論文 / 著書情報
Article / Book Information

題目(和文)	数学的モデルを用いたサルの道具使用能力の推定
Title(English)	Estimation of monkey's tool use abilities by mathematical models
著者(和文)	毬山利貞
Author(English)	Toshisada Mariyama
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第8150号, 授与年月日:2010年9月25日, 学位の種別:課程博士, 審査員:中村 清彦
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第8150号, Conferred date:2010/9/25, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

博士論文

数学的モデルを用いたサルの道具使用能力の推定

Estimation of Monkeys' Tool Use Abilities by Mathematical Models

東京工業大学　大学院総合理工学研究科
知能システム科学専攻
毬山　利貞
Toshisada Mariyama

目 次

第1章　序章

霊長類は自分の身体以外の物体を道具として使用する能力を有する (Köhler, 1927; Goodall, 1964; Nishida, 1973; Beck, 1980; Matsuzawa, 2001)。霊長類以外の生物が道具を使用することは稀であり，道具使用行動は霊長類の高次知能を示す象徴と考えられている。道具使用行動を理解することは，霊長類の持つ高次知能の理解につながる可能性がある。

道具使用行動はその課題の複雑さから霊長類の知能を表す指標として多く用いられている (Köhler, 1927; Beck, 1990)。より高度な知能を持つ種はより複雑な道具使用行動を獲得できると考えられている。道具使用行動に関する研究では，野性下の霊長類各種に対し野外観察を行いその行動を記録する手法と，飼育下においてコントロールされた実験を対象種に課しその認知能力や神経活動などを調査する手法が代表的である。

まず野外観察による研究の現状について説明する。Goodall は野生チンパンジーが枝から葉を取り除き，その枝を用いて蟻を食べるという行動を発見した (Goodall, 1963; 1964)。彼女の報告以前においても飼育下のチンパンジーが道具を用いることは知られていた (Köhler, 1927)。しかし当時，野生下においてはヒト以外の種は道具を使用できないと考えられていた。道具使用行動を自然に行なえることがヒトの定義ともなっていた程，道具使用行動は高度な行動だと考えられていた。彼女の報告によりその認識も変化し，チンパンジーの道具使用行動に関する様々な野外観察活動が行なわれるようになった。現在では，アフリカ各地に各国の調査チーム

が観察所を作り，野生下のチンパンジー・ボノボ・オランウータンなどの大型霊長類の道具使用行動が精力的に観察されている。その結果，木の枝を利用した蟻つり (Goodall, 1963; MacGrew, 1974; Goodall, 1986; Sugiyama et al., 1988; Boesch & Boesch, 1990; Sugiyama, 1995; Whiten et al., 1999; Humle & Matsuzawa, 2002)，木の葉を用いて木のうろに溜まった水を飲む行動 (Whiten et al., 1999)，木の葉をナプキンとして使用する行動 (Goodall, 1986; Whiten et al., 1999)，石を用いた種割行動 (Sugiyama & Koman, 1979; Hannash & McGrew, 1987; Kortlandt & Holzhaus, 1987; Boesch et al., 1994; Inoue-Nakamura et al., 1997) のように様々な道具使用行動が報告されている。これらの研究の中には，自然環境下にある物体をそのまま道具として用いるだけでなく，物体を加工し，道具を作成し，その道具を用いる行動も数多く報告されている (Beck, 1980; Sugiyama & Koman, 1979; Boesch & Boesch, 1990; Whiten et al., 1999; Matsuzawa, 2001)。これらの長年に渡る研究成果を基に Whiten らはアフリカ各地におけるチンパンジーの特徴的な行動の分布を整理した (Whiten et al., 1999)。この報告のように，野生チンパンジーの道具使用行動に関しては多くの情報が整理されつつある (Whiten et al., 1999; Matsuzawa, 2001; Nishida et al., 2009)。

一方チンパンジーのような大型類人猿と比較して，サルのような大型でない霊長類に関しては野生下における道具使用行動の報告はほとんど無い。ただし例外的に，海岸の岩で貝の殻を割る行為 (Carpenter, 1887; Malaivitnond et al., 2007) など，は存在する。

実験室内における行動実験研究は，野外観察と比較して観察環境や行動課題がコントロールしやすく，これまでに多く行われている手法のひとつである。これらの研究では，様々な行動課題を対象種に課し，その行動を分析し，その種のもつ認知能力を推定する。そのさきがけ的な研究として Köhler による研究があげられ

る。Köhler はチンパンジーを対象に様々な道具使用行動課題を課し，行動を詳細に記述した。この研究により，その後の多くの霊長類研究において，道具使用行動は知能の尺度と考えられるようになった。また野生下にて育ったチンパンジーと比べて，飼育下にて育ったチンパンジーは数多くの課題をこなせるようになる (Beck, 1980; MacGrew, 1992)。同じ理由からだと推測されるが，野生下では道具使用の報告事例がほとんど存在しないサルのような種も，飼育下では訓練により道具使用が可能になる (Iriki et al., 1996; Hauser, 1997; Hauser et al., 1999; Ishibashi et al., 2000; Hihara et al., 2003; Santos et al., 2005; 2006)。飼育下における霊長類の道具使用行動に関する代表的な研究事例として，スティックを用いた道具使用課題 (Köhler, 1927; Beck, 1990; Ishibashi et al., 2000; Hihara et al., 2003; Hauser, 1997; Santos et al., 2005)，箱を用いて餌を取る課題 (Köhler, 1927; Beck, 1990)，筒の中にある餌をとる課題 (Whiten et al., 2005)，数字ボタン押し課題 (Matsuzawa, 1985; Kawai & Matsuzawa, 2000; Beran, 2004)，石器を使用する課題 (Sumita et al., 1985; Hannach & McGrew, 1987; Hayashi & Matsuzawa, 2005) が挙げられる。また，各種の道具使用行動を分析し，その行動を整理，分類した研究も存在する (Torigoe, 1985; Takeshita et al., 1996; Matsuzawa, 2001)。

実験室内の研究では，これら行動研究に加え，脳内活動の計測に関する研究も近年盛んに行われている。具体的な手法としては，サルを対象とした電気生理実験 (Iriki et al., 1996; Maravita & Iriki, 2004; Rizzolatti & Craighero, 2004)，ヒトやサルを対象とした fMRI をはじめとするイメージング実験 (Imamizu et al., 2000; Obayashi et al., 2001; Johnson-Frey, 2004; Higuchi et al., 2007; Imamizu & Kawato, 2009; Menz et al., 2010) が挙げられる。一方，法的な制限によりチンパンジーなどの大型類人猿に関しては研究事例が限られている。これらの手法は霊長類の脳内活動を計測できる利点がある。その計算機構を考える上で有力な手法

であり，今後のさらなる発展が期待される。

以上のように，霊長類の道具使用行動に関する研究は実験・観察的な手法が主流である。そして各種が遂行可能な道具使用行動についても明らかになりつつある。

一方，霊長類の道具使用行動において，各種が用いるアルゴリズムを数学的アプローチから理解しようとする研究はほとんど無い。ただし，その手掛かりとなると思われる一般的な脳の数学的モデルや概念は存在する。

例えば内部モデルという概念がその 1 つである。内部モデルとは，自分の身体や外界の物体の入出力関係（e.g. 行動と物体の動き）を模倣する脳が持つモデルである。生物がこれら内部モデルを獲得し，予測や行動を行なっている可能性は高い (Ito, 1984; Kawato et al., 1987; Shadmehr et al., 1994; Wolpert & Kawato, 1998; Imamizu et al., 2000; Desmurget & Grafton, 2000; Kosslyn et al., 2001; Hesslow, 2002; Higuchi et al., 2007; Imamizu & Kawato, 2009)。元々は道具使用行動とは関係の無い研究であったが，道具使用行動を直接扱った研究や適用が可能な研究も存在する (Imamizu et al., 2000; Higuchi et al., 2007; Wolpert & Kawato, 1998)。この中で理論的な研究では，主に道具の位置制御などの力学的な問題を主眼としたスキーマを提案している (Wolpert & Kawato, 1998)。残念ながら霊長類各種がどのような道具使用行動が可能であり，不可能であるのかについては特に言及していない。

道具使用に限定せず，より一般的な行動系列学習として強化学習理論がある。Sutton らは TD 誤差学習法を利用した actor-critic モデルを提案した (Sutton & Barto, 1995; 1998)。このモデルは Houk らにより実際の脳内活動に対応付けされている。彼らは，actor と critic を大脳基底核内の 2 つの領域と対応させると，基底核内の神経回路が強化学習機構として動作しうるという仮説を提案した (Houk et al., 1995; Barto, 1995)。この実験的証拠として Schultz らは，行動課題中のサ

ルの脳内活動を計測し，中脳ドーパミン神経細胞の活動が TD 誤差学習法の報酬予測誤差を，大脳基底核線状体が Critic 部で計算される状態価値関数をそれぞれ符号化していることを示した (Schultz et al., 1993; 1997)。さらに，この Schultz の実験結果を説明するために，Montague らや，Suri らはドーパミン神経細胞からの入力を受けて行動系列を学習する神経回路モデルを提案している (Montague et al., 1996; Suri & Schultz, 2001)。その他にも，強化学習とそれを用いた行動系列の学習を議論する先行研究は数多く存在する (Doya, 1999; Hikosaka et al., 1999; Nakahara et al., 2001)。道具使用行動も行動系列学習の一種であるため，道具使用行動時に霊長類がこの学習方法を使用している可能性がある。ただし，強化学習理論はラットのような高次知能を有するとは思えないような種の行動も再現している。また，ランダムサーチを利用した学習方法である。霊長類がこの理論のみで道具使用行動を行なっているとは考え難い。

道具使用行動における各種のアルゴリズムを数学的モデルを基に理解するためには，上記のような一般的な脳の先行モデル・概念を基に，新たなアルゴリズムを構築する必要がある。

他の物理学の分野と同様，上記の豊富に行われてきた実験研究を数学的に記述し，その理解を深めることは，霊長類の持つ知能を厳密に深く理解する上で重要な手法である。数学的アプローチを道具使用行動に適用した場合，動物行動実験研究において導き出された結果に対しその結果の背景に潜む因果関係を厳密に説明できる可能性がある。つまり，ある種の道具使用行動に関する行動決定アルゴリズムを理解できるようになる可能性がある。本研究は，この数学的アプローチを霊長類の道具使用行動に適用し，その原理・原則を数学的に記述することを通じて，霊長類研究の更なる発展に貢献することを目的に行われた。

本研究は霊長類の中でも，ニホンザル，ワタボウシタマリンの二種の行なった

道具使用行動に対して，数学的モデルを用いてその行動を再現しその能力を推定した。道具使用行動の中でも，複数の道具の中から使用する道具やそれらの道具の使用順序を決定するアルゴリズムに本研究は焦点を当てた。

第 2 章では，サルの熊手使用行動を再現した数学的モデルを先行研究として紹介する (Mariyama et al., 2001)。このモデルはニホンザルの熊手使用行動の脳内活動 (Iriki et al., 1996) と，2 つの行動実験 (Ishibashi et al., 2000; Hihara et al., 2003) を再現対象としている。数学的モデルは，強化学習の一種である actor-critic モデルに，tool evaluator と呼ぶ機能を付け加えたモデルである。計算機実験を行った結果，提案したモデルが上述の 3 つの実験結果を再現した。その数学的モデルを用いて新たな実験を提案し，熊手使用時の脳内活動を予測した。

第 3 章では，各種のもつ内部モデルを推定するための実験を提案し，ニューラルネットワークを用いてその結果を予測した (Mariyama & Itoh, 2009)。内部モデルは霊長類などの生物が道具使用行動などの難しい課題を解く上で重要な要素の 1 つとして考えられている。一方，獲得可能な内部モデルは霊長類各種によって異なる可能性もある。各種のもつ内部モデルの構成に関する知見の取得と，その推定方法確立を目標に，我々は異なる内部モデルと道具使用行動との関係を調査した。具体的には，異なる内部モデルを持つ 2 つのエージェントに本研究にて提案する課題を課し，その計算機実験を行なった。ひとつめのエージェントは報酬量（各課題の成功または失敗）を予測する内部モデルを持ち，その計算を利用して課題を解くエージェントである。 もう一方のエージェントは環境の状態推移を予測する内部モデルを有し，その計算結果を利用して課題を解くエージェントである。本研究にて用いる課題は，道具使用行動の代表的な課題の 1 つである熊手使用行動課題に関する架空の課題である。課題は 2 種類から構成され，一方の基本的な課題を十分にこなせるようになった後に，もう一方の課題を課した。これらの課

題に対して上述の 2 つのエージェントを対象に計算機実験を行なった。その結果，後者のエージェントが前者のエージェントに対してより早く環境の変化に対して順応できることが示された。 もし仮に，上記のような内部モデルを持つ生物に提案した課題を行い，計算機実験から予測された結果と比較すれば，その生物が持つアルゴリズムを推定できる可能性がある。

第 4 章では，Santos ら (2005; 2006) によって行われたワタボウシタマリンの道具使用行動を再現する数学的モデルを作成した (Mariyama et al., 2010)。道具使用行動の数学的モデルを作成する上で，内部モデルの種類と，複数の道具を系列的に使用する際に内部モデルを用いて何手先の道具の動きまで予測できるか，に着目した。上記に関するアルゴリズムが異なる 5 つのエージェントを本研究は用意した。各エージェントに，Santos らによって行なわれた行動実験と同じ実験環境をコンピュータ上に再現し，シミュレーション実験を行った。そのシミュレーション実験結果とタマリンの行動実験結果とを比較した。その結果，タマリンは，本研究で定義した逆モデル（エージェントが望む非操作対象物の動きとその現在の状態から，使用すべき道具の状態と操作方法を求めるモデル）を用いて 1 つの道具の動きを予測する能力がある可能性を示唆する結果を得た。また，提案モデルからタマリンの持つ特性を考察した後に，提案モデルの妥当性を検証する実験を提案した。

第 5 章は，全体のまとめを行った。

なお本研究では，道具使用行動を「自分以外の物質を用い，第 3 の物体及び自分の状態（位置，形状など物質的・物理的特性）を変化させ，ゴール到達までの効率を向上させる行為である。」と定義する。道具使用行動の定義に関しては画一的な定義は存在せず，多くの場合各研究者が個別に定義している。例えば，van Lawick-Goodall は「the use of an external object as a functional extension of mouth or

beak, hand or claw, in the attainment of an immediate goal」と定義している (van Lawick-Goodall, 1970)。また Takeshita らは「the behaviors in which an individual uses a single or multiple detached environmental object(s) as an intermediary to efficiently change the environment in obtaining a goal」と定義している (Takeshita et al., 1996)。このように数多くの研究において，様々な道具使用行動の定義が存在する。これらの状況の中で，Beck は様々な研究者の定義を考察し，以下の定義を行った (Beck, 1990)；「the external employment of an unattached environmental object to alter more efficiently the form, position, or condition of another object, another organism, or the user itself when the user holds or carries the tool during or just prior to use and is responsible for the proper and effective orientation of the tool.」。現時点において，この定義は多くの賛同を得ている定義のうちの 1 つであるため，本研究の定義はこの Beck による定義を基に作成した。

第2章　数学的モデルを使用した先行研究の紹介

2.1　はじめに

本章では，博士前期課程にて行なったサルの熊手使用行動を再現した数学的モデルを先行研究として紹介する (Mariyama et al., 2001)。第 1 章でも述べたとおり，道具使用行動を数学的に記述した研究は数少ない。そのため，一般的な動物の行動や脳内活動を表現した数学的モデルを基に，新たにモデルを構築した。本章では，Iriki らが行なっていたニホンザルの熊手使用行動の脳内活動 (Iriki et al., 1996) と，2 つの行動実験 (Ishibashi et al., 2000; Hihara et al., 2003) を再現する数学的モデルを構築した。数学的モデルは，第 1 章でも紹介した actor-critic モデルに，後述する道具を評価するシステムを付け加えた。その数学的モデルを用いて，熊手使用時の脳内活動を予測した。

2.2　方法

2.2.1　先行研究

Iriki らはニホンザルに熊手を使って餌を取らせる行動課題を行わせ，その行動様式，及び脳内活動を調べてきた (Iriki et al., 1996; Ishibashi et al., 2000; Hihara et al., 2003)。行動課題は一段階行動課題と二段階行動課題の 2 つを用いた。以下にそれぞれの行動課題の内容とその結果を示す。

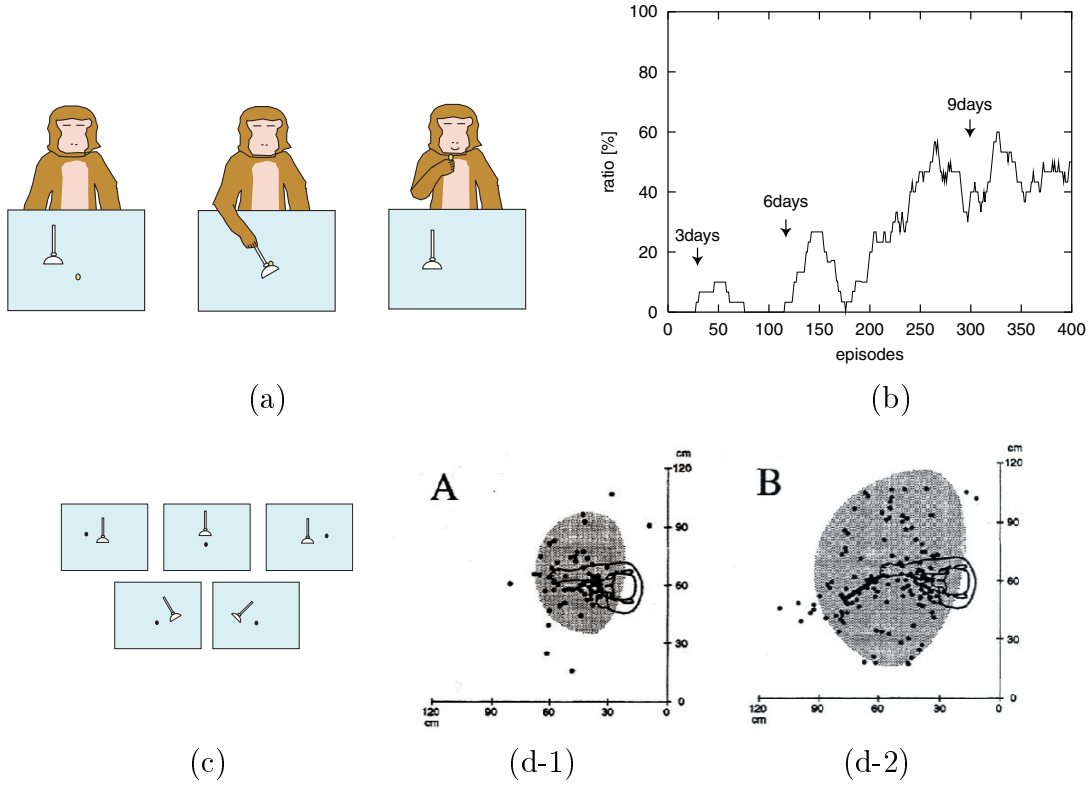


図 2.1: 一段階行動課題 (a) 課題内容 (b) 行動実験結果 (Hihara et al. 2003)：test エピソードの課題成功率の推移を示した。縦軸にはエピソード数を縦軸には該当するエピソードから 30 エピソード間の課題成功率を示している。3 日目，6 日目，9 日目の開始エピソードを矢印で示した。(c) テストエピソードの実験状態。d) 頭頂間溝の細胞応答 (Iriki et al., 1996) 餌を持った手を動かした時に，細胞が発火した場所を点で示した (d-1）素手の状態（d-2）熊手の使用直後の状態

一段階行動課題

一段階行動課題は手元に置かれた熊手を使って遠方の餌を取る課題である（図2.1a）。実験中サルは台座に固定されている。実験開始時，餌は手では届かない位置に，熊手は手で取れる位置に置く。よって，サルが餌を取るためには，熊手を用いて餌を手元まで引き寄せなければならない。

この課題は，トレーニングエピソードとテストエピソードの2種類のエピソードから構成されている。

トレーニングエピソードは，サルが熊手の使用方法を覚えるための練習用エピソードである。課題開始から数日間は，餌と熊手の先端がほぼ接触している。従って，この状況下ではサルは熊手をただ手前に引くだけで餌を取ることができる。この位置関係でサルが餌を取れるようになった後，徐々に餌を熊手から遠ざけて置いていく。課題遂行中には手本となる動きは一切見せない。よってサルは試行錯誤的に熊手の使用方法を学習していく。各エピソードの終了条件は，1）サルが餌をとれた時，2）サルが餌か熊手を遠方に飛ばしてしまい課題の続行が不可能になった時，3）約10秒〜20秒経っても餌が取れなかった時とした。

一方，テストエピソードはトレーニングエピソードの学習成果を計るエピソードである。テストエピソードで用いた熊手と餌の位置を図2.1cに示す。この5種類の配置を1セットとする。エピソードの終了条件はトレーニングエピソードと同じである。

課題開始当初，トレーニング課題を約50回毎にテスト課題を1セット行う。その後テスト課題での成功率の上昇に合わせて，テスト課題を行う割合を増やす。

図2.1bに一頭のサルのテストエピソードにおける成功率の推移を示した。横軸にテストエピソード数を，縦軸に該当するテストエピソードからその後の30テストエピソード間における課題成功確率を示している。なお，テストエピソードに

おいて，熊手の配置毎（図 2.1c) に成功率の推移は異なる。図 2.1b では，その平均の推移を見ている。

結果，サルは 6 日目 (約 120 エピソード）を過ぎたあたりから徐々に餌を取れるようになった。7〜10 日目（約 250〜400 エピソード）にかけて，約 40〜60 %の確率で餌を取る事が示された。なお，この一段階行動課題は Ishibashi らによって詳細に調べられている。彼らの実験においても，約 2 週間の訓練により課題成功率が 50〜60 %程度まで上昇する事が確認されている (Ishibashi et al., 2000)。その後の研究で，さらに 1ヶ月ほどのトレーニングにより，サルはほぼ確実に餌を取ることが出来るようになることが確認されている (Hihara et al., 2003)。

熊手を用いて餌を取れる段階で，頭頂間溝（IPS）におけるニューロンの応答が調べられている (Iriki et al., 1996)。頭頂間溝には，体性感覚性と視覚性の二つの受容野を持つニューロンが存在する。その中で肩のあたりに体性感覚受容野を持つ, 一つの代表的なニューロンの応答が図 2.1d に示されている。この実験では，まず実験者が手に餌を持つ。そして，固定されたサルの前を，実験者は自分の手を約 $0.5ms^{-1}$ の速度で様々な位置に移動させる。サルが素手の時に上述のニューロンが 3Hz 以上の発火頻度で発火した場所を図 2.1d-1 に，熊手を使って餌を取った直後に発火した場所を図 2.1d-2 にそれぞれ示されている。熊手を使って餌を取った後，約 1〜5 分経過すると視覚受容野は熊手を持っていても図 2.1d-1 になる。このニューロンの視覚受容野は，素手の状態では手の届く範囲であり，熊手使用時には熊手で届く範囲にまで拡大した事が示された (Iriki et al., 1996)。この結果は，サルは自分が届くと考えている領域を，手にもっている物体に応じて適切に変更している可能性を示唆している。

