



HOKKAIDO  
UNIVERSITY

# 講義「人工知能」 強化学習 1 エージェント

北海道大学大学院情報科学研究院  
情報理工学部門 複合情報工学分野  
調和系工学研究室 准教授 山下倫央

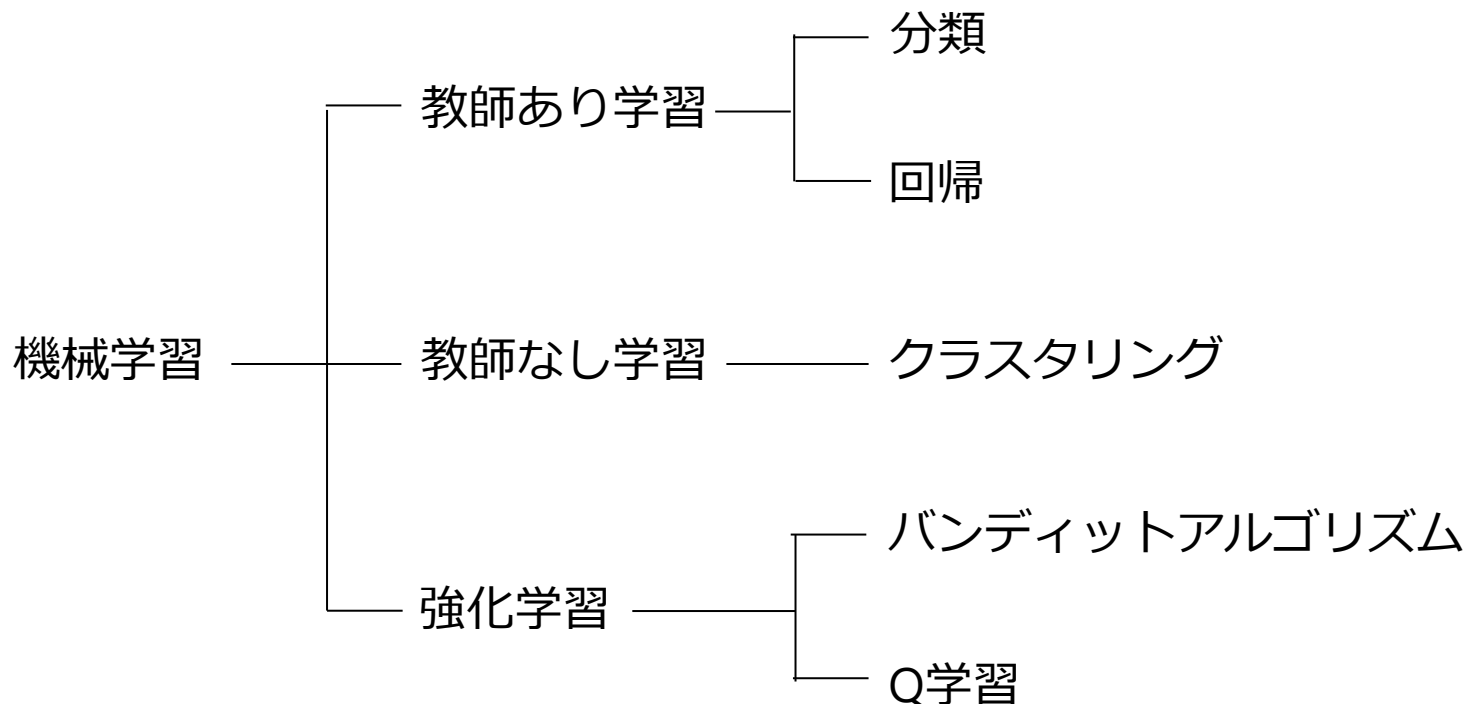
<http://harmo-lab.jp>  
[tomohisa@ist.hokudai.ac.jp](mailto:tomohisa@ist.hokudai.ac.jp)

2024年4月30日(火)

## ❖ 機械学習とは

- 人間が自然に行っている学習能力と同様の機能をコンピュータで実現しようとする技術・手法のことである。

(<https://ja.wikipedia.org/wiki/機会学習>)



## ❖ 教師あり学習

- 入力とそれに対応すべき出力を写像する関数を生成する。人間の専門家が訓練例にラベル付けすることで提供されることが多いのでラベルとも呼ばれる

## ❖ 教師なし学習

- 入力のみ（ラベルなしの例）からモデルを構築する

## ❖ 強化学習

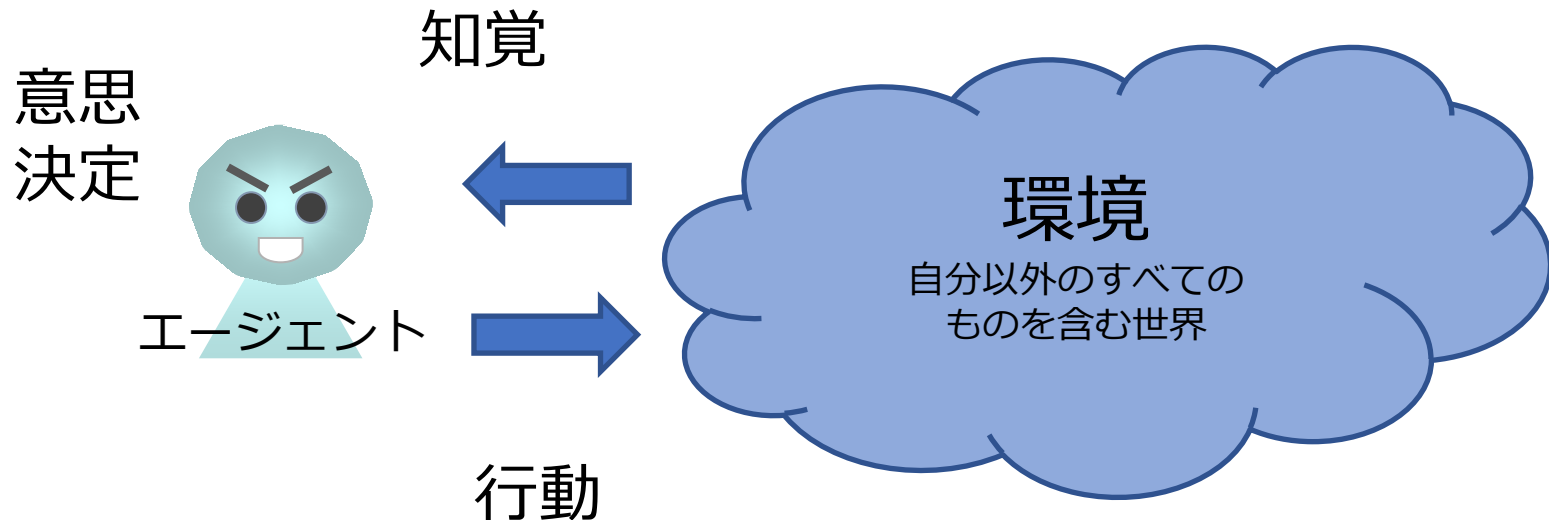
- 計算主体であるエージェントが周囲の環境を観測することでどう行動すべきかを学習する。環境に影響を及ぼす行動を取った結果、報酬を獲得する
- この報酬に基づいて、**エージェント**は数多くのエピソードを経験して試行錯誤的に期待報酬を多くするように学習を進める

