは来ない．あくまで，それに意味づけし「相手はこう考えているんだろうな」と推し量るのは解釈者の能動的な活動である．「相手の本当

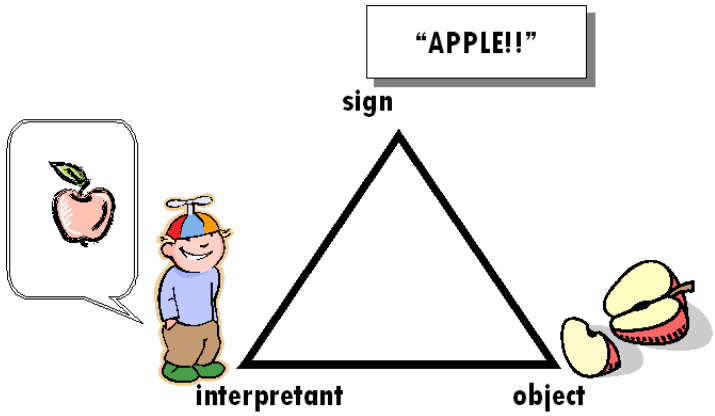


Fig.2.6: Peirce’s semiotic triad

に考えていることなんて神様でもないかぎりわからない」というのが実際である．ここに哲学でいう他我問題が存在する．バーロモデルのようなコミュニケーションが実際であれば他我問題は存在しないように思われる．しかし，実際には人間は言葉を通して相手の考えがある程度推測出来る．それゆえにバーロモデルは実際の現象を説明しているように思えてしまう．もしバーロモデルを信じたとしても，そこに一つの問題が残る．「誰が互いの送信機や受信機をチューニングするのか？」．人間が自律適応系であるという事実を受け入れると，どこにも送信機，受信機の差を見積もって調整できる人間なんて居ない．自分の送受信機をチューニングできるのは自分だけである．

Shannon-Weaver の意味を剥奪した後の情報観に立ち止まっている限り，人間と人工物の意思疎通は出来そうにない．よってそれに代わりえるものとして記号論，記号学の考え方を導入する．今までも何度となく C.S. Peirce，J. Uexküll の名は現れたが，再度ここで導入する．

まず，C.S. Peirce の記号論を参照しよう．C.S. Peirce はプラグマティズムの親である哲学者であり記号論 (semiotics) の創始者である．C.S. Peirce は記号を従来の指示されるモノ・指示するモノといった二項関係的な存在ではなく，サイン (sign)，対象 (object)，解釈項 (interpretant) の三項関係であることが本質であるとした (Fig. 2.6)．これは記号の指示関係が固定的な存在ではなく，解釈者の解釈活動を通して結び付けられるものであること示している．人間や動物におけるセミオーシス (記号過程:semiosis) とは，サインと対象とを解釈者が過去の経験や類推などによって仮説推

量的 (abductive) に結びつけて解釈していく過程であり，この動的な過程が記号を意味付ける [176]．例えば，医者は発疹から麻疹 (はしか) と診断するとき，記号 (サイン) は発疹，指示対象は麻疹であり，解釈項は医者の心の中でつくられる麻疹のイメージである．すなわち，医者は発疹から麻疹という「意味」を読み取るわけであり，それが「仮説推量 (abduction)」と呼ばれる思考の過程に他ならない．

日常生活の上では，文字や絵記号，音声といったサインの事を記号と呼ぶことが多いが[*7]，一つのサインには文脈や解釈者次第で様々な意味付けを行うことが可能であり，サイン自体が意味を直接的に伝達することは出来ない．故に，記号の持つ指標性をコミュニケーションにおいて利用する為には，その指示関係について前もっての規約が必要である．このような規約は天気の兆候や，病気の兆候といったように自然界からもたらされる場合もあれば，言語や風習といった形で人間社会が恣意的に設計する場合もある．いずれにせよ重要なのは意味生成が解釈者の内的なプロセスであり，それは主観的に決定されるということである．このようにパース記号論は，個人の思考過程や認知活動に着目して「意味」を捉えていった [178]．

これに対し，記号学 (semiology) の創始者である F. Saussure は，個人の記号過程というよりは集団の持つ言語体系に存在する恣意性に注目した．この恣意性には二種類あるが，特に重要なのはある記号 (言葉) によって何と何が区別されるかと言うことについてアプリオリな必然性は無く，その言語を持つ集団の文化によって恣意的に区別されているということである．つまり範疇化についての恣意性である[*8]．故に，自律系内部での範疇化についての恣意性は環境との相互作用の履歴やその自律系が属する社会に影響を受けるはずである．また，その範疇をどういう音韻やその他のサインで表現するかということがもう一つのラベル付けについての恣意性である．

記号学 (semiology) と記号論 (semiotics) は記号を見る角度の違いであり，本質的には相補的な存在である．C.S. Peirce の記号論は特に解釈者個々人に注目し，F. Saussure は共通言語を有した集団に注目したと言える．現在はこれら二つの流れはほぼ合流したと言ってよく，まとめて記号論と呼ばれる場合が多い．

C.S. Peirce と F. Saussure，記号論と記号学双方の要点を結合させると，以下のよ

[*7] 人工知能研究における記号も多くの場合この意味で用いられる．

[*8] ただし，Bateson が情報とは「差異を生む差異」といい [12]，西垣が「パターンを生み出すパターン」[178] というように，差異を解釈者の中に生み出すなんらかの存在が知覚する世界の中に存在しなければ，区別という操作は起動されずそこに情報は生まれないと言う点には注意すべきであろう．

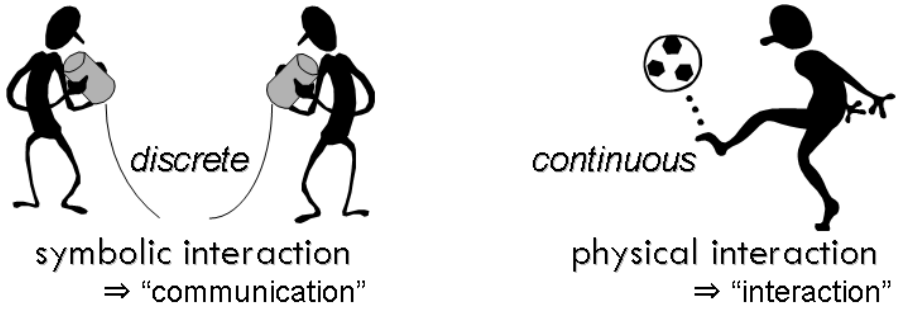


Fig.2.7: Differences between symbolic interactions and physical interactions

うな基本的認識が得られる[*9].

1. 記号とは三項関係であり解釈者無しには意味は生まれない． ゆえに，記号の意味は解釈者依存である．
2. 内的表象は解釈者の環境世界 (Umwelt) を恣意的に分節化しており，その分節には何ら普遍的・絶対的根拠は存在しない．
3. 環境世界の分節化は解釈者が所属する集団の持つ言語のおこなう分節化に影響を受ける．しかし，その分節には何ら普遍的・絶対的根拠は無い．

主に一つ目が C.S. Peirce の記号論の主張であり，二つ目，三つ目が F. Saussure の分節の恣意性についての主張を一部修正し二段階に分割したものである．F. Saussure は構造主義者らしく共時的な複数言語の間の差異体系の違いを問題にしたが，同じ言語を使う集団内の個人の間でも各成員が自己閉鎖的である故に世界の分節のずれはあるはずである．よって，J. Uexküll の言葉を用いて，その「世界の分節」を「環境世界の分節」と書き換える．個々の解釈者は作動的に閉じているので，他者の環境認識を知ることは出来ない．そのため「環境世界の分節」は，ずれているはずである[*10]．この考え方は記号学側に一部修正を迫るだけでなく，記号論側にも修正を迫る．C.S. Peirce は記号過程におけるダイナミクスについて仮説推量 (abduction) を通して説明したが，仮説推量の可能性は無限にあるはずであり，それが意味を成すためには記号過程が始まるには仮説が先行的に存在しなければならない．その仮説の体系こそ，

[*9] C.S. Peirce は記号過程を人間の認知プロセスにおける物，つまり解釈項が人間の解釈者内部にあるもののみならず，より広範な現象についても議論しているが，議論を明確にするため，記号現象の基本となる解釈者が人間を含めた自律適応系である場合を考える．

[*10] もとより環境世界がずれているのだが

F. Saussure が差異体系としてイメージしたものではないかと我々は考えている[*11].

自律適応系として生まれた解釈者は，初期は世界についての概念が未分化であり，環境世界やその環境世界に構造化された攪乱を与える他者との相互作用を通して自らの内部に差異体系を獲得していく．その差異体系を用いることで，刺激・信号を仮説推量的に解釈できるようになり，そこに記号過程が生まれ，ゆえに記号が生まれる[*12].

しかし，重要なのはこの差異体系がいかに自律適応系内部に浸透するか，もしくは構成されるかである．これについては記号論，記号学共に現在に至るまで強い論拠を持っていないというのが筆者の印象である．

ここまでで，Shannon-Weaver 流の見方から，解釈者の能動的な意味生成を基本にした記号論的な立場にシフトしてきた．この立場に立つと，Shannon-Weaver 的な「送信者」「発信者」の意味に対する平等な関係は消え，むしろ意味は「受信者＝解釈者」に独占される．この考え方でないととらえられない記号現象のかる一例を挙げよう．

「病気の母を持つ青年が秋の枯葉が散るのを見て悲しい思いに駆られる.」
この例では解釈者としての青年が枯葉を母の死の記号として解釈しているが，ここにそれを伝えようとしている送信者はいない．このように記号論は意味を解釈者の営みとして捉え，あくまでその延長線上に言語によるコミュニケーションを見るのである [176]．逆に「意味」という問題をとらえる時に，「意味」の発信者がいるような系のみを扱うのは狭量であり，このように拡大して考えた場合の方が適切な接近が得られると考えられる．

2.2.2　記号的相互作用と身体的相互作用の区別

人間と自律ロボット・自律エージェントとの相互作用を表現するときに，コミュニケーションに良く似た言葉にインタラクションがある．インタラクションという言

[*11] 分節には何ら普遍的・絶対的根拠は存在しないという表現は，分節化に何の基準も無いことを述べているのではない．あくまで，人間の存在とは独立に，それのみで世界を分節化する基準がアプリオリに存在することは無いという意味である．近年は認知意味論，認知言語学 [137] において，言語の成立に人間の認知的な要素が強く影響していることが明らかになりつつあり，故に差異体系の由来が人間の認知とそれを生み出す物質世界，身体性に強く依存していると考えられ始めている．

[*12] C.S. Peirce の意味では，記号過程，もしくは記号現象と記号は表裏一体の関係にある．解釈を伴わないサインは記号とは呼べず，解釈を伴う記号過程となって初めて記号と呼べるようになるからである [176].

葉は広範に利用され，様々な領域で重要なキーワードになっているが，人間・自律ロボットの相互作用を論じるにおいて，インタラクションとコミュニケーションの区別はうやむやになり易く，それが議論を不明瞭にしている場合がある．そこでこれらの言葉の使い方に差を導入する．

物理的，**身体的相互作用**には**インタラクション**を用い，言語的，**記号的相互作用**には**コミュニケーション**という言葉を用いたい[*13]．

ここで言う，コミュニケーションとは何か．まず，我々の自然言語や手話などを用いた会話をコミュニケーションに含むものとする．このときそれらは C.S. Peirce の意味においても，F. Saussure の意味においても記号を用いた相互作用に含まれる．そこでコミュニケーションと言った時，記号を用いた相互作用を指すものとする．これを記号的相互作用と呼ぶ．記号を用いるといったときに重要なのは，その媒体となるサインには，伝達したいと考える内容についての何の情報も含まれていないことにある．例えば「花」という言葉をいかに解析しても何かの花が咲き出すことは無い．それは視覚・嗅覚などという全く別の感覚モダリティを通して認知される花という対象を指示し，それに応じた解釈項をする想起させるサインなのである．もし解釈者が「花」という言葉を知らないか，「花」を見たことがなければ，この「花」と言う言葉は記号にすらならない．また，「花」を知っていたとしても我々一人一人の花についての経験が違うために想起されているものの相同性は保障されない．C.S. Peirce の考えに従えば，記号現象が起こらない限りそこに記号は無い．つまり，サインと解釈項の間に，アプリオリな結びつきは無い．それは恣意的であるが故に，自由な設計・解釈が可能になる．

これに対し，インタラクションは記号の持つような表象による指示関係を有さない情報のやりとりにより生じる相互作用を指す．例えば，バスケットボールをドリブルしている時に，通常はそれをボールと自分とのコミュニケーションだとは言わない．しかし，相互作用ではある．また，ビリヤードで球と球がぶつかり合うのもコミュニケーションではないが相互作用である．核融合炉内での原子同士の核融合も相互作用である．つまり，物理的な法則を埋め込まれたものがお互いに干渉しつつ入出力関係に基づいて状態変化を続けるとき，その相互作用をインタラクションと呼ぶ．必ずしも埋め込まれた法則が物理的である必要はない．重要なのは相互作用を通して入出力

[*13] ここでいう記号的相互作用は H.G. Blumer らのシンボリック相互作用論とは別であり，その内容は本章内の定義に従う．

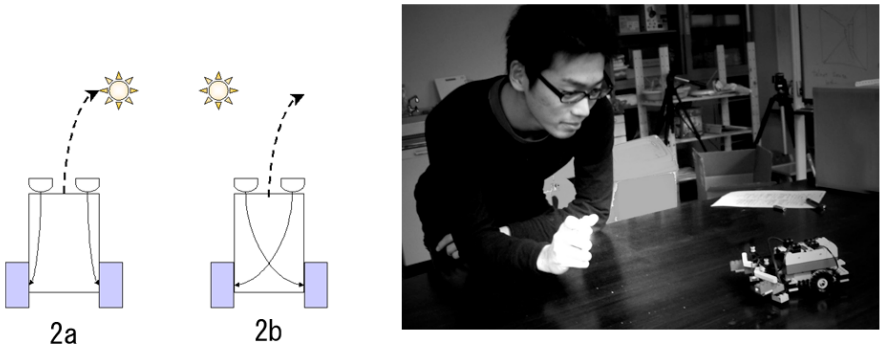


Fig.2.8: It's a simple interaction with a robot using a light, not a communication.

関係が不変に保たれ続けることである．ここに解釈者は存在しない．

ブライデンベルクビークルを例にとろう [15]．ブライデンベルクビークルとは二つの光センサと二つの車輪を備え付けられた簡易なロボットで，そのセンサ・モータの接続しだいで，光に向かっていく振る舞い，もしくは光から逃げる振る舞いを見せる．ブライデンベルグビークルについての議論では，その行動がまるで知的に見えるというという点から，「外部観察者から見た知性とは何か」という議論が展開されている [15]．ブライデンベルクビークルは簡単な刺激応答則を埋め込まれたロボットであるために，人間が光を当ててやるとそれに反応し振る舞いを変えるため，このような相互作用を人間―ロボットコミュニケーションの出発点に考える人もいるが，これは上で述べたバスケットボールに対して力を加えることと本質的に違わない．ロボットにとっては与えられた光の強さがそのままモータ系に直結し自らの運動を変化させたに過ぎない．つまり，ここでは与えられた光度情報がそのままモータ系に利用されているため，伝えられたものは記号的相互作用におけるサインではなく，モータ系を駆動させるための情報そのものであったといえる．外部観察者としての我々が「光を怖がってビークルが回避運動をしている.」といえばこの相互作用も記号的になるかもしれないが，このときの解釈者は観察者本人であり，ブライデンベルグビークル自身内部でそのような光をサインとした三項関係が成立しているわけではない．ブライデンベルクビークルが行っているのは，固定化され解釈の自由度を持たない二項関係的な刺激反応関係のみである．

2.2.3　デザインされた振る舞いはコミュニケーションを生まない

現在の多くの人間と社会ロボットの相互作用についての研究では，それら自律ロボットの振る舞いが設計者によって既にデザインされている．もちろん，その振る舞いは外部からの刺激で選択されて発火される仕組みを持っているために，ユーザの行為に対して応答することが出来る．しかし，その応答の枠組みは設計者により事前に与えられているために，ユーザとの相互作用を通じて変化することは無い．これがユーザに飽きをもたらしてしまうということに現在のエンターテインメントロボットの持つ一つの壁があることは，第１章でも述べたとおりである．ロボットの変化の無さがユーザに

「関わってもロボットが変化しないので無視されている気がする」

「一生懸命関わっても関わり損」

という疎外感を与える．人間-自律ロボットの相互作用においては「その場の楽しさ」という刹那的側面が前に押し出されることが多いが[*14]，コミュニケーションの本来の価値を意図伝達・共感に置くならば，コミュニケーションとはその与えた情報に対しロボットが即座に反応することではなく，むしろ，話者の表出したサインに対し解釈を行い，その話者の意図を汲みとり自らを変容させることにある．上に示したような外部からの刺激によりある種の動作を発火させるという形でデザインされた振る舞いは，本稿での分類ではインタラクションに過ぎず，記号的相互作用，コミュニケーションであるとはいえない．つまり，デザインされた振る舞いを基盤にした自律ロボットでは基本的に人間との記号的コミュニケーションを達成しえない．つまり，人と関わる上での相互作用の仕組みを設計者が**直接設計**したのでは，人間社会に溶け込む社会的な人工物は生まれないというのが筆者らの主張のひとつである．

2.2.4　創発する振る舞いこそコミュニケーションの基盤となる

これに対し現在，環境や他者との相互作用を通じて振る舞いを創発させる研究が注目されている．創発する振る舞い (emergent behavior) とは，デザインされた振る舞いに対し，人間や環境との相互作用をとおして自律ロボット自身が自発的に獲得していく行動を指す．上にも述べたように現在存在する多くの自律ロボットは基本的にそ

[*14] 実際は適応性がデザインできないために，刹那的側面を前面に押し出さざるをえない面もある．

の振る舞いを設計時に決定されている．これに対し，先にも述べたように人間をはじめ多くの生物は，発達の過程において様々な行為を能動的に獲得していく．

人間のような社会的な動物の発達過程においては，母親のような教師の存在が重要視されるが，すべての振る舞いの獲得において母親が必要なわけではない．社会的行動の獲得においては，母親のような社会との接点としての人 (これを社会ロボット研究では養育者 (caregiver) と呼ぶ場合が多い．) の存在が重要ではあるが，身体的な振る舞いの獲得はむしろ個人による自己閉鎖的学習の結果であると考えられる．なぜならば，我々が歩行，起き上がりなど身体的な振る舞いを獲得する上で相手にするのは物理的な力学 (physical dynamics) であり，その時に学ぶべきは自らの振る舞いに対し環境，対象がどう応答するのかという身体的相互作用に含まれた構造，因果関係のみだからである．このプロセスにおいて母親のような教師に出来ることは，幼児の行為や知覚を直接操作し，教師あり学習のように答えを教示することではなく，幼児が得る情報の源となる環境，対象の状態を整えてやることだけである．これを**学習環境設計**と呼ぶ．心理学的にはこのような学習者の自律性を重要視した立場は構成主義的な立場と呼ばれる．

振る舞いの創発という問題を考えたときに，それは振る舞いの学習とは異なった問題を含む．それは，主体があくまで自己閉鎖的な作動の中で主観的な判断に従い複数の振る舞いを累増的に獲得していかねばならないという問題である．振る舞いの創発においては，外部教示者が頭の中でデザインした振る舞いを自律ロボットに教え込んでいく作業ではなく，自律ロボット自身が自らの中で区別すべき振る舞いを意味づけ，獲得して行くという自己閉鎖的なプロセスが重要となる．そして，このような振る舞いにおける概念分化 (differentiation) を起こす能力こそ自律ロボットの認知発達の基盤となる．もし振る舞いの創発が不可能であればコミュニケーションはどうなるだろうか？

C.S. Peirce は解釈項に三種類の区分を設けている．それらは明確な区分とは言い難いが，情動的解釈項・活動的解釈項・論理的解釈項である．記号解釈の例で，解釈項として「林檎のイメージ」を示したが，このようなイメージに関わるものは一つ目の情動的解釈項でしかない．実際には，記号は，活動的解釈項を生み解釈者に何らかの行動を反射的に取らせたり，論理的解釈項として内的な記号操作による思考を生み出し，多くの場合，最終的に何らかの行動に帰着する．極端な例だが，麻疹の例の場合は，麻疹の治療という解釈項としての行動に落ち着くであろう．このとき，解釈者に向き合う人間にとって，相手の解釈の豊かさがわかるのはこの行動を通じてのみで

ある．また，この行動の可能性は仮説推量のための仮説として解釈者内部に保持されていなければならない．これより，振る舞いの多様さが無ければ，記号的相互作用の多様さも失われる．つまり，自律適応的に記号を獲得するには，振る舞いの創発がその基盤として不可欠なのである．

2.2.5　記号的相互作用の創発

ここまでで，コミュニケーションのイメージを Shannon-Weaver 型から Peirce 型に移行した後に，コミュニケーション (もしくは記号的相互作用) とインタラクション (もしくは身体的相互作用) の区別を行った．本稿での議論が適切であれば，記号的相互作用は本質的に，自律ロボット自らが環境や他者との相互作用を通して行為や知覚についての内部表象の自律的な獲得，組織化を行うことを要求するということになる．それが必要な理由は主に以下の二つである．

- 記号的相互作用は解釈系を必要とするが，どのような身体を持つ自律適応系にも普遍的なアプリオリな解釈系は存在しないため．
- 相互作用の中で変化していくことこそ，コミュニケーションの本質であるため．

解釈系の生成は相互作用の中での解釈系の変化に比べて，ややダイナミックに感じるが，この二つは別のようで繋がっている．その結果要求されるのは，感覚入力から内部系，運動出力へ流れる情報の流れから概念を組織化し，分化させていくという，主体内部での情報の自己組織化プロセスである．これを身体性に立脚した**記号創発**と呼ぶ．

さて，本節を通して，人工物とのコミュニケーションを可能にするためには人工物内部での概念の組織化が必要であるという結論に至った．つまり，コミュニケーションはその参加者が自律適応的な存在であることを元より求めているというのが，本研究での見解である．

2.3　人工知能における記号の歴史

さて，前節までで，本研究において設計と理解を目指す自律適応系がいかなるもので，それが行いえるコミュニケーションを実際にはどの様に捉えるべきかということについて議論してきた．ところで，知的な人工物を作ろうという研究は広い意味での

人工知能研究であり，当然のことながら当研究もその系譜の中にいる．

では，旧来の人工知能においては，人間と相互作用するための，もしくは自らの意思決定に用いるための記号をどう捉えていたか．その問題について本節で俯瞰し，GOFAI(Good Old Fashioned AI) と呼ばれた黎明期から 80 年代ごろまでの所謂古典的 AI における記号の持った意味と，MIT の R. Brooks らの批判によって始まった NewAI と呼ばれ，または行動主義的ロボティクスともばれる思想への展開について議論する．また，80 年代の NewAI は人工知能研究の上では既に古典となりつつある．R. Pfeifer は身体性認知科学という視点からこれらの考え方を整理し，さらに進めているが，単純な拡張では限界があると感じられる．

2.3.1　GOFAI と NewAI

古典的な人工知能研究においては知能は表象操作に限定され，人間の持つ知性を模擬するようなコンピュータの開発を目指していた．当時の研究者が無意識にもしくは意識的に抱いていた人間の知性についての概念とは，動物と人間を区別するものであり理性的で論理的な思考をさしていた．その能力は問題解決や定理の証明といったものに象徴される，所謂数学的・論理的思考をさした．それは記号処理を動作の基本とし，形式的な操作のみに閉じた系であった[*15]．これらの流れを汲み研究される人工知能のことを古典的人工知能つまり GOFAI(Good Old Fashoned AI) と呼ぶことがある．人間の知能がすべて論理的判断のみで構成されていれば，古典的人工知能の研究により人間の知能はほぼ模倣されえるはずであった．実際，多くの点で古典的人工知能は成果を上げている．例えば，チェスのプレイにおいては IBM 社の Deep Blue が人間の世界チャンピオンに勝利するに至っている．このように，もし人間の知能というものを形式化された世界における計算論的なものと限定するならば，機械はすでに人間の能力を超越しているのである[*16]．

しかし，事態はそれほど簡単ではない．問題は形式化された知能と実世界とのインタフェースにあった．ゲームのような形式的な世界では高度な能力を発揮する計算機が，実空間では余り上手く動かないのだ．これを機に古典的人工知能に対するより

[*15] ここでいう記号とは後に示すパースの三項関係としての記号ではなく，形式操作により扱われるサインとしての記号である．現在，諸分野で「記号」という言葉の使い方はこの二つの意味で混迷を極めているので注意せねばならない．

[*16] 余談であるが，すでに世界チャンピオンと互角の DeepBlue の子孫はノートパソコン上で動作する．

根本的で重要な問題点とそれに対する批判が出始めた．それは基本的に全て，形式化された知能と実世界とのインタフェースを問うものであり，フレーム問題などが有名であるが，その中で本論文においても特に重要な問題が記号接地問題 (symbol grounding problem) の問題である．その問題において主要な役割を果たすのが，古典的人工知能における身体性の欠如である．

古典的 AI に対し NewAI と呼ばれる考え方が 80 年代に台頭してきた．GOFAI の問題点は実世界で身体を持ったロボットに適用すると，ほぼまともに動かないということだった．例は文献 [63] に詳しい．これに対して，R. Brooks らは包摂アーキテクチャ (subsumption architecture) と呼ばれるアプローチを提案した．これは GOFAI の持つ，中央集権的な記号処理を廃し，分散的な刺激反応系だけから振る舞いを生みだすという考え方に基づいていた．これにより設計されたロボットは実空間で柔軟に動いた．それは振る舞いの面から見れば随分知的に見えた．この NewAI のアプローチは行動主義ロボティクスと呼ばれる．行動主義とはアメリカを中心に栄えた心理学の考え方で，人間を刺激反応系としてとらえ，行動に表出しないもの，例えば内部表象については不可知とする態度をさす．故に NewAI の態度は，徹底した記号系の廃棄という GOFAI に対する極端なアンチテーゼとなっている．

2.3.2 記号接地問題

現在もこのような知性を身体からボトムアップに捉えようという考え方は根強いが，NewAI 自体はすでに新しい考え方ではないと感じられる．実際に包摂アーキテクチャのような単純な仕組みでは振る舞いの面でも知的さに限界があり，また，刺激応答系として知能をデザインするために適応性という知性を持たせることができないという問題点がある．となると，どの様に自律的系に適応性を持たせるかという問題が自然と立ち表れてくる．しかし，適応性を持たせたとしても，人間とコミュニケーションを行わせるようなロボットの設計に内的な表象なしで立ち向かうのはあまりに無謀と感じられる．やはり，内部表象を完全に捨てるのはラディカルすぎるのである．

一方で，NewAI ほど極端な態度には出なくとも，GOFAI の問題点をどうにか乗り切らねばならないという問題は並列に存在し続けており，現在もなお解決には至っていない．その問題を端的に示すのが，**記号接地問題**である．記号接地問題とは S. Harnad が提唱した問題で，人工知能の持つ記号系をどう実世界と結びつけるかとい

う問題を記号接地問題と呼び，とり上げた [32, 163]．そこでは，トップダウンな記号設計とニューラルネットなどによるボトムアップな学習の接点を見つけることが重要だとされていたが，S. Harnad 自身はあまり具体的な解決は提案していない．

記号接地問題とは言うなれば，ロボットの内部にデザインされた記号をどう意味付けするかという，意味論的問題といえる．意味を扱うには工学的にはオントロジーや意味ネットワーク (semantic network) という表現が用いられることが多い [68]．しかし，それらの方法が捉ええるのは，あくまで記号同士の関係においての意味関係である．簡単に言えば，言葉を言葉で説明する行為がネットワークによる意味づけに他ならない．その作業はネットワークである以上，ノードが有限ならば意味の元をたどっていけば形式的な記号のネットワークを彷徨った挙句に意味を不問に伏す端点にたどり着くか永遠のループにはまるかである．この意味論は，いわば，言語の中で記号の持つ価値が，その言語内の他の記号との対立関係のみから決定されるという考え方になる．そこに実世界に存在する「対象」との関係は無く，「サイン」のネットワークがあるのみである．

GOFAI ではトップダウンに身体から切り離された意味をロボットに与えようとしていたように思われる．しかし，記号の本質とはそのように解釈者から切り離された静的ものではなく，解釈者内部で身体を用いた環境との相互作用をとおして組織化された解釈系を用いた記号過程によって生じる動的なものである．この矛盾に，記号接地問題の問題意識自体が影響を受けていることは，あまり認識されていない．先に述べたように，差異体系は自律適応系の環境世界との相互作用を通して形成されていく．ならば，その差異，もしくは仮説に相当する離散的な内的表象の数やその関係は前もって決めることが出来るだろうか．それすら徹底してボトムアップに立ち上がってくるべきもののはずである．とすれば，記号は接地されるというよりも創発すると考える方が自然である．

2.3.3　「自律知」の間接設計へ

GOFAI と NewAI を二項対立としてとりあげたが，その両者において問題があることを示してきた．では，果たしてこの中庸をとることが解決への道なのだろうか．

GOFAI も NewAI も知能をシステム外部の観察者の視点から捉えてきた．NewAI では振る舞いのレベルでの知能を問題にし，ロボット自身の内部構造の組織化という問題に特に関心を払わなかった．その行動は実世界で身体を用いて見た目に複雑で知

的な「動き」は見せることが出来た．しかし，その先に人間とのコミュニケーションなどの先が描けるかというと，悲観的にならざるをえない．GOFAI ではシステム内部の表象系を設計者が直接設計したために内部表象系はシステムの身体性[*17] に立脚しないものとなり実世界での意味を喪失した．しかし，文章生成のような複雑な認知プロセスを表現しえたという点で GOFAI が上であるし，実際に人間との言語的相互作用のデザインのためには現在もこのアプローチが主流である．よって，どちらがすばらしいというよりかは相補的な存在であるというのが実際であろう．

問題はこれら二つのアプローチが共に「知能の設計」を考えていた点にある．ここでいう設計はもちろん直接設計である．知能をスナップショット的な機能のように捉え，それを設計することにより知能を実現しようとしていた．これは，先に述べた「道具知」的な知能観であるといえる．このような立場では，「関わりあうことによって変わっていく」もしくは「相互作用を通して発達していく」ような知能，つまり「自律知」を実現することは出来ない．我々の立場では，相互作用の中で変化していくことの出来る，また知能を構成するブロックを一つ一つ構成し積み上げていく適応能力こそ知能の本質であることはこれまでに議論してきた．この立場では，これまでに人工知能の論争により語られて来た，内部表象を捨てる (NewAI) か捨てないか (GOFAI) の議論は二項対立として成立しない．むしろそれは知能の直接設計という立場では同一であり，自律適応設計という目標の対立項として捉えられる．記号接地問題の視点は，この意味では本研究の指針と相容れるものである．身体知と記号知は相容れないものではなく，共存すべきものである．さらに言えば，前節で述べたように記号接地問題という設定は不十分で，いかに徹底的にボトムアップに身体的な相互作用から上位の記号系が立ち上がってくるかという，身体知から記号知へ道筋，つまり**記号創発問題**こそ，知能研究におけるこれらの対立を止揚し総合するための重要課

[*17] 知能研究，ロボティクス，社会学などの諸領域において，身体性は重要なキーワードであるが，その意味するところは一様ではない．筆者の理解ではそれは以下の三つに分類できる [186].

1. 身体を持っていることが，社会的な存在感を与え，コミュニケーションの要素として重要である．この意味合いはコミュニケーション論や社会心理学において主に用いられる．
2. 身体を持つことで，知能が実空間との間に情報の関連性を持つことができ，これ無しにはあらゆる知能は実世界に設置されないという意味での身体性．言い換えれば実空間と認識世界のインタフェースとしての身体．本稿で用いるのは基本的にこの意味である．
3. 身体性認知科学や移動知においてよく言われる意味合い．身体を持つことで能動的な制御を使わなくても，その力学的特性そのものが身体的に知的に見える歩行や，移動に寄与する．受動歩行が顕著な例．R. Pfeifer はこれを morphological computation(形態による計算) と名づけている．

題なのである．この認識は多少のニュアンスの違いはあれど認知発達ロボティクス，移動知などの先端的な知能研究においても共有されている．次節では，ついに本研究の研究の核心である自律適応系における記号創発の問題へ進む．自律適応系は如何にして自らの閉じた環境世界における主観的情報のみから，それを分節化し，意味づけし，差異体系を獲得し，記号創発へと結び付けていくことが可能であるのかを議論する．

2.4　構成論的記号論 (constructive semiotics)

これまでで，自律適応系と記号論について話を進めてきた．問題は自律適応系という作動的に閉じた系が，如何にして現れるサインを解釈できるようになって行くかという問題であり，双方向的なコミュニケーションまで考えるのであれば，サインを設計できるようになるのかという問題である．このような問題を工学的な視点から追及することは非常にチャレンジングである．しかし，記号現象に今まで記号論研究で行われてきた以上に踏み込むためには，その挑戦が不可欠となる．記号現象の基礎ダイナミクスとしてのセミオーシスを理解するためには，記号論に対する構成論的アプローチが必須であり，この記号論に対する構成論的な接近を**構成論的記号論** (constructive semiotics) と名づける[*18]．

2.4.1　J. Piaget の認識論

記号解釈については，先に自律適応系が自らの環境世界についての差異体系を持つ必要があることを述べた．自律適応系はこの差異体系を通して世界を認識することになる．しかし，この差異体系がどこからやってくるのかについては，C.S. Peirce，F.

[*18] constructive semiotics を和訳すると構成主義記号論と訳すことも出来る．“構成論” は研究の方法論を指し，“構成主義” は哲学的立場を指すので表面上は区別すべき言葉である．本章で触れてきたように筆者の立場は，F. Varela，H. Maturana，J. Piaget といった先駆者と哲学的に共鳴する点が多い．これらの先駆者は構成主義者と呼ばれており，筆者の立場はその意味で構成主義的であるといえる．端的に言うと構成主義とは，「我々の認知する世界は我々自身の構成物である」という考え方を指す．本研究で行おうとしているのは，このような構成主義な主観的認識の形成プロセスを，構成論的に示そうということにある．構成主義の主張は構成のプロセス抜きには存在し得ないゆえに，その理解に構成論的な方法論を採用することは至極自然な成り行きであるといえる．この見地に立つと，日本語では二つの違う言葉になるが，英語表記では constructive semiotics という一つの言葉になる点は示唆的である．

Saussure らの記号論は答えを持たない．J. Uexküll の生物記号論は，生得的な認識系については述べるものの，犬や人間のような個体発生的に解釈系を構成していくプロセスについては説明を持たない．システム間の相互作用を構造的カップリングという言葉の元に済ましてしまったオートポイエーシスもしかりである．西垣の基礎情報学 [178] は我々の立場に最も近く，オートポイエーシスと記号論を結びつけることにより新たな情報観を得ようとしているが，アプローチが社会学的であるために，関心事である記号現象の基礎ダイナミクスとしての記号過程 (セミオーシス) の計算論的記述は持たない．しかし，その立場は我々と意見を共にする点が多く，相補的な存在である．

この点において示唆を J. Piaget の発生的認識論に求める．J. Piaget は哲学的課題である認識論を心理学の視点から個体発生の上で捉えなおそうとした．誰しもが，大人になるにつれて複雑な世界が見えてくるようになるが，J. Piaget はまさに，このような主観的な認識世界の現れ方を発達的に追おうとした．そのダイナミクスを心理学的実験によって実証的に追おうとしたのが J. Piaget の偉大な点であるといえる．

幼児はこの世に生まれてきたとき，右も左も知らない．知らないというよりむしろそのような差異が彼の認知システムの中には存在しない．つまり未分化な状態にある．最も始めには全てが中心化しており自他非分離な状態にある．その後に，内部の認知システムを形成するシェマというモジュール群が分かれ構造化し，幼児にとっての世界が差異化していくと J. Piaget は考えた．これが，オートポイエーシスのような閉鎖系の考えと符合していることは以下のことからも分かる．F. Varela の自律システム論をうけて P.M. Heji は，『入力を持たないシステムとは自己自身の状態に依存した外部の「表象」を持つシステムであり，その状態はシステムの外部表面が外部の事象からこうむる影響によって変更される．システムの変更された状態こそは，システムの諸経験や知覚の生物学的表象である.』と述べている [95]．これはシステムにとっての知覚が，外部からの攪乱としての入力によるのではなく，システム自らの内的な構造による解釈により成り立っているということだ．J. Piaget のシェマシステムはまさにこのような自律適応系の中で認識を生成しつつ，漸進的に構造変化していく認知システムであった．

例を挙げよう．乳児でも大人でも見ている映像はそれほどは違わないだろう．その意味での入力 (攪乱と呼ぶべきかもしれないが) は変わらない．しかし，それを同じものを知覚，もしくは認識しているというのだろうか．乳児においては環境認識は未分化であり [40]，そこに映る椅子も，書物も，ゲーム機も，母の乳房を除いては何も

意味を持たない．つまりそれは見えていないのだ．内的な構造による解釈とは，このレベルで見えることを指す．この前者と後者，つまり生理的，感覚的に見えているのと，知覚，認識として見えていることの違いを理解しなければ，環境認識の分化を通じた差異体系の獲得という話には進めない．Ⅱ 部以降の議論では一貫して前者を感覚，後者を知覚と呼ぶ．

R. Garcia によれば J. Piaget 理論の最も基本的な主張は以下の 3 つである [64].

1. 新生児における生物学な過程と，認知過程の始まりを示す組織化された行為のタイプとの間には連続性がある．
2. 生物学システムと認識論的システムの間には構造的に大きな差があるのに，そのどちらのシステムも類似した**同化**と**調節**を通して生物学的組織を環境に自らを適応させている．
3. 生物学的システムの進化は，認識論的システムの進化のように環境と相互作用する開かれたシステムの例．その二つには相似性がある．

これは物質の交換により自己組織化しつづける開放系としての生体システムと情報の交換により自己組織化しつづける開放系としての認知システムに同じ基本原理が働いているためだと J. Piaget が考えていたためである．自律適応システムの適応原理について論じるときも，このように外部との相互作用を通した情報の流れ (information flow) から自己組織化的なプロセスによりに知識が認知システム内部で生成されていくシステムを設計することが大切だと考えられる．それでは，J. Piaget はどのようにその具体的なシステムを考えていたのかを，本研究に関係する範囲で次に紹介する．

2.4.2　シェマシステム

生体が外界の諸現象を捉え，また外界に向けて働きかけていくためには，その生体自身の中に，まず何らかの構造が存在しなければならない．たとえば，外界に光の刺激があったとしてもそれを受容する視覚構造がなければ，光の刺激は何の意味もなさない．如何に低次なものでも高次なものでもそれが可能になるためには知覚に先行した一定の構造が必要である．生化学反応的な意味合いでもそうであるが，認知システムとしての生体としてもこの原理は同じである．J. Piaget はこの認知システムにおける構造をシェマ (schema) と呼んだ．つまり，認知システムの環境世界，世界の見

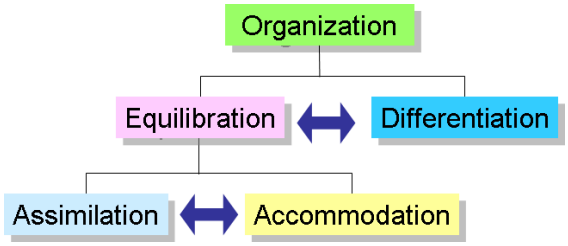


Fig.2.9: Adaptation dynamics of a schema

え方を決定するのがシェマである．

ただし，J. Piaget の呼ぶシェマと呼ばれる構造は曖昧であり，彼の著作の中でも徐々にニュアンスが変化していく．ただし変化はするものの，その底に貫かれている組織化のダイナミクスは一貫しており，シェマ自身はそのダイナミクスにより特徴付けられていると言ってよい．シェマは同化 (assimilation) と調節 (accommodation) によって特徴付けられる．同化とは，外界に働きかけ既存のシェマを外界に適用し，外界をそのシェマの内部に取り込む作用である．この取り込みにより認知システムは取り込んだシェマにより外界が表象しえることを知るが，通常そのシェマと実際に取り込んだ情報の間には小さな齟齬が残る．故により洗練されたシェマとして活動するためにはそれを修正する必要がある．このように取り込んだ情報により自らを変化させるダイナミクスが調節である．手持ちのシェマだけで外界のすべてが捉えられる (記述できる) わけではないので，それゆえ外界の新たな事象に出会ったときシェマは自らの修正をせまられる．これを調節と呼ぶ．また，シェマシステムを複数のシェマから構成される分散的な認知システムとして捉えるならば，要素レベルの調節とシステムレベルの調節を考慮する必要がある．J. Piaget はシェマシステムのダイナミクスを議論するうえで，同化，調節を繰り返すことにより均衡化がすすむが，それでも一つのシェマに同化できないほどに変更が蓄積されれば，新たなシェマを生み出しそのシェマに同化させていくと考えていた．ここで，本稿では新たなシェマを生み出し，役割分担をさせていくプロセスを**分化** (差異化：differentiation) と呼ぶことにする[*19]．

シェマシステムの根本に存在するダイナミクスは同化と調節である．この二つの基本的な適応のダイナミクスを通してシェマシステムは自然に漂流 (natural drift) す

[*19] この情報の流れの中で多様なシェマが形成されていく様子は，散逸構造におけるベナール対流のような動的な入出力関係の中で安定的にパターン形成的が起こっている状況を髣髴とさせる．

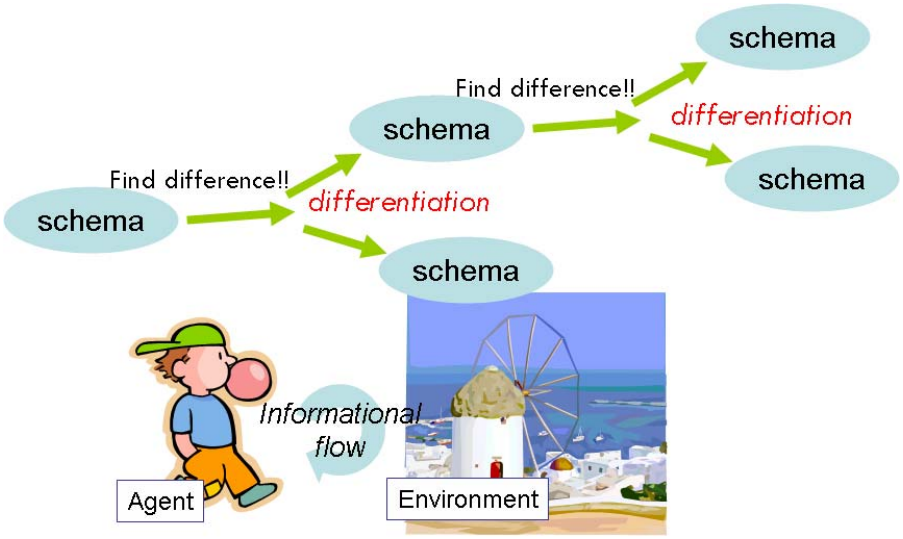


Fig.2.10: Differentiation process of schemata

るが，それはシェマと言う構造を通した記憶の組織化のプロセスとなる[*20]．

初めに均一なシェマの原初的な資源のみが認知システムに存在する．その資源を用いて認知システムが環境に対して働きかけていく間に，環境からの情報の流れを均衡化・分化を通して自らの中で組織化していく．その組織化していく仕組みが与えられれば，内部の認知モジュールとしてのシェマ自体は徐々に組織化されていく．これは，単純な情報のクラスタリングのようなものではない．もちろん，クラスタリングの側面もあるが，シェマはそのプロセスで同化・調節を繰り返し，その性質を環境に対して適合させていく．その結果として，様々な認識・行動が可能になっていく．単純なデータのクラスタリングではなく，これによりシェマは内部に形成された動的で自らの身体にとって意味のあるパターンとなる．

2.4.3　三項関係の基盤としての身体的相互作用に基づくシェマシステムの獲得

J. Piaget は発生的認識論者であるが，結局は発達心理学の中に含められることが多い．認知心理学が形成された認知システムの特徴を問題にする行動主義的な風潮が

[*20] オートポイエーシスのような自律適応系では答えを教えてくれる明示的な教師は存在しえないので，全ての学習はナチュラルドリフト (natural drift) として表現される．

色濃いのに対し，発達心理学には J. Piaget の影響が色濃く通時的な心的システムの形成を主体内部の変化として捉えようという姿勢の研究が多く，知能発達の構成的理解を構想する上で示唆に富む例が多い．知能が社会性を得，記号的相互作用が可能になるプロセスの理解についての示唆を発達心理学に求めよう．

C. Trevarthen らに拠れば子供・養育者・環境の関係は＜自己・他者＞もしくは＜自己・環境＞という独立した二項関係的なものとして始まるが，９ヶ月頃に＜自己・他者・環境＞という三項関係に移行する [126, 143]．この三項関係によって，他者の注意や活動を共有する中で他者の欲求や気持ちを理解していくことが出来るようになるという．この９ヶ月革命を期に様々な社会的能力を身に着けていくという．この時期は J. Piaget の発達段階では感覚運動期の第四段階である「二次的シェマの整合と，それらの新しい事態に対する適用」を行う時期である [40]．この時期にサインの利用を始め，サインに対する予期反応が，他の感覚運動型と同じように発達する．三項関係の理解が，記号の獲得と同期するのは，記号自体が三項関係を本質としていることを考えれば納得できる．

このような三項関係への移行は，時限爆弾のように突発的に９ヶ月の時点で起こるのだろうか．J. Piaget の考え方に従えば，そうではない．J. Piaget は感覚運動期を通して，その発達をシェマシステムの漸次的な構造化と関連させて説明する．故に，三項関係に移行するまでには，その基盤となるシェマシステムの形成が必要であるはずである．我々はこれより，三項関係に移行するためには，二項関係を通したそれまでのシェマシステム形成が不可欠であると考える．シェマシステムとは認識のための差異体系である．記号的相互作用においての議論から，我々は記号的相互作用には本質的に先行した解釈系としての差異体系が重要であると主張した．この二つの議論を比較すれば，一つの仮説にたどり着く．二項関係的な環境との相互作用を通して，幼児はシェマシステムを構成していき，これが 9 ヶ月革命において三項関係に移行し，社会的に意味づけされた様々な語彙を獲得していく基盤になるのではないか．

この考えを，本章始めに挙げた，工場環境から人間環境下に進出するロボットに投影してみよう．Fig. 2.12 では，環境の複雑さを物理的複雑さと人間要素を含めた複雑さとして，静的な環境から動的な物理環境を通って，人間社会へと進出するロボットを表している．これに対して Fig. 2.11 で表しているのは先天的な刺激応答だけで単項的に振舞う赤子が二項関係，三項関係へと移行していく場面を現している．この間に強いアナロジーが成立する．まず，工場環境下では直接機能設計をされたロボットがその与えられた機能だけを用いて刺激反応的に行動する．そこに，後天的な変化

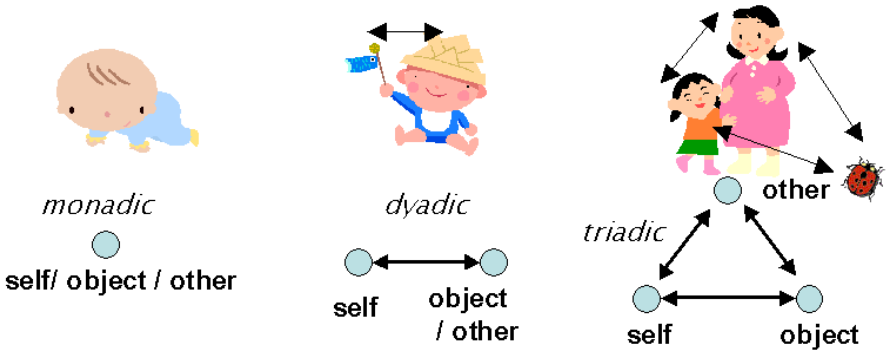


Fig.2.11: Infants become able to recognize triadic relationships.

はほぼ無い．次に動的な環境に現れると，そこではその空間固有のダイナミクスがあり，乗り越えるべき問題があるので，環境との二項的な相互作用のなかからロボットは学習し変容していかねばならない．しかし，人間社会に出ると，そこでは物理的な違いだけでは規定できない人間の主観による差異が存在する．例えば，バケツはごみを入れたらゴミ箱になる．物理的特性は殆ど変わらない．しかし，人間社会ではローカルな規約によってそれを区別することが殆どである．ロボットがバケツとの二項的な物理的相互作用を通して獲得できるのは，それに水や他の物質を入れれば溜まるということが限界である．これが，バケツなのかゴミ箱なのか，それとも他のものなのかは，あくまで三項関係に基づいて獲得せざるをえない．もちろん，この例は高度な例であるが，もっと低度なものでも基本的に同様である．これはパースの記号の Symbol の定義そのものである．パースは記号を Icon(類像記号)，Index(指示記号)，Symbol(象徴記号) に分類しているが，Symbol はそのサインが対象と如何なるアプリオリな結びつきもなく，恣意的に決まっていることがその特徴である．このような記号を学習するためには，二項関係的でなく三項関係的な相互作用の中にその情報を求めるより他ない．

ロボットが自律適応的な社会ロボットとして人間環境下で活動するためには，三項関係的な相互作用の中から学習をすすめる必要がある．しかし，その基盤としては二項関係的な相互作用の中から均衡化・分化を通して形成されたシェマシステムの存在が前提となるのである．**つまり，本研究の目的を遂げるための，第一歩は環境との身体的相互作用を通してシェマシステムを組織化し，三項関係へ結び付けていく認知発達のプロセスを構成論的に表現することであるといえる．**

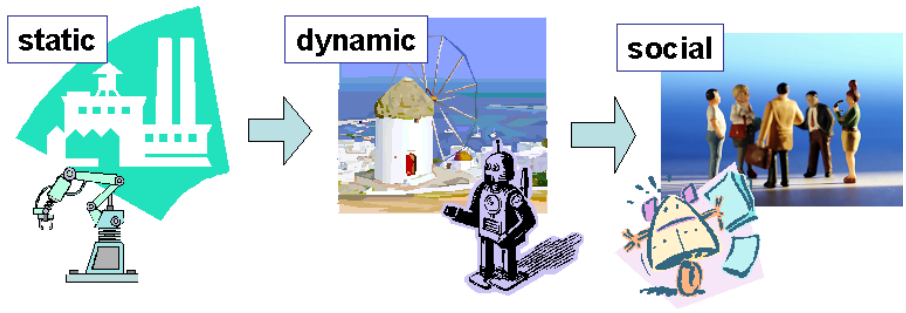


Fig.2.12: Three levels of complexity of robots' surrounding world

2.4.4　自律適応系の構成論としてのシェマモデル

本論文の目的は自律適応的に記号を獲得し，ユーザとコミュニケーションが出来るようなロボットを作れるかという問いに挑むことであった．工学的なアプローチはしばしば，現在ある道具で如何にそれを実現するかというアドホックなアプローチになりがちである．しかし，本研究ではアドホックな設計指針に沿うよりも，研究を自律適応系の構成論的研究として位置づけ，設計論と実在のシステムの理解を表裏一体として捉えることにより，領域横断的な諸学問の合流する場とする．具体的には，生命システム論，発達心理学，記号論，基礎情報学，生態学的認識論そして認知科学，人工知能，ロボティクスなどである．本章では余り触れなかったが，実際に人間の知能が形成されている場と見做されている，脳神経科学の領域ともかかわりを持つことが明らかになっていく．

本研究の特徴は心理学や記号論で論じられている諸現象を計算論的な構成論へと帰着させていく点にある．脳神経科学の領域では計算論的神経科学として一領域として築かれてきているが，我々の採る立場はより抽象的であり，特に神経修飾物質やニューロンのような物質科学的着地点を不可欠なものとは見做さない[*21]．このような立場は，脳内現象を力学系として抽象的に捉え，計算論的な認知理解を行う谷らの立場に近い [88, 89]．特に重要なのは諸学問で語られる認知発達の現象を構成できる計算論の構築であり，その数理的ダイナミクスを図式的に表現することにより，上に示した多数の領域で同じ意味なのに違う言葉を用いている現象や，違うものなのに感

[*21] もちろん第 IV 部で現れるような，同型関係が生じることは示唆に富み．興味深いことではあるが．

覚的に同じ言葉を当てはめてしまっている問題に対してダイナミクスベースの等号，不等号を導入することである．

最後に，本論文を通して提案されるシェマモデルについて説明する．具体的な定式化については次章以降において記述されるので，ここでは触れない．シェマモデルは J. Piaget のシェマシステムの発達を計算論的に表現するために提案される認知発達の構成論的モデルである．表象の獲得と，身体的運動学習という言わば，GOFAI の領域と NewAI の領域をボーダレスにかつ，適応的な学習により繋ぐことを目的としている．また，これは J. Piaget のシェマシステムの計算論的なモデルであり，このモデルを通して実際の概念分化のプロセスを考えることが出来る．実際に第 7 章では人間が複数の異なった道具を用いる際にそれらを「別の道具」として概念分化させるというプロセスがシェマモデルによって表現される．さらに，構成論と設計論の表裏一体性故に，運動学習を通して実際の自律ロボットを動かすことも出来る．学習理論としてはシェマモデルは時系列情報から各々プロトタイプを持つ複数の範疇形成を行って行く自己組織化学習の一種としてとらえられる．しかし，本論文を通して我々が到達しえる地点は J. Piaget の発達段階で言うところの感覚運動期の中期程度であり，J. Piaget の発生的認識論の構成論的モデルとしては稚拙な領域にとどまっていることについては付記しておかねばならない．しかし，その感覚運動期の中期とは二項関係から形成された内部構造を基盤にし，三項関係へ移行する転移点であり，構成論的に記号現象を考えるという点では重要である．本論文を通して幾つかのシェマモデルのクラスが導入されるが，その中で中心的な役割を果たすのが双シェマモデル (Dual-Schemata model) である．J. Piaget はシェマモデルにおいて母乳を吸うシェマや，ボールを蹴るシェマなど，表現は具体的であるが抽象的なシェマを様々に述べている．しかし，自律ロボットにおいて，内部表象としてシェマを計算論的に表現するためには，何らかの役割を与えなければならない．よって，Fig. 2.13 に示すように，環境や対象を表象するシェマとして知覚シェマ (perceptional schema) と行為の目的とそれを達成するための計画を表象する行為シェマ (intentional schema) という二種類のシェマを準備する．それぞれが，上に述べた同化・調節とそれにより形成される均衡化，ならびに分化のダイナミクスを通して組織化されていく．

大枠を Fig. 2.13 に示すように，第 II 部から第 IV 部の三部をかけて双シェマモデルの構成を説明していくことになる．第 II 部においては，行為シェマを事前設計した上での知覚シェマの均衡化・分化に焦点をあてる．第 III 部においては行為シェマの均衡化・分化に焦点を当てる．また，第 6 章において，行為シェマの分化の導入のた

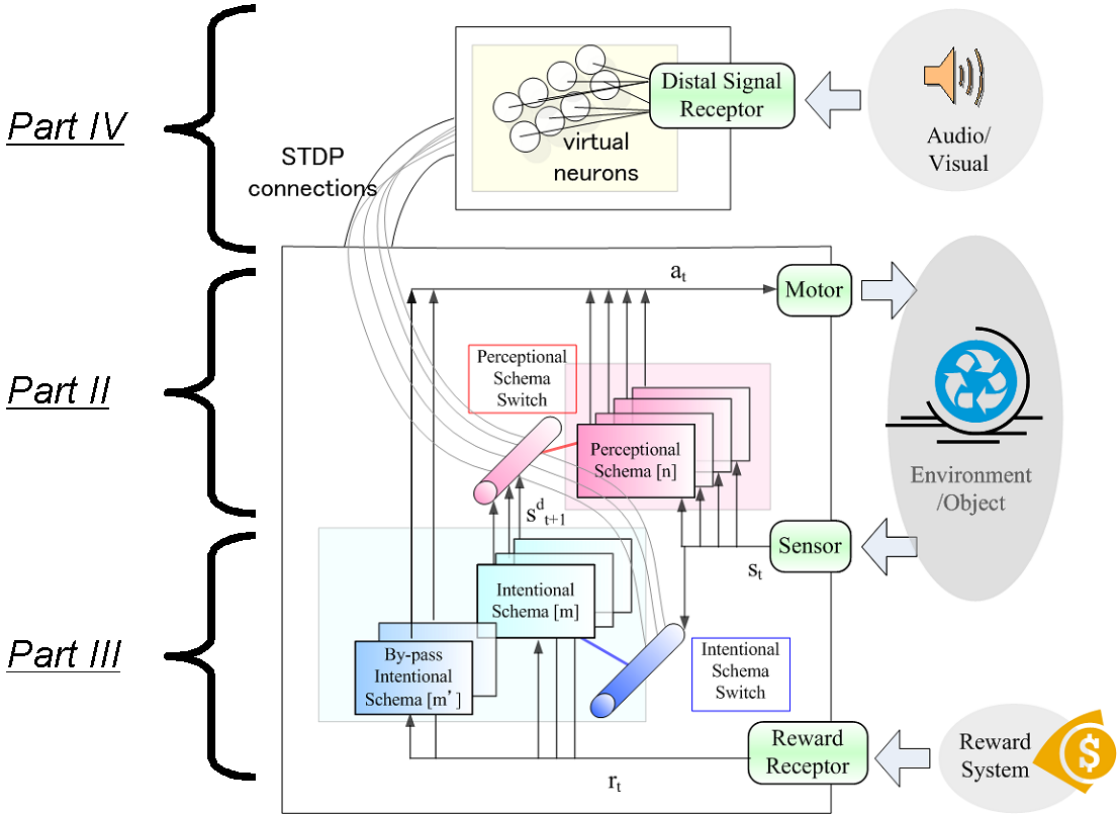


Fig.2.13: The Dual-Schemata model with STDP neuronal circuits

めに，強化学習シェマモデルというシェマモデルを提案する．また，第 7 章では知覚シェマの均衡化・分化の側面を非線形系に拡張し，正規化ガウスネットワークによるシェマモデル (NGSM) を提案する．これらのプロセスで内的表象群は豊富になっていくが，それが外的なサインとどう結びついて外的なサインを意味づけしていくのかというプロセスを脳神経科学に示唆を求め，STDP により可塑的にニューロン間の結合効率を変化させるニューロンがこの役割を担いえることを示す．これにより，パースの記号の三項関係がシェマモデルを通して実現されることとなる．

第 II 部

動的環境との身体的相互作用を通じた知覚表象生成

第 3 章

自律エージェントの為の自己組織化機械学習手法

第 II 部では自律適応系が外部からの明示的な指示なしに，如何に自らが関わりあう環境や対象の概念を自らの中に差異体系として組織化しえるかという議論を行う．まずその第一歩として，本章では，第 II 部，第 III 部を通して議論の中心的存在となる双シェマモデルを導入し，特にその内部において定義される知覚シェマの適応性について定式化を行う (Fig. 3.1).

3.1 はじめに

自律ロボットの設計において非常に重要な問題の一つに記号接地問題がある．2 章において述べたように，記号接地問題とは Harnad によって提唱された記号的な知識表現における問題である [32]．自律的な記号処理系による記号主義的な人工知能の設計論においては必ず，その記号の意味解釈を行う主体が人間の解釈に委ねられるため，エキスパートシステムのような所謂知的データベースの域を出ることが出来ない．つまり，自律ロボットを記号主義的にデザインした場合には，自律ロボット自身に適応的にその意味を状況に応じて解釈する能力が無いので，自律ロボットは実空間の中で上手く振舞うことができない．解釈は自律ロボットのセンサ・モータ系を理解した人間により記述的に与えられるしかないが，これはしばしば困難であるし，状況に応じた柔軟性を持たせることも難しい．これはデザインされた記号系の意味が自律ロボットの身体性と独立に設計されているためである [63].

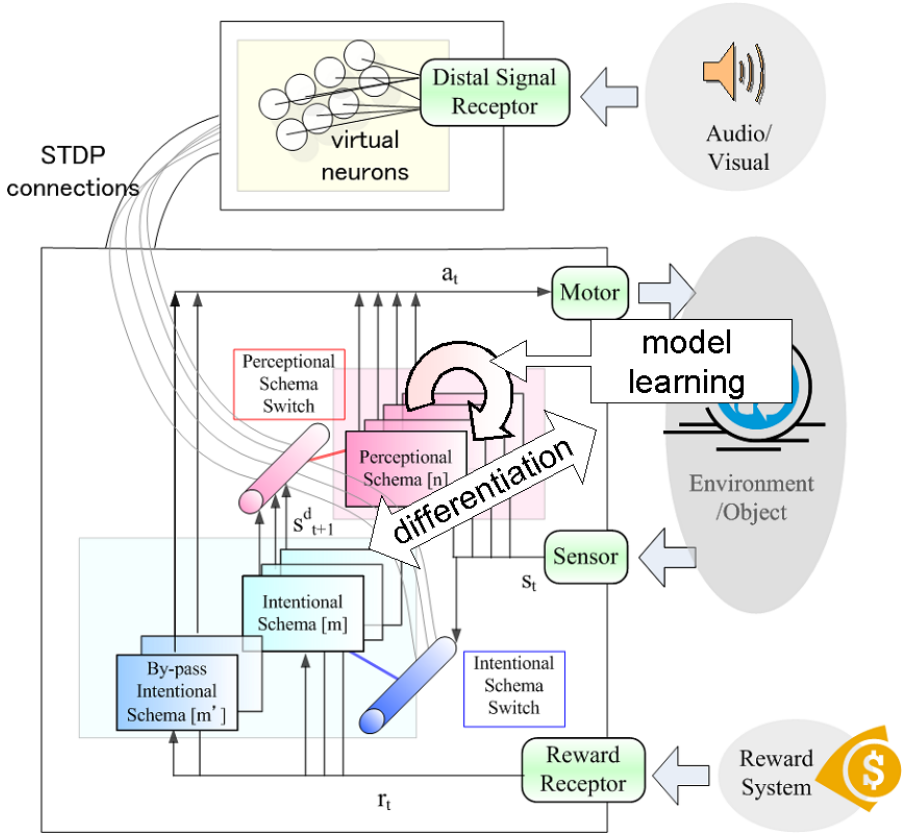


Fig.3.1: Perceptional schemata in the Dual-Schemata model with a STDP neuronal circuit

そこで現在，記号処理というトップダウンな知能デザインと，振る舞いレベルからのボトムアップな知能デザインとの接続をどうデザインするかが，次の世代の人工知能研究を切り開く重要なテーマとなっている．

しかし，２章で述べたように，記号接地問題はトップダウンにデザインされた記号の接地という立場から考えられやすいが，ロボットの身体性を無視してデザインされた記号系がロボットの環境世界 (Umwelt) の中で接地される保障はない．むしろ環境世界を自らの経験にしたがって分節化することにより，自らにとって有益な表象系を獲得していくと言う徹底したボトムアップのアプローチを我々はとる．

本研究においては，環境もしくは対象の範疇生成についての新たな構成論的モデルとして双シェマモデル (Dual-Schemata model) の導入を行い，特にそこにおける知覚シェマの適応ダイナミクスについて記述する．

当モデルを用いることにより，自律ロボットは環境との相互作用をする中で，環境

と自らの相互作用が成す因果関係を内化し (モデル学習)，与えられたタスクを達成することが出来るようになると共に，複数環境間の差異を自ら発見し，それぞれを表象する知覚シェマと呼ばれる内的表象を獲得することが出来る．

3.2 研究背景

3.2.1 記号接地問題と知識の自己組織化

古典的な人工知能研究 (GOFAI) においては，意味ネットワーク (semantic network) や Minsky のフレームに代表されるように，記号の意味を離散化された記号と記号の関係性の中で定義してきた [68]．そのため，記号の意味は記号系内部における関係性という形でしか存在せず，実世界との繋がりは解釈者としての人間を介さない限り存在しえなかった．このようなアプローチを自律ロボットの設計に用いる際には，人工知能の内部に埋め込むべき記号の意味は，客観的に定義され変容することはない．しかし，このような記号の解釈が環境において一定に保たれることが保証される世界は，網羅的に状態を定義できるようなゲーム環境や，人手により常に一定の環境が保たれ，産業用ロボットが正確に作動できる製造工場などのような人工的な環境に限定されることとなる．人間社会のような動的環境においては記号の意味とは常に不定なものである．そこでは，辞書のような記号系に含まれる言葉の意味記述のように解釈者の存在以前に存在するわけではなく，環境の動的状態変化に伴い解釈者としての主体そのものが生成しなくてはならない．

C.S. Peirce の記号論によれば，記号とは表意体 (sign)，対象 (object)，解釈項 (interpretant) の三項関係として定義される [61, 197]．これは，記号系は決して，それ自体，自律的な系ではなく，解釈者の参与なしでは存在しえない系であるということを意味している．もし，ボトムアップに記号系を生成する自律ロボットを設計するならば，この C.S. Pierce の記号観を受け入れ，記号の意味とは解釈者が何時何処の誰であるかということに関わらず客観的に存在するのではなく，人間やロボットの認知発達的なプロセスや，社会的なコミュニケーションを経て形成されるものであるとすることを認める必要がある．

また，M. Polanyi によれば，知識には二つの形があり，それらは明示的，形式的に示すことの出来る形式知と，言葉などにより形式的な知識としては表現できない暗黙知とに分けることが出来る [196]．この分類に従えば，古典的な人工知能研究は，知

識の意味を形式知の中のみに求めてきたといえる．スポーツ選手や巧みな技を持つ職人達の，匠な技は伝達しようにも言葉に出来ないとよく言われる．このように運動知能的なものは殆ど暗黙知である．このような知識は，自らが環境との相互作用の中で見出していかねばならない．また，言葉の意味の解釈はそれ自体が暗黙知である．

実世界における記号の意味や知識といったものは，常に行動主体の身体を通じて外界，環境と接続されているものであり，そこから与えられる「ゆらぎ」を排除するのではなく，むしろ自己を変容させることで自己の内部に取り込むといった作動的に閉じた活動であると言える．これのような常に解釈系を変容させながら自律適応的にシステムを形成していく活動は，生体における脳内情報処理や，細胞組織の処理系にも普遍的な活動とも言われている [128].