二段階行動課題

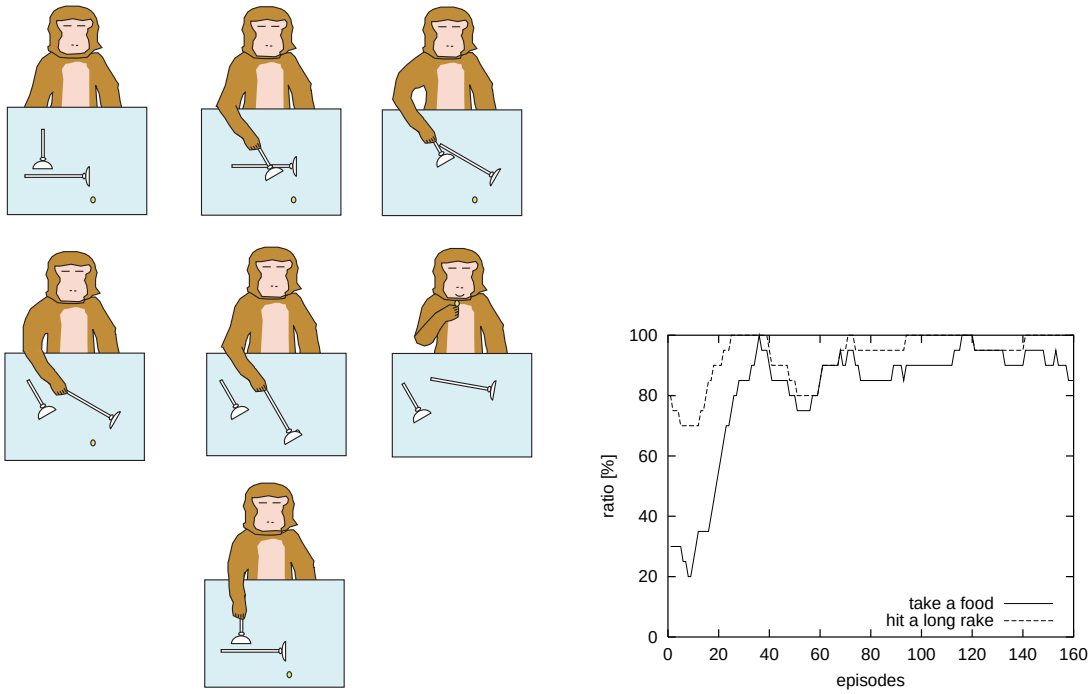
一段階行動課題を約1ヶ月間行い，サルが熊手を用いて十分に餌が取れるようになった後に，二段階行動課題を開始する（図 2.2a）。なお，この1ヶ月間のトレーニングにより，サルは一段階行動課題を高確率で解けるようになっている。この課題では長さの異なる熊手を 2 本用意する。短い方の熊手は手の届く範囲に，長い熊手は短い熊手を使わなければ届かない位置に，餌は長い方の熊手でしか届かない位置に配置する。従ってサルが餌を取るためには，短い熊手で長い熊手を手繰り寄せ，長い熊手を用いて餌を引き寄せなければならない。

二段階行動課題はテストエピソードのみで構成されている。

図 2.2b はその実験結果を示している (Hihara et al., 2003)。横軸にはエピソード数を，縦軸には該当するエピソードから，その後 20 エピソード間の割合を示している。図中の破線はサルが餌を取れた確率を，実線は餌の取得に関わらず，短い熊手を持ち熊手の先端を長い熊手の一部にでも当てた確率を示している。

短い熊手を用いて長い熊手を取るためには，短い熊手の先端を長い熊手の先端部分の後方に回り込まなければならない（図 2.2a の二番目の図）。そのため，課題開始から 50 エピソードまでは，短い熊手で長い熊手を取ろうとしても，長い熊手をはじいてしまうなどして，長い熊手を確実に取ることは出来なかった（グラフには記載されていない。）。だが，二段階行動課題を行う以前に，短い熊手と長い熊手を手に届く範囲に置き，長い熊手でしか届かない位置に餌を置いたところ，サルは常に長い熊手を用いて餌を取った。また，二段階成功課題で長い熊手を短い熊手で引き寄せると，即座に短い熊手を放し，長い熊手に持ち替えた。

そこで私は長い熊手を手に入れようする行動（図 2.1b 破線）は，餌を取るまでの手順をサルが計画できたと考えた。この考えに基づけば，サルは二段階行動課題開始当初から，約 80 %の確率（図 2.2b 破線）で餌を取るまでの正確な手順を計



(a) (b)

図 2.2: 二段階行動課題：(a) 課題内容 (b) 行動実験結果 (Hihara et al., 2003)：横軸はエピソード数を，縦軸は該当するエピソードから，その後 20 エピソードまでの確率を示している。実線は餌が取れた確率を，破線は短い熊手を手に取り，短い熊手を長い熊手に当てた確率を示している。

画できた事になる。ただし，熊手の操作方法（e.g. 角度，軌道）を正確に覚え，確実に餌を取ることが出来るようになるまでには，図 2.2b 実線が示すように約 40〜80 エピソードのトレーニングが必要になる。

まとめ

一段階課題において，サルは道具の使用方法を試行錯誤的に探索し，餌による報酬でその方法を強化している。

一方，二段階行動課題はサルにとって全く新しい状況であるにも関わらず，課題開始当初から適切な行動計画を行えた。

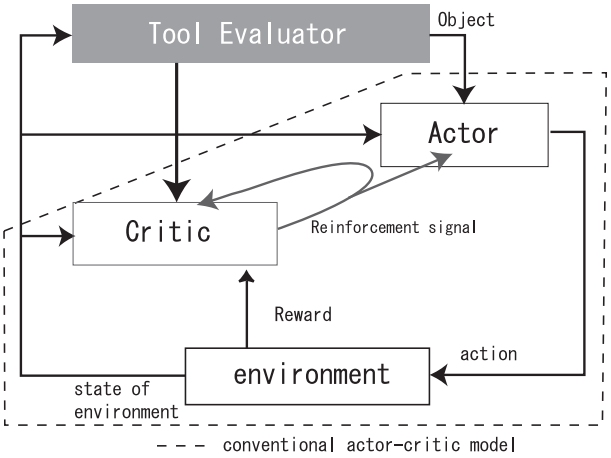


図 2.3: モデルの模式図

これらの行動はどのような脳内機構で生み出されているのだろうか。そして，このときの頭頂間溝皮質のニューロンはどのような形で行動に寄与しているのか。本研究では一段階課題で試行錯誤的に熊手の使用方法を学習し，二段階課題では一段階課題を応用し課題開始直後から課題をこなせる現象を再現するモデルを提案する。同時に頭頂間溝皮質のニューロンの応答をモデル上で一部再現し，その機能と行動との関係を説明する。

2.2.2 数学的モデル

本モデルは以下の仮説を基にして作った。

仮説 脳内には道具の価値を餌等の報酬量を基に計算する機構が存在する。

道具の価値を計算する機構を，本モデルでは道具評価システムと名付けた。この道具評価システムを強化学習で提案されたactor-critic モデルに付加したのが本

モデルである（図 2.3）。本モデルは，環境，道具評価システム，actor，critic の 4 つの部分から構成されている。

以下において，本モデルの各構成要素の詳細について述べる。なお，図 2.5 に本モデルの計算手順を簡易的にまとめた。

環境

本研究では環境を以下に示す有限オートマトン E を用いて表現した。

$$E = (E_K, E_\Sigma, E_\delta, E_q, E_F) \tag{2.1}$$

ここで E_K,E_Σ,E_δ,E_q,E_F は，それぞれ環境の状態，アルファベット，遷移関数，初期状態，終了状態を意味している。

環境の状態 E_K

E_K は x_f，x_s，x_l の 3 つの変数から構成されている。それぞれはサルの位置を原点とした，餌，短い熊手，長い熊手の位置座標（2 次元直交座標）を示している。ただし特殊な状況として，各物体が手の中にあるときは 0，台上に存在しないときは 1 と表記する。環境状態 E_K を数式で表記したのが式 2.2 である。

$$E_K = \{(x_f, x_s, x_l) | x_f, x_s, x_l \in Z^2 \vee \{0, 1\}\} \tag{2.2}$$

以後，f,s,l はそれぞれ餌，短い熊手，長い熊手とする。

アルファベット E_Σ，遷移関数 E_δ

アルファベット E_Σ は，他のシステムから環境 E への入力である。本モデルでは Actor の出力がこれに相当する。これは要素 $A_k(k = 0 \cdots 4)$ からなる。A_k の意味を表 2.1 に示した。

また，遷移関数 E_δ はアルファベット E_Σ によって生じる状態 E_K の推移を示している。すなわち

$$E_\Sigma = \{A_k | 0 \leq k \leq 4, k \in Z\} \tag{2.3}$$

$$E_\delta : (E_K, E_\Sigma) \mapsto E_K \tag{2.4}$$

なお，E_δ，E_Σ は実際のサルの行動に基づいて構成した。

表 2.1: アルファベットの内容

A_k	行動内容
A_0	何もしない
A_1	タスクを終了する
A_2	物体を手に取る。複数の物体が手に届く範囲内にある場合は，式 2.8〜2.10 で計算された価値の最も高い物体を手に取る。
A_3	手に持っている物を放す
A_4	手にしている熊手で物を手元までを引き寄せる。 複数の物体が熊手の届く範囲内にある場合，式 2.8〜2.10 で 計算された価値の最も高い物体を手元までを引き寄せる

初期状態 E_q，終了状態 E_F

初期状態 E_q は課題開始時の環境状態である。E_q には実際のサルの行動課題と同様，短い熊手 1 本で餌をとる課題，長い熊手 1 本で餌を取る課題，二段階行動課題の 3 つを用意した。

終了状態 E_F は，餌を手に取った状態，E_Σ が A_2（課題を終了する。）の場合と 1 エピソード中でのステップ数 p が決められた値 q を越えた場合とする。これらはサルの行動に則している。

$$E_F \quad = \quad \{E | x_f = 0 \vee E_\Sigma = A_2 \vee p > q\} \tag{2.5}$$

道具評価システム

このシステムでは，環境から x_f, x_s, x_l の入力を受け, 餌，短い熊手，長い熊手の価値 v_f,v_s,v_l を算出する。これらの価値を基に後述する目標物 $o \in \{f, s, l\}$ とその位置 X_o を actor-critic モデルに出力する。

道具の価値は餌の報酬量 r_f と道具の届く範囲を元に計算される。この道具の届く範囲を表現するために本モデルでは，関数 $\alpha_n(x)$，$\alpha_s(x)$，$\alpha_l(x)$ を導入した。

$\alpha_n(x)$，$\alpha_s(x)$，$\alpha_l(x)$ はそれぞれ，手，短い熊手，長い熊手に関して，サルが届くと考えている範囲を符号化した関数である。なお，入力 $x \in Z^2$ は任意の位置を示している。この 3 つの関数は任意の x に対して 0 から 1 までの値を取る。位置 x に対してこれらの関数が 0 の値を取る場合，該当する道具で x の位置にある物体を取れない，とサルが判断した事を示す。逆に 0 以上の値を取る場合は，該当の道具で x の位置の物体を取れる，とサルが判断した事を示す。この値が大きければ大きい程，その物体を取れるとサルは確信していく。なお，今後 n は手に何も持っていない状態を示す。

図 2.4a は，一段階行動課題開始前の $\alpha_n(x)$，$\alpha_s(x)$，$\alpha_l(x)$ を示している。一段階行動課題開始前ではサルは短い熊手，長い熊手を用いて餌を取る方法を知らない。よって，各道具で届く範囲を理解していないため，$\alpha_s(x)$，$\alpha_l(x)$ は任意の x に対して 0 の値を取る。一方，課題開始以前においても，素手で届く範囲は理解して

いるので，$\alpha_n(x)$ の値は届く範囲は 1 に届かない範囲は 0 になる。

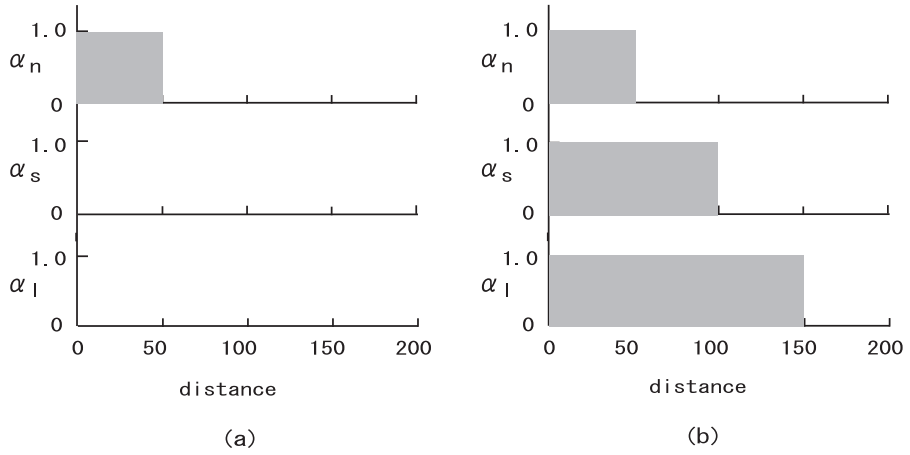


図 2.4: 各 $\alpha(x)$ の値：(a) 一段階行動課題開始前の値 (b) 一段階行動課題終了後の値：手の長さは 50，短い熊手の長さは 50，長い熊手の長さは 100 とした。従って，手で届く範囲，短い熊手を持った時の届く範囲，長い熊手を持った時の届く範囲はそれぞれ，サルの体を原点にとった半径 50，100，150 の円内部となる。

この関数の更新則を示す。手に持っている道具 $j \in \{n, s, l\}$ を用いて位置 $y \in Z^2$ においてある物体を引き寄せた場合に限り

$$\alpha_j(x) \Leftarrow \alpha_j(x) + \beta_1 1\{|\vec{Oy}| - |\vec{Ox}|\} \cdot (1 - \alpha_j(x)) \tag{2.6}$$

$$1\{z\} = \begin{cases} 1 & \cdots z \geq 0 \\ 0 & \cdots z < 0 \end{cases} \qquad z \in Z \tag{2.7}$$

とする。この更新則を用いた場合，一段階行動課題を十分に行わせた場合，各関数は図 2.4b になる。なお，今後 j は手に持っている道具とする。

この関数を用いて，各物体の価値を以下の式から求める。

$$v_f = r_f \tag{2.8}$$

$$v_s = \gamma_1 Max\{f_s(x_f, v_f), f_s(x_l, v_l)\} \tag{2.9}$$

$$v_l = \gamma_1 Max\{f_l(x_f, v_f), f_l(x_s, v_s)\} \tag{2.10}$$

$$f_s(x, v) = \alpha_s(x)v \tag{2.11}$$

$$f_l(x, v) = \alpha_l(x)v \tag{2.12}$$

$$f_n(x, v) = \alpha_n(x)v \tag{2.13}$$

ただし，r_f は餌の報酬量，γ_1 は 0〜1 までの値を取る割引率である。そしてこの道具の評価値を基に，現在手に持っている道具 j で取るべき物体 $o \in \{f, s, l\}$ とその位置 x_o，及びその価値 v_o を以下の式で求める。

$$(x_o, v_o) = f_j^{-1}[Max\{f_j(x_f, v_f), f_j(x_s, v_s), f_j(x_l, v_l)\}] \tag{2.14}$$

x_o の位置はサルから物体までの距離と熊手の長さを基に表 2.2 に示す 6 つの状態に分類できる。この x_o を表 2.2 に従って分類分けしたのが X_o である。x_o はそれぞれの物体の物理的な位置を示し，X_o はサルの考える o の位置状態を意味している。以上から求まった (X_o, o) を actor-critic 部へ出力する。

actor-critic 部

critic 部

critic 部は TD 誤差 δ を計算することにより，actor と critic 自身に環境を学習させるシステムである。

道具評価システムから目標物座標 X_o と現在目標としている物体 o（$o \in \{f, s, l\}$），環境から現在手に持っている道具 j （$j \in \{n, s, l\}$）と報酬量 r を入力として受け取り，Actor 部，Critic 部に TD 誤差を出力する。TD 誤差信号は状態評価関数 $V(X_o, o, j)$ を基に，式 2.15 から求まる。

$V(I)$，$a_k(I)$，$\alpha_j(x)$ を任意に初期化する。
環境状態 E_K の初期値を設定する。
各エピソードに対して繰り返し：
E_K を初期化
各ステップに対し繰り返し：
1) 道具評価システムで v_f，v_s，v_l，x_o，v_o をそれぞれ計算する。

$$\begin{aligned} v_f &= r_f \\ v_s &= \gamma_1 Max\{f_s(x_f, v_f), f_s(x_l, v_l)\} \\ v_l &= \gamma_1 Max\{f_l(x_f, v_f), f_l(x_s, v_s)\} \\ (x_o, v_o) &= f_j^{-1}[Max\{f_j(x_f, v_f), f_j(x_s, v_s), f_j(x_l, v_l)\}] \\ x_o &\mapsto X_o \end{aligned}$$

2)Actor で取るべき行動 A_k を決定する。

$$\pi(X_o, j, A_k) = \frac{\exp(\lambda a_k(X_o, j))}{\sum_{n=0}^{4} \exp(\lambda a_n(X_o, j))}$$

3) 行動 A_k を取り，報酬 r と次状態を観察し学習を行う。

$$\begin{aligned} \delta &= r + \gamma_2 V(X_o', o', j') - V(X_o, o, j) \\ \alpha_j(x) &\Leftarrow \alpha_j(x) + \beta_1 1[|\vec{Oy}| - |\vec{Ox}|](1 - \alpha_j(x)) \\ V(X_o, o, j) &\Leftarrow V(X_o, o, j) + \beta_2 \delta \\ a_k(X_o, j) &\Leftarrow a_k(X_o, j) + \beta_2 \delta \end{aligned}$$

E_K が終了状態なら繰り返しを終了。

図 2.5: 今回のモデルの計算アルゴリズム

状態評価関数 $V(X_o, o, j)$ は，(X_o, o, j) における将来手に入れられる総報酬量の期待値を意味している。この状態価値関数は式 2.15 で求めた TD 誤差を用いて式 2.16 に従って更新される。

表 2.2: 熊手の位置表記の変換

X	x
0	手の中にある
1	素手で届く距離にある $\alpha_n(x) > 0$
2	短い熊手で届く距離にある $\alpha_n(x) = 0 \wedge \alpha_s(x) > 0$
3	長い熊手で届く距離にある $\alpha_n(x) = 0 \wedge \alpha_s(x) = 0 \wedge \alpha_l(x) > 0$
4	長い熊手でも届かない位置にある $\alpha_n(x) = 0 \wedge \alpha_s(x) = 0 \wedge \alpha_l(x) = 0$
5	台上に存在しない

$$\delta \quad = \quad r + \gamma_2 V(X'_o, o', j') - V(X_o, o, j) \tag{2.15}$$

$$V(X_o, o, j) \quad \Leftarrow \quad V(X_o, o, j) + \beta_2 \delta \tag{2.16}$$

ただし，r は各状態における報酬量，(X'_o, o', j') は次状態を，γ_2，β_2 は 0〜1 までの値を取る定数をそれぞれ示す。

actor 部

actor 部は環境に対して行動を出力するシステムである。

道具評価システムから目標物座標 X_o，環境から現在手にしている道具 j，critic 部から TD 誤差の 3 つの値を入力として受ける。その結果，式 2.17 に示すように，入力 (X_o, j) に対し，行動 A_k を取る確率 $\pi(X_o, j, A_k)$ を行動優先度関数 $a(X_o, j)$ から計算する。この確率分布 $\pi(X_o, j, A_k)$ に基づいて行動 A_k を環境に出力する。

行動優先度関数 $a(X_o, j)$ は TD 誤差を用いて式 2.18 に従って更新する。

$$\pi(X_o, j, A_k) = \frac{\exp(\lambda a_k(X_o, j)}{\sum_{n=0}^{4} \exp(\lambda a_n(X_o, j))} \tag{2.17}$$

$$a_k(X_o, j) \Leftarrow a_k(X_o, j) + \beta_2 \delta \tag{2.18}$$

ただし λ は正の定数である。

以上，モデルの説明を終わりにする。

2.2.3 シミュレーション方法

シミュレーション課題の手順は，短い熊手と長い熊手の一段階行動課題を 3000 エピソード行わせ，次に二段階行動課題を行わせた。Hihara らの実験では，二段階行動課題を行う前にサルは一段階行動課題を十分に解けるように訓練されている。この実験設計に対応するために，つまり，エージェントが十分に一段階行動課題を成功できるようになるために，一段階行動課題を 3000 エピソード用意した。

終了条件で述べた 1 エピソード中のステップ数の上限値 q の値は，課題成功に必要最小値（一段階行動課題では 4，二段階行動課題では 7）の約 1.5 倍とした。つまり一段階行動課題では 6，二段階行動課題では 10 である。

行動優先度関数，状態評価関数の初期値は，実際のサルとの行動をあわせるため，手に届く位置に餌がある場合は餌を取る行動 $(E_\Sigma = A_2)$ を，何をすべきかわからない時は，タスクを終了させる行動 $(E_\Sigma = A_1)$ を（今回は 80 %の確率で）優先的に取るように調整した。それ以外の状況に対しては優先行動関数，状態価値関数のいずれも 0 の値を取る。

その他の学習パラメータを以下に示す。

$$\gamma_1 = 0.5, \beta_1 = 0.01, r_f = 2.0$$

$$\gamma_2 = 0.7, \beta_2 = 0.2, \lambda = 1.5$$

なお，図 2.4 と同様に，手の長さは 50，短い熊手の長さは 50，長い熊手の長さは 100 とした。従って，手で届く範囲，短い熊手を持った時の届く範囲，長い熊手を持った時の届く範囲はそれぞれ，サルの体を原点にとった半径 50，100，150 の円内部となる。

2.3 結果

図 2.6a-1 の実線は本モデルの一段階行動課題の 1 計算例を示している。破線は図 2.1b で示したサルの行動実験結果である。縦軸，横軸は図 2.1b と対応している。シミュレーションは 0〜50 エピソードまでは常に課題成功率が 0 %であったが，250〜400 エピソードには約 40〜60 %程度で課題を成功させた。この結果は，サルの実験で観察された，エピソード数と共に成功率が上昇していく現象を再現している。なお，ここには掲載してないが，このまま 3000 エピソードまで課題を続けると約 80〜90 %の確率で課題を成功する。図 2.6a-2 では，一段階行動課題について 10 個の異なるエピソード列の平均を示した。3000 エピソードまで課題を行えば，課題成功率が約 90 %になった。

図 2.6b-1 の実線は本モデルの二段階行動課題の課題成功に関する 1 計算例を示している。破線は図 2.2 の破線で示したサルの実験結果を示している。横軸，縦軸は図 2.2 に対応している。シミュレーションでの成功率は 1 エピソード目から約 90 %であり，二段階行動課題が課題開始と共に成功しているサルの実験結果を再現した。図 2.6b-2 では，一段階行動課題を 3000 エピソード行わせた後の二段階行動課題に関して 10 個の独立したエピソード列の平均を示した。1 エピソード目から 80 %以上で課題を成功させていた。以下では，本モデルの各関数の挙動を示す。図 2.7 に示しているのは，一段階行動課題を 3000 エピソード行わせた際の関数 $\alpha_j(x)$ ($x = (X, Y)$) の 1 計算例を示している。図 2.7a での XY 平面は，サルの位置を原

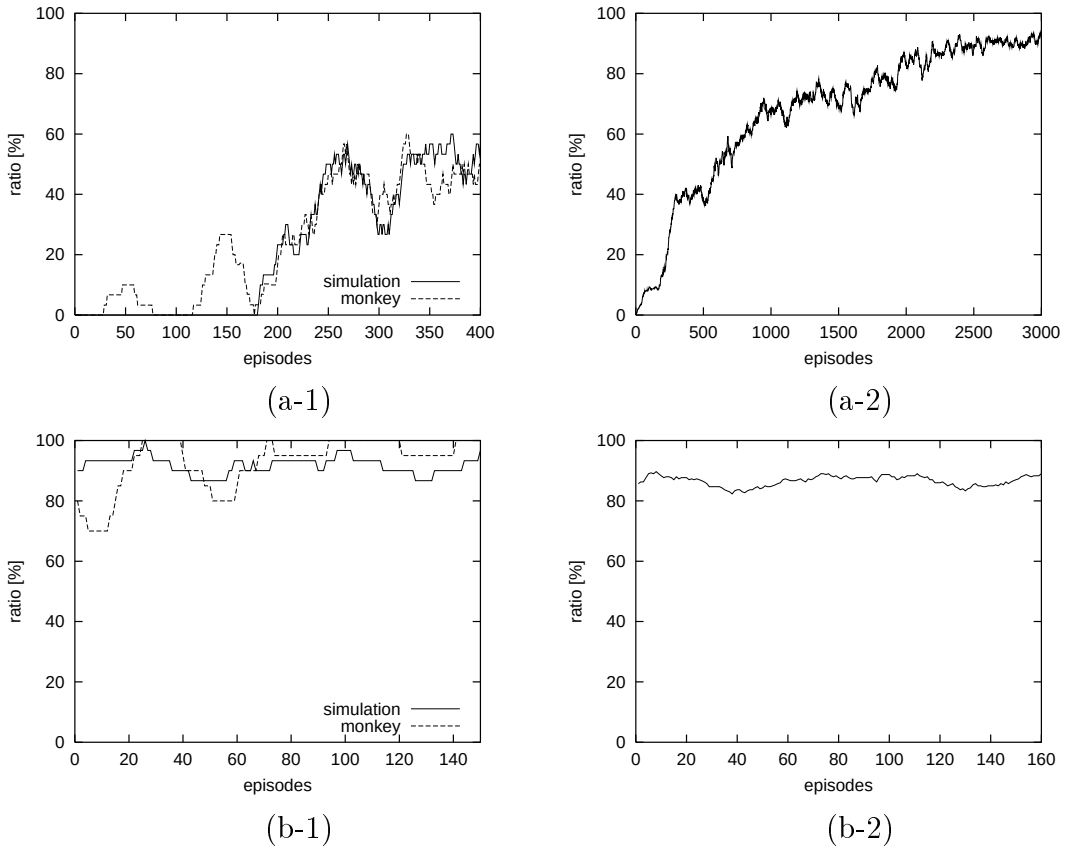


図 2.6: シミュレーション結果：(a) 一段階行動課題 (b) 二段階行動課題：(1) 実線はシミュレーションの 1 例を，破線はサルの行動実験結果 (図 2.1b) を示している。なお，二段階行動課題では，サルを対象とした行動実験，シミュレーション実験ともにテストエピソードのみで構成されている。一方，一段階行動課題では，サルの行動課題は，テストエピソードとトレーニングエピソードの二種類から構成され（図ではテストエピソードのみ掲載），シミュレーション実験ではテストエピソードのみから構成されている。(2) 独立した 10 個のエピソード列の平均を示す。a-1，b-1 の結果が偶然では無いことを示している。

