F-028

# 複数種類の報酬と罰に対応した意識的意思決定システムの提案
# Proposal of a Decision Making System Based on Consciousness in Multiple Rewards and Penalties Environments

宮崎和光 †
Kazuteru Miyazaki

## 1. はじめに

意識とは何だろう？意識の定義は種々であり [31, 26, 16] 「意識とは、種々の概念が混在したスーツケースワードである」[14] との見方もある．意識の研究には, ヒトの脳の機能の解明という側面 [25, 4, 5, 30, 10, 17, 12, 7, 29, 15] と, その機械 (コンピュータ上) への実装という側面 [8, 9, 33, 13, 3, 34, 35, 22, 24] が存在する．本稿では, 工学の観点から, 後者に注目するが, その実現には, 前者で得られた知見を最大限に活用する必要があるのは言うまでもない．一方, 逆に, 意識の機械による実現が, 脳における意識機能の解明に寄与する可能性も十分考えられる．これにより, 主観的と思われがちな「意識」を客観的に研究することが可能となる．

意識の機械による実現には様々なハードルが存在していると思われる．その中でも本稿では, 文献 [22] 同様, 「意思決定主体としての意識」を研究の対象とする．著者らは, これまでに, 文献 [24] において, 外界からの刺激に反応的に行動を行う系 (1 次系) に対し，1 次系を監視する系 (2 次系) を付加することの有効性を数値実験により確認した．そこでは, 1 次系は, 経験強化型学習 [21] (または強化学習 [32]) により学習を行い, 2 次系は, 1 次系の学習の成否を評価し, 学習がうまくいっていないと判断される場合には, 1 次系に対し再学習を促す役割を担っていた．

一方, 近年, ノーベル経済学賞受賞者である Daniel Kahneman は, 著書「ファスト&スロー」[6] の中で, 思考を「自動的」と「制御的」の 2 つに分ける考え方を示している．そこでは, 自動的な思考は, 「熟考や計画性を伴わずに素早く効率的に行われる意識的認識の外にあるもの」, 制御的な思考は, 「比較的ゆっくり行われる意図的・意識的なもの」とし, 2 種類の思考を区別している．また, 脳科学の観点からは, 甘利らは, 著書「精神の脳科学」[1] の中で, 記憶を「情動記憶」と「陳述記憶」に分け, それらの間の力動関係を,「日常の平穏ではあるが複雑な社会的・言語的環境においては陳述記憶の果たす役割が大きいが, いざ危急の事態が生じた際には情動記憶が迅速に反応して環境の変化に対応している」(「精神の脳科学」p.211 より引用）と述べている．甘利らの「情動記憶」が Kahneman の「自動的思考」に対応し,「陳述記憶」が「制御的思考」に対応するものと考えられる．

さらに, 甘利らは, これら 2 種類の記憶 (Kahneman の言葉では「思考」) の切り換え (力動性) は, 「扁桃体において生じているのかもしれない」(同 p.211) とも述べている．そのような扁桃体の機能については, 小野武年らの「知・情・意の神経機構」[27] (および文献 [28]) や Joseph LeDoux の「シナプスが人格をつくる」[11](p.319-p.325) においても詳しく述べられている．これらの知見は, 最終的な意思決定部位としての扁桃体の重要性を示唆するものであると考える．

以下, 本稿では, 上に述べたような情動記憶 (自動的思考) に対応する「素早い学習を行う部分」および陳述記憶 (制御的思考) に対応する「時間のかかる学習を行う部分」と「それらの間の調停を行う部分」の 3 要素を有するシステムを**意識的意思決定システム**と呼ぶ．著者らは, これまでに既に文献 [20] において, 意識的意思決定システムの 1 種と捉えることができる **MarcoPolo** を提案している．そこでは, マルコフ決定過程を対象にした強化学習問題において, 素早い学習を行う Profit Sharing[18] と環境の同定を行う k-確実探査法 [19] とを特別に設計された調停器によって切り替えることを提案している．

しかし, MarcoPolo は, マルコフ決定過程に限定された設計がなされており汎用性が乏しい．また, 学習のための教師信号である報酬に関しても, 最も単純な 1 種類の報酬のみが存在する場合を想定している．そこで, 本稿では, 複数種類の報酬と罰が存在するより一般的な環境における意識的意思決定システムを提案し,数値実験によりその有効性を確認する．

