

## 第7章

# 脳に宿る心：構成とミュレーション



パリ

第2章から第6章で心と脳の振る舞いやデータについて議論した。本章では、それを情報処理システムとして実現する、自立分散方式を提案する。

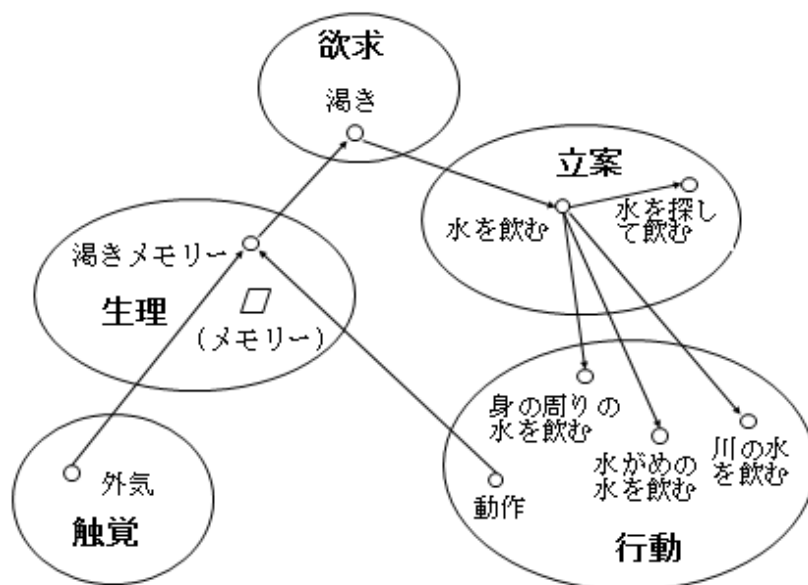
### 7.1 自律分散システム

ミンスキーは、それ自身心をもたない小さなエージェントの集合体として心をとらえる、心の社会論[46]を提案した。エージェントは必要に応じて自律的に動き始め、いくつかのエージェントと一緒にあって少し大きな仕事をする。小さな集合体がいくつか集まってもう少し大きな仕事をする。これを繰り返し、最終的には大きな機能を発揮する。処理ごとに参加するエージェントが異なり、異なる機能が発揮される。本書では、心の社会論に基づいてシステムを構成するが、エージェントと区別するため $\mu$ エージェントという名称を用いる。

$\mu$  エージェント群が互いに協力しながら一連の処理をこなす具体的なイメージを示そう。本節の最後に示すシミュレーションでは寓話の世界を舞台にして、主人公が暑さでのどの渇きを覚え、水を求めて探し歩くシナリオを用いる。そのシナリオに沿って以下のように $\mu$  エージェントが連続的に活性化する。なおシナリオでは、のどの渇きの度合いを保持する生理メモリーがあるものとしている。

[活性化連鎖]

- ① 渇きメモリーエージェントは、外気や主人公の動作量に応じ渇きの値を逐次更新する。
- ② それを監視する欲求領域の渇きエージェントは、値がしきい値を超すと、立案領域の水を飲むエージェントにのどの渇きを訴える。この訴えを受けて水を飲むという欲求が生じる。



大きな楕円は領域，小さな丸は $\mu$ エージェント，そして矢印は $\mu$ エージェントの活性化する順序を示す．4 辺形は渴きの度合いを示す生理メモリー．

図 7.1  $\mu$  エージェントの活性化の連鎖

- ③ 水を飲むエージェントは，まず身の周りの水を飲むという計画を選ぶ．この選択により具体的な物理目標が定まる．
- ④ 物理目標の実現に向けて立案木を掘り下げる．
- ⑤ 身の周りにある小川，池，水たまりを探すなど，さらに具体的な掘り下げを繰り返す，直ちに実行可能な所まで立案する．

.....