点とした位置座標を，Z 軸には各 XY 平面における $\alpha_j(x)$ の値を示す。図 2.7a-1 は素手，図 2.7a-2 は短い熊手を持った時である。α_j は熊手を持つことによって，熊手の届く範囲にまで正の値を取る領域が拡大した。この結果は，頭頂間溝皮質において熊手の使用と共に視覚受容野が拡大する現象に対応している。

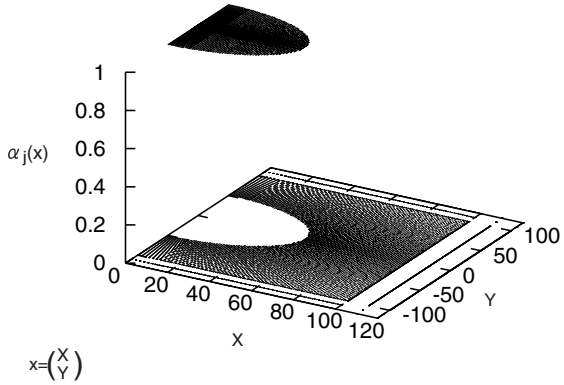
図 2.7b では，10 個の独立したエピソード列について各エピソード毎の α_s の平均の値を示している。横軸はサルからの距離を，縦軸は α_s の値を示している。エピソードと共に，α_s が図 2.4b に示した形に収束していく様子がわかる。

また，最後に図 2.8a，b では，一段階行動課題（3000 エピソード終了後）と二段階行動課題（1000 エピソード終了後）の，1 エピソード内の状態推移毎の状態評価関数の推移を示した。横軸はステップ数，縦軸は状態価値関数値を示している。通常の actor-critic モデルと同様に，報酬に近づくにつて状態評価関数の値が上昇していく。

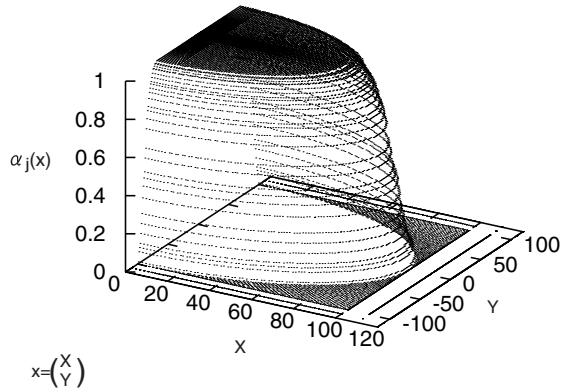
2.4 考察

前頭前背外側部 (Watanabe, 1996) 及び扁桃体 (Nishijo et al., 1988) に，与えられた餌の好みに応じて発火頻度が変化するニューロンが報告されている。本研究は，この実験結果をさらに発展させ，それ自体には報酬をもたない物体に対しても，状況に合わせてその価値を計算する脳内機構を仮定した。この仮定により，本モデルでは，道具を用いて異なる道具を引き寄せるという高度な行動課題を初体験において成功させる事ができた。

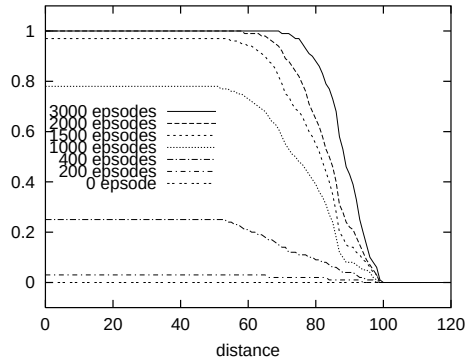
本モデルにおいて，道具の価値を計算する上で重要な役割を持つのが $\alpha_n(x)$，$\alpha_s(x)$，$\alpha_l(x)$，つまり各道具の届く範囲を符号化した関数である。この関数は Iriki らによって報告された頭頂間溝のニューロンに対応している (Iriki et al., 1996)。事実，図 2.7 で示したように $\alpha_j(x)$（$j \in \{n, s, l\}$）は道具を持ったときに視覚受容



(a-1)



(a-2)



(b)

図 2.7: シミュレーション結果：(a)α_j の値（1）素手の状態（2）熊手を持った状態 (b) 各エピソード時の $\alpha_s(x)$ を示す。

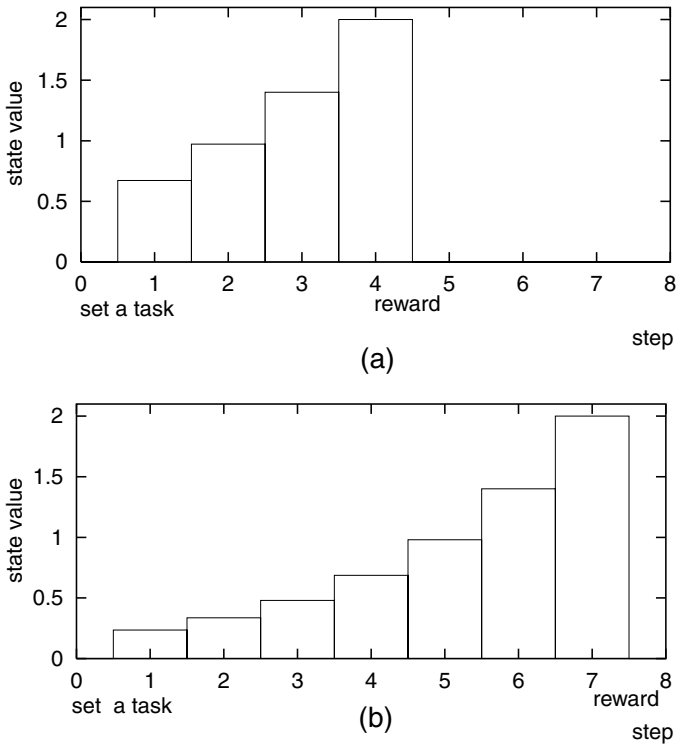


図 2.8: シミュレーション結果（a）一段階行動課題中の状態評価関数値（b）二段階行動課題中の状態評価関数値をそれぞれ示す。

野を熊手の届く範囲に拡大する。本研究の仮定した道具の価値を計算するニューロンは，式から考えて，この頭頂間溝皮質のニューロンからの神経投射を受けているものと考えている。

ただし，$\alpha_j(x)$ は Iriki の報告にある実験結果を全て再現していない。Iriki らの実験では，視覚受容野は単に熊手を持っただけでは拡大せず，道具を使用した時にのみ拡大した。本モデルの $\alpha_j(x)$ はその点は再現出来ていない。しかし，例えば次に定義した関数 $\alpha(x)$ を用いれば Iriki らによって報告されたニューロンと同じ応

答が再現できる。

$$\alpha(x) \quad = \quad \begin{cases} \alpha_n(x) & \cdots j = n \vee (j = s \wedge v_s = 0) \\ & \qquad \vee (j = l \wedge v_l = 0) \\ \alpha_s(x) & \cdots j = s \wedge v_s > 0 \\ \alpha_l(x) & \cdots j = l \wedge v_l > 0 \end{cases} \tag{2.19}$$

式 2.19 の構造から，頭頂間溝で発見されたニューロンは，$\alpha_n(x)$，$\alpha_s(x)$，$\alpha_l(x)$ を符号化したニューロンと，道具の価値を符号化したニューロンの神経投射を受けている可能性がある。現在，$\alpha_n(x)$，$\alpha_s(x)$，$\alpha_l(x)$ と同じ応答を示すニューロンは報告されていないが，その存在を期待している。

次に，今回仮定した道具の価値を符号化したニューロンの存在する領域を予測する。このニューロンは，モデル上の式から，報酬の評価を行う機構と熊手の届く範囲を符号化した機構の神経投射を受ける部位に存在するものと考えている。今までの研究 (Watanabe, 1996; Iriki et al., 1996) から，前者は前頭前野，後者は頭頂葉連合野が候補として挙げられる。この 2 つの領野には直接の相互神経投射経路が存在する事が示されている (Goldman-Rakic, 1988)。この点から，今回仮定したニューロンはこの 2 つののいずれかに存在するものと推定している。

一つの電気生理実験を提案する。図 2.9 に示すように，長い熊手と短い熊手を二段階行動課題と同じ位置に置く。この時サルは台座に固定されている。そして，餌を台の縁，つまり長い熊手を用いても届かない距離から徐々に近づけていく。この時，本モデルで計算した餌，長い熊手，短い熊手の価値の餌の位置毎の値を示したのが図 2.9 である。本研究で提案した仮説が正しければ，この縦軸の価値を発火頻度としたニューロンが存在することが予想される。現段階においてサルが今回提案した通りのアルゴリズム通りに餌を取っている証拠は無い。しかし，図 2.9 で提案した電気生理実験の結果によりその是非がわかると考えている。

最後に本モデルが再現できなかった現象を指摘する。サルは短い熊手を使いこなせるようになると，即座に長い熊手も使えるようになる。しかし，本モデルで

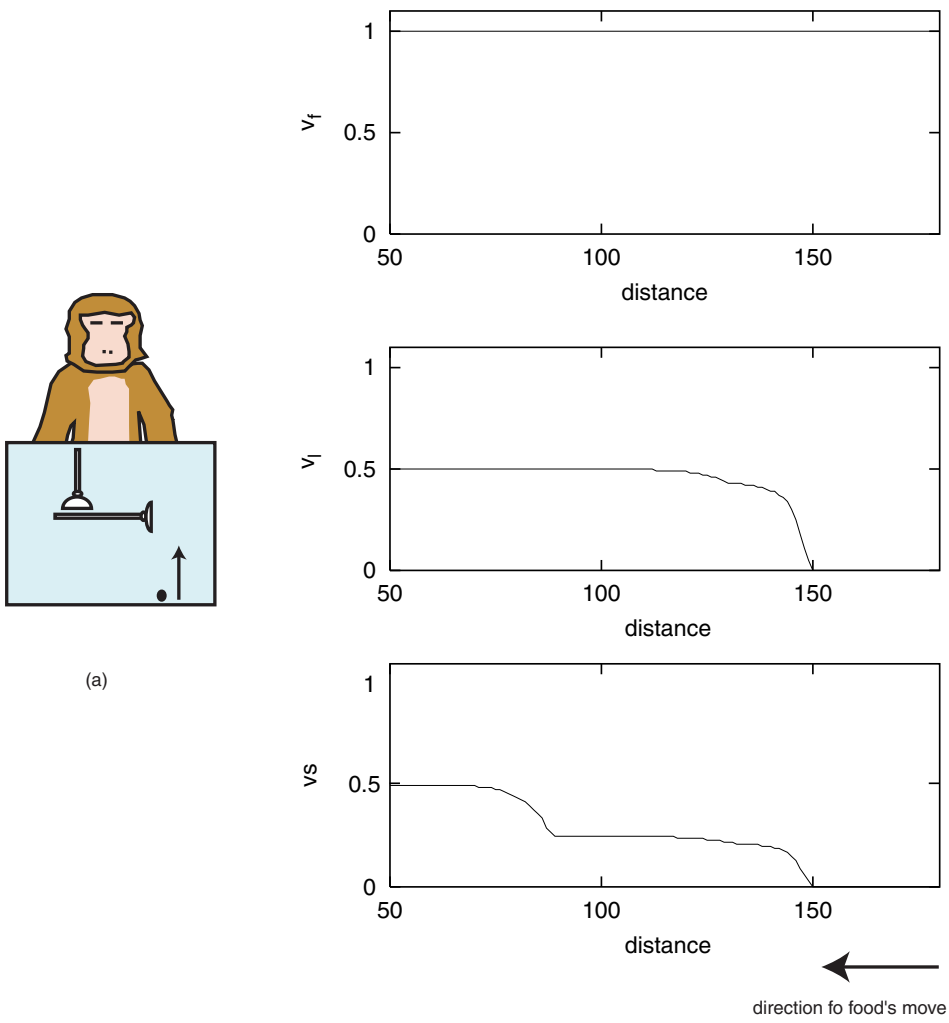


図 2.9: 電気生理実験の提案とその予測。(a) 電気生理実験の概略図：サルは台座に固定されている。二段階行動課題と同様，短い熊手は手の届く範囲に，長い熊手は短い熊手を使えば届く範囲に置く。餌を，長い熊手を使っても届かない台の縁から徐々にサルの方に近づけていく。(b) モデルからの結果予測：図は上から順に，提案した実験中における餌，長い熊手，短い熊手の本モデルから計算した各価値を示した（一段階行動課題 3000 エピソード終了後)。横軸にはサルから各物体までの距離を示している。従って，餌は矢印の方向に進む。本モデルから，縦軸を発火頻度としたニューロンの存在を予測した。

は，短い熊手と長い熊手は全く別の道具をして表記しているので，この点が再現できていない。課題として，この点を再現するようなモデルを構築する点が挙げられる。

謝辞

本研究を行うにあたり，実験データ及び有益なご助言を頂きました，入来篤史先生，石橋英俊先生，日原さやかさんに心より感謝いたします。

第3章　L字型スティック課題を用いたサルの内部モデルの推定方法の提案

3.1　はじめに

第3章では各種のもつ内部モデルを推定するための実験を提案し，ニューラルネットワークを用いてその結果を予測する。内部モデルとは自分の身体や外界の物体の入出力関係を模倣する脳内モデルである。内部モデルは霊長類などの生物が道具使用行動などの難しい課題を解く上で重要な要素の1つとして考えられている。例えば，Imamizuらは，ヒトが新しい道具を使いこなす際の脳内活動をfMRIを用いて測定した (Imamizu et al., 2000)。その結果，内部モデルの存在を示唆する新たな活動パターンが小脳において確認されている。一方，獲得可能な内部モデルは霊長類各種によって異なる可能性もある。Köhlerはチンパンジーに天井にバナナをぶら下げ，2つの箱を積み重ねた上に乗ってバナナを取る課題を課した (Köhler, 1927)。その結果，あるチンパンジーは長期間に渡って，1つの箱の上に，非常に不安定な斜向かいの角度で積み上げようと約1年以上もの間試みた。当然のことながら，箱は崩れ，チンパンジーはバナナを取ることができなかった。我々ヒトから見れば，このような努力は無駄であることが簡単に予想できる。この例のように，ヒトが見れば簡単に予測がつくが，チンパンジーには実際に試しても理解できない事例が数多く報告されている (Köhler, 1927; Beck, 1990)。このよう

な事例は，物理的なダイナミクスの予測性能がヒトとチンパンジーでは異なる可能性を示唆している。この現象を，内部モデルの観点から見た場合，ヒトが獲得している箱の内部モデルの精度が，チンパンジーの獲得した箱の内部モデルの精度よりも高いと解釈することが可能である。では，各種はどのような内部モデルを用いているのだろうか？各種のもつ内部モデルを推定するためには，どのような方法が考えられるのか？

上記の疑問に答えるために，本研究では内部モデルと道具使用行動との関係を調査した。具体的には，異なる内部モデルを持つ2つのエージェントに対し，本研究にて提案する課題を課し，計算機実験結果を行なった。ひとつめのエージェントは報酬量（各課題の成功または失敗）を予測する内部モデルを持ち，その計算を利用して課題を解くエージェントである。 もう一方のエージェントは環境の状態推移の代表値を予測する内部モデルを有し，その計算結果を利用して課題を解くエージェントである。

本研究にて用いる課題は，道具使用行動の代表的な課題の1つである熊手使用行動課題 (Hauser, 1997; Santos et al., 2005) を一部変更した課題である。なお，本研究で用いた課題は，実際の動物実験による検証は成されていない。あくまで架空の実験を想定している。

各エージェントにこの道具使用行動課題に関するシミュレーション実験を行う。課題は2種類から構成され，初めの基本的な課題を十分にこなせるようになった後に，もう一方の課題を課した。その結果，後者のエージェントが前者のエージェントに対してより早く課題環境の変化に対して順応できたことが示された。 もし仮に，上記のような内部モデルを持つ生物に，提案した課題を行なわせれば，計算機実験から予測された結果が期待される。以下，詳細について述べる。

3.2 方法

異なる内部モデルを持つ 2 つのエージェントを用意し，L 字型スティックを用いたコンピュータシミュレーション課題をそれぞれのエージェントに課した。課題は 3.2.1 にて，環境は 3.2.2 にて定義し，エージェントとその内部モデルは 3.2.3 にて定義した。

3.2.1 課題

本研究で使用する課題を説明する。エージェントは，始め図 3.1 A に示すように，左右の 2 つのトレイを同時に提示される。各トレイにはL 字型の熊手と餌が必ず 1 つ置いてある。熊手は垂直方向のみに引き寄せることが出来る。餌は熊手の内側に入っていないと手に入れることができない。左右のトレイのうち一方は必ず餌がとれるように，もう片方は餌が手に入らないように設定している。エージェントは左右のトレイの中から，餌が取れると思われるトレイを 1 つ選択する。選択した後は，左右のトレイをそれぞれ引き，その様子をエージェントは観測する。上記を 1 試行と呼ぶ。本研究では，図 3.1A に示された 6 つの課題を 1 試行ずつ順番におこない，これを 1 セッションと定義する。タスク 1 が十分に正解できるようになった後，図 3.1B に示す課題をおこなう。これをタスク 2 と呼ぶ。タスク 2 では，この 2 つの課題を 1 試行づつ順に行うことを 1 セッションと定義する。この課題は Hauser の研究を元に考案した (Hauser, 1997)。Hauser はワタボウシタマリンに上記に述べたのとほぼ同じタスクを行わせ，タマリンがこのタスクを成功させることを示した。Hauser はタスク 1 とタスク 2 を同時に行ったが，本研究では分けて順番に行った点が異なる。異なる課題を用意した理由は，事前に Hauser の課題をモデル 1，モデル 2 にて行わせたところ，内部モデルによる顕著な差が行動に現れなかったためである。

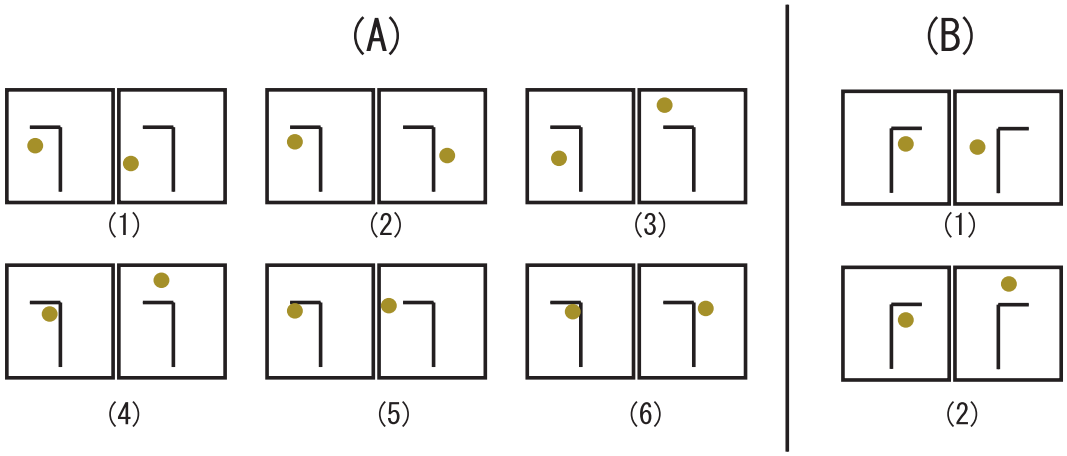


図 3.1: それぞれ (A) タスク 1(B) タスク 2 の模式図。タスク 1 が十分に成功できるようになった後に，タスク 2 を開始する。タスク 2 はタスク 1 と比較して，熊手の先端部分が反対になっている。

3.2.2 環境

環境は，有限オートマトン $E = (E_K, E_\Sigma, E_\delta, E_q, E_f)$ によって表現する。それぞれは，環境状態，アルファベット，遷移関数，初期状態，終了状態を示す。それぞれの詳細を以下に示す。

環境状態

環境状態 E_K を以下に定義する。各変数は，トレイ上の餌，熊手の左端，右端の位置を示している (図 3.2)。添え字の R と L は右と左のトレイを意味している。X 軸を図中の横方向，Y 軸を縦方向とし，F_X,F_Y は L 字型熊手の手持ち部分の先端を原点とした X 座標，Y 座標，T_{XR}, T_{XL} は熊手先端部の右端，左端の X 座標を，T_Y は熊手先端部の Y 座標をそれぞれ示す。

以後，$S* = (F^*_X, F^*_Y, T^*_{XR}, T^*_{XL}, T^*_Y), * \in R, L$ と定義する。

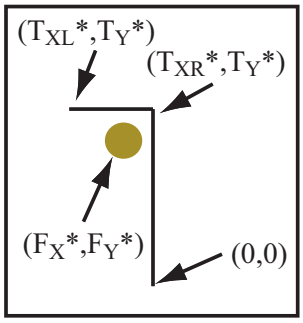


図 3.2: 環境の状態

アルファベット

アルファベットはエージェントから環境への入力を示す。$E_\Sigma \in \{A_R, A_L\}$ と定義する。ここで A_R と A_L はそれぞれエージェントが右のトレイを，左のトレイを引く行動を意味している。

状態遷移関数

状態遷移関数は，エージェントからの入力に応じた，環境の状態遷移法則を示している。今回は離散時間 $t = 0, 1, 2, \cdots$ を導入した。各課題は $t = 0$ から開始する。各変数は時刻 t と共に変化する。例えば $S^R(t)$ とは，時刻 t における右トレイ上の餌や熊手の各位置を意味している。

エージェントが行動 A_R を選択した後に，右トレイ上の餌と熊の位置は以下の通り変化する：

$$T_Y^R(t+1) = \begin{cases} T_Y^R(t) - 1 & \text{if } 1 \leq T_Y^R(t) \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

$$F_Y^R(t+1) = \begin{cases} T_Y^R(t+1) & \text{if } T_{XL}^R(t) \leq F_X^R(t) \leq T_{XR}^R(t) \\ & \text{and} T_Y^R(t) - 1 \leq F_Y^R(t) \leq T_Y^R(t) \\ F_Y^R(t) & \text{otherwise.} \end{cases} \tag{3.2}$$

なお，他のパラメータは変化しない。もしエージェントが A_L を選択した場合には, 上述と同様の方法で左トレイ上の餌と熊手が変化する。

初期状態

初期状態は，図 3.1 によって与えられる熊手と餌の状態に基づいて与えられる。$T_Y^* = 2$ とし， タスク 1（タスク 2）ではそれぞれ $T_{XL}^* = -1(or0), T_{XR}^* = 0(or1)$ とした。