## 2. 問題設定

未知環境下に置かれたロボットのような学習器 (エージェント) を考える．エージェントには, 環境の状態を知覚するための**感覚入力**および環境に働きかけ環境の状態遷移の原因となり得る**行動出力**が備えられている．なお, 環境には, エージェントの外部という意味の他に, エージェントの内部にエージェント自身が生成する**内部状態**の意味を含めることもできる．

1 章で述べたように本稿では「意思決定主体としての意識」(意識的意思決定システム) を研究の対象とする．そこで取り扱われる問題は, 各感覚入力に対し, 選択すべき行動出力を決定する問題として定式化できる．そのための教師信号として, 本稿では, **報酬**または**罰**の存在を仮定する．ここで, 報酬は, 目標となる感覚入力に遷移した場合に与えられ, 罰は, 遷移すべきでない感覚入力に遷移した場合に与えられる．エージェントの目的は, 罰を回避し, 報酬を得続けることにある．

時間は認識-行動サイクルを 1 単位として離散化され, 感覚入力は離散的な属性の種類ごとに, ユークリッド空間上の連続値として与えられるものとする．離散的な属性の種類を**次元数**と呼ぶ. 例えば, 視覚センサーと聴覚センサーをそれぞれ 1 個ずつ持つエージェントの次元数は 2 である．

通常, 計算機上では, 連続値で入力された情報は, 何

†独立行政法人大学評価・学位授与機構, NIAD-UE



Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

らかの形で離散化が施される．離散化された入力を**状態**と呼び, 状態の種類のことを**状態数**と呼ぶ. 行動は離散的なバリエーションの中から選ばれる．各状態に対し, 選択すべき行動を与える関数を**政策**と呼び, 状態-行動ペアを**ルール**と呼ぶ.

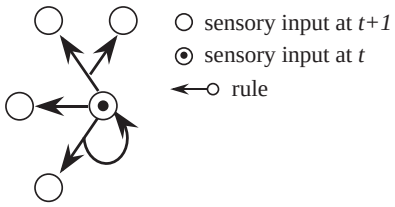


図 1: 環境の状態遷移の一例

環境は, 感覚入力を状態, 行動を状態遷移オペレータとする確率過程とみなすことができる．環境の状態遷移の一例を図 1 に示す．トークン付きのノードが現在の時刻 (時刻 t) の感覚入力を表す. 図 1 では, 時刻 t での感覚入力に対し, 3 種類の行動が選択可能, すなわち 3 種類のルールが選択可能となっている. 状態遷移は確率的なので, 同じルールを選択したとしても必ずしもつねに同じ感覚入力に遷移するとは限らない．矢印の枝分かれが, そのような状態遷移を意味している.

エージェントは, 環境の状態遷移に関する完全なる事前知識は有していないものとする. そのため, 環境との相互作用 (試行錯誤) を通じて, 政策の学習を進める必要がある．そのような「環境との相互作用に基づく目的指向の学習」は, 現在, **強化学習** (Reinforcement Learning)[32] および**経験強化型学習** (Exploitation-oriented Learning ; XoL)[21] において, 集中的に研究されている. 強化学習, XoL ともに, それぞれ対象とする環境のクラスを仮定することで, 報酬を得続けることが可能な政策の学習を保証することができる.

## 3. 意識的意思決定システムの全体構成

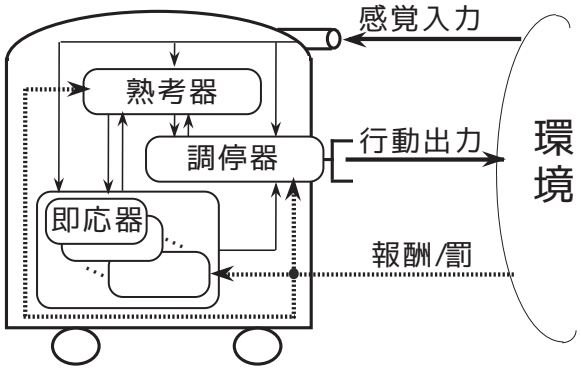


図 2: 意識的意思決定システム

図 2 に本稿で提案する意識的意思決定システムの全体構成図を示す．ここでは, 複数種類の**即応器**と単一の**熟考器**, および最終的に出力する行動を決定する部分としての**調停器**が存在する.

以下では, これらの各構成要素を**モジュール**と呼び, その定義と概要を述べる. なお, MarcoPolo では, 即応器として Profit Sharing, 熟考器として k-確実探査法が採用された手法であると考えることができる.