図 7.1 に活性化連鎖の様子を示している．

### 7.1.1 $\mu$ エージェント

まず，システムの基盤である $\mu$ エージェントを次のような形式でフレーム表現しよう：

- 名前(領域(所属する領域名)，
- 入力(入力メッセージ)，
- 実行(処理プログラム/NN)，
- 記述(処理の記述)，
- 出力(出力メッセージ))

一つの $\mu$ エージェントは、心の中のある領域に置かれ、通常停止している。ただし他の $\mu$ エージェントから不特定の $\mu$ エージェントに向けて発信されるメッセージは、常に監視している。これはいくつもあるラジオ局の放送に聴取者が関心を向けている状況に似ている。一つあるいは複数のメッセージが入カスロットの内容と一致すると、自ら動き始め、実行スロットの処理を実行する。記述には処理結果を示すメッセージが埋め込まれている。処理の終了後出力からメッセージを発信し、停止状態に戻る。なお実行スロットには、文書とかデジタル画像のような静的データのみが格納されることもあり、その場合 $\mu$ エージェントは単なるメモリーとなる。また処理中に他の $\mu$ エージェントと情報交換することもあり、その場合は返答が得られるまで待機状態となる。

メッセージは次の形式で表現する：

メッセージ(領域 1, 領域 2, 内容(値 1, 値 2, …))

すなわち領域 1 から領域 2 へ、それぞれ値 1, 値 2, …という成分をもつ内容を発信する。成分としては記号や数値に加え、 $\mu$ エージェントの実行結果を要約する格構造も許され、その場合は短い文に相当する内容が出力されることになる。

$\mu$ エージェントの実行スロットの処理について補足しておく。処理には抽象度に応じて低次の信号系を扱う NN と高次の記号系を扱うプログラムとがある。信号系は、述べるまでもなく、意識の上ののりにくい知覚的特徴や運動的特徴を扱う。特にプログラムの命令に相当する $\mu$ エージェントの処理は、NN で実現される。これにより抽象度の高いプログラムの、神経生理への接地が保証される。

次に、図 7.1 に沿って $\mu$ エージェントの具体例を示そう。渇きメモリーは外気や自身の動作量に応じて次のように値を更新する：

渇きメモリー(

領域(生理),

入力(メッセージ(触覚, 生理, 外気(温度, 湿度)),

メッセージ(行動, 生理, 運動(動作, 時間, 負荷))),

実行(読み取り(渇き\_前),

渇きの計算(渇き\_前, 動作, 時間, 負荷, 温度, 湿度),

書き込み(渇き\_後)),

記述(渇く(のど, 程度), ---),

出力(メッセージ(渇きメモリー, 欲求, 度合(渇き\_後)),

メッセージ(渇きメモリー, 言語, 記述(渇く(のど, 程度), ---))))

渇きメモリーは、入力メッセージとして一つは触覚領域から、もう一つは行動領域からの

発信を待つ。前者は外気に温度と湿度のデータ、後者は運動に動作、時間ならびに（荷物などの）負荷のデータをもたらす。二つのメッセージが届くと、実行スロットの三つの処理が順番に走る。まず読み取りが渴きメモリーから現在の値、渴き\_前を読み取る。次に計算がその値に基づいて、更新データを計算する。温度や湿度が高いほど、あるいは動作量、時間、負荷が大きいほど大きな値になる。計算結果は渴き\_後の値となり、最後の書き込みによってメモリーに書き込まれる。その後欲求領域と言語領域とに向けて二つのメッセージが発信される。言語領域に向けては、“喉がとても渴いた”という内容の格構造が発信される。

$\mu$  エージェントで用いている大文字は、変数を表す。入力スロットにメッセージを受信して特定のデータすなわち定数が得られると、 $\mu$  エージェントは活性化し、各変数には定数が割り振られて処理が進む。最終的には出力スロットの変数にも定数が割り振られてメッセージを発信する。メッセージの発信は、具体的には掲示板のような、システムにおける公共メモリーに書き込む。一般の $\mu$  エージェントは常にそこを見張っており、当該データが書き込まれると読み取って自己の入力とする。

