Vol.2013-MUS-98 No.16
Vol.2013-EC-27 No.16
2013/3/16

生物の基本原則の導入によるビデオゲームCOMプレイヤの「人間らしい」振る舞いの自動獲得

藤井 叙人[1,a)]　佐藤 祐一[1]　若間 弘典[1]　片寄 晴弘[1,b)]

概要：ビデオゲームのコンピュータ (COM) の振る舞いのデザインにおいて，強い COM の自律的獲得はゴールが見え始めた一方，人間にとって自然と映る COM をどう構成するかが，ゲーム AI 領域の課題になりつつある．本研究では，人間らしい振る舞いを表出する COM を，開発者の経験に基づき実現するのではなく，『生物の基本原則』の条件下で，自律的に獲得することを目指す．機械学習に対して『生物の基本原則』を導入するだけでなく，本論文では経路探索 (A*) で獲得された最適解をもつ強い COM に対しての適用可能性も検討し，振る舞い獲得の手法を問わず本研究が有用であることを示す．アクションゲームの “Infinite Mario Bros.” を学習対象とし，『生物の基本原則』として「身体的な制約：“ゆらぎ” “遅れ” “疲れ”」に加え，「生き延びるために必要な欲求 (本能)：“慣れと新奇に対する希求”」を新たに考慮し，強化学習 (Q 学習)，ないし，経路探索 (A*) の枠組みに導入することで，COM の人間らしい振る舞いの獲得を試みる．最後に，獲得された COM の振る舞いの『人間らしさ』に関する初期的検討を実施する．

1. はじめに

ビデオゲームは，人間プレイヤの相手 (協力関係ないし敵対関係) という視点から分類すると，人間が相手を務める場合と，COM が相手を務める場合の二つに分類される．後者においては，COM の振る舞いや戦略レベルのデザインが，プレイヤのゲームに対する印象に大きな影響を与えると言っても過言ではない．市販ビデオゲームでは，売上に直結する事項であるため，ゲームプログラマのヒューリスティックスに基づく綿密な作り込みにより実現されてきた．この実装に係る作業負荷の削減，及び，ゲーム AI 領域での興味から，国内外において COM の振る舞いの自律的獲得に関する研究が進められてきた [1], [2], [3]．結果，人間の熟達者を凌駕する**勝つための『強い』COM** の自律的獲得はゴールが見え始めた一方で，獲得された COM の振る舞いは過度に最適化され，人間プレイヤにとって機械的に映る，という問題が浮き彫りになった．

人間プレイヤを**楽しませるための『人間にとって自然』と映る COM** を試作検討するものとして，従来どおりプログラマがアドホックにデザインする手法の他に，人間プレイヤの振る舞いを記録し機械学習により模倣する手法 [4], [5],『強い』COM に対して意図的にエラーを導入することで自然な強くなさを実現する手法 [6] などが挙げられる．しかし，COM の振る舞いや戦略に関わるプレイスタイルのフレームワークは，ヒューリスティックに与えてやるほか無いのが現状である．

本研究では，**『人間にとって自然』と映る COM** の振る舞いを，開発者のヒューリスティックスで実現するのではなく，『生物の基本原則』の条件下での機械学習により，自律的に獲得する機構（自律的行動獲得機構）について検討する．また，既存手法 (経路探索)[1] で獲得された『強い』COM に対する，『生物の基本原則』の適用可能性についても検証する．アクションゲームの “Infinite Mario Bros.” を学習対象とし，『生物の基本原則』として，**「身体的な制約：“ゆらぎ” “遅れ” “疲れ”」**と**「生き延びるために必要な欲求 (本能)：“慣れと新奇に対する希求”」**という制約条件を，経路探索や強化学習のフレームワーク [1], [3], [7] に組み込む．これらの条件は，生物なら誰しもが持つ基本的な原則であるため，学習手法やゲームジャンルといった対象に限定されることなく，『人間らしい』COM の振る舞いを表出するフレームワークとして導入が可能である．

以下，第 2 章で，関連研究を紹介し，第 3 章で，学習対象とする “Infinite Mario Bros.” の仕様と，採用する機械学習手法について記述する．第 4 章で，自律的行動獲得機構のフレームワークについて述べ，第 5 章で，学習実験による COM の振る舞いの獲得状況について検証する．