### 3.1. 即応器

素早い学習を行う構成要素 (モジュール) を**即応器**と定義する．文献 [24] では, 即応器に相当するものとして, **1 次系**を想定している．そこでは, 1 次系は, (環境からの) 入力に反応して, (環境へ何らかの) 出力を行う系であると定義されていた．したがって, 1 次系は, 環境との入出力間の対応付け, すなわち, 2 章で定義した用語に従えば, 政策を学習する系として定義されていた.

文献 [24] では, 1 次系の学習に経験強化型学習 (または強化学習) を用いていたが, より一般的には, 学習の必要がないヒトに本能的に備わっていると考えられる「情動系」も即応器として捉えることができる．また, 条件反射に代表される反射的な行動も即応器と考えることもできる. そこで, 本稿では, これら種類の異なる意思決定方法を, それぞれ独立に即応器として記述することを提案する．以上より, 各即応器は甘利の「情動記憶」(Kahneman の「自動的思考」) に相当するモジュールであると考えることができる.

なお, 各即応器間の関係性を定義すれば, Brooks の包摂アーキテクチャ(Subsumption Architecure)[2] と類似したシステムを構成することができる. 一般に, 包摂アーキテクチャでは, 各モジュール間に階層を仮定し, 上位層に行くに従ってより抽象的なモジュールを構成する．また, 各層の目的は下位層の目的を包含している必要がある．しかし, 本稿では, そのような階層構造は仮定せず, あくまで, 各即応器は独立した存在として捉え, そらの間の競合解消は熟考器および調停器に任せる立場を取る.

### 3.2. 熟考器

即応器には, 情動, 反射, 政策の学習といった異なる種類のモジュールが含まれる. それら全体を統括, あるいは時間のかかる独自の計算を行う構成要素 (モジュール) を**熟考器**と定義する．すなわち, 高度な計算により, 即応器よりも適切な行動選択を実現することが熟考器の役割である.

熟考器は, ヒトの大脳新皮質に存在すると考えられているワーキングメモリに対応するものである．そのため, より下等な動物には, 熟考器に相当するものは存在せずに, 即応器のみで最終的な意思決定が行われている可能性もある. それに対し, ヒトでは膨大な新皮質領域を利用して即応器の内容を比較検討することで, より高度な意思決定を実現しているものと考えられる.

その他, 熟考器は, 即応器とは独立に独自の計算を行うことも許される. 以上より, 熟考器は「陳述記憶」(制御的思考) に相当するモジュールであると考えることができる.

### 3.3. 調停器

即応器および熟考器は, それぞれ行動の候補を提示するが, 一般に, 環境に対し出力できる行動は 1 種類である. 実際にどのモジュール (即応器または熟考器) の行動を環境に対し出力するかを決定するための構成要素



Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

(モジュール) を**調停器**と定義する.

文献 [24] では, 2 次系が, 調停器の役割を担っていた. そこでは, 2 次系は, 環境との間に入出力関係を持たず, 1 次系とのみ相互作用する系であると定義され, 1 次系の学習が安定 (収束) した後に, 何らかの不測の事態が生じた際に, 1 次系が獲得したことのすべてをリセットすることなしに, 不測の事態に対応可能となるために必要な系として 2 次系を捉えていた.

これは単純な 1 種類の 1 次系に対しては有効であると考えられ, 実際, 文献 [24] においても数値実験によりそれが確認されている. しかし, 本稿で扱うような複数種類の即応器が存在する場合に対しては, そのままでは適用できない. むしろ, 意思決定の詳細は, 即応器および熟考器で完了しており, 調停器は, その結果の評価, すなわち, どのモジュールの出力結果を採用すべきか (環境に対し出力すべきか), に注力すべきであると考える. そのためには, 調停器には, 環境との間の入出力関係とともに, 即応器内の各モジュールおよび熟考器からの入力が必須になる. それらの情報を勘案して, 調停器では, 最終的にどのモジュールの出力を採用すべきかを決定する機能が要請される. 以上より, 調停器はヒトの扁桃体が行う機能の一部を担うモジュールであると考えることができる.

## 4. 各モジュールの構成方法

### 4.1. 即応器

即応器は, 大きく分けて「情動」「反射」「条件反射」「学習」から構成される.

このうち, 「情動」や「反射」は生命体に予め組み込まれている, 例えば, 「恐怖からの回避」や「生命維持のための恒常性維持」などに関係する生命の根幹をなす (主として, 脳幹で処理されている) 部分である. そこで, 本稿でにおいても, これらに対しては, 事前の組み込み, すなわち, タスクに応じて予め規定しておく立場をとる.