### 7.1.2 活性化連鎖

ある $\mu$  エージェントが活性化すると関連する他の $\mu$  エージェントも活性化する。これには二つのモードがある：

- ・ モード1 メッセージモード

先に述べたメッセージの発信、受信に基づく。

- ・ モード2 埋め込みモード

フレーム表現は、いわゆる埋め込み構造をもち、親が活性化すると埋め込まれた子の $\mu$  エージェントも活性化する。

モード2の具体例を図7.1に沿って示す：

水を飲む(

領域(立案),

入力(メッセージ(欲求, 立案, 渴き(度合い, しきい値))),

実行(目標(水を飲む),

計画(タイプ2(身の周りの水を飲む),

タイプ3(住処に帰って水がめの水を飲む;

橋の下に行って川の水を飲む),

タイプ4(水を探して飲む);

水分の多い食べ物を探して食べる)),

記述(---),

出力(---))

前章の図 6.7 を, 2.4.2 項で述べた行動のタイプに沿って少し詳しくフレーム表現している. この  $\mu$  エージェントの目標はエージェント名と同じ, 水を飲むである. 各タイプに沿っていくつかの計画が示されており, それぞれは一つの  $\mu$  エージェントである. 親エージェントが活性化すると, 子エージェントも活性化する.

## 7.2 システムの振る舞い

システムの構成については, 既に 2.1 節の図 2.1 と 6.1 節の図 6.1 にそれぞれ機能と階層の観点から概念図を示した. 六つの心理機能と二つの生理機能はそれぞれ  $\mu$  エージェントの集合体によって実現され, 心と脳においてある領域を占める. 具体的な振る舞いについてはすでに第 2 章から第 4 章で詳しく述べた. 本節では, 今までの議論で必要にして十分な機能に目が行き届いているかを考察し, その上で特に注目すべきいくつかの振る舞いを活性化連鎖として記述しよう.

### 7.2.1 心理機能

六つの機能がそれぞれ必要な機能であることは明白である. いずれを欠いても心の働きの一部に支障をきたす. ではそれら以外に漏らしているものはないか. 認知科学の事典[27]を参考にすると, 次のような項目名が見られる:

*思考, 知識, 記憶, 学習, 発達, 推論, 問題解決, パターン認識, 言語, 読み書き, 神経生理, 運動, 感情, 視覚, 聴覚・音声, 遂行, 相互作用, 自己と性格, 意思決定, 注意, ……*

細かく観察すると, 同一の機能の異なる側面を強調する項目がある. 例えば読み書きは, 言語行為の一環で, 読むあるいは書く部分を強調している. そのため読み書きを言語に含めても大きな問題はない. 次に同義語あるいは類義語に気が付く. 例えば感情に対し, (上の項目群には含まれていないが) 情緒あるいは情動などの同義/類義語がある. それぞれにおいて微妙なとらえ方の相違があろうが, 共通な内容に注目して, 例えば情緒を代表と定める. このようにすると比較的固有な機能として認識, 表出, 情緒, 記憶, および言語が取り出せる. しかし思考, 問題解決あるいは意思決定のような総合的な機能をもつ項目が残る. これらに関し AI 研究が歴史的に課題としてきた項目を検討すると, 既知の事実から未知のことを推察する

ことと、新しく何かを考案することの二つの基本機能が見出せる。両者を合わせた推考を先の五つに加えると、例えば問題解決はまず問題を認識し、次に推考を加え、最後に表出する総合的機能ととらえることができる。このようにして総合機能を還元論的に把握すると、提案した六つになる。以上認知科学やAIの知見に基づくと、基本的な機能に対し目は届いていると言えるのではなかろうか。

最後に、基本機能の代表的な振る舞いを整理しておこう。

基本機能（

個別（（認識，特徴抽出），（推考（推論，考案）），（表出，素性分解），  
（情緒，（喚起，反応）），（記憶（一般化，特殊化），（言語，記述）），  
共通（パターン照合，探索，試行錯誤））