終了状態

終了状態は，$T_Y^R = 0$ または，$T_Y^L = 0$ を満たす状態とする。

3.2.3 エージェント

課題開始時（つまり $t = 0$ において） エージェントは環境状態を観測し，それぞれの熊手を引いた後トレイがどのように変化するのかを予測する。その予測結果に基づき，エージェントはより報酬が得ることができそうなトレイを選択する。(今回は $t = 0$ にエージェントが選択した行動は，その後の時刻 $t = 1, 2, \cdots$ においても変わらないものとした。)。本研究では，異なる内部モデルを用いて将来を予測する 2 種類のエージェントを用意した。1 つ目のエージェント（以後エージェント 1 と呼ぶ）は以後モデル 1 と呼ぶ内部モデルを持つ。モデル 1 は，報酬の有無を予測する，つまり熊手を用いて餌をとることができるのか否かを予測するモデルである。もう一方のエージェント（以後エージェント 2 と呼ぶ）はモデル 1 とは異なるモデル 2 という内部モデルを用いる。モデル 2 は状態推移の時刻ごとの遷移状態を予測する。本研究では内部モデルにニューラルネットワークモデル (Ramelhart et al., 1986) を用いた。ニューラルネットワークの構造はモデル 1，モデル 2 共に同じであり，教師信号のみが異なる。詳細は以下にて述べる。

モデル　1

本研究では，3 層構造（入力層 5 つ，中間層 5 つ，出力層を 1 つ）のニューラルネットワークを用いた。各ニューロンの入出力特性として，シグモイド関数を用いた。つまり，$f(x) = 1/\{1 + exp(-x)\}$ となる。ここで x は入力層への入力信号を示す。入力層 i 番目のニューロンへの入力信号を x_i，入力層 i 番目のニューロンと中間層 j 番目のニューロンとの結合強度を w_{ij}，中間層 j 番目のニューロンと出力層のニューロンの結合強度を v_j とする。この時，出力層の出力 z は $z = f(\Sigma v_j f(\Sigma w_{ij} f(x_i)))$ となる。モデル 1 において，ニューラルネットワークは，熊手を引き寄せた際の報酬（餌の取得の有無）を予測できるように学習を行う。$S^*(0)$ は課題開始時における熊手と餌の位置を示している（ただし，$* \in \{R, L\}$）。ここで，$O(S^*(0))$ を入力 $S^*(0)$ におけるネットワークの出力とし，O^* を教師信号として報酬の有無を表す（つまり，報酬時には $O^* = 1$，無報酬時には $O^* = 0$）。なおニューラルネットワークの学習には BP 法を用いた。

各試行における，課題開始時において，$O(S^*(0))$ を計算することにより，ニューラルネットワークは各トレイにおける報酬の有無を予測する。なぜならば，このネットワークの教師信号は報酬の有無を示しており，この学習が進めば，餌の取れるトレイの状態を入力すれば 1 を，餌の取れないトレイの状態を入力すれば 0 を，このネットワークがそれぞれ出力することが予測されるからである。エージェントはこの計算を左右の両方のトレイに適応し，その計算結果から，報酬量の大きいと予測されるトレイを選択する。上記の方法を実現する手法は様々考えられる（例えば，Soft-Max 法や，ϵ-greedy など）が，以下の確率でエージェントは右側のトレイを選択するようにした。

$$P(R) = \frac{O(S^R(0))}{O(S^R(0)) + O(S^L(0))} \tag{3.3}$$

モデル　2

モデル 2 はモデル 1 と同じ構造のニューラルネットワークを用いる。ただし，モデル 2 は，環境の各時刻における状態遷移を予測する。時刻 t において，モデルは時刻 t+1 における各トレイの状況を予測する。モデルの単純化のため，モデルは各時刻において，熊手が餌と接触するのか否かを予測させた。ネットワークへの入力は $S^*(t)$，学習信号 $O^*(t)$ は $T^R_{XL}(t) \leq F^R_X(t) \leq T^R_{XR}(t)$ かつ $T^R_Y(t) - 1 \leq F^R_Y(t) \leq T^R_Y(t)$ の場合は 1 を，そうでない場合には 0 を与える。注意しなければならない点が，モデル 2 は一試行中に $t = 0$， $t = 1$ の 2 回学習を行う点である。今回用意した環境では熊手の長さが常に 2 であり，手前側に引くまでに常に 2 単位時間かかる。エージェントはこの学習結果から，熊手が餌と接触すると思われるトレイを選択するようにした。この実現方法として，以下の計算式に基づいてエージェントは右側のトレイを選択するようにした。

$$P(R) = \frac{\Pi O(S^R(t))}{\Pi O(S^R(t)) + \Pi O(S^L(t))} \tag{3.4}$$

3.3　結果

図 3.3A にタスク 1 の結果を示す。タスク 1 に対し，それぞれのモデルで 5 試行おこない，平均値を求めた。細い線がモデル 1 の結果を，太い線がモデル 2 の結果を示す。縦軸には，餌をとることが可能なトレイを選んだ割合を示している。正解のトレイを選ぶ割合が 0.9 を超えるのに，モデル 1 は 3,830 セッション（22,980 学習回数），モデル 2 は 1,848 セッション（22,176 学習回数）であった。

ここで学習回数とは，ニューラルネットワークが入力信号と学習信号を得た後，学習を行った回数を示している。なお，前節で述べたとおり，モデル 2 はモデル 1 と比較して，同じ試行数においても 2 倍の学習回数となる。セッション数でみれ

ば，モデル 2 は半分程度のセッション数で正解を選べるようになったが，学習回数でみるとモデル 1 とモデル 2 では，ほとんど変わらない結果となった。

タスク 1 を十分こなせるように学習した後（それぞれ 20,000 セッションおこなった），タスク 2 をおこなった。その結果が図 3.3B である。細い線がモデル 1 を太い線はモデル 2 の結果を示している。正解のトレイを 0.9 以上の割合で選択できるようになるまで，モデル 1 では 2,222 セッション (4,444 学習回数) かかったのに対し，モデル 2 では 20 セッション (80 学習回数) で行えるようになった。

なお，タスク 1 のセッション数はモデル 1，モデル 2 共に約 20,000 セッションである。一方，学習回数はそれぞれ 120,000 回，240,000 回である。つまり，モデル 2 の学習回数がモデル 1 のそれよりも倍となる。その差がタスク 2 の結果に影響を及ぼす懸念を排除するため，モデル 2 にてタスク 1 を 10,000 セッションした後タスク 2 を行う検証実験も行ったが，上記と変わらない結果であった (ここには示していない)。

以上の通り，エージェント 2 がタスク 2 において，エージェント 1 よりも早く正解のトレイを選択する結果となった。この結果の理由を調べるため，結合強度の変化を図 3.3C，D に示した。これらの図は，タスク 2 においてのモデル 1 とモデル 2 の結合強度の変化量を示している。黒棒はタスク開始時から 30 セッションにおける，変化量の絶対値の総和を，白抜き棒は 30 セッション後とタスク開始時における結合強度の差を示している。図 3.3C はモデル 1 を 3.3D はモデル 2 をそれぞれ示している。モデル 1 において，黒棒は白棒よりも大きな値を示している。これは，各結合強度の変化量は多いが，その多くは，＋と－の値を繰り返しており，絶対値としての変化量は少なく，無駄の多い学習を行っていたためである。一方モデル 2 は，黒棒と白棒の値がほとんど変わっておらず，無駄のない学習を行っていた。そのため，結果として，エージェント 2 はタスク 2 に早く順応していた。

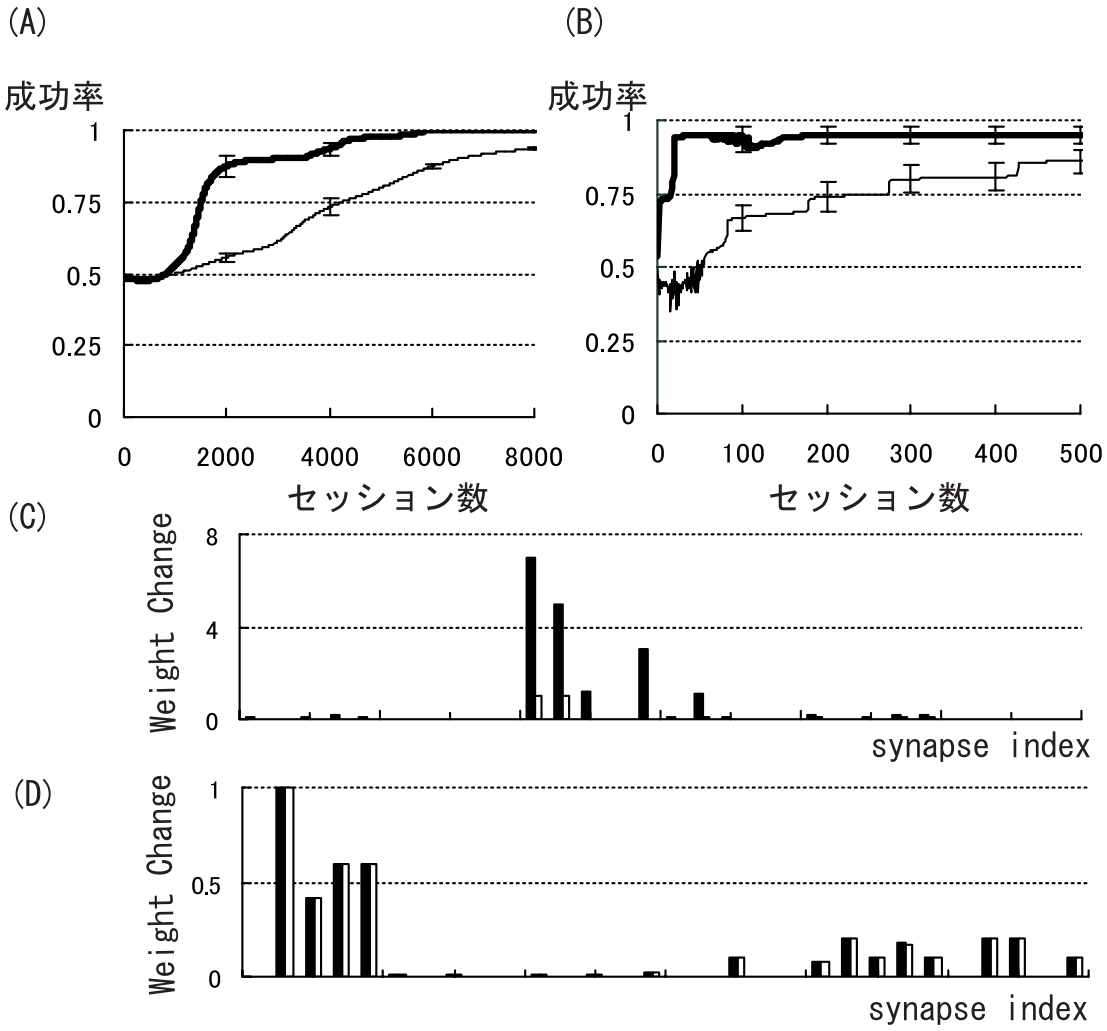


図 3.3: 実験結果：各エージェントの学習曲線 (A) 課題 1，(B) 課題 2。細線はエージェント 1，太い線はエージェント 2 の結果をそれぞれ示している。 横軸はセッション数を，縦軸は正解のトレイを選択した割合を示している。エージェント 2 はエージェント 1 よりも格段に早く課題 2 を成功させている。(C) と (D) は，それぞれモデル 1，モデル 2 の課題 2 開始 30 セッションにおける結合強度変化量を示している。黒棒と白棒はそれぞれ，変化量の絶対値の総和と 30 セッション後と課題開始時の差を示している。結合強度の変化はモデル 2 では効果的に変化しているが，モデル 1 では効果的に変化していないことがこの図からわかる。

3.4 考察

本研究では，L字型熊手という道具使用行動の中では比較的オーソドックスな課題を対象に， エージェント1；結果しか覚えないエージェント，エージェント2；環境のダイナミクス（餌とスティックの接触情報）を覚えるエージェントの2タイプを用意した。その結果，課題1にて熊手を固定して餌の位置だけかえた課題を学習した後，課題2において熊手の向きを反転した課題では，ダイナミクスを学習したエージェント2は20セッション程度の学習期間で正解を高確率で選択できるのに対し，結果を覚えるだけのエージェント1は，約100倍のセッション数を要することになった。

この原因は，学習を行っていた関数の差にあると推測される。モデル2は環境のダイナミクスを観測することにより，環境の遷移関数とほぼ同じ学習を行っていた。環境の遷移関数はタスク1と2で餌が熊手の先端部の直下にあれば熊手が餌とぶつかることが共通であり，この性質をエージェントは課題1で学んでいたため， 結果，課題2へ移行した際に結合強度の変更が少なくすんだ可能性がある。一方，モデル1は環境の遷移関数とは別のタスク1に特化した関数を覚えていた可能性があり，タスク2へ移行した際に結合強度を大幅に変更した可能性が高い。遷移関数を直接学習しているモデル2がモデル1よりも早くからタスク2をこなせるのはほぼ予想通りである。

この結果を支持するようなタマリンによる実験結果も報告されている。Santosらは今回のタスク1とタスク2を同じセッションにて行う課題をワタボウシタマリンに行わせ，その後熊手の条件を変えながらその行動様式を報告した(Santos et al., 2005)。その中で，熊手の柄の部分が2つの道具から構成され，片方は柄を引けば餌がとれるが，片方は柄部分で分離し餌をとれないような課題を行わせたところ，タマリンは長期間にわたって学習しなければ基準の正解割合をクリアする

には至らなかった。タマリンが，もしヒトのようにダイナミクスを理解していれば，ヒトがそうであるようにタスクを移行した際にすぐに順応できたはずである。このことは，今回のシミュレーション実験のように，タマリンがダイナミクスとは異なるタスクに特化した関数をなんらかの形で覚えていた可能性がある。

まとめると，本研究により，エージェントの持つ内部モデルの差，つまり，タスクの成功・失敗を予測する内部モデルと，ダイナミクスを予測する内部モデルの差が，異なった場面での道具使用行動の汎化能力の差に現れる現象を具体的に示した。今後の展開として，以下のことがあげられる。

第 1 に，本研究で報告した結果が発生する必要条件や十分条件を理論的に明らかにすることが必要である。例えば，本研究では熊手を反転するタスクを用いたが，他のタスクでどうなるかはまだ明らかではない。また，本研究では 3 層構造のニューラルネットワークに，バックプロパゲーションという，シンプルな構成のニューラルネットワークを用いることとした。これは，ネットワークの形態・学習法則にとらわれずに知りえる情報の差により行動の差を出したかったため，なるべくオーソドックスなモデルを用いたかったためである。神経系の研究において，脳内でバックプロパゲーションを用いているという結果はほとんどない。ニューロン同士の結合強化法則は Hebb 則に従うのが通例である。今回の結果は 3 層構造 BP 学習に依存した結果ではないと考えている。しかし，どのようなモデルならば同じ結果になるのかを明らかにする必要がある。

第 2 に，今回用いた以外の内部モデルについても調べる必要がある。今回はモデル 1，モデル 2 はそれぞれ極端な例であった。特にモデル 1 は一切ダイナミクスを学習しないエージェントを仮定している。しかし，実際のサルもある程度のダイナミクスは理解できる可能性はある。問題はどの程度までダイナミクスを学習できるかである。その点を考慮した学習信号を考えていく必要がある。チンパン

ジーの模倣に関する研究では，模倣相手のダイナミクスを学習しているのか，それとも道具のダイナミクスを学習しているのかを議論する場合が多い (Whiten et al., 2004)。今後この要素を組み込んだシミュレーション実験をすることにより，霊長類の持つ内部モデルについて理論的に議論することが可能になると思われる。

最後に，モデルの結果と行動実験結果を比較して議論する必要がある。現段階では本研究で仮想したタスクに対してシミュレーションを行ったにすぎず，実際の行動研究での対比が行われてはいない。実際に実験にて行われているタスクに対して，シミュレーションを行い，その結果から，例えばチンパンジーなどの認知能力に関して考察する必要がある。模倣に関する研究や，道具作成行動に関する研究に関して，内部モデルの構造により大きな差がでることが予想される。

このように，より研究を進めていく必要があるが，本研究のような研究が進めば，アルゴリズムと道具使用行動との関係が明らかになり，各霊長類がどのアルゴリズムを実際に獲得しているのかを推定できるようになる。現在，言語をもたない動物に，行動決定のアルゴリズムを実験的に決定することは非常に困難である。そのため行動の差を観察してもその差がなぜ生まれたかを知ることが難しい。それに対して本研究のような理論研究が大きな貢献を果たすと思われる。

謝辞

有益なコメントを頂きました査読者の方々に心から感謝いたします。 本研究の一部は科研費 (19700215) の補助を受けました。

第4章　ワタボウシタマリンの道具使用行動を表した数学的モデル

4.1　はじめに

第 4 章では，道具使用行動の数学的モデルを構築し，Santos らによって行われたワタボウシタマリンの道具使用行動実験 (Santos et al., 2005; 2006) を再現した。これによりタマリンが道具使用行動時に用いる行動決定アルゴリズムを推定した。

道具使用行動の数学的モデルを作成する上で，本研究は，道具の内部モデルの種類と，複数の道具を系列的に使用する際に内部モデルを用いて何手先の道具の動きまで予測できるかに着目した。

内部モデルとは制御対象の脳内シミュレーターであり（Wolpert & Kawato, 1998; Kawato, 1999; Grush, 2004），数多くの研究が内部モデルを生物の脳が獲得している可能性を支持している (Kawato et al., 1987; Shadmehr et al., 1994; Imamizu et al., 2000)。

内部モデルは順モデルと逆モデルに分けることができる (Kawato et al., 1999; Desnurget & Grafton, 2000; Grush, 2004)。順モデルは制御対象の入力から出力を予測する内部モデルである。被験者がある道具を用いて他の物体を操作した際に，次の瞬間の操作された物体の動きを予測するモデルを順モデルと本研究では定義する。遠方にある餌を熊手を用いて引き寄せる課題を例にする。手元にある熊手を用いて，餌を引き寄せるという行動に対して，餌の動きを予測するのが順モデルである。脳が順モデルを獲得している可能性は様々な研究から指摘されている

(Kosslyn et al., 2001; Hesslow, 2002)。

一方，逆モデルは制御対象の出力から入力を予測する内部モデルである。ある物体を特定の形状や状態に変化させる場合，十分となる道具の状態と行動を計算するモデルを逆モデルと本研究では定義する。先の熊手課題でいうならば，餌が手前側に引き寄せられる動きになるためには,「熊手が餌の近くにあること」と「引き寄せる」という行動を計算するのが逆モデルである。ヒトの運動制御に関する研究では，小脳において逆モデルが計算されている可能性が指摘されており，逆モデルを脳は獲得している可能性はある (Kawato et al., 1987; Wolpert & Kawato, 1998)。

内部モデルを用いて予測可能な道具の動きの数に着目した理由は，先に示した Matsuzawa の主張のように，課題中に使用可能な道具の数自体が知能の尺度として重要な指標となりうるためである。また，何手先まで行動計画を行えるのかについては，道具使用行動課題以外でも知能の尺度として用いられている。Lusiana らはロンドン塔課題を 4 歳〜8 歳までのヒトの子供に行わせた。2 手で正解できる課題では，課題達成までに要する時間は年齢に関して有意な差は無かった。3 手以上かかる課題では年齢に応じて有意な差が生じることがわかった (Lusiana & Nelson, 1998)。このように知能レベルを分類する上で，課題解決の手順数は有効な指標となりうる。

本モデルの学習機構としては，強化学習で提案されている actor-critic モデルを用いた (Sutton & Barto, 1998)。actor-critic モデルは報酬を手がかりとした学習方法の 1 つである。この actor-critic モデルで用いられている変数の 1 つが，脳内のドーパミンの放出量を表現しているとする報告が多数ある (例えば Schultz et al., 1997)。そのため，actor-critic モデルはサルやラットなどの動物の脳内で用いられていると考えられている有力なモデルの 1 つとなった (Barto, 1995; Montague et

al., 1996; Schultz et al., 1997; Suri & Schultz, 2001)。本研究でも，タマリンの行動のうち学習の部分についてはactor-criticモデルを用いてモデル化した。

本研究は内部モデルとして順モデルを持つエージェント（仮想的行動主体）と逆モデルを持つエージェントを用意した。また，各内部モデルに対して連続して使用できる道具の数を変えたエージェントも用意した。各エージェントはこれらのモデルが有効でないときactor-criticモデルに基づく学習を行うとした。なお，本研究では内部モデルは学習を行わない。各エージェントにSantosらによって行なわれた行動実験と同じ実験環境をコンピュータ上に再現し，シミュレーション実験を行なわせた。そのシミュレーション実験結果とタマリンの行動実験結果とを比較した。もしタマリンの行動実験結果と一致するエージェントが存在すれば，タマリンはそのエージェントが用いるアルゴリズムを使用している可能性がある。実験の結果，タマリンは逆モデルを用いて1つの道具の動きを予測する能力がある可能性を示唆する結果を得た。また，再現したモデルを用いて，タマリンの持つ能力を推定した。以下に詳細を報告する。

4.2 方法

4.2.1 先行研究

本研究の再現対象となるタマリンの2つの実験を先行研究として紹介する (Santos et al., 2005; Santos et al., 2006)。

タマリンは野生下では道具を使用しない種だと考えられているが，訓練により簡単な道具使用行動の獲得が可能であると指摘されている (Hauser, 1997; Hauser et al., 1999)。Santosらはタマリンに2つの熊手を用いて餌を取る課題を課した (Santos et al., 2005)。図4.1Aにその課題を示す。課題開始時，タマリンには左右2つのトレイが同時に提示された。トレイ上には熊手と餌が置いてある。片方のト

レイは垂直方向に熊手を引くことにより餌が取れ，もう片方は熊手を動かしても取れない。タマリンは左右どちらか一方の熊手を選択しなければならない。餌を取ることが可能なトレイを選択した場合，タマリンは餌が取れるまで自由に熊手を動かすことが出来る。一方，そうでないトレイを選択した場合には，タマリンが熊手を動かした後トレイを取り上げる。この場合，タマリンは餌を取ることができない。トレイを提示してから，熊手を選択し餌を取る（または失敗する）までを 1 試行とよび，20 試行を 1 セッションと定義する。

タマリンには，1 本の熊手を用いて餌をとる練習課題を通じて，熊手の基本的な使い方を覚えた後，図 4.1A に示された計 8 種類の課題を行なわせた。8 種類の課題は用いる道具の種類に応じて H-C(hook-to-cane) 問題と H-P (hook-to-post) 問題にそれぞれ分類される。各課題は 1 セッションにおける課題正解率が 2 セッション連続して 90% を越えるまで行なわれ，2 セッション連続して 90% を越えると次の課題に移る。課題の順番は H-C 問題（または，H-P 問題）の課題 1 から始まり，課題 2 → 課題 3 → 課題 4 と推移する。その後 H-P 問題（H-C 問題）の課題 1 から始まり課題 2 → 課題 3 → 課題 4，最後に H-C 問題と H-P 問題を複合させた最終テストを行なわせた。H-C 問題，H-P 問題のどちらから先に始めるのかは，個体に応じて変えておりバランスよく行なっている。

各課題について詳細に述べる。H-C 問題の課題 1 では両方のトレイに道具は 1 つしかなく，タマリンは餌が熊手の中にあるのか，外にあるのかだけ判定すればよい。課題 2 ではトレイ上に 2 つの道具を用意し (例外として，餌の取れないトレイでは 1 つの道具の場合もある。)，片方のトレイでは餌は必ず熊手の中にあり，もう片方のトレイでは餌は常に熊手の外にある。手前側の熊手は餌側の熊手に引っかかっているので，タマリンは餌側にある熊手の中に餌があるのかどうかを判断すれば正解が得られる。そのため，課題 2 は，課題 1 と同じ判断基準にてトレイを