それに対し「条件反射」は, 「情動」に基づいた行動に「学習機能」を付与したものであると考える. すなわち, 何を避けるべきかを後天的に学習することに相当する.

これら以外にも即応器には政策の「学習」も含める. これは経験強化型学習や強化学習で行われるような学習が相当する. なお, 学習が関与する部分には環境から得られる報酬/罰信号が大きな影響を与える.

### 4.2. 熟考器

熟考器は, 即応器が行わない「探索」, 「膨大なデータ間の比較」, 「時間のかかる学習」などコスト (高度な計算) の要する作業を担う. これは, ある種問題依存なので, 具体的な熟考器の構成方法は, 対象問題ごとに検討する必要がある

例えば, 即応器の各モジュールの内容を比較検討する場合には, どのモジュールがどれだけの確信度で実行可能かを評価する必要がある. これは, 例えば, 環境のモデルを自己内部に生成し, その下でシミュレーションを行うことで実現できる. モデルの生成・シミュレーションには多くの計算資源を要するので, 一般に, 迅速なる対応はできない. そのため, これは, まさに熟考器向けのタスクであると言える.

また, 学習に時間の要する作業を熟考器に任せることも考えられる. ある種の組み合せ探索問題や長いステップにわたる学習などがそれに相当する.

### 4.3. 調停器

調停器では, 環境に出力する行動の最終決定が行われる. 具体的な実装方法には, 問題に応じて, 種々のものが考えられるが, 例えば, 以下のような構成を考えることができる.

> 調停器が, 各モジュールからの信号の強さ (確信度) に従って意思決定を行う場合を考える. その際, その確信度の到達速度が重要となる. これは, 同じモジュールから出される出力の時間間隔に関係する量である. 確信度とその速度を総合的に勘案して調停器ではどのモジュールを選択すべきかを決定する. 確信度と速度のどちらを重視すべきかのパラメータは学習により随時更新される.
>
> また, 各モジュールからは, 期待される遷移先の分布情報が得られるものとする. その期待分布情報と実際の遷移先との差分, および環境から得られる報酬/罰信号により調停器は先のパラメータ値を学習により調整する. この学習には経験強化型学習や強化学習が利用できる.

次章では, より単純化した調停器を考え, 数値実験により, 図 2 に示す意識的意思決定システムの有効性を検証する.

## 5. 提案手法の有効性の検証

### 5.1. 検証方法

本稿では, 図 2 に示すような「即応器」「熟考器」「調停器」といった 3 種類のモジュールからなるシステムを提案している. 各モジュールをそれぞれ機能停止することでどの程度, 性能の悪化がみられるかで提案手法を評価する. すなわち, 以下のような実験を行う.

- 即応器のみの場合
- 熟考器のみの場合
- 提案手法 (即応器+熟考器+調停器)

以下では, こらら 3 種類の設定の比較検討を行う. なお, 構成としては, 即応器と熟考器の組み合わせも可能であるが, 最終的な意思決定を行う部分が存在しない場合, 行動をひとつに絞ることができないため比較対象から除外した.

テストベットとしては, 図 3 に示す環境を用いる. エージェントは, 内部に 4 種類の欲求レベル (A,B,C,D) を有しており, ひとつ行動を出力するたびに, 予め決められた量だけ, それぞれのレベルが減少する。エージェントは自身の欲求レベルは正しく観測できるが, 各欲求レベルの減り具合を規定する関数についての事前知識は有さないものとする。なお, 図 3 には示されていないが, s1 で行動 a を選択した場合には, 確率 0.5 で s3 へ遷移するが, 残りの確率 (0.5) では, s1 と s2 の間に設



Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

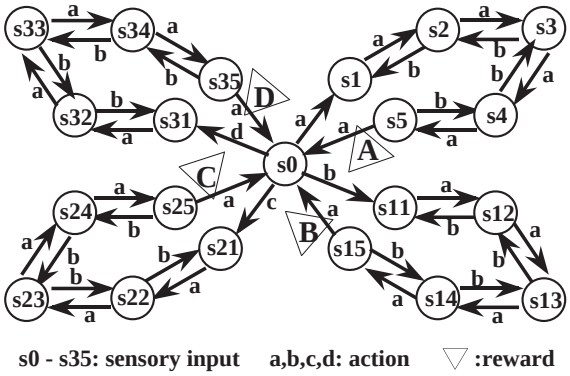


図 3: 実験に用いた環境

けられた新たな状態 s6 へ遷移する. 同様の不確実性は, s11, s21, s31 においても生じている.