### 7.2.2 いくつかの活性化連鎖

それでは領域内の、あるいは領域にまたがる小さな振る舞いはどうであろう。無数と言っていいほどの振る舞いがあり、網羅的なチェックは不可能に近い。しかし本システムでは小さな振る舞いを活性化連鎖として把握しようとしている。二、三の事例を取り上げて見よう。

#### ・ 夢

夢においては、眠っているにもかかわらず心が働く。先にふれた、知覚器にゲート機能があると考えると、ゲートが閉じているとき睡眠状態になる。にもかかわらず、心のどこかで活性化連鎖の生じているのが夢を見ている状態と解釈できる。夢の中の部分的な事物は記憶に基づいているため体験／見聞したエピソードに沿っている。しかし、睡眠中で制御が正常でないため、全体の組み立ては体験／見聞したエピソードと違ってくる。

#### ・ 神経症

心が閉じてしまい、同じことをどうどう巡りで思い悩む神経症に注目する。4.1節の神経生理、特に図4.3でループを持つリカレント型のNNを示した。通常はループになっていない回路が何かの事情でループを作ると、そこを信号がぐるぐる回って回路が疲弊してしまう。換言すると、活性化連鎖が閉じてしまい、同じ考えを繰り返す。電気回路ではこのような状態を発振状態と呼び、遂には回路が損傷してしまう。上記神経症は、発振状態のNNが損傷しつつある状態と解釈できる。

#### ・ 葛藤

複数の欲求が生じ、いずれを優先して処理すればよいか迷うことがある。通常欲

求はいくつも生じるが、ある一つが強い場合は他の欲求が抑制される。これは強い欲求の出力が、他の欲求には抑制性で入力され、自己には興奮性でフィードバックされる[46]。結果として強い欲求は、ますます強くなり集中的に処理される。しかし二つの欲求が同じくらいの強さだと、その選択に迷って集中的な処理に入れない。これが葛藤の状態である。

- ・ 自己意識

自己意識の枠組みはすでに述べた。一般に  $\mu$  エージェントは処理結果を不特定多数に向けて発信する。もし自己意識の管理人がそれらを傍受するなら、心のあちこちで生じている事柄を知り得る。傍受こそが自己意識の引き金を引く鍵である。

一連の活性化連鎖は、 $\mu$  エージェント自己意識から始まる。

- ① 自己意識は、欲求領域が発信する引き金メッセージを待っている。また実行中の活性化連鎖にも注目し、現在活性化している  $\mu$  エージェントのメッセージも傍受している。両方のメッセージが届くと自ら活性化し、割り込む。
- ② まず、直前に活性化していた  $\mu$  エージェントに現状を保持したまま待機状態に入るよう指示し、実行中の活性化連鎖を一時中断させる。
- ③ 次に、直前の活性化連鎖で用いていた作業領域のデータを複製し、それに基づいて、自己意識に関連する  $\mu$  エージェント群を活性化連鎖させる：
  - ・ 連鎖の内容は、割り込み直前の考えや行為を批評または正当化するなどである。
  - ・ 連鎖の規模は、作業領域の記憶容量と処理時間の制約のため、ごく少数の  $\mu$  エージェントが活性化する。
- ④ 活性化連鎖が終わると、自己意識は不特定多数に向けて結果を発信する。
- ⑤ これを受けて言語領域は、簡単な語句あるいは文を生成する。
- ⑥ 自己意識は、待機中の  $\mu$  エージェントに割り込み前の活性化連鎖に復帰するよう指示する。

では具体的に自己意識とそれにかかわる  $\mu$  エージェント群はいずれの領域に所属するのであろう。言語領域は最も抽象度が高く、心全体を見通しやすい。また意識内容を言語化するにも好都合なので、それらの  $\mu$  エージェント群は言語領域に所属すると考えるのが適切ではなかろうか。

以上により、提案システムは一応必要かつ十分な機能を満たしていると判断する。

### 7.3 イソップワールド

第2章から第4章で述べた基本機能に関しては、それぞれ個別の実験システムを

構築し、シミュレーションを重ねてきた。以下では、総合的なシステムイソップワールドによるシミュレーションについて述べよう [44], [45].