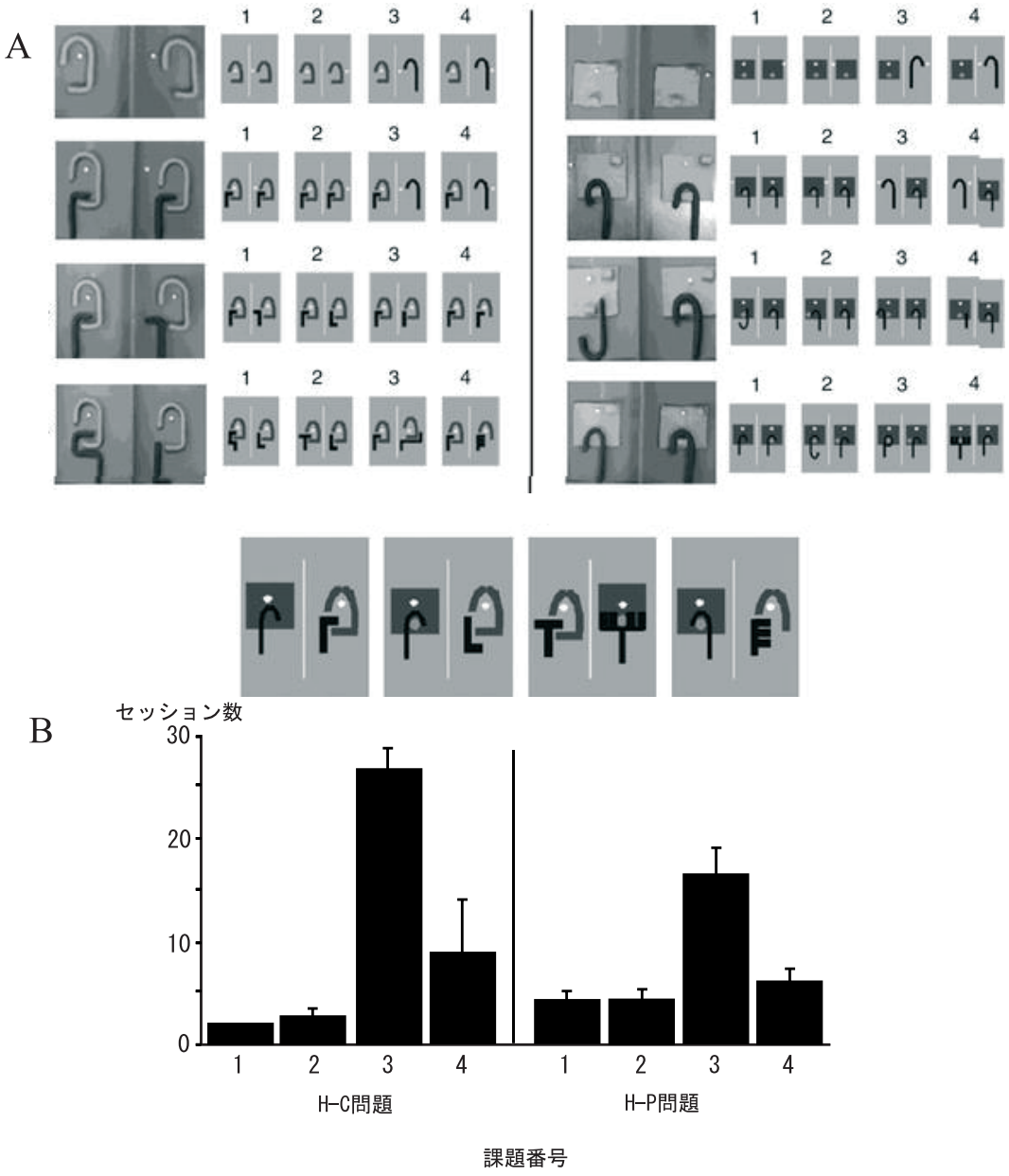


図 4.1: タマリンの実験課題とその結果：実験 1 (Santos et al. 2005 より抜粋) A) 実験の課題設定を示す。タマリンは熊手の基本的な操作が出来るようになった後に，H-C 問題 (H-P 問題) の課題 1 →課題 2 →課題 3 →課題 4 → H-P 問題 (H-C 問題) の課題 1 →課題 2 →課題 3 →課題 4 →最終テストの順に課題を行う。H-P 問題と H-C 問題のどちらを始めに行なうのかは個体によって異なり，バランスよく行なわれている。図には各課題について，4 種類の課題設定とそのうちの一つの実写真が示されている。図中，白い丸状の物体が餌を示している。B) 図中の縦軸はタマリンが各課題に対し 90% 以上の確率で連続して正解するまでに要したセッション数を示している。

選択しても，正解のトレイを選択することが可能である。課題 3 では，課題 2 とは異なり，どちらのトレイでも餌は熊手の中にある。片方のトレイでは，手前側の熊手で餌側の熊手を引き寄せられる。もう片方のトレイでは手前側の熊手を引いても，餌側の熊手を引き寄せることが出来ない。そのため，タマリンは手前側の熊手で餌側にある熊手を取ることが可能かどうかを判断しなければ，正解を得ることができない。なお，タマリンは熊手を引き寄せることは可能であるが，熊手をひっくり返すことは出来ない。課題 4 は課題 3 と同様に手前側の熊手で餌側の熊手を引き寄せることが可能かどうかを調べるタスクである。課題 3 と比べて，道具のバリエーションが豊富な点が異なる。

H-P 問題では，H-C 問題とは異なり，フック付きの板が道具として登場する。課題 1 はフック付きの板の上に餌が乗っているのか否かをタマリンが判定する課題である。課題 2 は H-C 問題の時と同様に課題 1 と同じ判定基準（餌が板の上に乗っているか否か）を用いれば餌のトレイを選択できるようになっている。課題 3 は，熊手がフック付きの板を動かすことが出来るかどうかを判定する課題である。課題 4 は課題 3 のバリエーションを増やした課題である。

上述のように H-P 問題と H-C 問題はどちらも，課題 1 は基本的な 1 つの道具の課題，課題 2 は課題 1 と同様の基準で正解が選べる課題，課題 3 は餌を取るためには課題 1，2 とは異なる基準を用いる必要のある課題，課題 4 は課題 3 のバリエーションを増やした課題である。

各課題には図 4.1A にあるように課題設定が 4 種類用意されており，各試行で用いられる課題設定はこの 4 種類の中からランダムに選ばれる。図 4.1B には各課題に対し，8 頭のタマリンが連続して正答率 90% に達するまでに必要としたセッション数の平均が掲載されている。検定の結果，H-C，H-P の両方の問題において，課題 1 と課題 2 の間にはセッション数に有意差がなく，課題 1, 2 に対して課題 3 は

有意に多く，課題 4 は課題 3 に関して有意に少ない結果となっている。また最終テストでは，1 セッション目での課題遂行成功率が 90 %を示していた。Santos らは，熊手の基本的な使用方法を訓練した後， 図 4.2 に示すような一部で分断させた熊手を用意し，タマリンが左右のトレイの中からどちらか一方のトレイを選ぶ課題を行った (Santos et al., 2006)。各課題において，以下の 4 種類の熊手を用意した。なお，タマリンが動ける範囲は決まっており，手の届く範囲にある道具のみを直接操作することが出来る。1）通常の熊手。2）根元の部分で 2 つに分断された熊手。ただしどちらの部分もタマリンから手の届く位置にある。餌側の熊手を使えば餌を取ることが可能である。3）先端部分が分断された熊手。これは先端部分にはタマリンは手が届かない。手前にある熊手を単純に引いただけでは餌を取ることができないが，熊手の引き方に工夫をこらせば餌が取ることができる。4）柄の中央が分断された熊手。餌側には手が届かないので，このトレイでは餌を取ることはできない。以上の 4 つの熊手を用いて，図 4.2 に示した計 6 種類の課題を用意した。課題 1 はすでにトレーニング済みの課題で，タマリンの課題に対するモチベーションを保つために用いた。課題 2〜4 は通常の熊手と分断された熊手のどちらを選択するのかを調べるための課題である。課題 5,6 は分断された熊手のどちらを好んで選ぶのかを調べるための課題である。以上の各課題に対し各 4 試行，合計 24 試行をタマリンに課した。結果は図 4.2 に示されている (課題 1 の結果はトレーニングで既に正解を得られることが確認されているため原典においても掲載されていない)。検定の結果，課題 2〜6 の全てにおいて，タマリンの選択は，ランダムな選択と有意差が無いことが示された。

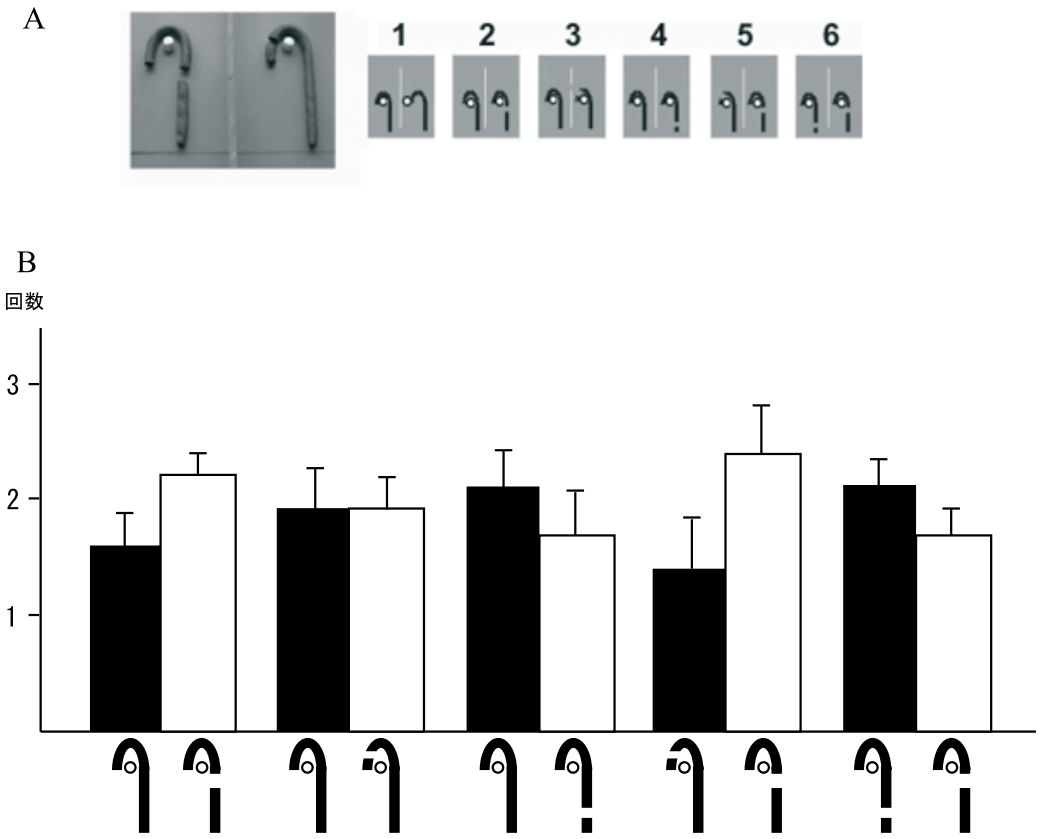


図 4.2: タマリンの実験課題とその結果：実験 2(Santos et al., 2006 より抜粋) 上の図は課題を，下の図はその結果をそれぞれ示している。タマリンは上図の 6 種類の課題をそれぞれ 4 試行（計 24 試行）行った。下図は各課題において 4 回中に左右のトレイをそれぞれ選択した回数を示している。課題 2〜6 に関しては左右の選択に有意な差はなかった。課題 1 は，モチベーションを保つための課題のため，論文中に結果は記載されていない。

4.2.2 実験方法

本研究では上述した 2 つの先行研究と同じ実験条件を計算機上に作成した。本研究では図 4.1 に示した課題 (Santos et al., 2005) の計算機実験を実験 1，図 4.2 に示した課題 (Santos et al., 2006) の計算機実験を実験 2 と呼ぶ。

実験 1 では図 4.1A に示された H-C 問題（H-P 問題）を課題 1 から順に行い，課題 4 終了後に H-P 問題（H-C 問題）を課題 1 から開始し，課題 4 終了後に最終テストを行う。H-C 問題，H-P 問題のどちらから先に始めるのかは，個体に応じて変えておりバランスよく行った。1 試行は左右のトレイが提示された時に開始とし，左右のトレイを選択して餌を得た（または得ることができない）時をもって終了とする。20 試行を 1 セッションと呼び，2 セッション連続して 90% 以上の確率で餌を取ることができれば，次の課題に移る。これらの手順は Santos らの実験に準じている。この課題を後述する 5 つのアルゴリズムで記述されたエージェントに行わせ，タマリンの結果と比較する。本研究のねらいは，様々なアルゴリズムを持つエージェントを用意し，各エージェントに Santos らと同じ課題を行なわせた際，どのエージェントが Santos らが示したタマリンの結果を再現できるかを調べることである。もし，あるエージェントのシミュレーション実験結果と Santos らの実験結果が一致すれば，そのエージェントが使用しているアルゴリズムをタマリンが用いている可能性がある。

実験 2 では図 4.2 で示したように，合計 6 つの課題に対してそれぞれ 4 試行を行わせる。なお 1 試行の定義は実験 1 と同様である。実験 1 同様，この課題を後述する 5 つのエージェントに行わせタマリンの結果と比較する。次に本研究で用いる数学的モデルについて説明する。

4.2.3 数学的モデル

環境

まず，Santos らの実験をシミュレーション上で再現するための，環境記述方法について説明する。環境は有限オートマトンで記述した。環境状態を E と表記し，$E = (E_{Left}, E_{Right})$ とし，各要素は，左右のトレイの状態を表した。なお，

$$E* = (food^*_x, food^*_y, tool1^*_x, tool1^*_y, tool1^*_{type}, tool2^*_x, tool2^*_y, tool2^*_{type})$$

$$* \in \{Left, Right\}$$

とし，餌の x 座標，y 座標，手前側の道具の下端の x 座標，y 座標，道具の種類，遠方の（餌側の）道具の下端の x 座標，y 座標，道具の種類をそれぞれ表す。H-C 問題，課題 1 の 1，左トレイでは，熊手の手前側左端部分が置かれている位置を原点とした。他のトレイでも，トレイ上のこの位置を原点とした。また，エージェント（タマリン）から見てトレイの水平方向を x 軸，奥行き方向を y 軸とした。$tool1^*_{type}$， $tool2^*_{type}$ は道具の種類を自然数で表し，形状の異なる道具は異なる種類とした。例えば図 4.1H-P 問題の，課題 1 の 3 と 4 の右トレイのように，同じ形状の道具でも左右反転のように方向の異なる道具は異なる種類の道具として扱った。なお，実験 1，実験 2 を通じて 26 種類の道具が存在する。

行動 $A \in \{L, R\}$ とし，それぞれ左のトレイを選択する，右のトレイを選択するという意味である。環境状態 E の遷移図を図 4.3 に示す。各エージェントは環境状態 E を観測し行動 A を計算する。この図に示すとおり，環境遷移則は道具の具体的な軌跡ではなく，熊手を引き終えた時点での各トレイの状態と報酬の有無を求めるようにしている。これは，後述する実験結果と各エージェントの学習方法からわかるように，各物体の軌跡を表現することは Santos らの実験結果を再現する上で必要がなく，選択した行動の結果，報酬が得られたか否かがエージェントの学習

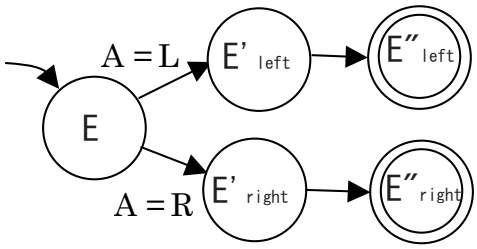


図 4.3: ○は状態，◎は終了状態を表す。E からスタートし，A = L の時は E 'left へ，A = R のときは E'_{right} へ遷移する。E'^{*}，$* \in \{Left, Right\}$ は E^{*} の状態から手元にある熊手を引き寄せた際のトレイ上の状態を示している。どちらのトレイを選んでも，一定時間後に終了状態 E''^{*} になる。E'^{*} がエージェントの手元に餌がある状態であれば，エージェントは餌を取得し，E''^{*} はトレイ上から餌が取り除かれた状態を示している。逆に E'^{*} の状態が手元に餌がない状態であれば，トレイ上には変化がなく $E'^{*} = E''^{*}$ である。簡単のため，トレイ選択後の状態遷移 (熊手を物理的に引っ張る過程) は詳細を計算せず，報酬が得られる状態 E^{+} あるいはもらえない状態 E^{-} へ推移するものとした。ただし片方のトレイは報酬 r が与えられる。図は左のトレイが正解の場合を示しており，この場合，エージェントは終了時に E''^{left} の時のみ報酬 r を与えられる。

に重要な情報であったためである。例えば，実験 1 の H-C 問題の課題 1 の 1 の E は $E = ((0.5, 0.5, 0, 0, 1, null, null, null), (-1, 0, 0, 0, 1, null, null, null))$ となっている。ここで，$null$ は存在しないことを意味する。$A = L$ の場合，この行動によって生じる状態 E 'は E '$= ((0, 0, 0, 0, 1, null, null, null), (-1, 0, 0, 0, 1, null, null, null))$ となる。餌あるいは道具の y 座標が 0 以下の時エージェントはそれをとれるものとした。このため次の状態 E "でエージェントは餌をとれることになる。また，H-C 問題の課題 3 の 1 の場合は $E = ((0.5, 1.5, 0, 0, 3, 0, 1, 1), (0.5, 1.5, 0, 0, 4, 0, 1, 1))$ で，$A = L$ の場合，E '$= ((0, 0, 0, -2, 3, 0, -1, 1), (0.5, 1.5, 0, 0, 4, 0, 1, 1))$ となる。E ' で餌の y 座標が 0 以下なので先と同様にエージェントは餌をとれることになる。オートマトンの環境遷移関数は，このように全て物理的法則に従うように設定されている。

上記の表記方法を用い，Santos らの行なった実験 1 と実験 2 を計算機上に再現した。なお，本研究では道具の形状の差に伴う，操作の難易や類似性は表現していないため，実験 1 を再現するにあたり，H-C 問題と H-P 問題は本質的に同じ問題となる。そのため，H-C 問題と H-P 問題のシミュレーション実験結果は同じ結果になることは自明である。しかし，最終テストを行なうためには両方の問題を行なわせる必要があるため，シミュレーション実験では H-C 問題，H-P 問題の両方を行なわせた。

エージェント

次に各エージェントのアルゴリズムについて説明する。本研究では，タマリンが 1) 内部モデルとして，順モデル，逆モデルのどちらを用いているのか，2) 連続して何個の道具の動きを脳内にて予測することが可能であるか，について数学的モデルを用いて検証することを目的としている。ここで本研究では，以下のアルゴリズムを持つ 5 つのエージェントを用意した。

エージェント 1: 内部モデルを用いない。

エージェント 2: 順モデルを用いて 1 つの道具の動きを予測する。

エージェント 3: 順モデルを用いて 2 手先の道具の動きまで予測する。

エージェント 4: 逆モデルを用いて 1 つの道具の動きを予測する。

エージェント 5: 逆モデルを用いて 2 手先の道具の動きまで予測する。

それぞれのエージェントの計算方法について以下に述べる。

エージェント 1　エージェント 1 は，内部モデルを用いず，強化学習で提案されている actor-critic モデルを用いた (Sutton et al., 1998)。actor-critic モデルは報酬を手がかりとした学習方法の 1 つである。餌を得ることができた場合，そのトレイの熊手，餌の配置を記憶し，次回以降，優先的にそのトレイを選択する。一方餌が取

れなかった場合には，そのトレイは段々と取る確率が減る。この actor-critic モデルを，サルやラットが脳内で使用している可能性が指摘されている (Barto, 1995; Montague et al., 1996; Schultz et al., 1997)。よって，タマリンも使用している可能性が高い。上記の内容を示した基礎式を式 4.1〜4.3 に示す。

$$\delta = r + \eta V(E) - V(E') \tag{4.1}$$

$$V(E) \Leftarrow V(E) + \beta_1 \delta \tag{4.2}$$

$$p(E, A) \Leftarrow p(E, A) + \beta_1 \delta \tag{4.3}$$

E は環境状態，E ’は次状態，δ は報酬予測誤差，r は報酬量， η は割引率， (β_1 は学習係数，$V(E)$ は環境 E 下の状態価値関数（報酬の期待値），$p(E, A)$ は環境状態 E 下において行動 A を取る確率の指標であり，$p(E, A)$ の値が大きければ行動 A を選択する確率が増える。環境状態 E 下において，左のトレイを選択する確率 $Pr(L|E)$ は以下の式に従う。

$$Pr(L|E) = \frac{exp(\alpha_1 p(E, L))}{exp(\alpha_1 p(E, L)) + exp(\alpha_1 p(E, R))} \tag{4.4}$$

ここで α_1 は係数である。

エージェント 2，エージェント 3　エージェント 2 と 3 のアルゴリズムの概略を説明する。これらのエージェントは，順モデルを用いて各道具を用いた際のトレイ上の状況の変化を予想し，その予測に基いて使用すべき道具を計算する。詳細な定義は後述するが，本研究では，被験者がある道具を用いて他の物体を操作した際に，次の瞬間の操作された物体の状態を予測するモデルを順モデルと定義する。

H-C 問題の，課題 1 の 1 を例にし，アルゴリズムの挙動を示す。まず，エージェントは左右のトレイの状態推移を予測する。左側のトレイについてエージェントは，順モデルにより，手元にある熊手を用いれば餌が手元まで引き寄せられると

予測する。一方，右側のトレイについて，熊手を引き寄せた場合，餌は熊手の外にあるため餌が手に入らないと，順モデルを用いて予測する。この 2 つの予測の結果から，左側のトレイを選択すれば餌が取れることが導き出される。このように，各トレイの状況から，各道具を使用した結果を予測し，餌を手に入れることができると予測されるトレイを，エージェント 2 と 3 は選択する。

エージェント 2 は順モデルを用いて 1 つの道具の動きを予測し，選択するトレイを決定する。エージェント 3 は，2 つの道具を動かした際の各物体の動きを予測し，選択するトレイを決定する。

もし，内部モデルの計算結果が不正解につながる場合が多いときには，actor-critic モデルを使用する。つまり，自分が予想したとおり（内部モデルの予測どおり）に環境が動かない場合には，正解のトレイ内の熊手の配置を単純に覚えること（actor-critic モデル）により，タスクの達成ができるようにした。言い換えれば，エージェントは内部モデルによる選択と actor-critic モデルによる選択のうち正解確率の高い方をより頻繁に選んでいる。この使用確率は課題が変わっても引き継がれるようにした。

以下にその詳細な数学的定義を述べる。まず，順モデル M を以下に定義する。順モデルは，ある道具 $tool$ を用いて操作対象 $object$ を動作 $Action_{tool}$ で操作した場合の，操作対象の動き $Action_{object}$ を予測する。この定義を定式化したのが以下の式である。

$$M(O_{object}, O_{tool}, Action_{tool}) = (O_{object}, Action_{object}) \tag{4.5}$$

ここで，O_i は物体 i のトレイ上の状態（x 座標，y 座標，形状）を，$Action_j$ は物体 j の動き (今回は $Action_j \in \{\downarrow, no\_move\}$ とした) を意味する。ここでは，$\downarrow$ 下方向へ動くことを，no_move は動かないことを意味する。つまり，本研究で定義された順モデルは道具 $tool$ を使って物体 $object$ を使った際に，物体が下に動くの

か，動かないかを判定するモデルである。

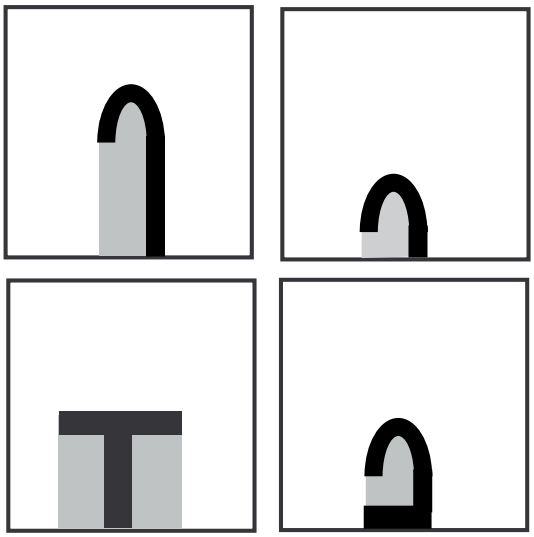


図 4.4: 順モデルの計算事例:順モデル（式 4.5）の計算事例を示した。入力条件を，状態 O_{tool} は各図に描かれた状態とし，$Action_{tool}$ は↓とした。図中，灰色の部分に物体 $object$ が存在すれば，モデルの出力 $Action_{object}$ は ↓となり，それ以外の部分に $object$ が存在した場合は no_move となる。一方，逆モデルは，物体の状態 O_{object} とその動き $Action_{object}$ が与えられた時に，$Object$ が $Action_{object}$ の動きとなるような道具とその動きを算出し，その結果を基に該当する台上の道具 O_{tool}，と動き $Action_{tool}$ を選択するモデルである。