エージェントが, 図 3 中の, 各目標 (図中に三角形で示した A,B,C,D のいずれか) に到達するごとに, 対応する (同じ文字で表される) 欲求レベルが, 予め決められた量だけ増加する。これが報酬に相当する。一方, 少なくともひとつの欲求レベルが 0 となった時点で, エージェントには罰が与えられ, 初期の欲求レベルにリセットされるとともに，状態 s0 に戻される。

この状況で, 状態 s0 でのエージェントの観測，すなわち, 各欲求レベルの値に基づき, いずれの目標を目指すべきかの学習問題を考える。この解は, 明らかに, 各欲求レベルの減り具合および目標到達時の各目標の増加量によって変化する。ここで，エージェントの目標は, 罰を得ずに報酬を得続けることを可能にする政策を獲得することにある.

### 5.2. 各モジュールの具体的設計方法

#### 5.2.1. 即応器

即応器としては, 欲求レベル最小の目標をつねに目指す**最小優先選択法 (MPS)**[22] を実装した.

MPS は, 多くの場合効果的な戦略となることが予想されるが, 例えば, 「A の欲求レベル＞ B の欲求レベル」となっていたとしても, 「A の欲求レベルの減り具合＞＞ B の欲求レベルの減り具合」の場合には、B ではなく A を優先して選択すべきケースが考えられる. その閾値は,A の欲求レベルの減り具合, B の欲求レベルの減り具合, A の欲求レベルと B の欲求レベルの差, 目標到達時の増加量によって変化する. 具体的にどうのような値の組み合せで MPS が不適切となるかは文献 [22] で調査済みなので, 本稿でもその知見を利用する.

#### 5.2.2. 熟考器

本稿では, 文献 [23] で提案されている**「回避リスト」**の学習を熟考器として採用する. すなわち, 罰による学習をより重視する立場から，罰により構成される,「回避リスト」，すなわち,「ある欲求レベルの組み合わせのときに選択すべきでない下位層」 を記憶する方法を用いる.

具体的には，罰を得たときに，罰を得る直前の選択‡での「欲求レベルが 0 になった（罰を得た）目標の欲求レベル」,「その時点で目指していた目標の欲求レベル」，および,「その時点で目指していた目標の種類」を「回避リスト」に登録する。この際，既に記憶している「回避リスト」に現在登録しようとしている「回避リスト」が含まれる場合は，重複しての登録は行わない。一方，既に記憶している「回避リスト」よりもより広い「回避リスト」が得られたならば,「回避リスト」の更新を行う。これは例えば,「その時点で目指していた目標の欲求レベル」を $a$,「欲求レベルが 0 になった（罰を得た）目標の欲求レベル」を $b$ としたとき,「回避リスト」内に $A > a$ なる $A$ および $B < b$ なる $B$ がそれぞれ相当する目標に存在するならば，その「回避リスト」の $A$ を $a$ に，$B$ を $b$ に置き換えることで実現される。

なお，欲求レベルが単調減少関数以外の関数に従い変化する場合には，それぞれの関数に応じて「回避リスト」の登録・更新手続きに修正を加えればよい。また，ここでは，必要最小限の条件である「欲求レベルが 0 になった（罰を得た）目標」と「その時点で目指していた目標」との関係のみを「回避リスト」に記憶させた。より一般的なすべての目標を考慮した「回避リスト」の有効性については今後の課題である。

#### 5.2.3. 調停器

本問題設定においては, 多くの場合 MPS が有効に機能することが予想できる. そこで, 調停器では, 不都合が生じない限り,「即応器」の出力をそのまま環境に出力する.

一方, ある環境で,「即応器」の出力に従って行動を行なった結果, 罰を得たならば, 同じ環境では次からは「熟考器」の出力を採用する.

なお, 同様の切り換えは, 文献 [23] でも行われているが, そこでは,「調停器」のような独立したモジュールは存在しないため, 環境の種類ごとに「即応器」と「熟考器」を切り換えるようなことはできない.

### 5.3. 実験方法

この問題に対し, 予め, 下位レベルの学習として, 各目標への到達方法 (最短経路) は学習できているものとする. すなわち, 各状態 (s0〜s35) から, 各目標 (A,B,C,D) への (最短の) 到達方法は既知である.