## ❖ エージェント(Agent)

- 環境の状態を知覚し，行動を行うことによって，環境に対して影響を与えることが出来る自律的主体

## ❖ 環境(Environment)

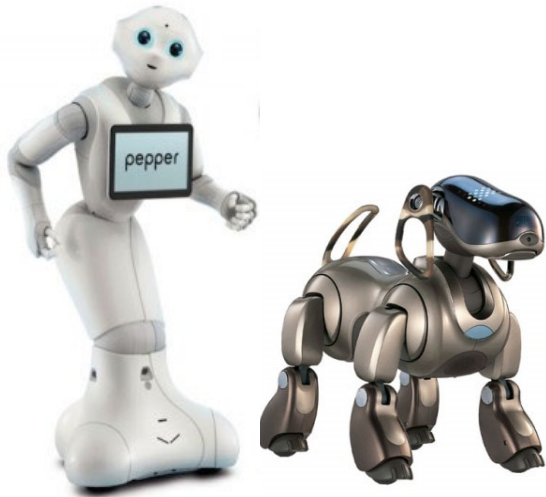
- エージェントの外部にあって，エージェントの意思によって直接変更できないものすべて



## ❖ ロボットエージェント

知覚

- ・ CCDカメラ, 赤外線センサ, 超音波センサ, ...

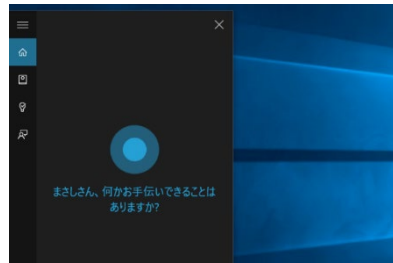


環境

行動

- ・ アクチュエータの制御

## ❖ ソフトウェアエージェント



知覚

- ・ ユーザの入力
- ・ ソフトウェアの動作状態, ...

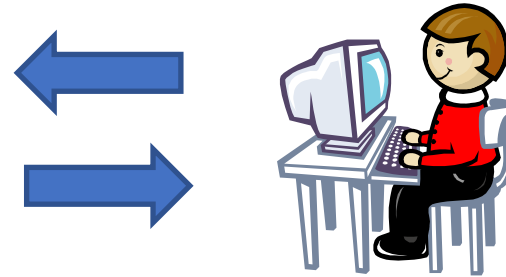
環境



siri



Google  
アシスタント



行動

- ・ アドバイスの表示
- ・ 適切なソフトウェアの起動, ...

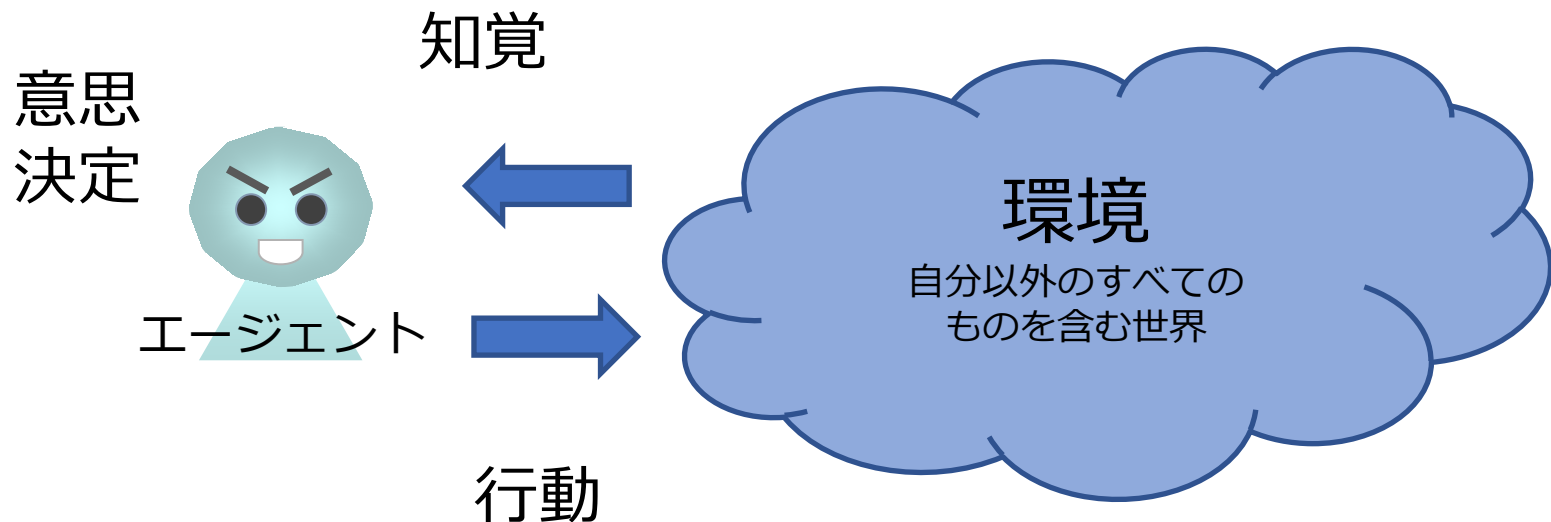
1. アクセス可能/アクセス不可能
2. 決定的/非決定的
3. エピソード的/非エピソード的
4. 動的/静的
5. 離散的/連続的