エージェント 2 はこの内部モデルを用いて，手元にある道具を用いた場合，餌を手に入れることが出来るのか否かを調べるエージェントである。式 4.6 がそれを示している。

$$M(O_{food}, O_{tool}, Action_{tool}) = (O_{food}, Action_{food}) \tag{4.6}$$

エージェント 2 は，式 4.6 を用いて，餌の状態 O_{food}，手元にある $tool$，その動き $Action_{tool}$ を入力とし，$Action_{food}$ を求める。例えば，実験 1，H-C 問題の課題 1 の 1 の左トレイの場合，式 4.6 は $M((0.5, 0.5, food), (0, 0, 1), \downarrow) = ((0.5, 0.5, food), \downarrow)$ となる。この式で表わされるモデルをエージェントがもつと，餌が下に動く可能性を予測できることになる。本研究では餌が下に動くことと餌が手元まで動いてく

ることを同一視するので，このモデルを持つエージェントは餌が取れる可能性を予測することになる。図 4.4 に順モデルを用いた計算事例を図示した。図中の道具が O_{tool} を示し，灰色の部分が，$Action_{tool} =\downarrow$ の際 $food$ が下に動く $(food^*_x, food^*_y)$ の領域を，白色の部分が $food$ が動かない領域を示している。このように，順モデルにより，道具を使用した後の餌の動きが計算できる。本研究における順モデルは道具を使って手前に動かすことが出来る領域を計算していると考えると，計算方法が理解しやすいと思われる。

式 4.6 の解が存在した場合，つまり餌を取ることが可能な道具とその動作が見つかった場合，本研究ではその道具をサブゴールと呼ぶ。その解が存在しない場合にはサブゴールは無しとする。このとき，各トレイの価値 $Value(*), * \in \{Left, Right\}$ を以下のように定義する。

$$Value(*) = \begin{cases} r & \text{サブゴールが存在する} \\ 0 & \text{サブゴールが存在しない} \end{cases} \tag{4.7}$$

r は餌の報酬量を示す。この式は，餌を引き寄せる方法が見つかった場合には，そのトレイの価値を報酬と同じとみなし，見つからなかった場合には価値が無かったものとみなすことを意味している。以上の手順にて，トレイの価値を求め，エージェントはトレイの価値が大きい方をより高頻度に選択する。なお，内部モデルの能力では正解が確実に計算できない場合は，actor-critic モデルを用いた選択に段々と切り替わるようにした。以上を表したのが式 4.8，式 4.9 である。

$$Pr(L|E) = \begin{cases} \frac{exp(\alpha_2 Value(L))}{exp(\alpha_2 Value(L)) + exp(\alpha_2 Value(R))} & \text{with probability:} \\ & \frac{exp(\alpha_3 p)}{exp(\alpha_3 p) + exp(\alpha_3 (1-p))} \\ \frac{exp(\alpha_1 p(E,L))}{exp(\alpha_1 p(E,L)) + exp(\alpha_1 p(E,R))} & \text{otherwize} \end{cases} \tag{4.8}$$

$$q \leftarrow \begin{cases} \beta_2(1-q)+q & \text{モデルを用いて成功} \\ & \text{または actor-critic で失敗} \\ \beta_2(0-q)+q & \text{otherwize} \end{cases} \tag{4.9}$$

$q(0<q<1)$ は内部モデルの計算結果の信頼度を示している。また，α_2 と α_3 はそれぞれ$0<\alpha_2, 0<\alpha_3$ を満たす係数である。式 4.8 はエージェントが左のトレイを選択する確率を示している。式 4.8 上段が内部モデルの計算結果に基いて左のトレイを選択する場合を示している。式 4.8 下段が actor-critic モデルに基いて左のトレイを選択する場合を示している。エージェントは$exp(\alpha_3 q)/\{(exp(\alpha_3(1-q))+exp(\alpha 3q))\}$ の確率で上段の式を用いて $Pr(L|E)$ を計算し，これ以外の場合に下段の式で $Pr(L|E)$ を計算する。したがって，q の値が大きければ，上段の内部モデルを用いた推測に基づきトレイを選択する確率が増加し，q の値が小さければ下段の actor-critic モデルに基づきトレイを選択する確率が増加することを意味している。q は，内部モデルの判断を用いて報酬を得た場合，あるいは actor-critic モデルの判断を用いて報酬が得られなかった場合には上がり，そうでなければ下がる（式 4.9)。なお，q の値は問題 (H-C 問題，H-P 問題) が切り替わった時に初期化する。このことは，エージェントが H-C 問題についての信頼度は H-P 問題には適用できないものであり，別に計算しなければならないと考えていると仮定したことを意味する。

本研究のモデル化対象であるタマリンは，事前の学習（課題を開始する前に行う，基本的な熊手の使い方を覚える練習課題）により熊手の簡単な使用方法を獲得できている。よって内部モデルは課題開始当初に適切に獲得できていると仮定しその学習は本モデルには含めない。actor-critic モデルの学習は式 4.1-4.3 に従い，actor-critic モデルの判断を利用した時のみ（式 4.8 にて上側の式を使用した時），学習を行っている。

同様に，2 つサブゴールが設定できるエージェント 3 は，以下の式を用いてサブゴールを決定する。

$$M(O_{food}, M(O_{object}, O_{tool}, Action_{tool}) = (O_{food}, Action_{food}) \tag{4.10}$$

式 4.10 にて，手元にある $tool$ の状態 O_{tool} と動き $Action_{tool}$，台上にある $object$ を入力とし，$Action_{food}$ =↓ となった場合，$tool$ をサブゴール 1，$object$ をサブゴール 2 と呼ぶ。そのような入力の組み合わせが存在しない場合には，サブゴールは存在しないとする。例えば，実験 1H-C 問題の課題 2 の 1 の左トレイの場合，式 4.10 は $M((0.5, 1.5, food), M((0, 1, 1), (0, 0, 3), \downarrow)) = ((0.5, 1.5, food), \downarrow)$ となっている。エージェントがこの式のモデルをもつと餌が下に動き餌が取れる可能性を予測できることになる。各トレイの価値は式 4.7 に従い，左側のトレイを選択する確率は，エージェント 2 と同様，式 4.8， 4.9 に従うものとする。

エージェント 4，エージェント 5　エージェント 4 と 5 は，逆モデルを用いるエージェントである。本研究で定義される逆モデルは，手に入れたい物体の状況を観測し，それに必要となる道具を計算するモデルである。エージェントはゴールとなる餌から，どの道具があればよいのかを逆モデルを用いて計算し行動を決定する。

実験 1，H-C 問題の課題 2 を例にする。左側のトレイに関して，餌の位置から，餌を取るためには餌側の熊手を使う必要があることが，逆モデルから求まる。一方，右側のトレイでは，餌が取れる熊手が存在しないことが逆モデルから計算できる。上記の結果を比較し，これらのエージェントは左側を選択すれば餌を取ることができる。

なお，エージェント 4 は 1 つの道具の動きを，エージェント 5 は連続して 2 つの道具の動きを内部モデルを用いて予測できるものとした。またこの計算方法では正解のトレイが選択できない状況が続けば，エージェント 2 および 3 と同様 actor-critic モデルが動くようにした。

以下にその詳細な数学的定義を述べる。逆モデル M^{-1} の入力は，手に入れたい物

体の状態O_{object}，その物体の動き$Action_{object}$であり，出力は，$object$を$Action_{object}$にするために必要な道具の状態O_{tool}とその動き$Action_{tool}$である（式4.11)。逆モデルとは，ある物体をある方向へ動かしたい時に，使うべき道具の状態とその動かし方を計算する内部モデルである。

$$M^{-1}(O_{object}, Action_{object}) = (O_{object}, O_{tool}, Action_{tool}) \tag{4.11}$$

例えば，実験1，H-C問題の課題1の1の左トレイで餌を下へ動かした場合，$M^{-1}((0.5, 0.5, food), \downarrow) = ((0.5, 0.5, food), (0, 0, 1), \downarrow)$で記述される。図4.4を例にすれば，動かしたい物体が斜線部分にあるような道具の状態を台上から探すのが逆モデルである。

今回の実験では，逆モデルによる計算の結果，O_{tool}が複数存在することは無かったが，もしそのような場合には，候補の中からランダムに選択する，もしくは状況ごとに強化するなどの方法が考えられる。

次に逆モデルを用いてサブゴールを決定する方法を説明する。一般的な場合を考え，n個のサブゴール$Sub(n)$を決定できるエージェントについて考える。ただし，$, k = 1, 2, \cdots, n$。ここで$hand$は手を，$null$は「存在しないこと」を，$tool_1, tool_2, \cdots$は台上に存在する物体を意味している。最初のサブゴール$Sub(1)$は，餌を手に入れるための道具とする（式4.12)。2個目以降のサブゴール$Sub(k)$は，1個手前のサブゴール$Sub(k-1)$を手に入れる道具とする (式4.13)。このサブゴールを基にトレイの価値$Value(*)$を決める。

トレイの価値$Value(*)$は，$Sub(k) = null$の場合には0を，そうでない場合には餌の報酬量rとする。この計算を$k = 1$からスタートし，$k = n$，または，$Sub(k) \in \{null, hand\}$の時に終了する。つまり，1) n個道具を使った場合でも，まだ餌が取れる可能性のあるトレイは餌の取れそうなトレイとみなす。2) 途中のk番目で，手を用いて$k-1$番目のサブゴールを手で操作可能ならば，そのトレイ

は餌が取れそうなトレイとみなす。3)$k-1$ 番目の道具を取る手段が無い場合には，そのトレイは餌がとれないトレイとみなす。以上を簡略化して表現したのが式 4.14 である。

$$(O_{Sub(1)}, Action_{Sub(1)}) = M^{-1}(O_{food}, \downarrow) \tag{4.12}$$

$$(O_{Sub(k)}, Action_{Sub(k)}) = M^{-1}(O_{Sub(k-1)}, Action_{Sub(k-1)}) \tag{4.13}$$

$$Value(*) = \begin{cases} r & Sub(k) \text{が存在する} \\ 0 & Sub(k) \text{が存在しない} \end{cases} \tag{4.14}$$

つまり，本研究で定義された逆モデルは，トレイ上の道具のうち，餌や道具を手元へ動かすことが可能な道具の状態を求めるモデルである。以上の手順で，各トレイの価値を求めたのち，そのトレイを取る確率 $Pr(L|E)$ を順モデルと同様に式 4.7 にて求める。順モデルと同様に，内部モデルでの選択に誤りが多い場合には，actor-critic モデルの判断を優先的に採択するようにした。

パラメータ

以上 5 種類のエージェントを用いて，Santos の行った 2 つの行動実験を対象に計算機実験を行った。今回の実験では，各パラメータの初期値を以下の通りに設定した。なお，この初期値の設定に関しては，考察で述べる。

$$\alpha_1 = 4.0, \alpha_2 = 4.0, \alpha_3 = 3.0, \beta_1 = 0.25, \beta_2 = 0.005, \eta = 0.75, q = 0.8, r = 5.0. \tag{4.15}$$

エージェントの前提条件

上述のエージェントを作成する上で，何点か大きな前提条件がある。本文中で既に記載されているが，重要な点であるためもう一度整理する。

まず，内部モデルは本課題中では学習せず，学習の際には actor-critic モデルを用いた。タマリンは問題前の練習課題において，基本的な熊手の使用方法を既に学習している。この練習課題において，熊手の内部モデルをタマリンは完成させていると仮定したためである。そのため，例えば，2 個以上ある熊手を 1 つの道具とみなし，新たに内部モデルを学習するという能力は現在のモデルでは想定していない。また，学習にあたって actor-critic モデルを用いたということは，タマリンは具体的な環境のダイナミクスを予測して行動を計画しているのではなく，単に報酬予測量と，環境状態，行動をペアで覚えていたという仮定を用いている。勿論，先の例で述べたように,「内部モデルが 2 個以上の熊手の動きを学習する」ようなモデルも作成することは可能である。しかし，タマリンの属する新世界サルは，道具を使うこと自体がほとんど観測されていない種であり，せいぜい使えたとしても 1 つの道具を使うのが精一杯だと考えられている (Matsuzawa, 2001)。このような背景の下，2 つの道具を組み合わせた内部モデルをタマリンが学習できる可能性よりも，actor-critic モデルを用いて，単に行動手順を覚える学習の可能性を先に議論すべきだと本研究は考えた。

また，内部モデルの入出力関係のパラメータについても，様々な可能性が残されている。例えば，今回定義した順モデルの入力は，道具の状態，道具の動き，操作対象の状態を，出力を操作対象の動きとした。しかし，これ以外の可能性も勿論ある。先に述べた内部モデルの学習と共に，この点を今後議論を進める必要がある。

最後に，確信度 q の値は H-C 問題から H-P 問題へ移った際に初期化される。ただし，同じ問題における課題の推移では初期化していない。つまり，本研究ではタマリンが各問題毎に確信度が異なると仮定している。

以上の過程を用いて数学的モデルを構築した。

4.3 結果

4.3.1 実験１

図 4.5 に実験 1 の結果を示す。図 4.5 は各エージェントが各課題に対して 90% 以上の確率で成功するまでにかかったセッション数と最終テストにて正解のトレイ (餌が取れるトレイ) を選んだ割合を示している。それぞれ乱数の種を変えて 100 回シミュレーションを行い平均と標準偏差を求めた。前述の通り，Santos らの実験では，タマリンが連続して 90% の正解率を出すまでのセッション数は，課題 1 と課題 2 との間には有意差がなく，課題 3 では課題 2 よりも有意に多く，課題 4 は課題 3 に対して有意に少なかった。以上の点をタマリンの結果の再現性を検証する上で用いる。

なお，後述の通り，いずれのエージェントも H-C 問題と H-P 問題の間に有意な差は見られなかった。今回のモデルでは，H-C 問題と H-P 問題は本質的に同じ数学的表現を用いたためである。そのため，H-C 問題の結果を中心に述べる。なお，本研究では，統計的処理にストゥーデントの t 検定を用い有意水準 $p < 0.01$ の場合に有意差があるとする。

まず，タマリンの行動実験結果を再現できたのはエージェント 4 のみであった。エージェント 4 の結果を図 4.5D に示す。H-P，H-C の 2 つの問題において，課題 1 と課題 2 の間では有意差はなかった ($p = 0.71$)。課題 3 は課題 2 に対し有意に多かった（課題 2：平均 $M = 2.13$，標準偏差 $\sigma = 0.46$，課題 3： $M = 20.24$， $\sigma = 2.9$；$p < 0.01$)。課題 4 は課題 3 に対して有意に少なかった（課題 4: $M = 5.20$, $\sigma = 2.16$; $p < 0.01$)。この結果から，タマリンは逆モデルを使用し，1 つの道具の動きを予測できる可能性があることが示唆された。

エージェント 4 は逆モデルを用いることによって，課題 1，2 では即座に正解のトレイを選択することが出来た。一方，課題 3，4 ではどちらのトレイを選ぶべき

か課題開始時に適切に判断することが出来なかった。この理由は，エージェント 4 は餌側にある 1 つの道具の動きしか計算できないため，正解のトレイを見抜くことができない点にある。課題 3，4 において連続して 90% 以上の確率で餌が取れるようになったのは，内部モデルによる判断で誤答が続き，actor-critic モデルの学習が十分に行われた後である。課題 3 終了時は，内部モデルよりも actor-critic モデルの判断を用いるようになる。この点は課題 4 開始時も続くようにモデル上なっている。課題 4 においても，内部モデルの判断では適切なトレイ選択が出来ない。課題 4 の方が課題 3 よりも早く正解のトレイを選択できた。

以下はその他のモデルの結果の詳細について述べる。

モデルを持たないエージェント 1 は約 3〜4 セッション程度で全ての課題を達成することが出来た。これは，今回の課題が左右のトレイの選択課題であり，試行錯誤的な戦略でも高々4 セッション程度で解くことが可能な難易度であることを示している。結果を図 4.4A に示す。一見してわかるように，エージェント 1 はタマリンの結果を再現しなかった。

順モデルを用いて 1 個の道具の動きを予測するエージェント 2 の結果を図 4.5B に示す。エージェント 2 は，課題 2 に多くのセッション（平均 20.59 回）を必要とした。この結果はタマリンが課題 2 は比較的早くから連続して正答率 90% を達成した結果と異なる。よってエージェント 2 は，タマリンの結果を再現できなかった。

エージェント 2 は手元にある道具を用いて餌が取れるか否かを判定している。計算できる道具の手順数が，課題に対して十分あれば，必ず正解が見つかるのだが，エージェント 2 は 1 つの道具の動きしか計算できない。課題 2 以降のタスクでは 2 つの道具の動きを予測する必要があるので，正解のトレイを見抜くことは出来なかった。そのため，エージェント 2 は課題 2 で，課題 1 より多くのセッション数を必要とした。課題 3 以降は内部モデルの計算結果よりも actor-critic モデルの計算

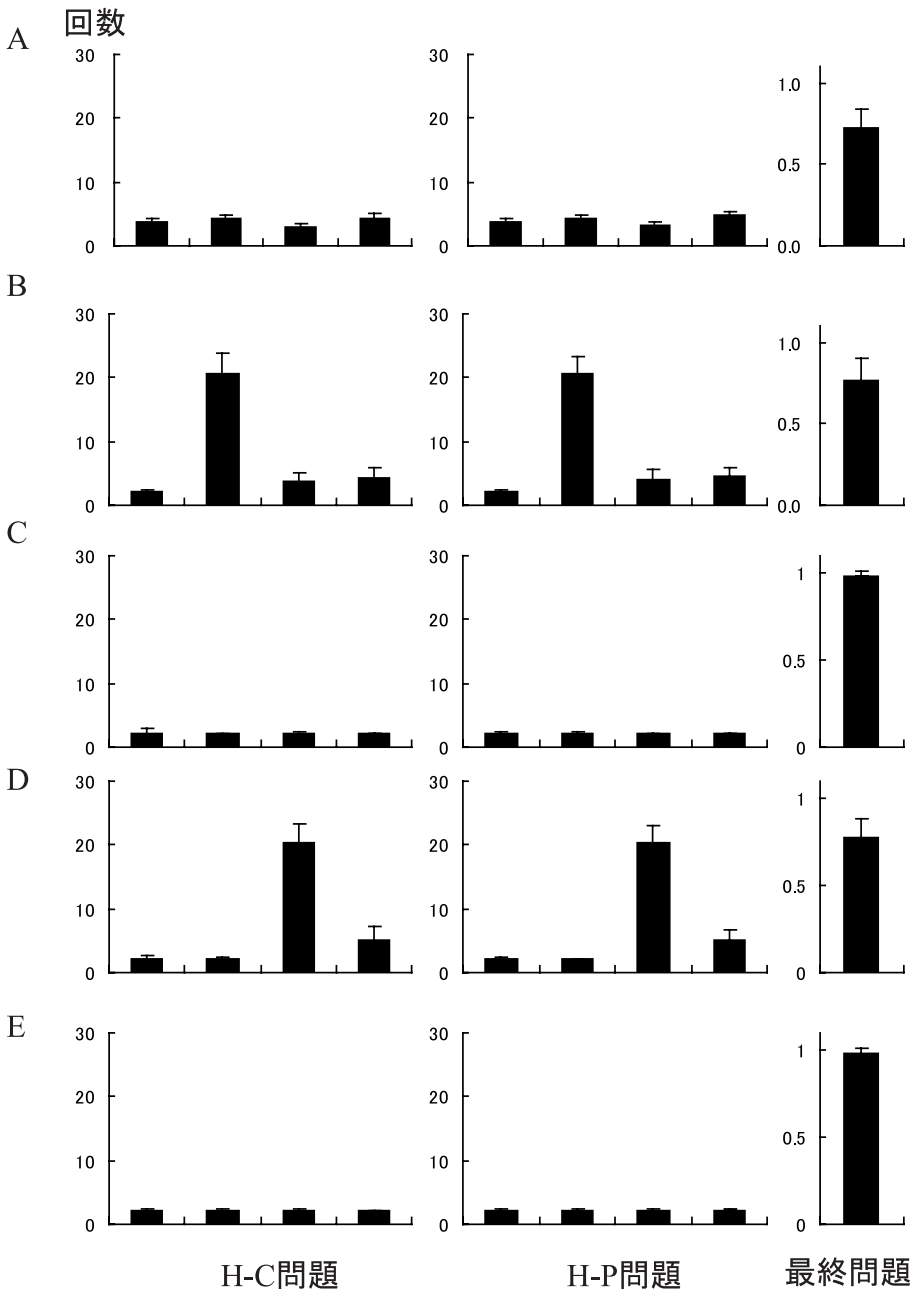


図 4.5: シミュレーション実験結果：実験 1 左段は H-C 問題の各課題においての 90% 以上の正解率を得るまでのセッション数を示している。右段は最終テストの正答率を示している。H-P 問題の結果は H-C 問題と有意差がなく，同一であるため割愛した。A) はエージェント 1 ，B) はエージェント 2 ，C) はエージェント 3, D) はエージェント 4，E) はエージェント 5 の結果をそれぞれ示している。 タマリンの結果を再現したのは D) エージェント 4 のみである

結果を信頼するようになったため，課題 2 に比べて課題 3，4 は少ないセッション数で 90% の成功率まで到達できた。

エージェント 3 と 5 は全ての課題に対して即座に正解に至った。その結果をそれぞれ図 4.4C，E に示す。一見してわかるように，エージェント 3 と 5 はタマリンの結果を再現しなかった。

今回の課題は高々2 つの道具を使った課題である。そのため 2 つの道具の動きを予測できれば餌が取れるかどうかは判定できる。よって，この 2 つのエージェントは actor-critic モデルに推移しなくても餌の取れるトレイを即座に選択出来た。

なお，最終テストの結果を図 4.4 (右列) にて示す。エージェント 1 は平均 72% ($\sigma = 12$)，エージェント 2 は平均 77% ($\sigma = 11$)，エージェント 3 は平均 98% ($\sigma = 2$)，エージェント 4 は平均 77% ($\sigma = 11$)，エージェント 5 は平均 99% ($\sigma = 2$) であり，どのエージェントもランダムよりも有意に高い確率で正解のトレイを選択し，Santos らの実験結果を再現した。

4.3.2 実験 2

図 4.6，表 4.1 に実験 2 の結果を示した。図 4.6 は，各エージェントが各課題において 4 回の試行中に左のトレイ（図中黒のバー），右のトレイ（図中白抜きのバー）を選択した回数を示している。表 4.1 は各エージェントが左のトレイを選択した回数の平均と分散の値を示している。各シミュレーションは 100 回行い，平均と標準偏差を求めた。前述の通り，タマリンを対象とした実験では，課題 2〜6 の全てに対して左右のトレイの選択に有意な差が見られなかった。この点について各エージェントと比較する。

エージェント 1 と 4 がタマリンの結果を再現した。それぞれ図 4.6A と D に示されている。どちらのエージェントも，全て有意差が無かった（$p > 0.01$）。エージェ

エージェント	1		2		3		4		5	
	平均	分散	平均	分散	平均	分散	平均	分散	平均	分散
課題 2	1.9	1.6	3.9	0.3	3.8	0.4	2.1	1.0	3.8	0.5
課題 3	2.1	1.7	2.1	0.9	2.1	1.0	2.0	0.9	2.1	0.4
課題 4	2.1	1.8	3.9	0.3	3.8	0.5	2.0	1.0	3.8	0.4
課題 5	1.9	1.3	3.9	0.3	3.8	0.4	1.8	1.0	3.9	0.3
課題 6	1.9	1.2	3.8	0.5	3.8	0.4	1.8	1.0	3.8	0.5

表 4.1: 実験 2 において，各エージェントが左のトレイを選択した回数の平均と分散を示した。

ント 1 では実験開始時，左右のトレイをランダムに選択し，学習が進むと正解のトレイを選ぶようになるアルゴリズムを用いている。よって，この結果は自明である。一方，エージェント 4 は，逆モデルを用いて，餌が中に入っている熊手を計算している。今回は実験 1 の課題 3 のように，課題 1 を除く全てのトレイにおいて餌が熊手の中に入っているため，どちらで餌が取ることができるのか判断がつかない。そのため，課題 2〜6 について全てランダムに選択せざるを得なかった。