### 7.3.1 実験のシナリオと装置

初めに、基本シナリオとして、イソップ寓話の狐とぶどうを選んだ。政治とか経済など複雑な社会システムとの関わりは避け、情緒を含む、素朴な児童レベルの心理活動や人間関係をシミュレーションするうえで適しているからである。のどの渇いた狐があれこれ推考、行動し、館の庭にたどり着く。ぶどう棚を見つけて何度も飛びつくが失敗し、最後に負け惜しみを言う。基本シナリオでは情緒の処理や対話の処理などに焦点を当て、それぞれ猟師を見かけてとっさに草むらに隠れる、友達の狸に出会って川原に誘われる、など変化形をいくつも準備した。さらに数台のパーソナルコンピュータ、画像装置、ならびに音声装置を結合したハードウェア上に自律分散システムを実現した。およそ一万程度の  $\mu$  エージェントの処理能力がある。狐の一連の行動や対話の状況は画面にアニメーションで表示され、心理的・物理的活動は音声装置から狐の独り言として発話される。特に情緒の心理状態は音楽に自動変換され、嬉しいときはリズムカルな、また悲しいときは陰鬱なメロディが背景に流れる。ただし狐の外界認識機能は個別実験のみで、また自己意識も議論のみで、イソップワールドには組み込まれていない。また生理レベルの処理はプログラムでシミュレーションしている。

### 7.3.2 実験結果

図 7.1 に主人公の狐がぶどうに飛びつくアニメーションの一場面を示している。詳しく見ると、独り言が画面下の方に表示されている。独り言は日本語あるいは英語で音声装置からも発話され、背景に悔しい思いを反映した音楽が流れている。次の文章は、考案の際の狐の独り言を示している：

[狐の独り言]

今日は暑いなあ。交差点から 300m のけもの道 1 にいる。のどがかなり渇いている。のどの渇きをうるおしたい。危険なことはせずうるおしたい。水を探して飲もうかなあ。水分の多い食べ物を食べようかなあ。住処に帰ろうかなあ。住処は遠い。住処に帰るのはやめた。橋の下に行こうかなあ。橋の下は遠い。橋の下に行くのはやめた。川原に行こうかなあ。川原に行くのはやめた。他の方法を考えようかなあ。水のある場所を探そうかなあ。池があった。池を探そうかなあ。池には B 池