我々は環境や対象と相互作用する中で，現前する対象の性質を見出し，意味づけをおこなっている．これはつまり，現在の経験を離散化し自らにとって意味のある記号として解釈していると言える．我々人間のような自律適応系は経験を離散化し，過去の記憶と同化 (assimilate) し，あるいはこれを受け入れることで形成してきた記憶構造を調整 (accommodate) することで，自らの活動にとって有意味なものとしていく機構を有している．それは自らの知覚と自らの行為のみによって閉じた自己閉鎖的なプロセスである．さらに，扱われる記憶とは感覚入力を範疇化するのみではなく，行為者として活動するための運動記憶といった身体を介した運動記憶もこめられる．このように同化と調節の循環により認知システムが構成的に環境適応していくという考え方は 2 章で述べたように，発達心理学者 J. Piaget の提唱したシェマモデルに由来している．J. Piaget の描いた認知機能の発達モデルは現代のシステム科学の言葉で言えば，認知システムを解放系と見なした時の自己組織化モデルであると言われており [64]，現代の複雑系科学におけるシステムの動的な視点に非常に親和性の高い発達観である．

3.2.2　Schema 理論

シェマとは schema のフランス語読みであり，英語読みではスキーマという．J. Piaget の流れを汲む発達心理学研究においてはシェマ，認知科学や人工知能研究においてはスキーマと訳される場合が多いが，両者とも語源的には同じ存在を指している．人工知能研究ではスキーマと読むことが好まれてきたが，本論文では敢えてシェマと呼ぶ．本節では先行研究において自律ロボット制御や機械学習において考えら

れたシェマもしくはスキーマと，本研究におけるシェマ理論の違いについて明確にする．

シェマシステムは動的に組織化を続ける認知構造として定義されるが，そこにおいて構成要素としてのシェマはある認知対象を表象する記号であると考えることができる [40]．J. Piaget のシェマ理論のほかにも心理学においては，F. Bartlett や U. Neisser，R.A. Schmidt の図式や運動スキーマという，記憶構造としての schema の心理学的モデルがある [152, 56, 189]．しかし，これらと J. Piaget のシェマ理論との決定的な違いの一つは，シェマの分化に代表されるドラスティックな構造化のダイナミクスを捉えているか否かにある．J. Piaget のシェマ理論は，シェマの同化・調節プロセスを通して，シェマの外界に対する対応が詳細化していく構成的プロセスを議論しようとしている．シェマを環境の状態を指示する内的表象とみなしたとき[*1]，シェマの均衡化・分化のプロセスは，対象との相互作用により生じた情報をカテゴリ化し記号化し体系化していくプロセスに相当すると考えられる．これは，本研究で目指すボトムアップな記号系の獲得のモデルとして適切である．

人工知能研究においても schema[*2]の考え方を応用したアプローチは既に多くなされている．形式的な知識表現では M. Minsky のフレーム理論が，また移動ロボットの行動についての知識表現では R.C. Arkin のモータ・スキーマが有名である [7]．これらのような，schema を用いた多くのアプローチでは環境や自らの行為についての離散的表象を計算機内の知識ベースやロボット内部に表現するために schema は利用されている．しかし，発達的なプロセスの中で schema を獲得していくモデルは多くない．

これに対し本章で提案する，双シェマモデルでは自律移動ロボットが環境に関する分節を与えない相互作用を通じて，教師無しで環境についてのカテゴリ表象群を獲得することが出来る．また，R.C. Arkin のモータ・スキーマが提供するような，実環境下での行動の方策をも提供する．我々は以上のような理由から J. Piaget のシェマ理論を基礎とし，その計算論的モデルとして双シェマモデルを提案する．

[*1] このようなシェマの役割を心像と呼ぶこともある [121].

[*2] この場合はスキーマと読む.

3.2.3　モジュール学習

一方，近年のロボットの運動学習や計算論的な脳科学研究において，制御器を分散的なモジュールで表現するモジュール型学習機構の研究が活発である [103, 47]. これらは schema という名称は用いていないものの，その特徴は J. Piaget や R.A. Schmidt の言う schema に似た側面を持っている．元々，schema という概念は行動主体の認知システムの中に存在する行動・認識のための分散的モジュールとして考えられてきた．故に，機械学習研究，脳科学という文脈から提唱されたモジュール学習という考えは，その性質からいって正に schema 的な構造を持っているといえる．

R.A. Jacobs らによって提案された mixture of experts[39] により代表されるモジュール型学習器は複数の学習器を配し，それらが時系列的な状況変化に応じて変化するゲーティング関数に役割分担を制御されながら学習を進めていく．D.M. Wolpert らが提唱している MOSAIC は行動主体の内部に複数のモジュールを持ち予測と制御を同時に行うモデルである [103]．その中では責任信号という変数を用いることにより各モジュールの制御と学習における分担割合を決め活動する．ここではゲーティング関数のようなモジュールそのものの他のゲーティング機構は必要としていないが，責任信号を平均化する操作などが必要となる．MOSAIC ではモジュールの初期パラメータに強く依存するということと，初めに与えられた数のモジュールを使って活動を行うため，環境との相互作用の状況に即して累増的にモジュールを獲得していくことが出来ない [193]．後者については mixture of experts や久保田らの MNN[47] も同様である．第二章で述べたように，本研究での焦点は寧ろ環境との相互作用を通じた認知モジュールの累増的な獲得にあてられ.

J. Piaget のシェマ理論においては学習主体が環境との相互作用を通じて内部のシェマを次第に構造化・複雑化させていくという動的プロセスが重要であり，先行研究の多くのモジュール学習はオンラインでの，このようなシェマの分化の可能性，つまり累増性を持たない．これに対し，提案する双シェマモデルは各モジュールの初期パラメータを設計する必要なく，また新たなモジュールが生成される時にも初期パラメータを設計する必要なく，要素となるモジュール，つまりシェマをを累増的に生成させていくことが出来るようになる．その時，分化の基準をどの様にするかが議論の的になる．アプリオリな誤差の基準を決めて，それ以上の誤差ならば分化するといった方法が考えられるが，このような方法では学習初期で必然的に生じる誤差でも分化

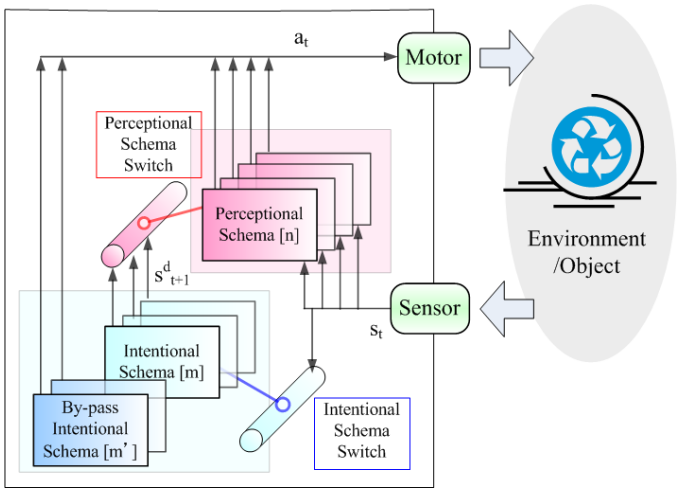


Fig.3.2: Dual-Schemata model

することなり，系が不安定になる．本章で提案するシェマモデルでは主観的誤差という無次元量を導入することで，学習器自体が分化のための誤差の閾値を適応的にスケーリングすることが出来るようになり，この問題を回避することが出来る．その詳細については 3.4 節において説明する．

3.3　双シェマモデルにおける均衡化のプロセス

本章で扱う，双シェマモデル (Dual-Schemata model) の概観を Fig. 3.2 に示す．この図に含まれる各要素について本節を通して説明していく．また，本節と次節を通じて特に知覚シェマ (perceptional schema) の適応ダイナミクスを導入していく．

双シェマモデルにおいて知覚シェマは，環境との相互作用を通じてその経験を同化し，その経験に応じて自らを調節することによって，環境との相互作用のモデルを内化していく．この経験には本質的に環境の性質のみならず自らの身体の情報も含まれているために，自らの身体が変化した時もそれは変わる．本節では同化とシェマ単体の調節により構成される均衡化のプロセスを計算論的に表現し，その適応プロセスを通して自律ロボットが静的な環境に対し適応的に活動出来るようになることを示す．

3.3.1　知覚運動のシステム

まず，自律ロボットのシステムは何らかの感覚系としてのセンサ類と行為系としてのモーター類を持っていて，それにより閉じた系であると仮定する．ここで言う，センサ類とは映像入力，音声入力，接触センサ入力，温度入力などであり，それらは何らかの特徴抽出フィルターを通して有限次元の実数値ベクトルとして内部の処理系に渡されるとする．そして，内部処理系における意思決定を経て，行為出力は数次元の実数値ベクトルとして出力され，それがモーター類において実際の動作出力として具現化する．このような古典的な仮定の下，自律ロボットにおいて，時刻 t に行為系に対し出力されるベクトルをモータ出力 a_t，感覚系から入力されるベクトルをセンサ入力 s_t と定義する．また，自律ロボットは直接的な外部からのセンサ入力の他に，短期記憶 (short-term memory) などの，予測の対象にはしないが，説明変数として利用できる補助入力を持つ場合がある．これを m_t とし，これにより拡張されたセンサ入力を $s_t^\dagger = s_t \oplus m_t$ とする．ここで，$\oplus$ は二つのベクトル空間の直和空間での自然な和を求める演算子である．ここで**行動の方策**とは時刻 t における拡張されたセンサ入力 $s_t^\dagger$ に対して，行為ベクトル a_t を決定する関数 $a_{t+1} = H(s_t)$ である．

ここで一定の環境下においては，自律ロボットと環境との相互作用にはある種の秩序が存在すると考える．秩序とは自律ロボット自身の行動と環境の応答における法則性，因果律を意味する．もちろん，その様な秩序が存在しない場合もあるが，そのような場合にはその環境においては再現性，再利用性のある知識を得ることは出来ないので，秩序が存在する場合のみを考えればよい．上の定義から，秩序とは状態 $s_t^\dagger$ と，その状況における行為 a_t，その結果 s_{t+1} の三項関係で表すことが出来る．この秩序は $u_t = (s_t^\dagger, a_t, s_{t+1})$ の三組のベクトル間の関数関係であると言える．この u_t を時刻 t における経験ベクトルと名づける．この三項関係を関数で表す方法は幾つかあるが，$F(s_t^\dagger, a_t) = s_{t+1}$ と $I(s_t^\dagger, s_{t+1}) = a_t$ の形が重要である．前者は予測モデルもしくは順モデル [41] とよばれ，将来の状態の変化を予測する関数である．また，後者は逆モデルと呼ばれ，次の時刻での状態を目標状態 s_{t+1}^d にしたいときには，$I(s_t^\dagger, s_{t+1}^d) = a_t$ を満たすような a_t を取ればいいことを教えてくれる [103]．これらの関数はその自律系にとっての環境の性質を教えてくれる．よって，この関数 F もしくは I を内的に獲得するということは，自身にとってのその環境に対応する表象を獲得したことになると考えれられる．これは Peirce がプラグマティズムの格率 [119]

として

『ある対象の概念を明晰に捉えようとするならば，その対象がどんな効果を，しかも行動と関係があるかもしれないと考えられるような効果を及ぼすと考えられるか，ということを考察してみよ．そうすればこうした効果についての概念は，その対象についての概念と一致する．』

と，言っていることに対し単純な関数描像を与えたものである．

J. Piaget は単一のシェマをもって，幼児の知覚から行動を生起するまでの行動の方策と捉えていたが，自律ロボット制御においては，意思決定と環境のモデルは分離したほうが理解しやすい．また，その方がより獲得概念の汎化性を保障できることも第 5 章の議論で明らかになる．そこで，行為シェマ IS_m，知覚シェマ PS_n の二つに分割し，これらの組み合わせによって生成されるシェマを双シェマ (dual schemata) と呼ぶことにする．行為シェマ IS_m は自身の行動の意思決定を表す内因的な側面を表す．具体的には自らの感覚入力としての s_t をどのようにコントロールしたいかの目標値 s^d_{t+1} を知覚シェマに出力する内部関数を持つ．本章では行為シェマは設計者によって与えられることとする[*3]．これに対し，知覚シェマ PS_n は環境と自身の相互作用の法則部分を表す外因的な側面を表現する内部関数を持つ．つまり，上に示したモデル F もしくは I を内化することにより，現在の状態 $s^\dagger_t$ から，自らが望む状態 s^d_{t+1} に移行するためにはどのような行為 a_t をとればよいか，もしくは適当に取った行為 a_t の結果 s_{t+1} を推定することが出来るようになる．知覚シェマ，行為シェマはそれぞれは自律ロボット内部で複数個存在し，その添え字は各シェマの通し番号を表す．また，行為シェマと知覚シェマと一つづつの組み合わせをシェマ状態と呼ぶ．双シェマの組み合わせ方については，行為シェマ，知覚シェマからそれぞれ任意に選び組み合わせることが出来る．自律ロボットは必ず知覚シェマ，行為シェマをそれぞれ一つづつ選択した状態にあり，その選択しているシェマ状態がその時点での自律ロボットの行動の方策を決定する．選択されるシェマは，シェマ活性度 $\mathbf{V}$ という変数によって決定される (詳しくは後述)．行為シェマ IS_m の内部関数は

$$s^d_{t+1} = G_m(s^\dagger_t) \tag{3.1}$$

[*3] 本論文第 III 部においてこの行為シェマをも適応的に獲得することができるようになる．

知覚シェマ PS_n の内部関数は

$$s_{t+1} = F_n(s_t^\dagger, a_t) \tag{3.2}$$
$$a_t = I_n(s_t^\dagger, s_{t+1}^d) \tag{3.3}$$

と表される．さらに，特殊な行為シェマとして，無知覚行為シェマ $IS_{m'}^-$ を定義する．その内部関数は

$$a_t = G_{m'}^-(s_t^\dagger) \tag{3.4}$$

となる．環境のモデル，つまり知覚シェマを利用する行為シェマは目標値ベースで行動を決定するが，そのような間接的な方法ではなく，環境のモデルに関係なく行為出力を決定するのが，無知覚行為シェマ (bypass intentional schema) であり，知覚シェマ獲得時の探索活動などに利用する．無知覚は，環境状態を知覚する知覚シェマを利用せずに行動決定を行うことことを意味する．以上から，行動の方策を定義できる．シェマ (IS_m,PS_n) を選択している時，行動の方策 H は

$$a_t = H(s_t^\dagger) = I_n(s_t^\dagger, G_m(s_t^\dagger)) \tag{3.5}$$

と決まる．また，IS_m が無知覚行為シェマの場合は知覚シェマを利用せず

$$a_t = H(s_t^\dagger) = G_m^-(s_t^\dagger) \tag{3.6}$$

と決まる．つまり，J. Piaget のシェマや，R.C. Arkin のスキーマが単独で行動の方策となるのに対して，双シェマモデルでは行為シェマと知覚シェマの組み合わせでもって行動の方策となる．

本章では議論を知覚シェマにのみ集中させるために，知覚シェマの組織化のみを扱い，行為シェマは設計者が埋め込む．知覚シェマについては後に述べる均衡化のプロセスにより，オンラインで獲得していくことが出来るため，前もって設計する必要はない．上の行動方策の定義でもわかるように，知覚シェマにとって本質的に重要なのは逆モデル I である．逆モデルの獲得には順モデルを求めてから逆モデルを求める順逆モデリング，間接法 [164] もあるが，本章では経験ベクトル u_t から直接関数近似により逆モデルを獲得する直接法を用いる．

3.3.2　記憶領域と主観的誤差の定義

本節ではシェマモデルにおいて中心的な役割を演じる無次元量の誤差，主観的誤差を導入する．本章では逆モデルの学習に直接法を用いるために，n 番目の知覚シェマ

の誤差計算に逆モデル誤差 E^n を用いて，それに基づいて主観的誤差 R^n を導入する．

$$E_t^n = a_t - \hat{I}_n(s_t^\dagger, s_{t+1}^d) \tag{3.7}$$

$$R_t^n = diag(\hat{\sigma}_{PS_{n(t)}})^{-1} E_t^n \tag{3.8}$$

ここで，$\hat{\sigma}_{PS_{n(t)}}$ は知覚シェマ PS_n が選ばれている間の逆モデル誤差の推定偏差ベクトルである．また，演算子 $diag(*)$ はベクトル $*$ を対角成分にとる対角行列である．

推定偏差の推定方法は特に限定しないが，これもオンラインで推定していく必要がある．後の章では順モデルの形を限定した上で，推定偏差の推定方法もより洗練されたものを紹介していくが，本章ではアドホックな方法を用いる．まず，各シェマの組み合わせ (IS_m, PS_n) 毎に，有限のキュー構造記憶領域 $\mathbf{U}_{IS_m,PS_n}$ を配置する．この記憶領域にはそれぞれのシェマを選択していた時に経験した経験ベクトル u_t が要素として格納されていく．この記憶領域は一定の長さを持っており，それ以上の要素が入った場合，最も古い要素から廃棄される．この記憶容量の有限性が忘却の役割を果たし，双シェマモデルの持つ可塑性に貢献する．

この記憶領域に格納されている経験ベクトル集合が，当該シェマの持つ内部関数として獲得される逆モデルに，どれほどの「ゆらぎ」を持って従っているかを $\hat{\sigma}_{PS_n}$ は示すことになる．本章では，$\bigcup_{IS_m} \mathbf{U}_{IS_m,PS_n}$（$PS_n$ が関わるキュー記憶構造の和集合）に格納されている経験ベクトルについての逆モデル誤差の各次元での値の絶対値を平均することで $\hat{\sigma}_{PS_n}$ を推定する．

3.3.3 均衡化プロセス

以上を踏まえた上で，均衡化のプロセスを説明する．均衡化は，同化とシェマ個体レベルでの調節の作用の循環により構成される適応過程である．まず，仮に時刻 t において IS_m，PS_n が選択されているとする．自律ロボットは入力 $s_t^\dagger$ に対し出力 $a_t = H(s_t^\dagger)$ を選択し，その結果として s_{t+1} を獲得した時点で，前述の経験ベクトル $(s_t^\dagger, a_t, s_{t+1})$ を獲得する．この時，主観的誤差 R_t^n が同化の基準を決める閾値，**同化係数** k に対して $R_t^n \leq k$ ならば，時刻 t における経験ベクトル u_t はその時刻に選択されているシェマ状態の記憶領域 $\mathbf{U}_{IS_m,PS_n}$ に取り込まれる．この作用が同化である．このとき，記憶領域 $\mathbf{U}_{IS_m,PS_n}$ はキュー構造をしているため，新しい経験ベクトルを取り込むと最も古いものが廃棄される．また，$R_t^n > k$ の場合には廃棄され，どの記憶領域にも組み込まれない．シェマ状態として選択された知覚シェマ PS_n に対応した環境では，十分高い確率で $R_t^n \leq k$ が成り立つように k を設定しておくので，

成り立たない場合は一時的にノイズが入ったもの見做し，記憶領域へ取り込まない．所謂，外れ値として扱う．ただし，そのような主観的誤差の大きい経験が暫く続けば，それは最早ノイズとは解釈されず，環境の相転移的変化だとシェマにより自律的に解釈される．この場合には知覚シェマ選択器はそのシェマを選択するのをやめ，他のシェマに同化をさせ始めるか，もしくは新たなシェマを生成する分化を行う．このプロセスについては後述する．

こうして経験ベクトルを同化して記憶領域にその環境との相互作用の経験情報を溜めると，次に知覚シェマは PS_n に関わる記憶領域 $\bigcup_{IS_m} \mathbf{U}_{IS_m,PS_n}$ に蓄えられた経験ベクトルを最小誤差で近似できるように内部関数 I_n を学習させる．この作用が調節である．具体的には逐次的に回帰を行わせることになる．この同化と調節の途絶えることのない循環により，知覚シェマは外界と主体との主観的相互作用に内包される規則的な構造，つまりダイナミクスを内化し続けることが出来る．本章では調節に逐次的な勾配法を用いる．

3.3.4　実験 1:運動球との相互作用を通した均衡化

前節で紹介した知覚シェマの均衡化プロセスにより自律ロボットが静的環境との相互作用を通してダイナミクスを内化することで，所望の行動を発現できるようになることを示す．上記のモデルを実装した自律ロボットが，プロジェクターにより壁面に映し出された青い運動球を追いかけることが出来るようになる実験を示す．まず，実験においては TRACLabs A division of Metrica 社製のカメラ制御装置 Biclops に市販の Logicool 社製の USB カメラ二基を取り付けたものを使用した (Fig. 3.3)．これを仮に顔ロボット (facial robot) と呼ぶ．顔ロボットは行為系として，Pan(横)，Tilt(縦) の二方向について合計二次元の自由度を持っている．感覚系としては二基のカメラがある．カメラの RGB 値による信号入力を各座標において，HLS 値に変換し，次に色相がいかに青に近いかで 0 ～ 1 の値に変換したものから，青球の中心座標のベクトルを出すために，画面の右上端を (1,1)，左下端を (0,0) とした x，y 座標について，その重心座標を求め，それを $v(s) = (v_x(s), v_y(s))$ として得る．ここで s は秒を単位とする実時間，t は顔ロボットの一回の動作を単位とした離散時間をあらわす．また，$s(t)$ で離散時間 t における実時間を表す．これを $10[Hz]$ のフレームレートで行った．また，青球が時刻 s においてロボットの視野内に存在しない場合は，その 1 フレーム前の $v(s-0.1)$ について $v_x(s-0.1) \geq 0.5$ ならば $v_x(s) = 1$,

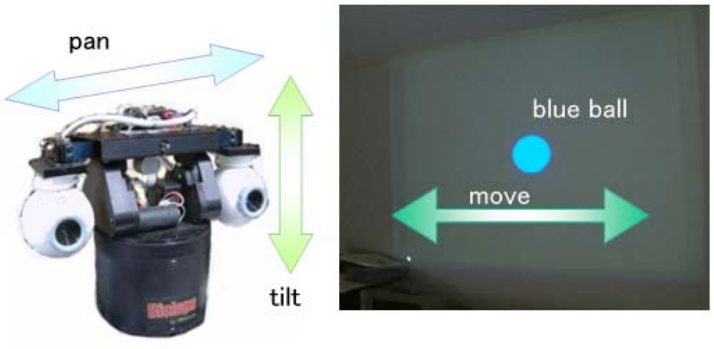


Fig.3.3: A facial robot and a moving blue ball projected on a wall

$v_x(s-0.1) < 0.5$ ならば $v_x(s) = 0$，とすることにより仮の値とした．また，$v_y(s)$ についても同様にする．また，t における行為ベクトル a_t として $a_t = (a_{x,t}, a_{y,t})$ の２次元ベクトルをとる．行為ベクトルは

$$\Delta tilt_t = GAIN \times (a_{x,t} - 0.5) \tag{3.9}$$
$$\Delta pan_t = GAIN \times (a_{y,t} - 0.5) \tag{3.10}$$

の式から実際の pan，tilt の角度の変化量に変換され，

$$\begin{cases} tilt_{t+1} = tilt_t + \Delta tilt_t \\ pan_{t+1} = pan_t + \Delta pan_t \end{cases}$$

の式を用いて顔ロボットの次の首の角度を決定する．時刻 t に次の $tilt_{t+1}$，pan_{t+1} への移動コマンドが送信され，一定速度で移動が開始される．pan と $tilt$ の絶対値の上限は 30 度とし，それぞれ 30 度を超えるときは，それ以内に収まる行為ベクトルを選択し直させた．これらのベクトルを利用し，時刻 t におけるセンサ入力 s_t として $v(s(t))$ の二次元ベクトルと，補助入力 m_t として，$v(s(t)-0.1)$，$v(s(t-1))$，a_{t-1} の３つの 2 次元ベクトルからなるベクトルを，よって $s_t^\dagger$ としては８次元ベクトルを採用する．センサ入力に前の時刻での行為ベクトル a_{t-1} を含めているため，自己回帰成分を含むことになる．逆モデルは上の $8(=4\times2)$ 次元ベクトルを基底とした線形関数により近似した．学習は逐次的に勾配法を使って線形関数の係数を更新した．

次に，ＰＣ上で青球を動かすプログラムを設計し，プロジェクターを使って壁面に投影することにより，顔ロボットが追うべき対象とした．運動状態としては，「静止」「横運動」「縦運動」「円運動」の４つの状態を準備した．この実験において，行為

Table3.1: Prepared intentional schemata for the facial robot

Intentional schema	Inner functions
IS_1: Random mode	$a_t = \hat{G_1^-}(s_t) \equiv 1\delta_t$
IS_2: Browse mode	$s^d_{t+1} = \hat{G}_2(s_t) \equiv 1\delta_t$
IS_3: Chase model	$s^d_{t+1} = \hat{G}_3(s_t) \equiv \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

シェマを Table 3.1 のように定める．Random mode のみ無知覚行為シェマであり，知覚シェマに関係なくランダムな行動をトリガする．このような行動は外部ダイナミクスを探索する上で重要である．Browse mode は目標値をランダムに出力する行為シェマである．Chase mode は青球を追跡するための行為シェマであり，カメラ画像の中で中心位置に次タイムステップで青球を持ってくることを常に目標値として出力しつづける．内因的な意思決定の機構の議論は難しく，行為シェマの選択については Table 3.1 のそれぞれを $1 \to 2 \to 3(\to 1)$ の順番で循環的に一定時間づつ遷移させつづけた．

3.3.5　実験結果

この結果，Fig. 3.4 に示すように顔ロボットは青球の 4 種類の各運動状態全てについて，それぞれ追跡することが可能になった．Fig. 3.4 において破線は運動する青球に対応して理想値だと考えられる首の角度を表し，実線は実際の首の角度を表す．実ロボットのため，サーボ系等の制約を受けいくらかの誤差が残るが，十分な追跡行動が可能になった．これは，一定の環境下において双シェマモデルにおける均衡化のプロセスが上手く機能していることを示している．

しかし，このようなオンラインでの学習は所謂，モデル学習，システム同定の類のものであり，新規性は特に無い．本モデルの最も重要な点である自己組織化型学習器としてのシェマ均衡化・分化により構成されるシェマ系の構成プロセスである．

3.4　双シェマモデルにおける分化のプロセス

本章では，知覚シェマにおける，分化のプロセスについて説明する．

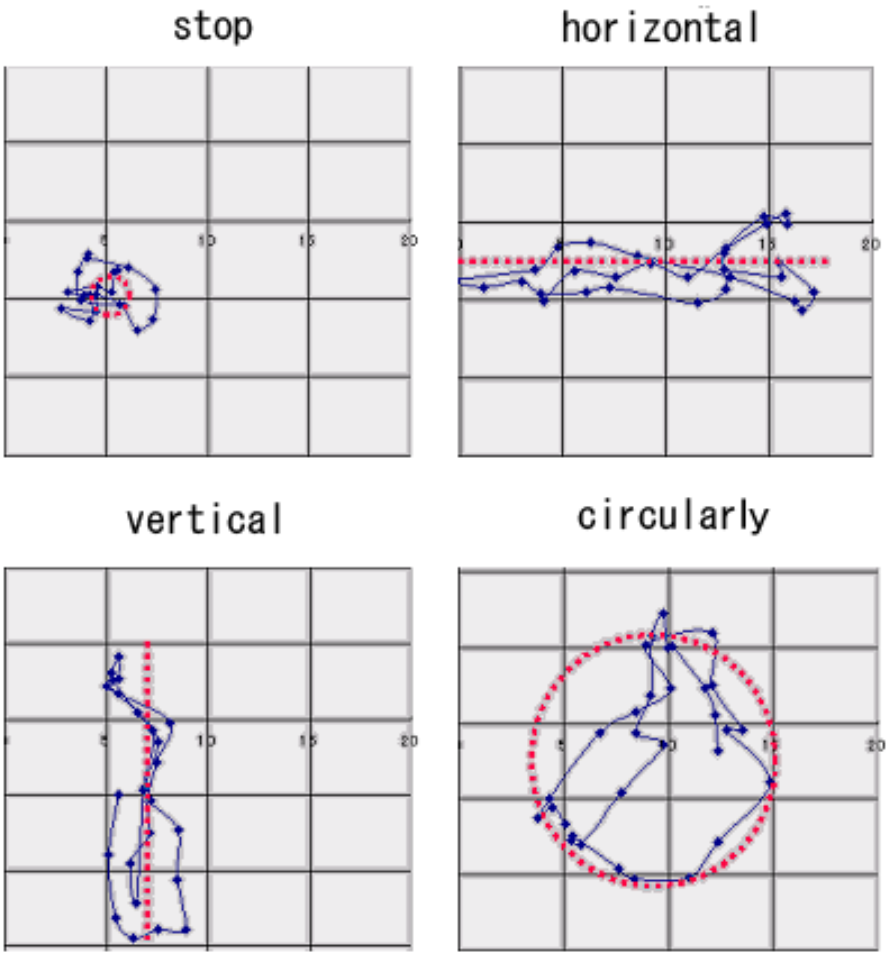


Fig.3.4: Trajectories of pan and tilt angles of the facial robot

3.4.1　分化プロセス

どのような知覚も先験的な概念無しには為しえず，既存シェマの駆動下で知覚は生起する．このようなシェマの役割は，新たな経験を既存の構造に同化する目的だけではなく，主観的な意味で，既存の概念に該当しない新規な経験であることを認識する目的のためにも重要である．ここで起こる知覚とは，センサ入力が入るという意味ではなく，時系列的な経験をどのシェマに同化するかと言う意味での，表象・認識レベルでの知覚である．ここで新たな経験であることを認識するためには，主観的な観測のみから外界の状態変化に気付き，自律的に新たな概念を作ることができなくては

ならない．シェマモデルにおいて，その自律的判断を含む部分が分化のプロセスである．我々が接する現象についての概念を分化させる仕方には，身体性や経験の仕方が大きく影響を及ぼしている．隣接するカテゴリ間の境界は，決して外界に客観的かつアプリオリに存在する境界線を内化することではない．複数の，もしく連続的な対象があった場合に，どこにカテゴリ境界が生じるかは，主体と環境間の身体的なインタフェース，つまり感覚行為系としての身体性と，その主体の経験してきた文脈に大きく依存する．この依存性は 2 章で述べたように C.S. Pierce の記号論に従い，記号化主体と解釈者の存在を肯定的に捉える以上，必然の帰結である．第 4 章ではその依存性について考察するが，その前に，本章では均衡化を通して得た知覚シェマを前提として生じる，知覚シェマの分化のプロセスについて，その計算論的モデルの提案を行う．

今，自律ロボットが新しい経験 u_t を獲得した時，この u_t が現在選択している知覚シェマ PS_n から「ずれ」ていると評価する場合に，原因として主に 3 つのことが考えられる．それは，

1. 現在の環境について学習がまだ，十分出来ていない．つまり PS_n が未だ均衡化不足である．
2. なにかしらのノイズが瞬間的に入ってきたため一時的に誤差が増加した．
3. 知覚シェマ PS_n が均衡化していた環境とは定性的に別の環境になった．

の 3 つである．ここで，上の「ずれ」ているという表現には二通りの意味が考えられる．一つは絶対値としての誤差 E_t であり，二つ目はシェマが過去に同化してきたものからのずれ，主観的誤差 R_t である．まず (1) だが，現在の環境において学習が完了していない場合，R_t^n の分母 $\hat{\sigma}$ も一般的に同じ次元で大きくなるので，分子 E_t^n の増加は相殺される．よって「ずれ」を主観的誤差 R_t^n を通して見るならば，このような原因は排除される．もし，誤差として自らの内部関数からの誤差 E_t だけを見るのであれば，それは概念のカテゴリ化に自らの経験の深さが反映されることにならない．逆に自らの過去の経験からの相対的な誤差としての R_t^n で測ることによって，客観的な基準からと言うよりかは，自らの経験からの「ずれ」を認識することが出来る．つぎに (2) だが，もし，突発的なノイズなのであれば，そのようなもののために新たな知覚シェマを用意することは有意義であるとは考えにくい．それが突発的なノイズなのか，概念分化すべき新たな定常的環境への移行を要請するものなのかを判別するためには，しばらくの間観察する必要がある．そのため，主観的誤差そのものを分化

の判断基準にはせずに，それにより駆動される変数，シェマ活性度 $\mathbf{V}_{PS_n} \in [0,1]$ を準備する．これは R_t^n の時系列により決定される量でなければならない．また，R_t^n が大きいときにはシェマ活性度は減少し，小さいときは増加する．このようなダイナミクスを表現するために，本章ではアドホックに以下のような更新則を設定する．

$$\mathbf{V}_{PS_n}(t+1) = f(\mathbf{V}_{PS_n}(t) + \Phi(R_t^n)) \tag{3.11}$$

$$f(x) = \min(1.0, \max(0.0, x)) \tag{3.12}$$

このとき，閾値 $\mathbf{V}_\alpha$ により，PS_n が現在の環境に反さないという仮説:H^n の真理値を

$$\mu(H^n) = sgn(\mathbf{V}_{PS_n} - \mathbf{V}_\alpha) \tag{3.13}$$

とする．ここで $sgn(*)$ は正の数に対し 1 を，非正の数に対して 0 を返すステップ状の関数である．これはファジィ真理値としてのシェマ活性度を脱ファジィ化している操作とも取れる．

この $\mu(H^n)$ を用いて，知覚シェマ選択器が PS_n を選択する確率は

$$P(n) = P(\bigwedge_{l=0}^{n-1} \overline{H^l} \wedge H^n) = \mu(H^n) \prod_{l=0}^{n-1} \mu(\overline{H^l}) \tag{3.14}$$

と定める．ここで，式を簡単にするために，0 番目知覚シェマは常に活性度がゼロのシェマ，ヌルシェマ (null-schema) として定義される．また，既存の知覚シェマの総数が N で，$P(k) = 0,\ \forall k \leq N$ であったときには，新たなシェマが生成され，PS_{N+1} となる．このとき新たなシェマにおける記憶容量は分散の非常に大きいランダムなサンプルで初期化される．生成されたばかりのシェマはこのように多様性の大きい状況である必要がある．このシェマ活性度を用いた一連のシェマの選択・生成の法則は，シェマを仮説とみなした一種の仮説検定と解釈することが出来る．実際に第 7 章においては仮説検定の厳密な立場から，これらの立式を再導出する．

このようなプロセスで自らの知覚・行為と言った身体活動を通じて外界との相互作用の中で生じる差異を分化された知覚シェマという形で離散的に内化することが出来る．

3.4.2 実験 2: シェマ分化を通した運動球概念の累増的獲得

前節で紹介した分化のプロセスで，顔ロボットが青球の運動状態に対応した知覚シェマを形成出来ることを示す．実験には前節で用いたものと，同じ pan-tilt 顔ロ

ボットを用いる．パラメータ類は新たに次のように設定する．まずシェマ活性度更新に使う関数はヒューリスティックに

$$\Phi(R) = \begin{cases} -1/3 & \text{if } 3 \leq ||R||_\infty \\ -1/6 & \text{if } 2 \leq ||R||_\infty < 3 \\ 1/6 & \text{if } 1 \leq ||R||_\infty < 2 \\ 1/3 & \text{if } 0 \leq ||R||_\infty < 1 \end{cases} \tag{3.15}$$

とする．また，同化係数 $k = 1.7$，仮説検定の閾値 $\mathbf{V}_\alpha = 0.33$，そして記憶容量を一つのキューにつき 150 サンプルとした．$||*||_\infty$ は L_∞ ノルムであり，各次元の中から絶対値が最大の要素をとってくる．

今回はプロジェクターで壁面に映し出した青球を「静止」→「横運動」→「縦運動」→「円運動」の順でそれぞれ 20 分間ずつ顔ロボットに提示した．その後に再び，10 分間ずつ順に提示した．この 20 分という数字は，先行実験において，実時間で動く顔ロボットが一つの種類の運動状態を追跡出来るようになるために必要であると経験測的に見積もられた十分な長さである．前節の実験と同じく，この間に一定間隔で，行為シェマを循環的に変化させた．後半の 10 分間ずつの再提示は，再び同じ運動状態を提示した時に過去にその運動状態を経験した知覚シェマが再び活性化し，運動状態を表象として再認し，そのダイナミクスに対し適切な行動を想起することが出来るのを確認するためである．

3.4.3　実験結果

実験結果として，青球の運動状態，各知覚シェマの活性度，顔ロボットがとった知覚シェマ状態，を各時間ステップでプロットしたものを Fig. 3.5 に示す．実験においては，まず顔ロボットが一つの知覚シェマ PS_1 しか持っておらず，顔ロボットの認識が未分化な状態から均衡化を開始する．行為シェマを切り替えながら，静止した青球についての情報を同化・調節していくうちに，静止球を追跡 (つまり凝視) 出来るようになる．これは前節において静止球の追跡を単独で学習させた場合と同じである．この均衡化を通じて，知覚シェマ PS_1 は静止球に対応した内的表象へと変化したことになる．静止球についての十分な概念が形成された後に，横運動を開始させる．この時，知覚シェマ PS_1 の順モデルで予測した結果 $\hat{s}_{t+1}$ と実際の結果 s_{t+1} の間にずれが生じ，その逆関数としての逆モデルにも誤差が生じ E_t^1 が増大する．もし十分に均衡化が進んでいない状態であれば，主観的にはその誤差は自らの学習不足によるものなのか，環境の相転移的な変化によるものなのか主観的には判断できない．

しかし，今回のように学習が十分に収束している場合，静止球の追跡が出来るほど知覚シェマは十分に均衡化を終えているので，もし大きな誤差が発生すればそれは外部環境の変化であると考えられる．ここでは学習の進行に伴い $\hat{\sigma}^1$ は十分に小さくなっているので，R_t^n は，分子の誤差 E_t^1 の増大に対し鋭敏に反応する．その結果として PS_1 のシェマ活性度 $\mathbf{V}_{PS_1}$ は急激に降下し，0 に達した時点で前節で定義したルールに従い分化がトリガされる．新たに生成された知覚シェマ PS_2 はまだ，何の情報も持たず，新たな環境 (横運動球) に対して均衡化を開始する．やがて，その均衡化も収束し，顔ロボットは横運動球を追跡することが出来るようになる．しかし，縦運動が開始されると，静止から横運動に変化したときと同じ事情により，再び分化がトリガされ知覚シェマ PS_3 が生成される．このような流れで円運動についての知覚シェマ PS_4 も獲得される．結果的に顔ロボットのとる知覚シェマ状態は青球の運動状態に対応しており，これは顔ロボットが球の運動状態を表象する内的表象を獲得できたことを示している．ここで重要なのはこれはパターン認識で多くあるように，センサ入力空間における距離関係の情報からカテゴリ化を行ったのではないということである．主体の概念生成という問題を考える時，それは自らの身体を通じた経験情報の自己組織化過程であるために，解釈者であり記号化主体である行為者の意味づけ無しには概念分化は起こりえない．ロボットに自律的に離散的なカテゴリを生成させる場合，そのカテゴリの意味とはカテゴリの境界が持っていると考えられる．そこではどう境界を引くかという問題が本質的である．当モデルにおいては，自律ロボットが自らの行為を環境の中で発現させる上で，どのように環境を分節化すればよいかということを自律ロボット自身に見つけさせている．これは，外部から環境の分節の基準をアプリオリに与えているのではなく，自律ロボット自身が自らの身体を通して行った環境との相互作用を通して，その環境世界を分節化する為の基準を発見している．故に，これは自律的な概念の分化のプロセスであるといえる．

さらに，この相互作用の結果生まれたシェマの活性度は先にも述べたようにファジィ集合におけるメンバシップ関数として意味を持つ．ここで，各時刻においての外部観測対象となる状態を全て異なる元として十分大きいサポート集合 W を定義しておく．離散時間 t における対象の状態を $w(t)$ とすると，4 つの知覚シェマについて十分に均衡化が完了した青球の二度目の静止以降のシェマ活性度は，そのもともとの意味から，その知覚シェマによって現在の外界がどれほどよく記述できているかを表す真理値として考えることが出来る．つまり，これは現在の外界がその知覚シェマが表象する部分集合に対する帰属度を示す真理値であるといえる．このメンバシップ関

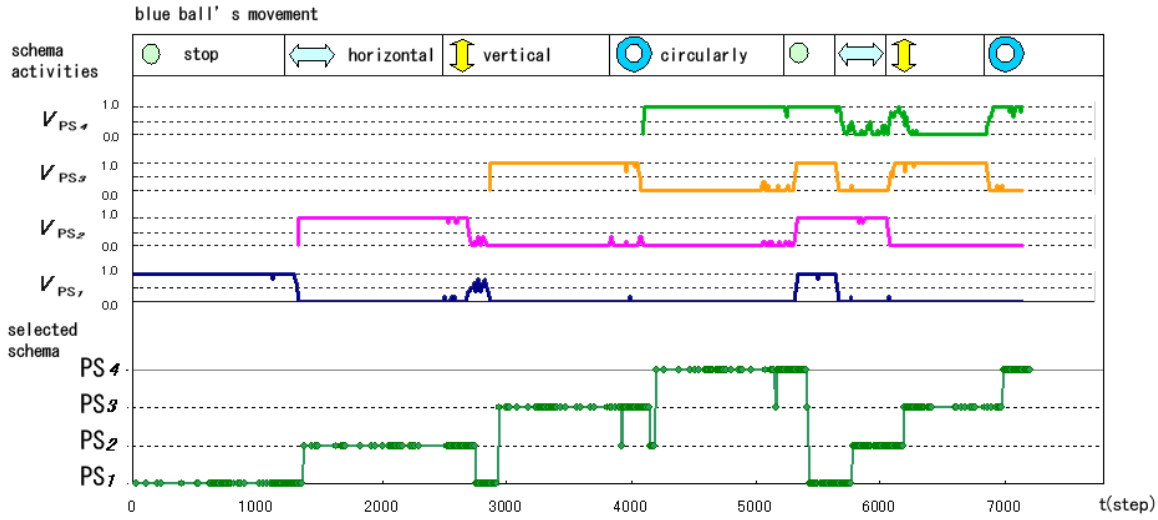


Fig.3.5: Transitions of activities of perceptional schemata and their differentiation processes

数 (Fig. 3.6) を用いることで Fig. 3.7 に示すようなファジィ論理の上での包含関係が成立することが観察された．これは，記号間の論理的関係が身体を通じた相互作用から組織化される可能性を示しており，我々の目指す徹底したボトムアップアプローチからの記号系創発に向けての第一歩といえる．また，これは一般に恣意的に決定するより他ないと考えられがちなファジィ理論におけるメンバシップ関数が自律ロボットの身体性と経験に根付いた形でボトムアップに立ち現れたものと考えることができ，ファジィ集合論の立場からも興味深い． 双シェマモデルでは，概念の分化というプロセスにより自律ロボットが記号を生成する過程を実現している．この実験結果は，一つ一つの運動状態について十分な学習時間を与えたため，分化が起きているが，もし，十分な時間を与えなかった場合には，例え同じ軌道を同じ順番で見せようと，概念分化は起こさない．この事実は非常に重要であり，カテゴリの境界が記号化主体の経験の深さに依存していることを示している．また，分化のプロセスで示した実験は一例に過ぎず，同じ 4 種類の運動状態の提示を行う場合でも，その順序，時間によって大きく異なった概念形成を示す．モデルの基本的性質から，均衡化が十分進むより前に，運動状態を変化させると全く分化が行われないのは自明であるが，その他にも，ファジィ包含関係が示唆するように，円運動に対し均衡化を行ったのちに，静止させても多くの場合分化は起きない．なぜならば，Fig. 3.6 に示したように静止は円運動

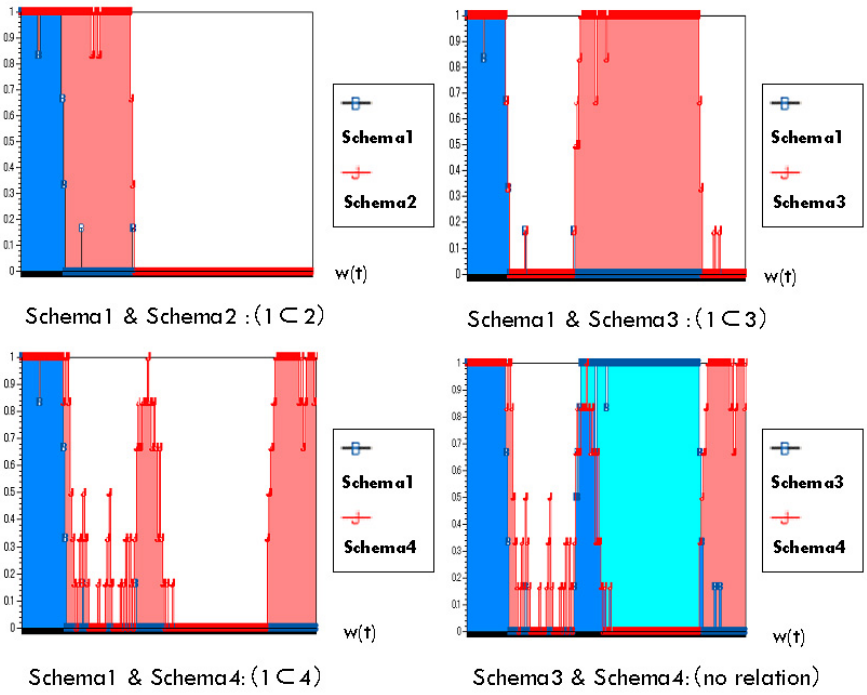


Fig.3.6: Fuzzy membership functions of the obtained perceptional schemata

の一種として捉えられるために，シェマ活性度が落ちないからだ．しかし，その後，十分長い時間，静止球に均衡化を行った後，円運動に戻すと今度は分化を行い，結果的には PS_1 が静止球，PS_2 が円運動に対応した知覚シェマとなる．これはより広いカテゴリであった円運動が静止球を同化するうちに，純粋な静止球の概念まで縮退し

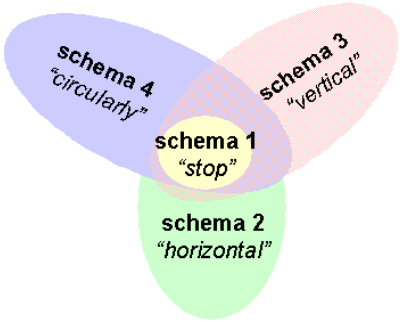


Fig.3.7: Relations of the four schemata represented by fuzzy sets

たためと考えられる．本稿においては実験結果をわかりやすくするために，外部観察者の立場からみて境界のはっきりした存在である四種類の球の運動状態を題材に概念分化を行わせたが，本来はこのように外部観察者にとって境界のはっきりした対象である必要はない．また，上にも述べたように外部観察者にとっての境界と自律ロボットにとっての境界が必ずしも一致するわけではない．しかし，そのような実験結果は説明が複雑になってしまうので，今回は説明しやすい分化の例を示した．当モデルによる，そのような分化の経験や身体性に対する依存性は一部次章で扱う．しかし，双シェマモデルにおける同化係数 k を始めとする各パラメータや，身体性と，シェマの分化プロセスの関係，および，獲得されたシェマの安定性については，本論文でも詳細には解析できておらず，今後の課題である．

3.5　考察とまとめ

本章では，自律ロボットのための内的表象の自己組織化モデルとして双シェマモデルにおける知覚シェマの適応ダイナミクスを導入した．さらに，当モデルを用いた自律ロボットが，教師信号や外界との連続的な相互作用を分節化するための特徴点などを必要とせずに，自律的な判断で対象の運動状態を切り分けられることを示した．このモデルを用いることで，自律ロボットは複数の異なる行動則を必要とする環境で，それぞれのための行動則を上書きすることなく獲得することが出来るようになる．例えば，歩行ロボットにとって平坦な路面と砂場のような場所では，まるで別の環境モデルを用いて行動しなければならない可能性がある．このような時，平坦な路面，傾斜，ぬかるんだ場所，滑りやすい場所などそれぞれの環境は連続に変形可能であり，どこからが別の環境モデルにしなければならないかは設計者の視点ではわからない場合がある．双シェマモデルを用いたアプローチではこのような場合に，歩行ロボット自身が自らの判断で足場についての知覚シェマを分化させ，それぞれの状況に応じた行動則をとることが出来るようになると考えられる．もちろん，歩行ロボットのような非線形で多自由度かつ時間遅れのあるような系への実際の適用自体は今後の課題であり，現段階は原理的な可能性を示唆する段階でしかない．

実験として，運動球の例を挙げたが，そこでは「静止」「横運動」「縦運動」「円運動」という，軌跡のパターンが違うものをとり上げたが，分化の機構はそのような対象に限定されたものではない．上に示した実験とは別に横運動のみを「通常速度」「1.5 倍速」「2 倍速」で提示するという実験を行った．ここでは，通常速度と 2 倍速

は分化するが，1.5 倍速は不安定に，二つのシェマ間で振動的に同化されるという結果が観察された．このように事前に与えられる機構は一緒であっても，環境の与え方次第で，内部で形成されるシェマの性質，個数はいかようにも変化する．この発達学習における柔軟さこそ第 2 章で言及した生態環境設計に基づいた自律適応系の可塑性であり，自律適応系設計論の第一歩として興味深い．

また，生態学的認識論における不変項の概念は行為と知覚の循環において不変なものととらえられ，物理学などで広く応用された群論 (変換群) の概念を以って解釈される場合もあるが [184]，本提案モデルにおいて知覚シェマが獲得しているのは，a_t という操作に対し，s_t がどの様に変化するかという変換構造である．シェマと不変項の概念的共通性が計算論的な立場から示唆される．

本章では双シェマモデルにおける知覚シェマのみの均衡化・分化を扱い，新たな行為シェマを獲得するフェーズについては記述しなかった．それらは第 Ⅲ 部において，均衡化，分化共に導入されるので参考にしていただきたい．

本章の実験では運動球の追跡といった特殊な例題について有効性を確認したが，双シェマモデルはより様々なタスクに適用可能なモデルである．これについても，次章以降で順次，幾つかの適用例を示していきながら，その適用可能性の範囲についても議論していきたい．

第 4 章

身体と環境の相互作用を通した記号創発：表象生成の身体依存性についての構成論

前章では，双シェマモデルにおける知覚シェマの適応ダイナミクスを導入した．それは経験をシェマに吸収させる同化とその経験に基づいてシェマが変化していく調節によって成り立つ均衡化と，現在のシェマ群がこれを同化できないときに高い多様性を持つ新たなシェマを生成する分化によって成り立っていた．さらに，これが完全に自律適応系の閉じた感覚運動の認知領域だけから生じることから，その概念生成はその自律適応系の生きてきた相互作用の歴史や，自律適応系と外界のインタフェースとなる感覚運動を規定する身体性に依存するはずである．また，身体は時間的感受性すら規定するときがある．例えば，視覚における時間解像度などがそうである．これらのことから本章では，環境と組み合わさることにより自律適応系にとっての経験を生み出す「身体」に注目し，身体性が自律適応系内部で生じる表象生成にどの様な影響を与えるかについて構成論的な接近を試みる．

4.1 はじめに

第 2 章で述べたように，自律ロボットの設計論においても，人間の認知活動についての研究においても，記号創発 (symbol emergence) は重要な研究テーマとなっている．GOFAI における表象操作を知能の基本としてとらえたアプローチは多くの場合，実世界では上手く機能しなかった．そのため，記号系を捨てて，知性の根源を身体性に求めた身体性認知科学的アプローチが提唱されている [63]．本研究では，むし

ろ身体性の重要性と記号系の存在を二者択一の対立項とはとらえず，それらを止揚することにより，表象・意図といった高次知能と反射・運動といった実空間に近い比較的低次と言われる知能とを連続的にとらえることを目指している．人間の知能について言えば，大脳新皮質の膨張により得られたとされる記号的な能力が，人間と他の動物とを分ける根本的特長であるとも根強く言われており [20]，記号系は知能研究の枠組みから外すわけには行かない．しかし，自律ロボットの身体性を無視した記号系のデザインは記号接地の可能性すら失う場合がある．元来，言語は神様がアプリオリに与えたものではなく，人間が適応的活動の中で発明してきたものである [125]．また，その発明は現代社会，日常生活においても停止することなく生じ続けている基本的なダイナミクスである．そう考えるとき如何に自律適応系が環境との相互作用を通して記号的認識を獲得していくのかと言う問題は重要である．

S. Harnad らが提唱した記号接地 [32] については現在様々な研究がなされているが，その多くは，デザインされた記号をパターン認識などの技術を通してある種のセンサ入力と関係づけるというものである [80]．これは，結局は設計者が記号の意味を与えているのであり，その記号概念は自律ロボット自身のものとは言えない．本研究では，このようなトップダウンに与えた記号を接地させるという記号接地の考え方に対し，人間が発達プロセスを通して自ら記号生成していく構成プロセスに注目するボトムアップな記号創発の考え方が注目され始めている [125, 6].

本章では前章において導入した自律ロボットに環境との相互作用を通して内的表象を累増的に生成させていくための累増的モジュール学習器である双シェマモデル (Dual-Schemata model) を用いてこの問題に接近する．特に身体性の異なる自律適応系が自らの記憶を組織化するプロセスにおいて生成する知覚シェマが，そのロボットの身体性にどう依存するのかについての実験及び考察を行う．

4.2　身体性と記号

本節では，記号論や生態学的認識論などの諸領域から身体性と記号の関係について考察する．一部は第 2 章においても言及しているが，より深める．

4.2.1 身体による世界の分節化

現代記号論の祖の一人である F. Saussure は記号の重要な性質として二段階の恣意性について述べた [124]．それは，言語が連続的で無限の現象が起こる現実世界をどう分節化しているかというカテゴリ化についての恣意性と，それをどのようなサインと結びつけているかという恣意性である．後者を扱う機械学習については古くからの人工知能研究において，様々なモデルが与えられてきたが，前者については暗黙的に設計者が与えるのが常であった．なぜならば，前者を扱うには，我々人間が何を基準にある対象とある対象が「違う」または「同じ」と認識するのかについての議論が必要であり，計算論的な記号処理だけに閉じた話ではなく，発生的認識論，脳科学，哲学などの領域へと広く開かれた議題だからである．つまり，後者は言語を共有する社会における拘束条件が優位であり，前者にこそ個体自身の身体性を含めた拘束条件が優位に働くものと考えられる．しかしこれらを科学的に取り扱うのは，内的表象がそれ自体，決して物理的に観測しえるものではないという制約のため，非常に困難である．

一方，J.J. Gibbson により開祖された生態学的認識論においては，対象の認識と自らの身体を用いた行為の可能性が非常に深く関係しているといわれている (アフォーダンス)[30]．例えばある巾を持った溝があるとき，ウマのような体の大きなものにとっては飛び越えやすいが，子犬にとっては越え難い障害となるであろう．巨大恐竜にとっては気づきさえしないものであるかもしれないし，鼠のようなさらに小さい動物にとっては飛び越える対象ではすでに無くなっていると考えられる [165]．このような行為主体内部に視点をおいて世界を見る考え方は 20 世紀前半に生物学者 J. Uexküll により生物記号論として提示されており，そこでは，ある生物から見た主観的な世界は Umwelt(環境世界) と呼ばれている．そこでは世界の分節が主体の身体に応じて変わる様が語られている．その分節は先天的なものばかりではなく，後天的な成長や自らの身体の状況に応じても変化する [94].

このようなことを踏まえると，今までの人間の持つ言葉 (記号) が行う世界の分節化の仕方を自律ロボットに一方的に覚えさせるという，記号接地の考え方はロボット自身の Umwelt を無視したものである．人間が持つ記号から始めるのではなく，自律ロボットの身体から始めてロボットにとっての記号が自身の記憶の組織化機構をとおして生成されていく必要があると考えられる．

4.2.2　言語進化と言語創発

内的表象はそれ自体，決して物理的に観測しえるものではなく，人間の認識プロセスを説明するため，または，人間の言語活動の心理的基盤として，仮想しているに過ぎない．しかし，そのようなものの存在を仮定することで，人間の認知活動・言語活動を説明しやすくなるために，その存在を多くの人々が信じている．にもかかわらず，それがどうやって内的に獲得されていくのかはその観測の困難性から，未解決な問題である．近年，言語学研究において，長きに渡り禁忌とされてきた言語進化の研究がなされ始めたが，その多くは社会または母親から子供への伝承時における言語の変化など，すでに分節化された知覚対象のラベル付けの共有についてのモデルが多い [80, 162]．これは記号処理的なプロセスでの，多主体間での言語共有を行っていることに近い．この意味で，言語進化は主に既に存在した言語が如何にして複雑化していったかという問題を主に扱っている．しかし，言語・記号が常に人間の身体を用いた認知活動と切り離せない存在であることに目を向けると，これ以外にも異なる身体性や経験をする主体がどのように異なる世界の分節化を生成するかというより意味論に寄った議論も必要である．記号創発はまさに記号が無い状況から始めに如何にして記号が立ち現れてくるかという問題である．現在の言語進化の議論を補完するためにも，身体を用いた記号創発，ひいては言語創発の研究が期待されている．

4.2.3　内的表象の必要性と記号生成の非線形仮説

我々は記号の必要性は二種類あると考えている．一つは状況を離散的に認識するためであり，もう一つはコミュニケーションのためである．ここでは仮に前者に相当する記号を内的表象と呼び，後者を外的表象と呼ぶことにしよう．本稿では特にこの内的表象の必要性に焦点を当てる．記号過程 (semiosis) は解釈者内部で組織化された事前的な仮説群としての差異体系なしには存在しえず，内的表象は外的表象の存在の基盤として捉えられる．ゆえに本来はその二つの相互作用も重要な論点ではあるが，外的表象の考慮はその議論に他者との協調といった社会的要因を含める必要があり，同時に扱うことは困難であるため本章では自己閉鎖的な系内部での内的表象の生成のみに焦点を当てる．記号的認識の存在理由は他者とのコミュニケーションにその存在理由が求められがちであるが，本章の議論で内的表象のみであっても，その獲得が個

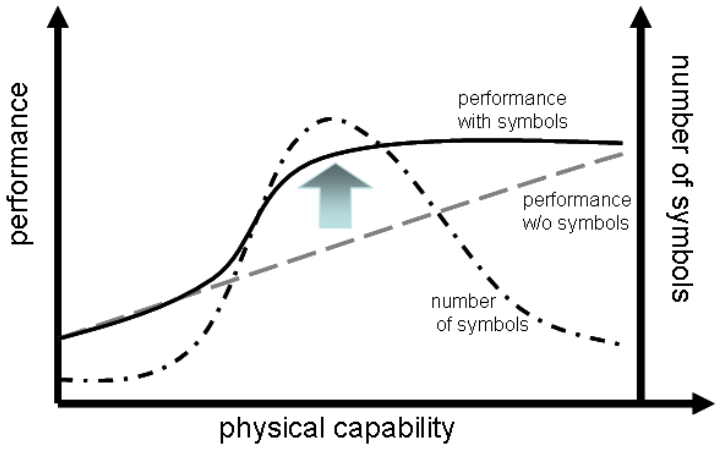


Fig.4.1: Conceptual graph of nonlinear hypothesis about symbol generation

体のパフォーマンスを向上させる場合があることが明らかになる．また，社会的な共有とまでは行かないが，外的表象の神経ネットワークを基盤にした獲得については第８章において扱う．

ところで，内的表象の必要性とは何であろうか？進化的・淘汰的観点から見ると，内的表象を持ち自らの Umwelt について離散的な認識をすることが何らかのパフォーマンスの向上に繋がり生体の生存能力を高められたのではないかと考えられる．実際，人間が異なる種類の道具を用いる時にはそれらの内部モデルを適当に分離して獲得する複数内部モデルの考え方が提唱されており [103]，そのような内的表象の構成機構が中枢神経系 (CNS) に存在することも示唆されている．内的表象による環境の離散的認識が生存のためのパフォーマンスを向上させるという仮定の下，本節では身体能力と記号生成の関係について考える．

ところで，内的表象の存在を仮定しなくとも生体はそれなりに活動することが出来る．自らの環境世界があるタスクを主体に求めてきたときに，身体能力の非常に高いもの，もしくは熟練を通して高い身体能力を獲得した玄人は離散的な認識をする必要も無くそのタスクをこなせるとするならば，そこに離散的認識の存在意義は無い．また，違いを認識してもパフォーマンスが一向に上がらないほどに身体能力の低い者にとっても離散的な認識を行ったところで利益はなく，離散的認識の必要性は無いと考えられる．

例えば，体の大きい大人がちょっとした段差を歩いていく時，それが段差だと意識せず踏破出来る．しかし，体の小さな子供や，足の不自由な老人にとってはそれが急

に区別しなければならない存在として際立って見えるということがよく言われる．逆にまだ歩くことすら出来ない赤ん坊にとってはその段差は何の意味も無い．

このような離散的識別の有用性という観点から，我々は「対象・環境に対して中程度の身体能力・サイズを持つものが最も記号的・離散的認識を行う」という，身体能力と記号生成についての非線形な関係についての仮説を持っている．Fig. 4.1 はこの仮説の概念図を表している．破線が表象生成能力を持たなかった場合のパフォーマンスを示してる．この時は単純に身体能力に比例してパフォーマンスは上昇する．実線が表象生成能力を持った場合のパフォーマンスを，一点鎖線がその時生成される内的表象の数を表している．身体的パラメータが中程度の領域で生成される内的表象の数が最大化され，その領域のみで行動主体のパフォーマンスが向上している．このような離散的認識，つまり範疇の形成が一定の素性群，つまり客観的な基準により決定されるのではなく個人間でゆらぐということは Rosch の族類似の考え方から指摘されているが [189]，その基準がどう主観的に生成されるかは難しい問いである．アフォーダンスにおける研究では古くから，人間の対象認識と身体の関係が議論されてきたが，範疇形成と身体のかかわりについて，解析的に妥当な認知実験系を組むのは非常に困難であり，研究事例も余り見られない．我々はこのような問題に対してこそ構成論的な接近が有効である考えている，本稿ではこのような疑問に対し，上の非線形仮説とともに構成的な接近を試みることにより，認知心理学研究への示唆を与えることをも目指す．

4.2.4　モジュール機構と言語

現在，上に述べたような心理学的，概念的考察とは別に脳科学や知能ロボットの領域から運動と言語の関係についての議論がなされている．D.M. Wolpert らは主に人間の運動をつかさどる小脳のモデルとして順逆モデル (forward-inverse model) が並列に複数個ならんだ MOSAIC という機構を提唱しており [103]，鮫島らはこの MOSAIC に強化学習を適用した MMRL を用いて他者の行動を分節化し自らの学習に利用する模倣学習のモデルを提案している [203]．模倣学習は言語の起源として注目されており，他にも稲邑らが，模倣と記号獲得についてのミメシスモデルを提案している [179]．通常言語は大脳皮質がその役割を担っていると考えられがちであるが，近年，小脳の役割が注目されており，身体を通した運動と言語の関係が研究されている [120].

