

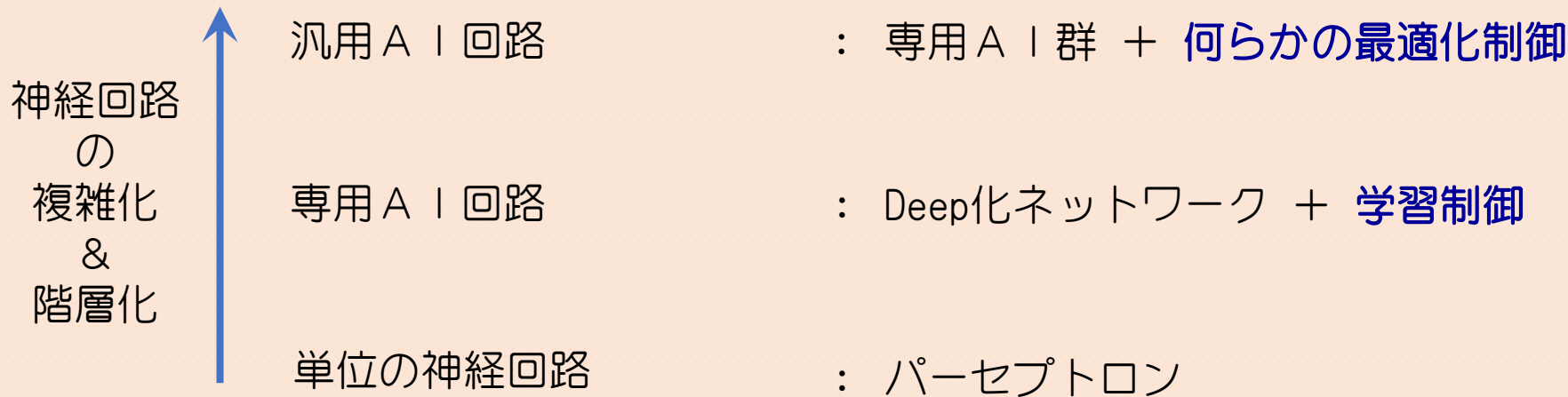
AGI用対話型計算機構の基礎検討について

(実装に適する回路から、汎用AIを考える)

情報統合技術研究合同会社

岡島義憲

出発点：専用AI群の統合 & 最適化により汎用AIにならないか



- 先ずは、最適化制御アルゴリズムを実装するプラットフォームを構想したい。
- 計算機構の全体像はどのようになるか？

ニューラルネットワークの回路表現の特徴

1) 確率的論理を表現できる可変な(プログラマブル)な回路は少ない

a) CPU : 状態遷移回路+演算回路+プログラム+カウンター+メモリ
(カスタムプロセッサ含)

b) FPGA : 接続スイッチ+真理値表

c) ハッシュ表(確率的真理値表) + ネットワークスイッチ

いずれも、「データで論理を表現する」ことで、論理を可変とする

2) ニューラルネットワークの単位回路のモジュール化は必須

- 全結合表現には困難が多い

- ある粒度でモジュール化した回路をカスケード化(多段化)させる方向は必須

専用AI と汎用AI の差異（直感的な理解に基づく）

- 専用AIの学習(Training) : Back-Propagation + 勾配降下法
- 汎用AIの最適化(進化) : 専用AIの学習 + 意識 + 言語活動 + α

対話や内省

ぼおっとしていては、発明も
発見もできないだろう！

↓ 戦略

先ず、「専用AI回路(群)を自律的に最適化する機能」の追加を考え、
(搭載する「改良アルゴリズム」は別途考える。)

更に、「どのような計算機構の追加が必要か」を考える。

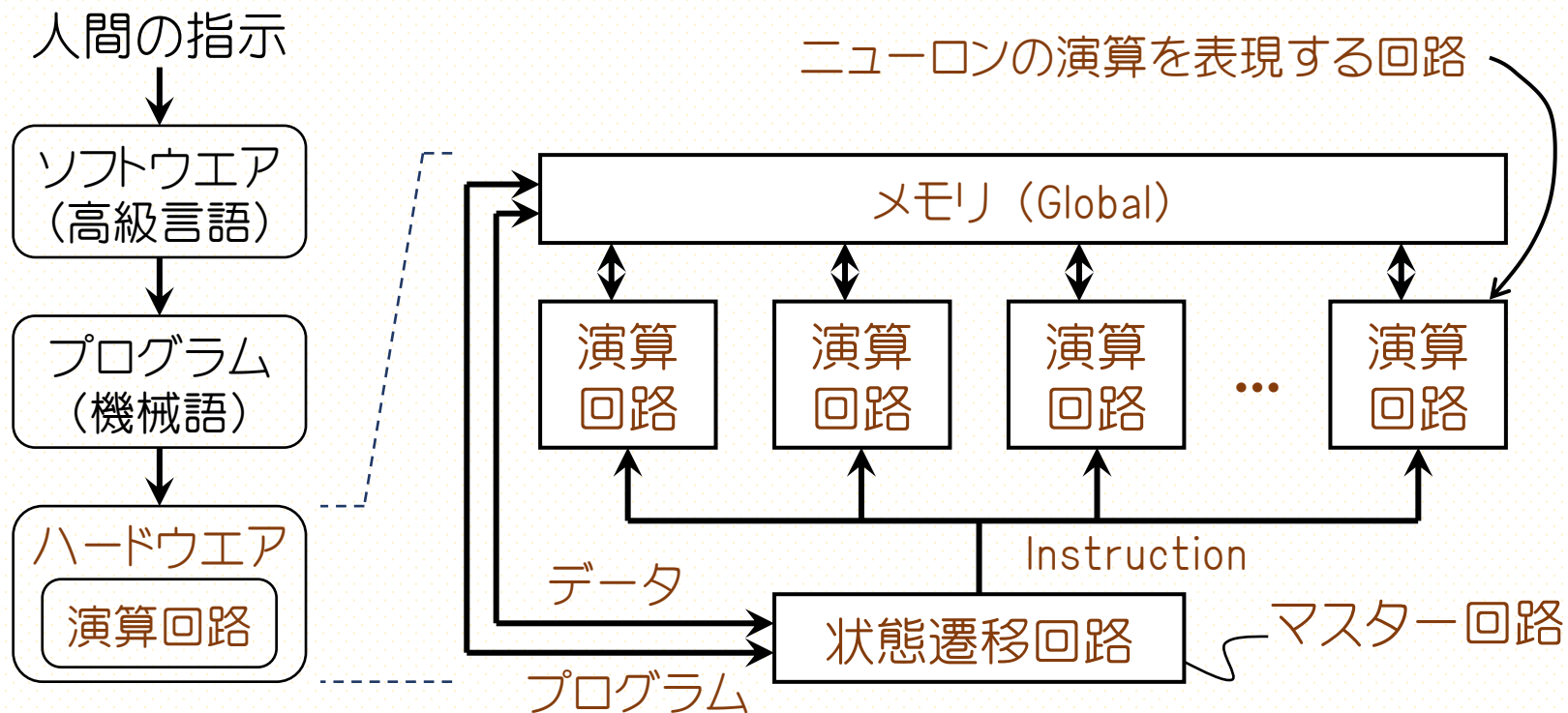
参考：Marvin Minsky [1961/1965] より (注) 意識あり