各欲求レベルの初期値はすべて 100 とした. その上で,「エージェントがひとつ行動を出力するごとに減らされる各欲求レベルの量」および「目標達成時の報酬の量」すなわち,「各目標に到達したときの到達した欲求レベルの増加量」を変化させる実験を行う. 具体的には報酬または罰の獲得回数が 30 万回に到達するごとに, 表 1 に示す環境 0,1,2 を順番に繰り返す. ここで, 文献 [22] の知見より, 環境 1 および環境 2 では MPS が有効であるが, 環境 0 では MPS では罰の完全なる回避は不可能であることがわかっている.

‡この登録に際しては，文献 [23] 同様,「目指してた目標」と「罰を得た目標」の相互関係を考慮して実行する。すなわち，罰を得る直前の選択よりも遡った形で「回避リスト」の内容が生成される可能性がある。



Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

表 1: 環境の種類

| | 欲求レベルの減少量 | | | | 欲求レベルの増加量 | | | |
|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | A | B | C | D |
| 環境 0 | 1 | 1 | 1 | 5 | 60 | 60 | 60 | 60 |
| 環境 1 | 1 | 1 | 1 | 1 | 80 | 80 | 80 | 80 |
| 環境 2 | 1 | 1 | 3 | 3 | 80 | 60 | 60 | 60 |

なお, 少なくともひとつの欲求レベルが 0 以下となった時点で, エージェントには罰を与えられるが, その際, エージェントは状態 s0 に戻されるとともに, すべての欲求レベルが初期値である 100 にリセットされる. また, 環境が変化した場合にも, 同様に, 欲求レベルを 100 に初期化し状態 s0 へ戻した.

ひとつの実験は,1 億 5 千万回行動が選択するまで実行した. そのような実験を 100 回行い, 各環境における報酬および罰の獲得回数の平均値で評価を行う.

### 5.4. 実験結果および考察

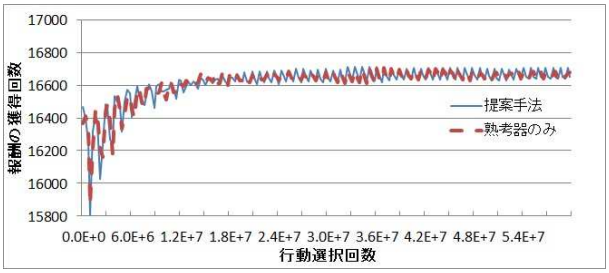


図 4: 環境 0 における報酬の獲得回数

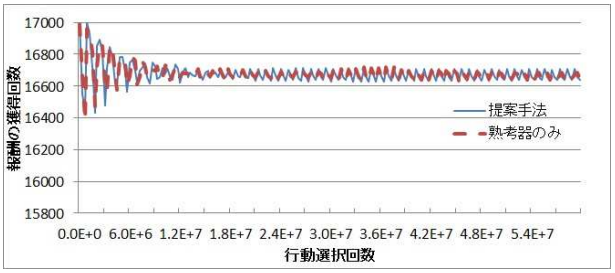


図 5: 環境 1 における報酬の獲得回数

即応器のみの場合は, 文献 [22] の知見通り, 環境 1 や環境 2 では完全に罰を回避できるが, 環境 0 では完全なる罰の回避は実現されなかった.

次に, 提案手法と熟考器のみの場合の比較を行う. 結果を図 4 から図 9 に示す. ここで, 横軸はすべて揃えてあるが, 図 7 から図 9 の縦軸は視認性向上のため揃えられていない点に注意されたい.

図 4 から図 6 より報酬の獲得回数については両手法間にほとんど差がない. 一方, 罰の獲得回数については, MPS で対応不可能な環境 0 ではほとんど差がない (図 7). それに対し, MPS で対応可能な環境 1(図 8) や環境 2(図 9) では, 提案手法では罰の獲得回数が 0 であるのに対し, 熟考器のみの場合では, 罰を獲得してしまっている.