このように運動制御のモジュール機構と記号的認識やコミュニケーションを結びつける考え方は一般的になりつつある．自律ロボットのためのモジュール型学習機構としては MOSAIC 以外にも，久保田らの MNN[47] や双シェマモデル，高橋らのモジュール型強化学習 [86, 138] などがある．また，明示的にモジュール化はせずにニューラルネットワーク内部にモジュールのゲーティング機能に相当する競合状態を作りだし，モジュール機構をニューラルネット内部のアトラクタとして分散的に表現する手法として谷らの RNNPB がある [88]．尾形らは，RNNPB を用いて実際に自律ロボットと人間との記号的相互作用を実現している [185]．しかし，その多くは前もって設計者がモジュールの数を与えてしまったり，モジュールの分担領域を決めるパラメータを制御するなどしており，自律ロボットが自らの経験のみを参照し累増的にモジュールを増やすものは少ない．その中で，RNNPB や双シェマモデルは累増的な内的表象の自己組織化を可能にするモデルであり，注目に値する．

4.3 双シェマモデル (Dual-Schemata model)

双シェマモデルは J. Piaget が人間の発達プロセスを説明するために提唱したシェマ理論 [40] を元にした累増型モジュール学習器である．これにより自律ロボットは環境との相互作用を通して環境・対象のモデルに相当するシェマと呼ばれる内部表象を獲得することが出来る．その内部における知覚シェマの適応ダイナミクスについては第 3 章で記述したため，詳細な説明は省略する．他のモジュール学習器との比較において特徴的な点は対象が変化した場合に，あらかじめ設計者によって決められた誤差の閾値では無く，自らが経験した過去の情報が含む多様性，つまり偏差との比較で決定される主観的誤差 (subjective error) と呼ばれる無次元パラメータを参照して，あくまでシェマに同化された長期的な過去の経験と，同化する短期的な経験の相対的な関係を基準にシェマの選択，分化を行うところにある．故に分化のための誤差の閾値自体は状態空間のスケールに依存せず適応的に変化する．さらに，どれだけの誤差ならば分化するかという閾値自体が，自律ロボットの身体性や環境との相互作用の歴史により変化・決定されると見做せるので，このような特徴こそが前章において述べたような，自律ロボット固有の Umwelt から内的表象を生成するという言語創発のためのモデルとして適当であると考えられる．

主観的誤差を計算する際には過去の一定期間にシェマが同化した経験サンプルがシェマの持つ内部関数に対して持つ分散を推定する必要があった．前章では，アド

ホックに過去の経験を記憶容量に溜め込んでおくことによりこれを計算した．しかし，これはモデルを複雑化させるために，双シェマモデルを複雑化させないためにはオンラインで逐次的に推定偏差を求める方法が必要である．本章で用いるモデルでは，一度，知覚シェマの内部関数を線形関数に限定することによって定式化する．これを簡易双シェマモデル (LDSM: Light Dual-Schemata Model) と呼ぶ．均衡化・分化という基本的ダイナミクスや Fig. 3.2 に示したフレームワーク自体は変化しないため，本質的には同様な働きを期待できる．以下で LDSM について導入する．

4.3.1　自律ロボットの定義

前章と同じく，自律ロボットは何かしらかのセンサ系とモータ系を持つものと仮定する．それぞれの入出力は毎時有限次元の実数ベクトルで与えられ，それぞれ s_t，a_t とする．また，自律ロボットは直接的な外部からのセンサ入力のほかに，短期記憶 (short-term memory) などの参照可能な補助入力を持ち，これを m_t とし，これにより拡張された入力を $s_t^\dagger = s_t \oplus m_t$ とする．また，自律ロボットはこれらの値を自らの持つ固有のサンプリングレートで取得，指定するものとし，この一周期を $\tau[s]$ とする．s_t のような添え字の t は自律ロボットにとっての離散時間を指し，実際にはこの 1step 間に $\tau[s]$ の時間が経過しているとする．

4.3.2　知覚シェマと行為シェマ

知覚シェマは内部関数として，以下のような順モデル F_n，逆モデル I_m を持つ．

$$s_{t+1} = F_n(s_t^\dagger, a_t) \tag{4.1}$$

$$a_t = I_n(s_t^\dagger, s_{t+1}) \tag{4.2}$$

自律ロボットにとっての環境のダイナミクスについての概念は，これらの関数によって表象される (三項の関係を正しく予測するかぎりにおいて)． これは，所謂システム同定 [169] と同じ作業になる．システム同定においては環境に内在する状態変数を直接観測できない問題が本質的であるために，観測できるものと環境の状態を別と見做すが，ここでは，あくまで，センサ変化とモータ出力の関係を内化するという簡単なシステム同定に限定する．全ての知覚シェマはこれら二つの関数をその内部に持ち，継続的にこれを更新し続ける．前章の知覚シェマでは逆モデルの獲得に直接法を

用いたため，内部関数としてあらゆる形の関数近似器を持つことを許したが (ニューラルネットワークや RBF などの非線形近似器)，このために逐次的な推定方法を設計できずアドホックな方法に頼った．また，直接法の問題点は川人らによって指摘されている [164]．このため，LDSM では順モデルを獲得してから逆モデルを生成する間接法，順逆モデリングを用いている．しかし，間接法では逆関数計算のために内部関数は少なくとも局所的に線形であることが求められる．本章では手法の複雑性を低減するために，その限定を受け入れる．よって，簡単のために LDSM では内部関数として用いるものは線形関数に限定することにする．非線形関数への拡張については第 7 章で NGSM という非線形系用のシェマモデルを本章の LDSM の考え方をベースにして提案する．LDSM では，F_n と I_n は以下のような線形式であらわされると仮定する．

$$s_{t+1} = F_n(s_t^\dagger, a_t) = F_n^{s\dagger} s_t^\dagger + F_n^a a_t + F_n^{const} \tag{4.3}$$

$$a_t = I_n(s_t^\dagger, s_{t+1}) = I_n^{s\dagger} s_t^\dagger + I_n^s s_{t+1} + I_n^{const} \tag{4.4}$$

これらの式の中で $F_n^{s^\dagger}$，F_n^a，$I_n^{s^\dagger}$，と I_n^s は実数行列であり，F_n^{const} と I_n^{const} は実数ベクトルである．

もう一方のシェマは行為シェマ (intentional schema) と呼ばれるが，知覚シェマが環境や行為対象の表象として働くのに対して，行為シェマは行為の表象として働く．行為シェマの立式については前章と変化が無いために，説明を省略する．知覚シェマと異なり行為シェマの内部関数は前章で導入したモデル同様 LDSM においても任意の関数をとることを許される．行為シェマの獲得は次章以降で説明する．本章では前章と同じく事前設計するものとする．

行為シェマと知覚シェマがそれぞれ選択されたときの行動方策 H の決定は前章と同様であるので省略する．

4.3.3 均衡化 (Equilibration)

本節では知覚シェマの持つダイナミクスの一つである均衡化のプロセスについて説明する．自律ロボットが時刻 t において三組のベクトル $s_{t-1}^\dagger$，a_{t-1}，s_t を獲得したとき，知覚シェマ $PS_n(t)$ はその経験を既存のシェマに同化 (assimilation) しようとする．ここで PS_n がその経験を同化するか否かは主観的誤差 (subjective error)R_t^n に

依存するが， 主観的誤差とは自らが内部に持つ多様性 $\hat{\sigma}_t^n$ と，あらたな経験と内部関数の誤差 E_t^n の相対的な関係において生じる値である．これを測るために，各知覚シェマは経験サンプルと内部関数の誤差を測る関数 E_t^n と，知覚シェマの持つ多様性を監視する関数 $\hat{\sigma}_t^n$ を持たねばならない．ここで知覚シェマが内部にもつ多様性 $\hat{\sigma}_t^n$ とは過去に自らが同化した経験の持つ多様性に他ならない．これは過去の経験が自らの内部モデルの間に生じた誤差の履歴から生じるものであり，対象・環境が変化しない場合にはこれを用いて次の経験に対し内部モデルがとる誤差の各成分絶対値についての期待値を予測することができると考えられる．つまり，シェマ内部の多様性を示す $\hat{\sigma}_t^n$ とは誤差 E_t^n を予測するための，誤差予測モデルとして定義できる．LDSM では順逆モデリングを行うために，順モデルを用いた所謂予測誤差を計測する．

$$E_t^n = s_t - F_n(s_{t-1}^{\dagger}, a_{t-1}) \tag{4.5}$$

また，推定偏差 σ_t^n の計算のためにシェマ状態毎に分散的にメモリを配置し，過去の誤差の平均値を用いていたのに対し，LDSM では各知覚シェマに誤差予測のための線形モデルを持たせる．これによりメモリを累増的に配置しなければならない複雑さを解消する．また，単純な定数とせず，線形モデルとすることで，前章での実験の random mode のようなセンサ値の変化が大きいときには比例して誤差量も大きくなるため誤差量が行為シェマの選択に依存するという問題を偏差推定モデルによって扱うことが出来るようになる．

$$\hat{\sigma}_t^n = \hat{\sigma}_n^{\dot{s}} |\dot{s}_t|_{each} + \hat{\sigma}_n^{const} \tag{4.6}$$

$$\dot{s}_t = \frac{1}{\tau}(s_{t+1} - s_t) \tag{4.7}$$

ここで $\hat{\sigma}_n^{\dot{s}}$ と $\hat{\sigma}_n^{const}$ はそれぞれ各成分非負の行列と実数ベクトルである．また $\tau[s]$ は離散時間 1 ステップに対する実時間の長さをあらわす．$\hat{\sigma}_t^n$ は順モデルが環境のダイナミクスを近似しきれていない度合いを示す．主に，$\hat{\sigma}_n^{\dot{s}}$ はモデル化誤差，$\hat{\sigma}_n^{const}$ は環境に含まれる定常的なノイズ等に相当する係数を同定する．近似は慣性項付きの勾配法によってまず線形項のみのモデルとして近似し，残差を定数項で近似しそれらを足し合わせることにより更新する．これらの関数を用いることにより，主観的誤差 $R_n(t)$ は以下のように算出される．これは新たな経験に対する誤差と過去の経験についての誤差の比で計算される無次元量である．ここで言う無次元量とはベクトルとして 1 次元という意味ではなく，物理量として単位がないという意味である．

$$R_t^n = diag(\hat{\sigma}_t^n)^{-1} E_t^n \tag{4.8}$$

$diag(*)$ は $*$ を対角成分とした対角行列である．もし $R_n(t) \leq k$ ならば知覚シェマ選択器により選択されていた知覚シェマ PS_n は時刻 t で経験を自らに同化する．同化されない経験は廃棄される．ここで k は同化のための閾値であり，同化係数 (assimilation threshold) と呼ぶ．上記の同化を繰り返す中で PS_n の内部で定義された順モデルは自らが同化した新たな経験に合うように更新し続ける．これが調節のプロセスである．これらの繰り返して行われる二組の操作は全体として均衡化 (equilibration) のプロセスと呼ばれる．この更新には慣性項付きの勾配法を用いる．慣性項が記憶の役割を果たす．逆モデル I_n は F_n が線形であるために F_n^a の擬似逆行列 (Moore-Penrose Matrix Inverse) を求めることによって算出することが出来る (7章参照)[130].

4.3.4 分化 (Differentiation)

LDSM での知覚シェマ活性度 $\mathbf{V}_n(t)$ の計算方法を示す．$\mathbf{V}_n(t)$ は主観的誤差に従って毎時以下の式によって更新される．

$$\mathbf{V}_n(t+1) = p\mathbf{V}_n(t) + (1-p)\exp(-\frac{1}{2}||R_t^n||_\infty^2) \tag{4.9}$$

$$= (1-p)\sum_{k=0}^{\infty} p^k \exp(-\frac{1}{2}||R_{t-k}^n||_\infty^2) \tag{4.10}$$

$||*||_\infty$ は L_∞ ノルムであり，各成分の内の絶対値の最大値をとる．固執係数 p はどれだけ過去の活性度に固執するかを表す．主観的誤差が大きいほど活性度は 0 に近づき，小さいほど 1 に近づく項と現在の活性度の重みつき平均により更新される．通常 $\mathbf{V}_n = 0$ は「現在の環境は PS_n に対応していない」ことを示す．また，$\mathbf{V}_n = 1$ は「現在の環境が PS_n に対応している」ことを示す．知覚シェマ選択器はこの知覚シェマ活性度を参照しながら以下の法則に従い知覚シェマを選択する．シェマの選択は前章のモデルと同じく，

$$\mu(H^n) = sgn(\mathbf{V}_n - \mathbf{V}_\alpha) \tag{4.11}$$

$$P(n) = P(\bigwedge_{l=0}^{n-1} \overline{H^l} \wedge H^n) = \mu(H^n)\prod_{l=0}^{n-1} \mu(\overline{H^l}) \tag{4.12}$$

とする．前章と同じく PS_0 は常に $\mu(H^0) = 0$ となるヌルシェマ (null-schema) である．$\mathbf{V}_\alpha$ は脱ファジィ化の基準であり，仮説検定の視点からは有意水準に類したもの

になる．仮説検定の視点とシェマモデルの関係については第 7 章に詳述する．現存するシェマすべてが $P = 0$ となった時に，新たなシェマを生成し，その σ を十分大きい値に設定する．ここで，H^m は m 番目の強化学習シェマが現在の環境に相反していないという仮説であり，$\bar{H}^m$ はその論理否定である．また，μ は真理値関数であり，H^n は PS_n が現在の環境の状況に適合しているという仮説である．このようなルールに従うことにより知覚シェマ選択器は知覚シェマを選択・作成していくことが出来る．新たなシェマは元のシェマの内部関数と十分に大きな $\hat{\sigma}$ が与えられる．$\mathbf{V}_\alpha$ は分化のための閾値となるが，活性度を動かす R が過去の同化した経験と現在の経験の間の相対的な値として決まるので,「誤差 E がどれほどならば分化するか？」という視点から見ると，分化のための閾値は過去の経験の履歴や自律ロボットの持つ身体性に依存して相対的に決定されることになる．故にこの選択のアルゴリズムは誤差最小のモジュールを選択するというものでは無い．しかし，この選択側が，誤差最小のモジュールを選択すると言う基準により選択されるモジュール型学習器よりも長期的な視点で見ると予測誤差が低減しうる (詳しくは第 7 章)．本章では $\mathbf{V}_\alpha = \exp(-\frac{1}{2}k^2)$ に設定する．

4.4　実験

本章では LDSM を前章と同じくボール追跡を学習する過程でシェマの分化を行うロボットに適用する．ただし，身体性の変化を実現するために，実空間ではなく仮想空間上のシミュレータに適用する．前章では双シェマモデルを実ロボットに適用した．このロボットは pan(横方向)，tilt(縦方向) の二自由度と二基の CMOS カメラを持っており，壁面に投影された青球を追跡するというタスクを与えられた．この実験のなかでロボットは 4 種類のボールの動き (静止，横運動，縦運動，円運動) を追うことが出来るようになった上に，これらの 4 つの動きに対応する形で自らの知覚シェマを 4 個に分化させた (第 3 章参照)．

しかし，前章の実験では，シェマの分化は人間のもつ 4 つの運動状態概念に対応するように生成されたのを観察したのみであり，自律ロボット固有の身体性から生じる Umwelt との関わりについては議論しえなかった．双シェマモデルで仮定しているような，自己閉鎖的な系が概念分化を行う時，それが何を区別して学習するかはその外部との相互作用の仕方に深く依存する．しかし，その相互作用は常に外界とのインタフェースとしての自らの身体の支配下にある．つまり，身体のサイズ・性能の違いが

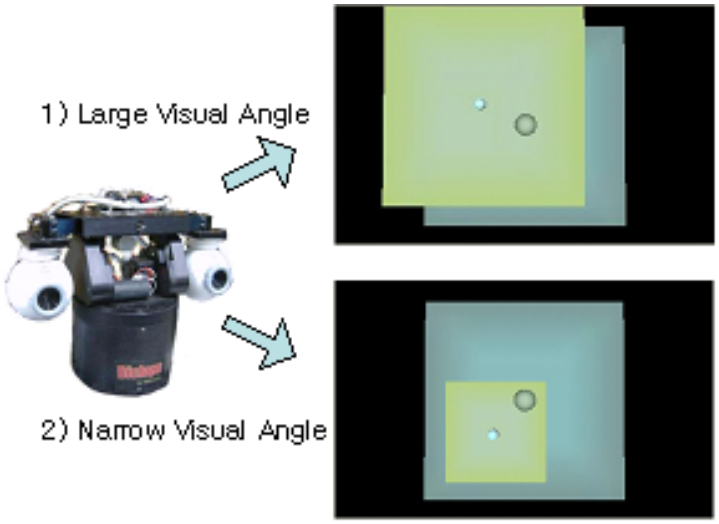


Fig.4.2: Physical environments of differently embodied robots

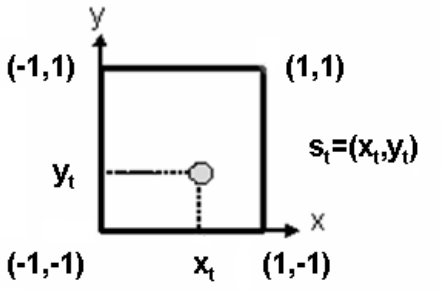


Fig.4.3: Sensory inputs for a facial robot

概念の概念分化に影響を与えるはずである．先に述べたように，我々現実の生体においても身体の違いが概念形成の違いを生み出すと考えられており，実験の結果が人間や動物の概念生成のプロセスと符合しているかが注目される．

本論文においては前章と類似した環境を計算機シミュレーション環境下に設計し，同様のタスクを異なる身体性を持ったロボットに与えることにより，身体性と知覚シェマ分化の関係を見る．

4.4.1 実験環境

シミュレーションについて説明する． Fig. 4.2 において小さいほうの球は壁面に射影されたロボットのカメラの焦点を示す．ここでその位置座標を (q^{pan}, q^{tilt}) と

する．次に薄いほうの四角形はカメラの視野を示す．その一辺は $2VA$[cm] とする．VA は視角 (visual angle) を意味する．また，実際の pan-tilt カメラは二つの自由度について有限の可動域を持つ．ゆえにそれに従い，ロボットはその焦点の動きも制限される．Fig. 4.2 において暗いほう (後方) の四角形は焦点の可動域を示す．その縦横はともに 200[cm] に設定し，その中心の座標を $(0, 0)$ とした．大きいほうの球は追跡対象の球を意味し，その座標は (B_x, B_y) とする．Fig. 4.3 はロボットが得ることの出来るセンサー入力を簡単に示している．ロボットは自らの視野内における主観的なボールの位置座標以外の情報は知ることが出来ない．もしロボットが動けば例えボール自身が静止していてもその座標が動くということは注意しなければならない．また，もし球が視野外に出た場合には，出る寸前のボール位置の方向にカメラの中心から $\sqrt{2}$ の距離にあるものとして与える．また，観測ノイズとして標準偏差 5.0×10^{-3} のガウスノイズを乗せる．次に，モータ出力を $a_t = (u_t^{pan}, u_t^{tilt})$ と定義する．また短期記憶，つまり 1step 前のセンサ・モータ情報 $m_t = (s_{t-1}, a_{t-1})$ を補助入力として与える．ロボットの焦点 (q_t^{pan}, q_t^{tilt}) は以下の式に従って移動する．

$$\frac{d}{dt}q_t^{pan} = gain * u_t^{pan}, \frac{d}{dt}q_t^{tilt} = gain * u_t^{tilt} \tag{4.13}$$

ここで $gain$ はモータ出力のゲインを表す定数である．本実験では 5[m/s] とする．行為シェマは設計者が Table 4.1 に従い与えることにする．ここで $\delta(t)$ は乱数ベクトルである．その各要素は標準偏差 1 のガウスノイズである．基本的に前章で扱ったものと同一である．本実験ではこのように与えたような行為シェマ群を用いた環境との相互作用の中で身体の変化が知覚シェマ形成にどう影響を与えるかを探る．

なお，この実験において同化係数 k は 2.3，$\mathbf{V}$ の更新に用いる p は 0.8 に固定する．同化係数は誤差が正規分布し，推定偏差 $\hat{\sigma}$ が真値に近いことを前提に考える場合，2 ∼ 4 程度に設定するとシェマは環境が変化しないときには分化せず，変化したときに分化を行うといった適度な適応性を持つ．

4.4.2　実験内容

本実験ではロボットに様々な身体的パラメータを設定することにより，ロボットの持つ身体性と記号創発の関係性を考察する．シミュレーション環境では様々な身体パラメータを持ったロボットを用意することが出来る．カメラの視野について二通りの身体パラメータを与えた例を Fig. 4.2 に例示した．まず，Table 4.2 に示すように対

Table4.1: Intentional schemata

Index	Name	Inner function
IS_0	Random mode	$a_t = G_0^- \equiv 0.5\delta_t$
IS_1	Chase mode	$s_{t+1}^d = G_1(s_t) \equiv (0,0)$
IS_2	Browse mode	$s_{t+1}^d = G_2(s_t) \equiv 0.5\delta_t$

Table4.2: Movement of target ball

Mode	Name	Dynamics
M_1	Stop	$(B_x, B_y) = (0,0)$
M_2	Horizontal	$(B_x, B_y) = (60\cos(\omega t), 0)$
M_3	Vertical	$(B_x, B_y) = (0, 60\sin(\omega t))$
M_4	Circularly	(B_x, B_y) $= (60\cos(0.5t), 60\sin(\omega t))$

象球の運動状態を定義する．なお，本実験では $\omega = 10[rad/s]$ に固定する．これらは前章において実空間で与えた物とほぼ同一のものである．これらは我々人間の目には異なるダイナミクスを持つとして捉えられる 4 つの動きである．しかし対象に関わる自律ロボットにとってのダイナミクスとは，あくまで主体の観測可能なセンサ情報とそれに働きかけられるモータ情報とのかかわりあいの中で見出されるものであるので，身体性が我々人間と異なるロボットが我々が行うのと同様な分節化を行うとは限らない．実験は 1000[s] を一つの単位とし，250[s] 毎に球の運動モードを M_1 から M_4 へと周期的に変更した．また，行為シェマについては 50[s] を一つの単位とし，20[s] の間 IS_0，次の 20[s] の間 IS_1，そして最後の 10[s] の間 IS_2 という動作を繰り返させた．この試行を 10000[s] 行い，最後の 2000[s] については均衡化のダイナミクスを止めて獲得されたシェマの性能について調べた．この時，様々な身体パラメータを持つロボットに対し，同じ環境の変化と行為シェマによる決定をさせているにもかかわらず，その結果として異なったシェマ分化が行われた場合には，それは身体性に原因していると考えることが出来る．

4.5　結果と考察

この実験において，変更することの出来るロボットの身体パラメータは多く存在するが，議論を発散させないためにカメラの視角 VA とロボットの行動周期 τ に限定して議論を行う．これらはそれぞれ空間的認知と時間的認知について自律ロボットが持つ身体的資源を決定する重要なパラメータである．追跡学習の収束過程についての議論は省略するが，均衡化自体は教師ありのモデル学習であるので，強化学習などに比べると非常に速い学習スピードを持ち，静止球の追跡については初めて chase mode をとってから数秒で追跡が可能になった．

4.5.1　シェマ分化のプロセス

毎 1000[s] の間に，対象球の運動状態毎に，その運動状態のときに自律ロボットがとった知覚シェマ状態のインデックス n についての時間平均値の一例を Fig. 4.4 に示す．もし，ある運動状態が常に n 番目の知覚シェマに受け持たれるようになった場合，この値は n になる．よってこのグラフは各運動状態が別々のシェマに受け持たれていく様子を表している．このグラフは $VA = 100$[cm] で $\tau = 0.1$[s] の時のものであるが，このグラフから，この時 4 つの運動状態が自律ロボット内でシェマの分化を通じて分散的に記憶されていくのが分かる．また，Fig. 4.5 に知覚シェマの内部関数である F の係数行列 $F^{s\dagger}$ の 1 行 1 列目の成分を 1000[s] までプロットしたものを示す．安定的な推移はしないが内部関数の係数パラメータの視点からも分化の様子が分かる．これらは前章の知覚シェマの適応ダイナミクスと同様のことが記憶容量を明示的に持たない LDSM においても実現できていることを示している．

4.5.2　身体性と記号生成

双シェマモデルにおける分化の仕組みに基づいた表象生成の身体依存性を調べるために，前章で述べたシミュレーションを $\tau = \{1/32, 1/16, 1/8, 1/4, 1/2\}$ 及び $VA = \{40, 60, 80, 100, 200, 400\}$ の 5×6 の 30 組の組み合わせについて行った (Fig. 4.6)．そのときそれぞれどれだけの知覚シェマが分化されたかについて考察する．しかし，知覚シェマの個数と一言で言っても，その中には生成されたものの $\hat{\sigma}$ が偶発的に小さくなりすぎて選択されなくなるものなど，作られたものの使われないで放置

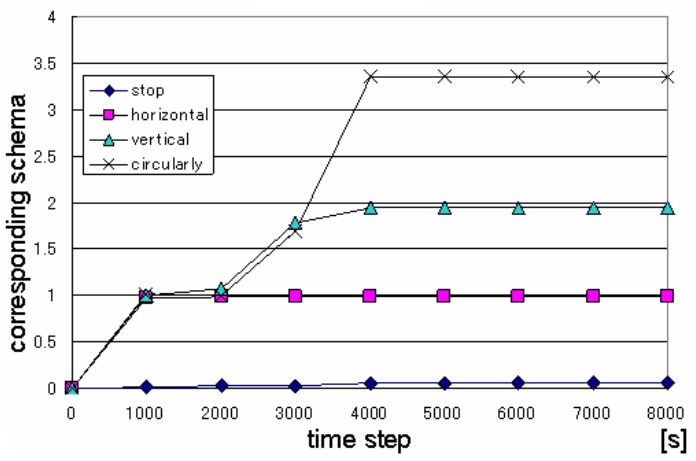


Fig.4.4: Differentiation process of perceptional schemata

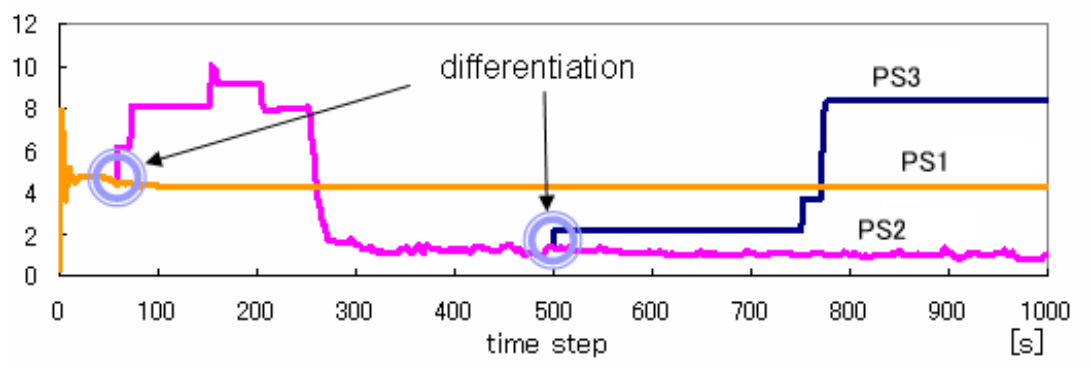


Fig.4.5: Differentiation process of perceptional schemata in terms of their inner function parameter

されるものも度々現れる．このようなものは数えずに，分化したシェマの中で利用されるものだけを数えることにする．これを測るために，生成されたシェマの個数に対応するパラメータを情報エントロピ [159] を用いて計算する．均衡化終了後のシェマ利用期間となる 8000[s]〜10000[s] の間で知覚シェマ選択器により PS_n が選択された割合を $\mathbf{P}(PS_n)$ とする．これを用いて，知覚シェマの選択についての情報エントロピ S と擬似シェマ数 (pseud-number of schemata)$\mathbf{M}$ を以下のように定義する．た

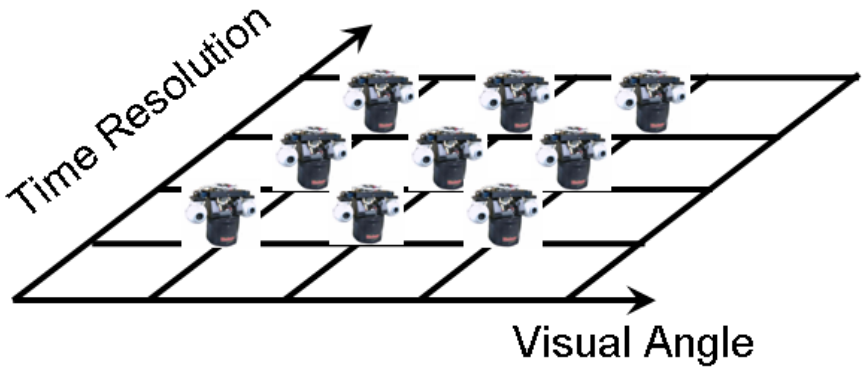


Fig.4.6: Facial robots with various embodiment

だし，この時シェマ選択のルールは 3.6 節で導入したルールをそのまま用いる．

$$S(PS) = \sum_{PS_n} -\mathbf{P}(PS_n)\log(\mathbf{P}(PS_n)) \tag{4.14}$$

$$\mathbf{M}(PS) = \exp(S(PS)) \tag{4.15}$$

こうすると $\mathbf{M}$ は各シェマが満遍なく使われた場合のシェマの数に一致する．これを分化したシェマの数のバロメータとして利用する．このとき，各身体パラメータについての $\mathbf{M}$ は Fig. 4.7 のようになる．ここでは，視野が小さいときにも大きいときにも分化したシェマは少なく，中程度で最も多い．また，時間解像度についても小さいときと大きいときにはシェマは多くは生成されず，中程度において最も多く生成される．具体的には何故このような結果が生まれてくるのかについて説明する．時間解像度が非常に高いときにはどのような運動球も静止しているも同然に見えるし，視野が非常に広い時は運動球の運動は相対的に小さくなっていくので静止しているも同然になっていく．故に，どのような運動状態の時もロボットにとっては区別する必要の無い存在と映るのである．また，視野が狭いとき，時間解像度が不足する場合は，情報が決定的に不足するために追跡行為を実現することが出来ず，縦運動であろうが横運動であろうが先の例とは逆の立場から区別しても仕方ない存在と映るのである．これは，ちょうど 4.2.2 節において我々が仮説として挙げた身体性と記号生成についての非線形な関係を表している．つまり，先に述べた記号生成についての仮説を認めるとするならば，双シェマモデルはその構成論的なモデルとなっているといえる．

しかし，Fig. 4.1 にあげた中での記号生成時のパフォーマンス向上についてはこの

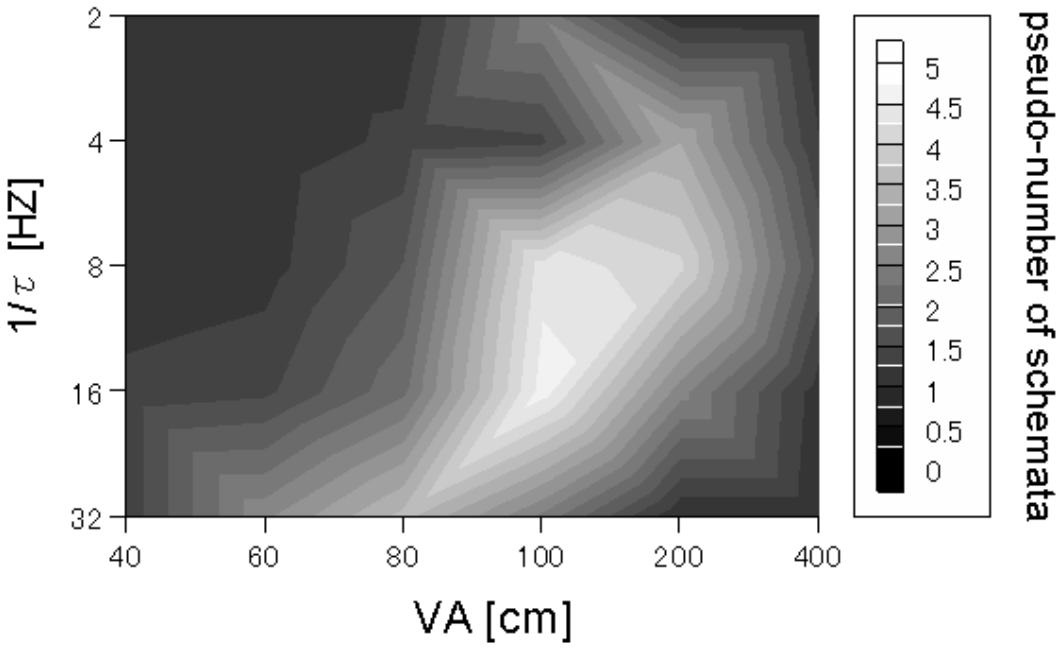


Fig.4.7: Relationship between spatiotemporal resources and pseudo-number of perceptional schemata

グラフは何も言っていない．次節では，シェマ分化とパフォーマンスについて考察する．

4.5.3 内的表象によるパフォーマンスの向上

モジュール化されない内部モデルは新たな環境に対面したときに，古い記憶が上書きされてしまうために，もとの環境に戻った時に一から再適応しなおさなければならないという非効率性がある．つまり，モジュール化による内的表象の利点とは，異なると判断したダイナミクスを別々のモジュールに割り当てることにより，記憶の保持を行う点にある．そこで，静止球についてこのような割りあてを知覚シェマ選択器が上手く行えているかどうかと，そのときの静止球追跡のエラーを比べることによって内的表象生成とパフォーマンスの関係について考察する．上のような割り当てが上手く出来ているかどうかを測るために，相互情報量 (相互エントロピー) を導入する [117, 159]．相互情報量は片方の情報を得ることで，もう一方の情報についてどれだけ情報を得られるかを表す．これを通信路に当てはめると，通信路が情報をどれだ

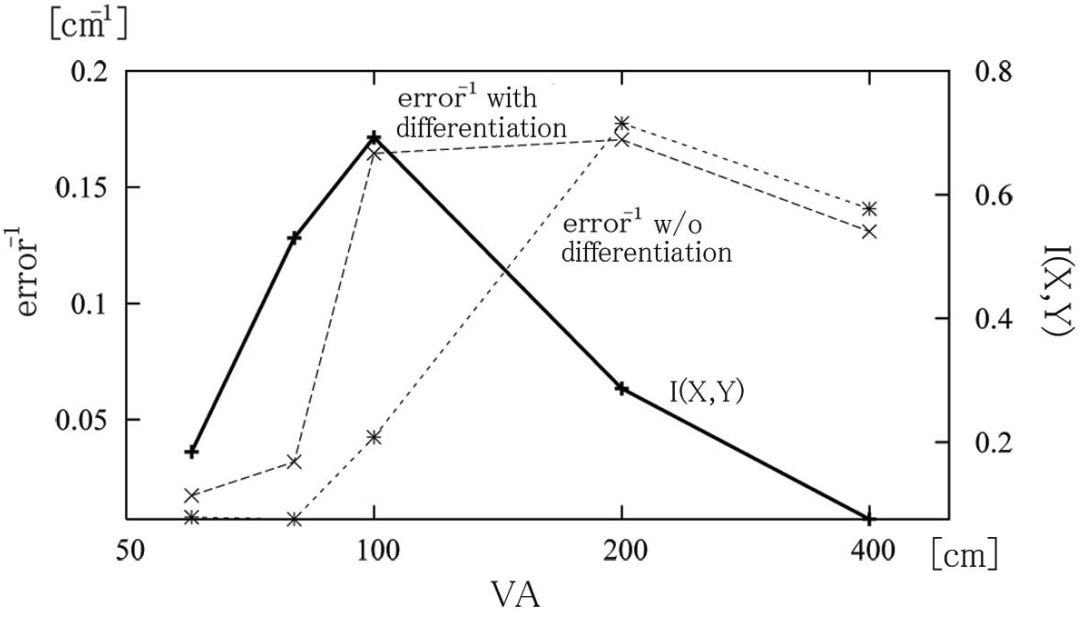


Fig.4.8: Broad line: mutual information between the ball's movements and selected perceptional schema; dotted lines: inverse values of averaged prediction errors of Dual-Schemata model and without differentiation.

け欠損せずに送っているかを調べる指標となる．今，集合 X，Y に対する確率分布 $P(x)$，$P(y)(x \in X, y \in Y)$ と同時確率分布 $P(x,y)$ があった時，相互情報量 I は

$$I(X,Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{4.16}$$

で定義される．$\mathbf{M}$ と同様に $\mathbf{N} = \exp(I)$ がその通信路の擬似的なチャネル数に相当する．今 $X = \{M_1, M_{other}\}$, $Y = \{PS_1, PS_2, ...\}$ として，パフォーマンスの指標となるカメラ焦点と対象球の距離の時間平均の逆数 $error^{-1}$ と共に Fig. 4.8 に示す．各身体性のセッティングについて 5 回ずつ実験を行いその平均値を示す．ちょうど静止球を他の球と識別する領域において，分化を行わない場合に比べてパフォーマンスが向上していることが分かる．このグラフ関係はちょうど Fig. 4.1 と重なり合う．Fig. 4.8 において，十分視野の広い領域においてパフォーマンスが落ちているのは，カメラの視野によるメリットよりもカメラ入力にのるノイズの大きさの方が大きな影響を与えたためと考えられる．これらのことより，双シェマモデルは我々の考える身体パラメータと記号生成についての非線形仮説を満たす構成論となっていることが分かる．

また，Fig. 4.8 に示した図を見ると内的表象生成機構の存在が運動球追跡のパフォーマンスの良い視野サイズの領域を視野のサイズの小さい方向へ拡大していると見ることが出来る．これは逆にとれば制限された身体性の上でも内的表象を獲得することにより，より広いクラスの環境に対して適応的に行動できることを示唆している．本実験結果の示すところは内的表象とは自らの身体パラメータが後一歩で及ばないようなクラスの環境条件に対して，内的表象を生成することにより適応することが出来るということである．

4.6 まとめ

本章では，身体性と自律ロボット内部での記号創発の関係について考察し，その構成論的モデルとして双シェマモデルをとり上げシミュレーション実験を通じて検証した．前章で導入した知覚シェマの内部関数を線形に制限することにより，簡易版である LDSM を導入し，様々な身体パラメータを持つ顔ロボットにシミュレーション空間内で実装し四種類の異なる物理的なダイナミクスを持った運動球と相互作用をするなかで知覚シェマの組織化を行わせた．その結果，身体パラメータが高すぎる場合も低すぎる場合もシェマの分化は殆ど行わず，中程度の時に最も多く分化するという非線形な結果を得た．単純に考えると，多くの情報を得ることが出来る身体を持っていれば多くの記号を獲得すると考えられがちであるが，必ずしも多くの記号を持つことが主体にとってメリットになるわけではない．主体は自らの身体に合った形で世界を分節化して初めてその恩恵を受けることが出来るのである．まさに，自らに意味のある表象のみが必要なのである．しかし，このようなことを構成論的な方法以外で追及することは，現時点では困難である．内的表象は計測されるものでは無く内観を通してしか認識されないものだからである．

自律系に認知される主観的な情報は身体と環境の相互作用の中で生み出される．本章の実験では，その身体の空間的，時間的特性を操作することにより，その情報の変化を生み出し，その結果として組織化される表象系が変化することを示した．もちろん逆も起こりえる．つまり，環境の空間的，時間的変化が表象生成の違いを生むと言うことである．特にこの環境における空間的な情報の違い，つまり，本実験ではボールの軌道など運動パターンを変化させればそれに応じて表象生成も変わるであろうことは容易に想像がつく．環境の変化が異なる表象生成を生み出すのは議論として新規性が無いと考え本章では議論しなかった．ただし，環境変化の時間的側面については

自明ではない．人間の目に異なると映る複数の運動パターンもそれらの切り替えの周期を細かくしていくと，まるで不規則な一纏まりの非線形な運動パターンのように見えてくる．このような現象も双シェマモデルにおける知覚シェマの分化によって実現される．本論文には具体的なデータは示さないが文献 [90] にて発表した．

第２章でも述べているように人間の行う記号の設計と解釈が適応的にどの様におこなわれているかは知能ロボット研究においても人間機械系の設計論においても重要であるし，認知科学，言語進化などの研究においても重要な横断的課題である．しかし，認知実験の系を考える上でも検証すべき合理的なモデルが無ければ始まらない．よって構成論的な研究を含め，今後とも多角的な研究が望まれる．

第 III 部

報酬設計に基づく社会的相互作用を通じた汎化行為概念生成

第 5 章

汎化行為概念の適応的獲得
-双シェマモデルベースの強化学習-

我々は獲得した行為を，その行為を獲得した状況でしか用いることができないだろうか．我々人間は環境変化に対して適応的に，行為を修正し実行することが出来る．これは，実際に我々人間が得ている行為概念がただのセンサ入力からモータ出力への条件反射的なものではなく，一段階汎化されたものであることを示している．本章では双シェマモデルにおける行為シェマ (Fig. 5.1) がこのような汎化行為概念に相当することを示し，さらにそれを強化学習 (reinforcement learning) を通して獲得するための手法を提案する．

5.1 はじめに

現在，機械システムの設計論は望みの状態を静的に維持することを目的にする安定性を志向した思想から，自律的に動的なパターンを生成することにより所望の振る舞いを生成したり，自律的に環境に適応することにより新たな機能を形成していくという動的な機能構成を目指した思想へと展開しつつある．[182, 167]．そこでは人間に隷属的な道具としての機械の設計という捉え方から，自律適応系としての機械の設計へという発想の転換が本質的である．

このような自律適応系は，動的に変化する環境下で様々なタスクをただ一つの身体を通して学習し実行していくということを求められる．しかし，そのような動的環境の性質を前もって設計者が全て想定することは不可能であり，故に，行うべきタスク

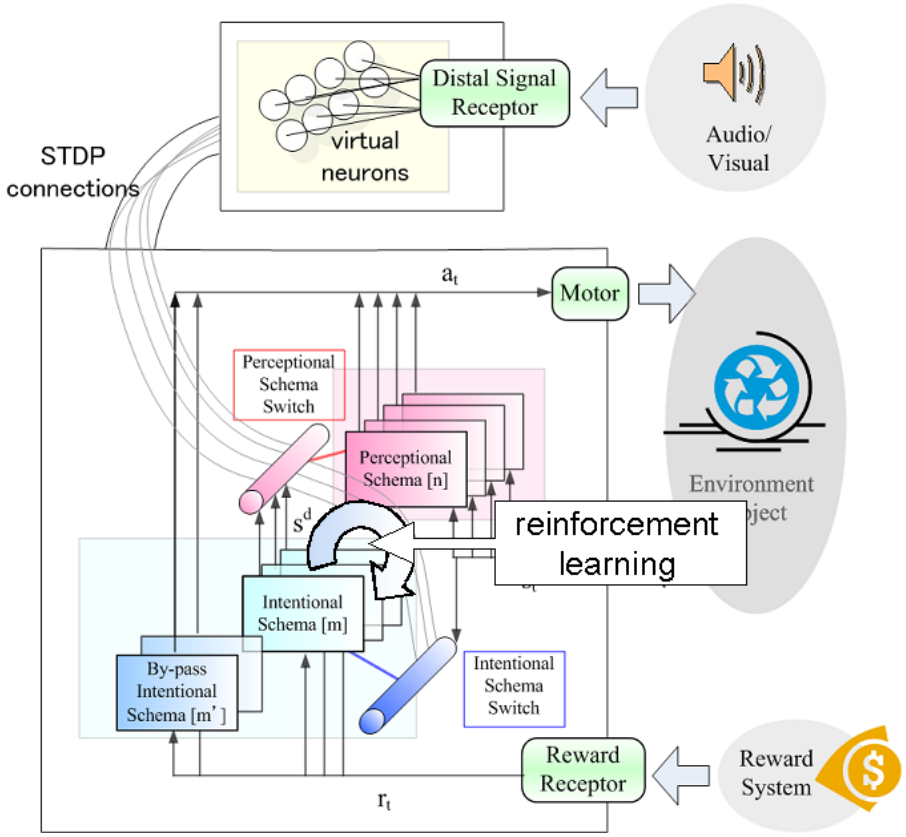


Fig.5.1: Intentional schemata in the Dual-Schemata model with a STDP neuronal circuit

もアプリオリには決定できない．このようなニーズを満たすためには，変化する環境や求められる振る舞いに対して適応的に機能生成を行っていくような発達学習機構が必要である．

このような適応性，発達性を実現する枠組みとして強化学習 [83] や模倣学習 [73] などが注目されている．そのような中，本章までに，自律ロボットが環境との相互作用を続ける中で，環境との相互作用の構造的変化に自律的に気づきモデル学習及び概念形成を行うためのモデルとして双シェマモデル (dual-schemata model) を導入してきた．しかし，前章まででは行為シェマ (intentional schema) として表象される行為概念は前もって設計者によって与えなければならないという制約があった．本章では，双シェマモデルにおける行為シェマに対し強化学習手法を通して行為概念を学習させる方法論を提案する．従来の強化学習においても行動方策としての行為概念を獲得させることは可能である．では，双シェマモデル内の行為シェマとして獲得される

メリットはどこにあるのだろうか．ここで獲得される行為シェマは通常の強化学習で獲得される方策とは異なり，環境ダイナミクスの変化に対する普遍性を持っている**汎化行為概念**となっているのが特徴となる．

本章では行為概念，及び汎化行為概念について議論した後に，汎化行為概念を獲得するための学習機構として双シェマモデル強化学習 (reinforcement learning on Dual-Schemata model) を提案する．

5.2 汎化行為概念

本章では本論文の主題でもある，汎化行為概念 (generalized behavioral concept) について議論を行う．そして，センサ空間のアトラクタとしての汎化行為概念の表現を提示する．

5.2.1 “行為” とは何か？

一般に強化学習理論等で “行為” という場合，それはモータ出力 a_t を指す場合が多い．しかし，グリッド空間で迷路抜けタスクを行うような離散的仮想空間で活動するエージェントではなく，実連続空間で活動するエージェントにとって瞬間的なモータ出力が何らかの意味を持つことは少なく，対象に対して働きかける一連の動作一纏まりで初めて “行為” として認識される．その点から見れば，ある目的をもった方策関数 $a_t = \pi(s_t)$ のような，センサ入力に対し時々刻々と出力を変化させていく関数関係によって示される制御則のことを “行為” と呼ぶ方が自然であると考える．また，その制御則によって表出される振る舞いをその制御則，つまり方策が表象する**行為概念**と呼ぶことにする．例えば，ある方策関数 (制御則) を用いたときに，観察者から見てロボットが歩行したとするならば，その行為概念は「歩行」であると名づけて構わない．

5.2.2 汎化行為概念

ところで，我々人間は歩く，投げる，叩くなどの様々な行動を表象する言葉，またはそれに相当する行為概念を持っている．しかし，例えば「投げる」という概念を例にとってみても，投げる距離や，投げる物，風の具合などによって，実際に筋肉に伝えられる筋指令はまるで変わってくる．つまり用いるべき方策 $a_t = \pi(s_t)$ 自体が変

化する．ロボットの言葉に置き換えてみると筋指令はモータ出力であり，姿勢角やボールの軌道がセンサ入力となる．そして，投げる物や風の具合といったものは環境や対象のダイナミクスとして行為の結果に影響を与える．このように，同一でないモータ出力の時系列にも関わらず我々は「投げる」いう一つの言葉でそれらを表象している．これは我々人間が環境や対象のダイナミクスには依存しない形である種の行為概念を持っていることを示唆している．しかし，Fig. 5.2 の右側のエージェントに示すように，一般に強化学習の枠組みなどを通じて獲得される行為は固定された環境ダイナミクス下で獲得され，他の環境下で応用することができない．このような行為概念は環境の変化に対する汎化性が殆ど無いと言える．今，ここで言う汎化性とは一般に学習器の汎化性と呼ばれるものとは異なる．学習器の汎化性とは経験していない状態量について正しい応答を返せるかということである．これに対して我々の述べている行為の汎化性とは，同じ状態量に対しても環境・対象次第で異なる応答を返さねばならないという点で，所謂学習器の汎化性より上位の概念として捉えられる．獲得した技能を新規環境で利用できるかという問題は新規性課題 (novelty problem) と呼ばれ人間の運動学習の研究においても古くから研究されている [67]． 本稿ではこのような環境変化に対する普遍性を持つ行為概念を**汎化行為概念**と呼ぶことにする．本章は Fig. 5.2 の左側のエージェントに示すような汎化行為概念を，自律ロボットに適応的に行為シェマとして獲得させることを目的としている．

5.2.3　行為主体にとっての行為

汎化行為概念は行為主体にとってどのようなものであろうか．行為主体は自らのセンサを通してのみ自らの行為の結果は観測できると考えるのが妥当であろう．つまり，行為主体は自らの行為をセンサ空間 (状態空間) 内の軌跡として捉えざるを得ない．そうしたとき，行為主体の行為概念とはセンサ空間内のアトラクタとして捉えることが出来る．目的志向的な行為は点アトラクタであり，歩行などの周期運動的な行為はリミットサイクルとして表現される [144]．このように考えたときのセンサ空間内でのアトラクタを形成する力学場は内部的なダイナミクスであり環境のダイナミクスには依存しない．よって，このようなアトラクタは汎化行為概念として異なった環境ダイナミクス間で再利用可能となる．我々はこのような視点から後に述べる双シェマモデル内で行為シェマにセンサ空間内でのアトラクタとしての役割を担わせることにより汎化行為概念の実現を行った．

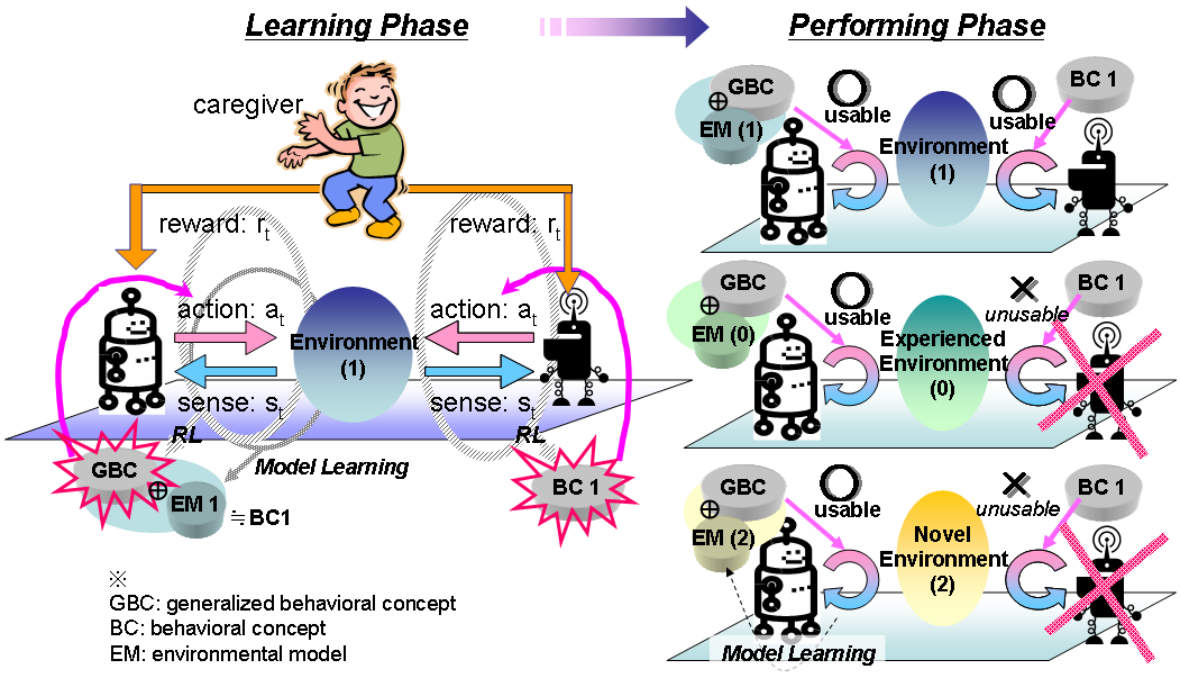


Fig.5.2: An overview of differences between a generalized behavioral concept and a specific behavioral concept

5.3　階層的モジュール型学習器

前節で述べたような汎化行為概念を自律ロボットに獲得させるには，通常の強化学習を用いて $a_t = \pi(s_t)$ の形で行為を獲得したのでは不十分である．このような欠点を補うためには，強化学習の研究から一歩踏み出した学習器の構造化についての研究が必要である．現在，モジュール性や階層性を導入した学習器の研究が盛んである．本章ではそれらの研究を紹介し，汎化行為概念の獲得におけるモジュール性，階層性の役割について考える．

5.3.1　モジュール型強化学習

先にも述べたが，近年，ロボット制御の手法や脳のモデルとして，適応的なモジュール型の制御器が注目を集めている [103, 89]．我々は環境が持つダイナミクスの相転移的変化に着目し，その変化を文脈依存的かつ主観的に発見し新たなモジュールを累

増的に作成していくモジュール型学習モデルとして双シェマモデル (Dual-Schemata model) を提案してきた．これは，環境のダイナミクスの変化に対応しつつ，自らの内部に多様性を創出していくことにより，構造化をすすめる，分化 (differentiation) を基本的なダイナミクスとした学習モデルである．

モジュール型の強化学習機構としては銅谷，片桐らのモジュール型強化学習 (MMRL) がある [25]．しかし，これは報酬とダイナミクスから最適方策を導出する強化学習の中心部分を，リカッチ方程式の代数的解法で求めるために，強化学習機構としては特殊なモデルであり，他手法との比較が難しい．また，一つの線形の予測モデルにつき一つの線形の制御器という考え方は，決して普遍的ではなく適用範囲が制限されてしまう．さらに，線形予測器により分割された状態間での報酬分配が考えられないため，大域的なタスク成功に到るとは限らない．これらを解決する手法は現在，杉本，鮫島らによって提案されている [155]．累増的にモジュールを獲得する強化学習アプローチには高橋らのモジュール型強化学習 [138, 86] がある．

これらのモジュール型強化学習では環境のダイナミクスをモジュールの分担により分節化することが出来る．故に，自律ロボットは，環境の変化に対しモジュールを切り替えることで，動的環境下においても適応的に振舞うことが出来る．

5.3.2　階層的強化学習

また，強化学習の研究において，学習を階層化する階層的強化学習の枠組みが議論されている．強化学習を階層化する利点は複数あるが，最も頻繁に強調されるのは上位層に全体のゴールに対してサブゴールを生成させることにより探索を効率化するというものである [55, 172]．しかし，この利点を利用するためには一般的には上位層を下位層より粗くセルで区切り離散化する必要がある．これは階層性そのものがもたらす利点というより，階層的強化学習の上位層と下位層の離散化の粒度の差，つまり分節の仕方の差から生じる利点である．

これに対し，我々は階層性の持つ本質的な利点を自律ロボットの獲得行為概念の汎化性の視点から捉えている．双シェマモデルでは下位層 (知覚シェマ) に環境のダイナミクスを任せることにより，上位層 (行為シェマ) においては環境のダイナミクスに依存しない行為表象を獲得することを目指している．しかし，本章で提案する方法では特に階層毎の粒度の違いは導入しないために，後の実験で示すように強化学習の高速化という面においての寄与は期待できない．獲得した行為概念の再利用という面

では港ら [133] の研究などがあるが，行為概念の再利用において保存される部分の決定があいまいであるといった点や，再利用時に変更部分を忘却してしまうために，再び同じ環境に戻ったときに再学習を要するなどの欠点がある．

次節以降で本章の主題である双シェマモデルにおける行為シェマの獲得手法についての説明を行う．

5.4 強化学習による行為シェマの獲得

本章では双シェマモデルにおいて，その行動学習のために強化学習を用いる手法を提案する．双シェマモデルにおける知覚シェマの適応ダイナミクスについては，第 2 章，第 3 章で説明したので参照していただきたい．本章の実験に用いる知覚シェマのダイナミクスとしては前章で導入した，LDSM を用いる．

5.4.1 強化学習の基礎

強化学習は設計された報酬を将来にわたり最大化するような方策をオンライン，つまり個体発生的プロセスを通して獲得するための手法である．広義には GA，クラシファイアシステムなどの系統発生を通した淘汰による進化的手法も含むが，ここでは TD 学習を中心とした個体発生的なプロセスで学習するオンライン学習の方法論のことを強化学習と呼ぶことにする [83]．強化学習では状態 $s_t \in S$ と行動 $a_t \in A$ が決定したときに確率的に次状態 $s_{t+1} \in S$ と $r_t \in R$ が決定することを仮定して一般的には定式化される[*1]．ここでは簡単のため微小なノイズを除いて決定論的な場合のみを扱うので $r_t = r(s_t, a_t)$ であらわされるものとする．方策 $a_t = \pi(s_t)$ の下での時刻 t における重みつき累積報酬の期待値 V^π は離散時間では以下のように定義される．

$$V^\pi(t) = E[\sum_{k=0}^{\infty} \gamma^k r_{t+k}] \tag{5.1}$$

ここで γ は割引率と呼ばれる定数である．強化学習では V^π を状態 s_t の関数 $V(s_t)$ として推定し，それに基づき方策 π を変更することにより最適方策を探索する．価値関数を推定する際の誤差は TD 誤差 δ_t と呼ばれ，状態行動価値 Q を用いて以下で定

[*1] 前章までは補助入力項 m_t を加えた $s_t^\dagger$ を入力として考えていたが，本章からは特に含めず定式化を進める．含めた場合への拡張も前章までと同様に自然に行うことが出来る．

義される．

$$\delta_t = r_t - \hat{r_t} = r_t - (Q(s_t, a_t) - \gamma V(s_{t+1})) \tag{5.2}$$

状態行動価値が推定されず，状態価値のみによる Actor-Critic 法では

$$\delta_t = r_t - \hat{r_t} = r_t - (V(s_t) - \gamma V(s_{t+1})) \tag{5.3}$$

となる．TD 誤差は上式より報酬の予測誤差と見做すことも出来る．ここで，価値関数が予測器としての機能も持っていることに注意する．TD 誤差を用いて価値関数や方策をどう変更していくかは Q-learning や Actor-Critic 法などの具体的な立式に依存する [83]．また，上では離散時間についての定式化について述べたが，連続時間への拡張も K. Doya，L.C. Baird らによって定式化されている [10, 23].

5.4.2　センサ空間のアトラクタとしての行為

双シェマモデルの枠組みにおいて，行為シェマは $(s_t^{\dagger}, s_{t+1})$ の二項関係を規定する．これにより，順次決定されるセンサ値が自律主体にとっての行為となる．これは自律主体にとって一纏まりのチャンク化された行為とは，受容される一連のセンサ値の系列で決められる状態空間内のアトラクタとして描かれることを前提とした定義である．行為シェマは自律ロボットが他者の助け無しに得ることの出来るセンサ空間の中のベクトル場として張られ，実際のモータ出力 a_t とはまったく無縁に定義される．自律主体の行為の自らにとっての意味とは，自らのセンサ入力のみから認識されるため，モータ出力 a_t がどのように出力されようとも，その影響が自らのセンサ越しに現れない場合，有意義な行為としては理解されえない．つまり，行為シェマ IS_m は作動上閉じた自律主体にとって理解されうる一纏まりの行為を表象している．行為シェマは知覚シェマと同様に並列に複数存在し，行為シェマ選択器によって選択されるが，本章では行為シェマに養育者 (caregiver) による報酬をあてがうことにより，養育者の意図する所望の振る舞いを強化学習により学習させる方法を提案する．説明は簡単のために離散時間で記す.

5.4.3　行為シェマの強化学習

行為シェマの適応プロセスを実装するにあたり，強化学習の理論としては，連続時空間上での Actor-Critic 法を用いる [23]．これは Actor-Critic 法を連続空間に拡張

したもので，価値関数 (critic)V と方策関数 (actor)G を基底関数の線形和として表現することにより，連続空間上へ拡張したものである．

$$V(s_t) = \sum_i \omega_i \mathbf{b}_i(s_t) \tag{5.4}$$

$$G(s_t) = \sum_j \theta_j \mathbf{a}_j(s_t) \tag{5.5}$$

また，方策関数における連続な粒度での探索を表現するために Gullapalli の確率的強化学習の手法 [31] を用いることにより，行動空間，状態空間共にセル状に離散化する必要を無くしている[*2]．さらに，連続時間上でメタパラメータを設定することによりロボットの行動周期の時間間隔の設定とは独立に全てのメタパラメータを調節できるようにしている．ただし，一定の行動周期に固定する場合においては，離散時間における立式と違いは無い．これらの定式化は連続な実空間，及び状態行動空間の適切な分節化が出来ないような環境で活動するロボットの強化学習を行う場合に非常に有用である．また，profit sharing を積極的に行うために，木村らの適格度トレースの枠組み [45] を導入した．これらの方法を用いた強化学習では通常，方策の出力としてモータ出力 a_t を割り当てるが，行為シェマの出力であるセンサ入力の目標値 s^d_{t+1} を割り当てることによりこれらの強化学習手法を行為シェマの適応ダイナミクスとして利用した (Fig. 5.3)．しかし，行為シェマにおける学習の場合，最終的に現れる真のセンサ入力としての s_{t+1} は知覚シェマのモデル化における精度が悪い場合は，行為シェマから目標値として出力される s^d_{t+1} とは異なる．よって，我々は次入力として得られる s_{t+1} を実際に行為シェマがとった行為と見なし，actor の学習に利用した．こうすることによって，行為シェマの出した目標値を知覚シェマが実現できない場合，実現できた値を基準として actor の学習が進み，おのずと系が実現しうる値のみを行為シェマは出すようになることが期待できる．また，局所的にしか与えられない報酬に対しても対処できるようにするために，適格度トレース (eligibility trace) を critic だけでなく，actor にも導入した．具体的な更新式は以下のようになる．

[*2] 近年，収束証明がなされたこともあり価値関数ベースでなくこのような方策ベースの強化学習が再評価されてきている [84, 46].

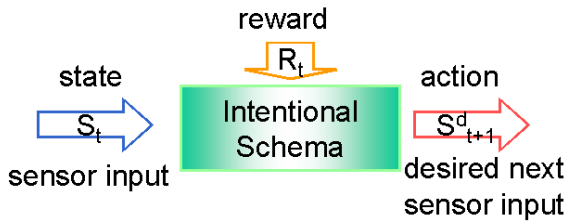


Fig.5.3: A desired sensory input as an action

価値関数の更新

Actor-Critic 法における価値関数とは時刻 t での状態 s_t を始点にして現在の方策 π に従った場合に将来にわたって獲得することの出来る報酬の重みつき期待値として定義される．基本的には学習はこの TD 誤差の二乗誤差を最小化するように慣性項つきの勾配法を用いて行うが，報酬の分配 (profit sharing) を考えるために以下で更新される適格度トレース (eligibility trace)$\mathbf{e}^{\omega_i}$ を導入する．

$$\mathbf{e}_t^{\omega_i} = \gamma\lambda\mathbf{e}_{t-1}^{\omega_i} + \frac{\partial\hat{V}}{\partial\omega_i} \tag{5.6}$$

ここで，λ は適格度トレースの減衰率パラメータである．この $\mathbf{e}^{\omega_i}$ を用いて毎時以下のように基底関数の係数を更新する．

$$\omega_i \leftarrow \omega_i + \alpha\delta_t\mathbf{e}_t^{\omega_i} \tag{5.7}$$

方策関数の更新

Actor-Critic 法においては探索のための項を方策内に陽に設計する必要がある．よって，実際の方策は標準偏差 1 のノイズ n_t を用いて以下のように書ける．

$$s_{t+1}^d = \mathbf{s}(\mathbf{s}(G(s_t)) + \sigma_t n_t) \tag{5.8}$$

また，実際の状態空間は各次元についての上界，下界を持つ場合が多い．そのために飽和量を超えたものを定義域内に納めるための関数 $\mathbf{s}$ を用いている．探索率 σ_t については Willson，木村らのように勾配法を用いて制御する方法 [45]，森本，銅谷らの

ように価値関数で制御する方法 [23, 55]，または時間の進行に対して単調に減少させていくフィードフォワード的な制御などが考えられるが，本研究では価値関数による制御を採用する．このどれを用いるかは本論文の主旨ではないので議論を避ける．

actor の学習は基本的にはノイズによる探索を通して実際に起こした行動が期待報酬よりも多くの報酬を得た場合にはその行動を強化するという仕組みで行われる．具体的な更新式は価値関数と同様に適格度トレースを用いて以下のように表される．

$$\mathbf{e}_t^{\theta_t} = \gamma\lambda\mathbf{e}_{t-1}^{\theta_i} + \frac{\partial G}{\partial \theta_i}(s_{t+1} - G(s_t)) \tag{5.9}$$

$$\theta_i \leftarrow \theta_i + \alpha\delta_t\mathbf{e}_t^{\theta_i} \tag{5.10}$$

以上は係数ベクトル ω，θ の更新に本質的には勾配法を用いたものであるが，学習のなだらかさを確保するために実際には慣性項を用いた更新を行う．

5.5　中心への位置制御の獲得

本章では前述のモデルを用いて行った実験について記す．前章までは，移動球追跡というタスクのみに双シェマモデルを適用してきたが，本章では新たに板上の球の制御課題を与える．

5.5.1　実験環境

本研究では振動を起こさずに加速や移動を行う台車に乗り，その上で転がるボールを平板の角度を変化させることにより制御するというタスクを行うエージェントを考える．論点を明確にするために，課題自体は非常にシンプルなものを用いている．具体的には Fig. 5.4 に示すような系を考える．ここで，平板を制御するエージェント (Balancer) と台車を制御するエージェント (Cart) は別であり，お互いの制御指令等は観測することが出来ないとする．Fig. 5.4 の Balancer 部と Cart 部に「目」の記号をつけているのはこの重要点を強調するためである．つまり，Balancer は制御しようとする球にかかる慣性力という形で Cart の行動の影響を受ける．それにも関わらず，Cart の行動を直接には観測できない環境で制御を行わなければならない[*3]．この系の y 方向についての物理量を Fig. 5.4 に示す．x 方向についてもほぼ同様である．Fig. 5.4 における平板の規格化された法線ベクトルを $\vec{n} = (n_x, n_y, n_z)$ とする．

[*3] 体育祭で行われるスプーン競争を想像していただきたい．

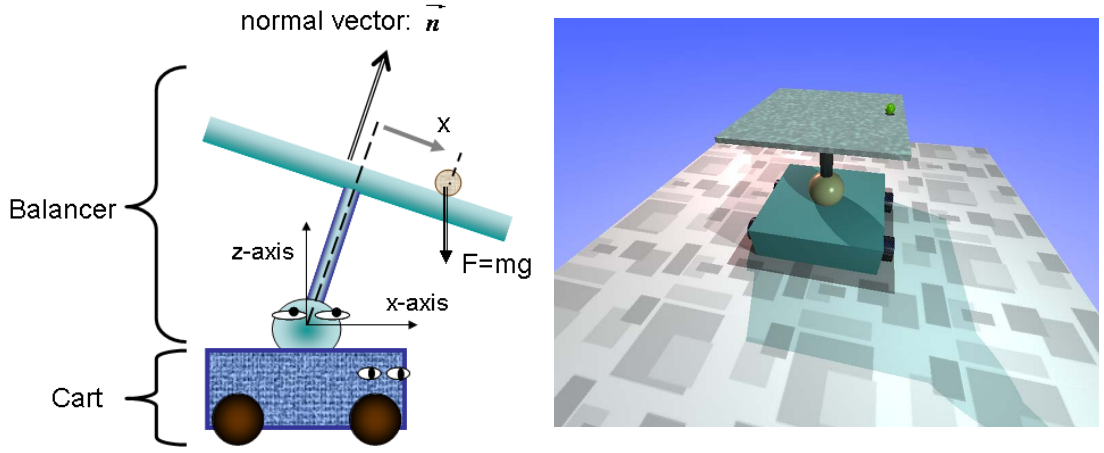


Fig.5.4: An overview of the agents used in the experiments

Balancer が動かしうる法線ベクトルの範囲には $-\frac{1}{2} \leq n_x, n_y \leq \frac{1}{2}$ の制限があるものとする．Balancer は二次元の行動出力 a_t を出す．$a_t = (a_t^x, a_t^y)$ により法線ベクトルは以下の式で直接的に制御される．

$$(n_x, n_y, n_z) = \frac{1}{2}(a_t^x, a_t^y, \sqrt{1 - (a_t^x)^2 - (a_t^y)^2}) \tag{5.11}$$

平板は一辺 2[m] の正方形で，その中心を原点とし x，y 座標を設定する．この時平板の上のボールの運動は近似的に以下のように表される．

$$(\ddot{x}, \ddot{y}) = (gn_x - \alpha_x, gn_y - \alpha_y) \tag{5.12}$$

$\alpha = (\alpha_x, \alpha_y)$ は Cart の運動や姿勢によって現れる外因的な項である．Cart の運動については詳述しないが，基本的には y 方向に加減速と左右方向への操舵が可能である．これに伴いボールには慣性力，遠心力がかかる．また，路面の傾きによっても α は変化する．ボールの位置と速度に微小なセンサノイズを乗せたものを Balancer のセンサ入力 s_t とし，Balancer は毎ステップこれを獲得することが出来る．

$$s_t = (x, \dot{x}, y, \dot{y}) + \epsilon_t$$

ボールが平板から落ちた場合には，次のステップで平板の上のランダムな位置にボールを置きなおす．この時，特にペナルティは与えない．今回のタスクではセンサ以外の入力 m_t は用いない．

行為シェマ内における actor の基底関数としては Gauss 型の RBF を各次元 4 個 (計 $4 \times 4 \times 4 \times 4 = 256$ 個) ずつ，critic には各次元 5 個づつ (計 $5 \times 5 \times 5 \times 5 = 625$

Table5.1: Intentional schemata

Index	Name	Policy
IS_0^-	move random	$a_t = G_0^- \equiv \mathbf{n_t}$ ($\mathbf{n_t}$ is a noise term)
IS_1	shake a ball	$s_{t+1}^d = G_2(s_t) \equiv \mathbf{n_t}$
IS_2	(for reinforcement learning)	$s_{t+1}^d = G_1(s_t)$

個) を等間隔に配置した．また，実験を通して強化学習のメタパラメータはそれぞれ $\gamma = 0.6, \lambda = 1.0, \alpha = 0.1$ に設定した．また，ロボットの行動周期は 2[Hz] とした．

5.5.2 中心への位置制御の獲得

本節では提案手法を用いて，ボールの平板上での中心位置制御が可能であることを示す．報酬系としては以下を用意した．これは単純にボールが平板の中央に来たときにエージェントに報酬を与える．

$$r_t = \exp(-\frac{1}{2}((\frac{x}{0.5})^2 + (\frac{y}{0.5})^2)) \tag{5.13}$$

行為シェマは初期状態では Table5.1 に示す 3 つのものを用意する．双シェマモデルでは行為シェマの強化学習の為の探索のみならず，知覚シェマのモデル学習のための探索も行う必要があり，複数の行為シェマを切り替えながら学習することが重要である．よって，タスクの目的である IS_2 以外についても適宜選択し利用する．ただし，IS_2 を選択するときのみ IS_2 は強化学習を行うものとする．10000[s] の間学習を行った時の報酬の短時間平均を Fig. 5.5 に示す．比較対象として，連続空間の Actor-Critic 手法で基底に同様の RBF を用いたものを示している．両者とも実験を 5 回行いその平均をとっている．この実験の間は Cart は静止しているために，Balancer にとって環境のダイナミクスは常に一定である．両者ともに制御は成功へと収束していくことが分かるが，提案手法の方が収束は遅くなっている．これは両者で行動空間の定義が異なり通常の Actor-Critic 手法では 2 次元の行動空間であったのが，双シェマモデルにおける行為シェマの行動空間は状態空間に等しくなるので 4 次元になり，方策の探索空間が広がったためと考えられる．ただし，タスクや基底のとり方によっては双シェマモデルの方が速く収束する例も報告されている [91]．但

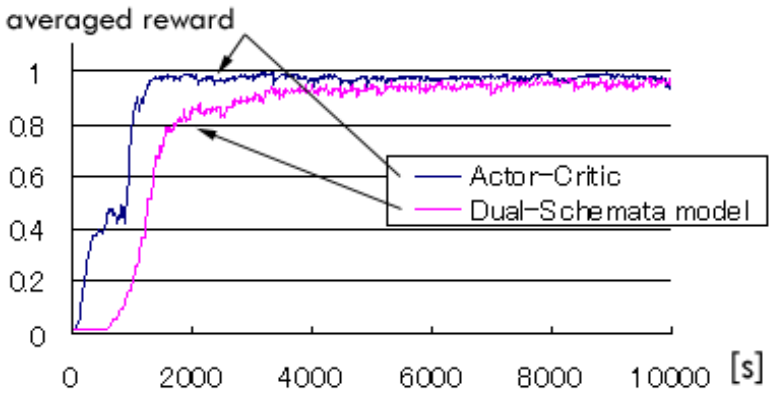


Fig.5.5: Transitions of the weighted averaged rewards in a simple reinforcement learning task.