## ❖ アクセス可能

- エージェントが環境の最新状態を完全に正確に把握可能

## ❖ アクセス不可能

- エージェントが環境の最新状態を完全に正確に把握不可能





## ❖ アクセス可能な環境

- コンピュータゲームのエージェント
- シミュレーション中のエージェント

## ❖ アクセス不可能な環境

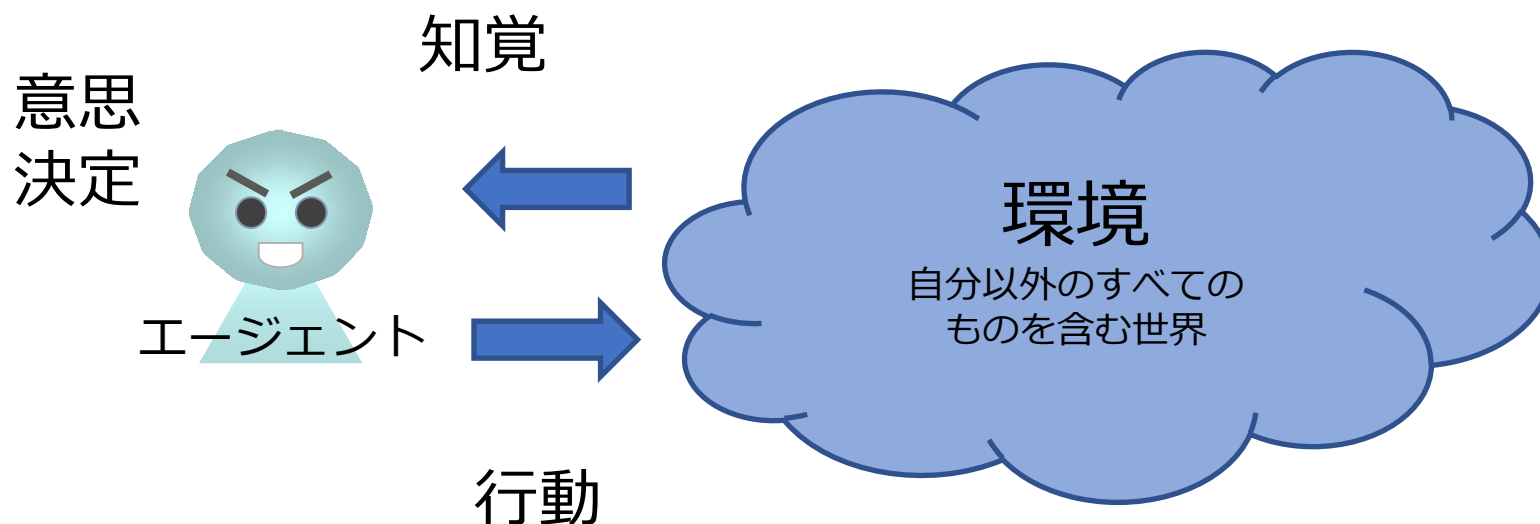
- お掃除ロボット

### ❖ 決定的な環境

- ある環境状態でエージェントが同じ行動を起こすとそれに応じた環境の変化が一意に定まる

### ❖ 非決定的な環境

- 自らの行動に対して確率的に次の状態が定まる



## 2. 決定的・非決定的な環境の例

11

### ❖ 決定的な環境

- チェス
- 将棋

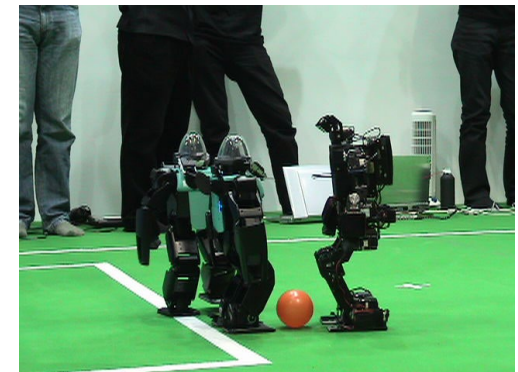
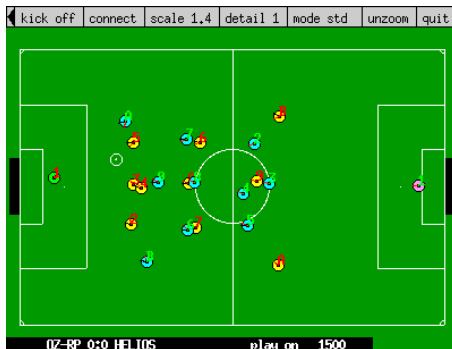
### ❖ 非決定的な環境

- 人生ゲーム
- ロボカップサッカー

## ❖ ロボカップ (RoboCup)

■ <http://www.robocup.or.jp/>

- ロボットと人工知能の新しい標準問題「2050年、人型ロボットでワールドカップ・チャンピオンに勝つ」を設定し、その研究過程で生まれる科学技術を世界に還元する事を目標としている
- 即応性・目的志向性・社交性を持ったエージェントを研究開発するためのテストベッドとしてとらえられる