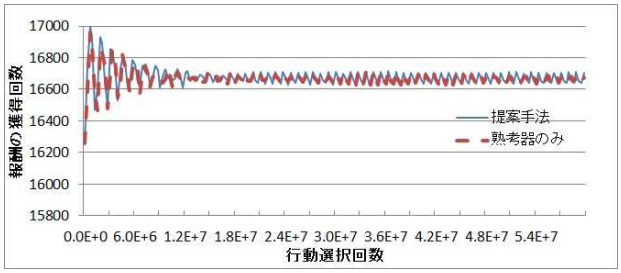


図 6: 環境 2 における報酬の獲得回数

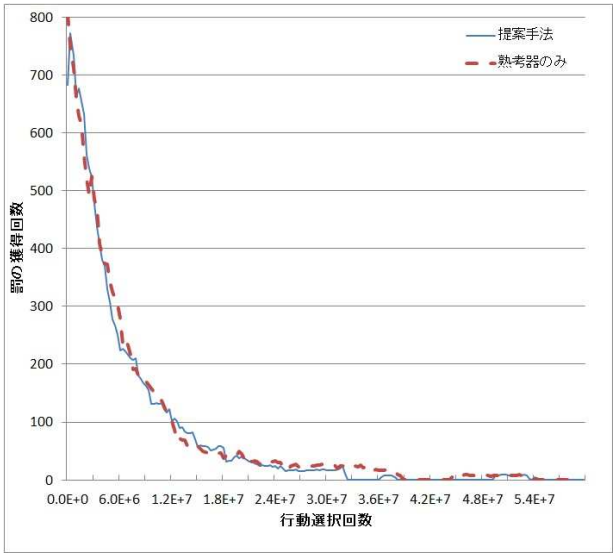


図 7: 環境 0 における罰の獲得回数

熟考器のみの場合では, MPS で対応可能なより簡単な環境に対しても, 熟考器による罰リストの学習を目指すため, 多くの罰経験が必要になったものと考える. それに対し, 提案手法は, 調停器により, 適宜, MPS と罰リストを切り替えることができ, 環境の難しさに応じて, 適切なモジュールの選択ができている. このことから, 「即応器」, 「熟考器」, 「調停器」 を有する提案手法の有効性を確認することができた.

## 6. おわりに

意識的意思決定システムとして「即応器」「熟考器」「調停器」から構成されるシステムを提案し, 数値実験により有効性を確認した. 今後は, 提案手法のマルチエージェント環境下での有効性や, 学習を導入した調停器に関する定理の導出などを行う予定でいる.

## 謝辞

明治大学理工学研究科の武野純一教授から意識システムに関する貴重なご助言をいただきました. ここに感謝の意を表します. また, 本研究は JSPS 科研費 26330267 の助成を受けたものです。



Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.


図 8: 環境 1 における罰の獲得回数

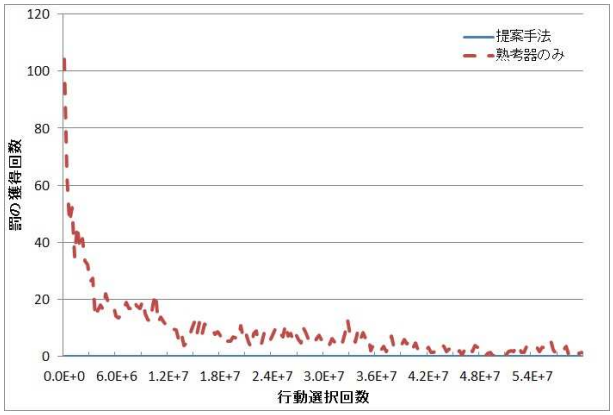


図 9: 環境 2 における罰の獲得回数

## 参考文献


[1] 甘利 俊一 (監修), 加藤 忠史 (編集): 精神の脳科学 (シリーズ脳科学 6), 東京大学出版会 (2008)

[2] Brooks, R. : A robust layered control system for a mobile robot, Robotics and Automation, **2**-1, 14/23 (1986)

[3] ジェフ・ホーキンス, サンドラ・ブレイクスリー, 伊藤 文英 (翻訳): 考える脳考えるコンピューター, ランダムハウス講談社 (2005)

[4] 深尾憲二朗: 他者を真似る自己, 講座生命 Vol.3, 中村雄二郎・木村敏 監修, 河合文化教育研究所 (1998)

[5] 深尾憲二朗: 自己・意図・意識—ベンジャミン・リベットの実験と理論をめぐって, 講座生命 Vol.7, 中村雄二郎・木村敏 監修, 河合文化教育研究所 (2005)

[6] ダニエル・カーネマン (著), 友野典男 (解説), 村井 章子 (翻訳): ファスト&スロー (上, 下): あなたの意思はどのように決まるか?, 早川書房 (2012)