し，本研究の主旨は学習の高速化には無いので，それを除けば学習能力としてはこの時点では両者大きな差は無い．このような静的環境下ではなく，動的環境下において行われる次章以下のタスクにおいて，この二つのモデルの差は明確になっていく．

5.6　汎化行為概念の獲得

さて，はじめにも述べたように，我々人間は異なる環境ダイナミクス下でも，同様の行為概念を用いた行動を行う．例えば，歩くという場合，坂を歩く場合でも，平地を歩く場合でも，砂地を歩く場合でも，その歩くという汎化された行為概念は変わらない．安定歩行時には足の関節角で構成された状態空間上でシステムは周期的な運動を行い，逸脱時にはその安定軌道へと引き込みを起こす非線形力学場，つまりリミットサイクルアトラクタとして “歩く” という汎化行為概念は表現される．行為シェマにより獲得された行動はこのようなものに近い．そのことを示すために，獲得された行為概念の環境の変化に対する再利用可能性を調べる実験を行う．

5.6.1　実験環境

実験のパラメータや系は前章と同じものを用いる．環境の変化としては Cart の路面の変化やコーナーでの遠心力を考える．加速する電車の中でバランスを保つためには，同じ立つという行動であっても，停車時や等速運動時とは違う力の入れ方をしなければならないという経験は誰しもがもっているであろう．実験のシナリオとしては

以下のものを考える．

1. 初め 10000[s] の間，Cart は停止し，Balancer はボール制御の強化学習を行う ($\alpha = (0, 0)$)．(状況 A)
2. その後 5000[s] の間，Cart が等速で約 10° の坂を上り始める ($\alpha = (0, -2)$)．Balancer は強化学習を続ける．(状況 B)
3. 獲得した行為概念の汎化性を調べるため，ここで Balancer の強化学習を止め，15000[s] から 20000[s] の間は，Cart は 200[s] 毎に状況 A と等価な等速運動，状況 B と同じ登坂を繰り返す．Balancer は獲得した方策を実行し続ける．(状況 C)
4. その後に Cart は半径 8[m] の円周上を秒速 4[m/s](時速 14.4[km/h]) で周回を始める ($\alpha = (2, 0)$)(状況 D)．

まとめると状況 A と状況 B の間に二種類の環境下で強化学習を行う．状況 C では一度学習経験のある環境での獲得された行動概念の再利用性を調べ，状況 D では中心位置制御というやるべきタスクは同じにもかかわらず，今まで経験していない環境に入っていった場合での獲得行為概念の再利用可能性，つまり汎化性を調べる．比較するのは以下の４つである．

エージェント 1	従来の Actor-Critic 手法
エージェント 2	従来の Actor-Critic 手法 (強化学習を止めない)
エージェント 3	双シェマモデル (知覚シェマの分化を考慮しない)
エージェント 4	双シェマモデル

ここでエージェント 2 は状況 C，D における強化学習の停止を行わず，強化学習を続ける場合の学習と比較検討する．エージェント 3 は LDSM による知覚シェマの分化の機構をはずしたものである．これによって，過去に経験した環境についての記憶を分散的に保持することが出来なくなる．知覚シェマの選択則については知覚シェマが分化しないために常に同じ知覚シェマを選択する．実験のパラメータについては前節と同様のものを用い同化係数については $k = 2.3$ に設定した．

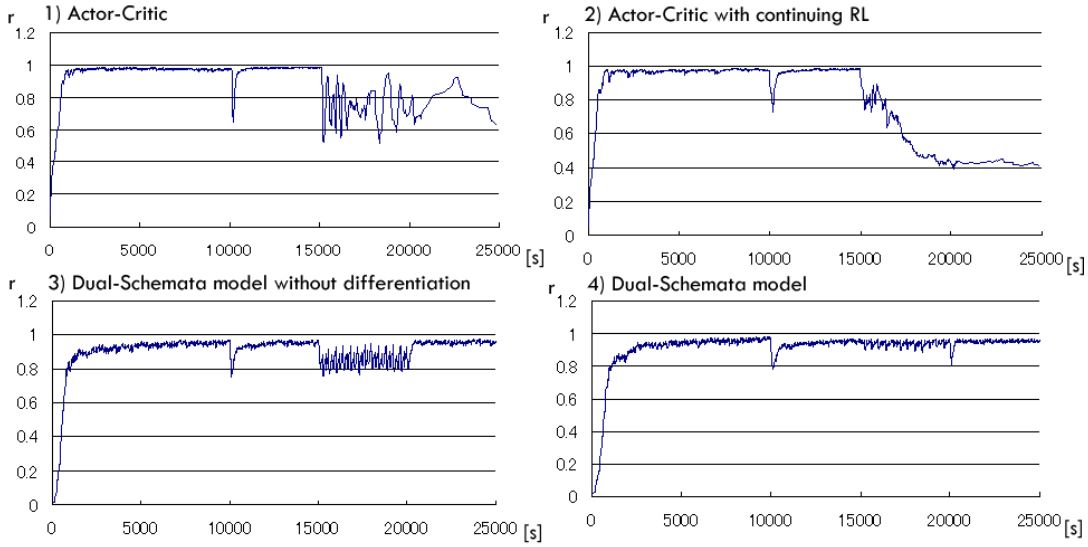


Fig.5.6: Transition of the weighted averaged rewards which were obtained by the four different learning architectures

5.6.2　実験結果

以上の４つの手法のもと得られた結果について，報酬の重み付き時間平均の推移を Fig. 5.6 に示す．それぞれ 6 回同様の実験をおこない，その平均をとっている．状況 A，B ではどのエージェントも十分な時間をかけてやると同様に環境への適応を果たす．平均報酬の変化だけを見ると大きな違いはみられないが，実際に各々のエージェント内で行われる学習は異なる．エージェント 1，2 では環境が変化したために，今までの方策が利用できなくなり修正を求められる．この修正を強化学習ベースで行っている．よって状況 A における方策は忘却される．グラフでは報酬の時間平均をとっているため，大きな差は出ていないが，エージェント 3 は強化学習より学習速度の速いモデル学習を行うために[*4]，強化学習はほとんど用いず，ここまでで獲得した行為シェマを維持したまま，知覚シェマの均衡化を通して環境に適応する．ただし，状況 A において学習した知覚シェマは忘れ去られてしまう．最後にエージェント 4 は環境の変化に伴い知覚シェマの分化を行い，この新たな知覚シェマをもって状況 B へと適応する．この時状況 A で獲得された行為シェマ，知覚シェマは忘却され

[*4] モデル学習は学習理論的には最も簡単な教師あり学習に分類される．

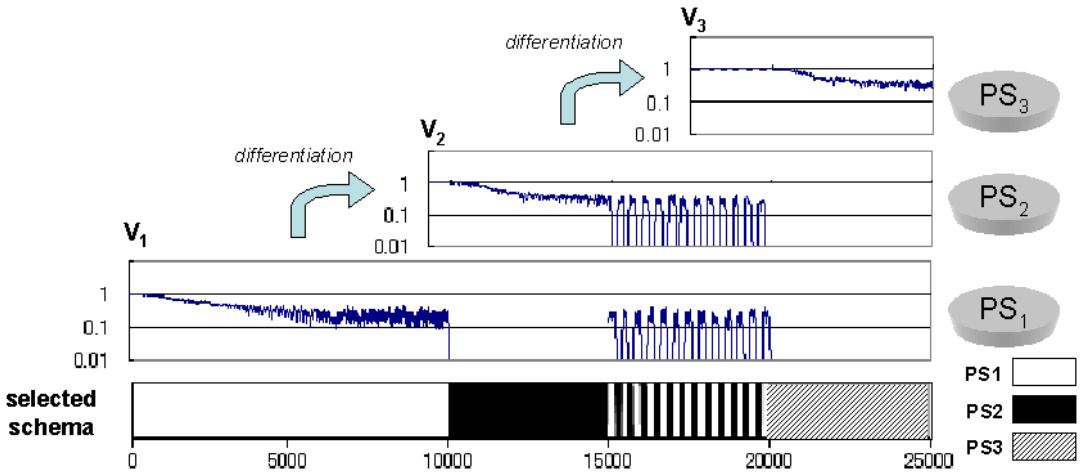


Fig.5.7: (Top) time course of schemata's activities and differentiation processes, (bottom) the selected schema at each time

ること無く保持される．

劇的な違いがあらわになるのは状況 C 以降である．エージェント 1 は状況 A での方策を既に忘却しているので，状況 B に相当する坂道に遭遇したときには上手くボールを制御できるが，状況 A に相当する状況 C 内で Cart が等速で平地をうごいている時には上手くボールを制御することが出来ない．よって，この二つのダイナミクスの切り替わりの中ではエージェント 1 は一貫した制御を行うことが出来ない．これに対しエージェント 3 は知覚シェマの均衡化のダイナミクスを通して随時適応していくことができるので，エージェント 1 に比べるとダイナミクスの切り替わりにもかかわらず，高い報酬を得続けることが出来る．しかし，知覚シェマの均衡化も時間を要するので，200[s] ごとの切り替わりには追従することが出来ない．これに対し，エージェント 4 は Fig. 5.7 のように 0〜15000[s] を通じて獲得，保持された状況 A 用の知覚シェマと状況 B 用の知覚シェマをスイッチングするだけで新たな環境に適応できるため，状況 C でも高い報酬をとり続けることができる．その後，状況 D になるとエージェント 3 は知覚シェマの均衡化，エージェント 4 は知覚シェマの分化及び均衡化を通じて，強化学習つまり行為概念の新たな学習を行わなくても適応することが出来る．これに比べてエージェント 1 は状況 B 用の方策を採り続けなければならないために，良い結果を出すことができない．最も悪いのは強化学習をダイナミクスがシフトしていくような環境下で行い続けた場合である．通常環境のダイナミクスが変化

した場合，HJB 方程式の解としての強化学習の最適解は変わってしまう [23]．よって，ダイナミクスが変化する中での強化学習はしばしば，片方のダイナミクス下だけで学習した結果を両方のダイナミクスに用いた場合より悪い結果を引き起こす [138]．今回の結果はそのような場合である．一定でないダイナミクスに支配された状況 C を通して崩壊させられた価値・方策は状況 D において新たな環境において強化学習を始めることにすら悪影響を与えているように見える．ただし，エージェント 1 と 2 の獲得報酬の優劣はどのようなダイナミクスの変化を与えるかということやタスクに依存し，優劣は逆転する場合もある [91]．

重要なのは通常の強化学習手法では環境が変わる毎に外部から与えられるたった 1 次元の報酬という情報を頼りに再学習を行う必要があるが，これは過去の行為概念を破壊し，再構築するプロセスに他ならない．報酬設計者の立場に立てば，環境が変化するたびに「ボールを中央に維持する」ということ自体を一から教えなければならないことになる．これに対し，双シェマモデルでは行為シェマは状況 A で獲得された後は，破壊されることはない．状況 A〜D を通して変わらぬ行為シェマが用いられる．つまり，これは環境変化に依存しない行為概念が獲得できたといえる．

5.6.3　行為シェマとセンサ空間上でのアトラクタ

主体にとっての一纏まりの行為とは，自らのセンサ空間内に描かれるアトラクタとして捉えられるのではないかという提案を 5.2.3 節において行った．アトラクタとは主に力学系のダイナミクスで一点に収斂していくもの (点アトラクタ) や，定常的な軌道へと収斂していくもの (リミットサイクル) を指すが，それぞれがリーチングなどのゴールのあるタスクと，歩行などの周期的な運動に対応している．リミットサイクルを形成する非線形振動子の結合によって，歩行を実現しようという研究は近年非常に盛んである [85, 192, 118]．本章ではこのセンサ空間内の力学系を行為シェマの内部関数において実現することにより，外部ダイナミクスに依存しない行為概念の内部表現を試みたわけである．

ところで，本稿の枠組みを用いて得られる汎化行為概念は，前節で述べたようなボールの中心への位置制御のように，ある固定値に安定化させるようなものだけではない．例として平面の中央で約半径 0.5[m] の円軌道に沿って周期運動を行なわせるような動的な周期運動を学習させる実験を行った．このための報酬系を以下のように定義する．基本的には平板の中心周りの角速度 ω を ± 1 で飽和させた項と半径 0.5 の

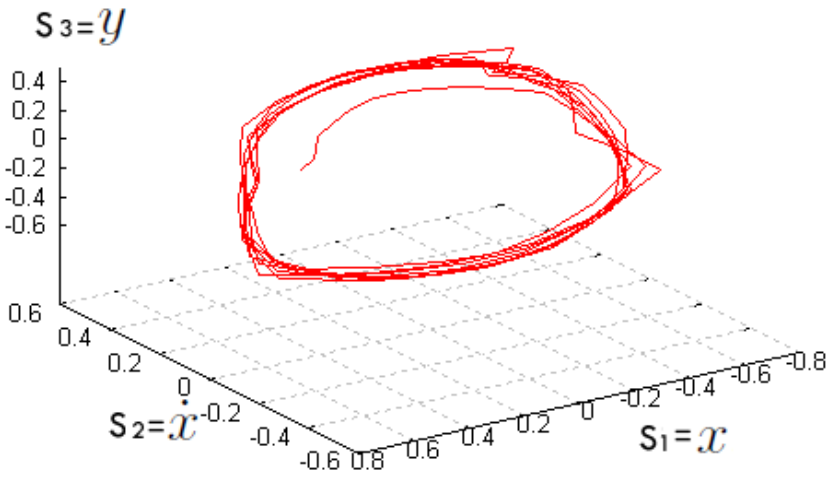


Fig.5.8: An obtained attractor in the agent's sensor space

円軌道上で最大となる項の積で構成されている．

$$r_t = \max(-1.0, \min(1.0, \omega)) \exp(-\frac{1}{2}(\frac{R-0.5}{0.3})^2) \tag{5.14}$$

この学習については次章においても再掲するが，結果的に，エージェントは前節同様に複数環境下で円運動を起こすことのできる行為シェマを獲得した．その結果，板上の適当な点に置いたボールは，獲得された行為シェマによりセンサ空間内のアトラクタに引き込まれていくことが確認された (Fig. 5.8)．このような周期運動へ引き込んでいく行為はセンサ空間上のリミットサイクルとして描かれていることが分かる．

5.7 まとめ

本章では双シェマモデルにおける行為シェマの学習に強化学習を導入し，階層的モジュール型強化学習モデル，双シェマモデル強化学習として提案した．強化学習器としては双シェマモデルが連続空間を前提としているために連続時空間の Actor-Critic を用いたが，Q-Learning や SARSA を用いても構わない．この学習モデルでは上位層に相当する行為シェマが環境のダイナミクスとは独立に求められる汎化行為概念をセンサ空間内でのアトラクタとして獲得するために，複数環境ダイナミクス間で利用可能な行為概念を獲得することが出来る．また，双シェマモデルの基本的特徴である，知覚シェマの分化の仕組みと組み合わせることにより，既に過去の経験からシェ

マを形成済みの環境であれば，知覚シェマ選択器が自律的に知覚シェマを選択するだけで，追加の学習なしに新たな環境に再適応できることを示した．また本稿では特に実験は示さなかったが，獲得された知覚シェマは中心への制御を行う行為シェマだけではなく，他の行為シェマによって再利用されることが出来る．双シェマモデルにおいて重要なのは，複数環境において行為シェマ (汎化行為概念) が再利用可能なだけでなく，複数行為間で知覚シェマ (環境との相互作用のモデル) が再利用可能だということである．動的環境下で活動する自律適応系にとって，このような二重の再利用可能性は非常に重要である．

しかし，知覚シェマが均衡化・分化を通して組織化されていくのに対し，行為シェマが強化学習を通じて形成されるプロセスは比較的静的で面白みに欠ける．行為シェマも知覚シェマと同様なプロセスを通して自己組織化的に環境との相互作用を通して形成できないだろうか．次章ではこれに答えるように，行為概念の分化を通して行為シェマを累増的に獲得する手法が導入される．

第 6 章

主観的報酬予測誤差に基づく行為概念の累増的獲得

第 II 部では自らが相互作用している対象の変化に対して概念を累増的に獲得する手法として知覚シェマの累増的シェマモデルを提案した．その最も基本的な形として LDSM を提案した．これに対して，累増的な行為概念の獲得はどのようにして実現されうるのかという疑問は自然と起こる．本章では，このような疑問に対して，強化学習の文脈で累増的な行為概念の獲得を実現する強化学習シェマモデル (RLSM: Reinforcement Learning Schemata Model) を提案する．それを，双シェマモデルにおける行為シェマの適応ダイナミクスとして適用することにより，双シェマモデルは知覚シェマのみならず行為シェマをも累増的に獲得出来るようになる．つまり，前章では行為シェマに均衡化のダイナミクスとして強化学習を導入したが，本章では Fig. 6.1 に示すように，行為シェマに分化のダイナミクスを導入する．

6.1 はじめに

人間との相互作用を前提とした社会ロボットの研究が行われている [17] が，それらの研究において用いられている社会ロボットの多くは設計者に埋め込まれた行動則により動くことしかできない．現在実際に存在する社会ロボットの代表格とも言えるペットロボットを例に取れば，このような適応性の無さが相互作用する人間に漠然とした “飽き” を感じさせていくということを 2 章においても指摘した．そのような中で自律ロボットが環境やユーザとの相互作用を通して累増的に行動則を獲得するこ

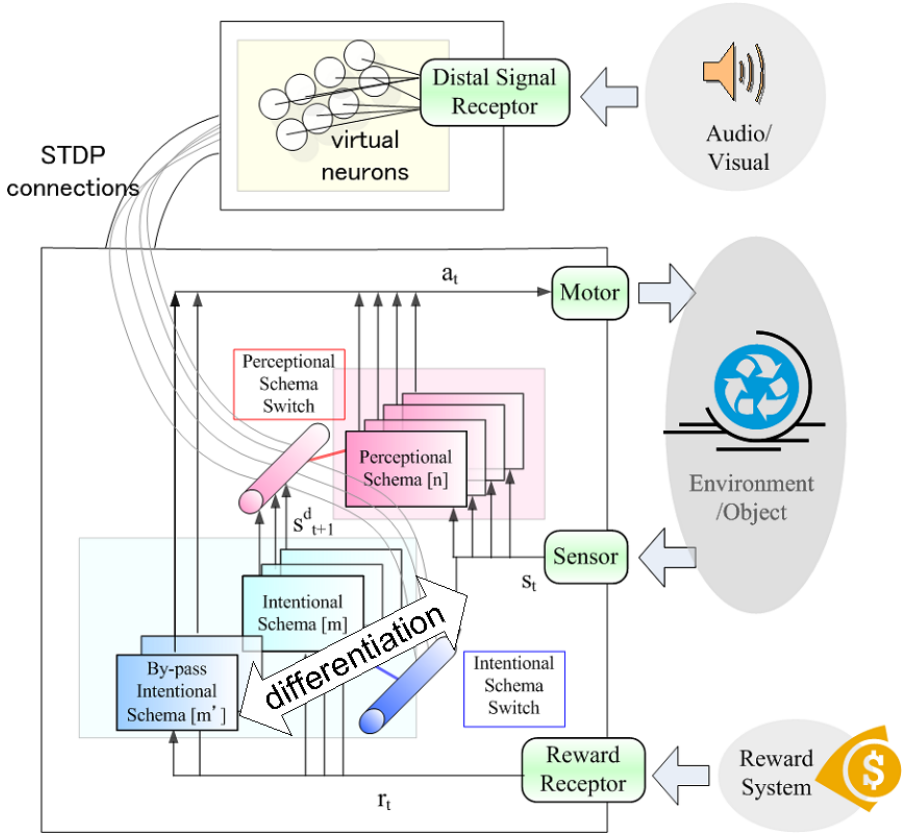


Fig.6.1: Differentiation of intentional schemata in the Dual-Schemata model with a STDP neuronal circuit

とにより，より豊かなコミュニケーションを行うことのできる枠組みが期待されている．

社会的なコミュニケーションでは sign を通して伝えた意図は，何らかの行為を通して実体化されなければ，解釈者は真に発話者の意図を汲み取ったとみなされない．この記号過程において行為として発動される解釈を C.S. Peirce は**活動的解釈項**と呼んだ．ここで，実際に行う行為は伝えられたメッセージの解釈と捉えることが出来る．ここで明らかなのは，社会ロボットが自らの行為を変容させたり，新規に獲得することが出来なければ interpretant の外在化としての活動的解釈項に多様性を持たすことが出来ないということである．ゆえに，コミュニケーションや環境との相互作用を通して社会ロボットが自らの行為を変更していける適応能力は対人間のコミュニケーションを豊かにする必要条件となる．

以上のように人間と社会ロボットのコミュニケーションをより豊かにするという目

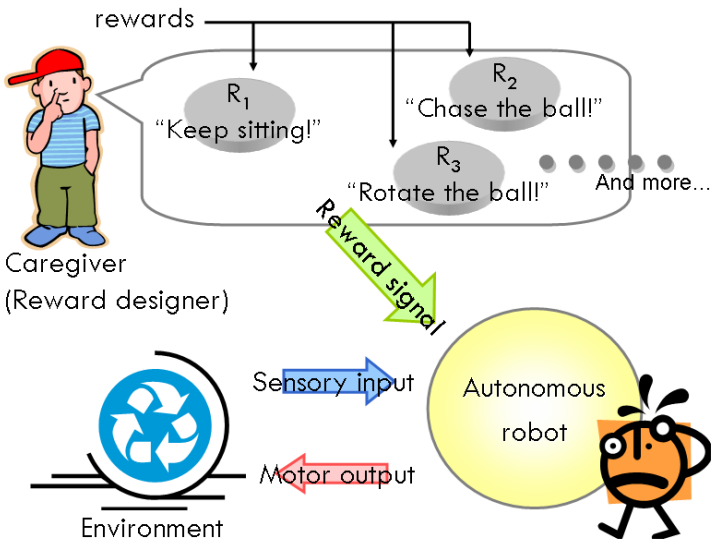


Fig.6.2: Interactions with an autonomous robot emerging by different designs of reward systems

的から，筆者らは環境やユーザとの相互作用を通じて自律ロボットに累増的な行為概念の形成を行わせる方法論の必要性を感じてきた．行為の獲得という点では強化学習における一連の研究 [83, 55] や，模倣学習における研究 [73] などが盛んであるが，それらの多くは一つの行為をいかに効率よく獲得させるかに終始しており，上に述べたような**行動多様性の獲得**を目指すものは少ない．

人間社会の中に参加した自律ロボットを自己閉鎖系として捉える時，自律ロボットは自らの感覚・行為系から得られる情報のみの環境世界 [94] の中に生きる．そのような情報を Fig. 6.2 に示すようにセンサ入力，モータ出力，報酬の三つと捉える事は自律ロボットの学習においては一般的である．この時，報酬はロボットに関わろうとする人間 (caregiver) が意図的に操作することの出来る唯一の情報チャンネルとなる．人間が犬など実際のペットを育成する中で，様々な芸を覚えさせようとするときには，この報酬を通じて状況に応じた様々な行為をペットに獲得させるものと考えられる．人間はそれぞれの芸に対応した報酬系を複数準備し，ペットの試行に対して適宜，報酬や罰を与えている (Fig. 6.2)．しかし実は，ここで重要なのはペット自身がそれぞれの報酬を区別して認識する能力である．もしペットがこれらの報酬を渾然一体として捉えてしまった場合，ペットが認識する報酬は無秩序になり，人間が欲するいずれの行為をも獲得できなくなる．このような視点から，我々は自律ロボットが多

様な行為概念を次々と獲得するためには，人間から与えられる報酬を自律ロボット自身が適宜，対応する記憶構造に同化し，必要であればそれらを分化させ新たな行為概念を獲得するような発達的な構造化能力を持つことが必須であると考えた．

これらを実現する為に，本章では強化学習を行う中で，知覚シェマの学習における主観的予測誤差に相当する主観的報酬予測誤差という値に基づいて行為概念を累増的に獲得させる方法を提案するこれにより自律ロボットは方策を強化学習のプロセスの中で累増的に行為概念を獲得し，それらを状況に応じて使い分けることができるようになる．次節では，強化学習の文脈から本章におけるアプローチの必要性を議論する．次々節ではシェマ理論に基づく行為概念の分化の方法論を導入する．次に単純な Q 学習においてこれを適用した例を簡単に示す．さらにそれを前章で提案した双シェマモデル強化学習に適用した例を示し，最後にまとめる．

6.2　強化学習と複数行為の獲得

本節では次節で提案する手法を従来の強化学習研究と比較するために，複数行為の獲得という視点から関連研究の考察と議論を行う．

6.2.1　強化学習における方策の上書き

報酬を与えることにより，自律ロボットは価値関数を変容させ方策を獲得していく．しかし，多くの研究事例のように目標となるタスクが唯一で，それに対する方策さえ獲得すば終了といったようなゴールありきの発達を実現する自律ロボットならそれで十分であるが，前節で述べたような動的環境下で次々に行為概念を獲得していく自律ロボットのような未来に開いた発達を実現していくシステムの設計論としては問題がある．自律ロボットが方策を獲得した後に，新たな行為を獲得させようと与える報酬系 $r(s_t, a_t)$ を別の報酬系 $r'(s_t, a_t)$ に変化させると，自律ロボットは自らの価値関数 V と方策 π をともに壊したのちに新たな報酬系 r' に対応した価値関数 V' と方策 π' を獲得していかねばならないだけでなく，既に学習済みの方策 π' が方策 π の再獲得を妨げることさえある．その結果，過去に獲得した方策 π は上書きされて忘却されてしまう．その後に再び過去に学習させたタスクを実行させようとしても，またロボットは一から方策 π を学習しなおさなければならない．これはあまりに非効率である．我々人間は様々な行為概念を持ち，一つの身体を通してそれらを状況に応

じた形で発火させることができる．このような問題を乗り越えるためには J. Piaget が提唱したシェマ系 [40] のような分散的な記憶構造が必須であると我々は考えている．しかし，盛んな強化学習研究の中でもこのような一個体による複数行為のオンラインでの獲得についての研究は多くない．本章では，一個体が強化学習をとおして累増的に異なる行為概念を複数獲得し，分散的に記憶領域に保持する為の枠組みを提案する．

6.2.2 強化学習におけるタスク分割と複数行為

異なる行為を複数累増的に獲得させる研究は多くは無いが，一つの大きなタスクを小さな単位に分解し，サブタスクを複数獲得させる研究は多くある．S.P. Singh らはマルコフ決定タスクの複合で表現されるタスクを自律的に複数のサブタスクに分解させ学習させる手法として CQ-Learning を提案した [76, 77]．CQ-Learning は mixture-of-experts[39] の強化学習への拡張と捉えられ，複数の Q-Table の組み合わせで一つの合成タスクを実行することが出来る．現在，このような離散的な状態行動空間を取り扱った研究のみならず，連続状態行動空間への強化学習の適用例も報告されている．連続な状態行動空間では，それらを上手く扱う為に，上位層でタスクを粗く分割し，下位層のサブタスクの複合によりタスクを達成する階層型強化学習のアプローチが多くとられている．森本らは上位層でサブゴールを設計者が与え，下位層で連続空間の強化学習をサブゴール毎に設計された学習器に対し用いることにより 2 リンクロボットの起き上がり動作を獲得させた [55]．杉本らは予測モデルにより状態を分割する強化学習において複数の内部報酬，つまりサブゴールを事前にエージェントに与えることにより，その組み合わせで複雑なタスクを実現させている [155]．このようなアプローチに共通して言えるのは，複数行為を獲得させるのはあくまでその上位にある単一タスクの獲得の為である為に，そのおおもとの報酬系 r が変化すれば，前節で述べたような方策の忘却を起こす点である．これらの研究は，複数行為を獲得してはいるものの，前節で述べた複数行為概念の獲得とは異なる．

6.2.3 複数報酬と複数行為の獲得

では逆に複数行為を環境との相互作用を通して獲得させるにはどうしたらいいか．高橋らは自律的に複数の目標を階層的に自己生成し，並列に学習する枠組みを提唱

している [172]．このように，ロボットが自ら目標を設計し内部生成的な報酬によって学習する枠組みが考えられる．しかし，このような方法では，外部教示者が自律ロボットに取らせたい行動を伝える為の情報チャンネルとしての報酬の持つ役割は無視されざるを得ない．これに対し，強化学習における報酬と獲得行為の一対一対応の関係を前提とすると，複数の報酬のチャンネルを用意しそれぞれについて並列に学習を行わせるといったアプローチが考えられる．これは，数理的には問題ないが，我々の研究の目的である人間機械の報酬を通した社会的相互作用のなかでの複数行為の獲得という局面（Fig. 6.2）を考えると，報酬の種類を前もって獲得させたい行為の種類だけ用意するという考え方は不自然である．よって，本研究では唯一つの報酬のチャンネルを通して，複数の行為を自律ロボットに累増的に獲得させる枠組みについて研究を行った．本章では報酬系変化に対して獲得した行為概念を忘却せずに累増的に行為を獲得していける強化学習の枠組みを提案する．この時，報酬系の変化を明示的にロボットに教えることはしない．あくまで行為概念の新規生成は自律ロボット自身により判断される．この手法は後に示すように行為概念のみならず，双シェマモデル強化学習において定式化された汎化行為概念の累増的獲得にも利用可能である．また，前章で述べたように，行為概念の累増的獲得は双シェマモデルにおける知覚概念の累増的獲得（知覚シェマの分化）と同型的な仕組みで設計される．

6.3　シェマの分化を考慮した強化学習

シェマモデルとは J. Piaget が提唱したシェマシステムに示唆を受けた, 自律ロボットのための時系列情報の組織化手法である．筆者はこれを人間や他の自律適応系の認知発達の基盤となる動的な分散的記憶システムの発達理論と考えている．J. Piaget の理論では感覚運動期といわれる発達段階でのシェマは主に同化 (assimilation)・調節 (accommodation) を通して発達していくが，シェマ単体の同化・調節を均衡化 (equillibration)，そしてシステムレベルでの調節，つまり，新たなシェマの生成と役割分担の決定プロセスを分化 (differentiation) と区別し，これらの二組作用で適応を考える．第 II 部までで提案してきたシェマモデルは，知覚・認識レベルでのシェマ分化を議論した研究であった．これまでの双シェマモデルにおいては知覚の面においてシェマを累増的に獲得されるものの，行為の面においてはシェマを累増的に獲得させることは出来なかった．これに対し，本節では双シェマモデルにおけるシェマの均衡化・分化と同型的なメカニズムを行為概念獲得のための枠組みである強化学習に導入

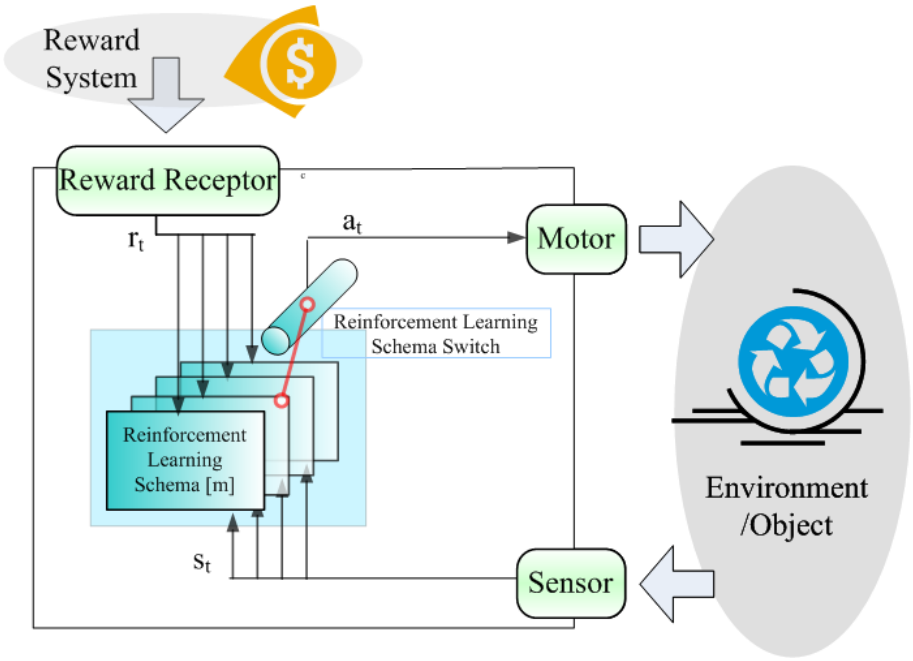


Fig.6.3: RLSM: Reinforcement Learning Schemata Model

することにより，行為レベルでのシェマ分化を実現するシェマモデルを提案する．これを強化学習シェマモデル，RLSM(Reinforcement Learning Schemata Model) と名づけ，その概観を Fig. 6.3 に示す．

6.3.1　強化学習シェマの均衡化

シェマは自らに近い経験のみを同化し，また，同化したものにあわせて自らの内部関数とその内部多様性を調節していく．この繰り返しによって構成されるのが均衡化のプロセスである．このとき同化するかどうかを判断する為に，シェマが自らの内部に獲得された性質と新たな経験とが近いかどうかを測る必要がある．強化学習においては，状態 s_t と行動 a_t の結果として結果 s_{t+1} と報酬 r_t を獲得する．この中で，報酬設計者が想定している獲得させたい行動は一連の報酬 r_t により表現されていると考える．とすれば，r_t に一貫性を求めるのは自然である．よって，シェマは報酬の予測誤差を以って，自らが獲得した性質と，その報酬により表現されている caregiver(Fig. 6.2) が獲得させようとしている行動が，どこまで適合しているのかを評価するものとして定式化するのが自然であろう．ここで報酬の予測誤差を用いるこ

とは，双シェマモデルや MOSAIC，高橋らのモジュール型強化学習において，状態の予測誤差を用いていたのと対照的である．

ところで，この報酬の予測誤差は前節で述べたように，強化学習 [83] における TD 誤差 δ_t に他ならない．TD 誤差は価値関数の誤差であり，TD 誤差とシェマモデルにおける予測誤差を対応付けることによって，強化学習をシェマ単体の基礎ダイナミクスとするシェマモデルを構築することが可能になる．以下，強化学習をシェマ単体の基礎ダイナミクスとするシェマモデルを定式化する．

強化学習を行うシェマをここでは強化学習シェマ (reinforcement learning schema) と呼ぶことにする．また，シェマは並列に複数存在するので，m 番目の強化学習シェマを RS_m と呼ぶことにする．以下の式の添え字 m は RS_m についてのパラメータであることを意味する．

$$R_t^m \equiv \delta_t^m / \hat{\sigma}_t^m \tag{6.1}$$

$$\mathbf{V}_{t+1}^m = p\mathbf{V}_t^m + (1-p)\exp(-\frac{1}{2}(R_{t+1}^m)^2) \tag{6.2}$$

$$= (1-p)\sum_{k=0}^{\infty} p^k \exp(-\frac{1}{2}(R_{t-k}^m)^2) \tag{6.3}$$

を上式で定義する．それぞれ RS_m の主観的 TD 誤差 R^m，強化学習シェマの活性度 $\mathbf{V}^m$ である．また，p はどれだけ過去の認識に固執するかを示すパラメータである．シェマの活性度とは，現在 caregiver が自律ロボットに獲得させようと意図している行動とそれぞれのシェマが受け持っている行動とがどれだけ近いかを表すパラメータである．第 3 章で示したようにシェマ活性度はエージェントによって自発的に生成されたファジィメンバシップ関数として解釈できる．

知覚シェマの時と同様，強化学習シェマにおいてもこのシェマが過去に経験した経験サンプルについての推定偏差 $\hat{\sigma}_t^m$ をどの様に獲得するかが問題になる．強化学習はブートストラップ学習であり，測られる誤差は真の価値関数からの誤差とは言えない．また，Q 学習のような off-policy 型の強化学習ではなく，on-policy 型の Actor-Critic 法や SARSA などでは TD 誤差が方策に依存する為に，$\hat{\sigma}_t^m$ も方策に依存する．よって，$\hat{\sigma}_t^m$ 推定の為の厳密なモデルを立てる事は困難である．よって，本章では，アドホックに経験した TD 誤差の分散を見積もることで推定する．離散状態行動空間における Q 学習については，価値関数の二次統計量を考慮することにより，状態行動対毎に推定偏差を出すモデルを立てる事も可能であるが (補遺参照)，連続空間での Actor-Critic 法や連続空間での Q 学習ではこのような厳密な手法は必ずしも

適用可能ではなく，本章で用いる安価な手法の方が上手く働く場合が多い．

$$\hat{\sigma}_{t+1}^m = \sqrt{\eta(\delta_t^m)^2 + (1-\eta)(\hat{\sigma}_t^m)^2} \tag{6.4}$$

ここで η は報酬予測モデルの学習率であり，同化の判断の為に，同化係数 k を導入する．経験を同化することが出来るのは強化学習シェマ選択器 (reinforcement learning schema switch) により選択されたシェマのみである．選択されたシェマは $R_t^m \leq k$ を満たす経験のみを同化する．選択の仕方については次節で説明する．同化された経験はシェマの調節に利用される．調節は具体的には，同化された経験を用いた強化学習（つまり価値関数と方策の更新）と上に示した推定偏差 $\hat{\sigma}_t^m$ の更新からなる．本章では簡単に時系列的に決定するスカラー値として推定偏差を扱う．同化しない場合，つまり，TD 誤差の絶対値が今まで TD 誤差に比べて大きすぎる報酬は，外れ値として扱い強化学習に利用しない．

6.3.2 強化学習シェマの選択と分化

自律ロボットは環境と相互作用する中でセンサ入力 s_t とモータ出力 a_t，および報酬 r_t を経験するが，それをそれぞれの強化学習シェマの内部でコヒーレントに同化することにより，既に獲得した行為概念を破壊せずに新たな行動を獲得することが出来る．そのような仕組みを導入することで報酬系の変化による累増的な行為概念の獲得が可能になる．その為，自律ロボットは双シェマモデルにおける知覚シェマの選択則と同様のルールで，強化学習シェマをシェマ活性度 $\mathbf{V}_t^m$ を基準にして選択する．これにより，エージェントは外部からの報酬の変化を自律的に読み取り，新たな行為概念を累増的に獲得していくことが出来るようになる．

$$\mu(H^m) = sgn(\mathbf{V}^m - \mathbf{V}_\alpha) \tag{6.5}$$

$$P(m) = P(\bigwedge_{l=0}^{m-1} \overline{H^l} \wedge H^m) = \mu(H^m) \prod_{l=0}^{n-1} \mu(\overline{H^l}) \tag{6.6}$$

現存するシェマすべてが $P = 0$ となった時に，新たなシェマを生成し，その σ を十分大きい値に設定する．ここで，H^m は m 番目の強化学習シェマが現在の環境に相反していないという仮説であり，$\bar{H}^m$ はその論理否定である．μ はその真理値関数である．$P(m)$ は m 番目の強化学習シェマの選択確率になる．0 番目のシェマは式を簡単にするためのヌルシェマ (null-schema) であり常に $\mu(H^0) = 0$ となる．本稿では $\mathbf{V}_\alpha \equiv \exp(-\frac{1}{2}(k^2)$ に設定する．これは誤差が正規分布すると仮定した場合，偏差

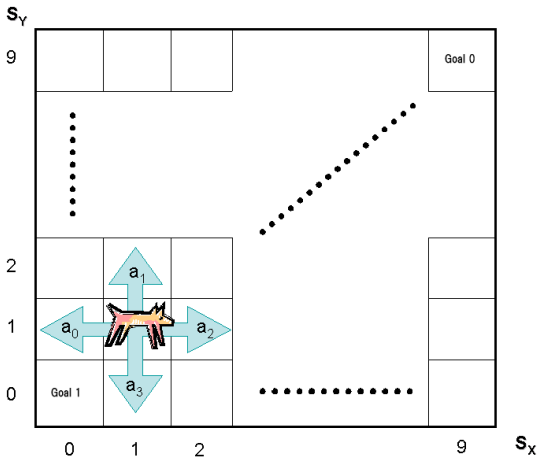


Fig.6.4: The grid world and the agent in the experiment 1

1 に正規化された正規分布で $R > k$ が起こり続けた場合に分化するという規範である．これらの枠組みは Actor-Critic 法や Q 学習といった強化学習の具体的な立式に依存しない．TD 誤差をベースに学習を行うあらゆる TD 学習に基本的には適用できる枠組みである．実際，次節の実験では Q 学習に適用し，その次に我々の提案している双シェマモデルを用いた階層的モジュール型強化学習に適用する．後者は本質的には Actor-Critic 学習により構成されている．

6.4　実験 1 :Q-table の累増的獲得

本節では単純な Grid 空間における Q 学習にシェマ分化を考慮した強化学習の枠組みを適用した実験とその結果を示す．

6.4.1　実験環境と課題

Fig. 6.4 に示すような単純な Grid 空間を考える．状態は縦横 10 のグリッドに仕切られている．また，行動空間は Fig. 6.4 に示すように上下左右の 4 方向への移動へと割り当てる．上下左右の終端において，それより外に移動する行動をとった場合は移動せず同じ場所にとどまることにする．強化学習機構としては Q 学習を用いた．Q 学習の詳細については省略するが，行動選択にはボルツマン選択を用い，逆温度定

数 β の制御にはエピソード毎の平均報酬 $r_{average}$ を用いて制御した．

$$\beta = (1-\zeta)\beta + \zeta \exp(r_{average}) \tag{6.7}$$

定数 ζ は 0.01 に設定した．逆温度ははじめは 0 に設定する．シェマが分化したときにも新たなシェマに対して逆温度を 0 に設定する．また，学習には適格度トレースは用いない．

一つ目のタスクとして右上に行くほど多くの報酬を与える報酬系 r^1 を，二つ目のタスクとして左下に行くほど多くの報酬を与える報酬系 r^2 を準備した．ともに報酬の最大値は 1 になるようにしてある．

$$r^1 = \frac{s_x + s_y + 2}{20} \tag{6.8}$$

$$r^2 = 1 - \frac{s_x + s_y}{20} \tag{6.9}$$

1 episode を 100step として実験を行った．episode の開始時点ではエージェントを Grid にランダムに初期配置する．はじめ報酬系は r^1 からはじめ 500episodes 毎に二つの報酬系を切り替えた．通常の強化学習の場合，報酬系を切り替えると獲得した価値関数と方策を一度崩して，再び再学習しなおさなければならない為，獲得報酬は振動的に推移することが予想される．合計 2500episodes を行った後に終了した．

6.4.2　実験結果

実験の結果として通常の強化学習とシェマ分化を考慮した強化学習のエピソード毎の報酬和を比較したものを Fig. 6.5 に示す．それぞれ 5 回試行を行い，その平均をとっている．また，図を見やすくする為に 10 区間の移動平均をとり平滑化している．また，このときの強化学習シェマの活性度 $\mathbf{V}_m$ の変化の一例を Fig. 6.6 に示す．500episodes が終了した時点で新たなシェマが分化により作り出され，それ以降は報酬系が切り替わるごとにそれぞれを担当するシェマが活性化し役割分担をが自律的に行われていることがわかる．また，報酬については，単純な強化学習器の性能が報酬系が変わる毎に激しく悪化しているのに対し，シェマの選択切り替えに要する数 episodes を除き，ほとんど再学習の必要なく，獲得した行為を切り替えて出力していることがわかる．全エピソードでの報酬の平均は単純な強化学習で 68.6，シェマ分化を考慮した強化学習で 80.8 となった．このタスクでは最適な場合に得られる全エピソードでの平均報酬は 97.3 と期待され，ランダムに行動しても約 50 の平均報酬を手

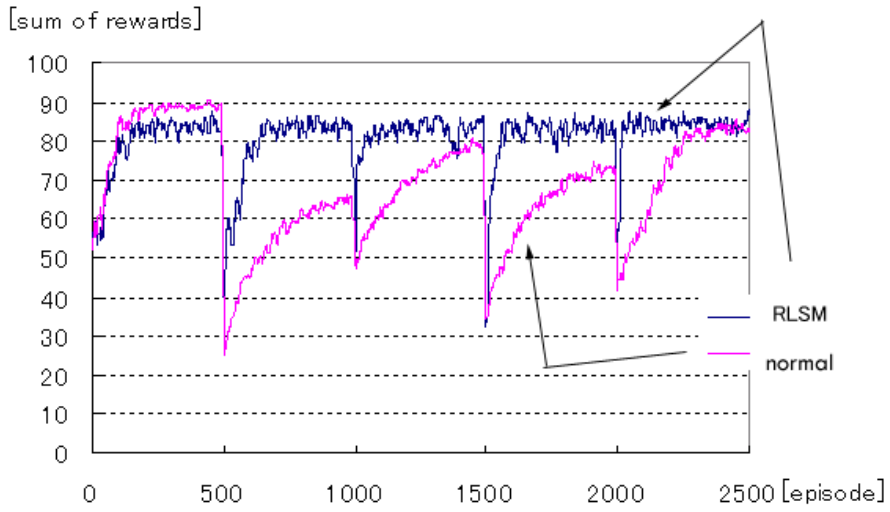


Fig.6.5: Transitions of the sum ofrewards in each episode

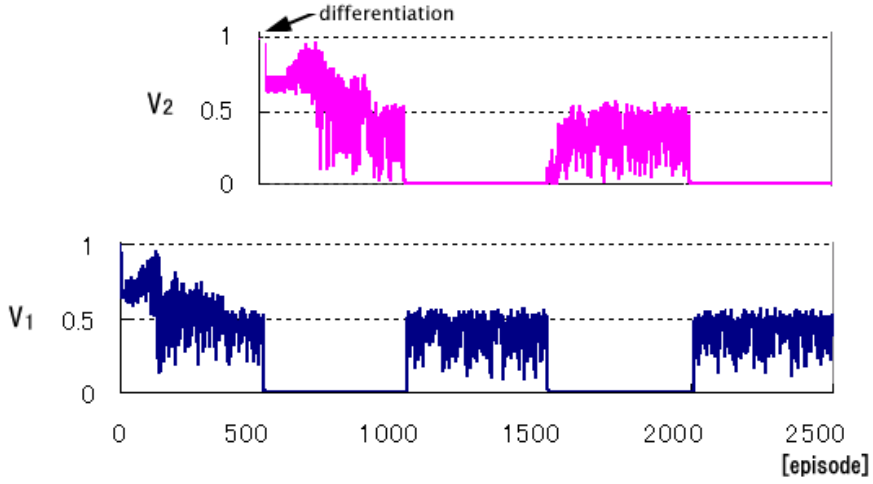


Fig.6.6: Transitions of the activities of the reinforcement learning schemata

に入れることが出来る．これらを考慮に入れると，このように報酬系が切り替わる系において，これら二つのスキームの性能差は十分に優位であると考えられる．

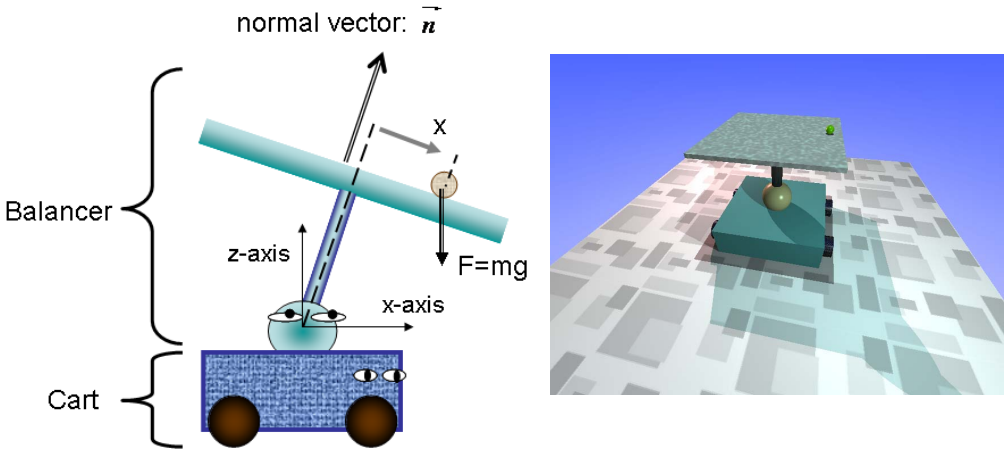


Fig.6.7: An overview of the agents used in the experiments

6.5　実験２：汎化行為概念の累増的獲得

本節では上のような枠組みを使って，強化学習の方策として表現される行為概念だけでなく，第５章において導入された双シェマモデルにおいて行為シェマとして表現される汎化行為概念 (generalized behavior concept) をも累増的に獲得することが出来ることを示す．

6.5.1　実験環境

前章で提案した双シェマモデル強化学習は，基本的に TD 学習である Actor-Critic 法に則っているので，本稿で提案した累増型の強化学習手法が利用可能である．実験環境としては第５章で用いたものと同じく自分の加速や移動をする台車エージェントに別のエージェントが乗り，その台車の上で台車上のエージェントが平板上を転がるボールを平板の角度を変化させることにより制御するとい パラメータや状態空間の設定は第 5 章における設定と同じなので，説明を省略する．

複数行為概念の累増的獲得を問題にするわけだが，獲得させる行為の一つは前章で獲得させたものとと同じ平板の中心位置へのボール制御であり，もう一つは前章の最後に言及した平板の上で平面の中央で約半径 0.5[m] の円を描かせるという動的なタスクである．このような運動は xy の直交座標系では最適な方策が循環的な軌跡を描くので，K. Doya らのモジュール型強化学習など (MMRL) で利用される状態価値関

Table6.1: Intentional schemata

Index	Name	Policy
IS_0^-	move random	$a_t = G_0^- \equiv \mathbf{n_t}$ ($\boldsymbol{n_t}$ is a noise term)
IS_1	shake a ball	$s_{t+1}^d = G_1(s_t) \equiv \mathbf{n_t}$
IS_2	——	$s_{t+1}^d = G_2(s_t)$

数 V の傾斜に基づいた状態価値勾配法 (value gradient based policy)[25, 23] を基本にした方法論では学習できないタイプの問題である．

ボールを中心位置への静止させることを学ばせる為の報酬系を報酬系 A として以下のように設定する．

$$r_t = \exp(-\frac{1}{2}((\frac{x}{0.5})^2 + (\frac{y}{0.5})^2)) \tag{6.10}$$

次に円運動を獲得させる為の報酬系を報酬系 B として定義する．基本的には平板の中心周りの角速度 ω を ± 1 で飽和させた項と半径 0.5 の円軌道上で最大となる項の積で構成されている．

$$r_t = \max(-1.0, \min(1.0, \omega)) \exp(-\frac{1}{2}(\frac{\sqrt{x^2+y^2} - 0.5}{0.3})^2) \tag{6.11}$$

また，通常の強化学習にくらべて，前章で導入した双シェマモデル強化学習の有利な点は，環境ダイナミクス変化が変化した際にその変化を知覚シェマの均衡化もしくは，分化によって生成された複数の知覚シェマによる分担によって吸収する点にある．この知覚シェマ分化による環境ダイナミクスの変化に対する追従能力の有利さも同時に検証するため，以下のような複数の状況を用意する．

1. 状況 A：Cart は停止している．$(\alpha = (0, 0))$
2. 状況 B：Cart が等速で約 10° の坂を上り始める．$(\alpha = (0, -2))$
3. 状況 C：Cart は半径 8[m] の円周上を秒速 4[m/s]（時速 14.4[km/h]）で周回する．$(\alpha = (2, 0))$

行為シェマは初期状態では表 1 に示す 3 つのものを用意する．双シェマモデルでは行為シェマの強化学習の為の探索のみならず，知覚シェマのモデル学習の為の探索も行

Table6.2: The scenario of the experiment

時間 [s]	状況	報酬系
0-5000	A：カート停止	A：中心に制御
5000-7500	B：坂道上昇	A ：中心に制御
7500-10000	C：周回運動	A ：中心に制御
10000-18000	A：カート停止	B：ボールを回す
18000-20000	B：坂道上昇	B：ボールを回す
20000-22000	C：周回運動	B：ボールを回す
22000-24000	A：カート停止	B：ボールを回す
24000-27000	A：カート停止	A：中心に制御
27000-30000	A：カート停止	B ：ボールを回す

う必要があり，複数の行為シェマを切り替えながら学習することが重要である．よって，外部報酬系に対して，最適な行為シェマを選択する場合と，探索的に行為シェマを選択する場合を作る．具体的には一定時間ずつ $IS_0^{\times}$ と IS_2 をとった後に，報酬従属モードをとるということを繰り返す．報酬従属モードとは，外部から入る報酬によって変動するシェマ活性度に従って強化学習シェマの選択を行う，いわば最適な行為シェマを選択するモードである．また，$IS_0^{\times}$，IS_1 はすでに完全に収束した行為とみなし，両方 $\hat{\sigma} = 0$ という設定で行う．

6.5.2 実験の為の環境設計

この実験に用いる報酬系を含めた環境の変化は表２のように設定した．一定環境下での学習を取り扱う従来の強化学習研究などに比べて，本研究ではどのように環境の変化という時系列情報が如何に複数の強化学習器として組織化されていくかという点が問題点となる．つまり時系列的な環境設計 [177] の側面が重要であり，その環境変化をシナリオとして設計することにより，そのシナリオ下での自律ロボットの環境適応性を知ることが出来る．

前章における実験と同じく，はじめの 10000[s] で環境のダイナミクスを変えながら中心への位置制御を学習する．この過程において知覚シェマの分化が進むと同時に一つ目の行為シェマの形成が進む．次に報酬系を切り替えて 10000[s]-18000[s] の間

で新たに円運動を学習させる．この後に再びダイナミクス（状況）を 3 回変化させるその後 24000[s] 以降では状況 A に状況を固定し，再び報酬系を切り替えエージェントの振る舞いを観察する．また，このとき行為シェマの同化係数は $k = 2.4$ に設定した．

6.5.3　実験の結果

前節で述べたように実験は表 2 のような環境下で行った．その結果，行為シェマの活性度は Fig. 6.8 のように変化した．図に示すように，行為シェマは 10000[s] において報酬系が A から B に変化した時点で分化し，新たな行為シェマ IS_3 を生成している．図中の行為シェマ活性度のグラフは 8 回同様の実験を行った時の平均値である．知覚シェマの分化については図を省略するが，前章における実験と同様に初め 5000[s] と 7500[s] に於いて状況を A → B → C と変化させた時に，環境ダイナミクスの変化に自発的に気づき，知覚シェマを分化させた．その後は状況が変わるたびに獲得した 3 種類の知覚シェマが活性化し，シェマ選択則に従い選択することにより，環境ダイナミクスの変化に対しては柔軟に対応することが出来た． 8 回の実験について平均をとれば上のようになった．しかし，行為シェマの分化自体には全ての試行で成功したものの，実験 8 回の内，半数ほどは 24000[s] に於いて報酬系を A に戻した際に IS_2 を再選択することをしなかった．つまり，想起には失敗した．想起とはある状況もしくは報酬系に対して生成・分化したシェマが，同様の状況において再選択されることをさす．なぜ想起に失敗したかについては後で考察する．また，想起に成功した 4 つの実験結果についてその報酬の重み付き移動平均をとり Fig. 6.9 に示す．また，同時に環境の状況，caregiver が与えた報酬系についても示している．更に，それらの変化に対して自律ロボットがどのシェマを選択したかを示す．通常の強化学習では方策の上書きにより，報酬系についても環境のダイナミクスについても，どちらかが変化した時点で報酬平均は急激に降下するが，本手法では環境ダイナミクスや報酬系の変化に自発的に気づきシェマを分化させることにより，過去の方策を分散的に保持する為に一度学習した環境ダイナミクスや報酬系に関してはシェマの切り替えのみで対処することが出来る．さらに，約 15000[s] において，二つの行為シェマ（中心静止と円運動）と三つの知覚シェマ（等速運動時と坂道登坂時と周回運動時 ）を全て獲得し終えた後の，(状況, 報酬) $= (B, B), (C, B)$ については，その組み合わせとしては新規な環境であるにもかかわらず，過去に獲得された行為シェマと知覚シェマを

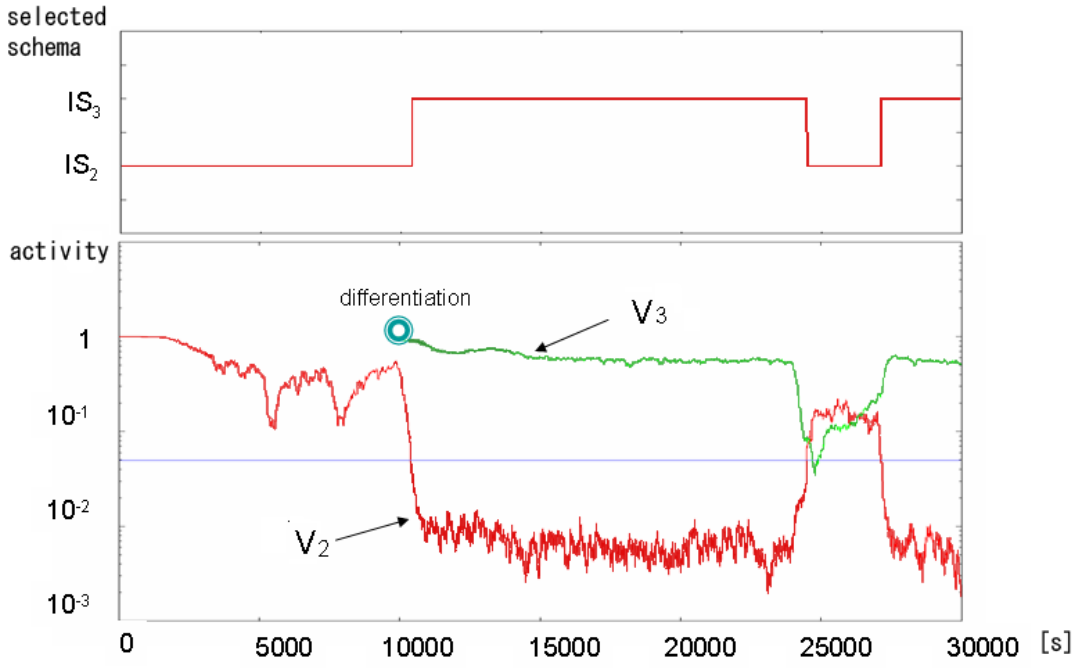


Fig.6.8: Top: the selected intentional schema, bottom: time course of the intentional schemata activities

用いることにより実行可能であることが分かる．

6.5.4　考察

まず，前節の実験のうちの数回に於いて想起に失敗した理由について考察する．この理由は行為シェマ活性度変化を起こしている TD 誤差が Actor-Critic 手法では方策依存 (on-policy) となってしまうために [83]，異なる種類の方策を用いていた場合，正確な TD 誤差が強く方策の変化に依存することに起因する．報酬系を A に戻したときには方策は報酬系 B に対応した方策が選択されており，その影響から報酬系 A に相当する方策をとっていた場合よりも報酬系 A に対応するシェマが多くの報酬予測誤差を出してしまったことによる．この誤差のために，IS_2 が同化すべきと想定された報酬を同化しなかったものと考えられる．この問題を克服する為には Q 学習などの価値関数が方策に依存しない強化学習 (off-policy) を行為シェマの適応メカニズムとして用いる必要がある．しかし，Actor-Critic 法に比べ Q 学習は連続系には適用

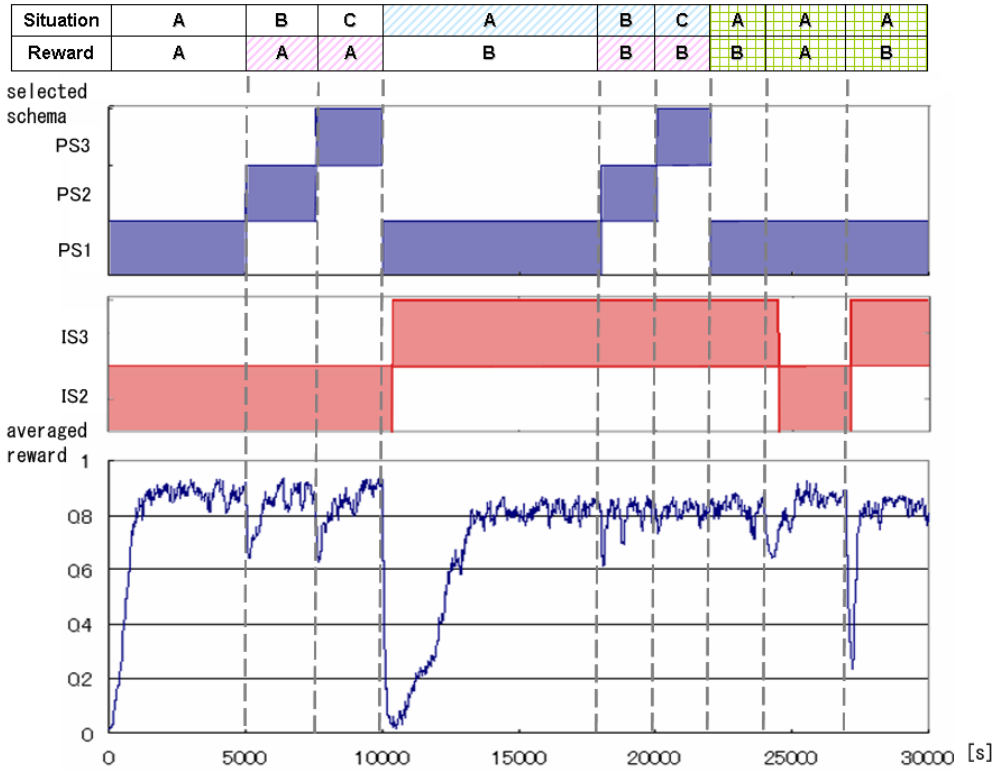


Fig.6.9: From top to bottom 1）the agent's situations and the reward systems given by the caregiver. The shaded areas denote that either its reward system or situation has been experienced. The plaided areas denote both its reward system and situation has been experienced. 2）the selected perceptional schema; 3）the selected intentional schema; 4）time course of the smoothed reward.

しにくいという欠点がある[*1]．よって，Q 学習の効率のよい連続系の拡張を考え適用するか，もしくは Actor-Critic 法を用いた際の上のような方策変化による TD 誤差を緩和する機構を導入することが求められる．D. Precup らは重点サンプリングを用いることでエピソード単位で更新を行う類の強化学習では on-policy 学習を off-policy 化できることを示している [66]．内部らは CLIS[93] の枠組みにおいて，重点サンプリング法を用いて on-policy 強化学習器である複数の Actor-Critic 法や SARSA を

[*1] 連続空間には空間をグリッドで分割する代わりに，状態行動空間上に動的基底核などの基底群を配置し Q 関数を近似するという方法が取れるが，これでは Q 学習の長所である off-policy 性や収束性が保証されなくなる欠点がある．residual gradient[9] や重点サンプリング法 [66] によって解決する面もあるが，学習速度や安定性の面で従来の Q 学習の良さを失ってしまう．

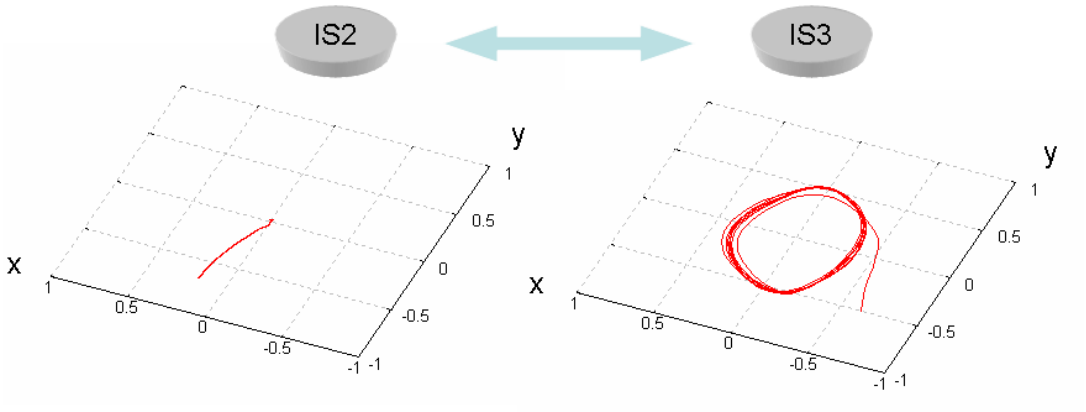


Fig.6.10: Change of behaviors corresponding to the changes of reward systems

用いた学習器を並列学習させられることを示している．この方法論を用いることによっても上の問題を解決できる可能性がある．

また，自律ロボットが環境や caregiver との相互作用を通して活動する中で，累増的に多様な行動を獲得することができたが，さらに，各行為概念を獲得させた後には，別の行動をとっている自律ロボットに対して学習済みの報酬を短い時間与えるだけで，caregiver が意図した行動に行動を切り替えさせることが出来た．Fig. 6.10 に報酬系を切り替えたときに観測された板の上のボールの軌跡を示す．実際，一度目の学習に於いて与える報酬の量，つまり一度目に目的の行動をとるようになるまでの報酬を与える時間に比べて，想起の時点で目的の行動をとるようになるまでに再び報酬を与える時間は非常に少ない (Fig. 6.9)．報酬のこのような想起を促すという作用は，一般の強化学習では現れてこないが，報酬の新たな役割として興味深い．このように利用されうる報酬という情報の流れは人間と自律ロボットとの相互作用における新たな手段となる可能性を持っている．つまり，新たな行為を獲得させるためだけでなく，一度獲得させた行為を引き出すための情報として利用可能になる．しかし，本章では人間との相互作用を通した行為創発により生み出される行動多様性や，報酬系による想起の機能が人間機械コミュニケーションにどのような影響を与えることができるかについてはまでは議論を進めることは出来なかった．しかし，我々はこのような仕組みがより豊かな人間機械コミュニケーションを生み出して行く基盤の一つになると考えている．

本章で提案したシステムの問題点としては，この様な報酬系を含めた力学的な相互作用の中で行為シェマ，強化学習シェマの切り替えを適切に行おうとすると，必然的

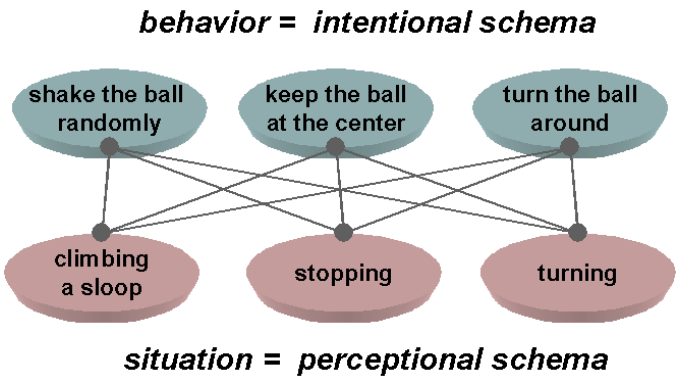


Fig.6.11: Compositionality of perceptional and intentional schemata

にシェマ切り替えに時間遅れが付きまとう．これを解決することが記号の存在理由の一つではないかと考えられる．この動機をもとに第８章ではこのシェマ切り替えを可塑的に補助する機構として神経回路網の STDP 学習則に注目し記号創発の機構を提案する．

6.6　まとめ

本章ではシェマの均衡化・分化の機構を強化学習に導入することにより，累増的に学習器を配置していくことのできる強化学習シェマモデル (RLSM) を提案した．また，双シェマモデルにおける行為シェマ強化学習に分化の機構を加え，累増的に行為シェマを獲得することのできる仕組みを提案した．この手法は離散値の Q 学習や双シェマモデル強化学習のみならず，あらゆる TD 型の強化学習に用いることの出来る手法である (ただし on-policy である場合には想起性能が落ちる)．これに加えて，双シェマモデルにおいては，知覚シェマの分化機構との相乗効果により，その恩恵を通常以上に享受することが出来る．双シェマモデルでは上位層に相当する行為シェマが環境のダイナミクスとは独立に求められる行為をセンサ空間内でのアトラクタとして獲得されるために，複数環境ダイナミクス間で利用可能な行為概念を累増的獲得することが出来る．また，双シェマモデルの基本的特徴である，知覚シェマの分化の仕組みと組み合わせることにより，既に過去の経験からシェマを形成済みの環境であれば，知覚シェマ選択器が自律的に知覚シェマを選択するだけで，追加の学習なしに新たな環境に再適応できることを示した．これは行為概念の環境概念に対する汎化性

と，環境概念の行為概念に対する汎化性という双方向的な汎化性の実現を意味している．また，このような双方向的な汎化性故に，これらのシェマは Fig. 6.11 に示すように，基本的には自由に接続することが出来る．この結合可能性 (compositionality) は双シェマモデルの基本的性質である．認知的なレベルにおけるモジュール間の結合可能性と，原初的な言語における文法の始まりとの関係を考える上でもこの結果は興味深い．我々は基礎単位を担うモジュールを行為シェマ，知覚シェマとして埋め込み，その種類を累増的に獲得させたが，Y. Sugita らは逆に言語の構造と認知におけるダイナミクスの構造の同型性に注目し，この結合可能性を RNN 内部に分散的に獲得させている [81].