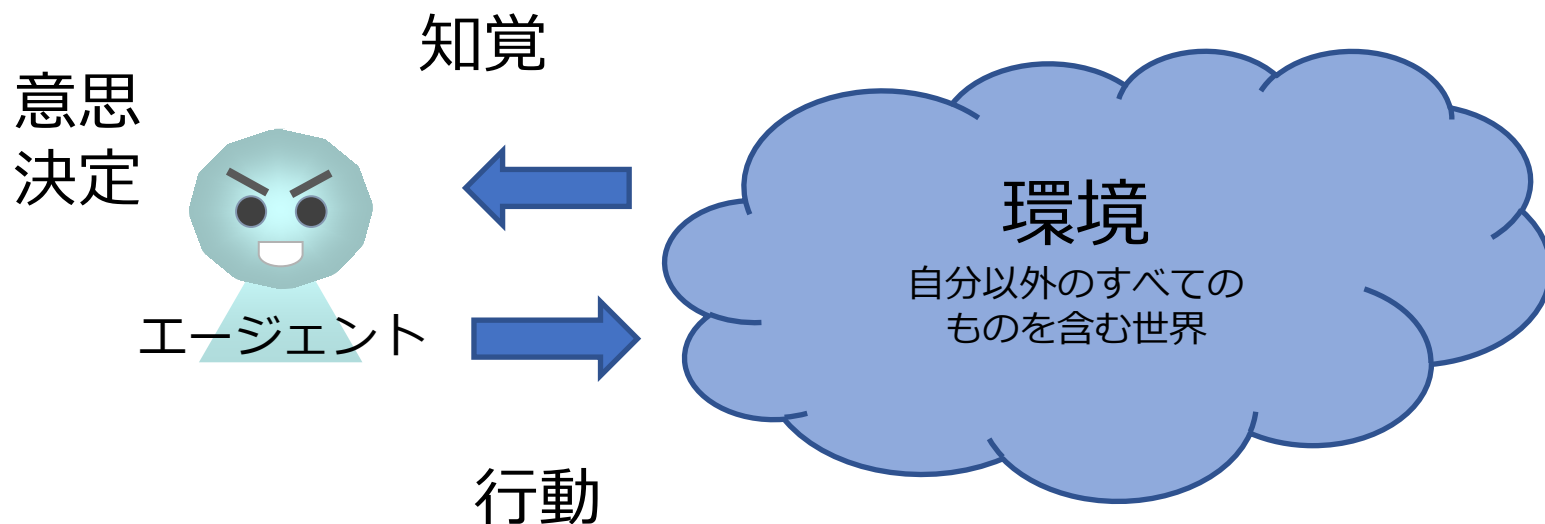


#### ❖ エピソード的な環境

- エージェントの行動がエピソードと呼ばれるある種の単位に分割できる場合，開始状態と終了状態が明確なタスク達成型の環境

#### ❖ 非エピソード的な環境

- 分割できない環境



## ❖ エピソード的な環境

- 将棋やチェス
- 人生ゲーム

## ❖ 非エピソード的な環境

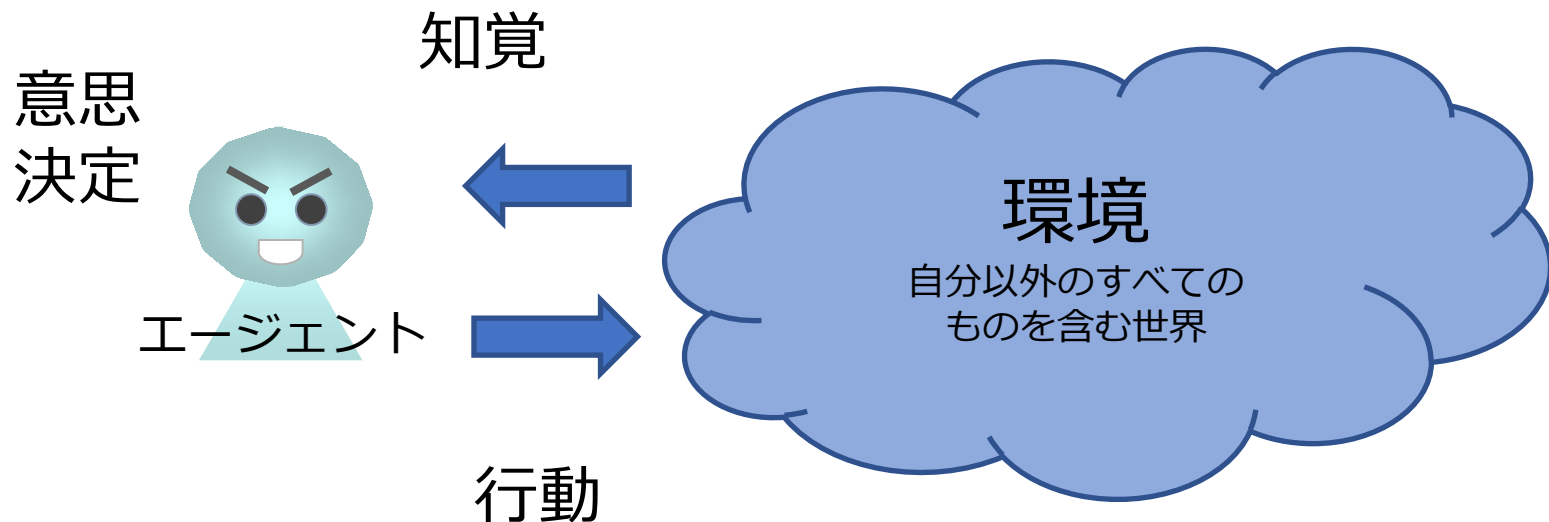
- ロボカップサッカー
- 火星探査

### ❖ 動的な環境

- エージェントの行動以外の何らかの作用によって環境が変化

### ❖ 静的な環境

- エージェントの行動のみに基づいて環境が変化



### ❖ 動的な環境

- 掃除機ロボット

### ❖ 静的な環境

- 将棋やチェス

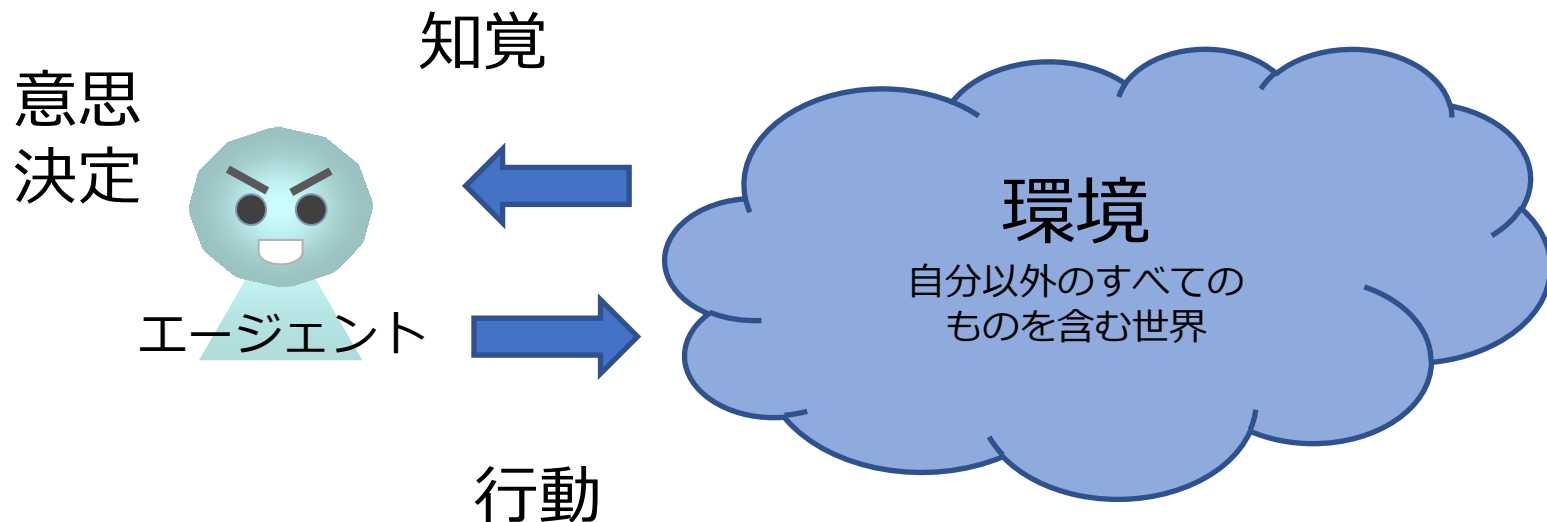