その他のエージェントは，課題 2，4，5，6 では左のトレイを有意に多く選択した。どちらのトレイでも餌が取れる課題 3 では有意差はなかった。

エージェント 2 は順モデルを用いて，現在選択可能な道具から順々に使用する道具を選択していく。実験 2 は実験 1 と異なり，1 つの道具を用いて餌を取ることが可能である。そのため，1 つの道具の動きが予測できれば正解のトレイを選択することが可能である。その結果，全ての課題において正解のトレイを選択することが出来た。2 つの道具の動きを予測できるエージェント 3 と 5 は，内部モデルの種類に関係なく正解のトレイを選択できた。以上のように，タマリンの行動を両方の実験で再現したのはエージェント 4 のみであった。

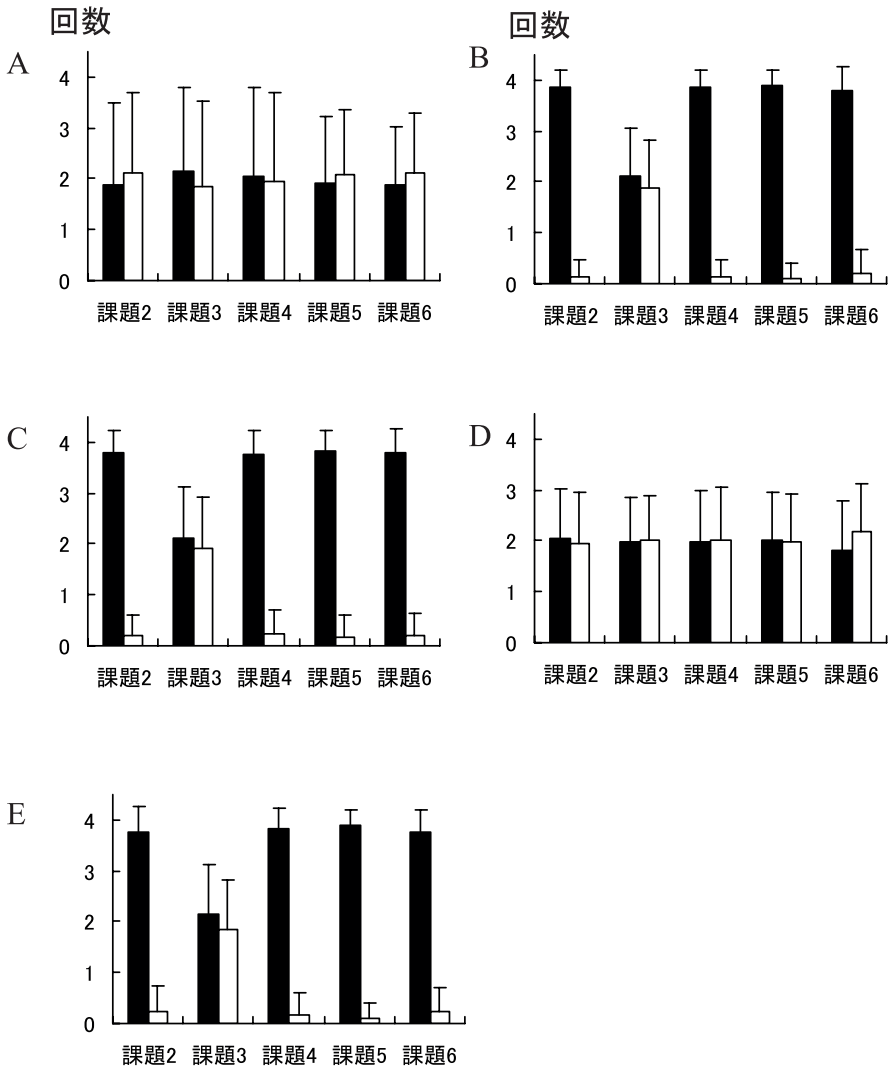


図 4.6: シミュレーション実験結果：実験 2 各図は，各課題に対し黒棒が左のトレイを選択した回数，白抜き棒が右のトレイを選択した回数を示す。100 回のシミュレーションの平均を記載した。エラーバーは標準偏差を示す。A) はエージェント 1，B) はエージェント 2，C) はエージェント 3，D) はエージェント 4，E) はエージェント 5 の結果をそれぞれ示している。タマリンの結果を再現しているのは，A) エージェント 1 と D) エージェント 4 であった。

4.4 考察

本研究では，Santos らによって行なわれたタマリンの 2 つの熊手使用実験を数学的にモデル化し，その能力を 1) 内部モデルの計算方法，2) 連続して道具を使用する際に，内部モデルを用いて何手先の道具の動きまで予測できるか，の 2 点から検討し，数学的モデルの挙動からタマリンの能力を推定した。その結果タマリンは，順モデルを用いず，逆モデルを利用して 1 個の道具の動きを予測できる能力を有する可能性を示唆した。

4.4.1 実験結果から推測される，タマリンの道具使用行動について

まず，内部モデルについて考察を行なう。本研究において，内部モデルを持たず actor-critic モデルのみを用いるエージェント 1 は実験 1 における全ての課題をランダムに解いており，Santos らによって示されたタマリンの行動を再現できなかった。サルが actor-critic モデルを脳内にて使用している可能性は様々な論文から支持されている (Barto, 1995; Montague et al., 1996; Schultz et al., 1997; Suri & Schultz, 2001)。今回の結果から，タマリンは actor-critic モデルだけでなく，その他のアルゴリズムも用いて Santos らの課題を解いている可能性がある。

本研究では，actor-critc モデル以外のアルゴリズムとして，順モデルと逆モデルに焦点を絞り，タマリンの能力を推定した。以下，本研究の実験結果における順モデルと逆モデルの差を説明する。その際，図 4.5 に示すように実験結果に大きく差が現れたエージェント 2 とエージェント 4 を対象に話を進める。まず，実験結果を振り返る。順モデルを持つエージェント 2 は実験 1，実験 2 ともにタマリンの行動を再現しなかった。エージェント 2 は，実験 1 では課題 2 を解くために多くのセッション数を必要とし，実験 2 においては，全ての課題に対し餌を取ることが可能なトレイを正確に選択した。この結果は，タマリンが実験 1 において，課題 2

を少ないセッション数で正解のトレイを選択できた一方，課題 3 において多くにセッション数を必要とした結果 (Santos et al., 2005)，実験 2 において餌の取れるトレイを見分けることが出来なかった結果 (Santos et al., 2006) と異なる。一方，逆モデルをもつエージェント 4 は，実験 1，実験 2 の全ての課題に対して，Santos らによって示された行動実験結果を再現した。

次に，上記実験結果に対して考察する。実験 1 に関して，順モデルを用いるエージェント 2 は課題 2 を解くために多くのセッション数を必要とした。一方，逆モデルを用いるエージェント 4 は課題 2 については即座に解けている。この差は，順モデルが手で動かすことができる道具から状態推移を計算している一方で，逆モデルは餌を引き寄せるために必要な道具から求めている点から生じている。若干不正確な言い方ではあるが，本研究での順モデルはスタートから状態遷移を求めていくが，逆モデルはゴールから求めていくと考えればこの結果を理解しやすいと思われる。以下，それぞれのエージェントに関して，具体的に説明する。エージェント 2 の場合，まず手元側の道具を動かした状況を順モデルから計算する。エージェント 2 は，エージェント 3 とは異なり内部モデルで計算できる道具の数が 1 つしかないため，複数の道具の動きまでは計算できないように設定されている。よって，エージェント 2 は手元側の道具で餌が取れるのかどうかの判定しか計算できない。しかし，課題 2 は，一見して明らかなように，手元側の熊手のみで正解のトレイを判定することは出来ない課題である。そのため，順モデルの判断を用いるエージェント 2 は不正解なトレイを選択し続け，図 4.5B と図 4.6A が示すように時間をかけて actor-critic モデルにて正解のトレイを学習していく結果となった。一方，逆モデルを用いるエージェント 4 は，餌を動かすことが可能な配置にある道具を求める。言い換えれば，エージェント 4 は餌側の道具が餌を動かせるか否かを判定することになる。課題 2 の場合，餌側の熊手で餌がとれるか否かを判断す

れば，正解を求めることができる。そのため，エージェント 4 は課題 2 において即座に正解のトレイを判定できた。一方，課題 3 では，手元側の道具が餌側の道具を引き寄せられるのか否かを判定しなければ正解のトレイは判断できない。エージェント 5 と異なり，エージェント 4 は内部モデルを用いて計算できる道具の数が一つであり，手元側の道具の動きを内部モデルで計算できないため不正解が続き actor-critic モデルにて時間をかけて正解のトレイを学習した。

実験 2 では，手元側の道具で餌を取ることが出来るのかについての判断だけで正解のトレイが判定できる。そのため，エージェント 2 のような順モデルで手元の道具の動きを計算すれば，正解が判断できた。一方，この実験では，全てのトレイにおいて餌が熊手の中にあり，その熊手を引き寄せることが可能であれば餌が取れてしまう。そのため，エージェント 4 はどのトレイでも餌が取ることが可能だと判断し，全ての課題において左右のトレイを約 50% の確率で選択した。この結果は，Santos らによって示されたタマリンの行動を再現している。

仮にエージェント 2 に関して，例えば以下の能力的拡張を行なえば，実験 1H-C 問題の課題 2 を即座に解くことができ，順モデルを用いても Santos らが示したタマリンの実験 1 の結果を再現することが可能であると推測される：候補 1）餌側，手元側 2 つの道具を一つの道具としてみなし，新たに内部モデルを作成する能力，候補 2）順モデルの計算を餌側の熊手にも行なう能力。しかし，仮にこのような機能的拡張をした場合でも，実験 2 の結果が説明できない。図 4.2 に示したとおり，タマリンは実験 2 に対して，全ての課題において左右のトレイをランダムに選択している。しかし，図 4.6 に示したとおり，エージェント 2 は，全ての課題において，正解となる餌が取れるトレイを有意に高い確率で選択している。機能的拡張を行なえば能力の向上が予想されるが，機能的拡張していないエージェント 2 が，既に全ての課題において正解のトレイを選べ，タマリンの能力よりも高い正解率

を得ている。実験 1，実験 2 の結果を総合的に考えると，タマリンが本研究にて定義した順モデルを使用してる可能性が低く，逆モデルを用いて求めている可能性が高いと推測される。

なお，本稿にて定義した逆モデルは，餌の状況から使用すべき道具の形状・動作を計算するモデルである。その理由（e.g. 何故この道具で餌が取れるのか。など）や動作原理（e.g. 道具と餌の動き，接触状態など）を計算過程に組み込んでいない。Santos によって行なわれたタマリンの実験結果をただ再現するだけであれば，上述の能力は必要ないことが本研究から示唆された。

次に内部モデルを用いて何個の道具の動きが予測できるかについて考察する。本研究では 2 つの道具の動きを予測できるエージェント（エージェント 3 と 5）は，実験 1，実験 2 を通じて全ての課題に対し即座に正解のトレイを選択できていた。この結果は，一見してわかるように，Santos らによって示されたタマリンの行動実験結果をいずれも再現しない (Santos et al., 2005; 2006)。よって，タマリンは道具を使用する際に，2 個以上の道具の動きは予測できない可能性が示唆された。また，先に述べた通り，actor-critic モデルを用いるエージェント 1 も Santos らの結果を再現できなかった。ことから，何らかのモデルを獲得し道具の動きを予想している可能性がある。エージェント 4 が Santos らの実験結果を再現した点からも，モデルを用いて 1 つの道具の動きを予測できる可能性が高いと考えられる。

4.4.2 数学的モデルのパラメータから推測されるタマリンの特性について

本研究の数学的モデルは，課題開始時には内部モデルの計算結果を用いてトレイを選択するが，内部モデルを用いて不正解が続いた場合，actor-critic モデルをより高い確率で用いる。本モデルでは，この内部モデルと actor-critic モデルの切

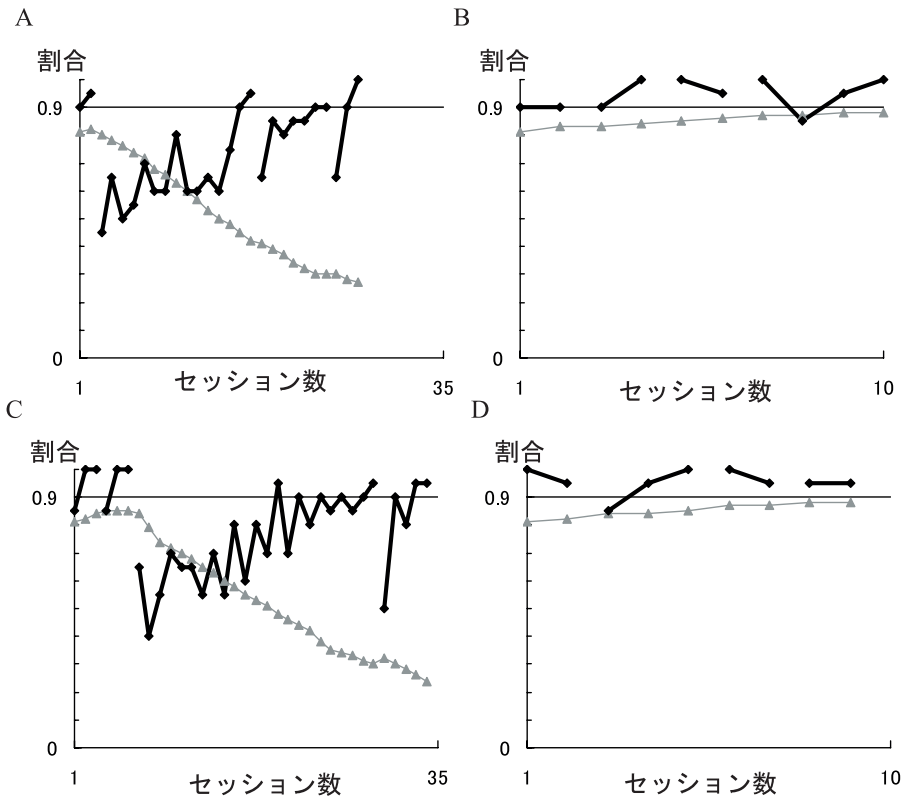


図 4.7: H-C 課題における，各エージェントの成功割合（黒線）と確信度 q の値を示した。成功割合は各セッション（20 試行）で餌を取ることが出来た割合を示し，確信度 q は各セッション終了時の値を用いた。A）はエージェント 2，B）はエージェント 3，C）はエージェント 4，D）はエージェント 5 の，それぞれ代表的な 1 計算結果を示している。

り替え頻度を確信度 q というパラメータで表現している。確信度 q が高ければ高頻度で内部モデルの計算結果を用い，低い場合には actor-critic モデルを高頻度で用いる。

図 4.7 に実験 1H-C 問題，課題 1 から課題 4 における各エージェントの各セッションごとの正解率とセッション終了時の q の値を示した。エージェント 2 が課題 2 から q の値が下がる一方で，エージェント 4 は課題 3 から下がり始めるという差がある。先の考察で述べたとおり，エージェント 2 は順モデルを用いているため，内部

モデルからは課題 2 における正解のトレイを正確に計算できない。そのため課題 2 においてに q の値が下がり，actor-critic モデルにて正解のトレイを学習していく。今回，結果を示していないが，エージェント 2 は課題 3，4 についても内部モデルでは正解のトレイを計算できないが，図 4.7 に示したとおり課題 3，4 の開始時には q の値が下がっており，課題開始時から actor-critic が有意に高い確率で選択される。actor-critic モデルにとって今回の問題は，行動選択が 2 つ（左右のトレイの選択）であり，図 4.5A からもわかるように高々3 セッション程度あれば適切な行動選択が出来る問題である。そのため，課題 3，4 は actor-critic モデルとほぼ同様のセッション数にて学習が終了している。一方，エージェント 4 は逆モデルを用いているため，先の考察にて述べたとおり，課題 2 は内部モデルを用いて正解のトレイを即座に求めることができるが，課題 3 は内部モデルでは解くことができない。課題 3 に関しては，actor-critic モデルの計算結果を高頻度で使用するようになるまで，つまり q の値が適当な値まで下がるまで正解のトレイを選択できる。課題 4 に関しては，先のエージェント 2 にとっての課題 3，4 と同様，課題開始時に q の値が低いため早く学習が終わる。

道具使用行動ではないが，内部モデルのようなモデル有り学習機構と，actor-critic モデルのようなモデルフリー学習機構がモデルの信頼性に基いて切り替わっている機構は既に Daw らによって提案されている (Daw et al., 2005)。彼らは，ラットの行動実験結果を上記のモデルにて再現し，前頭皮質を含む系がモデル学習機構を，大脳基底核を含む系がモデルフリー学習機構をおこない，それぞれの予測信頼性に基づいて各学習機構が選択されるという仮説を立てている。彼らの主張する予測信頼性は，本研究の確信度 q に相当する。本研究において，q の初期値によってはエージェント 4 においても Santos らの実験結果を再現できない場合が存在する。図 4.5 においては q の初期値を 0.8 としたが，図 4.8A に示すように q の

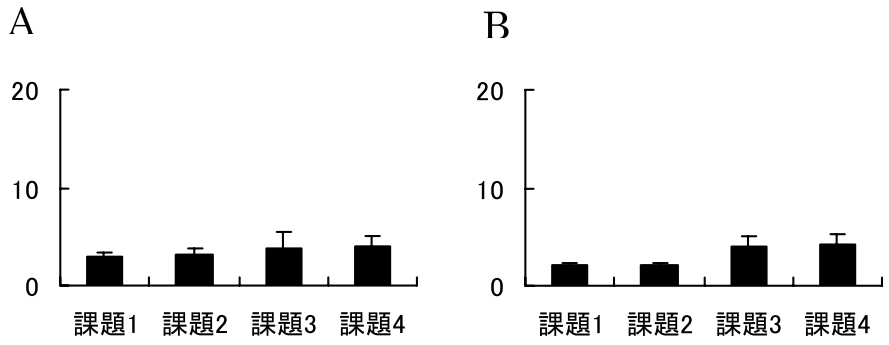


図 4.8: パラメータによる挙動の変化図 4.5 と異なるパラメータのエージェント 4 が実験 1 H-C 課題を解いた際の挙動。図の横軸，縦軸は図 4.5 と同じ。A) 確信度 q の初期値を 0.2 とし，その他のパラメータは図 4.5 と同じにした際の挙動。B)β_2 を 0.25 とし，その他はパラメータは図 4.5 と同じにした際の挙動。

初期値を 0.2（actor-critic モデルを内部モデルよりも有意に多く使用する）とすると，課題 3 と 4 の間に有意差がなく ($p = 0.49$)Santos らの実験結果を再現できなかった。

つまり，タマリンは Santos らの実験のような状況では，初見の問題に対し，内部モデルを優先的に用いている可能性が本モデルから推測される。また，数学的モデルの説明でも述べたとおり，Santos らの実験結果を再現するためには，確信度 q の値は H-C 問題から H-P 問題へ移った際に初期化する必要があった。その一方，H-C 問題内では，q の値は課題が変わっても引き継がれる必要があった。つまり，actor-critic モデルと内部モデルの切替は，環境状態（特に環境に存在する道具の種類）に依存している可能性がある。確信度 q の初期化を必要とする環境の変化の条件については，本研究から主たる知見は得られていないため，今後の課題となる。

actor-critic モデルの学習速度に対して，確信度 q の学習速度が比較的小さな値

である必要があった。図 4.8B に確信度の学習係数 β_2 を 0.25 としたシミレーション結果を示している（図 4.5 では $\beta_2 = 0.005$ としている）。この結果では，課題 3 と 4 の間に有意差がなかった（$p = 0.36$）。学習係数が小さいということは，仮にモデルを用いて不正解が続いても同じモデルに固執する傾向が強いという特徴を持つ。例として，チンパンジーが長期に渡って自分の信じた方法に固執した例（Köhler, 1925）や，前頭前野の疾患患者がウィスコンシンソーティングテストのような途中でルールの変更を求める課題に対して苦手とするといった現象 (Fuster, 1997) を想起させる。前頭前野がヒトと比べて比較的小さいタマリンがルール変更を行うことが得意でないことが，課題 4 が課題 3 に対して有意に少ないという結果に現れている可能性がある。逆に β_2 が高い値を持つ生物，つまりモデル選択が柔軟に変更できるような生物に対して，Santos と同じ行動実験をおこなうのであれば，図 4.8B のような実験結果がモデルから予想される。

4.4.3 モデルの妥当性を検証する実験

今回のモデルの中でエージェント 4 は Santos の示したタマリンの 2 つの行動実験結果を再現することが出来たが，この結果から即座にタマリンがエージェント 4 を用いているとは限らない。更なる検証実験を行ない，その妥当性を高める必要がある。例えば次のような実験が考えられる。図 4.9 に示すような実験 1 課題 2，課題 3 に更に 1 つ道具を増やしたような実験を行なった場合，タマリンがもしエージェント 4 で記述した計算機構を使用しているのであれば，元の実験と同じように，図 4.9A では即座に出来，図 4.9B の実験では正解を得るまでに時間のかかることが予想される。もし，予想どおりの結果であれば，タマリンが本研究において定義した逆モデルを利用し 1 つの道具を決定している可能性が高まる。また，実験 1 課題 3 で内部モデルから actor-critic モデルへの遷移が起こるという仮説を検

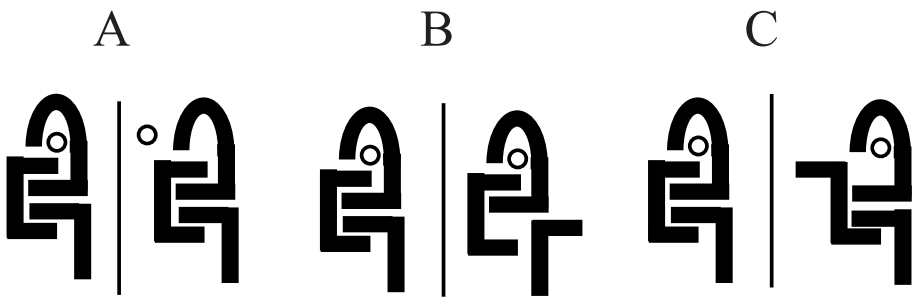


図 4.9: 検証実験:本研究の妥当性や他の生物の能力の推定に対応するための実験例。○は餌を示している。

証する実験として，課題 3 と課題 4 の順序を入れ替えた実験が考えられる。予測される結果は，課題 3 と課題 4 が同質の課題であるので課題 4 の成功率向上に多くのセッション数を要し，課題 3 ではセッション数が有意に減少することになる。

4.4.4 数学的モデルから予測される実験結果

今までの考察で述べたとおり，Santos らの結果を再現するためには，内部モデルの種類，内部モデルで計算可能な道具の数，特定のパラメータ条件などを満たす必要がある。これらの条件はタマリンの能力特性を意味する。逆に，異なる生物に同じ課題を行なわせた場合，異なる結果が出てくる可能性がある。それらの結果を再現するパラメータやモデルの種類を特定することが可能であれば，各生物の特性を比較しより詳細な能力比較が可能になると考えられる。例えば，本研究では再現対象としなかったが Santos らは実験 2 に対してベルベット・モンキーに同じ行動課題を課し，タマリンの行動と比較している。ベルベットの結果と本研究図 4.7 とを比較すると，ベルベットはエージェント 2，3，5 に似た行動結果となっている。さらにベルベットの能力を推定したい場合には，実験 1 を行なわせてみることも有効な方法であると考えられる。もし仮にベルベットが順モデルを使用している場合には，図 4.5B または C のような結果が予想される。また，エー

ジェント 5 のように逆モデルを用いて道具を 2 つ決定する場合には，図 4.5E の結果が予想される。図 4.5C または図 4.5E のような場合には，2 つの道具の動きを推測する能力を有する可能性がある。この場合には更に異なる実験をしなければならないが，例えば図 4.9A，B，C のように道具の数を増やすことにより特定が可能であると推測される。なお，エージェント 5 であれば，A と C は即座に解くことができるが，B は解くまでに時間がかかることが予想され，エージェント 3 であれば，全ての課題を解くのに時間がかかることが予想される。また仮に図 4.5D の結果になった場合，順モデルと逆モデルを両方状況にあわせて使い分けている可能性が高いと予想される。この理由は課題 2 は逆モデルで即座に解くことができるが，課題 3，4 は順逆いずれの内部モデルでも解くことが出来ないので，図 4.5D のように課題 3，4 にて多くのセッションを使用することが予想されるからである。ただ，現状ではあくまで推測の段階であり，この主張をする場合にはシミュレーションなどにより厳密に調べる必要がある。