さらに，環境ダイナミクスの変化とは別に，外部から与える報酬の与え方の変化に自律的に気づき，行為シェマを分化させる機構を加えることにより，複数の行為概念を累増的に獲得出来ることを示した．これによって，双シェマモデルは知覚シェマ，行為シェマの両シェマを自律的かつ累増的に獲得することが出来る枠組みとなった．これで，第 II 部および第 III 部を通して導入を行ってきた双シェマモデルについての議論は一段落となる．

補遺：Q 学習におけるより正確な推定偏差の獲得

本章では TD 学習全般に対して定式化を行うために，推定偏差の学習については，アドホックに重み付き時系列移動平均的な近似を与えることとしてきた．しかし，Q 学習の場合に限定すれば，より，安定した推定法が存在する．補遺として，その手法について付記する．

基本的な考え方としては行動価値関数 Q の二次統計量 $Q^{(2)}$ を TD 推定し，二次統計量 $Q^{(2)}$ から一次統計量 Q の二乗値を引き平方根をとることにより $\hat{\sigma}(s_t, a_t)$ を求める．

$$\delta_t = r_t + \gamma V(s_{t+1}) - Q(s_t, a_t) \tag{6.12}$$

$$\begin{aligned}\delta_t^{(2)} &= r_t^2 + 2\gamma r_t V(s_{t+1}) + \gamma^2 Q^{(2)}(s_{t+1}, a_{t+1}^*) \\ &\quad - Q^{(2)}(s_t, a_t)\end{aligned} \tag{6.13}$$

$$V(s_t) = Q(s_t, a_t^*) \tag{6.14}$$

$$a_t^* = \arg\max_a Q(s_t, a) \tag{6.15}$$

ここで γ は割引率である．それぞれの関数はこれらの誤差を用いて

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\delta_t \tag{6.16}$$

$$Q^{(2)}(s_t, a_t) \leftarrow Q^{(2)}(s_t, a_t) + \alpha\delta_t^{(2)} \tag{6.17}$$

$$\hat{\sigma} = \sqrt{Q^{(2)} - Q^2} \tag{6.18}$$

が得られる．これにより，$\hat{\sigma}$ を各状態行動対に対して求めることが出来る．この Q 値を確率変数とみなし，その二次統計量を求めることで推定偏差を得るアイデアは [21] に拠っている．強化学習シェマモデルは第 8 章においても再び取り上げるが，その時にはこの手法を用いて推定偏差を求める．

第 IV 部

構成論的記号論の脳神経科学的基盤とシナプス可塑性を通じた記号創発

第 7 章

シェマモデルに基づいたモジュール型非線形内部モデルの累増的獲得

第 II 部，第 III 部と自律適応系内部における内的表象生成がどのように構成論的に表現できるのかという点から，自らの身体を用いた環境との相互作用において，その環境下でより良く振舞っていくための学習を通して時系列情報組織化する機構，シェマモデルを提案してきた．それらはただ，時系列情報を自己組織的にクラスタリングするのみならず，それを通して身体的を用いた動的環境下でのタスク性能を向上させるということも同時に示してきた．さて，実際にこの様な計算論は人間の相互作用を通じた環境認識の発生を説明しうるのだろうか．元来，シェマとは幼児における認識の発生を議論するための道具であった．その考えを踏まえ,「静止球」「円運動球」「坂道」などの状況や対象を指示する対象概念としての知覚シェマ,「追跡する」「円運動させる」などの行為概念としての行為シェマを組織化することが出来る計算論的枠組みを既に提示してきた．しかし，そのような認識生成のダイナミクスが実際に人間の内部で駆動している保障は無い．

これに対し, 計算論的神経科学 (computational neuroscience) ではより具体的な脳科学の立場からどの脳部位がどのような計算を行っているかという議論が展開されている．第 IV 部では我々のシェマモデルと脳科学の関係を議論する．まず，第 7 章となる本章では小脳の複数内部モデル仮説をとりあげ，その計算論的モデルとして，第４章で提案した LDSM における知覚シェマを非線形系に拡張した NGSM(Schemata Model with a Normalized Gaussian network) が現在提案されている MOSAIC の計算モデルよりも適切なモデルで有り得る事を示す．ここでは，シェマモデルで議論

したような対象概念の生成が実際に人間の認知系において行われている可能性を脳科学研究の知見との結合により示唆する[*1].

また同時に，本章では統計学的視点から，仮説検定とシェマモデルの関連性を捉えながら，シェマモデルの作動原理を仮説検定理論を援用することで明確にしていく．前章までのシェマモデルの定式化はアドホックな面を常に含んでいたが，この定式化によりシェマモデルの数理的基盤が明確化される．

7.1　はじめに

近年，小脳皮質において，環境や扱う道具のダイナミクスについての内部モデルが獲得されるという，所謂内部モデル仮説が信じられてきている．また，それに加え，内部モデルがモジュール状の構造で並列に複数獲得されているという考え方が支配的になってきている．この内部モデルは計算論的には双シェマモデルにおける知覚シェマの内部関数に等しい．D.M. Wolpert らは “the Modular Selection and Identification for Control (MOSAIC)” と呼ばれるモデルを提案した [103, 102]. fMRI を用いた生理実験においても，この考え方を指示するような知見が得られている [38, 37, 36]．一方で，このモジュール型の内部モデルを理解するための計算論的なモデルも同時に提案されている [103, 25, 33]．計算論的なアプローチにおいて問題となるのは，小脳がどのように内部モデルを獲得し，また，それをさまざまな制御タスクにおいてどのように利用しているかである．内部モデルの利用という点では，一連のモジュール型内部モデルについての計算論的な研究において，幾つかの有用な利用方法が提案されてきた．特に二つの基本的な考え方が提案されている．ひとつは Multiple Paired Forward-Inverse Model (MPFIM) [103] であり，もう一つは Multiple Model based Reinforcement Learning (MMRL) [25] である． MPFIM では順モデルと逆モデルという二種類の内部モデルが獲得される．一度，適切な複数の逆モデルが獲得されると，人間はそれらを適宜選択して用いることにより，望みの状態を達成できるようになる．一方，MMRL においては，強化学習もしくは動的計画

[*1] 本研究における「自律適応系」という抽象的なシステム論的枠組みの中では，脳神経系という物質的な接地点を不可欠な要素として求めることはしない．なぜなら，人工生命が地球上の炭素ベースの生命以外の生命の可能性を探し，また人工知能が人間以外の知能の可能性を探したように，我々はより抽象化された自律適応系における発達的機能構成の枠組みを求めているからである．しかし，その具体的な一例としての人間の認知系の理解には，脳神経系の理解が重要であるのも事実である．

法を基礎においており，報酬関数が局所的に二次形式で与えられると仮定すると，それぞれの順モデルと報酬関数の対に対して，局所的には最適な制御器を獲得することができる．K. Doya らは，MMRL がどのように小脳と大脳基底核に渡って実現されているかについての概観を説明している [22]． これらの複数の順モデルについての利用方法は様々なシミュレーションタスクで上手く動くことが報告されているのである程度合理的な計算論モデルであるといえる．ただし，MMRL については局所的な問題は最適に解決するものの，大域的な制御問題には言及できないために，それを解決するための階層的な手法が提案されている．[155]

しかし，もし複数の順モデルが適切に獲得され，選択されることが保障されなければ，これらの計算論的な複数順モデルの利用方法は全く使えないにもかかわらず，MOSAIC において提案されている複数順モデルの獲得手法は実は不安定であることが明らかになってきている [193].

一方，多くの MOSAIC についての計算論的な研究において，それぞれの順モデルは線形と定義されている [25, 33]． 次章で議論するように，この線形の仮定は非常に合理的でかつ有用であり，時には必要条件となる．D.M. Wolpert らは勾配法にのっとって複数順モデルをオンライン学習し，モジュール選択はベイズ推定によって行うモデルを提案しているが [103]，K. Doya らはこの定式化に空間局在性 (spatial locality) と時間連続性 (temporal continuity) という二つの概念を追加することにより，非線形で時変な系でも学習が不安定になりにくくするための補正を行った [25]．しかし，それでも MOSAIC の学習則は非線形で時変な系において安定して良い学習結果を生み出さないことが多い．その学習結果は各モジュールの初期状態に強く依存し，もしランダムな初期値から始めると，殆どの学習結果は多く存在する局所最適値に落ちてしまう．これは学習器としては致命的な欠点である．

この不安定さに加え，MOSAIC の機構はもう一つの問題を持っている，それは順モデルの数が学習の開始以前に決定されて，学習プロセスを通して不変であるという点である．人々は新しい道具を使い，それに相当する新たな内部モデルを次々と累増的に獲得していく．しかし，MOSAIC はそのような内部モデルの累増的な獲得を計算論的に表現できない．しばしば，冗長なモジュールをもとより準備しておけば，その余ったモジュールを新たな道具に割り当てることにより，累増的な獲得を表現できるという考えが示されるが，過剰なモジュールの存在は MOSAIC の学習能力を下げるので，この考えは適切とは言えない [25]．ゆえに，新たなモジュールを必要に応じて割り当てる，もしくは生成するという考え方は重要である．

これら，モジュールの初期値依存性と累増的獲得能力の欠如を克服するために，本章では新しいモジュール型学習機構を提案する．本章で提案する “Schema Model with a Normalized Gaussian network (NGSM)”(正規化ガウスネットワークを用いたシェマモデル) は二つの基本的な理念に基づいている．一つは，モジュールの選択と生成を仮説検定に基づいて行うという点である．MOSAIC や他の多くのモジュール型学習機構がベイズ推定に基づいて選択を行うのに対して対照的である．もう一つは，空間方向のモジュールと時間方向のモジュールを明示的に区別するという点である．これは LDSM で線形に限定した知覚シェマの非線形系への拡張となっている．

次節では空間方向のモジュール分割と時間方向のモジュール分割の違いについて議論する．次々節ではシェマモデルに基づいて NGSM を導入する．最後に，非線形時変な系において NGSM を実装したエージェントに複数の順モデルを獲得させる実験を行い，その有用性を検証する．

7.2　空間方向のモジュールと時間方向のモジュール

時変で非線形な系を学習・制御対象とする場合には二種類のモジュール分割が考えられる．それは，空間方向のモジュールと時間方向のモジュールである．時変といえども，完全に無秩序な場合は一貫した学習が完全に不可能なので，道具を持ち替えたときにダイナミクスが変わるといったような，一定時間続く複数の定常状態，つまり時不変な系が切り替わっていくような系を考える．これら二つのモジュール分割は MOSAIC では特に区別されていなかった．しかし，本研究ではこの二つが定性的にかなり異なる由来を持っていると考えている．故に，NGSM においては階層的にこれらを区別して管理することを考える．

7.2.1　空間方向のモジュール：非線形ダイナミクスの線形ダイナミクスへの分解

非線形なダイナミクスは制御において非常に取り扱いにくい．故に，制御理論では非線形系はしばしば局所的な線形系に分解されて扱われることが多い．これは線形系が制御を簡単にする様々な特質を持っているためである．故に，大域的な非線形系を局所的に線形モデルによって近似し，複数の線形の内部モデルとして獲得するのは有用であり，合理的である． また，実際に人間を用いた実験においても内

部モデル学習時にその学習が空間的に局所的な領域にしか反映されないことが報告されている [29]．この種のモジュール分割を “空間方向のモジュール分割 (spatial modularity)” と呼ぶ．次に何故，基本単位となるモジュールが線形であったほうが良いか，MPFIM と MMRL の場合について議論する．

MPFIM: Multiple Paired Forward-Inverse Model

MPFIM は D.M. Wolpert と M. Kawato [103] らによって提案されたオリジナルの MOSAIC モデルである． この機構は複数の対になった順逆モデルを持っている．一般的に連続時間において，順モデル，逆モデルは以下のようにかける．

$$\dot{x} = F(x, u) \tag{7.1}$$
$$u = I(x, \dot{x}^d) \tag{7.2}$$

離散時間では以下となる．

$$x_{t+1} = F(x_t, u_t) \tag{7.3}$$
$$u_t = I(x_t, x^d_{t+1}) \tag{7.4}$$

これは前章まで知覚シェマに獲得させていた順逆モデルに等しい．線形の場合は z 変換により，また，そうでない場合は近似的にオイラー近似により連続と離散の順モデルは互いに変換することが出来るのでどちらの表現を用いても等価である．ここで，$x(\in X)$ は状態空間 $X(= R^n)$ における状態変数であり，$u(\in U)$ は行動空間 $U(= R^m)$ における行動出力 (トルクなど) である．$\dot{x}^d$ は望まれる状態変化を表す．別の言葉でいえば指令値である．順モデル F に対応する，I は以下の残差 e を最小化するように求められる．

$$e = ||\dot{x}^d - \dot{x}||^2 = ||\dot{x}^d - F(x, I(x, \dot{x}^d))||^2 \tag{7.5}$$

これが最小化されなければ，獲得された逆モデルは望まれた軌道を実現することが出来ない．F が非線形の場合には一般的には逆モデルを求めることは困難であるが，もし順モデル F が下のように線形であれば，対応する逆モデルは簡単に得ることができる．

$$F(x, u) = F_x x + F_u u \tag{7.6}$$
$$I(x, \dot{x}) = I_x x + I_{\dot{x}} \dot{x} \tag{7.7}$$
$$I_x = -F_u^{\#} F_x \, I_{\dot{x}} = F_u^{\#} \tag{7.8}$$

$F_u^{\#}$ は行列 F_u の Moore-Penrose (MP) 逆行列をさす [130]．もし，U の次元が，X の次元より小さく，e を常に 0 にすることが出来ないときは，この逆モデルはその中でも e を最小化すること保障する．一方で，U の次元が X の次元より大きいときには，多くの逆モデルが e を 0 にする．このとき，上の逆モデルは $||u||^2$ を最小化する逆モデルとなる．これは MP 逆行列の性質による．

しかし，このような理想的な逆モデルの存在は F が非線形であると保障できない．さらに，場合によっては，逆モデルは多値関数になってしまう場合すら現れる．このような場合の逆モデルの出力値はその平均となってしまい，無意味な出力を生成する場合がある [41]．これは一般の関数において，逆関数の存在が保障されないのと同様の問題である．

故に，MPFIM においては，空間的に線形分割された順モデルを獲得するというのは非常に理にかなった方略である．

MMRL: Multiple Model-based Reinforcement Learning

K. Doya らは非線形制御のためのモジュール型の強化学習機構を提案した [25]．これは内部モデルのもう一つの利用方法である．MMRL においては Multiple Linear Quadratic Controllers (MLQC) がその中心的な役割を果たす．もし，局所的なダイナミクスが i 番目の順モデル F^i によって線形に近似され，局所的な報酬関数 r_i が二次形式で与えられれば，LQC(Linear Quadratic Controller) による局所的な最適制御器を得ることができる．

$$r_i(x,u) = -(x-x_i)^T Q_i (x-x_i) - u^T R_i u \tag{7.9}$$

Riccati 方程式を解くことにより最適制御器 u_i^* を得ることが出来る．これが MMRL の基本的な考え方となる．ゆえに，局所的なモジュールが線形であるというこが MMRL の基本的な仮定なのである．このように空間方向のモジュール分割はこれらの枠組みを用いる上で必要となる．

7.2.2　時間方向のモジュール

人々は時変な環境を相手にしなければならない．たとえば，ある人が使っていた道具を別の道具に持ち替えた時，即座に新たな道具のための内部モデルを利用し始めなければならない．つまり，その人は一つの力学モデルから新たな力学モデルへと自

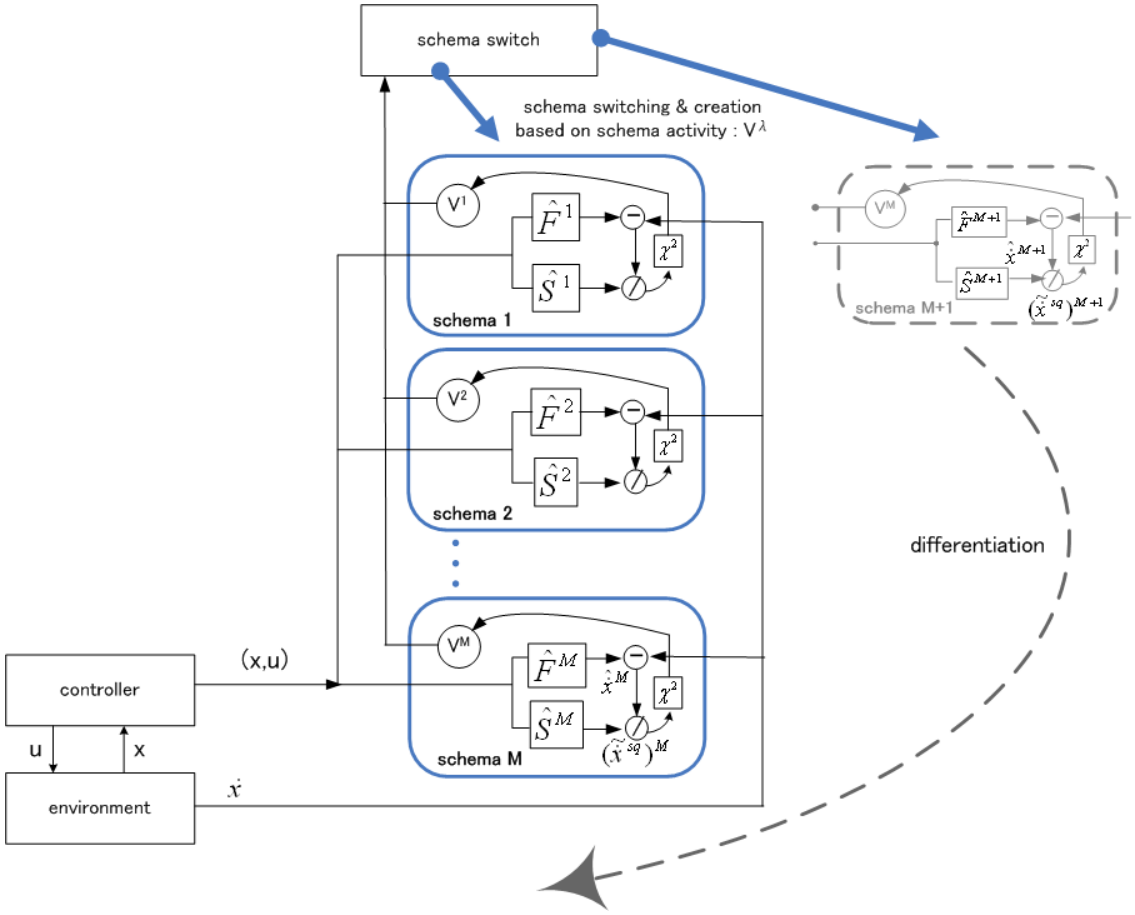


Fig.7.1: An overview of the schema model with a normalized Gaussian network

らの認識をシフトしなければならないのである．もし，時間方向のモジュール分割が無かったら，道具についての内部モデルは道具を持ち替えるたびに上書きされ，過去に利用した道具についての熟練は消滅することであろう．つまり，もし時間方向のモジュール分割がなかったら，人々は将来の利用のために学習した環境 (道具を含む) のダイナミクスを保存できなくなる．これは，非常に非効率な適応戦略である．このように，時間方向のモジュール分割は明らかに必要である．

7.3　NGSM: Schema Model with a Normalized Gaussian network

Fig. 7.1 は本章で提案する NGSM の概観を表している．これまでに述べてきたように，我々人間の認知システムは多くのシェマを持ち，それらは環境と相互作用する中で並列に活動する．同化・調節，均衡化・分化といったシェマの初期の発達ダイナミクスについては今までの章で述べているので割愛する．本稿では前章までに記号創発や，ロボットの自律的知能発達の観点から累増的なモジュール型学習機構として，双シェマモデル (Dual-Schemata model) や強化学習シェマモデルといった，一連のシェマモデルを提案してきた．シェマモデルは上の分類では時間方向のモジュール分解を行うためのモデルとなっている．これらの機構は Piaget のシェマモデルと呼ばれる自律分散的な機構にそのアイデアを得ていたが，それらは均衡化，分化のプロセスを計算論的に表現することを可能にしたが．その考え方が計算論的脳神経科学における内部モデルの獲得プロセスにおいても重要な役目を果たしえることを示す．

7.3.1　二種類のモジュールの階層的管理

MOSAIC とは対照的に，NGSM においては，空間方向と時間方向のモジュール分割を明示的に区別する．さらに，時間方向のモジュール分割に MOSAIC が拠るところのベイズ推定ではなくシェマモデルを適用することにより，モジュールの累増的な獲得を可能にする．この区別は Fig. 7.2 に示すような，二種類のモジュール分割の階層的な管理を要求する．ここで上位のモジュール，つまり各モジュールが時不変の非線形なダイナミクスを獲得する時間方向のモジュールを**シェマ**(schema) と呼び，下位の局所的な線形モデルを担当するモジュールを単純に**モジュール**(module) と呼ぶことにする．

7.3.2　シェマによる時間方向のモジュール化

本節では NGSM がどのように時間方向のモジュールであるシェマを累増的に獲得するかについて示す．一般的に道具や環境のダイナミクスは線形ではない．線形システムはあくまで理想化された系である．ゆえに順モデルは非線形関数として獲得され

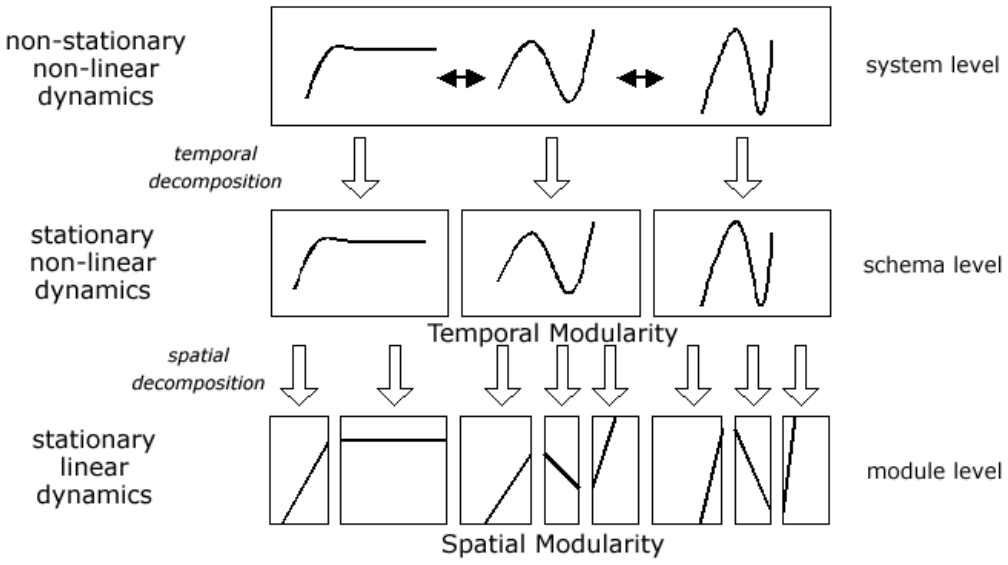


Fig.7.2: Abstract image of hierarchical management of two types of modularity. Each curve represents non-linear dynamics. Non-stationary and non-linear dynamics are decomposed into stationary linear dynamics hierarchically.

なければならない．空間方向のモジュール化と異なり，時間方向のモジュール化では各モジュールが線形でなければならない理由は無い．ここで，提案する時間方向の累増的モジュール分割の方法は前章までに提案したものと基本的には同じである．鍵となる考え方は同化と調節，そして均衡化と分化である．さらに本節では今までアドホックに与えてきた同化，分化のためのシェマ活性度に統計学的に明確な意味づけを与える．

シェマの内部関数の獲得

一般的に対象となるダイナミクスは以下のようにかける．このダイナミクスは非線形で時変である．

$$\dot{x} = F(x, u, t) + n(t) \tag{7.10}$$

ここで $n(t) \sim N(0, \Sigma^F)$ はノイズ項である．一つのシェマは少なくとも二つの内部関数を持つ．一つは非線形の順モデル $\hat{F}^\lambda$ であり，もう一つはその誤差の分散推定を

行う誤差予測モデル $\hat{S}^{\lambda}$ である．それぞれのモデルはパラメータベクトル w を持つ．また，$\hat{\cdot}$ は推定値をあらわす．

$$\hat{\dot{x}^{\lambda}} = \hat{F}^{\lambda}(x, u; w^{F^{\lambda}}) \tag{7.11}$$

$$(\widehat{\tilde{\dot{x}}^{sq}})^{\lambda} = \hat{S}^{\lambda}(x, u; w^{S^{\lambda}}) \tag{7.12}$$

ここで $\lambda \in \Lambda$ はシェマの番号である．また，$\tilde{\dot{x}}^{\lambda}$ は λ 番面のシェマの持つ順モデルの予測誤差をあらわす．また，演算子 sq はベクトルの各次元の値を二乗値したベクトルを返す．

$$\tilde{\dot{x}}^{\lambda} = x - \dot{x}^{\lambda} \;\; = \;\; F(x, u, t) - \hat{F}^{\lambda}(x, u; w^{F^{\lambda}}) \tag{7.13}$$

$$sq : x = (x_1, x_2, .., .x_n) \longmapsto x^{sq} = (x_1^2, x_2^2, ..., x_n^2) \tag{7.14}$$

$\hat{F}^{\lambda}$ と $\hat{S}^{\lambda}$ は共に，環境との相互作用を通してだんだんと変化していく (均衡化)．このときの目的関数は $\hat{F^{\lambda}}$ と $\hat{S^{\lambda}}$ それぞれについて以下のように定義される．$\hat{F}^{\lambda}$ と $\hat{S}^{\lambda}$ は λ 番目のシェマが環境との相互作用を通じて経験を同化するたびに，それぞれの目的関数を最小化するように変更されていく (調節)．

$$O_F = ||\dot{x} - \hat{F}^{\lambda}(x, u; w^{F^{\lambda}})||^2 \tag{7.15}$$

$$O_S = ||\tilde{\dot{x}}^{sq} - \hat{S}^{\lambda}(x, u, w^{S^{\lambda}})||^2 \tag{7.16}$$

これらの最適化問題は明らかに，最小二乗問題となる．つまり，もし適当な基底群 $\{b_i\}$ と共に一般化線形近似器を下のように用意すれば，最小二乗法が最適パラメータ $(w^{F^{\lambda}})^*$ と $(w^{S^{\lambda}})^*$ の存在を保証する．ここで $*$ は最適値を意味する．

$$\hat{F}^{\lambda} = \sum_i w_i^{F^{\lambda}} b_i^{F^{\lambda}}(x, u) \tag{7.17}$$

$$\hat{S}^{\lambda} = \sum_i w_i^{S^{\lambda}} b_i^{S^{\lambda}}(x, u) \tag{7.18}$$

しかし，我々が問題にする学習はオンラインで行われるので，最小二乗法は用いることが出来ない．そのかわりに，オンラインで最小二乗問題を解くための様々な代替手法を用いることにより，十分に適当な w を推定することが出来る．これらの解法には逐次最小二乗法や勾配法などを含む．

シェマの選択と生成

本小節では，シェマの選択と生成のアルゴリズムを示す．これは同化と分化のダイナミクスに相当する． まず，主観的誤差 R^λ を以下のように定義する．

$$R^\lambda(t) = \tilde{\dot{x}^\lambda} \cdot diag(S^\lambda(x,u))^{-\frac{1}{2}} \tag{7.19}$$
$$= (\frac{\tilde{\dot{x}^\lambda}_1}{\sqrt{(\tilde{\dot{x}}_1^2)^\lambda}}, \frac{\tilde{\dot{x}^\lambda}_2}{\sqrt{(\tilde{\dot{x}}_2^2)^\lambda}}, ..., \frac{\tilde{\dot{x}^\lambda}_n}{\sqrt{(\tilde{\dot{x}}_n^2)^\lambda}}) = (\frac{\tilde{\dot{x}^\lambda}_1}{\hat{\sigma}_1^\lambda}, \frac{\tilde{\dot{x}^\lambda}_2}{\hat{\sigma}_2^\lambda}, ..., \frac{\tilde{\dot{x}^\lambda}_n}{\hat{\sigma}_n^\lambda})$$

ここで $diag(S^\lambda(x,u))$ は λ 番目の順モデルの予測誤差が各次元で独立であったとした場合の分散共分散行列である．簡単のために，$\tilde{\dot{x}}^\lambda$ は各次元独立であるとする．これは分散共分散行列が対角行列であることを意味する．また $diag(*)^{-\frac{1}{2}}$ はベクトルを対角行列に変換した後に全ての対角成分を平方根の逆数にする関数である．

多くの場合，残差ベクトル $\tilde{\dot{x}}^\lambda$ は正規分布する．故に，時不変系において $R^\lambda(t)$ は多次元標準正規分布をすると考えられる． これは $R^\lambda(t) = (R_1^\lambda, R_2^\lambda, ..., R_n^\lambda)$ の各成分が標準正規分布をとるということを意味する．

この仮定により，λ 番目のシェマが観測されたサンプルベクトルを同化するかどうかの基準を**仮説検定理論**に基づいて導入することができる．確率変数 X_i が標準正規分布をとるとすると $Z = \sum_{i=1}^n X_i^2$ は自由度 n の χ^2 分布をとる．故に，主観的誤差の各成分の二乗値の平均は $\frac{1}{N}\sum_{i=1}^N R^\lambda(t_i)^{sq}$ の分布が求まる．しかし，オンライン学習の文脈ではこのような平均値はほとんど用いられず，重み付き平均が好んで用いられる．ここで，主観的誤差の重みつき平均を導入する． p $(0 \leq p < 1)$ は固執率とよび， 1 に近ければ近いほど，長時間平均となり， 0 の時にはその瞬間の値となる．

$$\overline{R^\lambda(t)^{sq}} = (1-p)\sum_{k=0}^{\infty} p^k \cdot R^\lambda(t-kdt)^{sq} \tag{7.20}$$
$$= (1-p)\cdot R^\lambda(t-dt)^{sq} + p \cdot \overline{R^\lambda(t-dt)^{sq}}$$
$$\sim \int_0^\infty \frac{1}{\tau_p} exp(-\frac{s}{\tau_p}) R^\lambda(t-s)^{sq} ds \tag{7.21}$$

ここで，dt は観測のタイムステップである．この重み付き平均は連続時間でとらえると，積分表現が得られる ($\tau_p = \frac{dt}{1-p}$)．上式より，重み付き平均ならば明示的に過去のサンプルを記憶することなしに逐次的に求めることができる．この平均操作は MOSAIC における時間連続性の概念に等しい [25]． しかし，重み無し平均の分布が

解析的に求まるのに対して，重みつき平均の分布はガンマ分布の再生性 [181] を利用できないために，簡単な解析的解は求まらない．よって，分散の等しいカイ二乗で近似することを考える．$\overline{R^{\lambda}(t)^{sq}}$ に $\frac{1+p}{1-p}$ を掛けた確率変数は近似的に自由度 $\frac{1+p}{1-p}$ の χ^2 分布にかなり高い精度であらわされる．本章で提案する枠組みでは，それぞれのシェマが仮説 $H^{\lambda} : (F = \hat{F}^{\lambda})$ を各次元独立に χ^2 検定を行うことにより，経験を同化するかどうかを判断する． ここで，H_i^{λ} を現在の環境が λ 番目のシェマの i 次元目の予測と合っているという仮説とする．仮説 H^{λ} と H_i^{λ} の関係は以下で表される．

$$H^{\lambda} = H_1^{\lambda} \wedge H_2^{\lambda} \wedge \cdots \wedge H_n^{\lambda} \tag{7.22}$$

$$H_i^{\lambda} \Longleftrightarrow F_i = \hat{F_i^{\lambda}} \tag{7.23}$$

$$\Longleftrightarrow \overline{R_i^{\lambda}(t)^{sq}} = 0 \tag{7.24}$$

仮説検定の枠組みにおいては仮説を採択するか否かは，人為的に定められる有意水準 α を用いて判断される．ここで，単純な閾値関数 sgn を用いることによって以下のファジィ真理値関数 μ を定義することが出来る．χ^2 分布の累積密度関数を $\chi^2(x, n)$ であらわすと

$$\mu(H^{\lambda}) = \mu(\bigwedge_{i=1}^{n} H_i^{\lambda}) \tag{7.25}$$

$$= \min_{i=1}^{n}(\mu(H_i^{\lambda})) \tag{7.26}$$

$$= \min_{i=1}^{n} sgn(\chi_c^2(n_p \cdot \overline{R_i^{\lambda}(t)^{sq}}, n_p - \alpha) \tag{7.27}$$

$$= sgn(\min_{i=1}^{n}(\chi_c^2(n_p \cdot \overline{R_i^{\lambda}(t)^{sq}}, n_p - \alpha) \tag{7.28}$$

$$= sgn(\chi_c^2(n_p \cdot \max_{i=1}^{n} \overline{R_i^{\lambda}(t)^{sq}}, n_p - \alpha) \tag{7.29}$$

$$sgn(x) = \begin{cases} 1 & if \quad x > 0 \\ 0 & otherwise \end{cases} \tag{7.30}$$

となる．ここで，$n_p = \frac{1+p}{1-p}$ かつ $\chi_c^2(x, n) = 1 - \chi^2(x, n)$ である．この式を **シェマ活性度**と呼ばれる値を導入することで簡略化する．このシェマ活性度は各時刻の主観的誤差から逐次的に求まって行くという点で前章までのシェマ活性度と本質的に同じである．

$$\mathbf{V}^{\lambda}(t) = \chi_c^2(n_p \cdot \max_{i}^{n} \overline{R_i^{\lambda}(t)^{sq}}, n_p) \tag{7.31}$$

これを用いることによってシェマが経験を同化するか否かの基準が最終的に得られる． ここでシェマ同化の基準となる有意水準を $\mathbf{V}_\alpha$ とする．

$$\mu(H^\lambda) = sgn(\mathbf{V}^\lambda(t) - \mathbf{V}_\alpha) \tag{7.32}$$

この仮説採択の判断は閾値関数をシグモイドのような連続関数に変えることによりファジィ化することができる．本章では真理値がクリスプ，つまり二値的である場合のみを扱う．ただし，この真理値は MOSAIC における責任信号のような，λ 番目のシェマの選択確率ではない．なぜなら，複数のシェマが真と見做される場合があるからである．この真理値を用いて選択確率を定義しなければならない．先に述べたように，シェマモデルは累増的なモジュール獲得を可能とする機構であるので，選択則と生成則が一体となって表現されなければならない．ここでは選択ルールと生成ルールを一体として一つの式で定義する．ここで，式を簡単にするため 0 番目のシェマをその活性度が常にゼロであるヌルシェマ (null-schema) として定義する．$P(\lambda)$ は λ 番目のシェマが選択される確率である．本章では μ が二値的であるので，この選択則は決定論的となる．

$$P(\lambda) = P(\bigwedge_{l=0}^{\lambda-1} \overline{H^l} \wedge H^\lambda) = \mu(H^\lambda) \prod_{l=0}^{\lambda-1} \mu(\overline{H^l}) \tag{7.33}$$

この選択則は 1 番目から $(\lambda - 1)$ 番目のシェマが全て仮説を採択せずに，λ 番目のシェマが採択された場合にのみ，λ 番目のシェマが選択される．もし，存在するすべてのシェマ ($\forall\lambda \in \Lambda$) が活性化されなかった場合には，新たなシェマが生成されて $(\#(\Lambda) + 1)$ 番の番号が振られる．あらたなシェマが作られた時その $\hat{S}^{\#(\Lambda)+1}$ は新規な環境における経験を同化するために，十分大きな値に初期化される．また，順モデル $\hat{F}^{\#(\Lambda)+1}$ はランダムに初期化される．これらのアルゴリズムは具体的に内部に持つ学習器とは独立であり，原理的には様々な学習器でもって実現することが可能である．この仮説検定理論に基づく累増的モジュール型学習器をシェマモデル (schema model) と呼んでいる．シェマモデルでは経験が仮説を生成し，仮説が経験を検定するという自己言及的なループの中で構造化が進められていく．

7.3.3 シェマ内部における空間方向のモジュール表現

先に指摘したように，その有用性のためには要素となる最小単位の順モデルは線形となる必要がある．MOSAIC モデルにおいては K. Doya らが，現在の状態量 x に

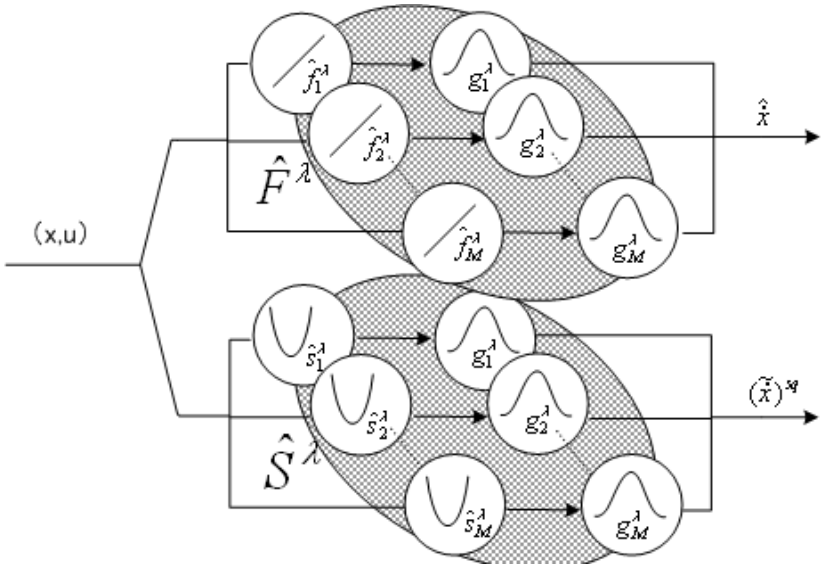


Fig.7.3: Spatial modularity inside a schema

応じてモジュール割り当ての事前確率を決定する空間局在性 (spatial locality) の概念を導入した [25]．この考え方は locally weighted regression (LWR) [8] の考え方に非常に類似している．LWR ではそれぞれの局所モデルが空間的に局所的な事例のみから学習される．もし，全ての重み関数が正規化ガウス関数で，また，すべての局所モデルが線形であった場合には，しばしばそれは正規化ガウス関数ネットワーク (NGnet) と呼ばれる [54] [71]．言い換えると，正規化ガウス関数ネットワークとは正規化ガウス関数をゲーティング関数に用いることにより，非線形関数を複数の線形関数に分割して近似する近似器である．NGSM では空間方向のモジュール分割を単純に NGnet を関数近似器として利用することで実現させる (Fig. 7.3)．

ゲーティング関数 $\mathbf{g}_k^{\lambda}(x)$ はガウス関数 G を用いて以下のように定義される．ここで k は局所モデルに振られた番号である．

$$\mathbf{g}_k^{\lambda}(x) = \frac{G(x; c_k^{\lambda}, \Sigma_k^{\lambda})}{\sum_{l=1}^{M} G(x; c_l^{\lambda}, \Sigma_l^{\lambda})} \tag{7.34}$$

$$G(x; c, \Sigma) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} exp[-\frac{1}{2}(x-c)^T \Sigma^{-1}(x-c)] \tag{7.35}$$

c_k^{λ} はゲーティング関数の中心を表すベクトルであり，Σ_k^{λ} は分散共分散行列である．各シェマが持つ順モデル F^{λ} と誤差予測モデル S^{λ} は幾つかの空間局所的なモデル

$\hat{f}_k^\lambda, \hat{s}_k^\lambda$ にそれぞれシェマ内部において分解される．

$$\hat{F}^\lambda(x,u;w^{F^\lambda}) = \sum_{k=1}^{M} \mathbf{g}_k^\lambda(x)\hat{f}_k^\lambda(x,u;w^{f_k^\lambda}) \tag{7.36}$$

$$\hat{f}_k^\lambda(x) = \hat{A}^{f_k^\lambda}(x - c_k^\lambda) + \hat{B}^{f_k^\lambda}u \tag{7.37}$$

$$w^{F^\lambda} \equiv [w_1^{F^\lambda}, w_2^{F^\lambda}, \cdots, w_M^{F^\lambda}]$$

$$w_k^{F^\lambda} = [\hat{A}^{f_k^\lambda}, \hat{B}^{f_k^\lambda}]$$

S^λ の局所モデルは線形ではなく，各次元において (x,u) についての二次形式としてあたえられる．これは局所モデル f_k^λ の推定がその領域の中央では正確だが，外部では不正確になるためである．また F^λ と S^λ は同じゲーティング関数群を共有する．これは局所モデルの誤差は局所線形化点を離れるほど大きくなると考えられるためである．

$$\hat{S}^\lambda(x,u;w^{S^\lambda}) = \sum_{k=1}^{M} \mathbf{g}_k^\lambda(x)\hat{s}_k^\lambda(x,u;w^{s_k^\lambda}) \tag{7.38}$$

$$\hat{s}_k^\lambda(x) = \hat{A}^{s_k^\lambda}(x - c_k^\lambda)^{sq} + \hat{B}^{s_k^\lambda}u^{sq} \tag{7.39}$$

$$w^{S^\lambda} \equiv [w_1^{S^\lambda}, w_2^{S^\lambda}, \cdots, w_M^{S^\lambda}]$$

$$w_k^{S^\lambda} = [\hat{A}^{s_k^\lambda}, \hat{B}^{s_k^\lambda}]$$

LWR はしばしば新しいアプローチだと言われるが，もしゲーティング関数の学習を行わない場合には，それは結局古典的な一般化線形近似器でしかない．

$$\hat{F}^\lambda(x,u;w^{F^\lambda}) = \sum_{k=1}^{M} \mathbf{g}_k^\lambda(x)(\hat{A}^{f_k^\lambda}(x - c_k^\lambda) + \hat{B}^{f_k^\lambda}u) \tag{7.40}$$

$$= \sum_{k=1}^{M} \begin{bmatrix} \hat{A}^{f_k^\lambda} & \hat{B}^{f_k^\lambda} \end{bmatrix} \begin{bmatrix} \mathbf{g}_k^\lambda(x)(x - c_k^\lambda) \\ \mathbf{g}_k^\lambda(x)u \end{bmatrix}$$

$$= \sum_{k=1}^{M} w_k^{F^\lambda}\mathbf{b}_k^{F^\lambda}(x,u) \tag{7.41}$$

$$\hat{S}^{\lambda}(x,u;w^{S^{\lambda}}) = \sum_{k=1}^{M} \mathbf{g}_k^{\lambda}(x)(\hat{A}^{s_k^{\lambda}}(x-c_k^{\lambda})^{sq} + \hat{B}^{s_k^{\lambda}}u^{sq}) \tag{7.42}$$

$$= \sum_{k=1}^{M} \left[\begin{array}{cc} \hat{A}^{s_k^{\lambda}} & \hat{B}^{s_k^{\lambda}} \end{array} \right] \left[\begin{array}{c} \mathbf{g}_k^{\lambda}(x)(x-c_k^{\lambda})^{sq} \\ \mathbf{g}_k^{\lambda}(x)u^{sq} \end{array} \right]$$

$$= \sum_{k=1}^{M} w_k^{S^{\lambda}} \mathbf{b}_k^{S^{\lambda}}(x,u) \tag{7.43}$$

これらの等式はゲーティング関数を固定した NGnet は M 個の基底を持った単純な一般化線形近似器に過ぎないことを示している．先に述べたように一般化線形近似器は回帰において最適なパラメータ w^* を持つことが知られている．もしバッチ学習を行うならば，最小二乗法 (LSM) を用いることができ，この最適値を得ることが出来る．しかし，本稿ではあくまでオンライン学習を問題にするので，そのかわりに，勾配法に基づいた学習法を提案する．MOSAIC においても基本的な学習ダイナミクスは勾配法によって導入されているので，この導入により後に公平な視点から比較することが出来る．もちろん，逐次最小二乗法や他のオンライン学習の手法に基づいて定式化することも可能である．その場合，(7.13) (7.14) に示した目的関数を最小化するように定式化する．$(x,u,\dot{x})$ が観測されたとき

$$\tau_F \frac{dw_k^{F^{\lambda}}}{dt} = -(\dot{x} - \hat{F}^{\lambda}(x,u)) \cdot (b_k^{F^{\lambda}}(x,u))^T \tag{7.44}$$

$$\tau_S \frac{dw_k^{S^{\lambda}}}{dt} = -(\tilde{\dot{x}}^2 - \hat{S}^{\lambda}(x,u)) \cdot (b_k^{S^{\lambda}}(x,u))^T \tag{7.45}$$

に従ってパラメータは更新される．ここで τ_F と τ_S は学習の時定数である．この定式化は適応的に非線形な順モデルを適応的に生成し選択・学習する機構を導く．さらに，その局所モデルは線形であるという条件も満たしている．

7.4　実験

我々は提案手法である NGSM の有効性を検証するためにシミュレーションを使って実験を行った．単振子を振り上げるというタスクは非線形系の制御タスクとしてよく用いられる系である．本実験でもこれを対象として用いる．K. Doya らは MMRL を単振子系に適用し，振り上げタスクについて，よいシミュレーション結果を得た [25]．しかし，この機構で追実験を行うと複数の順モデルを利用して振り上げタス

クを実行する制御面では高い性能を示すものの，順モデルを獲得する段階においては各モジュールの適切な初期配置とメタパラメータの設定なしでは，頻繁に悪い実験結果に陥った．この結果は特に振子の長さが時々切り替わるような時変系 (これは例えば人が長さの違う道具を持ち換える場合などに対応しているが) において顕著であった． 故に本章では，時変系における複数順モデルの獲得過程に焦点を絞ることにする．

7.4.1 実験環境

我々はエージェントがある期間の間，一定の単振子の振り上げの練習を行い，それを時々別の力学特性を持つ単振子に持ち換えるという状況を考える．Fig. 7.4 に対象とする単振子を簡単に示す．

$$\ddot{\theta} = -\frac{g}{l}\sin(\theta) - \frac{\mu}{mgl^2}\dot{\theta} + \frac{1}{mgl^2}u \tag{7.46}$$

$$\text{pendulum } 1 = \{l = 2, m = 0.5, \mu = 0.1\} \tag{7.47}$$

$$\text{pendulum } 2 = \{l = 1, m = 1, \mu = 0.1\} \tag{7.48}$$

$$\text{pendulum } 3 = \{l = 0.5, m = 4, \mu = 0.1\} \tag{7.49}$$

重力定数は $g = 10$ とし，探索のための制御器は事前に，

$$\frac{du(t)}{dt} = -cu(t) + k\frac{dB_t}{dt} \tag{7.50}$$

と設定する．制御器は簡単なノイズ生成器として定義される．これはブラウン運動に減衰項が加わったもので，オルンシュタイン-ウーレンベック過程 (Ornstein-Uhlenbeck process) と呼ばれる [57]．振子のダイナミクスを探索し，その道具に馴れるために，エージェントはランダムに道具を振り上げることを試みる．ここで $c = 1$, $k = 1$ と設定するが，このとき上の確率過程の定常状態における標準偏差は $\sigma_u = 1$ となる．さらにトルクを $u \in [-T^{max}, T^{max}](T^{max} = 3)$ に制限する．複数の順モデルを適切に獲得し終えた後は，これらを用いてエージェントは新たな制御器を獲得することが出来る．複数順モデルの利用方法については他の文献に譲るので [25, 103]，本章では著さない．第 5 章で提案した双シェマモデル強化学習はこの利用法の一つである．状態空間は $x = (\theta, \dot{\theta}, 1)$ の三次元とする．ここで，三次元目は線形回帰の定数項を表すための次元であり，本質的には状態空間は二次元である． θ は $\theta \in [-3\pi, 3\pi]$ に制限される．θ がこの領域の外に出た場合には θ は初期値へとリセットされる．初期値は $\theta \in [-\pi, \pi]$ の範囲にランダムに設定される．

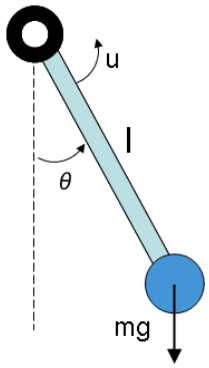


Fig.7.4: Pendulum

ゲーティング関数のパラメータは $c_k^\lambda = (k-3)\pi$, および $\sigma_k^\lambda = diag(1,1,1)$ と設定する．ここで $k = (0, \cdots, 6)$ である．これらの設定は NGnet の一般化線形近似器としての基底を決定する．

学習のための時定数は $\tau_F = \tau_S = 20$ および $p = 0.98$ $(\tau_p = \frac{dt}{1-p} = 5)$ とした．τ_p の値は文献 [38] における fMRI 実験に基づいている．文献 [38] においては内部モデルの切り替えに約 5 秒かかることが観測されており，上の時定数はそれに合わせている．学習のための時定数と，切り替えのための時定数における時間差は一般的なモジュール型学習器において本質的な点である．本来は 4 倍程度ではなく，より大きな時間差をとるべきであるが，シミュレーションにおける学習時間の短縮のために学習の時定数を小さく設定している．シェマの選択，生成に用い，シェマシステムの粒度と安定性を決定する有意水準は $\alpha = 0.005$ に設定した．エージェントは一つ目の振子を 2000[s] の間利用し続けることで熟練し，次にそれをもう一つのものに持ち替え 2000[s] の間利用する．次に 3 つ目の振子に持ち替えて 2000[s] 用いる．これら三つの振子を学習した後にこれら 3 つの振子を次々に持ち替えて利用する．順序依存性が無いことを示すために，このとき持ち帰る順番は学習時と同一でなく，$1 \rightarrow 3 \rightarrow 2$ とする．

7.4.2　結果

Fig. 7.5 の下図はシェマ活性度の時間変化を示している．それぞれのシェマは順々に活性化しているのがわかる．Fig. 7.5 の上図はシェマの分化と選択が時系列的に起こっていく様を示している．相互作用を通して，シェマは 4 つに分化していった．こ

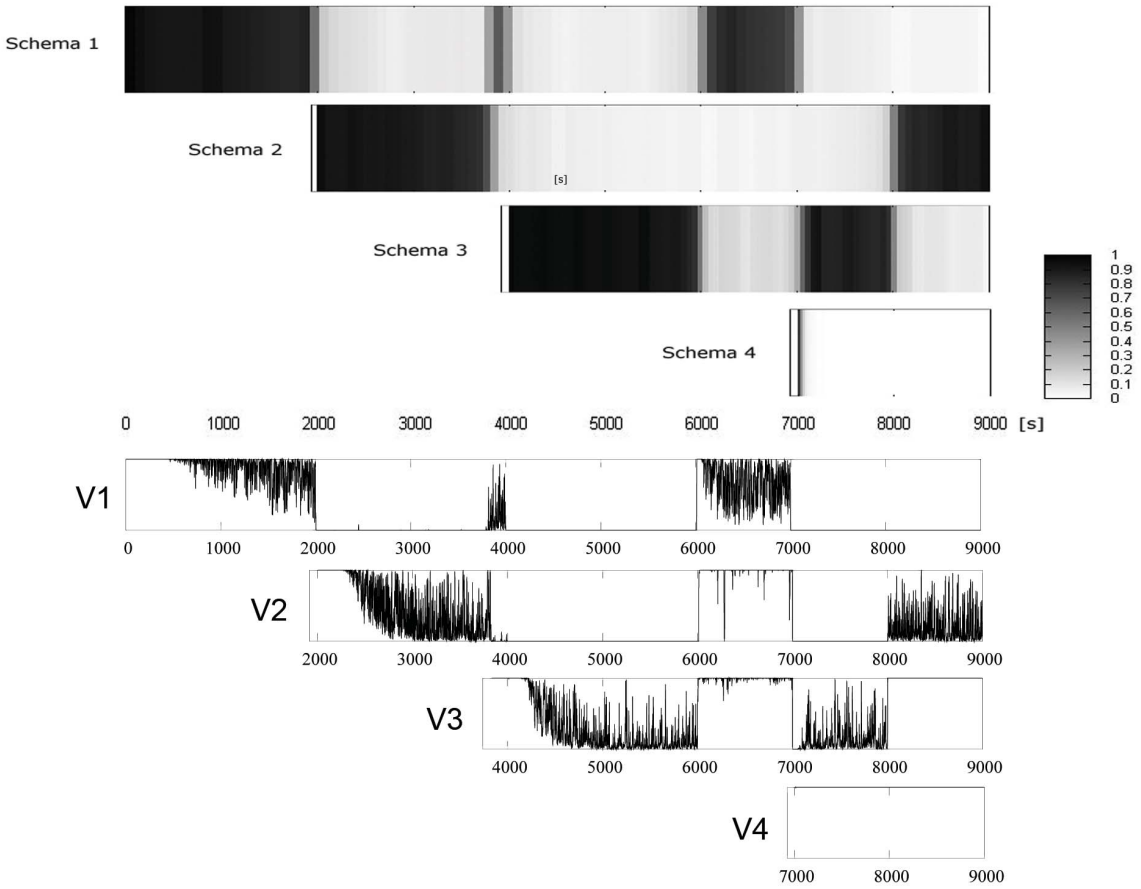


Fig.7.5: Top: ratio in which each schema was selected in each 10[s] timestep, bottom: time course of schema activities

れらは NGSM が変化する環境 (道具) に対して，適当な量のモジュールを配置し，順に選択していくことが出来たことを示している．

Fig. 7.6，Fig. 7.7 はそれぞれの時間でそれぞれのシェマ内部に獲得された順モデルの $(\dot{\theta}, u) = (0, 0)$ での断面図を示している．それぞれのシェマは時変系の時不変部分を分担し，その内部において線形分担する局所モジュールの組み合わせにより適切に非線形関数を近似出来ている．

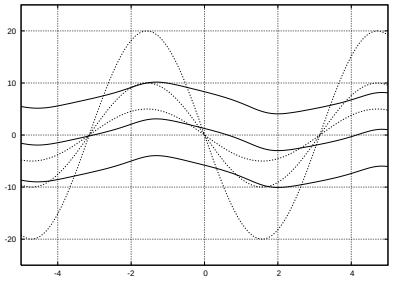
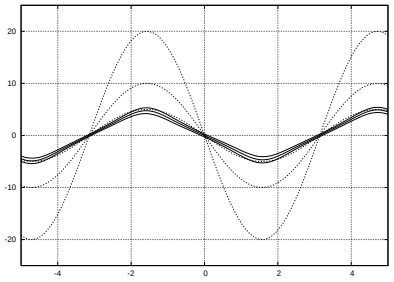
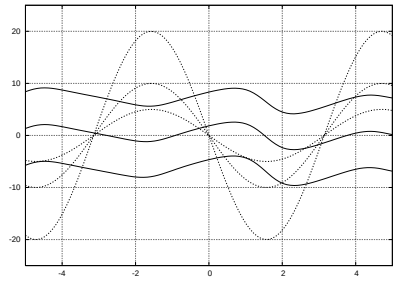
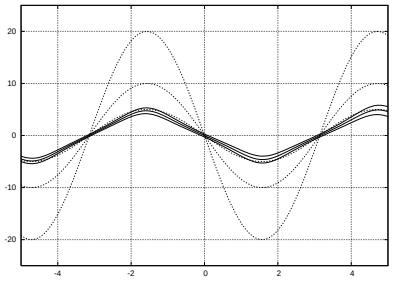
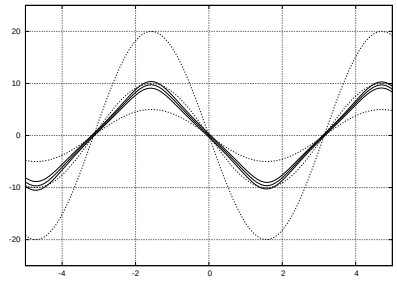
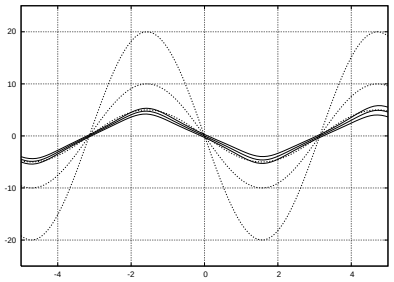
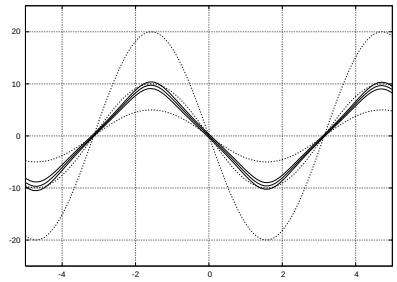
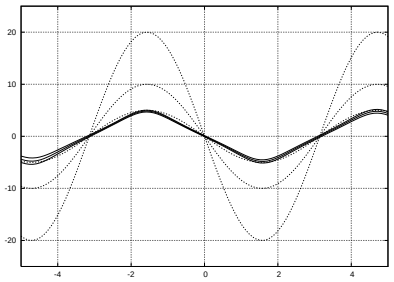
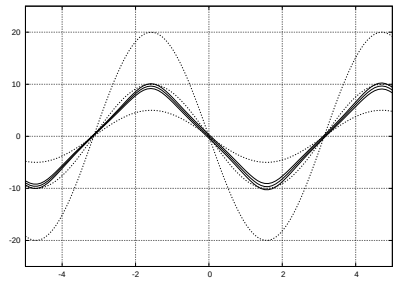
t	schema 1	schema 2
0		*does not exist*
2000		
4000		
6000		
9000		

Fig.7.6: Obtained forward models for respective schemata: dotted curves are three target dyanamics , middle curve represents $\hat{F}^{\lambda}$, and other curves represent $\hat{F}^{\lambda} \pm \hat{\sigma}^{\lambda}$ respectively.

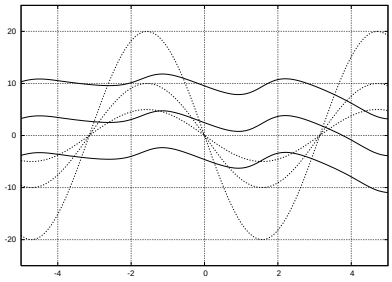
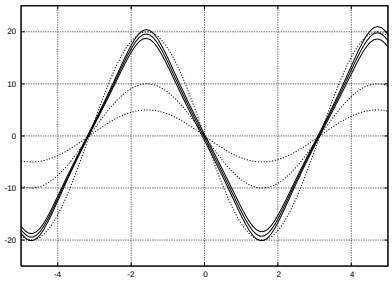
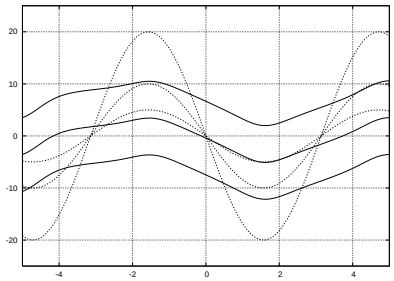
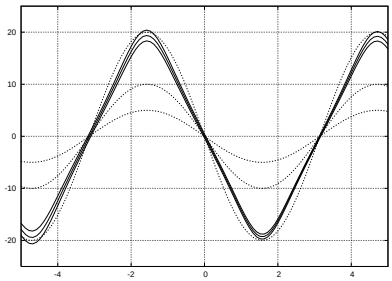
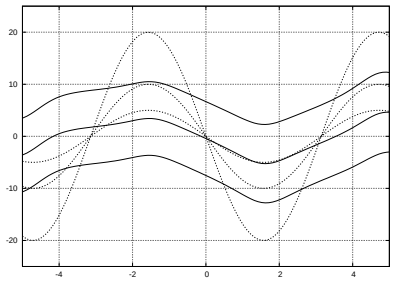
t	schema 3	schema 4
0	*does not exist*	*does not exist*
2000	*does not exist*	*does not exist*
4000		*does not exist*
6000		
9000		

Fig.7.7: Obtained forward models for respective schemata: dotted curves are three target dyanamics, middle curve is $\hat{F}^{\lambda}$, and other curves represent $\hat{F}^{\lambda} \pm \hat{\sigma}^{\lambda}$.