## ❖ 離散的な環境

- エージェントの知覚と行動を明確に有限個に区別できる

## ❖ 連続的な環境

- エージェントの知覚と行動を明確に明確に区別できない



### ❖ 離散的な環境

- チェス, 将棋, トランプ

### ❖ 連続的な環境

- 実環境で稼働するロボット

- ❖ 下記の特性を備えるエージェントを「知的エージェント」と呼ぶ
- ❖ 即応性
  - 環境の変化に対して即座に反応して行動をすることが可能
- ❖ 目的志向性
  - 目的に向かって積極的な行動をすることが可能
- ❖ 社交性
  - 他のエージェントと通信を行うことが可能

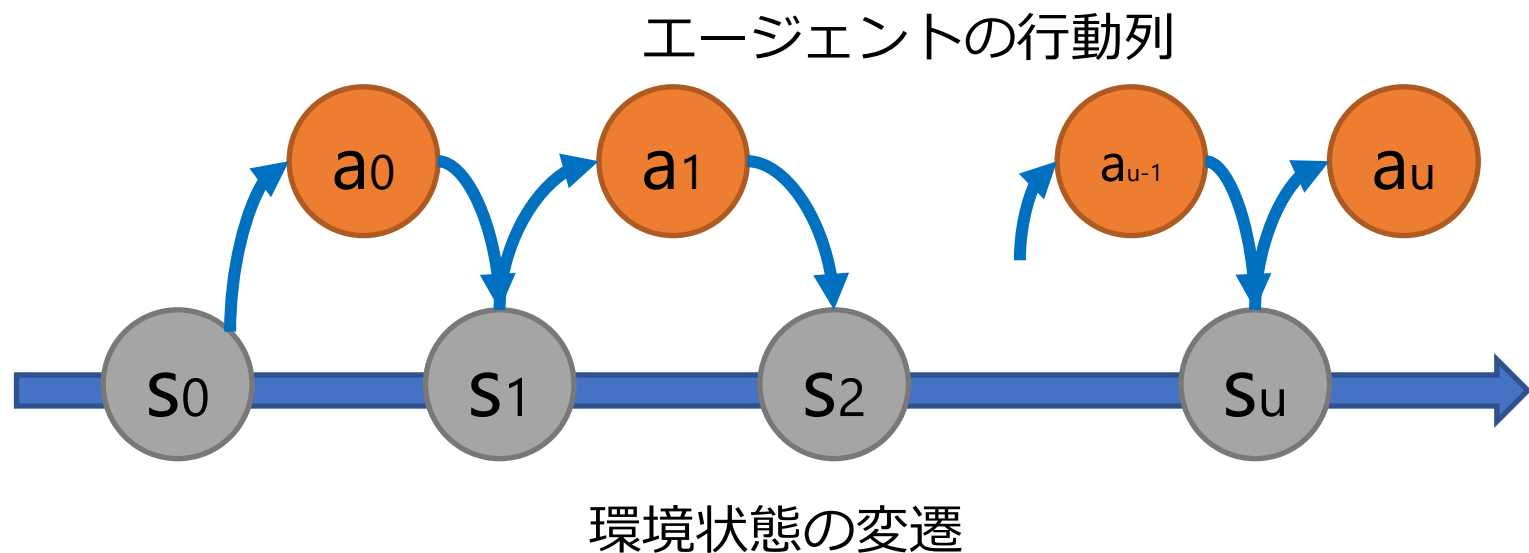
## ❖ 設計対象

- 行動関数と環境関数
- エージェントの知覚
- エージェントの内部状態

## ❖ 環境の場合分け

- 1. アクセス可能
- 2. アクセス不可能な場合
  - 2-1. アクセス不可能で内部状態を持たない場合
  - 2-2. アクセス不可能で内部状態を持つ場合