- Self-Improvementを行うには、改良されるSelfの外側に、改良するSelfが必要だろう。
- 人間には、物事の改善点を見つけることができる。トライアルをある程度行くと、改善方向が分かって来るからだ。AIにも、問題の本質や究極のゴールはなかなか定義できないだろうが、「進歩の方向」は、発見できる筈だ。
- 「勾配の大きい方向に山を登る」というアルゴリズムを頼りにすると、最善箇所ではない局所的ピークにトラップされてしまう可能性がある。局所最適から逃れるには、全体を調べるしかない。従って、その調査の総量を最小化する方式が鍵となる。
- 知性は、多数のサブMachineによって制御されて自律動作するModelsと、Mindとの間の内的対話(Introspection)によって発現すると考えると理解しやすい。

Minsky [1961] ; “Steps toward Artificial Intelligence”

Minsky [1965] : “Matter, Mind and Models“

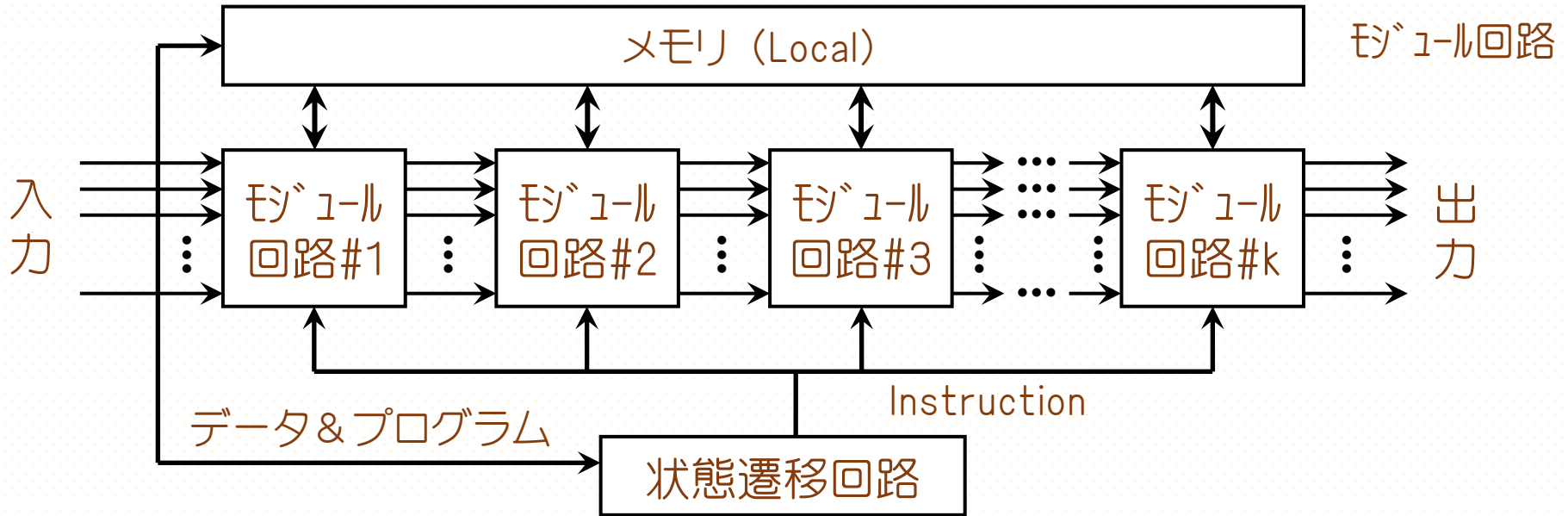
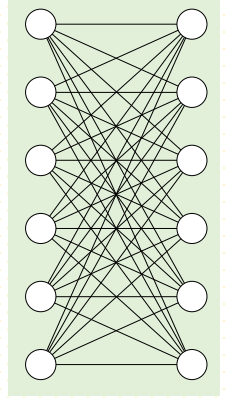
一般的な計算機構 (チューリングマシン)



(注) 図中の回路は、可変性(論理や分岐)が担保されたハードウェアに実装されたソフトウェアを表現しているが、説明中には、各回路が、ハードウェアであるかのようなイメージで表現しています。