[7] 兼本浩祐: 心はどこまで脳なのだろうか (神経心理学コレクション), 医学書院 (2011)

[8] 川人光男: 脳の計算理論, 385/403, 産業図書 (1996)

[9] 喜多村 直: ロボットは心を持つか—サイバー意識論序説, 共立出版 (2000)

[10] クリストフ・コッホ, 土谷尚嗣 (翻訳), 金井良太 (翻訳): 意識の探求—神経科学からのアプローチ (上, 下), 岩波書店 (2006)

[11] ジョセフ・ルドゥー (著), 谷垣 暁美 (翻訳): シナプスが人格をつくる 脳細胞から自己の総体へ, みすず書房 (2004)

[12] ニック・レーン, 斉藤隆央 (翻訳): 生命の跳躍—進化の 10 大発明, 345/386, みすず書房 (2010)

[13] 前野隆司: 脳はなぜ「心」を作ったのか—「私」の謎を解く受動意識仮説, 筑摩書房 (2004)

[14] マーヴィン・ミンスキー, 竹林洋一 (翻訳): ミンスキー博士の脳の探検 —常識・感情・自己とは—, 共立出版 (2009)

[15] 岡野憲一郎: 脳から見える心—臨床心理に生かす脳科学, 岩崎学術出版社 (2013)

[16] スーザン・ブラックモア, 山形浩生 (翻訳), 守岡桜 (翻訳): 「意識」を語る, NTT 出版 (2009)

[17] スーザン・ブラックモア, 筒井晴香 (翻訳), 信原幸弘 (翻訳), 西堤優 (翻訳): 意識, 岩波書店 (2010)

[18] 宮崎和光，山村雅幸，小林重信: 強化学習における報酬割当ての理論的考察，人工知能学会誌，**9**-4, 580/587 (1994).

[19] 宮崎和光，山村雅幸，小林重信: k-確実探査法 : 強化学習における環境同定のための行動選択戦略，人工知能学会誌，**10**-3, 454/463 (1995)

[20] 宮崎和光，山村雅幸，小林重信: MarcoPolo : 報酬獲得と環境同定のトレードオフを考慮した強化学習システム，人工知能学会誌，**12**-1, 78/89 (1997)

[21] Miyazaki, K. and Kobayashi, S.: Exploitation-Oriented Learning PS-r#, Journal of Advanced Computational Intelligence and Intelligent Informatics, **13**-6, 624/630 (2009)

[22] 宮崎和光: 複数報酬環境下における意識的意思決定方法に関する研究, 第 39 回知能システムシンポジウム, 95/98 (2012)

[23] 宮崎和光: 複数種類の報酬と罰に対応した経験強化型学習の提案と設計指針に関する研究,666 平成 24 年 電気学会 電子・情報・システム部門大会, 559/564 (2012)

[24] 宮崎和光, 武野純一: 意識システムにおける 2 次系の必要性に関する一考察, 第 3 回コンピューテーショナル・インテリジェンス研究会, 59/63 (2013)

[25] 苧阪直行: 意識とは何か - 科学の新たな挑戦, 岩波書店 (1996)

[26] 日経サイエンス編集部: 脳科学のフロンティア 意識の謎 知能の謎 (別冊日経サイエンス 166), 日経サイエンス (2009)

[27] 小野武年, 西条寿夫: 知・情・意の神経機構, BRAIN and NERVE, **60**-9, 995/1007 (2008)

[28] 小野武年: 情動と記憶, 中山書店 (2014)

[29] ワイルダー・ペンフィールド (著), 塚田 裕三 (翻訳), 山河 宏 (翻訳): 脳と心の神秘, 法政大学出版局 (2011)

[30] ベンジャミン・リベット, 下條信輔 (翻訳): マインド・タイム 脳と意識の時間, 岩波書店 (2005)

[31] 下條 信輔: 「意識」とは何だろうか—脳の来歴、知覚の錯誤, 講談社 (1999)

[32] Sutton, R. S. & Barto, A. G.: Reinforcement Learning: An Introduction, A Bradford Book, MIT Press (1998)

[33] 武野純一: ロボット・意識・心—人工意識の構築へ向けて (現代理工学大系), 日新出版 (2004)

[34] 武野純一: 心をもつロボット—鋼の思考が鏡の中の自分に気づく！, 日刊工業新聞社 (2011)

[35] Junichi Takeno: Creation of a Conscious Robot: Mirror Image Cognition and Self-Awareness, Pan Stanford Publishing (2012)




Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.