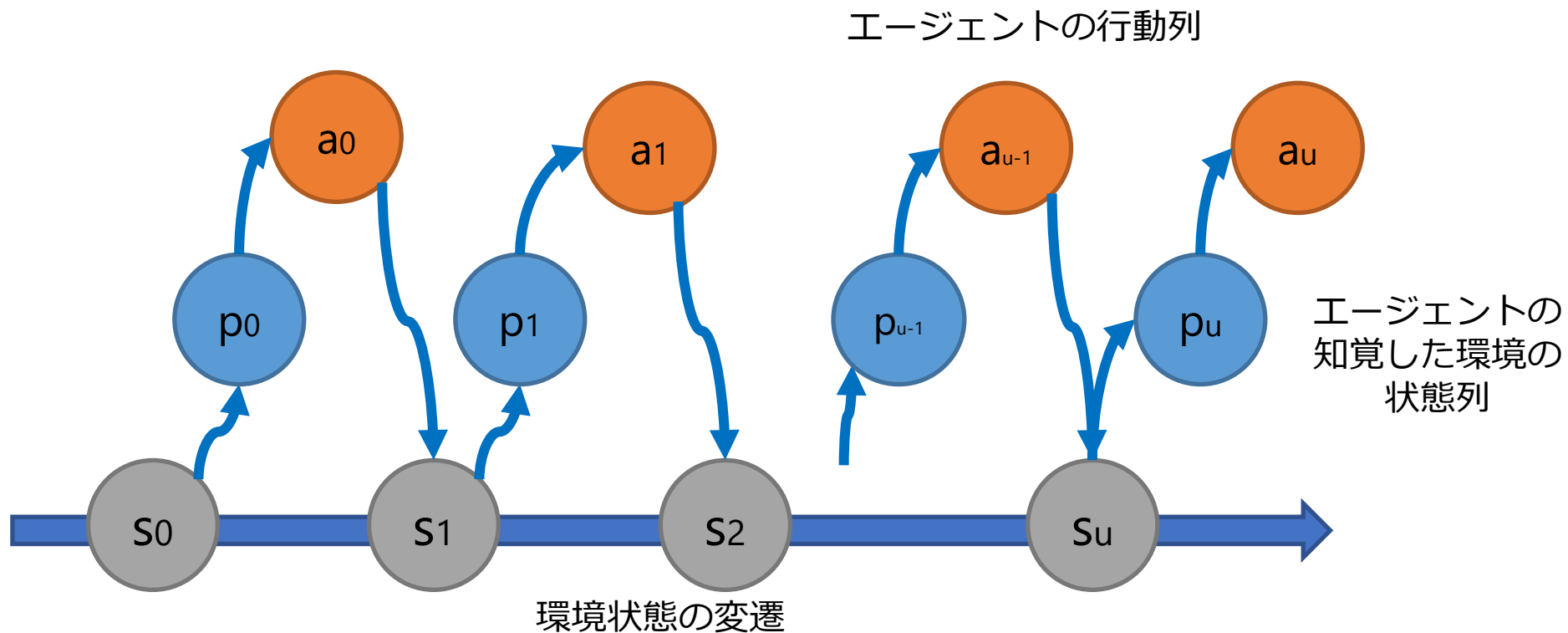
- ❖ 環境の状態集合： $S=\{s1,s2,...\}$
- ❖ 環境の状態列すべての集合： $S^*$
- ❖ 状態  $S$  のべき集合： $P(S)$
- ❖ エージェントの行動集合： $A=\{a1,a2,..\}$
  
- ❖ 環境関数  $env: S^* \times A \rightarrow P(S)$
- ❖ 行動関数  $action: S^* \rightarrow A$



- ❖ 環境の状態集合 :  $S = \{s_1, s_2, \dots\}$
- ❖ 環境の状態列すべての集合 :  $S^*$
- ❖ 状態  $S$  のべき集合 :  $P(S)$
- ❖ エージェントの行動集合 :  $A = \{a_1, a_2, \dots\}$
- ❖ エージェントが知覚可能な環境集合 :  $P = \{p_1, p_2, \dots\}$
- ❖ 知覚可能な環境の状態列すべての集合 :  $P^*$
  
- ❖ 環境関数  $\text{env}: S^* \times A \rightarrow P(S)$
- ❖ 知覚関数  $\text{see}: S \rightarrow P$
- ❖ 行動関数  $\text{action}: P^* \rightarrow A$

## 行動関数と環境関数 (2-1. アクセス不可能・内部状態無し)

- ❖ 環境関数  $\text{env}: S^* \times A \rightarrow P(S)$
- ❖ 知覚関数  $\text{see}: S \rightarrow P$
- ❖ 行動関数  $\text{action}: P^* \rightarrow A$