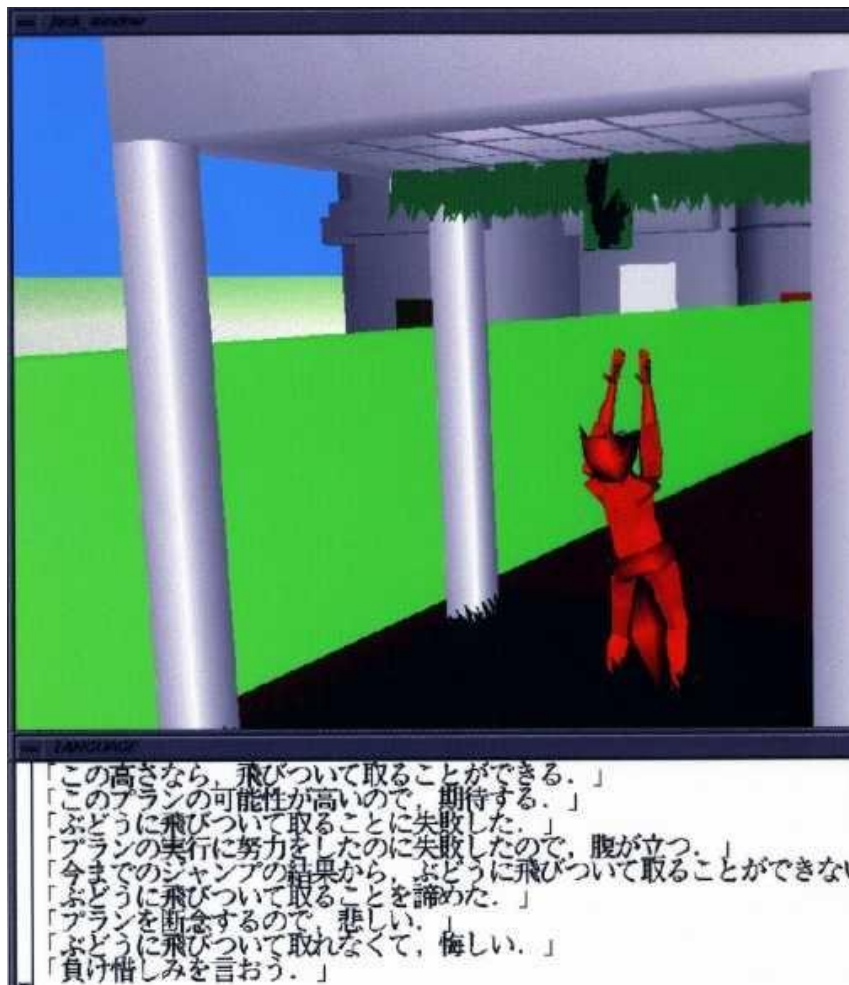


図 7.1 イソップワールド：狐とぶどうの一場面

があった。B池に行こうかなあ。B池はイソップワールドにあった。B池の近くには  
 猟師小屋があった。猟師小屋には猟師がいるだろう。猟師は人間だ。人間は危ない。  
 B池に行くのは危ない。-----

水分の多い食べ物を食べようかなあ。果物を探して食べようかなあ。野いちごを  
 食べようかなあ。野いちごは春になった。季節は今夏だ。季節は野いちごを食べる  
 のに良くない。野いちごを食べるのはやめた。りんごを食べようかなあ。りんごは  
 秋になった。季節はりんごを食べるのに良くない。りんごを食べるのはやめた。山  
 に行つてぶどうを食べようかなあ。ぶどうは夏になった。季節は今夏だ。季節はぶ  
 どうを食べるのに良い。ぶどうはぶどうの木にあった。ぶどうの木はD山にあった。  
 D山に行こうかなあ。D山は遠い。ぶどうを食べるのはやめた。-----

言語処理技術が未熟なため単調な繰り返しが多くなっているが、主人公の知識を背景に立案の過程が伺える。水を飲む $\mu$  エージェントの活性化に沿った文脈も見受けられる。記述スロットに埋め込まれた飲む、行く、探す、計画するなどが、活性化の後メッセージとして言語領域に発信される。これを受けた言語領域が上記のような日本語あるいは英語の文章を生成する。

次は、友人の狸との対話の例である：

[狐と狸との対話]

K1: やあ.

T2: やあ.

K3: どこへ行くの?

T4: 川へ魚を捕りに.

K5: いいねえ.

T6: それで君は?

K7: 館の池へ水を飲みに.

T8: 館は危ないよ.

K9: どうして?

T10: だってさっき前を通ったら人声がしたよ.

K11: 本当? どうしようかなあ.

T12: 川へ一緒に行かない?

K13: でも遠いよ.

T14: あそこの水は冷たくておいしいよ.

K15: 分かった. じゃあ, 行こう.

この対話は、3.3節の言語で述べた友人同士の対話の例を応用したものである。コンビニが館に、またスーパーが川に代わっているだけで、友人のよしみとか勧誘の戦略は同様である。

以上の結果から理解されるように、イソップワールドでは第2章から第6章で意図した諸機能や常識がかなり実現できている。また認識の個別実験では、物、事象、属性などに関し2.2節の議論を裏付ける結果が得られている。自己意識に関しても、最も重要な自己を観察する機能は独り言の内容からも認められる。これらの結果を総合的に判断して、図2.1および6.1のシステムの原型が自律分散システムとして実現できているものと認められる。