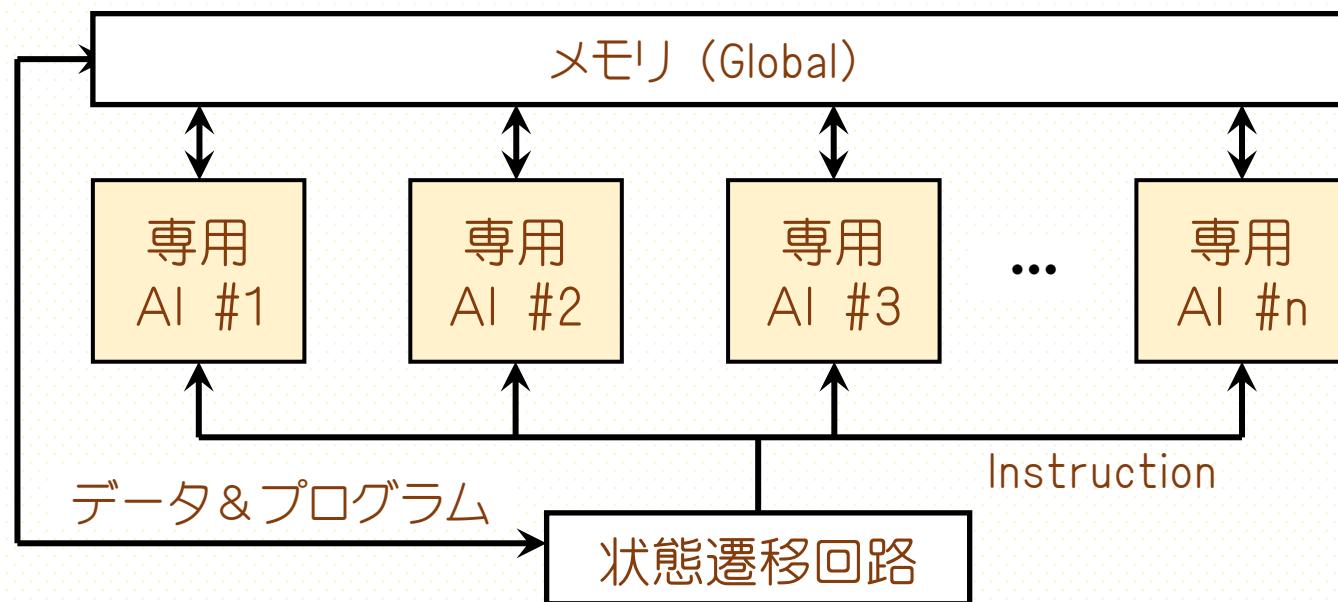
専用 AI 回路モデル : Deep Neural Network

- 演算回路にニューラルネットワークを表現する用いることで、パラメータ調整で論理を修正できるようにする。



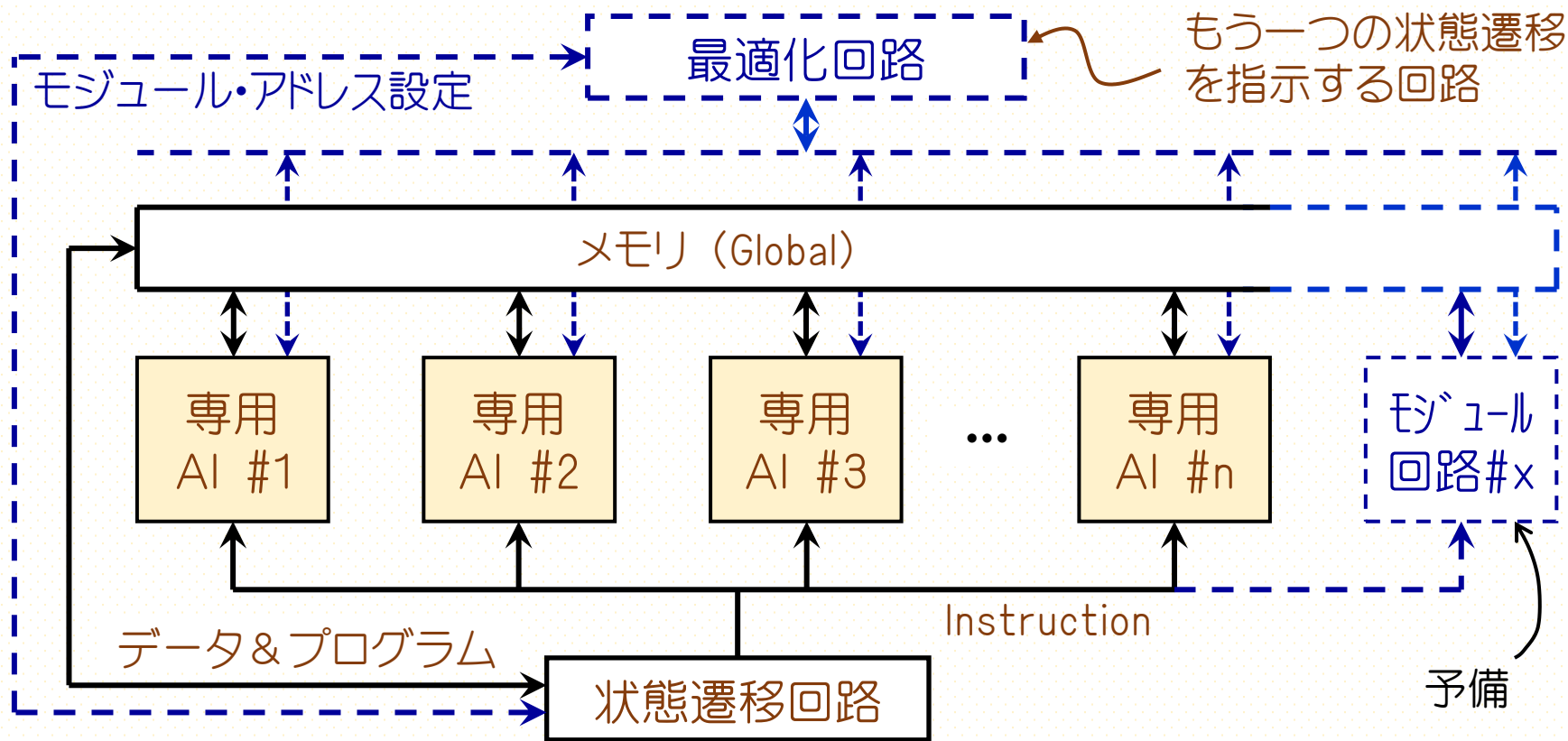
(注) 図中の回路は、可変性(論理や分岐)が担保されたハードウェアに実装されたソフトウェアを表現しているが、説明中には、各回路が、ハードウェアであるかのようなイメージで表現しています。