7.4.3　MOSAIC との比較

ここで提案手法である NGSM と従来手法である MOSAIC の勾配法に基づいた学習手法とを比較する．MOSAIC 研究においては幾つかの学習手法が提案されているが，K. Doya らによって提案されたもっとも典型的なものを用いる [25]．MOSAIC では大域的な順モデルは局所的な線形順モデル $\hat{f}^\lambda$ の重みつき線形和によって表現される．その責任信号 (responsibility signal) と呼ばれる重み L^λ は以下のベイズ推定に基づいて算出される．$\mathbf{g}^\lambda$ は NGnet の場合と同じく各局所線形モデルの空間局所性を決定するゲーティング関数である．ここでは NGnet の場合と同じ関数を用いる．

$$\hat{\dot{x}} = \sum_{\lambda=1}^{\#(\Lambda)} L^\lambda(t)\hat{f}^\lambda(x,u;w^{\hat{f}^\lambda}) \tag{7.51}$$

$$L^\lambda(t) = \frac{\mathbf{g}^\lambda \exp(-\frac{E^\lambda(t)}{2\sigma^2})}{\sum_{i=1}^{\#(\Lambda)} \mathbf{g}^\lambda \exp(-\frac{E^i(t)}{2\sigma^2})} \tag{7.52}$$

$$E^\lambda(t) = (1-p)\sum_{k=0}^{\infty} p^k \cdot e^\lambda(t-kdt) \tag{7.53}$$

$$e^\lambda(t) = \|\dot{x}(t) - \hat{f}^\lambda(x,u;w^{\hat{f}^\lambda})\|^2 \tag{7.54}$$

MOSAIC における一つのモジュールは，NGSM におけるシェマ内部の局所モジュールに相当する．これより MOSAIC における責任信号 L^λ に相当するものを NGSM において定義することができる．λ 番目のシェマにおける k 番目の局所モジュールの責任信号を L_k^λ とすると

$$L_k^\lambda(t) = \mathbf{g}_k^\lambda(x)P(\lambda) \tag{7.55}$$

となる[*2]．これは結局，NGSM と MOSAIC の計算論的な違いは責任信号の計算方法の違いによるということを示している．

次の疑問は，この計算論的な相違が NGSM と MOSAIC の複数順モデルの獲得過程にどのような影響を与えるかということである．我々は比較のために MOSAIC を実装したエージェントを同じ環境下でシミュレーションを行った．ここではシェマモデルと公平な比較をするために $28 = 4 \times 7$ のモジュールを事前に配置した．この数

[*2] 学習の式において完全な同一性を保たれないが，その差は微小でありほぼ同一視できる．

は NGSM のシミュレーションにおいて最終的に獲得された局所モジュールの数に等しい． また，我々は NGSM におけるゲーティング関数に相当する MOSAIC モデルにおける空間局所性を $c^{\lambda} = (\lambda_{(7)} - 3)\pi$ かつ $\sigma^{\lambda} = diag(1, 1, 1)$ として導入した．ここで $\lambda = (1, \cdots, 28), \lambda_{(7)} \equiv \lambda \ (mod\ 7)$ である．

ここで，メタパラメータを $p = 0.1$ かつ $\sigma = 0.1$ と設定した．我々は初め NGSM における実験と同じく $p = 0.98$ と設定したが，この時 MOSAIC は適切に分化された複数内部モデルを獲得することが出来なかった．これは MOSAIC が従来時定数の違う空間方向の局所性と時間方向の局所性を渾然一体として扱っていることに起因する．また，責任信号計算で予測誤差をスケーリングする σ が大きすぎた場合，同じ空間局所性を共有する 4 つのモジュールがほぼ全ての実験において融合してしまった．そこで我々はこれらの目標ダイナミクスを適切に分化してくれるメタパラメータを探索した．我々の探索では適切なメタパラメータの領域は非常に狭く，その値は $\dot{x}$ の値域の大きさに依存した．なぜなら，MOSAIC におけるメタパラメータ σ は次元を持つ量であるからである．これに対して，NGSM においてシェマの粒度を決定する $\mathbf{V}_{\alpha}$ は無次元量であり，時間局所的な時不変系においては誤差が正規分布するという仮定を除けば実際のタスクに拠らないので調整が容易である．我々の実験では $\mathbf{V}_{\alpha} \sim 0.0001$ 程度の値に設定すれば，他の系に対しても，適当な学習結果を示している．

Fig. 7.8 は NGSM と MOSAIC の予測誤差の時間発展を平滑化したものを示している．これは 10 試行の平均である．これは NGSM が MOSAIC より高精度な順モデルを獲得したことを示している． MOSAIC のモジュール選択，学習則が予測誤差最小化という目的から導かれているのに対して，NGSM のシェマ選択側が予測誤差を最小化するという規範とは直接関係しない式から導かれているにもかかわらずより小さな予測誤差を出すのは驚くべき点である．これは，予測モデルを導く統計的学習が通常一価関数を前提としているため，時変的な系を相手にする場合，予測誤差を最小化するまえに，そのまま扱えば渾然一体となり多価関数となる問題を一価関数の問題群に分割する必要があるためと考えられる．そのため，予測誤差最小化に基づいたモジュール分割よりも，過去に学習したデータ群と相反するデータをそのモデルの学習に加えないという基準でのモジュール分割，つまりベイズ推定に基づく MOSAIC よりも仮説検定に基づくシェマモデルの方が合理的であるためと考えられる．Fig. 7.9 は単振子の角度についてのゲインを表す項 ($\{A_{21}\} = \frac{\partial \ddot{\theta}}{\partial \theta}$) の変化を示している．原点 $(x, \dot{x}) = (0, 0)$ 付近のダイナミクスに注目し，その付近に空間局所性

をもつ MOSAIC のモジュール $\hat{f}^4, \hat{f}^{11}, \hat{f}^{18}$, と $\hat{f}^{25}$ のパラメータと NGSM の原点付近での局所モジュールの係数変化について示している．この図は NGSM が累増的かつ，MOSAIC より安定的に複数の順モデルを獲得していっていることを示している．

傾向としては MOSAIC では頻繁に環境を切り替えた方が比較的安定的に分化が進んだ．これに対して，NGSM では頻繁な環境の切り替えは簡単にそれらを一つの環境とみなし脱分化が進む．実際の人間を用いた実験では，視覚や聴覚による提示なしの頻繁なダイナミクスの切り替えは被験者が複数の内部モデルを獲得することを妨げることが報告されている [59, 16, 3]．故に，分化を導くためには，十分長い間ひとつの一価関数で表されるようなダイナミクスを提示し続けることが重要であると考えられる．統計学的な視点からすると，複数の関数から出てきた出されたサンプルや多価関数から出されたサンプルを関数近似することは無意味である．関数近似が有意味であるためには，対象となるサンプルはホワイトノイズを除いて，一貫して一つの関数から出力される必要がある．NGSM におけるシェマの切り替え則はこの一貫性を保障することを狙っている．Fig. 7.8 は短期的な誤差を減らすよりもこの同化する対象の一貫性を維持することが長期的な誤差を減らす上ではより効果的であることを示している．

7.5　結論

本章では NGSM を提案した．これは複数の順モデルをオンラインで獲得する手法の一案である．NGSM は従来手法である MOSAIC の様々な利点を保持している．例えば，モジュールの分担を決定する責任信号が局所的な計算によって決定される点や，最小単位のモジュールが持つ順モデルが線形であるという点がそれである．それに加え，NGSM は二つの新しい長所を有している．それは仮説検定に基づいた累増的なモジュール生成と，一般線形回帰に基づいた安定的な学習プロセスである． これらの特徴を利用するために，空間方向のモジュール分割と時間方向のモジュール分割を区別し階層的に管理する必要性について述べた．

モジュール型の学習制御機構が小脳の計算論的モデルとして提案されており，この考えは fMRI を用いた生理学的研究や認知科学的研究などによって支えられている．しかし，そのような実験は特定の計算論的な枠組みを限定するまでには至っていない．複数の順モデルが競合しつつ学習するモジュール型学習器としては mixture of experts や MOSAIC が主に考えられて来た．この二つの計算機構の大きな違い

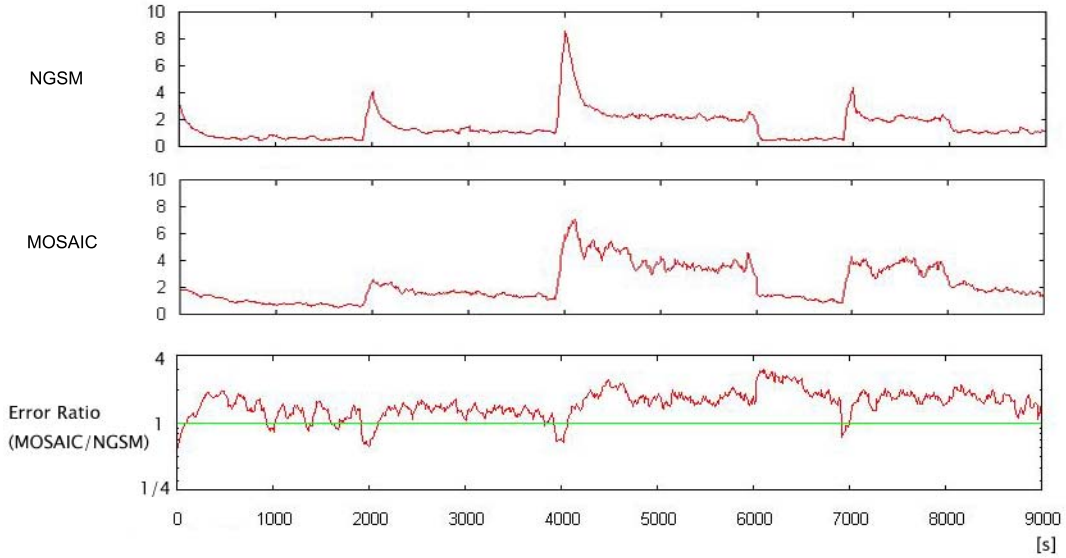


Fig.7.8: Top: time course of prediction error with NGSM, middle: time course of prediction error with MOSAIC, bottom: ratio of these prediction errors.

としてはモジュール群の外部にゲーティング関数が在るか無いかが挙げられる．H. Imamizu らは fMRI による実験を通して，モジュールのスイッチングのタイミングと脳部位活性関係を調べることで計算機構を特定しようとする実験を行った [38]．しかし，MOSAIC と本章で提案した NGSM は共にゲーティング関数を持たず局所的な計算のみでモジュール分担を決定する故に，そのような実験を用いてもどちらの計算が実際に行われているのかを判別することは出来ないと考えられる．NGSM は MOSAIC の計算機構を支持する脳科学的な多くの知見と矛盾せず，MOSAIC の持っていない累増性，安定性を持つ新たな計算機構であると言える．逆に言えば，今まで MOSAIC の存在を示唆してきた知見の多くは，シェマモデルの人間の認知システム内部における存在を示唆する知見として援用する事が可能になる．

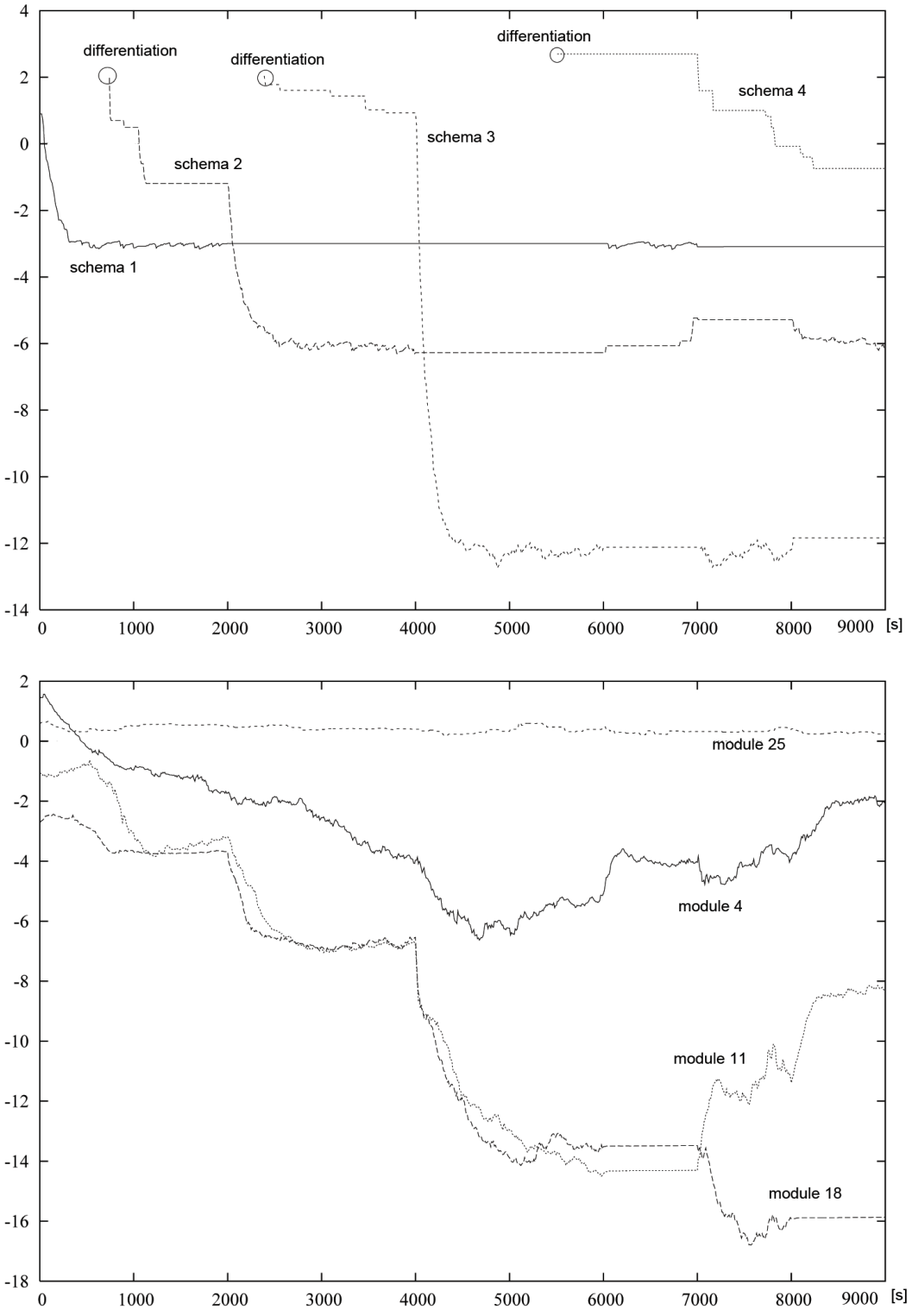


Fig.7.9: Differentiation and adaptation processes in NGSM (top) and MOSAIC (bottom)

第 8 章

スパイクタイミング依存のシナプス可塑性とシェマモデルの結合による三項関係の創発

第 IV 部はシェマモデルと脳神経科学の知見とを融合させることが狙いの一つであるが，前章では中枢神経系 (CNS) における複数内部モデルの獲得においてシェマモデルを適用することで従来手法より，より安定した内部モデル獲得が可能であることを示した．次に，本章では大脳新皮質や海馬といった部位でのダイナミクスとの関係を考える．大脳新皮質は進化の上で人間に多く備わった部位であり，言語つまり我々の関心事である「記号」とのつながりが強く認識されている [20]．近年，この部位において STDP(Spike Timing-Dependent Plasticity) と呼ばれるスパイクタイミング依存の可塑性が発見されており [49]，その役割が注目されている．第 8 章では，これまで提案してきた手法により組織化された自律システム内部に生成された内的表象を聴覚，視覚といった身体的相互作用に直接参与せず，遠隔的な情報を提供するチャンネルから入ってくる感覚刺激と連合させることにより，C.S. Peirce の三項関係に示される記号の解釈過程を自律エージェント内に発現させることが出来ることを示す．このとき，記号過程は第 6 章末で問題提起したシェマ選択における時間遅れの問題を解決するように働く．Fig 8.1 に示すように STDP 神経回路網をこれまでに提案してきた双シェマモデルにおける知覚シェマや行為シェマの選択機構に接続することにより，C.S. Peirce の記号における三項関係を自律適応系の閉じた作動のみから組織化する機構が完成することになる．

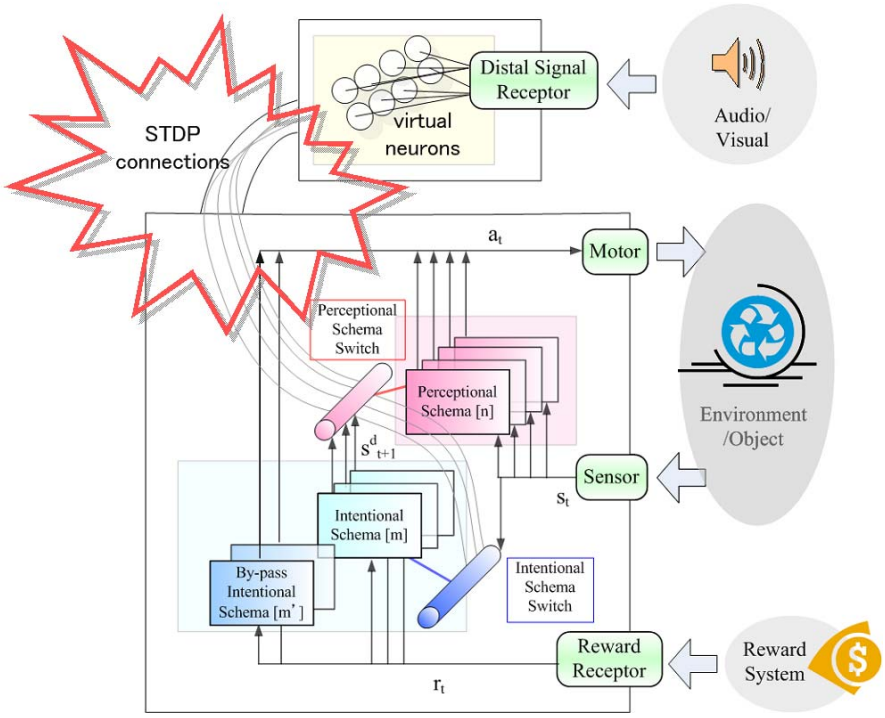


Fig.8.1: A STDP neuronal circuit connected to the Dual-Schemata model

8.1　はじめに

人間以外の多くの動物で，言語に類したコミュニケーション様式が発見されているにも関わらず，現在でもなお言語は人間の知性を特徴付ける脳の高次機能であるという考え方は支配的である [20]．しかし，なぜ言語という記号を通じて人はコミュニケーション可能なのか，また意味共有が可能なのか，という疑問は未だ解決されていない．19 世紀に C.S. Peirce が創始した記号論は言語を標識や病気の兆候などを始めとする記号の一部として捉えることにより，記号全般における解釈過程の理解に努めて来た [61]．しかし，その深い洞察と議論にもかかわらず，哲学的，社会学的な議論は記号現象を分類し，実際の社会における記号の使われ方と照合することが限界であり，その動的な解釈過程を諸科学のように解析的に理解し，実験的に再現させるには至っていない．これは，言語哲学における L. Wittgenstein や S. Kripke, D. Davidson らの議論においても同様である [200, 150]．ただ，記号を三項関係として理解するといった C.S Peirce の立場は記号の意味作用の本性を的確にとらえてい

ると言える．本論文第２章で，我々はその意味作用を自律適応系内部で生じさせるためには，自律適応系による環境との相互作用を通した差異体系，もしくは仮説群の獲得が必須であるという考えを示した．その構成論として，第 II 部，第 III 部を通して双シェマモデルを中心とした幾つかのシェマモデルの計算論を提案してきた．これらは，「違い」を明示的に指示する教師無しに，また，事前にいくつに分類しなければならないという命令無しに，自らの環境との相互作用を通して記憶を組織化する中で複数のシェマと呼ばれる学習器を獲得し，環境を分節化していく．実際に，第７章では個々のシェマが「仮説」そのものであることが，仮説検定理論の立場から示された．

前章までは，環境との二項的な相互作用[*1]の中で内的表象を獲得してきたが，これは人工知能研究の上での “記号” であっても，C.S. Peirce の記号論の視点からすれば記号とは言えない．記号過程は基本的には解釈者が自らの中の仮説を起動し，外在するサインを対象と関連付けて解釈する仮説推論 (abduction) によって作動するため，そのようなサイン，対象，解釈項の三項関係が存在しなければ記号と呼ぶことは出来ないからである[*2]．本稿で提案してきた計算論的なシェマモデルはシェマとしての仮説群[*3]を二項的な相互作用の中から生成することに成功してきた．これを基にして C.S. Peirce の三項関係に基づくセミオーシス (記号過程) を表現するためには，サインをシェマの想起に abductive(仮説推論的) に接続する機構が必要になる．しかし，どのようにそれを接続するのが妥当であるのだろうか？また，その接続は行動主体としての自律系にどのようなメリットを得させるのだろうか？本章まで新たな機構を導入する際に系として何らかのパフォーマンスの向上を伴うことを常に示してきた．パフォーマンスの向上を示すことは設計論として重要であるだけでなく，そのような機構が実際に人間の認知システムに存在するかどうかを考える上で構成論的にも重要である．三項関係により記号過程が実現されるフェーズでもそれは例外ではない．では，記号過程は形成されたシェマシステムにおける，どのような問題を解決することに貢献するのであろうか．それは，シェマモデルにおける時間遅れの問題である．

記号はしばしば行動主体が未来に遭遇する状況を前もって気付かせる役割を担う．高速道路での道路標識はまだ見えないパーキングエリアの存在を気付かせてくれるし，複数人で机を持ち上げるときの掛け声は実際に触覚から相手の力の変化が伝わっ

[*1] 2.4.3 節参照

[*2] 内在的な解釈項が再びサインとなる，思考のプロセスとしての無限のセミオーシスというダイナミクスも論じられるが，ここでは最も単純な形式のみを扱う [197].

[*3] 統計学的な意味でも仮説と呼べることが前章で明らかになった．

てくる前に状態変化を予測させてくれる．本章ではこのような状況変化の先見情報こそ記号の存在理由であると考える．

シェマモデルは環境の変化に相互作用を通して気付き，適切なシェマを選択，学習することが出来た．しかし，それはあくまでも相互作用を行い，その後に気付くのであって，それ以前に気付くことは出来ない．この性質上シェマ切り替えにはある程度の時間遅れが必然的に存在する．実時間環境で生きる行動主体にとってこのように相互作用をしてから対象の性質を察知していたのでは遅い場合が多い．例えば，第 III 部で取り上げた盤上での球の制御問題において，平地から坂道に移った時，自律的にその変化に気付き知覚シェマを切り替えることで，報酬のロスを少なく抑えつつ適応出来ることを示したが，そこにも時間遅れは存在した．実際には数度，球を盤上から落としてしまうことになる．この時間遅れは実際に状況の変化によって予測からの誤差が生じ，その誤差が一定レベルまで累積した場合にシェマを切り替えるという論理から当然の帰結である．

これに対して我々人間は何らかの表示を事前に与えられれば，それにより新たな状況へ踏み込む前に自らの態度を変えることが出来る．例えば，目の前に坂があることを視覚で認識すれば，歩容を変化させて転倒することなく踏破することが出来る．逆に，この能力を積極的に利用する為に山道では事前にカーブや坂道の標識が提示される場合がある．この事前の提示こそ，その新たな状況についての標識，つまりサインと呼べるのではないだろうか．

本章ではこの文脈から行動選択に利用可能なサインを自律的に環境から切り出し記号過程を構成させるための構成論的モデルの提案を行う．その実現のために近年，新皮質や海馬において発見されている STDP 学習則を導入する．STDP 学習則は，主体の行動選択，つまりシェマモデルにおいてはシェマ選択に利用することの出来るサインを，教師なしで切り出すことが出来る．

本章ではまず，シェマモデルにおけるシェマ選択の時間遅れの問題を明確化し，次に，近年，海馬や大脳新皮質において発見されている**スパイクタイミング依存のシナプス可塑性** (STDP: Spike Timing-Dependet Plasticity)[49, 198] を導入する．その後にシェマモデルの一例としての強化学習シェマモデル (RLSM) に STDP 神経回路を結合させたハイブリッドモデルを提案する．次にシミュレーション空間上の二輪の移動ロボット Khepera[98] にそれを実装させることにより，実際に音声刺激のようなスパイク状の刺激を，自らが累増的に形成した強化学習シェマと結合させることによって，外部他者の発する記号を解釈し，行動選択に利用するとが出来るようになる

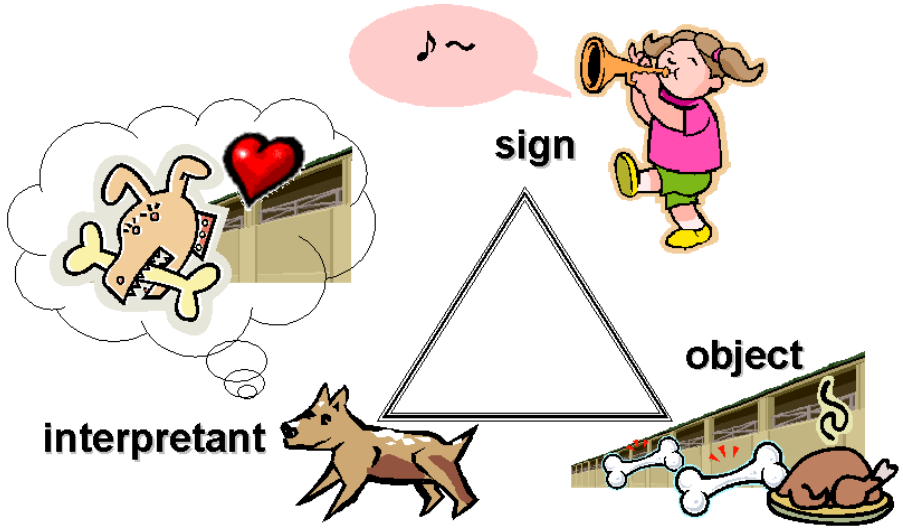


Fig.8.2: Peirce's semiotic triad with energetic interpretant

ことを示す．この学習全体に教師あり学習が一切含まれない点は特筆すべき点であり，全体としての学習器は自己組織化型学習器と見做すことが出来る．本章では代表して STDP 神経回路網の強化学習シェマへの接続を議論するが，本質的には双シェマモデルにおける知覚シェマや行為シェマへの接続も可能である[*4]．

8.2　セミオーシスの構成論

セミオーシスとはサインと対象とを解釈者が仮説推量的に結びつけて解釈する過程である．この解釈は無秩序な任意性を持つのではなく，多くの場合は連合学習や，帰納学習により解釈者内部に解釈系が獲得され，それに則った記号解釈をする場合が殆どである．Fig. 8.2 では犬が，飼い主のラッパによる合図 (サイン) を，壁際に餌が置かれたこと (対象) と解釈し，そちらに向かって行って食べようとする (解釈項) 一連の記号過程が描かれている．このような記号過程が成立するためには，それ以前にラッパが鳴った後に壁際に餌が置かれているのを発見したといったような経験が必須となる．これは古典的条件付けやオペラント条件付けといった心理学的な学習理論の図式 [204] に等しい．つまり，第 2 章でも述べたようにサインの解釈のためには，解

[*4] さらに，誤差の重み付き時間平均を用いてモジュール選択を行うクラスのあらゆるモジュール型学習器に利用可能であり，そのクラスには前章でシェマモデルと比較した MOSAIC も含まれる．

釈に先立った仮説群としてのシェマシステムが不可欠である．よって，セミオーシスの構成論を展開する為には少なくとも解釈者内部におけるシェマの組織化と，サインを仮説群に関係付ける学習の二段階の学習が必要となる．

サインから解釈される意味とは何かと問われれば，それは解釈項である．C.S. Peirce によれば解釈項は三つに分類することができる．それらは情態的解釈項，**活動的解釈項**，論理的解釈項と呼ばれる [61]．第 6 章でも述べたように，活動的解釈項は記号解釈の結果として解釈者が行う行動であり，Fig. 8.2 を例にとると，ラッパの合図を聞いて壁の方向に向かって走り出すのが活動的解釈項となる．他の例では楽譜 (サイン) を見て行う演奏 (活動的解釈項) などがそれにあたる．記号の表象する対象を例示するとき，行為ではなく第 2 章で図示したように「林檎」のような対象物を想定し，名詞的なサインの意味解釈を例示する場合が多いが，コミュニケーションにおいては相手にある行動を期待する合図や，足を止めさせる赤信号など，結果として行動に結びつく記号は多く存在する．本章で扱うのは，この活動的解釈項の組織化である．ここで注意しなければならないのは，解釈の結果行う行動についての情報はサインに記述されているわけでも，アプリオリに決まっているわけでもなく，主体自身が自律的な学習の中で獲得しなければならないということである．つまり，サインはまだ見えない報酬，被害などの可能性を示唆してくれるだけであり，そこから逃げるという行動はそれはそれ自体として学習しなければならないし，その行動の判断は自分自身で行う必要がある．

8.3　シェマ選択における時間遅れ

本論文を通じて提案してきたシェマモデルには必ず時間遅れがある．本章ではその点を明確にする．

8.3.1　シェマの選択・生成

前章までに提案してきた，シェマモデルは選択について以下の形式を持つ．主観的誤差がスカラーかベクトルかで記述は若干変わるので，スカラーである RLSM の場合で記述するが，基本的な構造はどのシェマモデルでも変わらない．シェマは並列に複数存在し，第 7 章で導入した χ^2 分布に基づいたシェマ活性度の定義を利用すれば，λ 番目の強化学習シェマについての主観的誤差とシェマ活性度は以下のようにな

る[*5].

$$R^{\lambda}(t) \equiv |\delta_t^{\lambda}/\hat{\sigma}_t^{\lambda}|^2 \tag{8.1}$$

$$\breve{R}^{\lambda}(t+\triangle t) = (1-p)R^{\lambda}(t) + p\breve{R}^{\lambda}(t) \tag{8.2}$$

$$\sim \int_0^{\infty} \frac{1}{\tau}\exp(-\frac{s}{\tau})R^{\lambda}(t-s)ds \tag{8.3}$$

$$\mathbf{V}^{\lambda}(t) = \chi_c^2(\frac{1+p}{1-p}\breve{R}^{\lambda}(t), \frac{1+p}{1-p}) \tag{8.4}$$

$\chi_c^2(x,n)$ は自由度 n のカイ二乗分布の片側確率 $P(X>x)$ である．それぞれ λ 番目のシェマの主観的誤差 R^{λ} とその重み付き平均 $\breve{R}^{\lambda}$，シェマの活性度 $\mathbf{V}^{\lambda}$ である．シェマの活性度とは，現在の環境において経験される状態，行動，報酬などの関係が，それぞれのシェマが受け持ってきた経験にどれだけ近いかを表すパラメータである．p はどれだけ過去の認識に固執するかを示すパラメータである．連続時間では時定数 $\tau = \triangle t/(1-p)$ がそれに相当する．$\triangle t$ は離散時間の 1 step に相当する実時間である．細かな議論は，6 章と 7 章を参照していただきたい．$\breve{*}$ は

$$\breve{f}(t) = \int_0^{\infty} \frac{1}{\tau}\exp(-\frac{s}{\tau})f(t-s)ds \tag{8.5}$$

で定義される積分演算子であり，後にも用いる[*6]．これらの部分の定式化は，ほぼこれまで提案してきたシェマモデルに共通した枠組みと言ってよい[*7]．シェマモデルではシェマ活性度 $\mathbf{V}^{\lambda}$ を基準にしてシェマの選択，分化を行う．シェマモデルにおいてシェマ活性度関数 $\mathbf{V}^{\lambda}$ と有意水準 $\mathbf{V}_{\alpha}$ により，λ 番目のシェマが選択される確率 $P(\lambda)$ は

$$\mu(H^{\lambda}) = sgn(\mathbf{V}^{\lambda}(t) - \mathbf{V}_{\alpha}) \tag{8.6}$$

$$P(\lambda) = \mu(H^{\lambda})\prod_{k=0}^{\lambda-1}(1-\mu(H^k)) \tag{8.7}$$

で定義される．ここで $\mu(\lambda)$ は λ 番目のシェマが $\breve{R}^{\lambda}$ に関わるサンプルを表現できるという仮説:H^{λ} の真理値となる，$sgn(x)$ は x が正なら 1，それ以外で 0 を返す関数である．H^0 はここで式を単純にするために定義した常に偽の仮説である．真理値関

[*5] 記述を簡略化するために，6 章における主観的誤差の二乗値を，本章では R で記述している．

[*6] 7 章ではこの演算子を $\overline{*}$ で表記したが，本章では $\breve{*}$ で統一する．

[*7] RLSM においても LDSM においても誤差の正規分布を仮定すれば χ^2 分布を前提としたシェマ活性度を導入することが出来る．

数として $sgn(*)$ を用いたため，上式では選択確率は必ず 0 もしくは 1 になり決定論的な選択になるが，$sgn(*)$ の代わりにシグモイド関数のような連続的な関数を用いることも可能である．その場合には上の選択式により確率的なシェマ選択が行われることになる．本章では簡単のため，決定論的な場合のみを扱う．

8.3.2　状況認識における時定数のトレードオフ

モジュール型の学習器では現在相互作用している環境に対して有効とされるシェマ (モジュール) により現在の状況を特徴付けることができる (状況認識の機構). MOSAIC[103] やシェマモデルのようにモジュール選択を予測誤差の時系列に従って行う場合，重み付け平均の時定数 τ の長さ (離散時間では p の大きさ) の決定に際しては学習の安定性と即応性のトレードオフが必ず存在する[*8].

シェマモデルではシェマ活性度に基づいてシェマ選択・生成が行われるが，シェマ活性度は $\chi^2_c(nx, n)$ が x についての単調減少関数である事を考えれば，$\breve{R}^\lambda$ に基づいていると考えてよい．$\breve{R}^\lambda$ は主観的誤差 R^λ の重み付き時間平均であるので，必ずしも現在の環境と λ 番目のシェマの関係における主観的誤差の期待値にはなっていない．重み付き平均を連続時間で表現した式 8.3 は，$\breve{R}^\lambda$ が R^λ を Fig. 8.3 左上図のような窓関数 $\mathbf{W}_-$ で積分することによって平均していることを示している．この窓関数の重心は $-\tau$ にあるため，$\breve{R}^\lambda$ は大体 τ 秒前の環境と λ 番目のシェマの関係を表している．つまり，シェマモデルを通した環境認識には必ず τ に比例した時間遅れが存在する．よって，効率のよいシェマの生成・選択の為には出来るだけ τ は小さい方がよいが，時定数 τ が小さすぎる場合にはシェマ選択はこれが，シェマモデルにおける学習の安定性と即応性のトレードオフである．

特に強化学習シェマモデルにおいては，切り替えのために用いることの出来る情報が TD 誤差という確率的な一次元情報に限られるために，この時定数 τ をかなり大きくとる必要がある．本章末で示す実験においても $p = 0.999$，1step を $0.32[s]$ として，$\tau = 320[s]$ 程度となり，結果的にモジュール切り替えのために約 $800[s]$ 近い時間の観察がシェマの生成選択の安定化の必要であった．しかし，動的な実世界で時間遅れなく行動する為には，可能ならば求めたいのは現在の予測誤差，もしくはモジュールを切り替えて利用する未来の予測誤差である．センサ値は現在の情報しかとること

[*8] mixture of experts や RNNPB のような手法においても形は違えど，このような時定数は必ず存在する．

は出来ないので，原理的に将来の予測誤差を求めることは不可能である．そこで，代替案として先見情報に基づいて予測誤差を見積もることを考える．

まず，窓関数 $\mathbf{W}_{\triangleright}$ と $\mathbf{W}_{-}$，$\mathbf{W}_{+}$ を以下のように定義する．

$$\mathbf{W}_{\triangleright}(t) = \begin{cases} \frac{1}{\tau}\exp(-|t|/\tau) & if\ t > 0 \\ -\frac{1}{\tau}\exp(-|t|/\tau) & otherwise \end{cases}$$
$$\mathbf{W}_{-}(t) = \begin{cases} 0 & if\ t > 0 \\ \frac{1}{\tau}\exp(-|t|/\tau) & otherwise \end{cases}$$
$$\mathbf{W}_{+}(t) = \begin{cases} \frac{1}{\tau}\exp(-|t|/\tau) & if\ t > 0 \\ 0 & otherwise \end{cases}$$

この三つの窓関数は $\mathbf{W}_{-} + \mathbf{W}_{\triangleright} = \mathbf{W}_{+}$ を満たす．明らかに R^{λ} を窓関数 $\mathbf{W}_{-}$ で積分したものが $\breve{R}^{\lambda}$ となる．さらに仮説推定パラメータ $a \in [0, 1]$ を導入した和 $\mathbf{W}_{-} + a\mathbf{W}_{\triangleright}$ の窓関数としての重心は $(2a-1)\tau$ となる．よって，もし時刻 t において積分値

$$\int_{-\infty}^{\infty} \mathbf{W}_{\triangleright}(s)R(t+s)ds \tag{8.8}$$

を正確に推定できれば，これと $\breve{R}^{\lambda}$ のパラメータ a による重み付き和を取ることで，その重心が現在もしくは未来にある窓関数で主観的誤差を重み付き平均した値の推定値をシェマ選択に用いる事が出来る (Fig. 8.3).

実際にある行動を選択する前に，視覚や聴覚のような遠くの情報を取れるモダリティによる先見情報として得られる事象 $\varkappa$(合図や標識など) に関係付けて上の値を推定することを考える．$\mathcal{T}_{\varkappa}$ を事象 $\varkappa$ が起こった時間の集合とすると，その事象 $\varkappa$ 発生下での式 8.8 の期待値は

$$\Omega_{\varkappa} = E[\int_{-\infty}^{\infty} \mathbf{W}_{\triangleright}(s)R(t_i+s)ds | t_i \in \mathcal{T}_{\varkappa}] \tag{8.9}$$

となる．これより，$\Omega_{\varkappa}$ を計算し，保持しておけば事象 $\varkappa$ 発生下では現在もしくは未来の主観的誤差の推定に基づいて時間遅れなくシェマを切り替えることが出来るようになる．しかし，このような値を推定することはどのような機構で可能なのだろうか？次章で導入する STDP 学習則を用いることで，この $\Omega_{\varkappa}$ をニューラルネットワークの結合強度 w に獲得させることができるようになる．

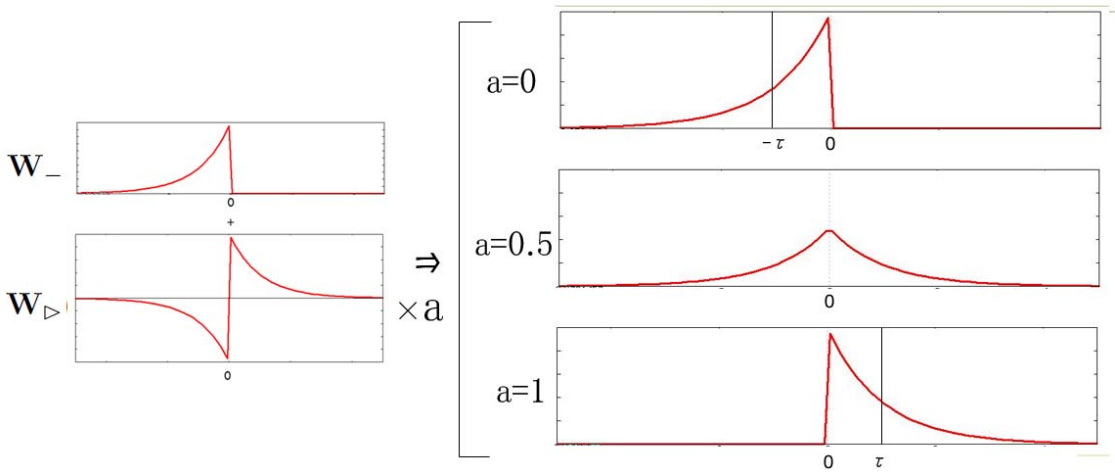


Fig.8.3: Time virtual windows used in the integral calculus of prediction errors. When one obtains some predictive information $\varkappa$, the window can be modified by changing parameter a.

8.4　STDP 学習則

本節では，シェマモデルの選択・生成についての時間遅れを解決し，セミオーシスを計算論的に構築するために STDP 学習則を導入する．

8.4.1　セミオーシスとシナプス可塑性

セミオーシスは，サインから解釈項を導き出すプロセスであり，所謂連想のプロセスに非常に近いと考えられる．脳科学の視点からこの現象を捉えた時，連想記憶を支える基本的なシナプスの可塑性は Hebb 則と呼ばれる学習則であると考えられてきた．Hebb 則は同時発火，近傍発火をシナプスの結合強化につなげる考え方である．そこでは，pre-post シナプスの発火順序は考慮されない．移動ロボット研究においても Hebb 則を用いて連合学習を行わせる研究事例は多い [63]．また，Hebb 則の持つ連想記憶装置としての性能は統計熱力学のモデルとの同型性から数理的な基盤が形成されてきた [141]．しかし，近年その神経科学的な基盤が疑問視されはじめてきている．Hebb 則とは異なるシナプス可塑性が発見されたのである [49]．Hebb 則は pre シナプスと post シナプスの同時発火をトリガーに結合を強めるが，発見され

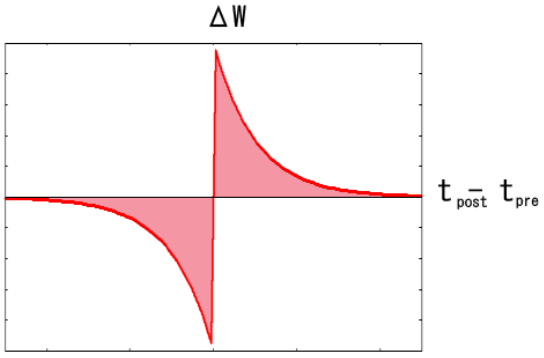


Fig.8.4: Asymmetric synaptic efficiency modulation function in spike timing-dependent plasticity

たシナプス可塑性，STDP(Spike Timing-Dependent Plasticity) は pre，post の発火の順序に依存して増強させたり減弱させたりする [5]．この法則を認めると，Hebb 則で実現される様々な機能が実現されないことが徐々に明らかになり，問題視されている [198]．これに対して筆者らはこの STDP の持つ順序依存性が，むしろ記号の持つ，遠隔，未来の情報の伝達という機能と密接に結びついていると考えている．本章では，この STDP 則を用いて聴覚情報のようなインパルス刺激を行動選択に利用する方法を提案する．

8.4.2 STDP 則による $\Omega_\varkappa$ の推定

STDP 学習則は大脳新皮質や海馬において発見されているシナプスの発火タイミングに依存した学習則である [198, 5](Fig. 8.4)．具体的な STDP 学習則については様々な形が提案されているが，その設定に応じて性質が様々に変わる．また，STDP 学習則を固定した場合にランダムなパルスが入ってきた場合の結合強度の変化についての解析的研究は進んでいるが [198, 79]，STDP の持つ機能的側面についての報告は未だ少ない．連想記憶においては Hebb 則と同様な効果が得られる場合もあれば得られない場合もある．我々は Hebb 則によって実現されていたパターン想起という形ではなく，モジュール選択という立場から連想記憶を捉えると STDP 則を自然に用いることが出来ることを発見したので本章で報告する．一般的に STDP 学習則は以

下のようにかける [198].

$$\triangle w = \sum_{i,o} W(w, t_o^{post} - t_i^{pre}) \tag{8.10}$$

$$W(w, \triangle t) = \begin{cases} \frac{S_+(w)}{\tau_+} \exp(-|\triangle t|/\tau_+) & if\ t > 0 \\ -\frac{S_-(w)}{\tau_-} \exp(-|\triangle t|/\tau_-) & otherwise \end{cases}$$

w は pre-post シナプス間の結合強度であり，t_i^{post} は post シナプスの i 番目の発火時刻，t_j^{pre} は pre シナプスの j 番目の発火時刻である．$W(w, \triangle t)$ は Fig. 8.4 に示すように左右非対称の関数になっており，発火タイミングが pre-post で入れ替われば，結合の増強 (LTP) と減弱 (LTD) が切り替わる．また，τ_+ と τ_- は STDP 学習則の時定数であり，S_+ と S_- は学習のゲインである．一般の STDP 学習則ではゲインは結合強度に依存する形で定式化されるが，本章では定数である場合のみを考える．時定数の非対称性は考慮せずに $\tau_+ = \tau_- = \tau$ とする．また，この学習則では規則的な入出力を pre-post で行わせると w は発散するので，上限と下限を設ける [79] か減衰項を設計しなければならない [5]．本章のモデルでは入力スパイクが入った時に結合強度に比例して減衰する項を設計する．

$$\triangle w = \sum_i -S_{decay} w(t_i^{pre}) \tag{8.11}$$

S_{decay} は減衰項のゲインである．ここで，$\tau_1 = \tau_2 = \tau$ およびゲインが定数であると仮定して，STDP 則を微分表記すると以下のようになる (補遺 1 参照)[5].

$$\dot{w} = S_+ I\breve{O} - S_- O\breve{I} - S_{decay} wI \tag{8.12}$$

ここで pre シナプスからの入力スパイク列 I と post シナプスからの出力スパイク列 O を以下のように定義している．

$$I(t) = \sum_{t_i \in \mathcal{T}_I(T)} \delta(t - t_i) \tag{8.13}$$

$$O(t) = \sum_{t_o \in \mathcal{T}_O(T)} \delta(t - t_o) \tag{8.14}$$

$\mathcal{T}_I(T), \mathcal{T}_O(T)$ は時刻 T までの pre，post のスパイクの発火時刻の集合であり，$\delta(t)$ はスパイクをあらわす Dirac のデルタ関数である．I についてはデルタ関数の和で表されることを仮定して議論を進めるが，O については一般の関数であっても構わな

い．さて，このように STDP 則を設定した時に何の情報がネットワークの重み w に獲得されるのだろうか？

今，前章の最後で導入した以下の期待値計算を考える．式 8.9 で $t < T$ までのスパイクに制限し，I がデルタ関数の和であることを考慮し変形すると

$$\sim \frac{1}{\#(\mathcal{T}_I(T))} \int_{-\infty}^{T} \int_{-\infty}^{T} I(t) \mathbf{W}_{\triangleright} O(s) ds dt \tag{8.15}$$

$$= \frac{\int_{-\infty}^{T} O\breve{I} - I\breve{O} dt}{\int_{-\infty}^{T} I dt} \tag{8.16}$$

が得られ，これを時間で微分すると

$$\dot{w} = \frac{1}{\#(\mathcal{T}_I(T))} (O\breve{I} - \breve{O}I - wI) \tag{8.17}$$

が得られる (補遺 2 参照)．これは，$S_+ = S_- = S_{decay} = 1/\#(\mathcal{T}_I(T))$，$\tau = \tau_+ = \tau_- = 1$ とした STDP 則そのものである．ここでは，STDP のゲイン S が入力スパイクの経験数 $\#(\mathcal{T}_I(T))$ によって焼きなまされていくが，生理実験においてはそのような示唆は今のところ報告されていない．これに対し，既存の STDP 則を導出するためには期待値計算式を，過去を忘却するように重み付き期待値計算に変更すると，入力スパイクが断続的に入り続けているという条件下では，以下の更新則が得られる (補遺 3 参照).

$$\dot{w} = \frac{1}{\nu} (O\breve{I} - \breve{O}I - wI) \tag{8.18}$$

これは，$S_+ = S_- = S_{decay} = \frac{1}{\nu}$ とおいた時の STDP 則に他ならない．つまり，post シナプスが自励的に主観的誤差 $R^{\lambda}(t)$ を発振し続けているとすると，減衰項を付与した STDP 則は前章末に示した $\Omega_{\varkappa}$ を逐次計算しネットワーク重みにコーディングする仕組みに他ならないと解釈できる．ここで，先見情報 $\varkappa$ とは pre シナプスからの入力スパイクのことに他ならない．よって，モジュールの予測誤差を強制的にスパイクとして出力する仮想的な post ニューロンに聴覚情報などをスパイク状に入力するニューロンを接続した際には，ネットワークの重みは STDP 則に従うことで，自動的に予測誤差の将来予測を立てるための情報 $\Omega_{\varkappa}$ を自らにコーディングしていく．これにより聴覚や視覚から入力されるサインをシェマ選択に効果的に利用することが期待できる．

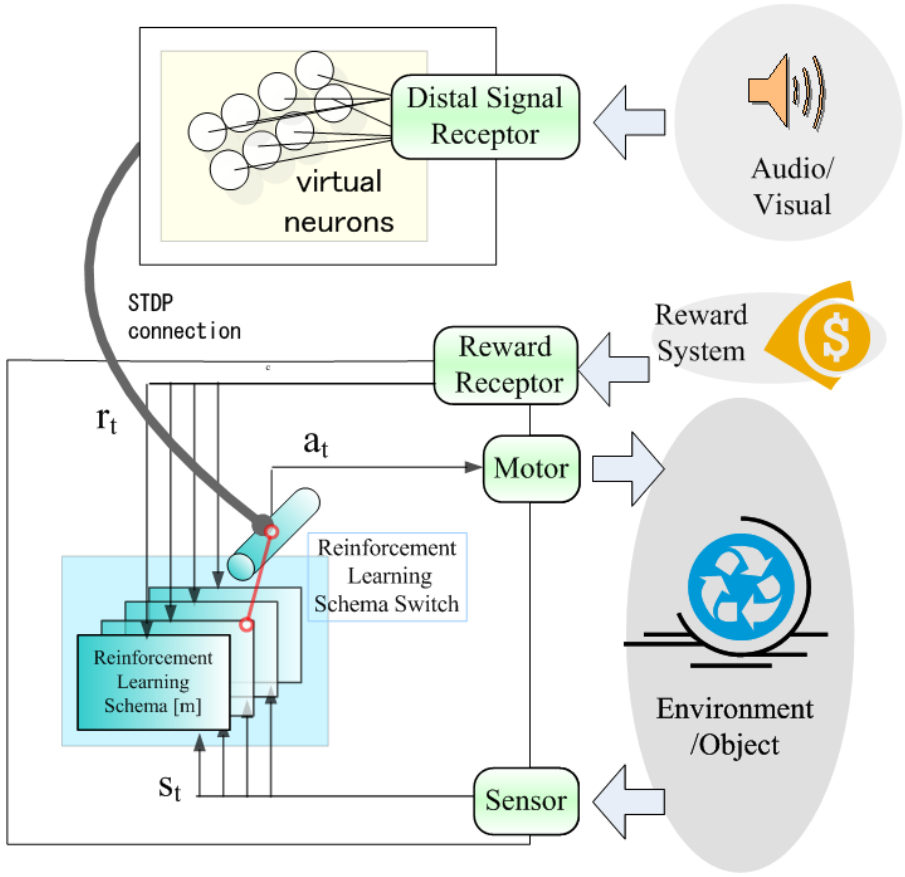


Fig.8.5: The reinforcement learning schemata model with a STDP neuronal circuit

8.4.3　強化学習シェマモデルへの STDP 神経回路網の接続

前節で提案した STDP 則により可塑的に結合重みを変化させるニューラルネットを主観的誤差を観測するたびに強制発火するシェマ群に接続したネットワークを考える (Fig. 8.6)．λ 番目のシェマは内部状態として $\check{R}^\lambda$ を持つが，前章で述べたように，これに STDP で結合強度に学習した結果を $a \in [0,1]$ の係数をかけて上乗せすることにより，主観的誤差の未来予測を得ることができる．本節で特に強化学習シェマモデル (RLSM) に接続する場合を考える (Fig. 8.5)．以下の内部状態を考える．

$$\dot{\Psi}^\lambda = \sum_i (w_i^\lambda - \Psi^\lambda) sgn(w_i^\lambda - w_{dz}) I_i(t) - \frac{1}{\tau}\Psi^\lambda \tag{8.19}$$

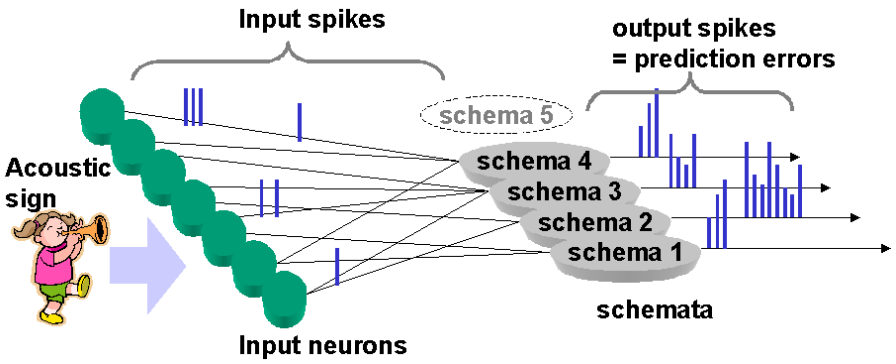


Fig.8.6: A STDP neuronal circuit connected to RLSM

ここで w_i^λ は i 番目のニューロンから λ 番目のシェマへの結合強度をあらわす．I_i は i 番目のニューロンから発された入力信号である．右辺第一項が基本的に最近に入った入力スパイクの通る結合の強度 w_i^λ を反映するものであり，右辺第二項がその影響を時間的に減衰させる．STDP で学習された重みのうち入出力に時相関が無いものは $w \to 0$ に収斂していくが，それらの入力による有意でない未来予測を Ψ^λ に反映させないためには，不感帯 w_{dz} 設定する必要がある．主観的誤差の時系列 O^λ がその平均値を中心に分散 $var[O^\lambda]$ のガウス分布をしていると仮定して構わない場合には[*9]，ランダム入力に対するそのネットワーク重みは分散 $var[w] = 2var[O^\lambda]S_i^\lambda$ のガウス分布になるので，適当な正の係数 k によって $w_{dz} = k\sqrt{var[w]}$ と設定する．k が大きければ大きいほど相関の無いニューロンからの入力は無視されるが，相関のあるニューロンからの入力も無視される可能性が高くなる．S_i^λ は i 番目のニューロンから λ 番目のシェマへの結合の学習率である．実際には主観的誤差の時系列は正確なガウス分布をしない場合が多いので $var[w]$ は過大評価されがちになるので k による適当な補正が必要である．

$\varkappa[i]$ を i 番目のニューロンを発火させる感覚刺激を起こす事象であるとすると，主観的誤差の変化に十分な相関を持つ事象存在下では上の式によってもとまる Ψ は $\Omega_{\varkappa[i^*]}$(ただし $i^* = \arg\max_i(t_i)$) の推定値に減衰項 $exp(-\frac{t-t_{i^*}}{\tau})$ を掛けたものになっている．よって

$$\breve{R}^\lambda(t) + a\Psi^\lambda \tag{8.20}$$

[*9] ここで O^λ の分散とは，サンプリング周期毎に積分した値の数列の分散のことである．

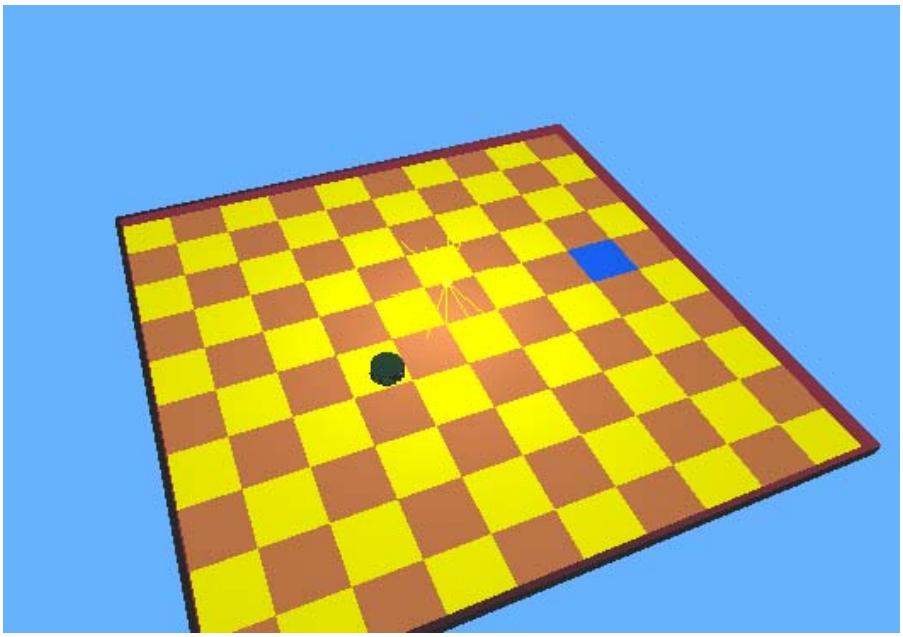


Fig.8.7: Simulation environment

は重心が $(2a-1)\tau$ 秒後になる窓関数でシェマ切り替えを駆動する主観的予測誤差を積分，つまり重み付き平均した値の推定値となる．これを $\breve{R}$ の代わりに用いることにより，エージェントは聴覚や視覚等の事前情報からモジュール切り替えへの連想関係を時間遅れなく記憶・想起することができるようになる．

8.5　実験

前章で提案したハイブリッドモデルを用いて活動的解釈項を自律的に組織化させる実験をおこなう．

8.5.1　実験環境

シミュレーションは cyberbotics 社の Webots を用いて行った [98]．二輪の小型移動ロボット Khepera を一辺 2[m] の壁で囲われた正方形状の領域を移動するという系を考える (Fig. 8.7)．領域の中心には高さ 10[cm] に光源を置いた．Fig. 8.8 に示すように，Khepera はモータ系として左右の車輪を持っており，毎 step においてその回転速度を指定することが出来る．また，センサ系として前面に赤外線センサと光セン

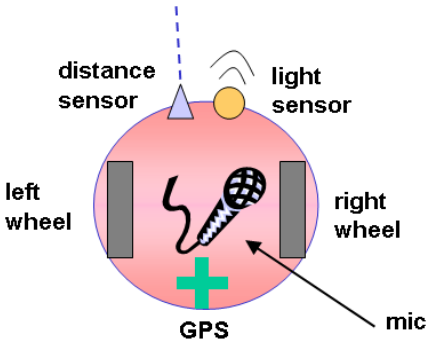


Fig.8.8: Khepera's overview and its sensor-motor system

サ，および GPS を備え付けた．また，仮想的な音声入力を捉えるために，仮想的なマイクをつけた．通常 Q 学習を用いる際には適当に状態行動空間を離散化する必要がある．状態空間は GPS から得られる Khepera の x，y 座標，角度をそれぞれ 6 等分し $216 = 6 \times 6 \times 6$ 個状態に離散化した．また，行動空間は前進，後退，静止，左右旋回の 5 個の行動に離散化した．前進後退は 1step で約 30[cm]，旋回は約 60 度の超信地旋回を行う．赤外線センサの値 $(0 \leq ds \leq 1)$ と光センサの値 $(0 \leq ls \leq 1)$ は報酬を計算するためだけに，マイクはニューラルネットワークへの入力スパイクを獲得するためだけに利用する．近接赤外線センサは通常 0 で，壁に接近したときのみ 1 に近い値を出す．

報酬は以下の 3 種類を用意した．

$$r^1 = ds \tag{8.21}$$
$$r^2 = 1.5ls \tag{8.22}$$
$$r^3 = 1.8v_{forward} \tag{8.23}$$

$v_{forward}$ は静止，後退，旋回時は 0，前進したときは最大を 1 とし前進した距離に比例して与えられる値である．ここで報酬を Khepera にとってのエネルギー源もしくは栄養分と読み替えて意味づけすると，上の状況は以下のようにできる．

1. r^1 の時間帯では壁の近くに行けば栄養分が漏れ出していて，壁に近づくほどエネルギーがもらえる．
2. r^2 の時間帯では光源から栄養分が出ていて光源に近づくほどエネルギーがもらえる．
3. r^3 の時間帯では地面から栄養分が湧き出していて，走り回るほどエネルギーが

もらえる．

ただし，今，どの時間帯であるかは，Khepera には直接的にはわからず，動き回って，実際にエネルギーが得られたり得られなかったりする中で初めて知ることが出来る．また，学習に用いるメタパラメータはそれぞれ $\alpha = 0.2$, $\gamma = 0.8$ とした．STDP 学習則については，忘却項を設定せず，学習のゲインが焼きなまされていく方法 (式 (8.17)) を用いた．力学系のシミュレーションは Webots に備え付けられたものを用い，神経系のダイナミクスは $1step$ を最小単位とするオイラー近似により行った．

8.5.2　実験 1：行為の累増的獲得

報酬関数がシフトする動的な環境下で，強化学習シェマモデルが複数の行為概念を獲得，想起できることを示す．実験は 200step を 1episode として行う．行動は Q 値を用いてボルツマン選択を行うが，episode の前半の 100step は逆温度定数 $\beta = 0$ として探索行動 (exploration) をとらせる．次の 50step は $\beta = 1$ として行動させ，最後の 50step は $\beta = 3$ に設定し知識利用 (exploitation) な行動を選択させた．実際にはシミュレーション空間上で 1step は 0.32[s] であるが，簡単のため 1step を時間の単位として取り扱う．実験は以下のように環境の一部である報酬系が以下のようにシフトしていく系にエージェントを相互作用させる．このとき外界の変化を明示的に教えることはしない．ロボットは環境との相互作用から自発的にその変化に気づかなければならない．

1. 1〜750episode　→ 状況 r^1
2. 750〜1500episode　→ 状況 r^2
3. 1500〜2250episode → 状況 r^3
4. 次に 250episode づつ，$r^1 \rightarrow r^2 \rightarrow r^3$ と変化させる．

この時，シェマの活性度 $\mathbf{V}^\lambda$ は Fig. 8.9 上図のように変化し，その過程で初めは一つだった強化学習シェマが分化のプロセスを通じて，それぞれの報酬系に対応した三つの強化学習器，結果として行為概念を組織化した．また，このシェマ活性度によって，Fig. 8.9 下図のように各シェマが選択された．これより，それぞれのシェマが各報酬系に対応する形で得られていることがわかる．

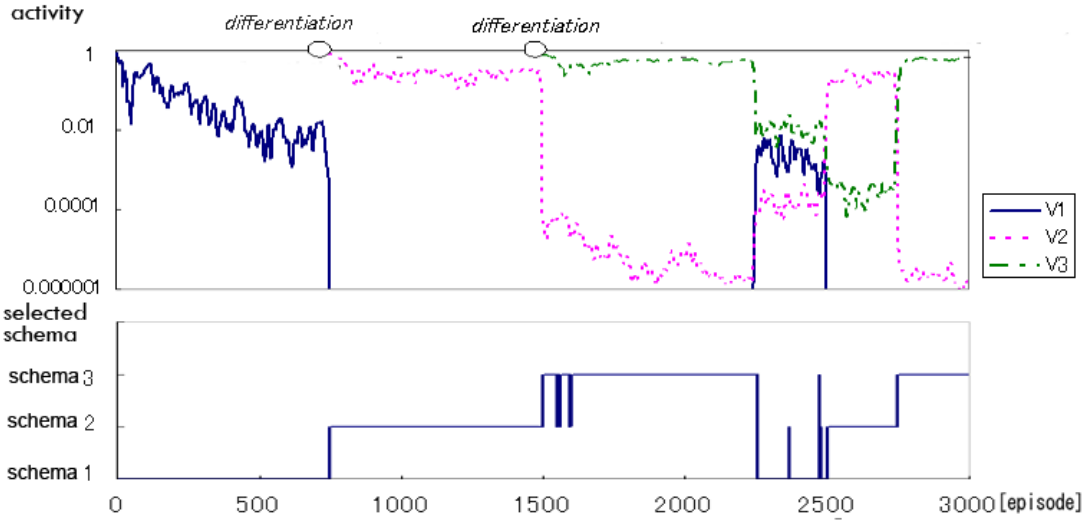


Fig.8.9: Top：time course of schema activities and differentiations, bottom： selected reinforcement learning schema.

8.5.3　実験 2：シンボルの獲得

次に，報酬系を切り替えながら切り替えるタイミングで合図となる音を発することにより，解釈系の学習を行わせる．前節では明示的な指示なしに，報酬系を含めた環境との相互作用を通して複数の行動器を獲得し，状況に合わせて切り替えることが出来た．しかし，その切り替えは実際に予測誤差が発生してから，それを十分に観測し現在選択しているシェマと有意に異なっていれば切り替えるという段取りを踏むために，平均約 2000step もの時間遅れがあった．その間に，実際には報酬において損失を出している．例えば，光を見る行動をとっていた時に，光を見ていても報酬が得られない状況であることを，光を見続けていても予想よりも報酬が得られないことを通して段々気づいていき，その結果その行為をやめて，他の行動を選択し始めるといった流れである．そこで，本章で導入した STDP 則にしたがって可塑的に先見情報をモジュール切り替えに利用する枠組みを強化学習シェマモデルに付与して，追加実験を行う．このフェーズでは，先の実験で獲得した Q 値を用い，強化学習自体による追加学習は行わなず，STDP 学習のみを実行する．

擬似聴覚ニューロンは音の高さと音色のような特徴量に相当する二次元ベクトルを

適当に離散化し，現在鳴っている音に相当するニューロン発火するというものを考える．本実験では各次元 6 つに分割し $36 = 6 \times 6$ 個のニューロンを用意し，報酬系を 1 から 2 に切り替える瞬間に $(2, 4)$，報酬系を 2 から 3 に切り替えるときに $(3, 6)$．報酬系を 3 から 1 に切り替えるときに $(1, 2)$ のベクトルが受け持つ音を鳴らす．さらに，ノイズ環境下でも報酬系の切り替えに関係する音のみを連合学習できることを示すために，他のニューロンに相当する音を時相関無くランダムに鳴らす．イメージを湧きやすくするために，二つの特徴量は (音色, 音程) であるとし，6 つの音程にはド〜ラの音が対応していることにする．この時，レの音が壁面から，ファの音が光源から，ラの音が地面からエネルギーが供給され始める合図としてデザインされている．

まず, 比較的長い時間間隔で切り替わる系において活動する場合について実験する．25episode 毎に 3 つの報酬系を順次切り替えるという実験を行った．3×25episode を 20 回繰り返した．STDP 則による先見情報の寄与率 a が 0 の場合と 1 の場合のについて，シェマの選択までの時間遅れ (時間ロス) の違いを求めた．STDP ありでは初期学習フェーズでは $a = 0$ にして，10episode が終わった時点で $a = 1$ に増加させた．この時，モジュール切り替えまでの時間ロスは Fig. 8.10 のように推移した．実験は 5 回行い平均を取っている．右軸は報酬系の切り替え回数を，縦軸には報酬系を r^1 に切り替えてから，1 番目のシェマが選択されるまでの時間を step 数で表している．学習則により，モジュール切り替えのための時間のロスが低減されていいくことがわかる．これは，RLSM における概念分化によって獲得されたシェマをベースにして，擬似音声情報をサインとして STDP 学習則によって意味づけすることにより，先見情報として利用することが出来るようになっていっていることを意味している．

次に，比較的短い時間間隔で切り替わる系を相手にする場合について実験する．本実験では 1episode 毎に報酬系を順に切り替える．このとき 1episode は 200step であるので，先見情報が無い場合通常 2000step 程度を要する切り替えを行うことは出来ない．しかし，擬似音声を STDP による可塑性を通して記号として利用することが出来るようになり，リアルタイムでのシェマ切り替えが可能になる．Fig. 8.11 に示すように，十分 STDP による学習が進んだ後では，強化学習シェマモデルの持つ時定数より細かい報酬系の切り替えに対しても，追従できるようになった．

また，このときの報酬の初期値を 0 にした重み付き移動平均は Fig. 8.12 ように推移していった．音声情報を記号として利用し，活動的解釈項としての行動に結び付けていくセミオーシスが報酬によって評価される性能の面でもエージェントのパフォーマンスを向上させていることがわかる．

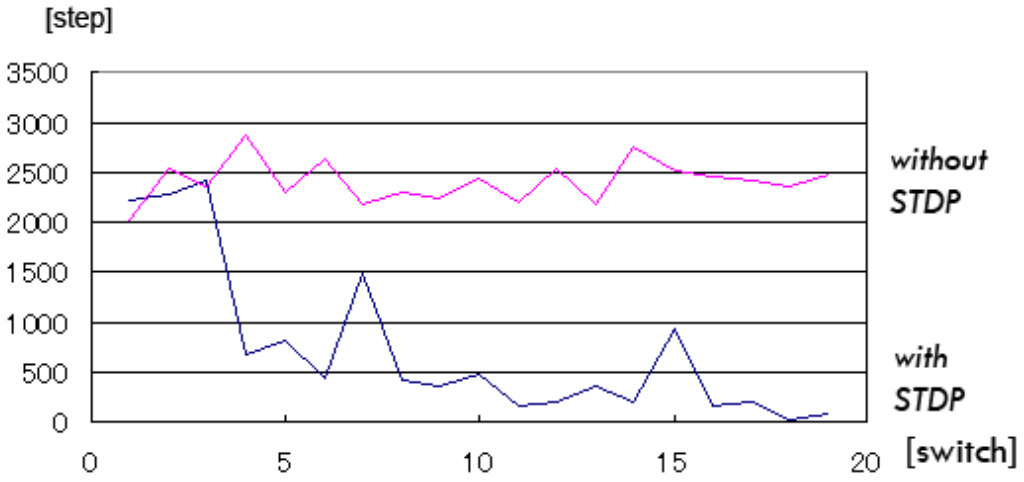


Fig.8.10: Time course of time delays in switching of schemata

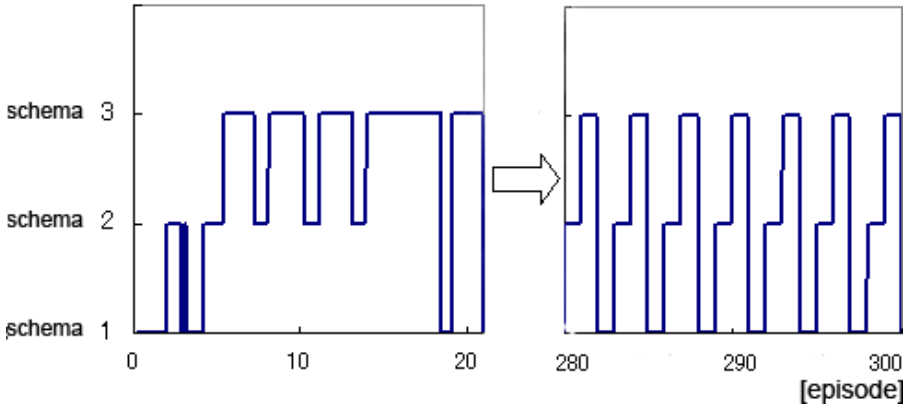


Fig.8.11: STDP makes it possible for the RLSM to switch schemata so frequently.