4.4.5 今後の課題

第 1 に，本研究で提案した計算方法以外の可能性について検討する必要がある。本研究では 5 つのエージェントを用意したが，その他の可能性も考えられる。例えば，先の考察の最後に紹介した順モデルと逆モデルを両方持つエージェントや，順モデルを手元にある道具から適用するのではなく，どこからでも計算するようなエージェントである。また，今回の内部モデルでは入出力関係を一通り仮定し，学習は行わない物と仮定した。これ以外の可能性も十分考えられるため，今後はこれらの可能性に関しても研究する必要がある。

第 2 に今回と同様の結果を再現するパラメーター群を求めることがあげられる。今回はタマリンの結果を再現するパラメーターの 1 組を求めた。しかし，図 4.8 で

示したようにパラメーターによっては，Santos らの実験結果を再現できない場合もある。今後は再現するためのパラメーターの条件を探る必要がある。これによりタマリンが持つ能力をさらに詳細に調べることが可能となる。

第 3 に，エージェント 4 の得意な環境や不得意な環境を調べることが考えられる。このことによって，数学的モデルからタマリンに関する新たな実験パラダイムを提案することが可能になる。

第 4 に今回再現の出来ていない現象について検討が必要である。実験 1 において，タマリンは最終実験で 1 セッション目にて 90% の確率で正解を選択していた。これに対し，エージェント 4 はランダムな選択より有意に多く正解のトレイを選んでいたが，タマリンの結果である 90% までには至らなかった。エージェント 4 にとって，最終テストは内部モデルでは解くことは出来ない課題であり，actor-critic モデルで正解を選ばなければならない。actor-critic モデルにとって最終課題は各トレイは経験済みでも，トレイの組み合わせ自体は未学習な課題である。そのため最終課題が始まってから，正解の学習を始めたため，高い正答率にまでは至らなかった。この問題を回避する上で様々の方法が考えられるが，数学的モデルが複雑になるのを避けるため今回は検討を見送った。今後はこの問題を解決でき，かつシンプルな形の数学的モデルの検討をする必要がある。

最後に，数学的モデルの対象を他の種にも広げることがあげられる。タマリンはヒトやチンパンジーと比較すると道具使用行動は得意ではない種である。チンパンジーの道具作成行動など，他にも高度な道具使用行動がある。これらの道具使用行動結果と計算機実験結果を比較することにより，各霊長類間の知能の差が詳細に表現できるものと考える。チンパンジーやヒトの道具使用行動に関しても同様に数学的モデル化を行う必要がある。

4.4.6 まとめ

本研究ではタマリンの道具使用行動の結果を元に，数学的モデルを用いてその能力を推測した。タマリンは自然界では道具使用行動を行わない種であるが，そのような種に対しても道具使用行動課題を調べることにより，その能力を数学的に議論することが可能であることを本研究において示した。道具使用行動はチンパンジーのような大型類人猿に関しては数多くの研究がなされている。特にチンパンジーは道具使用行動の中でも高度とされる，道具作成行動も報告されている (Whiten et al., 1999)。数学的モデルを用いた研究を行うことで，これらの高度な道具使用行動に関しても，そのアルゴリズムを厳密に議論することが可能である。これによりチンパンジーなどの高度な知能の設計原理に対して考察できる可能性がある。今後は，本研究のアプローチを他の道具使用行動を得意とする霊長類にも発展させることが 1 つの大きな課題となる。

謝辞

有益なコメントを頂きました査読者の方々に心から感謝いたします。 本研究の一部は科研費 (19700215) の補助を受けました。

第5章　終章

本稿にて紹介した研究は，道具使用行動を手がかりに，霊長類各種の持つアルゴリズムの推定し，霊長類の持つ高度な知能の解明を目的に行われた。また，本研究では，霊長類の道具使用行動の研究としては通常行なわれていない数学的な手法を用いた。そのため，各種の持つアルゴリズムを推定することと同時に，その手法の確立も本研究の大きな目的である。道具使用行動は多岐にわたるが，本研究ではその中でも代表的な課題の1つであるスティック課題を，ニホンザル (Iriki et al., 1996; Ishibashi et al., 2000; Hihara et al., 2003) やワタボウシタマリン (Santos et al., 2005; Santos et al., 2006) を対象にした研究に焦点を絞った。

第2章での研究は，主に博士前期課程に行ったものであり，当時共同研究を行っていた入来篤史先生らの研究グループから頂いた実験データを再現する数学的モデルを作成した。既にあった3つの実験データ (Iriki et al., 1996; Ishibashi et al., 2000; Hihara et al., 2003) を，既存の強化学習モデルをベースに道具の価値というパラメータとその計算機構を付与することにより全て再現することができた。特に Hihara らの行なった実験結果を再現するためには，従来から提案されている強化学習では再現することが出来ないため，なんらかの形で変更したモデルを用意する必要があった。数ある変更方法の中の一手法として，本研究では道具の価値というパラメータを導入した。このパラメータを用いて，餌の位置と道具の価値の値を基に電気生理による実験を提案し，その結果を予想することが出来た。

予想された結果は，モデルを作ってしまえば自明な結果であり，また他の再現

の仕方も当然存在する。例えば，道具の価値の計算方法は，Iriki らの結果から得られた神経細胞応答に基づいて作成したが，その計算方法は様々な方法が考えられる。本研究では，餌からスタートしてその価値を求めたが，手元の道具から環境推移を予測する方法も考えられる。この点を考慮していなかった点が第 4 章の問題提起につながった。ただし，これまでの研究で道具使用行動を焦点に当てた研究は数少なく，その後の研究を行うにあたって，基本的な数学的モデルの構造と，その問題提起が行えたことは，大変貴重な経験でもあり有益な研究であった。

第 3 章では，L 字型のスティックを用いて，各行動主体の持つ内部モデルの構造を推定するための実験を提案し，その実験結果をニューラルネットワークを用いて予測した。提案したタスクは，はじめのタスクでは L 字の向きが固定されており，その後行なわれるタスクでは L 字の向きが反転する。もし仮に，エージェントが我々ヒトと同じように，L 字型のスティックを引き寄せるためには L 字の先端部分が餌と接触する必要があることを学習していれば，反転したとしても特に苦も無く課題を成功することが出来ることが予想される。しかし，例えばサルのような場合には，必ずしも接触情報を見ているとは限らないような実験事例なども報告されている (Ishibashi et al., 2000; Hihara et al., 2003)。種によっては，接触情報を見ているのではなくて，報酬量と単に餌の位置がスティックの左側にあるという情報のみを覚えている可能性もある。サルなどの種が何を手掛かりにして道具使用行動課題を解いているのかを理解することは，サルが野生下で道具を使った事例がほとんど存在しないという事実に密接に関連していると考えこのような課題を考えた。

シミュレーション結果は，スティックと餌との接触状態を学習しているエージェントが新しい課題にも対応することが出来た。ただし，この結果は用いたニューラルネットワークに依存する可能性がある。異なるネットワーク構造，学習法則

を用いた場合には異なる結果が生まれる可能性がある。もし仮に，脳の数学的モデルが完全に理解されており，その分野が確立しているのであれば，このようなシミュレーション結果も価値があるように思われるが，残念ながら現状は，脳の数学的モデルは様々な可能性がり，実際に実験をしなければ，その妥当性を検証できないというのが実情である。この点が第 3 章の研究の反省点であった。架空の実験のみを提案し，その結果を数学的モデルで予測するというスタンスは現状ではかなり問題があり，第 2 章で紹介したような，ある実験結果を再現し，更に新たな実験を予測するという研究が現状では現実的で意味がある研究だと深く認識した。ただし，この第 3 章では，2 つのエージェントを用意して，実験結果を比較し，そのアルゴリズムを推定するという研究手法を取った。もし，既存の実験に対して，複数のアルゴリズムを用意しその再現性を検討すれば，実験を再現するための条件を絞ることにはなる。この手法は第 4 章において大きく役に立った。

第 4 章では，第 2 章，第 3 章，の研究手法の反省を基に，再現すべき既存の実験を選び，複数のアルゴリズムを用意し，その結果が再現できるためのアルゴリズムの条件を絞り込んだ。また，アルゴリズムも，モデルの種類とその計算の階層の深さという 2 種類の要素を入れることにより，再現対象の結果がアルゴリズムから自明でないような形を極力取った。その結果，ワタボウシタマリンが本研究で定義する「逆モデル」を用いて，1 つの道具のシミュレーションを用いている可能性を示唆した。また，このモデルの特性からタマリンが以下の能力を持つ可能性を示唆した；1）タマリンは課題開始時には actor-critic モデルよりも内部モデルを高頻度に利用して課題を解いている。2）actor-critic モデルの学習と比較して，確信度 q（actor-critic モデルと内部モデルの切替変数）の学習係数が小さい値を取る必要がある。3）確信度 q が問題の種類によっては，初期化する必要がある。また，上記の仮説の検証を行なうための実験を提案し，その妥当性が検討できるように

した。なお，第 4 章では，逆モデルや順モデルという用語が頻繁に出るが，一般的な順モデル，逆モデルではなく，本研究で定義した「逆モデル」と「順モデル」を比較した際に，タマリンは「逆モデル」を用いているとの意味である。この入出力関係に関しては，本研究で提示した以外の関係も十分あり得るため，今後の議論が必要である。

この章では再現すべき実験データとして Santos らによる実験 (2005; 2006) を選択した。Santos らが行なった研究は，ワタボウシタマリンに関して「できる課題」，「できない課題」が分類されており，能力を推定するという観点からすれば大変優れた研究であった。本来，タマリンは自然環境下では道具を使用する事例が報告されておらず，道具使用行動の対象として適切であったかどうかは議論の余地が残る。しかし，道具使用行動を霊長類の知能の尺度とみなす観点からすれば，自然環境下では道具を使用できない種だとしても，各種を訓練した場合，どの程度の道具使用行動が出来るのかを調べることはその認知能力を調査する上で価値がある。事実，今回の研究から示唆されることは，タマリンは餌の状態から必要な行動や道具の状態を推定しており，身におかれた環境から行動後の環境の変化を正確には予測していない可能性を示唆している。第 3 章でも述べたとおり，道具の使用方法や環境推移を正確に知ることがなくても，道具は使用することが可能である。この場合，問題はどの程度道具の状態遷移を理解しているのかを知ることであり，この点に関して実験・理論の両面から研究を進める必要性を感じる。

最後に，第 3 章，第 4 章で紹介した論文を通したことにより博士論文を提出することが出来たが，書き終えた後に強く思ったことは，現時点にてようやく霊長類研究の入り口に立つことが出来たという事である。道具使用行動の主役的実験の対象は，チンパンジーやヒトのような種であり，今回のニホンザルやワタボウシタマリンは自然界で道具使用行動を行なっていない。第 4 章にて用いた手法は

様々な実験データに適応することが可能だと考えられる。この手法をチンパンジーやヒトに適用することにより，より厳密に各種の能力を推定できると考えている。今後は，今回の Santos の事例に留まらず，更に多くの事例をモデル化する必要がある。具体的な形で数学的アプローチの有効性を見せ，多くの研究者がその必要性を感じ，活用できるようになる手法の土台を作ることが今後の目標である。

以上，各章の結果をまとめた。現在，数学的アプローチを用いた霊長類の研究は数少ないが，確実に有効な手法の 1 つである。今回の知見によって，霊長類の知能を理解する上で何かしらの貢献ができれば幸いである。

謝辞

本研究を進めるにあたり，長年に渡りご指導を頂きました東京工業大学　大学院総合理工学研究科　知能システム科学専攻　中村清彦教授に感謝いたします。日頃の議論を通じて多くの知識や示唆を頂き，ご助力いただいた佐賀大学　理工学部電子工学科　伊藤秀昭講師に感謝いたします。博士論文の審査にあたり有益なご助言を頂きました東京工業大学　大学院総合理工学研究科　知能システム科学専攻　小池康晴教授，青西亨准教授，木賀大介准教授，宮下英三准教授に感謝いたします。研究を進めるにあたって，ご助言を頂いた中村研究室の皆様に感謝いたします。

本論文を完成させるにあたり，ご配慮を頂きましたダイフク研究・研修センターの皆様に感謝いたします。

そして，長年に渡り応援して下さった両親，義父母，利広君，祖父母に感謝いたします。最後に，本研究を進めるにあたって常に影で支えてくれた美保さんと利晴君に感謝いたします。

2010 年 9 月　毬山　利貞

参考文献

[1] Barto AG (1995) In:Houk JC, Davis JL, Beiser DG(eds) Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, 215-232

[2] Beck BB (1980) Animal tool behavior: the use and manufacture of tools by animals. Garland STPM Press, New York

[3] Beran MJ (2004) Long-term retention of the differential values of Arabic numerals by chimpanzees (Pan troglodytes), Anim Cogn 7:86-92

[4] Boesch C, Boesch H (1990) Tool use and tool making in wild chimpanzees. Folia Primatol 54:86-99

[5] Boesch C, Marchesi P, Fruth B, Joulian F (1994) Is nut cracking in wild chimpanzees a cultural behavior? J Hum Evol 26:325-338

[6] Carpenter A (1887) Monkeys opening oysters. Nature 36-53

[7] Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral straital systems for behavioral control. Nat Neurocsci 8:1704-1711

[8] Desmurget M, Grafton S (2000) Forward modeling allows feedback control for fast reaching movements. Trends in Cogn Sci 4:423-431

[9] Doya K (1999) What are the computations of the cerebellum, the basal ganglia and cerebral cortex. Neural Networks 275:1593-1599

[10] Fuster JM (1997) The prefrontal cortex. Lippincott-Raven Publishers, New York

[11] Goldman-Rakic PS (1988) Topography of cognition: Parallel distributed networks in primate association cortex. Ann Rev Neurosci 11: 137-156

[12] Goodall J (1963) Feeding behaviour of wild chimpanzees: a preliminary report. Symp Zool Soc Lond 10:39-48

[13] Goodall J (1964) Tool-using and aimed throwing in a community of freeliving chimpanzees. Nature 201:1264-1266

[14] Goodall J (1986) The chimpanzees of Gombe: patterns of behavior. Harvard University Press, Cambridge, Mass

[15] Grush R (2004) The emulation theory of representation: motor control, imagery and perception. Behavi Brain Sci 27:377-396

[16] Hannah A, McGrew W (1987) Chimpanzees using stones to crack open oil palm nuts in Liberia. Primates 28:31-46

[17] Hauser MD (1997) Artifactual kinds and functional design features. What a primate understands without language. Cognition 64:285-308

[18] Hauser MD, Kralik J, Botto-Mahan C (1999) Problem solving and functional design features: experiments on cottontop tamarins (Saguinus oedipus). Anim Behav 57:565-582

[19] Hayashi M, Matsuzawa T (2003) Development of orienting manipulation in infant chimpanzees. Anim Cogn 6:225-233

[20] Hayashi M, Mizuno Y, Matsuzawa T (2005) How does stone-tool use emerge? Introduction of stones and nuts to naive chimpanzees in captivity. Primates 46:91-102

[21] Head H, Holms G (1911) Sensory disturbance from cerebral lesions. Brain 34:102-254

[22] Hesslow G (2002) Conscious thought as simulation of behaviour and perception. Trends in Cogn Sci 6:242-247

[23] Higuchi S, Imamizu H, Kawato M (2007) Cerebellar activity evoked by common tool-use execution and imagery tasks: An fMRI study. Cortex 43(3):350-358

[24] Hihara S, Obayashi S, Tanaka M, Iriki A (2003) Rapid learning of sequential tool use by macaque monkeys. Physiol Behav 78(3):427-434

[25] Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, Doya K (1999) Parallel neural networks for learning sequential procedures. Trends Neurosci 10:464-471

[26] Houk JC, Adams JL, Barto AG (1995) In Houk JC, Davis JL, Beiser, DG(eds) Models of Information Processing in the Basal Ganglia. 249-270, MIT Press, Cambridge.

[27] Humle T, Matsuzawa T (2002) Ant-Dipping Among the Chimpanzees of Bossou, Guinea and Some Comparisons With Other Sites. American Journal of Primatology 58:133-148

[28] Imamizu H, Miyauchi S, Tamada T, Sasaki Y, Takino R, Putz B, Yoshioka T, Kawato M (2000) Human cerebellar activity reflecting an acquired internal model of a new tool. Nature 403:153-154

[29] Imamizu H, Kawato M (2009) Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions. Psychol Res 73:527-44

[30] Ishibashi H, Hihara S, Iriki A (2000) Aquisition and development of monkey tool-use: behavioral and kinematic analysis. Can J Physiol Pharmacol 78:958-966

[31] Iriki A, Tanaka M, Iwamura Y (1996) Cording of modeified body shema during tool use by macaque postcentral neurons. Neuroreport 7:2325-2330

[32] Iriki A (2006) The neural origins and implications of imitation, mirror neurons and tool use. Curr Opin Neurobio 16:660-667

[33] Ito M (1984) The cerebellum and neural motor control. Raven Press, New York

[34] Johnson-Frey SH (2004) The neural bases of complex tool use in humans. Trends Cogn Sci 8:71-8

[35] Kawai N, Matsuzawa T (2000) Numerical memory span in a chimpanzee. Nature 403:39-40

[36] Kawato M, Furukawa K, Suzuki R (1987) A hierarchical neural-network model for control and learning of voluntary movement. Biol Cybern 57:169-185

[37] Kawato M (1999) Internal models for motor control and trajectory planning. Curr Opin Neurobiol 9:718-727

[38] Köhler W (1927) The mentality of apes. Brace and Company, New York

[39] Kortlandt A, Holzhaus E (1987) New data on the use of stone tools by chimpanzees in Guinea and Liberia. Primates 28:473-496

[40] Kosslyn SM, Ganis G, Thompson WL (2001) Neural foundations of imagery. Nature Reviews Neurosci 2:635-642

[41] Luciana M, Nelson CA (1998) The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. Neuroposchologia 36:273-293

[42] Malaivitnond S, Lekprayoon C, Tandavanittj N, Panha S, Cheewatham C, Hamada Y (2007) Stone-Tool Usage by Thai Long-Tailed Macaques(Macaca fascicularis). American J of Primat 69:227-233

[43] Maravita A, Iriki A (2004) Tools for the body (schema). Trends Cogn Sci 8:79-86

[44] Mariyama T, Hihara S, Ishibasi H, Iriki A, Nakamura K (2001) A reinforcement learning model with tool-evaluation for macaque monkey tool-use. Neuroscience Research, Supplement 25:S200

[45] Mariyama T, Itoh H (2009) Towards a Comparative Theory of the Primates ' Tool-Use Behavior. Koppen M et al. (Eds.): ICONIP 2008, Part I:327-335

[46] Mariyama T, Itoh H, Nakamura K (2010) A mathematical model for Tamarin's tool-use behavior. Primate Research 26(1):13-33 (in Japanese)

[47] Matsuzawa T (1985) Use of numbers by a chimpanzee. Nature 315:57-59

[48] Matsuzawa T (2001) Primate foundations of human intelligence: A view of tool use in nonhuman primates and fossil hominids. In Matsuzawa T (ed): Primate origins of human cognition and behavior.pp 3-25, Springer, Berlin Heidelberg, New York,

[49] McGrew WC (1974) Tool use by wild chimpanzees in feeding upon driver ants. J Hum Evol 3:501-508

[50] Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. J Neurophysiol 16:1936-1947

[51] Menz MM, Blangero A, Kunze D, Binkofski F (2010) Got it! Understanding the concept of a tool. Neuroimage 51:1438-44

[52] Nakahara H, Doya K, Hikosaka O (2001) Parallel cortico-basal ganglia mechnisms for acquisition and execution of visuomoter sequences - a computational approach. J Cogn Neurosci 13:626-647

[53] Nishida T (1973) The ant-gathering behavior by the use of tools among wild chimpanzees of the Mahale Mountains. J Human Evol 2:357-370

[54] Nishida T, Matsusaka T, McGrew WC (2009) Emergence, propagation or disappearance of novel behavioral patterns in the habituated chimpanzees of Mahale: a review. Primates 50:23-36

[55] Nishijo H, Ono T, Nishino H (1988) Single neuron responses in amygdala of alertmonkey during complex sensory stimulation with affective significance. J Neurosci 8:3570-3583

[56] Obayashi S, Suhara T, Kawabe K, Okauchi T, Maeda J, Akine Y, Onoe H, Iriki A (2001) Functional brain mapping of monkey tool use. Neuroimage 14:853-861

[57] Oztop E, Kawato M, Arbib M (2006) Mirror neurons and imitation: A computationally guided review. Neural Network 19(3):254-271

[58] Ramelhart DM, Hinton GE, Willams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DM, McClelland JL (eds) Parallel distributed processing: exploration in the microstructure of cognition. Foundations, vol. 1,318-362. MIT Press, Cambridge

[59] Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169-192

[60] Santos LR, Pearson HM, Spaepen GM, Tsao F, Hauser MD (2006) Probing the limits of tool competence: Experiments with tow non-tool-using species (Cercopithecus asthiops and Saguinus oedipus). Anim Cogn 9:94-109

[61] Santos LR, Rosati A, Sproul C, Spaulding B, Hauser MD (2005) Means-means-end tool choice in cotton-top tamarins (Saguinus oedipus): finding the limits on primates ' knowledge of tools. Anim Cogn 8:236-246

[62] Schultz W, Apicella P, Ljungberg T, Romo R, Scarnati R (1993) Reward-related activity in the monkey striatum and substantia nigra. Prog Brain Res 99:227-235

[63] Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599

[64] Shadmehr R, Mussa-Ivaldi FA (1994) Adaptive Representation of Dynamics during learning of a motor task. J Neurosci 12:3208-3224

[65] Sugiyama Y, Koman J (1979) Tool-using and -making behavior in wild chimpanzees at Bossou, Guinea. Primates 20:513-524

[66] Sugiyama Y, Koman J, Bhoye Sow M (1988) Ant catching wands of wild chimpanzees at Bossou, Guinea. Folia Primatol 51:56-60

[67] Sugiyama Y (1995) Tool-use for catching ants by chimps at Bossou and Monts Nimba. Primates 36:193-205

[68] Sumita K, Kitahara-Frisch J, Norikoshi K (1985) The acquisition of stone-tool use in captive chimpanzees. Primates 26:168-181

[69] Suri RE (2001) Anticipatory responses of dopamine neurons and cortical neurons reproduced by internal model. Exp Brain Res 140:234-240

[70] Suri RE, Schultz W (2001) Temporal difference model reproduces anticipatory neural activity. Neural Comput 3:841-862

[71] Sutton RS, Barto AG (1995) In Gabriel M, Moore J(eds) Learning and Computational Neuroscience:Fundation of Adaptive Networks pp 497-537, MIT Press, Camdridge, Massachusetts

[72] Sutton RS, Barto AG (1998) Reinfoecement Learning, MIT Press, Cambridge, Massachusetts

[73] Takeshita H, Walraven V (1996) A comparative study of the variety and complexity of object manipulation in captive chimpanzees (Pan troglodytes) and Bonobos (Pan paniscus). Primates 37:423-441

[74] Takeshita H (2001) Development of combinatory manipulation in chimpanzee infants (Pan troglodytes). Anim Cogn 4:335-345

[75] Takeshita H, Fragaszy D, Mizunoc Y, Matsuzawa T, Tomonaga M, Tanaka M (2005) Exploring by doing: How young chimpanzees discover surfaces through actions with objects.: Infant Behavior and Development 28:316-328.

[76] Torigoe T (1985) Comparison of object manipulation among 74 species of Non-human primates. Primates 26:182-194

[77] Van Lawick-Godall J (1970) Tool-using in primates and other vertebrates. In Advances in the study of behavior, vol.e, eds., D. Lehrman et al., pp. 195-249. New York: Academic.

[78] Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C (1999) Cultures in chimpanzees. Nature 399:682-685

[79] Whiten A, Horner V, Waal FBM (2004) Conformity to cultural norms of tool use in chimpanzees. Nature 437:738-740

[80] Watanabe M (1996) Reward expectancy in rimate prefrontal neurons. Nature 382:629-632

[81] Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. Neural Networks 11:1317-1329

[82] Wood JN, Glynn DD, Phillips, BC, Hauser MD (2007) The Perception of Rational, Goal-Directed Action in Nonhuman Primates. Science 317:1402-1405