汎用 AI 化の出発点：専用 AI 回路群を Integration



(注) 図中の回路は、可変性(論理や分岐)が担保されたハードウェアに実装されたソフトウェアを表現しているが、説明中には、各回路が、ハードウェアであるかのようなイメージで表現しています。

改良されるSelfの外側に、改良するSelfを付加する



(注) 図中の回路は、可変性(論理や分岐)が担保されたハードウェアに実装されたソフトウェアを表現しているが、説明中には、各回路が、ハードウェアであるかのようなイメージで表現しています。

最適化回路が、自律的に演算回路の動作を調べ改良する。

もう一つの状態遷移
を指示する回路

最適化回路

人間の指示

ソフトウェア
(高級言語)

プログラム
(機械語)

ハードウェア
演算回路

最適化
回路

データ収集 & 論理の自己改良

データ
プログラム

状態遷移回路

メモリ (Global)

演算
回路

演算
回路

演算
回路

演算
回路

Instruction

(注) 図中の回路は、可変性(論理や分岐)が担保されたハードウェアに実装されたソフトウェアを表現しているが、説明中には、各回路が、ハードウェアであるかのようなイメージで表現しています。

そもそも、改良されるSelfは何か？

- 出発点で自己改良するとした対象は、専用AIが獲得するとした論理
 - 世界Model
 - 生成Model
 - 辞書
 - 等
- 汎用AIとして望まれる機能はもっとある
？

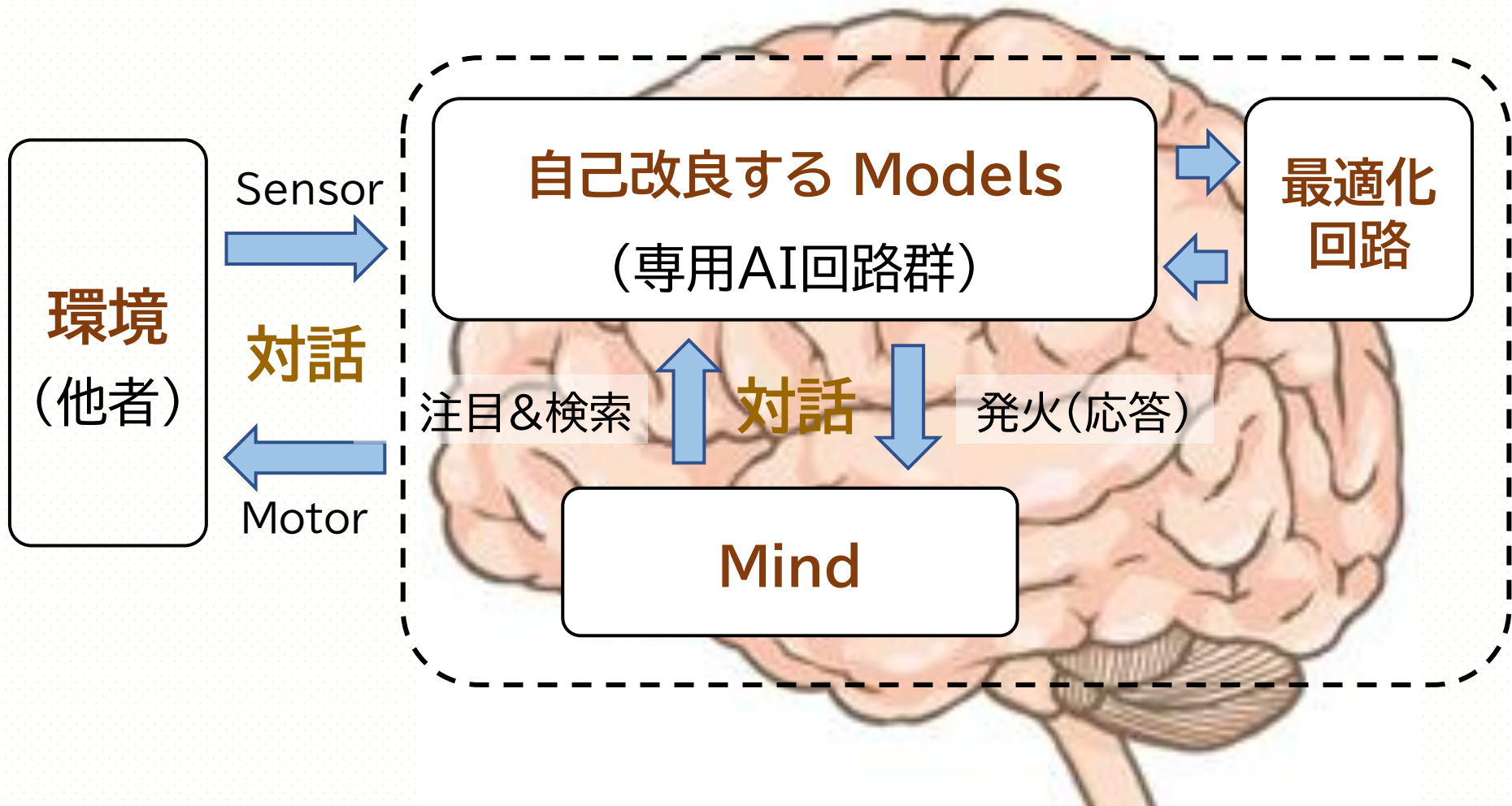
参考：Marvin Minsky [1961/1965] より (注) 意識あり