# Perceptual aliasing Problem

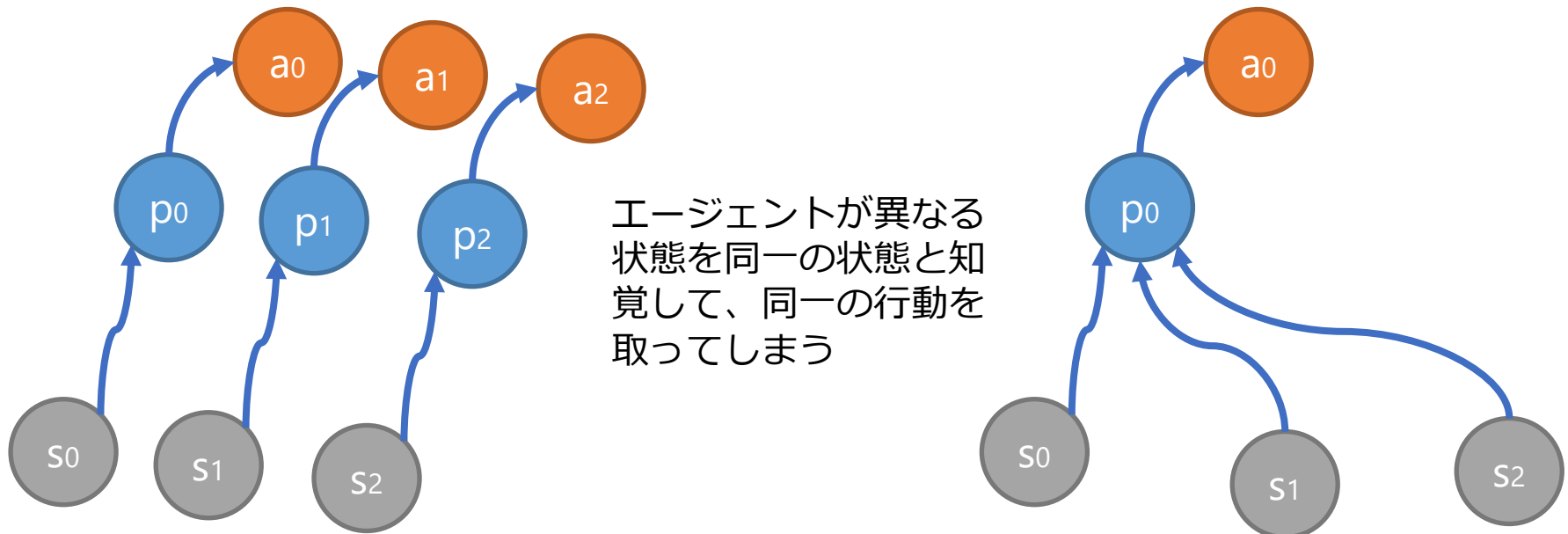
## ❖ 異なる状態を同一の状態とみなしてしまう

### ■ 知覚の見せかけ

- すべての状態を区別したいが、区別できない場合に発生
- 複数の状態に対して同じ行動しか決定できない

### ■ 知覚関数 $see: S \rightarrow P$ に起因

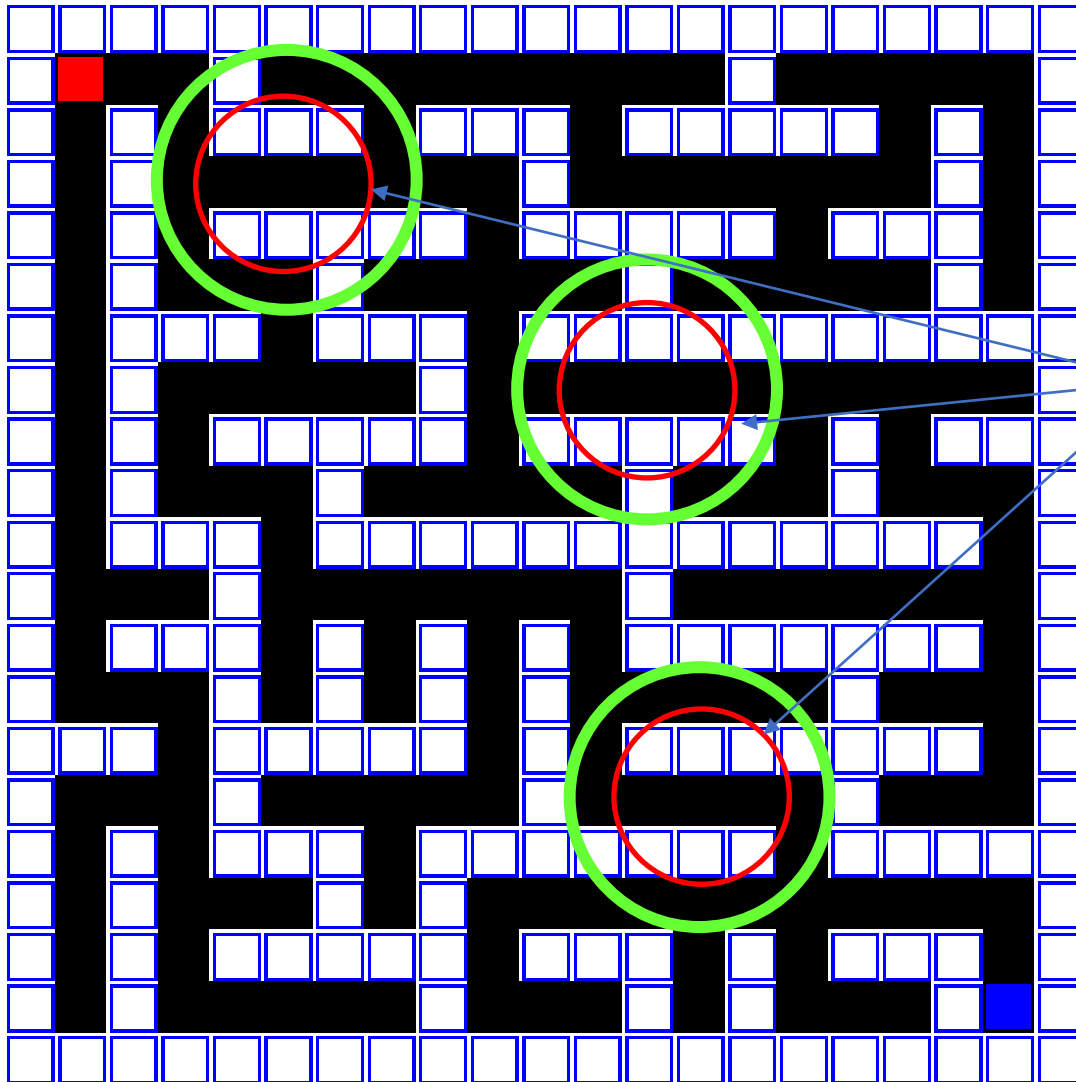
- センサの精度・メモリ容量等が原因





# Perceptual aliasing Problem の例

## ❖ 迷路における状態認識

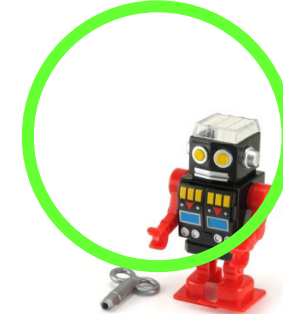


低性能ロボットの視界



低性能のロボットは異なる状態を  
同一の状態と知覚してしまう

高性能ロボットの視界



## 行動関数と環境関数 (2-2. アクセス不可能・内部状態有り)

- ❖ 環境の状態集合 :  $S = \{s_1, s_2, \dots\}$
- ❖ 環境の状態列すべての集合 :  $S^*$
- ❖ 状態  $S$  のべき集合 :  $P(S)$
- ❖ エージェントの行動集合 :  $A = \{a_1, a_2, \dots\}$
- ❖ エージェントが知覚可能な環境集合 :  $P = \{p_1, p_2, \dots\}$
- ❖ エージェントの内部状態の集合 :  $I = \{i_1, i_2, \dots\}$
  