最終的に得られた，音を表す座標値である各ニューロンから各シェマへの結合の強さを表すマップを Fig. 8.13 に示す．それぞれのシェマを想起させる音からの結合が負の値を示し，黒く塗られている．音が鳴った後にその状況では無くなるという，逆の因果関係を持つ音については白く塗られている．また，それ以外の音については実験中報酬系切り替えと時相関無く継続的にランダムに入力したが，主観的予測誤差と因果関係を持たないので結合重みが 0 付近に収束していることがわかる．これは音という特にアプリオリには境界の存在しないチャンネルがエージェントの活動を通じて差異化されていったことを示している．解釈者としての Khepera の立場に立つ

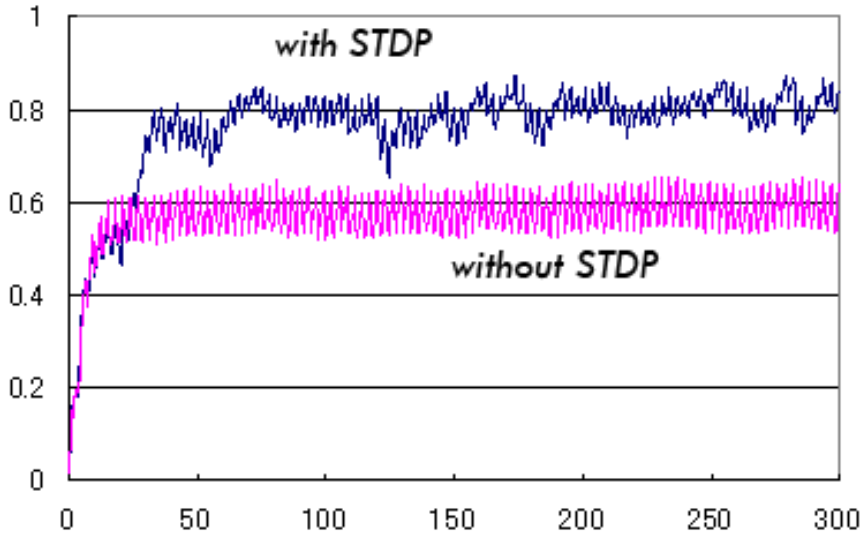


Fig.8.12: The real-time switching increases the averaged reward.

と，灰色領域の音を聞いたときには何も解釈しないが，白い領域はその行動を抑制する意味を持つ．黒色の領域の音を聞いたときには，近い未来の報酬系が予測されるので，その報酬を最大化するような行動をとろうとする．例えば，レの音が鳴った時には，壁側に向かう行動を担うシェマの予測誤差に減弱のバイアスがかかる．これは，レの音が鳴った後にやってくると予期される状況に対して壁側に向かう行動が適当だという予測が立ったことを意味している．それにより壁際に向かう強化学習シェマが選択され Khepera は壁際に向かう．これはまさに，Fig. 8.2 において，ラッパの音 (サイン) を聞いた犬が，それを壁際に餌が置かれる (対象) であろうことを予測して，壁際に走り出す (解釈項) のと同様である．故にこれは活動的解釈項としてとらえられ，Khepera は C.S. Peirce の三項関係としての記号を獲得することが出来たと考えられる．

8.6　まとめ

人間の認知する世界を満たす記号には他者からの意図的な伝達のほかにも多くの機能がある．R. Thom は C.S. Peirce の Icon，Index，Symbol という記号の三分類

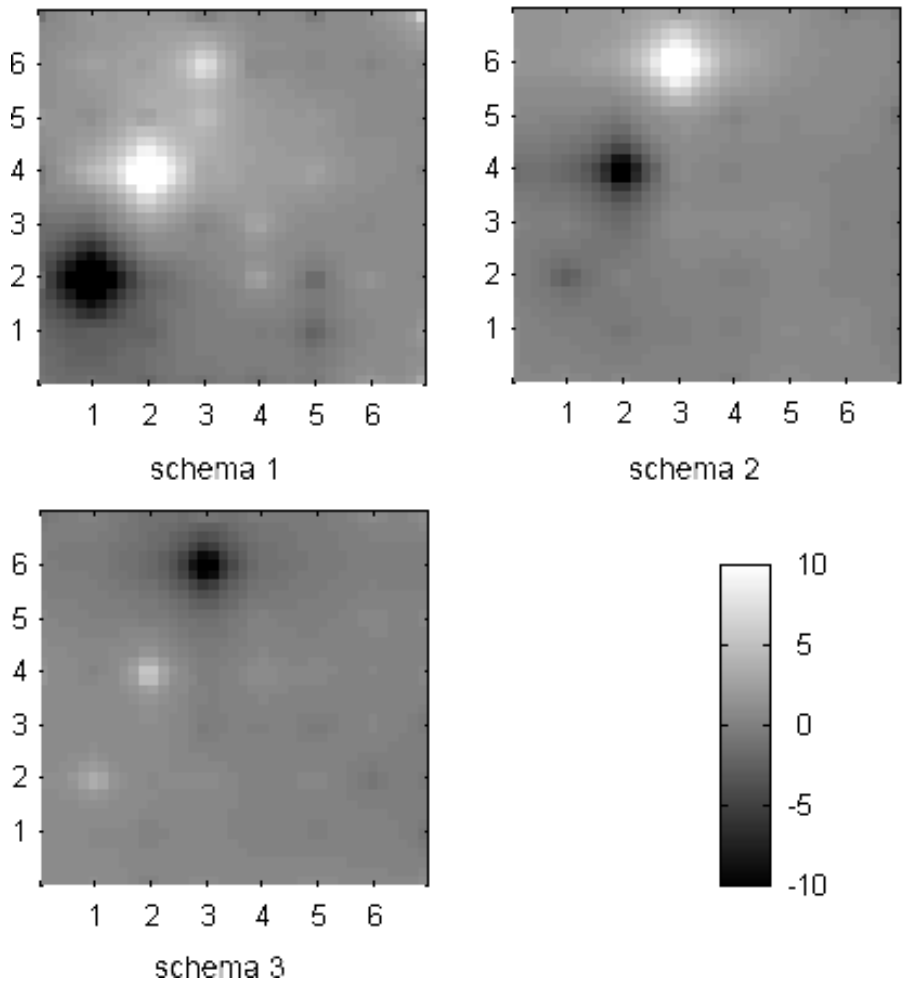


Fig.8.13: Efficacy of synaptic connections from each virtual auditory neuron to each schema

[61] を時間・空間的な視点から再解釈する際に，Symbol はサインが現在に存在し対象が未来に存在するものであるという解釈を示している [135, 92]．記号の存在意義は「そこに無いもの」についての情報をサインで表象することにより，解釈者に情報を引き出させることにある．この引き出された情報が解釈項として新たな記号を生み出すプロセスこそがセミオーシスである．記号の原初的な姿は，捕食や逃走といった生存競争において中心的となる近距離的な力学的な相互作用とは異なった，映像や，音声といった，遠くの情報が取れるチャネルを用いた情報利用であると考えられるが，その記号の意味は多くの場合アプリオリに定められるものではなく，解釈者が構成するものである [176]．そこでは，解釈者がその現前する映像や音声をどう分節化

し，それらからどのような意味を引き出すかは，解釈者がそれまでに獲得してきた解釈系に強く依存する．つまり，本来，人間のような自らの感覚運動系で閉じた自律的閉鎖系は，そのような解釈系自体を環境や他者とが関わる中で構成していかねばならない [40]．RLSM は環境との相互作用を通じて，再利用可能な運動を累増的に学習していく枠組みである．しかも，シェマの数すら時系列情報を組織化することで得られるために，RLSM では，全ては，エージェント自身の身体性と環境との相互作用によって生み出される．エージェントが，その獲得された構造をベースにして自らの知覚情報を STDP の持つ時間非対称な因果性抽出機構を用いて関連付けていくプロセスは，まさに，人間が知覚情報を自ら生成した解釈系で解釈し，自らの記号体系を組織化していくプロセスと重なって見える．

行動主義心理学の立場から言えば，実現された実験結果は条件付け学習そのものであるように見える．条件付け学習では学習を刺激反応対が形成されるという描き方がされるが，本章の枠組みでは聴覚刺激の先見情報としての利用は重み a を 0 から 1 に動かすことによって，仮説推論により得た情報を利用することもしないこともできるので，刺激反応対が形成されたことにはならない．しかし，条件付けのプロセスによる一種の連合学習は記号過程が生まれるための基本的な学習能力と考えられる．その点においては，本章での議論は条件付け学習を記号論の枠組みから論じたと言うことも出来る．

本章では RLSM の枠組みと大脳新皮質や海馬で見つかっている STDP 学習則を組み合わせた記号創発のためのハイブリッドモデルを提案した．これを用いることにより，エージェントは教師なしで学習していく中で，自律的に行動生成を行い，さらに因果的にそれと関係しそうなサインを擬聴覚刺激から切り出すことが出来るようになった．そして，そのサインを仮説推論的に前もって形成していたシェマに連合することによって，時間遅れなく適切な行動を活動的解釈項として想起するセミオーシスが実現された．

さらに本稿はそれぞれについて詳述はしないが，双シェマモデルにおける行為シェマや知覚シェマの選択にも形式的同型性から明らかに同じ枠組みが利用可能である．つまり，この枠組みは行為へのセミオーシスだけでなく知覚へのセミオーシスをも導くことが出来る．これにより，本章冒頭の Fig. 8.1 に図示した STDP 神経回路網を接続された双シェマモデルが完成することになる．

人間は記号的思考をする生き物だと言われるが，人間がその進化の過程で大脳新皮質が特異に大きくなったことは有名であり [20]，そこにおいて発見された STDP 則

が記号創発と計算論的に繋がったことは興味深い．しかし，このように計算論化された解釈の動態は記号論で考えられている記号の持つ性質のほんの一部に過ぎず，人間の記号論的な特質のほんの一部に構成論的なメスを入れたことにしかならない．また，個々で得られた記号は外界から入ってくる音声情報を意味づけるという意味では外に開いているが，その情報は個人の中で閉じて生成されるものであり，記号の重要な利用法である他者とのコミュニケーションの問題を含んでいない．今後はこの自らの感覚運動系に閉じた現象学的領域で組織化された記号がいかにして他者と共有していくことが出来るのかという問題へと進んで行きたい．

補遺 1：STDP の微分表記化

STDP 学習則は式 8.10 に表されるように，多くの文献ではバッチ処理的な表記がなされる [79, 78, 198, 5]．しかし，これはオンライン学習で利用するには不都合である．オンライン学習に用いる場合，力学系として微分表現される方が利用しやすい．よって，本節では式 8.10 を変形して微分表記を導く．ここでは本文中と同じくゲイン S_+，S_- が定数であり，時定数が等しい ($\tau_+ = \tau_- = \tau$) 場合のみを扱う．

まず，時刻 T までの重みの変化量 w を Dirac のデルタ関数を用いて書き直すと，

$$w = \sum_{t_i \in \mathcal{T}_I(T)} \sum_{t_o \in \mathcal{T}_O(T)} [\iint_{-\infty}^{T} \delta(s - t_o)\delta(t - t_i)W(s - t)dsdt] \tag{8.24}$$

$$= \iint_{-\infty}^{T} \sum_{t_o \in \mathcal{T}_O(T)} \delta(s - t_o) \sum_{t_i \in \mathcal{T}_I(T)} \delta(t - t_i)W(s - t)dsdt \tag{8.25}$$

$$= \iint_{-\infty}^{T} O(s)W(s - t)I(s)dsdt \tag{8.26}$$

ここで，W について増強の領域 $S_{LTP} = \{(s, t)|s > t\}$ と減弱の領域 $S_{LTD} =$

$\{(s,t)|s \leq t\}$ の二つに分解する．

$$= \iint_{S_{LTD}} -\frac{S_-}{\tau}\exp(\frac{s-t}{\tau})O(s)I(t)dsdt + \iint_{S_{LTP}} \frac{S_+}{\tau}\exp(\frac{t-s}{\tau})O(s)I(t)dsdt \tag{8.27}$$

$$= -\int_{-\infty}^{T} I(t)\int_{-\infty}^{t} \frac{S_-}{\tau}\exp(\frac{s-t}{\tau})O(s)dsdt + \int_{-\infty}^{T} O(s)\int_{-\infty}^{s} \frac{S_+}{\tau}\exp(\frac{t-s}{\tau})I(t)dtds \tag{8.28}$$

$$= \int_{-\infty}^{T} S_+O(t)\breve{I}(t) - S_-I(t)\breve{O}(t)dt \tag{8.29}$$

これを T で微分して減衰項を足せば式 8.12 を得る．減衰項の微分化については自明のため省略する．(証明完了)

補遺 2：期待値計算からの STDP 学習則の導出

式 8.15 から式 8.16 への変形は補遺 1 と同様の計算を行えば変形できる．式 8.17 への変形を行う．

$$\frac{dw}{dT} = \frac{d}{dT}\{\frac{\int_{-\infty}^{T} O\breve{I} - I\breve{O}dt}{\int_{-\infty}^{T} Idt}\} \tag{8.30}$$

$$= \frac{(O\breve{I} - I\breve{O})(\int_{-\infty}^{T} Idt) - (\int_{-\infty}^{T} O\breve{I} - I\breve{O}dt)I}{(\int_{-\infty}^{T} Idt)^2} \tag{8.31}$$

$$= \frac{O\breve{I} - I\breve{O}}{\int_{-\infty}^{T} Idt} - \frac{1}{\int_{-\infty}^{T} Idt}\frac{\int_{-\infty}^{T} O\breve{I} - I\breve{O}dt}{\int_{-\infty}^{T} Idt}I \tag{8.32}$$

$$= \frac{1}{\#(\mathcal{T}_I(T))}(O\breve{I} - I\breve{O} - wI) \tag{8.33}$$

これは 8.17 に等しい．(証明完了)

補遺 3：重み付き期待値計算からの STDP 学習則の導出

ゲイン項が定数となる STDP 学習則を重み付き期待値計算式の微分から導出出来ることを示す．

式 8.16 に過去のスパイクほど忘れるように減衰の重み項をつけた重み付き期待値計算式を以下のように定義し，順に変形する．

$$w = \frac{\int_{-\infty}^{T} \frac{1}{\nu} \exp(-\frac{1}{\nu}\int_{t}^{T} I(s)ds)(O(t)\breve{I}(t) - I(t)\breve{O}(t))dt}{\int_{-\infty}^{T} \frac{1}{\nu} \exp(-\frac{1}{\nu}\int_{t}^{T} I(s)ds)I(t)dt} \tag{8.34}$$

$$= \frac{\exp(-\frac{1}{\nu}\overline{I}(T))\int_{-\infty}^{T} \frac{1}{\nu} \exp(\frac{1}{\nu}\overline{I}(t)(O(t)\breve{I}(t) - I(t)\breve{O}(t))dt}{\exp(-\frac{1}{\nu}\overline{I}(T))\int_{-\infty}^{T} \frac{1}{\nu} \exp(\frac{1}{\nu}\overline{I}(t))I(t)dt} \tag{8.35}$$

$$= \frac{\int_{-\infty}^{T} \frac{1}{\nu} \exp(\frac{1}{\nu}\overline{I}(t)(O(t)\breve{I}(t) - I(t)\breve{O}(t))dt}{\int_{-\infty}^{T} \frac{1}{\nu} \exp(\frac{1}{\nu}\overline{I}(t))I(t)dt} \tag{8.36}$$

ν は忘却の時定数である．ここで,

$$\overline{I}(t) \equiv \int_{0}^{t} I(s)ds \tag{8.37}$$

$$\lim_{t\to-\infty} \overline{I}(t) = -\infty \tag{8.38}$$

である．負の無限大における境界条件は入力スパイクが入り続けている条件から求められる．式 8.36 の分子を q，分母を p とおいて T について微分すると，ほぼ補遺 2 と同様な変形が出来る．

$$\frac{dw}{dT} = \frac{q'p - p'q}{p^2} \tag{8.39}$$

$$= \frac{q'}{p} - \frac{p'}{p}\frac{q}{p} \tag{8.40}$$

$$= \frac{1}{p}(q' - wp') \tag{8.41}$$

$$= \frac{(\frac{1}{\nu}\exp(\frac{1}{\nu}\overline{I}(T))(O\breve{I} - I\breve{O}) - w\frac{1}{\nu}\exp(\frac{1}{\nu}\overline{I}(T))I)}{\exp(\frac{1}{\nu}\overline{I}(T)) - \exp(\frac{1}{\nu}\overline{I}(-\infty))} \tag{8.42}$$

$$= \frac{1}{\nu}\{O\breve{I} - I\breve{O} - wI\} \tag{8.43}$$

これは式 8.18 に等しい．(証明完了)

第Ⅴ部

まとめと展望

第 9 章

議論と考察

第 V 部では本論文をまとめ，展望を述べる．本章では次章で結言を述べる前に，第 I 部で示した問題提起と課題設定に対して，第 II 部〜第 IV 部までで提案してきた自律適応的認知系の構成論がどこまで答えられてきたのか，また，それによりどのような現象理解が可能になったのかについて議論する．また，本論文は自律適応的に発達する機械システムを実現するために，J. Piaget のシェマ理論に示唆を受けて一連のシェマモデルを提案してきた．このシェマモデルは結局はどういう存在であるのか議論と考察を行う．

9.1 自律適応系におけるセミオーシスの発現

序論で論じたように本研究が目指すのは統語論的に豊かな自然言語を利用するロボットの設計論ではなく，貧しい単語，文法しか持たなくともコミュニケーションにおいて意味作用を起こせるロボットの設計論である．そこでは，西垣が指摘するように，相互作用を通じた変容こそ根本的な必要条件となる．

意味のあるコミュニケーションとはなにか？我々はこの問いに工学的に答えるために二章において Shannon-Weaver のコミュニケーション理論から脱却し，C.S. Peirce の記号論にその本質を求めるべきだと考えた．なぜならば，我々人間が作動上閉じた自律適応系であることは明らかであり，コミュニケーションにおいて，符号器 (送信機)・復号器 (受信機) を直接観測しそれらをチューニング出来ることを前提に立つ Shannon-Weaver のコミュニケーション理論から安易に拡張されたモデルでは人間同士のコミュニケーションの最も肝心な部分を議論できないからだ．しかし，

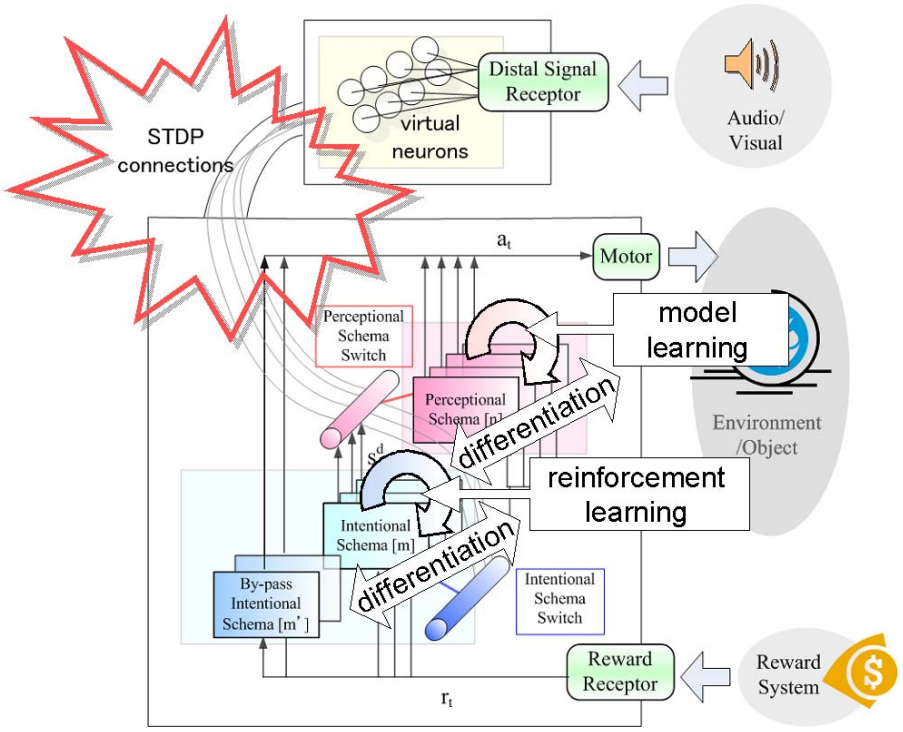


Fig.9.1: Adaptation dynamics in the Dual-Schemata model with a STDP neuronal circuit

C.S. Peirce の記号論の立場に立ち「意味」の解釈者依存性を認めても，その場その場での記号過程 (セミオーシス) がどのように起こるのかは不問にふされていた．記号過程が本質的に仮説推論 (abduction) である以上，自らの主観的世界，つまり環境世界 (Umwelt) に対する可能性，仮説を解釈者がセミオーシスの発現に先行して持たねばならない．この議論から，記号過程の前に内的に組織化された記憶構造が必要であると考え，環境との相互作用を通した記憶システムの構造化こそセミオーシスの構成論のための第一歩であると考えた．そのような環境との相互作用を通した自律主体内部での記憶の構造化を構成論的に表現するために，J. Piaget のシェマ理論から示唆を得第 II 部から第 III 部にわたって双シェマモデル (Dual-Schemata model) の定式化を行った．また，第 8 章ではこれに視覚・聴覚情報のような先見的情報を接続するための STDP 学習則を加えた．結果的に本研究では Fig. 9.1 に示すように二種類のシェマの同化・調節，均衡化・分化が絶えることなく進行し，また，その選択の時間効率を向上させるためにシェマ選択器に接続される神経回路の自己組織化型学習が

進むという，多様な学習が走る統合的な学習器を構築した．

それらは大きくわけて

1. 知覚シェマ系
2. 行為シェマ系
3. シナプス結合系

の三要素に区分できる．知覚シェマと行為シェマは環境や養育者が設計した報酬との二項的相互作用を通して環境を分節化しながら学習を進め内部シェマ系を構築していく．知覚シェマと行為シェマは学習器として独立性を保ちながらも，６章において示したように任意の組み合わせで結合することで実際にその行為を学習した環境下以外にも，その行為を汎化することや実際にその環境を学習したときに行っていた行為以外の行為にその環境概念を応用するといった再配置可能性 (relocatability) や結合可能性 (compositionality) を実現した．Y. Sugitai[81] らは RNN 内部でトップダウンに結合可能性を有したシンボル入力を与え，またボトムアップに行動学習を行わせることで，その間に系内部の力学系の拘束性から行為の結合可能性が獲得されることを示している．双シェマモデルの得た結合可能性はこのような適応的に形成されたものというより知覚シェマと行為シェマの定義の自然さから得られている．実際には創発された結合可能性というよりかは設計された結合可能性である．しかし，統語論的に結合されることで適切な行動方策となる行為シェマと知覚シェマは環境との相互作用の中で累増的に組織化していくことが出来た点は注目に値する．

また，二章において，幼児の９ヶ月革命の説と共に二項的な自律系・環境の相互作用の中から差異化された環境や行為の概念が社会的，記号的相互作用である三項関係の基盤となるという仮説を立てた．そして，この説を実際に計算論的に表現した．それが第８章で導入した STDP シナプス可塑性に基づくセミオーシスの構成論である．ここでは実際に環境との相互作用を通して自律適応系内部に構成されたシェマ系の想起をサインの仮説推量的利用によって加速させることで，実際の環境の変化が起きる前でも，そのシェマ，つまり記憶を想起させることが可能になった．これが，適切に与えられれば動的環境下で自律適応系がより高いパフォーマンスを発揮することが示されたが，その本質はセミオーシスが構成論的に実現されたことにある．しかも，このセミオーシスの実現までのプロセスは完全に教師なし学習に分類される類のものであり，シェマモデルの学習と STDP 学習則は共に自己組織化型学習に分類される．シェマモデルではシェマ単体の学習ダイナミクスはモデル学習，強化学習といった具

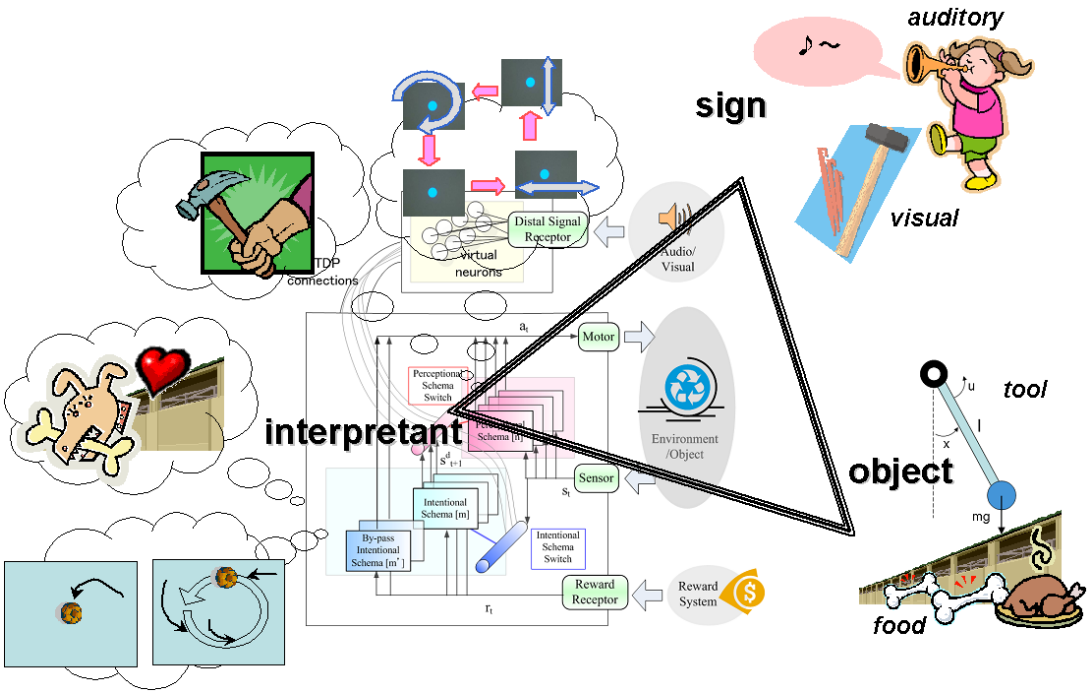


Fig.9.2: Semiotic triad over the Dual-Schemata model with a STDP neuronal circuit

体的なダイナミクスで描かれるために個別学習器としてのシェマ単体は自己組織化型学習を行わないが，シェマモデル全体としては自己組織化型学習を行っていると言える．これについては後に考察する．

Fig. 9.2 に双シェマモデルにより実現された C.S. Peirce の三項関係 (semiotic triad) のイメージ図を与える．聴覚や視覚といったチャンネルから獲得されるサインをシェマという仮説に abductive に帰着させる，つまりエージェントにおける記号過程を起こす議論については，８章における強化学習シェマモデルと STDP の接続でしか議論していないが，シェマモデル間の同化・調節，均衡化・分化というフレームワークの同型性から明らかに双シェマモデル内の知覚シェマ・行為シェマの選択にも利用可能である．Fig. 9.2 の各解釈項は各章で獲得されたシェマを表している．それらは，その訪れを告げるサインを STDP シナプス可塑性を通じてボトムアップに発見することによって，逆にサインから想起されることが可能になる．ここに徹底したボトムアップなプロセスによる自律適応系におけるセミオーシスが実現されるのである．

本研究において我々は動的環境下において主観的な環境との相互作用から差異を発

見し，シェマとしての概念を分化させ，それに基づいて聴覚刺激，視覚刺激を意味づけすることにより記号を獲得する自律適応系の設計論を得た．これは Peirce の記号論で論じられる解釈者における記号過程の，最も基本的なダイナミクスを計算論的に表現したものである[*1]．

では，これによって人間とロボットとの適応的なコミュニケーションは可能になるのだろう？

記号論は，その意味作用について解釈者に重きを置く．これは意味が作動上閉じた自律適応系内部でしか存在しえないものであるからであるが，コミュニケーションにはやはり双方向性が無くてはならない．つまり，本稿では記号の解釈についてはその構成論的モデルを提案したが，未だ記号の設計については論じていない．よって，人工物との自律適応的なコミュニケーションを実現するのはもう少し先の話だといわざるをえない．

ただし，本論文で導入した様々な適応性だけでも，より変化に富んだ社会ロボットの設計論となる可能性は十分あり，このような指針を元に設計されたロボットが関わるユーザにやはり飽きをもたらすのか，それとも直接設計された社会ロボットよりも豊かな相互作用を生むのかは興味深い問題である．実際に実機に実装する為には学習器の安定性や，自律的行動選択機構の埋め込みなど，克服すべき問題がいくつか存在する．これらを克服することは今後の課題と言える．

9.2　本研究における恣意的な埋め込み

第二章において F. Varela の作動上の閉じから，自律適応系をその閉鎖性を通して定義し，本論文を通して自ら閉じているにもかかわらず自律適応系同士のコミュニケーションがいかに可能になりえるかという点について構成論的に接近してきた．しかし，本研究で双シェマモデルを通して扱ってきた自律適応系は二章で定義した自律適応系として十分なものであっただろうか？本節では，我々が本論文の議論のなかで事前に埋め込んだものを明らかにしてまとめる．

[*1] 記号論の論じる記号過程では，類推や比喩染みた，高度な記号過程も論じられ，我々が計算論的に表現したものは，その最も簡易なものでしかない．やはり，条件付けの色彩が強いのは否めない．しかし，そのような表現すら今までの記号論ではまるで見られなかったことである．

9.2.1　状態空間と基底群の事前設計

まず一つは状態空間と関数基底群の事前設計がある．F. Varela は入力型の記述と閉鎖型の記述としてシステムの表現を区別したが，本論文第 2 章では，後者を以って自律適応系の前提とした．確かに観察者視点という面では，シェマモデルは常に自律系内部に主体の視点を置き，その観測可能な情報のみから記憶を構造化し，情報を分節化していくという構図になっている．しかし，そのシェマモデルが作動する上でのセンサ入力，モータ出力の次元や選択については，設計者がかなり意図的に設計したという点は否定できない．このようにシステムの外部の存在である設計者や観察者が状態空間を指定することは，まさにシステムを入力型の記述として描くことに他ならず，実験で取り扱った系は完全に閉鎖型の記述によるものとは言いがたい．もちろん，各実験で与えられたようなセンサ入力とモータ出力しか得られないシステムなのだと定義して議論を進めてきたが，第 3 章の実験では a_{t-1} をセンサ入力に加えるなど，かなり設計者の視点が色濃い事前設計が入っていた．

真の自律適応系ではより生のセンサ入力，モータ出力から意味のあるサブセットを抽出し，自律的に状態空間を構成していくといったダイナミクスが不可欠である．Olsson らは情報理論を用いて大量のセンサ情報からそこに隠れた構造を自律的にどう発見していくかについての議論を展開している [58]．また，杉浦らもセンサに注目し，行動とのバランスの中からいかに適切な状態空間を構成していくかという議論を行っている [82, 153]．これらの情報理論に基づいた研究が期待される．

状態空間が決まっても，それが対象のダイナミクスや目標の方策などが非線形である場合には単純な線形近似器を配置するだけでは十分に機能せず，適切な基底の設計が必要となる．第 7 章における NGSM の議論でも局所モジュールの配置は事前設計し，また第 5 章，第 6 章における強化学習でも RBF を基底として事前に配置した．基底の配置は高次空間への写像として捉えられ，状態空間構成の問題と連続的に捉えることも出来る．これについても現在，様々な角度からのアプローチがなされている．Schaal は学習データから局所的な近似器を次々と構成していく方法論を提案している [72]．また，NGnet においてゲーティング関数をも学習させる場合，それらが重なりあうことによって学習結果が悪化する場合が経験的に多い [149]．これに対し黒木や関野らはゲーティング関数による空間分割にボロノイ境界を明示的に用いたアプローチを行っており興味深い [140, 160]．他にも基底の配置の議論については多

くの先行研究が存在し，それだけでも十分に大きな領域を形成している．しかし，その多くはモデル選択と呼ばれオフラインでの手法となる．特に我々が問題にするようなオンライン学習における適応的基底設計は困難である．

当研究ではこれらの先行研究の存在を踏まえながら，これらの研究とは一線を画するかたちで，適切な基底決定した後の，しかも一つのシェマの学習が収束しえるという状況の上で，更なる上位の構造化がオンラインでどう形成し得るかという問題に絞って研究を行った．ただし，第４章で触れたように，環境と系を結ぶインタフェースとしての身体が変わればその後のシェマ形成にも影響を与える．この意味では状態空間構成もシェマの議論と独立ではなく，今後はそれらの相互作用も考慮して議論を進めて行かなければならない．

9.2.2　自律的行動選択の機構

自律ということは自ら意思決定を行い行動を決定することが基本である．本論文を通した実験では養育者や設計者によって直接は操作はされないものの，常に，“何ステップごとに行為シェマを切り替える．”といった自律ロボットのタイマー駆動的な行動選択がなされてきた．強化学習シェマモデル (RLSM) では外在する報酬の切り替えに応じて行動を変える行動選択機構が実現されたが，この行動選択は “報酬従属モード”という言葉の定義が示すように外因的であり，主体的な意思決定が実現されているとは言い難い．

感情や動機をモデル化することにより自律ロボットの行動選択の機構を設計する社会ロボットの研究も行われてはいるが [17]，現時点ではその性能を評価をする尺度が無く研究の枠組みが成立しがたい状況にある．故に実際のエンターテイメントロボットの製作というようなニーズがない学術的な場での議論は余り進んでいないのが現状である．実際，我々もそのような取り扱いがたい問題はとりあえず自律適応系研究から外して取り扱い，代わりに行動選択の機構をアプリオリに埋め込んだ．

自律主体のより自然な行動選択機構の設計論への道筋は全く無いのだろうか？自律適応系の目的を本稿で扱ったようなシェマ分化を通した “環境世界のより詳細な認識” にあるとすれば，必ずしも行動選択機構が学習・適応と独立であるとは言えない．J. Piaget は幼児の発達プロセスの中で「同化遊び」という興味深い現象について述べている [40]．それは，幼児が一つの道具，対象に対し，繰り返しぶつけたり転がしたり，執拗な相互作用を求めるという現象である．それは正に遊んでいるように見え

るが，それは様々な行為を投げかけることによって対象を自らのシェマに同化していっているプロセスであるといえる．本稿で提案した知覚シェマの計算論的枠組みから言えば，より詳細な環境認識の獲得，つまりシェマ分化のためには，その前段階としての知覚シェマの収束が必要である．この「同化遊び」のプロセス一つの知覚シェマにどんどん情報を同化させることにより，その学習を収束させていっているプロセスではないかと考えられる．シェマの収束なしには，その他の対象との区別は芽生えず，このように一つの対象に固執することによって，シェマシステムの構造化のプロセスを効率化しているのではないかという仮説が考えられる．もしこの仮説が正しいとすれば，シェマモデルにおける各シェマの収束状態と行動選択は独立ではなく，シェマシステム内部からの内因的な行動選択の指標を導くという設計指針も考えられる．強化学習の領域では探索重視 (exploration) と知識利用 (exploitation) の対立が有名であり，行動選択が強化学習の視点から議論される場合も多い [21].

実際に本研究における提案手法を社会ロボットの中で生かすにはこのような自律的な行動選択の機構との連繋も重要であり，議論を進めるべきである．

9.2.3　＜自己 $\neq$ 他者＞概念の分化

本稿で提案した双シェマモデルは全く均一なシェマから複数のシェマへと分化していくことが特徴的であった．しかし，初めの段階ではシェマモデル内は完全に未分化であったかというと，そうではない．我々は分化をシステムレベルでの調節として定義したが，J. Piaget はしばしば分化という言葉をより高次のものとして用いている．それは特に自己も他者もない中心化された世界から自己他者概念の分化を指す場合が多い．双シェマモデルでは知覚シェマと行為シェマという二種類のシェマを用意し，これらに結合可能性が成立することを 6 章でも見た．実は，知覚シェマは対象概念であり「他者」の概念にあたる．これに対し，行為シェマは自分が何を実行したいかを表象し「自己」の概念にあたる．つまり双シェマモデルは，自他の分化が完了した状態からスタートしているシェマシステムであるといえる．実際この二つが分離することが，第 III 部で示したような様々なメリットを生むわけである．行為と知覚も分かれていない真の中心化した状態からの分化による双シェマモデルの獲得といった，より基礎的なレベルからの内的構造の形成が計算論的に可能かどうかという問題は興味深い問題である．この初めの分化を完了した段階から議論を進めたために，この知覚シェマと行為シェマの分轄という制約も恣意的に埋め込んだ点であると言えよう．

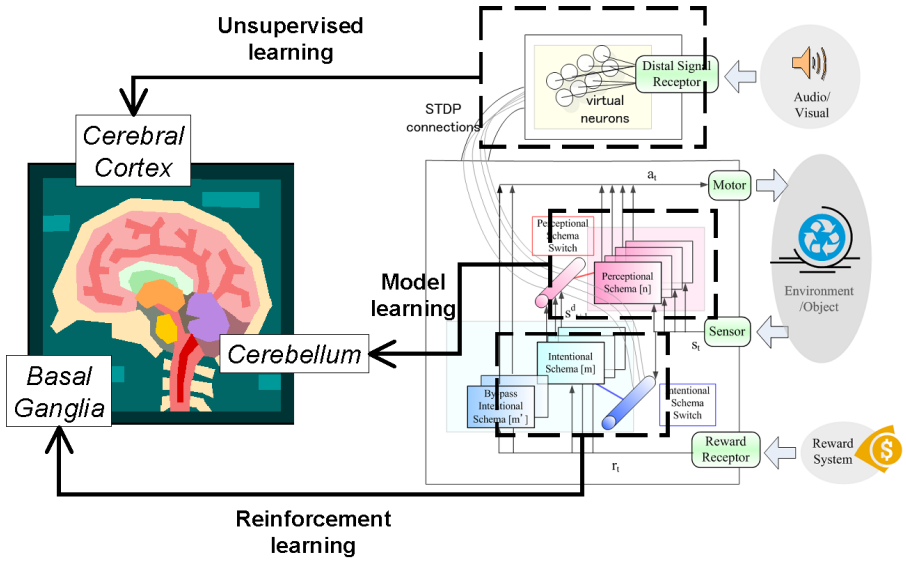


Fig.9.3: The computational similarity between human brain and the Dual-Schemata model

9.2.4　攪乱の意味づけと脳科学的知見

Fig. 2.4 に示したよう自律性を作動上の閉じから議論するならば，外界からの入力は自律システムにとってはただの攪乱に過ぎない．この攪乱に対して，状態空間を事前に設計し埋め込んだという恣意性については先に述べたが，もう一つ埋め込んだものがある．それは，入力の特徴付けである．第 II 部ではセンサ入力だけだった入力は，第 III 部で報酬入力が，第 IV 部で擬似聴覚入力が付け加えられた．これは，与えられる情報が設計者によって事前に意味づけされたようにみえる．特に報酬入力は，自律系の行動志向に特権的にアクセスする経路として存在してしまう．強化学習を行う自律系に対し，外部からの報酬値を直接的にコントロールするという考え方が実際の自律系の強化学習の構成論的モデルとして果たして正しいかどうかについては，強化学習の学習速度の遅さや並列学習の必要性からも議論され始めている [11]．　自律適応系における認知発達が全くの未分化で完全に中心化され対称的である状態から，分化を通じて対称性を崩していかねばならないならば，このような批判はもっともである．しかし，認知発達はあくまで生物学的発生のプロセスに対し，システム的に上位の階層で駆動するダイナミクスであると考えれば，このような批判は必ずしも

正当な理由をもたない．脳の計算機構は元来，視覚，聴覚，触覚などで異なることが考えられ，その意味で生物学的発生の段階で外部からの攪乱がアプリオリに特徴づけられるているという主張は必ずしも根拠不足ではない．また，学習の計算論的役割も，その分担が解剖学的に脳内でなされているという主張が K. Doya によってなされている [22]．K Doya はモデル学習が小脳 (Cerebellum) で，強化学習が大脳基底核 (Basal Ganglia) で，教師なしの自己組織化型学習が大脳皮質 (Cerebral Cortex) で行われているという，大枠での計算論的な学習理論の分類と脳の部位の対応関係についての仮説を提唱している．この仮説に基づくと，STDP 神経回路網を接続した双シェマモデルにおける各部位と人間の脳の部位の間にの Fig. 9.3 のような対応関係を考えることが出来る．これらの議論と相互作用しつつ，実際に人間が認知において攪乱をどう先天的な機構で特徴づけ，後天的な機構で意味づけるのかについての議論を進める必要がある．

本節では，本稿の議論において設計者が恣意的に埋め込んだ部分について議論した．次節では本稿を通して提案したシェマモデルの性質について議論する．

9.3　自己組織化学習シェマモデルについての議論

本論文では双シェマモデル，強化学習シェマモデル (RLSM) を初めとして幾つかのシェマモデルを提案した．大きく分けるとモデル学習[*2]をシェマの均衡化ダイナミクスとする知覚シェマ，LDSM，NGSM と，強化学習をシェマの均衡化ダイナミクスとする行為シェマ，RLSM に分けられる．ここでは前者をまとめてモデル学習シェマモデルと呼ぶ．これらに共通した “シェマモデル” という名前は，どのようなクラスの学習器を指すのかを本節では議論していきたい．シェマモデルについての解析的な議論は未開拓であり，本稿では示すことは出来ないが，他のモジュール型学習器との経験則的な比較も後に行う．まず，シェマモデルがある種の自己組織化学習のクラスを作ることを示す．

[*2] モデル学習は大きくは教師あり学習に含まれる．

9.3.1　シェマモデルの基本条件

シェマモデルとは経験が仮説を生み，仮設が経験を検定することによって構造化されていく記憶システムであるが，その基本的な枠組みは以下の通りである．

1. 各シェマが何らかの学習器をもっている．
2. 各シェマは誤差 E_t^λ を測ることができ，また，標準偏差推定 $\hat{\sigma}_t^\lambda$ を行うことができる．
3. 誤差を推定偏差で割ることにより，主観的誤差 R_t^λ を割り出し，シェマ活性度 V_t^λ に基づいてシェマを選択・生成する．

この時，記憶システムとしてのシェマシステムの次元からは各シェマが経験を同化するかどうかが問題であり，調節のダイナミクス自体はシェマ単体の学習則として扱われる．よって上の枠組みを守る範囲では，今後様々なシェマモデルを議論することが可能である．そのシェマモデルという学習器の行う学習はシェマ分化を通した情報のクラスタリングである．この意味でシェマモデルは明示的な教師を必要としないで，作動上閉じたナチュラルドリフトを通じた内部構造としてのシェマ群の組織化を行う．

本論文の各章でモデルの提案と実験による検証を行ってきた．しかし，第 7 章における統計学的な整理にも関わらず，シェマ活性度の変化を担う時定数 τ_p や同化の基準 V_α の設定方法と，そのパラメータ設定に依存したシェマ形成や安定性などの解析的議論を進める事は出来なかった．その解析的議論を行う前段階の準備を本節で行い，今後の方針を述べる．

しかし，解析的議論を行う為には入力に a_t を含むモデル学習や誤差がブートストラップ法で与えられる強化学習は複雑であり，解析的議論を行うには難度が高い．よって，より内部の学習則を簡略化したより低次なシェマモデルから解析的議論を進める事が適切な接近であると考えられる．シェマモデルに普遍な構造をよりよく理解するために，低次なクラスのシェマモデルとして ARSM(AutoRegressive Schemata Model) と MESM(Mean Estimation Schemata Model) を導入し簡単な議論を行う．これにより隠れマルコフモデルと (HMM: Hidden Markov Model) と呼ばれる確率モデルとの関係が明らかになり，統計的学習理論との接点が見えてくる．

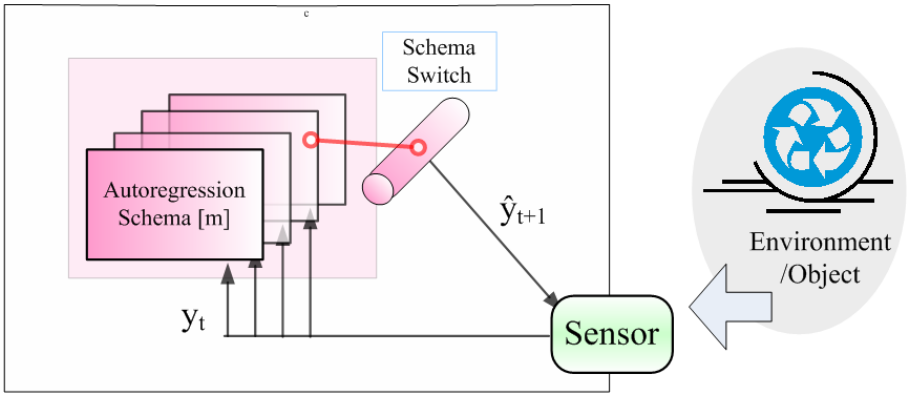


Fig.9.4: ARSM: AutoRegression Schema Model

9.3.2　低次なシェマモデル

シェマモデルとは同化・調節を通じ，仮説を生成し，その仮説により新たな経験を検定する再帰的なループにより，構成される自己組織化型オンライン学習器である．シェマモデルの一つの効果は与えられた時系列データをクラスタリングすることにあるが，バッチ的なクラスタリング手法と違い，過去の学習データに応じて現在の学習データが分類されるために，自己組織化マップや連想記憶ネットワーク [141] に近い学習則となる．ただし，重要なのはそのクラスタリングがシェマ単体内部での同化する情報のコヒーレントさ維持に繋がり，シェマ単体レベルでの学習器の性能を複数シェマの協調により上昇させる傾向があることである．本節ではモデル学習シェマモデル，強化学習シェマモデルよりも簡易な学習則をシェマ単体の基礎ダイナミクスとしてもつ簡易なシェマモデルを導入する．

自己回帰シェマモデル

ARSM(AutoRegression Schema Model) は時系列予測モデルである AR モデルをシェマの内部関数としてシェマモデルである．概観を Fig. 9.4 に示す．簡単に考えるとモデル学習シェマモデルからモータ出力の a_t 項を無くしたシェマモデルと考えら

れる，q 次の AR モデルは以下の式で表される．

$$\hat{y}(t+1) = \sum_{k=0}^{q-1} \hat{a}_k y(t-k) \tag{9.1}$$

である．ここで時刻 t での観測値を $y(t)$ 予測誤差は $e(t) = y(t) - \hat{y}(t)$ となる．これを用いて推定偏差 $\hat{\sigma}$ は以下のように近似されていく．

$$\hat{\sigma}(t+1) = \sqrt{(1-\eta)\hat{\sigma}(t)^2 + \eta e(t)^2} \tag{9.2}$$

AR モデルの係数 a_k の学習は勾配法を用いればよい．シェマの主観的誤差は $R(t) = e(t)/\hat{\sigma}(t)$ で求まるので，この後はシェマ活性度の計算，及び選択・生成は 7.3 節の法則に従えばいいことになる．

モジュール型の学習器で時系列予測のずれを広い意味でのモジュール切り替えの指標とし，自律系内部での認識を生成する研究には N. Kubota や [47] や J. Tani[88, 89] のものがある．これらは状態行動対 (s_t, a_t) を時系列情報と見做し，それを RNN（再帰型ニューラルネットワーク）に学習させることにより，環境認識や行動生成を行わせている．しかし，この枠組みでは自らが行動を変化させることによる時系列情報の変化と，環境の変化による時系列情報の変化を区別することが出来ない．これを解決することが，我々が双シェマモデルとして二つの質的に異なる学習モジュールを配置した理由であるが，学習器の構造としては時系列予測の方がモデル学習より多少簡単な為に解析は容易であると考えられる[*3]．

ただ，時系列予測をシェマモデルの基礎ダイナミクスとした場合には，基本的には観測を通じて分節化を行うのみであり，行為者としての自律適応系にとっての情報を分節化するメリットをパフォーマンス面から考えることが出来ない．あくまで現実世界の行為者としての自律適応系内部の表象生成に注目する為，本稿ではモデル学習シェマモデル，強化学習シェマモデルといった行動を含んだシェマモデルのみを扱った．しかし，数理的な視点から言えば，モデル学習は時系列予測を拡張したものであり，シェマモデルの解析の目的からも重要なモデルといえる．ARSM を用いた実際の研究については今後の課題となる．

[*3] 非線形近似器である RNN の解析は複雑であるので，解析を目的にする場合には線形近似器を前提とする AR モデルの方が適当であると考えられる．

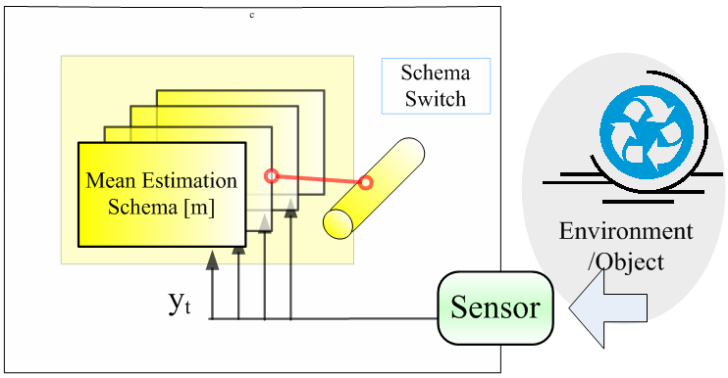


Fig.9.5: MESM: Mean Estimation Schema Model

MESM(**平均推定シェマモデル**)

MESM(Mean Estimation Schema Model) は最も単純なシェマモデルである．推定対象は定点を中心としたベクトルを発振するホワイトノイズ源であり，内部関数はただ，経験として入ってくる値の中心のベクトルを推定する．つまり，対象として以下のものを想定する．

$$y(t) = m + \epsilon(t) \tag{9.3}$$

ここで ϵ はホワイトノイズである．ここではシェマの持つ学習器は定数ベクトルでしかないので，その更新はただ，得られた経験との重み付き平均を取ることによって更新されていく．1 次元では

$$\hat{m}(t+1) = (1-\alpha)\hat{m}(t) + \alpha y(t) \tag{9.4}$$

$$\hat{\sigma}(t+1) = \sqrt{(1-\eta)\hat{\sigma}(t)^2 + \eta e(t)^2} \tag{9.5}$$

で更新される．推定偏差の更新は ARSM と同じである．この後はシェマ活性度の計算，及び選択・生成は 7.3 節の法則に従えばいいことになる．

MESM は ARSM の入力ベクトルを無くした 0 次の ARSM と見做すことが出来る．よって，モデルとしては MESM→ARSM→ モデル学習シェマモデルと，複雑化する構図がある．モデル学習シェマモデルの解析が困難であるならば，より低次で簡易なモデルである MESM から接近を試みるのが妥当であろう．また，上の定義から MESM では解析対象を Fig. 9.6 のようにガウス分布のノイズ源の切り替わり (shift)

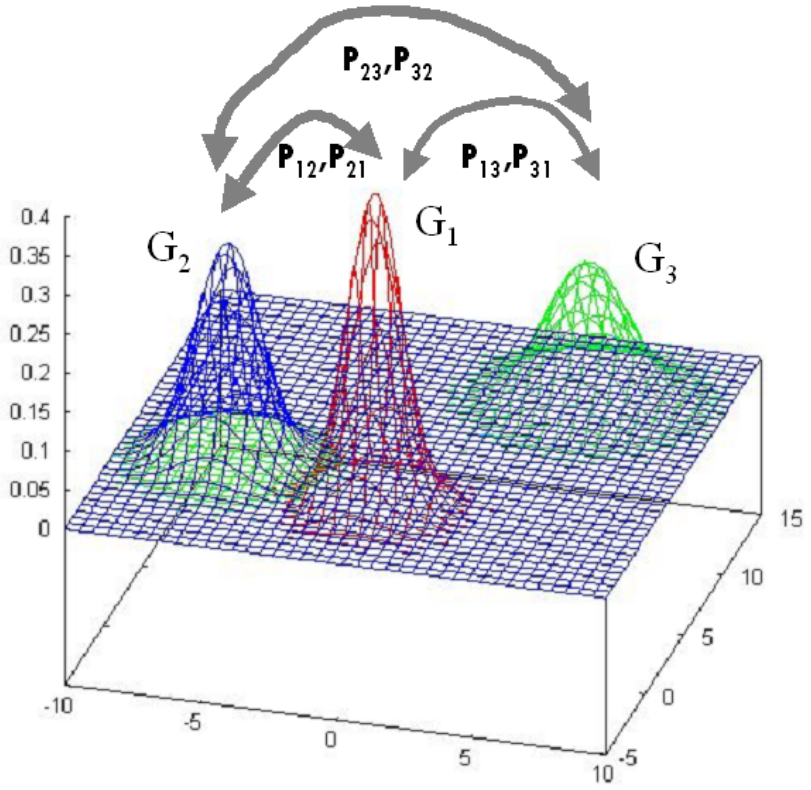


Fig.9.6: Shifting Gaussian distribution as a HMM

が時相関を持つ混合ガウス分布，つまり HMM(隠れマルコフモデル) として扱うことが出来る．HMM については統計的学習理論においても，近年，音声認識課題を中心とした領域で理論の整備が進んでおり，理論的に接地されたシェマモデル解析の足場の形成を期待することが出来る．

9.3.3　シェマモデルの機能と時定数

前節では，シェマモデルの本質を見えやすくする為に，より低次元な二つのシェマモデルを導入した．本節では最も低次元な MESM を例にとることで，本稿で扱ったシェマモデルに共通した性質を考える．

HMM，POMDP 同定問題としてのシェマモデル理解

MESM は対象が，ある定数を中心にガウス分布のホワイトノイズを出す系を対象にする．そして，そのノイズ元の中心が時々変化するという状況を想定する．それは HMM(hidden Markov model) の仮定に他ならない．HMM とは，複数のノイズ源複数存在し，隠れ状態 λ_t により λ 番目のノイズ源が選択され，そこから時々刻々出力が出されるというモデルである．隠れ状態 λ_t はマルコフ過程に従い遷移するが，それ自体は観測出来ないということが，隠れマルコフモデルと呼ばれる所以である．MESM はこのような一つのノイズ源がガウス分布であるような HMM を対象にするが，ARSM やモデル学習シェマモデルの対象は一定の入出力関係の周辺にホワイトノイズを持つノイズ源を隠れ状態の単位として持つ HMM であると考えることが出来る．

7 章でシェマモデルとは原理的には異なるが，同様の対象を学習する学習器として MOSAIC を比較対照としたが，MOSAIC では対象を明示的に HMM と見立てることにより学習則を導くという研究が既に存在する [33]．ただし，シェマモデルが対象とするのは HMM の中でもモデル間の切り替わりの間隔が十分に長く，しかも切り替わるときには瞬時であるようなものである．MESM の場合は非常にイメージしやすい．Fig. 9.6 のように 2 次元空間中にガウス分布が複数存在して，それらが，時々切り替わる，具体的にはガウス核 λ からガウス核 λ' に毎ステップ確率 $P_{\lambda'\lambda}$ で移る状況を考えよう．この時，シェマは一つのノイズ源の中心を同定していく．これは時系列情報の平均値を取ることで簡単に求まる[*4]．と同時に分散推定も行うので，自然とガウスノイズの概形を推定することとなる．その後にノイズ源が別のガウスノイズに切り替わればその主観的誤差は大きくなりる．一回の主観的誤差の増加では変化はしないが，ある一定の時間観察すると，その二乗平均は各次元 1 に収束するはずなので，そこからのずれが有意であると仮説は棄却されて，新たなシェマが作られる．よって，ノイズ間の切り替わりが時間的に緩やかで離散的である場合にはこれらのガウスノイズに相当する平均推定シェマが獲得されることになる．モデル学習シェマは HMM の中でも各状態が点周りのガウス分布ではなく，順モデルとしての関数周りのガウス分布であるものを扱っていたが，この描像から理解するとその本質が見えやす

[*4] 重み付き平均なので不偏推定であることは望めないが

くなる．

また，強化学習シェマモデルでは隠れ状態のある MDP(Markov decision process) を扱っていたと言える．隠れ状態のある MDP については POMDP(Partly Observable Markov Decision Process) という言葉があるが，我々の扱っているのは隠れ状態間の遷移が離散的に，さらに低頻度で起こる POMDP の中の特殊なクラスであると考えられる．

今後はこの視点に立ちシェマ単体ではなく，シェマシステムの安定性や分化の粒度についての解析的な議論を進める必要がある．その為には先に挙げた ARSM や MESM といった低次元のシェマモデルが重要な役割を果たすと考えられる．

短い時定数と長い時定数

本論文では詳細な解析的議論を行わなかった為多くは言及していないが，シェマ切り替えにおける一定時間観察の固執率 p と，学習率 α，η の関係がシェマモデルの作動の安定性や分化を解析する上では本質的となっている．モジュール型学習機構では必ず二つの時定数が存在するといって良い．それは，モジュール切り替えのための時定数と，学習のための時定数である．この二項対立は相互作用の文脈からモジュール数自体を決定していくシェマモデルではより明確になっていく．具体的には固執率 p と，学習率 α，η の対立である．完全なオンライン学習の場合は説明が込み入るので，半分バッチ的な学習を行う場合を考える．説明の題材には第 3 章で取り上げたキュー記憶構造を明示的に持つモデル学習シェマモデルが適切である．上の二つの時定数の対立は，言い換えると

1. 何個のデータサンプルをセットにして同化を判断するか？
2. 一つのシェマの学習にいくつのデータサンプルを用いるか？ (つまりキュー記憶構造のサイズをどれだけに設定するか？)

という対立になる．統計的学習では学習とはデータセットからパラメータ空間 (もしくは，モデル空間) への写像として表現される．つまり，データセットが学習結果そのものを意味しているといっても過言ではない[*5]．そこでは，各シェマが学習に寄与するデータ群をどれだけストックしておくかが本質的となる．第 3 章ではこれを明示

[*5] 最小二乗法などは正にこの写像を構成する．

的にキュー記憶構造を置いて表現したが，LDSM 以降のモデルでは学習の勾配 (α, η など) に間接的に表現されることとなった[*6]．このキュー記憶構造を一杯にするのにかかる時間を学習の時定数 τ_α とする．これに対し，モジュール切り替えにおいても，第 8 章で特に議論したように，ある程度の時間の観測が必要になる．そして，それが余りに長いとモジュール切り替えの時間遅れが激しくなり，また短いと不安定になる．この時定数を τ_p とする[*7]．この二つの関係がシェマモデルでは重要になる..

もし，$\tau_p = \tau_\alpha$ ならば学習が完了するのとモジュール切り替えにかかる時間が同じになり，モジュール切り替えは意味を成さない．モジュール切り替えの時定数は学習の時定数よりも短時間であってこそ意味がある．この視点に立つと，モジュール型学習とは必ずしも学習を効率化させるものではない．統計的学習理論の立場に立つと学習とモジュール選択は

1. 学習は無限に存在するパラメータ群から適切なパラメータを同定すること．
2. モジュール選択は少数に限定されたパラメータ群から適切なパラメータ，つまりモジュールを同定すること．

と出来る．共に正しい同定を行いたければ時定数を長くしないといけない．しかし，時定数を長くすることは実空間での行動を妨げる．つまり，前節で示したような HMM や POMDP のような，その背景に差異を持った構造が隠れており，モジュール，もしくはシェマとして少数の仮説に限定することが出来る場合のみ，モジュール型学習は功を奏する．候補が少数に限定されれば，パラメータ空間は劇的に縮約されるので，選択にかかる時定数は非常に短くで済む．よって，一般的には $\tau_p < \tau_\alpha$ と設定するのが普通である．つまり，あくまでもモジュール型の学習は時間制限の強い系に対する適応方策であり，時間的制限の少ない環境，つまり最適解が一定で，その計算にどれだけ時間をかけてもいいような系では意味を持たない．つまり，第 2 章で挙げたロボットの社会進出への 3 段階 (Fig. 2.12) で考えると，シェマモデルのようなモジュール型学習器は工場環境下のような静的環境下では利用価値がなく，実環境下で意味を持つということになる．

[*6] たとえば MESM では $\alpha = 0.01$ だと約 200 個のデータサンプルが一つのシェマの内部に蓄えられると考えることが出来る．ただし，MESM より複雑なシェマモデルではこのような換算は一般的には難しい．

[*7] τ_p の定義については 7 章参照

一般のモジュール型学習器では選択の問題だけであるので上の議論で済むが，シェマモデルはこの時定数によってシェマの生成さえも支配される．つまり時定数 τ_α によってシェマの推定偏差の学習も進行するために，短い時間スパンで操作する道具の持ち替えや，状況の変化が起きれば分化は起きない．認知実験では比較的この立場が擁護されているように感じられるが [16, 3]，その数理的解析は進んでおらず，どのような条件が実際に必要であるのかはこれからの研究課題であるといえる．また，シェマが形成された後の安定性についても，シェマ単体の学習則ではなく，シェマシステムとしての自己組織化学習が安定的であるのか，どのような外乱が加わればそれが崩壊する可能性があるのかなど検証すべき点は多い．

9.3.4　主要モジュール学習器との比較

本論文を通じてしばしば，他のモジュール型学習器に言及してきた．その中で主なものとして mixture of experts[39, 96, 89] と MOSAIC[103, 34, 25] と本稿を通して提案してきたシェマモデルを比較する．

モジュール型学習器では無いが時系列的な情報を少数のシンボルとして組織化するモデルとして J. Tani らの RNNPB が近年注目され始めているが，明示的にモジュールを配置しない仕組み[*8]のために構造の違いが大きく比較はしない．また，岡田らは予測誤差に基づいて自己組織化マップ上に並べたモジュール群を学習・組織化させるという DBSOM[147] を提案してる．これは離散的なモジュール像と連続的な表現の中間的な表現であり非常に興味深い．

Table 9.1 に比較を示したが，概説を行う．まず，基礎となる理論であるが，シェマモデルが唯一仮説検定を基礎においているのに対して，他の方法はベイズ推定に拠っている．このような，隠れ変数推定として形式化される問題に対してはベイズ推定を適用するのが主流であり，シェマモデルは亜流である．しかし，ベイズ推定が必ずしも優れているわけではないことは第８章の MOSAIC との比較実験でも明らかである．この仮説検定を基礎におくことで，シェマ内部に同化した学習サンプルの一貫性を保つことが出来，そのコヒーレンスを重視したモジュール選択が可能になっている．MOSAIC は予測誤差最小につとめ，mixture of experts は尤度最大化を評価関数に，各モジュールとは独立した時系列としてのゲーティング関数を逐次変化さ

[*8] J. Tani らはこのようなシンボル表現を分散表現と呼んでいる．これに比してモジュール型学習機構が局所表現とよばれる．

Table9.1: Comparison of three modular learning schemes

学習器	シェマモデル	MOSAIC	mixture of experts
基準となる変数	シェマ活性度	責任信号	ゲーティング関数
選択のための理論	仮説検定	ベイズ推定	ベイズ推定
選択基準	コヒーレンス優先	予測誤差最小	尤度最大化
重ねあわせ	なし	あり	あり (ほぼなし)
モジュール追加	○	×	△
滞在時間	長	短	中
応用例	双シェマモデル, RLSM	MPFIM[103], MMRL[25]	mixture of RNN[89], CQ-learning[77]

せることによって重みを決定している．また，各モジュールの出力の取り扱いであるが，シェマモデルでは差異の獲得を前面に押し出しているためもあり本稿の枠組みでは少なくとも出力を重ね合わせることは行っていない．これに対し，他の二つのアプローチは重ね合わせを許すが，我々の試行によると mixture of experts では尤度最大化が通常，一つのモジュールが選択されることによって達成されるので winner take all 的な選択になる．よって，重ね合わされるのはその移行期のみである．これに対して，MOSAIC では頻繁に重ね合わせが生じ，これが MOSAIC をモデルとして不安定にしている要因のように感じられる．また，新たなモジュールの追加では，シェマモデルがその選択則の中に生成則を一体として表現しているのに対して，他の二つは選択肢の和を固定するベイズ推定に基づいているために，その生成則が表現されない．ただし，mixture of experts の研究では，階層的に新たなモジュールを構築していく手法などが提案されており [96]，それらを考慮することによって累増性を表現することが出来る．最後に，滞在時間[*9]であるが，これはあくまで筆者の経験則からの感想である．モデルが潜在的に持っている特性としてこのようなものが考えられる．シェマモデルは先に述べたように分化に十分な推定偏差の収束を求めるために，その構成・生成に十分な長さの連続した経験がいる．しかし，獲得後の切り替え則は

[*9] これは上手く定義された言葉では無いが，その学習手法が環境の離散的変化の間隔をどのくらいの長さであることを前提としているかである．

MOSAIC とほぼ同様に主観的予測誤差の重みつき時間平均によって行うために短くなる．MOSAIC はその時刻その時刻での予測誤差最小で単純にモジュールを切り替えるために前提とする状態遷移の間隔は非常に短く設定できる．これに対し mixture of experts では切り替えのためのゲーティング関数の時系列的変化の立ち上がりが指数関数の立ち上がりのような動きをするために，反応は MOSAIC に比べても遅くなる．

概略的にこのような違いが各モジュール型学習則には存在し，説明する対象や，問題に対してしか優劣はつけられない．本研究では発達における概念分化のようなプロセスを表現しようとしたのでシェマモデルに辿り着いた．問題をオンライン学習による概念の安定的かつ累増的な獲得に置くならば，mixture of experts や MOSAIC に比べてもシェマモデルが有利な点を持っていると考えられる．しかし，その数理的な振舞いは未だ理解しきれておらず，これらの学習則の間の関係性や，特にシェマモデルの動特性についての理解は前節でも述べたように今後の課題である．

9.4 “自律適応系”の構成論的研究

本節では本論文題目にも含まれる自律適応系の構成論的研究という，システム論の研究としての視座から本研究の位置づけと今後の展開について議論する．

9.4.1 システム論研究

“自律適応系”とは作動上の閉じを前提にしつつ，自らの閉じた認識領域に現れる情報のみからナチュラルドリフトを通じて情報を組織化することで適応していくシステムである．第 2 章において自律適応系という言葉を定義したが，そこでは本論文で中心的に扱った認知システムのみならず，生体システム，社会システムといった異なったシステム階層に属する自律系をも議論の場に含むことを明言した．つまり，我々は自律適応系としての一般のシステム理解を模索している．本論文では認知システムの内的表象系の組織化とそれに基づいた C.S. Peirce の意味での記号の獲得，つまりセミオーシスの実現に焦点を絞り議論を行ってきた．システム論の議論に慣れた方には，この議論が自律適応系の研究としては具体的過ぎる議論として映るかもしれない．この本研究のスタンスはオートポイエーシスやオートレファレンス (自己参照) をキーワードとするような，近年のシステム論研究 [115, 116] は抽象度の高さは

持ちつつも，現代思想が陥りがちな “言葉遊び” の様相を呈しつつある現状に対する反省に拠っている点が大きい．言葉の意味とは不定であり，言葉で突き詰めるのみのシステム論の議論には限界がある．システム論にインパクトを与えた N. Wiener や Y. Prigogine，H. Haken らには実際に数理的な立場から，みずからの主張するシステム観を描いて見せたという事実があり，このことが彼らの言説に説得力を与えている重要な要因となっている．

近年，定常解や安定性などの解析を得意とし動的なプロセスを扱う事を苦手とする数学に加え，動的なプロセスを実際に表現できる高速な計算機を得て，複雑系科学が生まれ，構成論的研究が可能になった．システム論の発展の為に現在求められるのは抽象度の高いシステム論の議論というより，むしろ，個々のシステムにおける動態，適応，発達能力を，他のシステムの議論にも転用できるような一般性を保った形式で数理的な立場から捉えることである[*10]．現在，社会システム研究でもエージェントベースで構成論的に社会の動態を理解しようという立場が興ってきている [26, 1]．また，生体システム研究においても構成論的な研究が注目されている [191, 131].