- Self-Improvementを行うには、改良されるSelfの外側に、改良するSelfが必要だろう。
- 人間には、物事の改善点を見つけることができる。トライアルをある程度行くと、改善方向が分かって来るからだ。AIにも、問題の本質や究極のゴールはなかなか定義できないだろうが、「進歩の方向」は、発見できる筈だ。
- 「勾配の大きい方向に山を登る」というアルゴリズムを頼りにすると、最善箇所ではない局所的ピークにトラップされてしまう可能性がある。局所最適から逃れるには、全体を調べるしかない。従って、その調査の総量を最小化する方式が鍵となる。
- 知性は、多数のサブMachineによって制御されて自律動作するModelsと、Mindとの間の内的対話(Introspection)によって発現すると考えると理解しやすい。

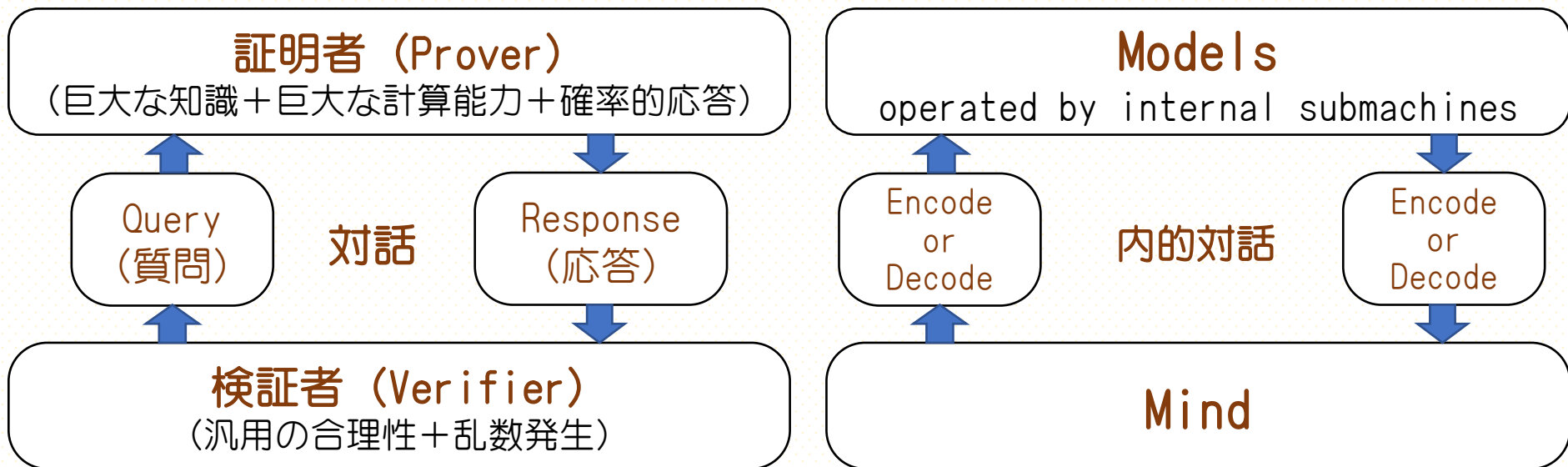
Minsky [1961] ; “Steps toward Artificial Intelligence”