- ❖ 環境関数  $\text{env}: S^* \times A \rightarrow P(S)$
- ❖ 知覚関数  $\text{see}: S \rightarrow P$
- ❖ 内部状態関数  $\text{next}: I \times P \rightarrow I$
- ❖ 行動関数  $\text{action}: I \rightarrow A$

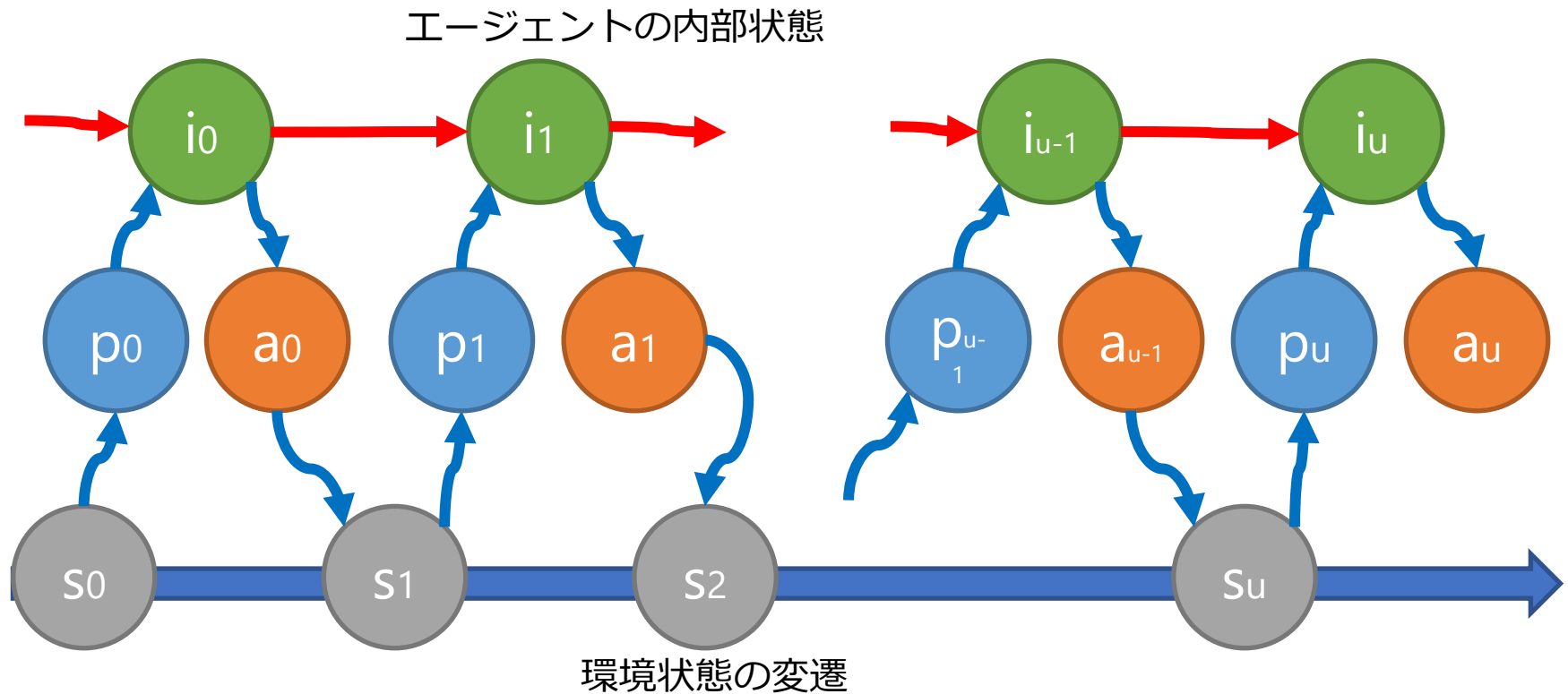
## 行動関数と環境関数 (2-2. アクセス不可能・内部状態有り)

❖ 環境関数  $\text{env}: S^* \times A \rightarrow P(S)$

❖ 知覚関数  $\text{see}: S \rightarrow P$

• 内部状態関数  $\text{next}: I \times P \rightarrow I$

• 行動関数  $\text{action}: I \rightarrow A$



# マルチエージェントシステム

- ❖ 対象となる環境中に多くの自律的なエージェントが存在し、環境とエージェント間の相互作用を通して複雑な様相が現れる系
- ❖ 創発
  - 相互作用を通して現れる複雑な様相



- ❖ ゲーム理論
  - マルチエージェントシステムの分析ツール
    - 利害が対立するプレイヤー(=エージェント)の合理的行動を表現する数理モデルを発見し、数学原理から合理的行動の一般的特性を導出
  - ミクロ経済学で利用