理想的なシステム論は各システムの議論を我々がアナロジーを用いて「人間は細胞の社会」と呼んだりするように，異なるシステムにおける現象の同型性に基づき，あるシステムの現象理解から別のシステムの現象理解を生む枠組みを整備することであると，筆者は考えている．これは形式的な数学において圏論が群，や位相空間といった異なった数学的実体の間に同型関係を見つけることで，その間の関手 (functor) を構成していったことに近い [48].

生体システムや認知システム，社会システムは，それぞれに時定数や研究者がどれだけ実証研究をできるかという条件が異なる．物質科学は人間の生活周期程度で反応を見れる系については多くの知見を得ることに成功したが，同じ実験・結果・考察の枠組みが社会システムや認知発達システムの議論に当てはめられるわけではない．しかし，社会システムの理解は我々の生活に密接に関係し必要不可欠である．実験可能な他のより時定数の小さい系で理解できた事実が，システム間の変換を通してより時定数の大きい社会システムのような対象の理解に役立てば，非常に有益であると考えられる．我々は使えるシステム論を構築することを目的としている．そのシステム論へのアプローチの次のキーワードとして注目されるのがセミオーシスの概念である．

[*10] 一般性を保つとは，そのダイナミクスの本質が “要素” 自身に無いことを指す．“要素” はシステムの階層に固有だが，要素に依存しないダイナミクスは他のシステム理解にも応用できる．

9.4.2 セミオーシスによるシステム論の展開

セミオーシスとは，何度も説明したように C.S. Peirce の用語で記号過程を指す．椹木は複雑適応系を捉える新たな視点としてセミオーシスの重要性を説いている [207, 208]．これは今までの科学が要素が構成する系を外部観察者として捉えてきたことに対し，要素内部に視点を持つことに積極的な意味を見出し，要素内部から見える世界の姿について議論することが重要となる．本稿冒頭に挙げたペットロボットを含めた人間と機械との相互作用の場でも，客観的な意味でのどのような仕組みが対象に存在するかは人間機械の相互作用の要点ではなく，人間がその機械との相互作用の中でどのような意味を見出していくかが重要となる．人間が含まれた系における相互作用では，人間が必ず解釈を行う生き物である為に往々にして記号的である．第 2 章でインタラクションとコミュニケーションの違いについて議論したが，そこで重要だったのは，記号と意味とが物理的因果関係のように一対一対応するようなものはコミュニケーションとは映らず，C.S. Peirce の三項関係のようなサインと解釈項の恣意的な関係付けが基礎ダイナミクスとして存在してこそ我々がまさに “意味” を感じ取るコミュニケーションが成立しえるということであった．しかし，その記号過程のプロセスは科学的，数理的議論の場では捉えられきれていない．故に我々はその記号過程に対する構成論的接近こそ自律適応系の研究のためにまず行うべきことだ考え本研究を行った．

このような記号論によるシステムへの接近は何も人間機械系におけるコミュニケーションの実現だけが目的ではない．生命分野においても現在セミオーシスの議論が行われている．J. Hoffmeyer らは「記号は人間社会ばかりでなく，生命界にも等しく満ち溢れている」[128] として生命記号論を提唱している．この議論は生命システムの要素としての細胞の内部に視点を移動した生命活動の議論をも含む．

本論文を通して “分化” というキーワードが議論の中心的役割を果たし，それが “記号” と結びついていく議論を行ったが，“分化” は生命研究においても，現在盛んである幹細胞研究と結びついて重要なキーワードとなっている．生命研究における分化とは例えば肝細胞から骨細胞や上皮細胞へと細胞自身が変化していくことを指す．細胞自身が役割分化をしていくことで全体としての組織を形成していくのである．

細胞は DNA からタンパクを作り出すメッセージを読み出し，それを生産することで活動する．また，その相互作用はイオンチャンネルやタンパクなどを用いて行われ

るとされるが，このような物質科学的説明で説明が足りるのだろうか？

この疑問が湧く一つの理由に細胞がどうやって自分の現在の位置・役割を把握しているのかという有名な疑問がある．もし，個体発生がチューリング機械がプログラムを読み出すように，DNA のプログラムを中で次々読み出すプロセスのみで構成されているいるとしたら，それは恐ろしいほどのフィードフォワード的な機構である．J. Neuman のセルオートマトン [108] のような形式的な世界での発生ならばそれで済むかもしれないが，実際の外乱環境下ではそのようなフィードフォワード的な機構が上手く駆動するとは，工学的な感覚では考えにくい．フィードバック的な情報の流れがあるはずであり，それ無しには安定した個体発生と生命の維持は考えがたい．そのような調節機構が様々な形で存在することは分子生物学の知見でも分ってきているが，重要なのは細胞自身が変化していくという自律適応系としての姿を持っている点である．

現に，第 2 章でも言及した ES 細胞が力学的環境次第で様々な機能細胞へと分化するという山本らの実験結果はは細胞が力学的環境を読み取ることで自らを変化させていく環境認識能力を持っていることを示している [177, 136]．幹細胞の分化誘導研究において誘導因子を分子レベルで特定する研究が大半を占める中で，これは明らかに要素としての細胞，もしくは集団としての生体組織が周りの状況を読み取り自らを変容させる自律適応性を持っていることを示している．さらに，その分化のプロセスを通じて，細胞にとってのタンパクの記号としての意味自体が，次々と変化してしまうということも注目に値する．それは何を意味するのだろうか？

完全に細胞を分類し，その範疇ごとの相互作用の意味を議論する上では物質科学的な分類学が功を奏するが[*11]，分化現象のような自律適応系のナチュラルドリフトを考える上では相互作用の細胞にとっての意味作用の変化のダイナミクスを考慮する必要がある．つまり，細胞間のコミュニケーションを記号論的に捉えざるを得ないということである．これが生命記号論の要点の一つであると筆者は考えている．

一方で，J. Piaget は初めは生物学者であり，その時に得た「生体の適応性と認知システムの適応性には連続性が存在する.」という信念が有機的なシェマ理論の形成を支えている．では，逆の連続性は無いのだろうか．筆者らは上のような視点から，分子レベルの生命理解から離れて，細胞の分化を細胞の解釈系の変化であるととらえ

[*11] その辞書は恐ろしいほど膨大になってしまう．この点からも従来のゲノム研究は批判されている [190].

る自律適応系の視点から生命現象を理解する道筋はないかと思索している．

通常理解したい対象があるときには，詳細に対象を見ることはしばしば理解を妨げ，時には粗く見ることで適当な議論の粒度を設定しなければならない．これを粗視化と呼ぶが，生体システムにおいても認知システムにおいてもそのシステムとしての理解には分子単位からの疎視化された抽象的な議論が必要になると考えれる．生命現象を力学系としてみる疎視化の方法が提案されているが [42, 191]，生命記号論のように記号現象をベースにして理解する疎視化も候補の一つではないかと考えられる．筆者らは本研究において認知システムにおけるセミオーシスに構成論的に接近したわけだが，これと同型な接近様式が生命におけるセミオーシスにも存在し得ないかと考えているが，未だ模索段階でしかない．

9.4.3 新たな情報観の基礎づけへ

本研究で示した実験を振り返り，一体 “情報” もしくは “意味” とは何なのだろうかと問う．3 章で示した実験では，環境との相互作用を通して顔ロボット内部に複数の知覚シェマが形成された．そして，その知覚シェマに対応する運動状態を提示すると，そのときに限って対象球の追跡方法を想起し追跡が可能となった．顔ロボットはそれらの運動球についてのシェマ形成の後にはにある運動球を見せるとちゃんとその運動球シェマが活性化したのだ．ここでは経験が新たな経験の意味を見出している．また，セミオーシスを構成した 8 章の実験では，全く作動上閉じた自己組織化学習からエージェントはラッパの音を意味づけ，自らの行動を切り替える術を覚えた．この時，ラッパの音の意味とは何処にあったのだろうか？それはラッパを鳴らす人間が設計したのかもしれないが，際立った因果関係を持つ差異として自律適応系が活動する環境にデザインしたに過ぎない．例えば，ラッパの音ではなく，チーターの足音を逃げる行動に結びつける動物がいたとしてもよい．その場合チーターは足音を記号としてわざわざデザインしているわけではなく，その動物が自らその足音を，そこから逃げる活動的解釈項に自律的結びつけたのだ．エージェントは報酬を含んだ環境との相互作用を通して自らの身体と経験に基づいて差異を発見し，シェマを分化させ，その切り替えに時相関を持つサインを発見しそれを自ら仮説推量的に利用した．この全てが，エージェント自身の主体的な活動として計算論的に表現できることが明らかになった．数理的に見ても “意味” が自己の中に完全に閉じて形成しうることを見た．

20 世紀末にインターネットを含めた情報技術の急速な発展があり，現在も情報産業は日進月歩に発展している．その中で，様々な領域で情報の “意味” が問われ直している．それ以前は過去には情報は足りないことの方が多く，価値のあることが前提のようであった．現在は情報は基本的に氾濫しており，特にインターネットの情報は何が正しく何が正しく無いか分からず，情報の量は多くても，“意味が無い” とよく言われる．情報量という言葉はは Shannon-Weaver の情報理論に基づく情報観に基づいて定義される量であるが，そこには当然意味論は含まれない．21 世紀は生命と情報の世紀だといわれたりするが，我々に意味を感じさせる情報とは何なのか．その本性は良く分からない．それが良く分からないままに，情報工学ばかりが発展し「意味」の意味を置き去りにしたままで情報という言葉が蔓延している現状がある．このような背景から西垣は基礎情報学を唱え，“意味” を抜いた Shannon-Weaver 流の情報を超えた意味論的な基礎づけを求めている [178]．第 2 章でも論じたように西垣の立場は我々の意味への接近に最も近い立場である．本論文を通して，我々は解釈者にとっての最小限のセミオーシスの構成論を提示した．そこでは主体が自らの身体に基づき環境を分節化し，サインをその分節に基づいて意味づけている．西垣も意味の根源を記号論の中に求めたが，我々はさらにその記号論を計算論に結び付けていくことにより，情報の意味解釈のプロセスを説明する最小限の構成論的表現を得ることが出来た．これを手がかりにより確かな意味作用への接近を試みて行きたい．

情報とは既存のシェマに基づいてしか意味を持つことは出来ず，常に漸増的である．近年，情報公開が叫ばれ，専門的な情報が官公庁のホームページ等で公開されることが多いが，それを理解できる，つまりそこから意味を得ることの出来る一般人は一握りであり．にもかかわらず，情報開示社会には情報の意味作用の仕組みから原理的な限界があることを認識する人は少ない．なぜならば，意味作用が未だ理解されていないからである．

また，近年,「ゆとり」教育と「つめこみ」教育の二項対立が良く議論される．一時の「ゆとり」教育のながれから，最近は「つめこみ」教育への揺り戻しが盛んである [122]．しかし，自律適応系の議論からすれば，この二項対立には核心が欠落している．まず,「つめこみ」教育は情報を生徒に注入できるという図式の上に成り立っており，学習者自身による意味作用を無視しているように見える．実際に教育現場において体験を含まない教科書を通してのみの一方向的な伝達では，教科書が意味する本当の意味を伝えるのは不可能であり,「つめこみ」教育ではこの面が頻繁に叩かれてきた．しかし,「つめこみ」には別の意味合いもある．人間は既存のシェマ，つまり

認知構造に基づいてしか対象を認識できず，発達は常に漸増的に起こる．よって，早い段階で「つめこむ」ことはその後の学習を支えることにもなり非常に重要である．問題は基本的な認知構造を「つめこむ」のに，いわゆる「つめこみ」教育でよいのかということである．「つめこみ」教育は自律系内部での意味作用を無視した情報観であり，「つめこみ」の必要性を論じるのは更なる学習，認識の基盤としての自律系内部での認知構造の先行的な獲得を重要視する作動上の閉じを前提とした自律適応系における意味作用を認識した情報観である．つまり，「ゆとり」との二項対立が立つ前に，「つめこみ」の内部に二項対立が立ってしまっているのである．このような問題を議論する上でも，自律適応系における意味作用の理解が重要である．

このように，システム論をとりまく状況の中で各システムの階層においてもセミオーシス，つまり作動上閉じた自律適応系を前提とした意味の研究が重要な意味を持ち得る．しかし，その問題に接近しようとする動きはまだ少ない．実証的な研究としては捉えにくい問題であるが故に，科学的にストイックになれば，如何なる接近をも不可能に思われる．これに対し，複雑系科学において始まった構成論的研究は，科学哲学的にはどのように位置づけられるのかについては議論が尽きないが，既存の科学的なストイックな方法に無い説明能力と，物質科学的な接地点を持たない不確かさの両面を持っているように思われる．科学とは本来，人間の知りたい心を満たす為の営みであり，厳格な科学的方法が通じないならば，その制約を緩くしてでも接近するしかない．その緩め方の一つが構成論的研究であると筆者は考える．そのような視点から本研究では自律適応系のコミュニケーションの根幹を担うと考えられる“意味”の問題に対して，構成論的接近を試みた．

第 10 章

結言

本論文を結ぶにあたっては，初めに示した問いに回帰するのが筋だろう．

「自律適応的に言葉を獲得し，人間と意思疎通出来るようになる人工物を作ることは可能だろうか？」

始めに問いを立てた以上，筆者としては如何に難しい問いであろうと，これに答えなければならない．答えは

「将来的に可能であると考えるに足る根拠の一つを得た．」

である．まるで政治家のような回りくどい表現だが，実際にそのような人工物を作ったわけでないので断言するわけには行かない．しかし，第 2 章で我々が最もこの問題の核心部分だと考えた，環境との相互作用を通して差異体系を獲得し，それを元に C.S. Peirce の三項関係 (semiotic triad) を実現するという部分については構成論を組むことが出来た．しかも，その構成論を支える一段一段のブロックは，時間的制約のある動的環境において従来の学習理論の性能を向上させることが示された．この点は，その存在意義を信じる上で重要である．ここで，本稿での議論を振り返りたい．

第 2 章では，人間環境下で活動する社会ロボットの設計のためには，直接的機能設計は問題があり，環境との中から機能を構成していく自律適応系設計を考えるべきであると指摘した．また，人間とのコミュニケーションを考える上では，この適応性の保持がより重要であることを議論した．自律適応系設計を掲げた時に，「自律適応系とは何か？」という疑問が生まれる．これに対して一定の概念与えるために

F. Varela の作動上の閉じという概念を引用し．自律適応系の最前提とした．その後に，人間と相互作用するために必須な，ロボット自身による記号の意味理解を議論する為に，コミュニケーションとインタラクションの差別化をはかり，コミュニケーションの特徴づけを行うために C.S. Peirce の記号論を導入した．しかし，C.S. Peirce の記号論はセミオーシスと呼ばれる解釈のダイナミクスについて，その基盤を明らかに出来ていなかった．よって我々はこの基盤の形成を問題とし，J. Piaget のシェマ理論にその示唆を求めた．第 3 章以降では順次その構成論を論じていった．

第 3 章では自己組織化型学習手法として双シェマモデルを提案し，その内部にもたれる知覚シェマについて，その同化・調節，均衡化・分化のダイナミクスを計算論的に提案した．そして，これを実際に顔ロボットと名づけられた自律ロボットに実装し自律系内部においてシェマを形成する実験を行った．顔ロボットは提示される運動球の運動状態を知覚シェマという形で内化し，運動球のダイナミクス変化という形であらわれる環境の変化に呼応する形でそれらのシェマを想起することが出来るようになった．また，ボトムアップに組織化されたこれらの内部表象にはファジィ集合としての包含関係が存在することが示された．

第 4 章ではこのような自律適応系内部での表象生成が，あくまでその自律ロボット自身のセンサ・モータ系に閉じた領域，つまり現象学的な領域での出来事であり，客観的な表象の境界は存在し得ないことを確認する議論を行った．筆者はアフォーダンスや生物記号論の示唆の元，身体パラメータと内的表象の生成には非線形な関係があるという仮説を持っていた．これを構成論的に検証するために，第 3 章で提案した双シェマモデルの改訂版である LDSM をシミュレーション空間内の顔ロボットに実装し実験を行った．具体的には，それぞれのロボットに様々に時間解像度やロボットの視野といった身体パラメータを振り同一環境下で活動させた．この実験を通して，双シェマモデルがこの非線形仮説を満たす構成論となっていることが明らかになった．

第 5 章では行為概念の獲得に焦点を当てた．環境や対象の概念が名詞的なものであるとするならば，これは動詞的な意味を持つ内的表象の生成に焦点を当てたものである．特に自律的な学習を実現する為，強化学習を導入して議論を行った．まず，人間が獲得しうる行為とは実際のモータ出力などに縛られない一種汎化されたものであると考え，環境ダイナミクスの変化に依存しない汎化行為概念を定義した．その上で，双シェマモデル内での行為シェマは汎化行為概念となっていることを示し，それを強化学習で獲得する双シェマモデル強化学習の枠組みを提案した．実験としては運動球の制御問題をとりあげ，台車が動くことによる慣性力変化などの環境変化に対し，汎

化された行為概念を知覚シェマを使って具体化することで環境適応可能であることを示した.

第 6 章では強化学習シェマモデル (RLSM: Reinforcement Learning Schemata Model) を導入し，三章で導入した双シェマモデルにおける環境概念の分化および累増的獲得と同様に行為概念の累増的獲得が可能なことを示した．また，この手法は第５章で導入した双シェマモデル強化学習にも適用可能であり，これによって双シェマモデルは適応的に知覚シェマ及び行為シェマを累増的に獲得する事が出来るようになった．また，知覚シェマと行為シェマは双シェマとして任意に接続可能であり，それぞれの接続が様々な環境における行動の方策として適切な意味を持つために原初的な文法のようなルールが観察されることが示された.

第 Ⅳ 部ではこのような記号生成のプロセスが実際の人間の脳におけるどの様な可塑性と同型的であるかについて計算論的脳科学の立場との相関から考えた．構成論的な枠組みでは実際に記号生成にどの部位が関係しているというような，科学的な結論を導出することは出来ないが，計算論的に等しい構造を見つけ出すことが出来れば，我々の定式化する記号現象と実際の脳内現象との関連についての示唆を得ることが出来る．第 7 章では小脳のモデルとして提案されている MOSAIC に対して，それと交換可能と考えられる計算論的なモデルとして NGSM(正規化ガウスネットワークを用いたシェマモデル) を提案した．これは第４章において提案した LDSM を非線形な系への拡張となっている．実験を通して，MOSAIC が元々持っていた不安定性や累増性についての制限を NGSM が取り払うことが示された．また，同時にシェマモデルの定式化を厳密化することで統計学における仮説検定理論との接点を明確にした.

第 8 章において，近年，大脳新皮質や海馬で発見されたスパイクタイミング依存のシナプス可塑性 (STDP: Spike Timing-Dependent Plasticity) とシェマモデルを結合することにより，自律系内部に獲得された表象と外的な聴覚・視覚情報（遠隔情報）を連合させ適応的に行動できるようになることを示した．STDP 則は予測誤差の将来的な推定積分値を自己組織化的にシナプス結合にコーディングすることが出来ることを示し，これが，通常必ず時間遅れを伴うシェマ選択を加速させうることを示した．これにより，第 Ⅱ 部，第 Ⅲ 部を通して内的に獲得できるようになった内的表象が外界の信号と連合することにより，サインからの想起が可能となった．これは Peirce の三項関係 (semiotic triad) の構成論に他ならず，自律適応系のナチュラルドリフトとしての学習から，記号の三項関係が立ち上がり得る事を構成論的に示すことが出来た．このときもやはりサインの利用により自律ロボットのパフォーマンス自体

が時間的制約の強い動的環境下では向上することが確かめられた．

第 9 章ではこれらをまとめ，自律適応系におけるセミオーシスについて論じた後に，我々が当研究で前提として埋め込んでいたものを反省的に明らかにした．また，本稿の中心部では一貫してシェマモデルを提案してきたが，シェマモデルは特殊なタイプの隠れ状態推定の方法であるという視点を提供した．さらに，他の有名なモジュール型学習と対比することによりシェマモデルの持つ統計的学習法としての性格を考察した．その後に，より領域横断的なシステム論の立場から，これからのシステム論における記号論の役割と今後の展望を述べ，若干の議論を行った．

本研究では，社会ロボットの設計論という工学的問題に端を発しながらも，その問題が原理的に含んでいる．コミュニケーションにおける意味作用と，そのコミュニケーションを行う主体が作動上閉じているという本質的問題に敢えて踏み込んだ．これに踏み込みうるのは，複雑系科学が開拓した構成論的研究手法が存在するからである．我々は，厳密な実験データによらず，哲学，発達心理学，記号論といった数値化されない傍証を集めることにより，ラディカルに数理モデルを立ち上げ，実際に計算機，ロボットを用いてダイナミクスを走らせることにより，その基礎ダイナミクスを考えるという構成論的アプローチを採った．これは，まさに記号現象，そしてそれを問題にする記号論への構成論的アプローチであり，**構成論的記号論** (constructive semiotics) と呼ぶことが出来る．

20 世紀の物質科学はマクロレベルでは宇宙の最果てまで明らかにし，ミクロ，ナノレベルではクオークのレベルまで明らかにしてきた．両端はすばらしいが，最も身近な真ん中のレベルはどうだろう？私たちは私たち自身のことをどれだけ理解できているのだろうか？物質科学の発展に比べれば，我々自身の心や知性やその発達という，誰もが身近に触れられる領域の問題については，まだまだ分からない事だらけだと言わざるを得ないだろう．

物質的豊かさと精神的豊かさという言葉が二項対立のように言われるが，20 世紀を通して物質的に豊かになって来た事は誰しもが認めるにちがいない．そしてそれは明らかに物質を扱う科学，もう少しゆるめて，物質を知ることに拠ってきた．ならば逆に，精神的豊かさを得るために我々に出来るのは精神を知ることしか無いはずだと，筆者は考える．

これに対し，S. Freud に始まる精神分析や，J. Piaget の流れを汲む構成主義的な

発達心理学を除けば，20 世紀の心理学，心の科学は人の心を外部から物質的，静的に物質科学の対象として扱う行動主義的な見方に支配されてきた感がある．これは，もともと科学の根幹を支える数理的な枠組みが発達や適応といったダイナミクスを捉えることを不得手としていたという理由に負う面も確かにある．しかし，現在では，計算機というダイナミクスの表現を得意とする研究の道具が現れたために，複雑系科学に代表される，複雑であったり動的であったりする対象への新たなアプローチが可能になってきた．我々の研究は正に，このような背景に裏打ちされた人間を始めとした自律的，適応的要素への接近に他ならない．

マスコミでヒューマンエラーが叩かれ，ちょっとした予測や情報を提示するにも科学的根拠が問われる科学原理主義的な時代であるが，人間要素を始めとした自律適応系要素が未だに解明できない以上，世の中は静的な物質科学的な情報のみを頼みに絶え間ない意思決定を行っているように見える．そのアドホックな意思決定が，更なる人間要素の排斥を生む社会を構成し始めている気がしてならない．この居心地悪い世界を形成する流れに逆らうためには，むしろ，人間の多様性や適応性を陽に認めたシステム理解と人間の個別性や意味作用のレベルを含めた情報の基礎づけこそ，人工物と人間が共存する住み良い未来社会形成の為に現代のシステム科学研究者が負った，最も重要な仕事ではないかと，筆者は考えている．

より人間の個性（多様性）や有機的側面を生かした住み良い社会を作り出すためにも，上のような時代の流れをひっくり返すような説得力のある共通理解の形成を欠くことは出来ない．筆者は本論文による個々人の主観性と適応性を陽に認めた「意味」に対する工学的，構成論的アプローチが，そのような多様性と適応性をプラスと捉えるようなシステム理解に対する “小さくても確かな一歩” となっていることを信じている．

ペットロボット一匹の話から，思いもかけぬほど大きな話に広がったように思われるかもしれない．いや，実際に広がったわけだが，本論文を終えた後も更に前進する必要があるので，それを再び小さな話に戻すこと無しに，広げたままで，本論文を閉じる事にする．

著者関連文献

学術論文

1. 谷口 忠大, 椹木 哲夫, “双シェマモデル”, 人工知能学会論文誌, Vol. 19(6), pp.493–501, (2004)
2. 谷口 忠大, 椹木 哲夫, “身体と環境の相互作用を通じた記号創発：表象生成の身体依存性についての構成論”, システム・制御・情報学会誌,18(12), pp.440–449, (2005)
3. 谷口 忠大, 椹木 哲夫,“汎化行為概念の適応的獲得 —双シェマモデルベースの強化学習—”, 計測自動制御学会論文集 2006 年 3 月掲載予定, (2006)
4. 谷口 忠大, 椹木 哲夫,“報酬設計を通した社会的相互作用による行為概念群の構築：シェマ理論に基づいた累増的強化学習”, 知能情報ファジィ学会誌, (submitted)
5. T. Taniguchi, T. Sawaragi, “Incrimental acquisition of multiple nonlinear forward models based on differentiation process of schema model”, Neural Networks, (submitted)
6. 谷口 忠大, 椹木 哲夫,“強化学習シェマモデルと STDP シナプス可塑性に基づいた記号創発”, Cognitive Studies, (submitted)

国際会議

1. T. Taniguchi, T. Sawaragi, “Assimilation and Accommodation for Self-organizational Learning of Autonomous Robots: Proposal of Dual-Schemata Model”, IEEE International symposium on CIRA 2003, pp. 277–282, (2003)
2. T. Taniguchi, T. Sawaragi, “An Approach of Self- Organizational Learn-

ing System of Autonomous Robots by Grounding Symbols through Interaction with Their Environment”, SICE Annual conference 2003, pp.2259–2264, (2003)
3. T. Taniguchi, T. Sawaragi, “Design and Performance of Symbols Self-Organized within an Autonomous Agent Interacting with Varied Environments”, IEEE International Workshop on RO-MAN CD-ROM, (2004)
4. T. Taniguchi, T. Sawaragi, “Self-Organization of Inner Symbols for Chase: Symbol Organization and Embodiment”, IEEE International Conference on SMC 2004 CD-ROM, (2004)
5. T. Taniguchi, T. Sawaragi, “Adaptive Organization of Generalized Behavioral Concepts for Autonomous Robots: Schema-Based Modular Reinforcement Learning”, IEEE International symposium on CIRA 2005, in cdrom, (2005)
6. T. Taniguchi, T. Sawaragi, “Incremental acquisition of behavioral concepts through social interactions with a caregiver”, Artificial Life and Robotics (AROB 11th ’06), (2006)

国内会議

1. 谷口 忠大, 椹木 哲夫, “人間とロボットのコミュニケーションにおける 記号行為と関係性構築のダイナミクス”,Human Interface Smposium 2002,pp.785–788, (2002)
2. 谷口 忠大, 椹木 哲夫, “記憶シェマの動的構築による運動観察からの ファジィ概念獲得”,Fuzzy System Symposium 2003,pp. 611–614, (2002)
3. 谷口 忠大, 岡田 敬信, 椹木 哲夫, “協同作業場における言語創発のダイナミクス”,Human Interface Symposium 2003,pp. 749–752, (2003)
4. 谷口 忠大, 椹木 哲夫, “相互作用からの意味生成： 自閉的カテゴリー生成と行為分化の共時性”, SICE 自律分散システム講演会 2004, (2004)
5. 谷口 忠大, 椹木 哲夫, “自律ロボットの動的環境との相互作用を通じた主観的記号生成”, 日本人工知能学会 2004 年全国大会 CD-ROM, (2004)
6. 谷口 忠大, 椹木 哲夫, “身体的相互作用知に基づいた記号的相互作用についての考察”, 人工知能学会若手の集い MYCOM2004,Web Proceedings, (2004)

7. 谷口 忠大, 椹木 哲夫, “顔ロボットの移動物追跡のための運動記憶の動的構成による環境適応”, 計測自動制御学会システム・情報部門講演会 (SSI2004), pp.185–190, (2004)
8. 林 寿和, 椹木 哲夫, 谷口 忠大, “需要と供給の共創場ダイナミクスに基づく消費購買行動の構成論的解析”, 第 32 回知能システムシンポジウム資料, pp.127–132, (2005)
9. 小川 賢治, 谷口 忠大, 椹木 哲夫, “複数の Q-table を用いた多様環境下での状況弁別型強化学習”, 第 21 回ファジィシステムシンポジウム講演論文集,pp.724–727, (2005)
10. 谷口 忠大, 椹木 哲夫,“シェマ分化に基づいた非線形予測モジュール群の累増的獲得手法”, 日本神経回路学会第 15 回全国大会 (JNNS2005), pp.194–195, (2005)
11. 谷口 忠大, 椹木 哲夫,“報酬設計による社会的相互作用に基づいた多様行為概念の自律的獲得”, 第 15 回インテリジェント・システム・シンポジウム (FAN2005), pp.365–370, (2005)

参考文献

[1] 21 世紀 coe プログラム「エージェントベース社会システム科学の創出」. http://www.dis.titech.ac.jp/coe/.

[2] Aibo official site. http://www.jp.aibo.com/.

[3] F.A. Mussa-Ivaldi A, Karniel. Does the motor control system use multiple models and context switching to cope with a variable environment? *Exp Brain Res*, Vol. 143, pp. 520–524, 2002.

[4] H.D.I. Abarbanel, R. Huerta, and M.I. Rabinovich. Dynamical model of long-term synaptic plasticity. *PNAS*, Vol. 99(15), pp. 10132–10137, 2002.

[5] L.F. Abbott and S.B. Nelson. Synaptic plasticity: taming the beast. *nature neuroscience supplement*, Vol. 3, pp. 1178–1182, 2000.

[6] C. Angelo. Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on evolutionary computation*, Vol. 5(2), pp. 93–101, 2001.

[7] R.C. Arkin. Motor schema - based mobile robot navigation. *The International Journal of Robotics Research*, Vol. 8(4), pp. 92–112, 1989.

[8] C. Atkeson, A. Moore, and S. Schaal. Locally weighted learning. *Artificial Intelligence Review*, Vol. 11, pp. 11–73, 1997.

[9] L. Baird. Residual algorithm: Reinforcement learning with funtion approximation. In *Armand Prieditis & Stuart Russell, eds. Machine Learning: Proceedings of the Twelfth International Conference*, 1995.

[10] L.C. Baird. Reinforcement learning in continuous time: Advantage updating. In *Proceedings of the International Conference on Neural Networks*, 1994.

[11] A.G. Barto, S. Singh, and N. Chentanez. Intrinsically motivated learning of hierarchical collections of skills. In *In International Conference on Development and Learning (ICDL 2004)*, 2004.

[12] G. Bateson, (佐藤良明 訳). 精神の生態学 改訂第 2 版. 新思索社, 2000.

[13] G.Q. Bi and M.M. Poo. Synaptic modification in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, Vol. 18(24), pp. 10464–10472, 1998.

[14] O. Bock, S. Schneider, and J. Bloomberg. Conditions for interference versus facilitation during sequential sensorimotor adaptation. *Exp Brain Res*, Vol. 138, pp. 359–365, 2001.

[15] V. Braitenberg, (加地大介 訳). 模型は心を持ちうるか. 哲学書房, 1987.

[16] T. Brashers-Krug, et al. Consolidation in human motor memory. *nature*, Vol. 3382, pp. 252–255, 1996.

[17] C L Breazeal. *Designing Sociable Robots (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2004.

[18] G. A. Carpenter and S. Grossberg. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, Vol. 37, pp. 54–115, 1987.

[19] P. Dayan and G.E. Hinton. Feudal reinforcement learning. In *Advances in Neural Information Processing Systems 5*, 1993.

[20] T.W. Deacon. *The Symbolic Specoes: The co-evolution of language and the brain*. W W Norton & Co Inc (金子隆芳　訳「ヒトはいかにして人となったか」, 新曜社), 1997.

[21] R. Dearden, et al. Bayesian q-learning. In *AAAI-98*, pp. 761–768, 1998.

[22] K. Doya. What are the computations of the cereblllum, the basal ganglia, and the cerebral cortex? *Neural Networks*, Vol. 12, pp. 961–974, 1999.

[23] K. Doya. Reinforcement learning in continous time and space. *Neural Computation*, Vol. 12(1), pp. 219–245, 2000.

[24] K. Doya. Metalearning and neuromodulation. *Neural Networks*, Vol. 15(4), pp. 495–506, 2002.

[25] K. Doya, et al. Multiple model-based reinforcement learning. *Neural Computation*, Vol. 14, pp. 1347–1369, 2000.

[26] J.M. Epstein and R. Axtell. Growing artificial societies –Social science from the bottom up–. the Brookings institution, 1996. (服部, 木村 訳,「人口社会–複雑系とマルチエージェント・シミュレーション–」, 共立出版, 1999).

[27] L. Fadiga, L. Fogassi, G. Pavesi, and G. Rizzolatti. Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neuropysiology*, Vol. 73(6), pp. 2608–2611, 1995.

[28] C.L. Fancourt and J.C. Principe. Modeling time dependencies in the mixture of experts. In *Proceedings of IJCNN'98, 1, pp. 2324-2327*, 1998.

[29] F. Gandolfo, et al. Motor learning by field approximation. *Proc. Natl. Acad. Sci.*, Vol. 93, pp. 3834–3846, 1996.

[30] J.J. Gibson. *The Ecological Approach to Visual Perception,*, Vol. Company. Houghton Mifflin Company (古崎, 辻, 村瀬訳: 「ギブソン 生態学的視覚論」サイエンス社 (1985)), 1979.

[31] V. Gullapalli. A stochastic reinforcement learning algorithm for learning real-valued functions. *Neural Networks*, Vol. 3, pp. 671–692, 1990.

[32] S. Harnad. The symbol grounding problem. *Physica D*, Vol. 42, pp. 35–346, 1990.

[33] M. Haruno, D.M.Wolpert, and M. Kawato. Mosaic model for sensorimotor learning and control. *Neural Computation*, Vol. 13, pp. 2201–2220, 2001.

[34] M. Haruno, D.M. Wolpert, and M. Kawato. Hierarchical mosaic for movement generation. *International Congress Series*, Vol. 1250, pp. 575–590, 2003.

[35] Y. Horiguchi. *Design of Co-Adaptive Interface System for Supporting Joint Task by Human and Machine Autonomies.* PhD thesis, Kyoto University, 2005.

[36] H. Imamizu, et al. Human cerebellar activity reflecting an acquired internal model of a new tool. *nature*, Vol. 403, pp. 192–195, 2000.

[37] H. Imamizu, et al. Modular organization of internal models of tools in the human cerebellum. *Proc Natl Acad Sci USA*, Vol. 100, pp. 5461–5466, 2003.

[38] H. Imamizu, et al. Functional magnetic resonance imaging examination of two modular architectures for switching multiple internal models. *The Journal of Neuroscience*, Vol. 24(5), pp. 1173–1181, 2004.

[39] R. A. Jacobs, M. I. Jordan, et al. Adaptive mixtures of local experts. *Neural Computation*, Vol. 3(1), pp. 79–87, 1991.

[40] J.H.Flavell. *The Developmental Psychology of Jean Piaget*. Van Nostrand Reinhold (岸本 訳:ピアジェ心理学入門, 明治図書 (1969)), 1963.

[41] M.I. Jordan and D.E. Rumelhart. Forward models: Supervised learning with a distal teacher. *Cognitive Science*, Vol. 16, pp. 307–354, 1992.

[42] S. Kauffman. 自己組織化と進化の論理 宇宙を貫く複雑系の法測. 日本経済新聞社, 1999.

[43] M. Kawato. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, Vol. 9, pp. 718–727, 1999.

[44] M.P. Kilgard and M.M. Merzenich. Cortical map reorganization enabled by nucleus basalis activity. *science*, Vol. 279, pp. 1714–1718, 1998.

[45] H. Kimura and S. Kobayashi. An analysis of actor/critic algorithms using eligibility traces: Reinforcemen learning with imperfect value function. 15th International Conference on Machine Learning pp.278–286, 1998.

[46] V.R. Konda and J.H. Tsitsiklis. Actor-critic algorithms. In *Advances in Neural Information Processing Systems 12, pp. 1008-1014.*, 1999.

[47] N. Kubota, et al. Fuzzy and neural computing for communication of a partner robot. *Journal of Multi-Valued & Soft Computing*, Vol. 9, pp. 221–239, 2003.

[48] F.W. Lawvere and S.H. Schanuel. *Conceptual Mathematics*. Cambridge University Press, 1997.

[49] H. Markram, et al. Regulation of synaptic efficacy by coincidence of post-synaptic aps and epsps. *SCIENCE*, Vol. 275, pp. 213–215, 1997.

[50] H. Maturana, F. Varela, (管 啓次郎 訳). 知恵の樹. ちくま学芸文庫, 1997.

[51] H.R. Maturana, F.J.Varela, (河本英夫 訳). オートポイエーシス 生命システムとは何か. 国文社, 1991.

[52] M.C.W. van Rossum, G.Q. Bi, and G.G. Turrigiano. Stable hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience*, Vol. 20(23), pp. 8812–8821, 2000.

[53] T. Mirata, et al. Discrete stochastic process underlying perceptual rivalry. *NeuroReport*, Vol. 14(10), pp. 1347–1352, 2003.

[54] J. Moody and C. J. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation*, Vol. 1, pp. 281–294, 1989.

[55] J. Morimoto and K. Doya. Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. *Robotics and Autonoumous Systems*, Vol. 36, pp. 37–51, 2001.

[56] U. Neisser, (古崎敬，村瀬旻　共訳). 認知の構図　人間は現実をどのようにとらえるか. サイエンス社, 1978.

[57] B. Oksendal, (谷口説男 訳). 確率微分方程式　入門から応用まで. シュプリンガーフェアラーク東京, 1999.

[58] L. Olsson, C. L. Nehaniv, and D. Polani. Sensor adaptation and development in robots by entropy maximization of sensory data. In *proceedings of CIRA2005*, 2005.

[59] R. Osu, et al. Random presentation enables subjects to adapt to two opposing forces on the hand. *Nature Neuroscience*, Vol. 7(2), pp. 111–112, 2004.

[60] C. Panchev and S. Wermter. Spike-timing-dependent synaptic plasticity: from single spikes to spike trains. *Neurocomputing*, Vol. 58-60, pp. 365–371, 2004.

[61] C.S. Peirce, (内田種臣 (編訳)). パース著作集２　記号学. 勁草書房, 1986.

[62] C.S. Peirce, (米盛裕二 編訳). パース著作集１　現象学. 勁草書房, 1985.

[63] R. Pfeifer and C. Scheier. *Understanding Intelligence*. The MIT Press (石黒章夫，細田耕，小林宏訳:「知の創成」，共立出版 (2001)), 2001.

[64] J. Piaget, R. Garcia, (芳賀純，能田伸彦 翻訳). 意味の論理 意味の論理学の構築について. サンワコーポレーション, 1998.

[65] J. Piaget, (滝沢武久，佐々木明 訳). 構造主義. 白水社, 1970.

[66] D. Precup, R. Sutton, and S. Dasgupta. Off-policy temporal-difference learning with function approximation. In *Proceedings of the Eighteenth Conference on Machine Learning*, 2001.

[67] R.A.Schmidt. Motor and action perspectives on motor behaviour. In *Meijer, OG and Roth, K (Eds), Complex Movement Behaviour, Elsevier Sci. Pub.*, pp. 3–44, 1988.

[68] S. Russell, P. Norvig (古川康一監訳). エージェントアプローチ人工知能. 共立

出版, 1997.

[69] Y. Sakai, K. Nakano, and S. Yoshizawa. Synaptic regulation on various stdp rules. *Neurocomputing*, Vol. 58–60, pp. 351–357, 2004.

[70] K. Samejima and T. Omori. Adaptive internal state space construction method for reinforcement learning of a real-world agent. *Neural Networks*, Vol. 12, pp. 1143–1155, 1999.

[71] M. Sato and S. Ishii. On-line em algorithm for normalized gaussian network. *Neural Computation*, Vol. 12, pp. 407–432, 2000.

[72] S. Schaal and C.G. Atkeson. Constructive incremental learning from only local information. *Neural Computation*, Vol. 10(8), pp. 2047–2084, 1997.

[73] S. Schaal, A. Ijspeert, and A. Billard. Computational approaches to motor learning by imitation. *Philosophical Transaction of the Royal Society of London: Series B, Biological Sciences*, Vol. 358, pp. 537–547, 2003.

[74] R.A. Schmidt. A schema theory of discrete motor skill learning. *Psychological Review*, Vol. 82(4), pp. 225–260, 1975.

[75] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal,*, Vol. 27, pp. 379–423, and 623–656, 1948.

[76] S.P. Singh. The efficient learning of multiple task sequences. *Neural Information Processing System*, Vol. 4, pp. 251–258, 1992.

[77] S.P. Singh. Transfer of learning by composing solutions of elemental sequential tasks. *Machine Learning archive*, Vol. 8(3-4), pp. 323–339, 1992.

[78] S. Song and L.F. Abbott. Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, Vol. 32, pp. 339–350, 2001.

[79] S. Song, et al. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *nature neuroscience*, Vol. 3, pp. 919–926, 2000.

[80] L. Steals. Evolving grounded communication for robots. *Trends in Cognitive Science*, Vol. 7(7), pp. 308–312, 2003.

[81] Y. Sugita and J. Tani. Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adaptive Behavior*, Vol. 13(1), pp. 33–52, 2005.

[82] K. Sugiura, M. Akahane, T. Shiose, and O. Katai. Exploiting interaction

between sensory morphology and learning. In *Proceedings of the International Conference on Systems, Man and Cybernetics*, pp. 883–888, 2005.

[83] R. Sutton and A.G. Barto. *Reinforcement Learning : An Introduction*. The MIT Press, 1998.

[84] R.S. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Technical report, AT&T Labs– Research*, 1999.

[85] G. Taga. A model of neuro-musculo-skeletal system for human locomotion. *Biological cybernetics*, Vol. 73, pp. 97–111, 1995.

[86] Y. Takahashi, et al. Modular learning syatem and scheduling for behavior acquisition in multi-agent environment. In *RoboCup 2004 Symposium papers and team description papers, CD-ROM*, 2004.

[87] S.C. Tanaka, K. Doya, G. Okada, K. Ueda, Y. Okamoto, and S. Yamawaki. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *nature neuroscience*, Vol. 7(8), pp. 887–893, 2004.

[88] J. Tani, M. Ito, and Y. Sugita. Seif-organization of distributedly represented mulyiple behavior schemata in a mirror system:reviews of robots using rnnpb. *Neural Networks,*, Vol. 17, pp. 1273–1289, 2004.

[89] J. Tani and S. Nolfi. Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems. *Neural Networks*, Vol. 12, pp. 1131–1141, 1999.

[90] T. Taniguchi and T. Sawaragi. Design and performance of symbols self-organized within an autonomous agent interacting with varied environments. In *IEEE International Workshop on RO-MAN proceedings in CD-ROM*, 2004.

[91] T. Taniguchi and T. Sawaragi. Adaptive organization of generalized behavioral concepts for autonomous robots: Schema-based modular reinforcement learning. In *IEEE International Symposium on CIRA 2005 proceedings in CD-ROM*, 2005.

[92] R. Thom. L'espace et les signes. *Semiotica*, Vol. 29, pp. 193–208, 1980.

[93] E. Uchibe and K. Doya. Competitive-cooperative-concurrent reinforcement learning with importance sampling. In *Proceedings of International Con-*

ference on Simulation of Adaptive Behavior From Animals and Animate, pp. 287–296, 2004.

[94] J.V. Uexküll, G. Kriszat(日高, 野田訳, 思索社)“生物から見た世界”. *Streifzüge Durch die Umwelten von Tieren und Menschen Bedeutungslehre*. Verlag GmbH, 1970.

[95] H. Ulrich, G.J.B. Probst, (徳安彰　訳)（編）. 自己組織化とマネジメント. 東海大学出版会, 1992.

[96] S. R. Waterhouse and A. J. Robinson. Constructive algorithms for hierarchical mixtures of experts. In *Advances in Neural Information Processing Systems*, Vol. 8, pp. 584–590. The MIT Press, 1996.

[97] C. Watkins and P. Dayan. Technical note: Q-learning. *Machine Learning*, Vol. 8, pp. 279–292, 1992.

[98] Webots. http://www.cyberbotics.com. Commercial Mobile Robot Simulation Software.

[99] N.M. Weinberger. Specific long-term memory traces in primary auditory cortex. *nature reviews / neuroscience*, Vol. 5, pp. 279–290, 2004.

[100] R.J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, Vol. 8, pp. 229–256, 1992.

[101] D.M. Wolpert, K. Doya, and M. Kawato. A unifying comuputational framework for motor control and social interaction. *Phil Trans R Soc Lond B*, Vol. 358, pp. 593–602, 2003.

[102] D.M. Wolpert, Z. Ghahramani, and M.I. Jordan. An internal model for sensorimotor integration. *science*, Vol. 269, pp. 1880–1882, 1995.

[103] D.M. Wolpert and M. Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, Vol. 11, pp. 1317–1329, 1998.

[104] L. Xu, et al. An alternative model for mixtures of experts. In & T.K. Leen In G. Tesauro, D.S. Touretzky, editor, *Advances in neural information processing system*, chapter 7, pp. 663–640. MIT Press, Cambridge, 1995.

[105] V.P. Zhigulin and M.I. Rabinovich. An important role of spike timing dependent synaptic plasticity in the formation of synchronized neural ensembles. *Neurocomputation*, Vol. 58-60, pp. 373–378, 2004.

[106] けいはんな社会的知能発生学研究会（編）. 知能の謎 認知発達ロボティクスの

挑戦. 講談社ブルーバックス, 2004.
[107] ダニエル・デネット, (土屋俊 (訳)). 心はどこにあるのか. 草思社, 1997.
[108] J. フォンノイマン. 自己増殖オートマトンの理論. 岩波書店, 1975.
[109] 宇多田ヒカル. ”can you keep a secret?” in utada hikaru single clip collection vol.2. DVD, 2001.
[110] 伊庭崇, 福原義久. 複雑系入門　知のフロンティアへの冒険. NTT 出版, 1998.
[111] 伊藤一之, 松野文俊, 五福明夫. 強化学習による冗長ロボットの自律制御に関する研究　-身体像を考慮した強化学習-. 日本ロボット学会誌, Vol. 22(5), pp. 672–689, 2004.
[112] 内部英治, 銅谷賢治. 強化学習における適切な報酬関数の選択. 第 22 回日本ロボット学会学術講演会, 2004.
[113] 内部英治, 銅谷賢治. 複数報酬のもとでの階層強化学習. 日本ロボット学会誌, Vol. 22(1), pp. 120–129, 2004.
[114] 岡田敬司. 「自律」の復権. ミネルヴァ書房, 2004.
[115] 河本英夫. オートポイエーシス 2001　日々新たに目覚めるために. 新曜社, 2000.
[116] 河本英夫, 松野孝一郎ほか. 現代思想：システム論 (内部観測とオートポイエーシス). 青土社, 1999.
[117] 大矢雅則, 松岡隆志. 状態のフラクタル次元を用いたクレータおよび河川の複雑さの解析. 電子情報通信学会論文誌 A, Vol. J79-A(9), pp. 1590–1599, 1996.
[118] 矢野雅文, 富田望. 実環境における 2 足歩行の創発的リアルタイム制御. 日本ロボット学会誌, Vol. 23, pp. 11–16, 2005.
[119] 笠松幸一, 江川晃. プラグマティズムと記号学. 勁草書房, 2002.
[120] 今水寛ほか. 運動と言語，認知科学の新展開 (3). 岩波書店, 2001.
[121] 大浜幾久子（編）. ピアジェの発生的心理学. 国土社, 1982.
[122] 三田紀房. ドラゴン桜. 講談社, 2003–.
[123] 門内輝行. 人間-環境系としての生活環境の設計における創発システムの記号論. システム制御情報, Vol. 49(12), pp. 469–474, 2005.
[124] 丸山圭三郎. ソシュールを読む. 岩波書店, 1983.
[125] 橋本敬. 言語進化とはどのような問題か? ? 構成論的な立場から. 第 18 回日本人工知能学会全国大会予稿集，in CD-ROM, 2004.
[126] 鯨岡 和子 翻訳, 鯨岡峻. 母と子のあいだ 初期コミュニケーションの発達. ミネ

ルヴァ書房, 1989.

[127] 木村元, 小林重信. 確率的 2 分木の行動選択を用いた actor-critic アルゴリズム: 多数の行動を扱う強化学習. 計測自動制御学会誌, Vol. 37(12), pp. 1147–1155, 2001.

[128] J. Hoffmeyer (松野孝一郎, 高原 美規 (訳)). 生命記号論 宇宙の意味と表象. 青土社, 1999.

[129] 吉川弘之. 一般設計学序説 –一般設計学のための公理的方法–. 精密機械, Vol. 45(8), pp. 906–912, 1979.

[130] 吉川恒夫. ロボット制御基礎論. コロナ社, 1988.

[131] 山本浩司, 富田直秀ほか. 生態環境設計のための状態遷移モデルの作成. 第 32 回知能システムシンポジウム資料　pp.117–120, 2005.

[132] 川上浩司ほか. ユニバーサルデザインと不便益との関係に対する考察. SICE 知能システムシンポジウム, pp.361-366, 2005.

[133] 港隆史, 浅田稔. 環境の変化に適応する移動ロボットの行動獲得. 日本ロボット学会誌, Vol. 18(5), pp. 706–712, 2000.

[134] 倉田耕治, 和田浩司, 酒井裕. 自己組織化モデルの新しい方向. 計測と制御, Vol. 44(5), pp. 319–325, 2005.

[135] 山口昌哉. 複雑系科学の哲学に不足しているもの. 日本ファジィ学会誌, Vol. 9(5), pp. 603–610, 1997.

[136] 山本浩司ほか. 骨軟骨再生のための生体環境設計. システム・情報部門学術講演論文集.pp. 197–202, 2004.

[137] 山梨正明. 認知言語学原論. くろしお出版, 1997.

[138] 枝澤一寛, 高橋泰岳, 浅田稔. 複数学習器を用いたマルチエージェント環境における行動獲得. 第 22 回日本ロボット学会学術講演会 in CD-ROM, 2004.

[139] 上田修功, 中野良平. 混合モデルのための併合分割操作付き em アルゴリズム. 電子情報通信学会論文誌 D-II, Vol. J82-D-II(5), pp. 930–940, 1999.

[140] 黒木秀一. 競合連想ネットの漸近最適性と非線形関数の逐次学習への応用. 電子情報通信学会論文誌 D-II, Vol. J86-D-II(2), pp. 184–194, 2003.

[141] 西森秀稔. スピングラスと連想記憶. 岩波書店, 2003.

[142] 中西淳, A.J. Ijspeert, S. Schaal, G. Cheng. 運動学習プリミティブを用いたロボットの見まね学習. 日本ロボット学会誌, Vol. 22(2), pp. 165–170, 2004.

[143] 小嶋秀樹, 高田明. 社会的相互行為への発達的アプローチ　-社会のなかで発

達するロボットの可能性-. 人工知能学会誌, Vol. 16(6), pp. 812–818, 2001.

[144] 小林誠ほか. 二足歩行運動の動的安定性. 信学技報, Vol. MB,E97-48,1997-7, pp. 63–69, 1997.

[145] 徳安彰. システム理論の構成主義的意味論. 社会・経済システム, Vol. 16, pp. 115–120, 1997.

[146] 岡田昌史, 中村大介ほか. 力学的情報処理におけるアトラクタ設計法に関する研究–逐次設計による可塑性の導入と階層化設計による大規模化–. 日本ロボット学会誌, Vol. 23(5), pp. 583–593, 2005.

[147] 岡田昌史, 中村大介, 中村仁彦. 力学的情報処理を用いた自己組織的記号獲得と運動生成. 人工知能学会論文誌,, Vol. 20(3), pp. 177–187, 2005.

[148] 小嶋祥三. ミラーニューロンと言語の起源. 科学, Vol. 69(4), pp. 404–408, 1999.

[149] 石井信, 佐藤雅昭. オンライン em アルゴリズムによる動的な関数近似. 電子情報通信学会技術研究報告,NC97-142, pp.43-50, 1997.

[150] 森本浩一. デイヴィドソン–「言語」なんて存在するのだろうか. NHK 出版, 2004.

[151] 森本淳, 銅谷賢治. 階層型強化学習を用いた 3 リンク 2 間接ロボットによる起立運動の獲得. 日本ロボット学会誌, Vol. 19(5), pp. 574–579, 2001.

[152] 神宮英夫. スキルの認知心理学. 川島書店, 1993.

[153] 杉浦孔明ほか. 行動に基づく情報獲得に向けた形態と制御系の同時設計. 第 32 回知能システムシンポジウム資料　pp.105–110, 2005.

[154] 銅谷賢治 川人光男杉本徳和. 複数の状態予測と報酬予測モデルによる強化学習と行動目標の推定. 電子情報通信学会ニューラルコンピューティング研究会 1 月報告 (NC2001-118), 2001.

[155] 杉本徳和ほか. ダイナミクスの線形性に基づいて状態空間を分割する階層型強化学習. 電子情報通信学会 NC 研究会 6 月報告,NC2003-16, 2003.

[156] 杉本徳和, 鮫島和行, 銅谷賢治, 川人光男. 複数の状態予測と報酬予測モデルによる強化学習と行動目標の推定. 日本神経回路学会 第 12 回全国大会, 2002.

[157] 杉本徳和, 銅谷賢治, 川人光男. 教示者の行動目標を推定する見まね学習. 電子情報通信学会ニューラルコンピューティング研究会 10 月報告, 2003.

[158] 杉本徳和, 銅谷賢治, 川人光男. マルチエージェント環境における共通なシンボルの生成. 電子情報通信学会 NC 研究会 10 月報告, 2005.

[159] 豊田正. 情報の物理学. 講談社, 1997.

[160] 関野正志, 片上大輔, 新田克己. 基底関数間の相互作用に基づく状態空間自己組織化. In *The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 1D3-04*, 2005.

[161] 北村正晴, 木村逸郎（編）. 安全の探求　「人・社会と巨大技術が構成するシステムの安全学とその実践」. ERC 出版, 2001.

[162] 中村誠, 橋本敬, 東条敏. 言語動力学におけるクレオールの創発. *Cognitive Studies*, Vol. 11(3), pp. 282–298, 2004.

[163] 斉藤康己. シンボル・グラウンディング問題とは何か. 言語, Vol. 23(8), pp. 58–65, 1994.

[164] 川人光男. 脳の計算理論. 産業図書, 1996.

[165] 川人光男, 佐々木正人ほか. 運動　岩波講座認知科学 (4). 岩波書店, 1994.

[166] 川人光男, 柴田智広ほか. 特集　「川人学習動態脳プロジェクト」. 日本ロボット学会誌, Vol. 19(5), pp. 1–47, 2001.

[167] 浅田稔, 石黒浩, 国吉康夫. 認知ロボティクスの目指すもの. 日本ロボット学会誌, Vol. 17(1), pp. 2–6, 1999.

[168] 竹西素子ほか（編）. 別冊ロボコンマガジン　愛知万博　最新ロボットガイド. オーム社, 2005.

[169] 足立修一. ユーザのためのシステム同定理論. 計測自動制御学会, 1993.

[170] N. Wiener (池原止戈夫他訳). サイバネティクス　第 2 版　動物と機械における制御と通信. 岩波書店, 1962.

[171] 高橋泰岳, 浅田稔. 実ロボットによる行動学習のための状態空間の漸次的構成. 日本ロボット学会誌, Vol. 17(1), pp. 118–124, 1999.

[172] 高橋泰岳, 浅田稔. 複数の学習器の階層的構築による行動獲得. 日本ロボット学会誌, Vol. 18(7), pp. 1040–1046, 2000.

[173] 高橋泰岳, 浅田稔. 階層型学習機構における状態行動空間の構成. 日本ロボット学会誌, Vol. 21(2), pp. 164–171, 2003.

[174] 高橋泰岳, 浅田稔. 行動学習に基づく環境の階層的表現. 第 32 回知能システムシンポジウム資料 pp.111–116, 2005.

[175] J. Piaget(滝沢武久 訳). 発生的認識論. 白水社, 1972.

[176] 池上嘉彦. 記号論への招待. 岩波新書, 1984.

[177] 富田直秀, 中村桂子. 機能設計から生体環境設計へ　「安心」を育てる科学と医

療. 丸善, 2005.

[178] 西垣通. 基礎情報学. NTT 出版, 2004.

[179] 稲邑哲也, 中村仁彦, 戸嶋巌樹, 江崎英明. ミメシス理論に基づく見まね学習とシンボル創発の統合モデル. 日本ロボット学会誌, Vol. 22(2), pp. 256–263, 2004.

[180] 電気学会（編）. パターン・記号統合 (基礎と応用)〜ペットロボットのペットらしさを求めて. 丸善, 2004.

[181] 東京大学教養学部統計学教室（編）. 自然科学の統計学. 東京大学出版会, 1992.

[182] 21 世紀 COE プロジェクト動的機能機械システムの数理モデルと設計論. http://cme.coe21.kyoto-u.ac.jp/index.html.

[183] 海保博之, 黒須正明, 原田悦子. 認知的インタフェース コンピュータとの知的つきあい方. 新曜社, 1991.

[184] 三嶋博之. ギブソン知覚理論の根底：刺激情報，特定性，不変項，アフォーダンス. システム制御情報学会誌, Vol. 46(1), pp. 35–40, 2002.

[185] 尾形哲也, 菅野重樹, 谷淳. 動作プリミティブを介した人間とロボットの協調. 第 22 回日本ロボット学会学術講演会 in CD-ROM, 2004.

[186] 岡田美智男, 三嶋博之, 佐々木正人（編）. 身体性とコンピュータ. 共立出版, 2000.

[187] 浜田寿美男. ピアジェとワロン. ミネルヴァ書房, 1994.

[188] 鈴木敏, 上田修功. 混合回帰モデルのための smem アルゴリズム. 電子情報通信学会誌 D-II, Vol. J83-DII(12), pp. 2777–2785, 2000.

[189] J. R. Anderson(富田, 増井, 他訳:認知心理学概論, 誠信書房, (1982)). *Cognitive Psychology and its Implications*. W.H.Freeman and Company, 1980.

[190] 金子邦彦. 生命とは何か [複雑系生命論序説]. 東京大学出版会, 2003.

[191] 金子邦彦, 津田一郎. 複雑系のカオス的シナリオ 複雑系双書. 朝倉書店, 1996.

[192] 冨田望, 矢野雅文. 大脳基底核ー脳幹系のモデル化による二足歩行運動の創発的リアルタイム制御. 第 16 回自律分散システムシンポジウム資料, 2004.

[193] 坂東麻衣, 中西弘明. リアプノフの方法によるモジュール型強化学習に関する研究. 第 32 回知能システムシンポジウム資料　pp.121–125, 2005.

[194] 木村元, 小林重信. Actor に適正度の履歴を用いた actor-critic アルゴリズム：不完全な value-function のもとでの強化学習. 人工知能学会誌, Vol. 15(2), pp. 267–275, 2000.

[195] C. Belsey (折島正司訳). ポスト構造主義. 岩波書店, 2003.

[196] M. Polanyi (高橋 勇夫訳). 暗黙知の次元. ちくま学芸文庫, 2003.

[197] 有馬道子. パースの思想 記号論と認知言語学. 岩波書店, 2001.

[198] 酒井裕. Stdp 学習則が導く現象–神経系数理における新展開. 数理科学, Vol. 3(501), pp. 46–52, 2005.

[199] 酒井裕, 吉澤修治. Stdp によるシナプスパターンの競合と調節のメカニズム. 信学技法 NC, Vol. 21, pp. 55–60, 2003.

[200] 飯田隆. クリプキ–言葉は意味をもてるか. NHK 出版, 2004.

[201] 吉永良正. 「複雑系」とは何か. 講談社現代新書, 1996.

[202] 戸田山和久. 科学哲学の冒険　サイエンスの目的と方法をさぐる. NHK ブックス, 2005.

[203] 鮫島和行, 銅谷賢治, 川人光男. モジュール競合による運動パターンのシンボル化と見まね学習. 電信情報通信学会論文誌 D-II, Vol. J85-D-II(1), pp. 90–100, 2002.

[204] 鹿取廣人, 杉本敏夫（編）. 心理学. 東京大学出版会, 1996.

[205] 椹木哲夫. 人間を内部に含む系の創発 –インタラクションの内包する複雑性. システム/制御/情報, Vol. 45(11), pp. 615–622, 2001.

[206] 椹木哲夫. インタラクションにおける記号行為. 第２９回知能システムシンポジウム講演論文集, pp.157–162, 2002.

[207] 椹木哲夫. 「セミオーシス」って，なんどすねん？ システム/制御/情報, Vol. 48(4), pp. 36–37, 2004.

[208] 椹木哲夫. セミオーシス：記号過程がかもしだす知覚世界の動的複雑系. 21 回ファジィシステムシンポジウム講演論文集 in CD-ROM, 2005.

[209] 椹木哲夫, 北村正晴, 稲垣敏之. 自動化によるヒューマン・システム・インタラクションの諸相. ヒューマンインタフェース学会誌, Vol. 2(1), pp. 30–39, 2000.

謝辞

本論文を結ぶにあたり，ご指導やご援助，またはそれらの言葉では表現できない類の様々な影響を与えてくださった方々に感謝の意を表したいと思います．

まず，自分の価値観で至って自律的に勝手に行動する私を，広い視野と広い心で見守り，導いて下さった指導教官である京都大学工学研究科機械理工学専攻の椹木哲夫教授に感謝いたします．京大らしい放任主義ともいえる文化の中で，あくまで学生の個性を尊重しつつ，ポイントを抑えたご指導，ご教示を頂いたお陰で本研究は存在しえたと思っています．修士課程，博士課程学生として５年間研究室に在籍する中で，公私共々に大変お世話になりました．心より感謝の気持ちを申し上げます．

また，同研究室の堀口由貴男助手には日ごろの研究活動や，研究室生活において様々にお世話になりました．数少ない先輩として，また同研究室において椹木教授を除いた上での唯一のスタッフとしての堀口助手が居られたからこそ，日々の研究室においての研究活動を行うことが出来たと考えています．心から感謝いたします．

また，同研究室の秘書，湊忍女史には日ごろの研究室生活において多大なご支援を受けました．厚く御礼申し上げます．男ばかりの荒んだ研究室にあってポジティブな紅一点の存在がいかに研究室に潤いをもたらしていたか計り知れません．また，日々の研究生活では様々な煩雑な事務作業を代行していただきまして非常に助けられました．また，折に触れて学生に振舞ってくださった紅茶やお菓子は行き詰った研究の局面を踏破するエネルギーを与えてくれました．多面的なご援助に心から御礼申し上げます．

そして，私と共に研究生活を送る中で様々なことを教えたり教えていただいたり，邪魔したり邪魔されたりする中で影響を与え与えられ，この五年間の時間を刻み，当研究への向かう日々の環境を形成してくれた旧デザインシステム論，現機械システム創成学研究室の歴代メンバーに感謝します．読み手には伝わるべくもありませんが，本論文は私にとってはその日々の「記号」として存在することと思います．

京都大学情報学研究科の片井修教授には，その先進的なシステム観において多大なる影響を受け，折に触れてお話いただく奥深い世界観に私の興味と研究は強く方向付けられていった気がします．また，片井研究室という存在はディスカッションの場としても私の研究生活をより充実したものへと変化させてくれました．合わせて御礼申し上げます．

また，椹木研究室の先輩でもある，京都大学情報学研究科の塩瀬隆之助手には，その出会いが私が椹木研究室に入るきっかけとなっただけではなく，その研究者としては特殊とも言える社交的な活動スタイルに強く影響を受けました．感謝いたします．

「博士課程というのは孤独なもの」と聞きます．しかし，私の博士課程の研究生活はそうでもありませんでした．特に椹木研と片井研において自律適応的なシステムの理解に興味を持つ学生と，我々の専攻がその一翼を担う文科省２１世紀 COE プロジェクト「動的機能機械システムの数理モデルと設計論」に所属する博士課程の学生を中心に組織された有志研究会，京都自律適応系研究会のメンバーとの定期的な会合は，自らの研究室に閉じていては気づけない新たな視点や，情報を与えてくれました．その場において様々な影響を与えてくれた京都大学情報学研究科博士課程 杉浦孔明氏，嶋本正範氏，同工学研究科博士課程 青井伸也氏，山本浩司氏，坂東麻衣女史を始めとする参加メンバーの皆様に感謝します．特に杉浦氏，嶋本氏，坂東女史には本論文の校正作業においても御協力をいただきました．重ねて御礼申し上げます．

プライベートでの友人である京都府立大学農学部付属演習林の美濃羽靖助手にも校正作業にて御協力を頂きました．感謝いたします．

研究とは個人の思想，こだわりに強く動機付けられてこそ結実するものと思います．その意味では私は今までの人生で私に良い意味でも悪い意味でも様々な影響を与えてくれた家族・友人・先輩後輩・先生・生徒皆々にお礼を言わざるを得ません．特に，思春期に絶えずくだらないことで口論をした兄，谷口和大に下の意味で感謝を言いたいと思います．

兄 「お前は非常識やねん！」
私 「和君の常識ってなんやねん！そんなん知らんわ！」
兄 「常識は常識やろ，屁理屈ばっかり言うな！」

今から考えると，何度も繰り返された，このはたからみれば馬鹿らしい口論が，私の中で「人間は本質的に他人が頭の中で当たり前だと思うことを共有できない.」とい

う本研究の中で取り扱ったコミュニケーションにおける根源的問題に固執する契機となっていった気がします．

また，私が博士課程に進学するにあたり強く背中を押し，また，我侭な私に絶大な援助と暖かな心配りを絶え間なく与えてくれた両親に心より感謝します．こうして一編の作品として博士論文をまとめることが出来たのは間違いなく両親の支えのお陰であったといえます．

最後に博士課程が始まった頃に出会い，今は妻として私の生活上，精神上の支えである妻，宏美に感謝の気持ちを伝えたいと思います．

Everybody, Thanks a lot!