Minsky [1965] : “Matter, Mind and Models“

脳内では、複数エージェントが対話しているのか？



Models-Mind構造 と Interactive Proof System

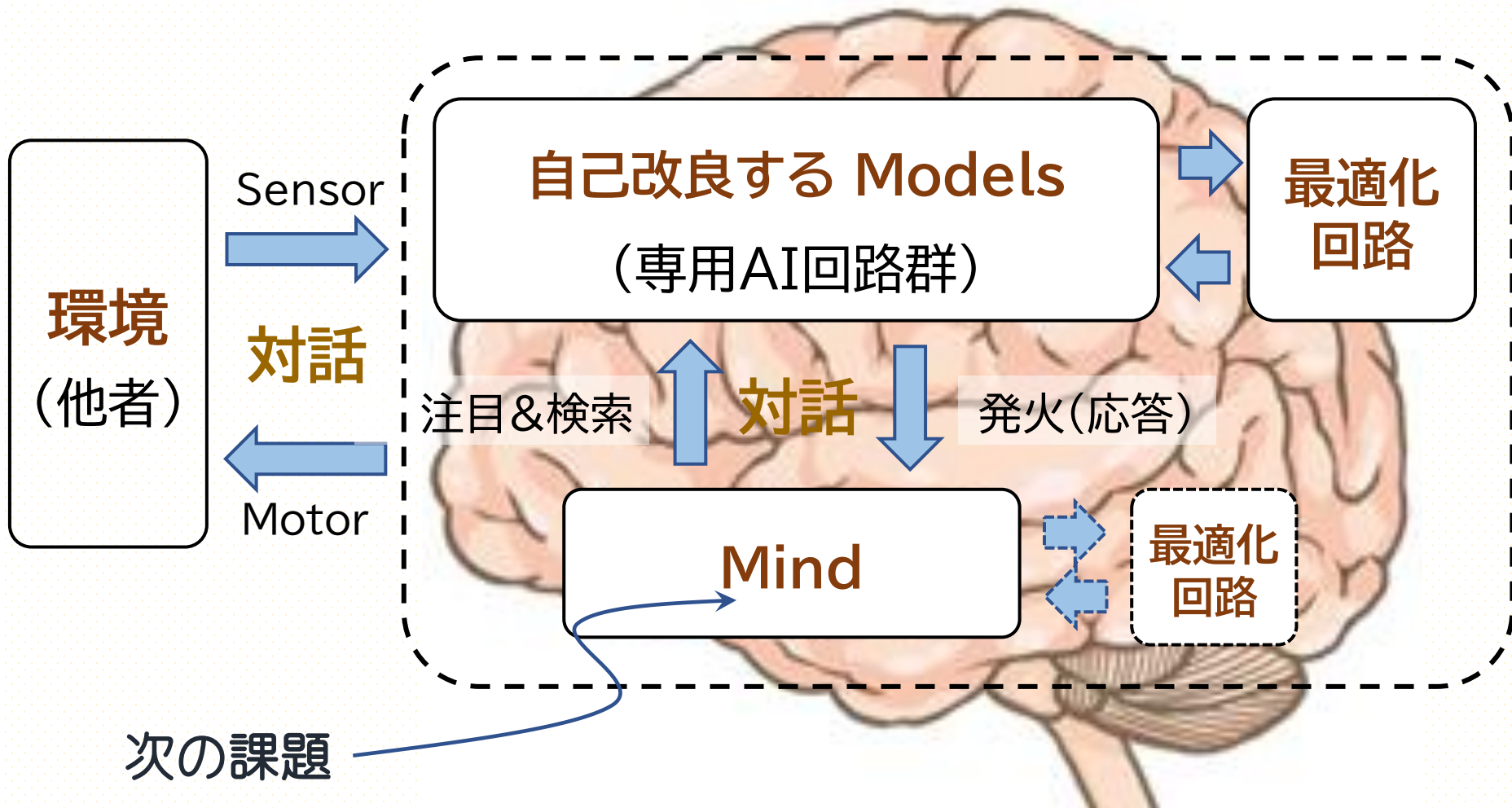


(A) Interactive Proof System

(B) MinskyのIntelligence発現構造

Goldwasser, S., et al. [1985] ; “Knowledge Complexity of interactive proof systems”.
Babai, L.[1985] ; “Trading Group Theory for Randomness”,

脳内では、複数エージェントが対話しているのか？

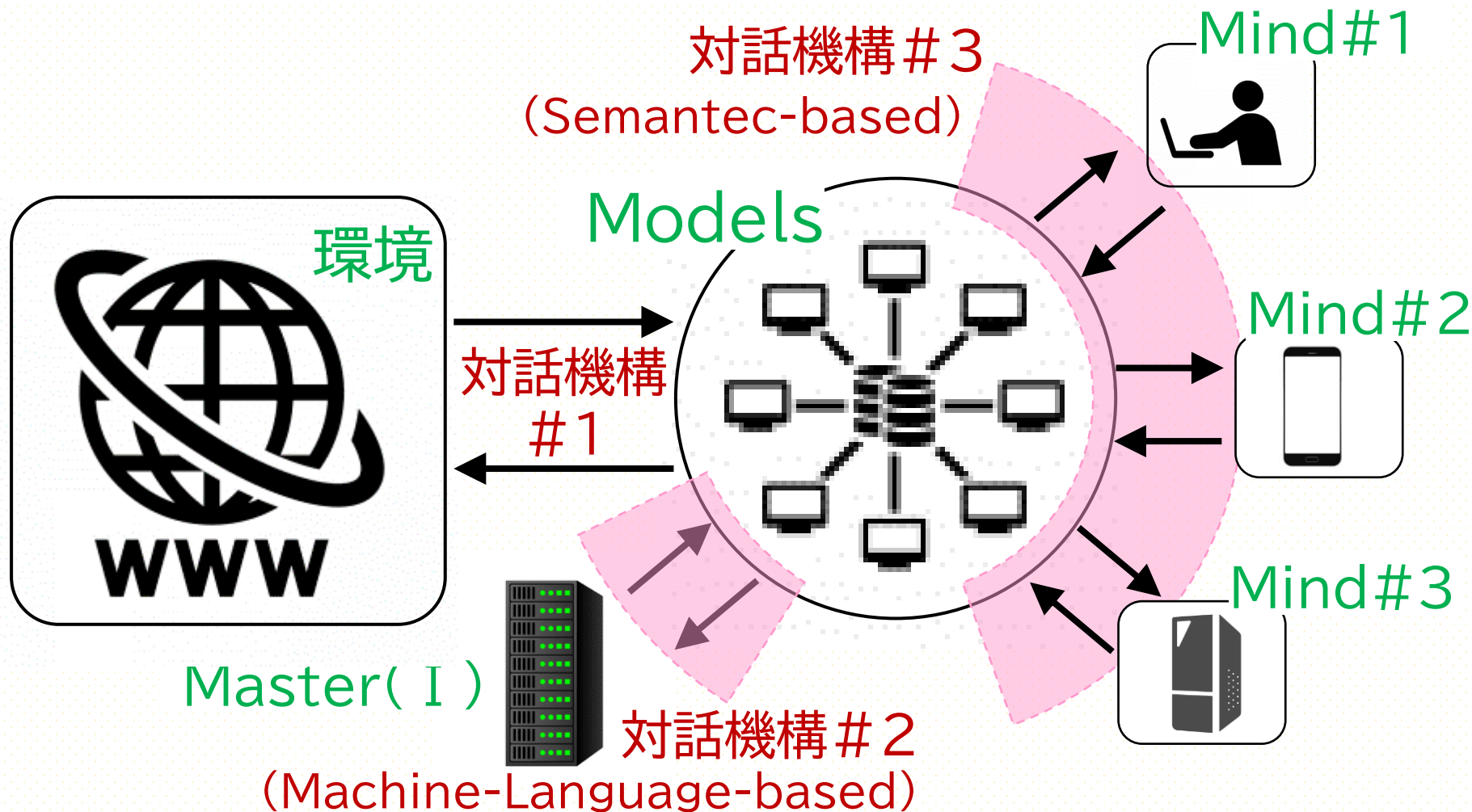




ご清聴ありがとうございました。

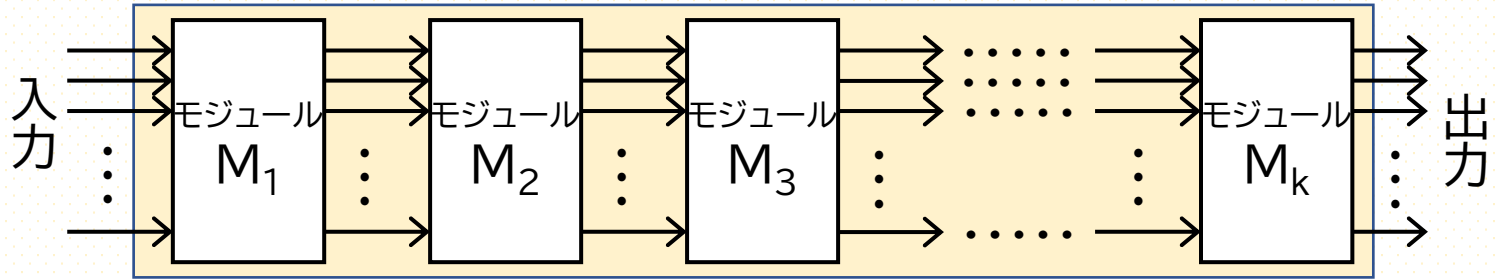
議論や情報交換を行って頂ける方、
論文原稿に掲載したE-Mailアドレスにご連絡下さい。

AGIの異形：世界モデルを提供するセンター

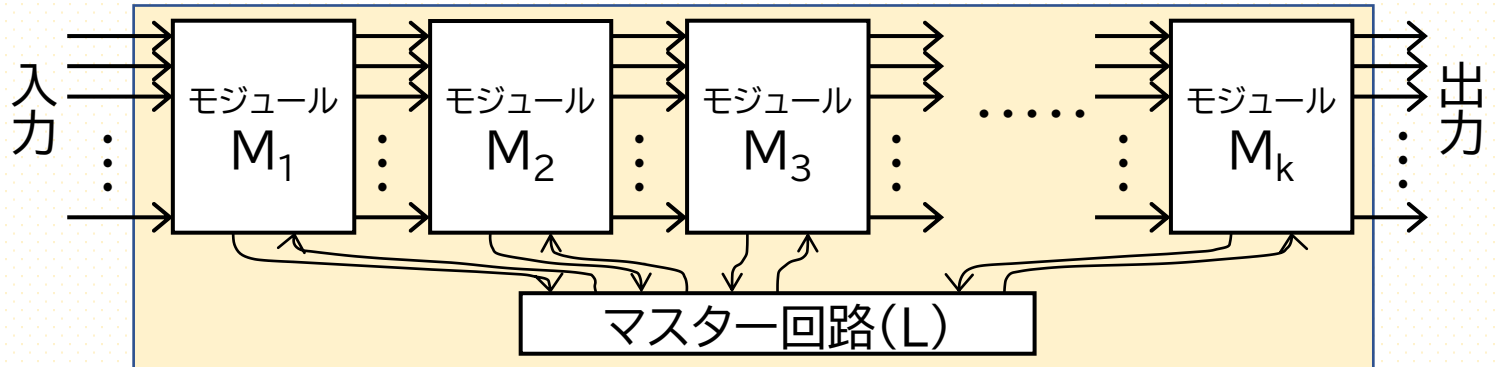


専用AIのモデル

2a:
専用AIの原型

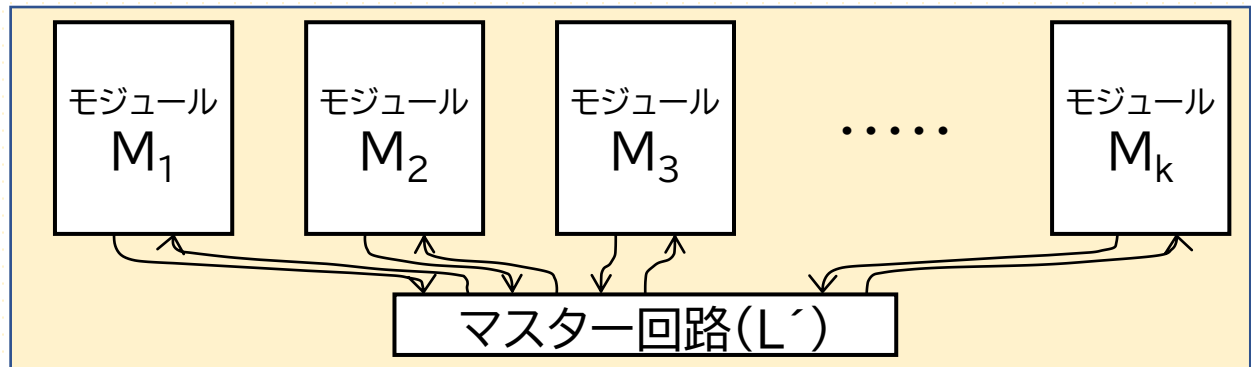


2b:
専用AIの変形

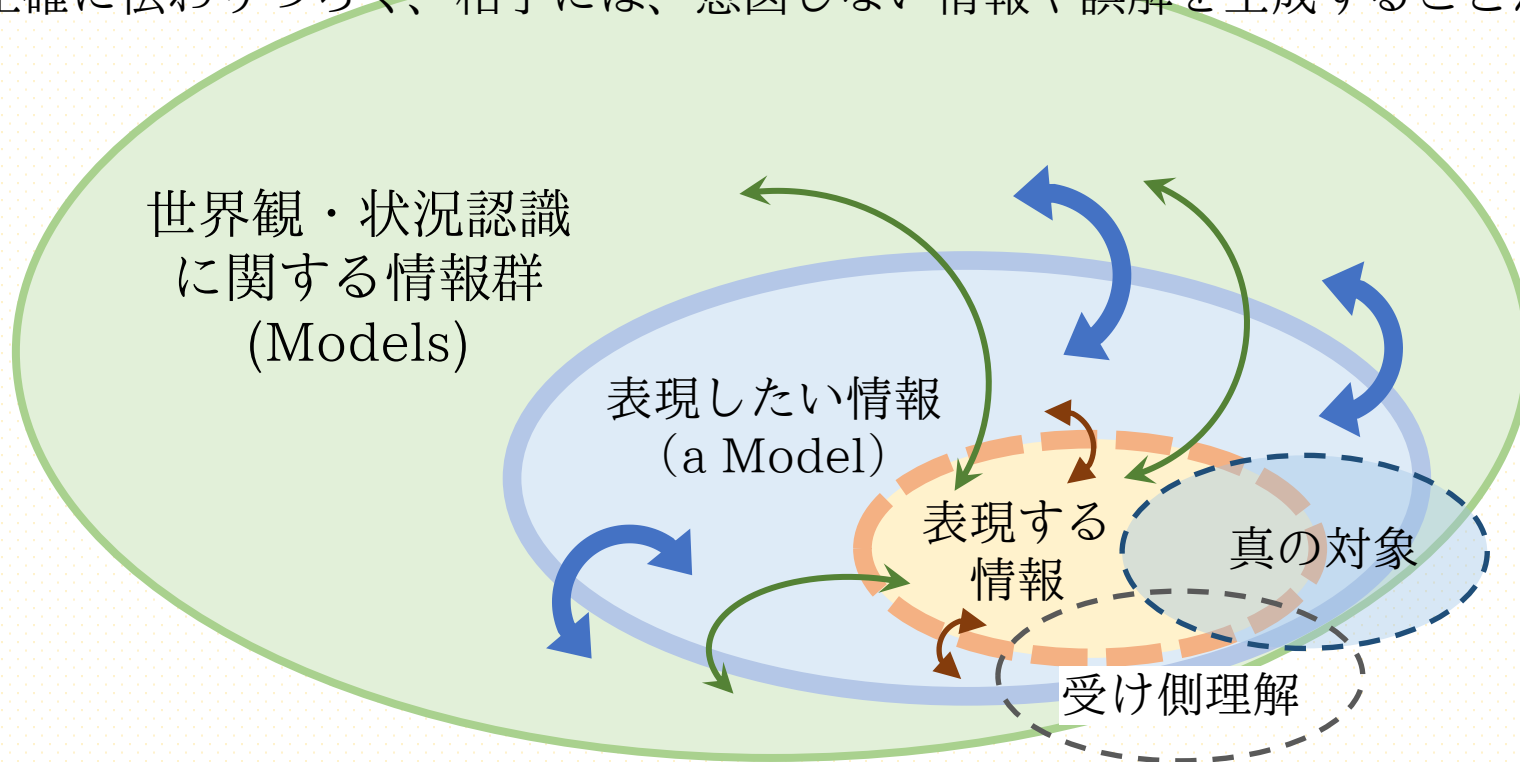


2c:
専用AIの変形

(注) 2cでは、入出力を省略



表現は、表現したい内容ができるだけ正確に相手に伝わるように計画されているが、伝わるには、「言語」が共有されている必要がある。
しかし、「言語」は、正確に共有されていることは稀であるため、実際には、正確に伝わりづらく、相手には、意図しない情報や誤解を生成することがある。



チューリング・マシンのタイプ

- 否定派 : 多くは、DTM、NTM、MPTMをベースとしている。
 肯定派 : 未だ存在していない、より高度なマシンを想定している。
 . . . MIP*実現迄待つ必要あるか? MIPやIP(AM) で実装可能か?

計算複雑性 (計算量) 大 ⇐

計算時間	想定される仮想コンピュータのタイプ	
二重の指数関数時間 (NEEXP) Doubly Exponential Time	MIP *	複数の量子コンピュータ間で量子もつれ現象が可能とした「複数証明者」を持つ対話型証明システム (Interactive Proof system of using Multiple Quantum Computers and shared entangled qubits.)
指数関数時間 (NEXP) Exponential Time	MIP	複数の証明者と検証者からなる対話型証明システム (Multi-prover Arthur-Maline Interactive Proof System)
多項式時間 (P) Polynomial time	IP (AM)	証明者と検証者からなる対話型証明システム (Arthur-Maline Interactive Proof System)
	NPTM	非決定性確率的動作のチューリングマシン (Non-deterministic & Probabilistic Turing Machine)
	NTM	非決定性チューリングマシン (Non-deterministic Turing Machine)
	DTM	決定性チューリングマシン (Deterministic Turing Machine)