[1] 関西学院大学大学院 理工学研究科 人間システム工学専攻
Department of Human System Interaction, School of Science and Technology, Kwansei Gakuin University
[a)] nobuto@kwansei.ac.jp
[b)] katayose@kwansei.ac.jp

Vol.2013-MUS-98 No.16
Vol.2013-EC-27 No.16
2013/3/16

2. 関連研究

本章では，振る舞いや戦略を自動的に獲得する関連研究として，COM の『強さ』を追求した研究例と，『人間にとって自然』と映る COM を検討した研究例を紹介する．

2.1 COM の『強さ』の追求

振る舞いや戦略を自動的に獲得する手法として，経路探索問題に帰着する手法 [1] と，人間のプレイデータや試行による機械学習を用いる手法 [3] がある．

経路探索問題によるアプローチのうち代表的なものとして，Robin は，2009 年の Mario AI Competition において，A*アルゴリズムに基づいたエージェントを構築し優勝している [1]．Mario AI Competition とは，Infinite Mario Bros.[8] （ランダムに生成されるマリオライクなステージを，制限時間中に攻略するアクションゲーム．詳しい仕様は第 3 章で説明する．）を対象としたエージェント評価コンテストである [9], [10]．マリオや敵キャラクタの動きを事前に学習・解析し，A*アルゴリズムを用いたルート探索によってステージを攻略することで，敵キャラクタを可能な限り避け，ステージをより早く，より遠くまで攻略することが可能となっている．

機械学習によるアプローチの代表的なものとして，藤田らは，カードゲームの Hearts を題材とし，Q 学習を用いて戦略獲得に成功している [3]．カード 52 枚を使用するため巨大な状態空間となること，相手の所持するカードは観測できないため部分観測状況となること，4 人対戦のマルチエージェントゲームであること，の 3 つを Hearts における戦略学習の困難性と考察している．その上で，解決手法として，パーティクルフィルタ，相手の行動予測器，状態を評価する状態価値関数，ゲームの特徴に基づく次元圧縮を提案している．計算機実験として，提案手法に基づく学習エージェントと，人間の熟達者とを対戦させた結果，人間の熟達者よりも優れた戦略を得ることに成功している．

経路探索問題や機械学習を用いて獲得されたこれらの COM の振る舞いは，極めて最適であるが故に人間プレイヤにとっては機械的であり，人間プレイヤの代替として扱うことは難しい．

2.2 『人間にとって自然』と映る COM の検討

人間プレイヤを楽しませるために，COM の『人間にとって自然』と映る COM を検討した関連研究として，Jacob らは，2012 年の The 2K BotPrize[11] において，大会史上初となる，人間よりも人間らしいと評価されるエージェントの構成に成功している [4]．The 2K BotPrize とは，2008 年から開催されている，FPS(一人称視点シューティングゲーム) を対象とした，エージェントの人間らしさを競う評価コンテストである．人間プレイヤの動作をトレースしたデータベースを基に，人間らしいと思われる動作を決定論的に定義し，ニューラルネットにおける制約として適用することで，対戦相手のプレイヤから人間らしいと評価される振る舞いの獲得を可能としている．また，池田らは，コンピュータ囲碁を対象に，既存の『強い』COM に意図的に人間らしいミスをさせることで，手加減と思われない程度の『強くなさ』の初期的検討を実施している [6]．現在の局面における予測勝率と候補手の選択確率を用いた形勢の制御，楽観派や悲観派といったプレイスタイルによる獲得戦略の分析をしており，レベルデザインにおける一アプローチとして提案している．

図 1 “Infinite Mario Bros.” のゲーム画面
エージェントは，制限時間内に進んだ距離や倒した敵の数から算出されるスコアを競う．

上記の手法は，『人間にとって自然』と思われる振る舞いを，開発者が恣意的に定義したものであり，学習における作業負荷の軽減，学習フレームワークの汎用性の確保という視点から見ると，適しているとは言いがたい．

3. 学習対象ゲームと学習手法

『生物の基本原則』を機械学習エージェントに組み込み，『人間にとって自然』と映る COM の振る舞いの自動獲得を目指すにあたり，学習対象とするゲームと，採用する機械学習手法を検討する必要がある．本章では，学習対象とするアクションゲーム “Infinite Mario Bros.” の仕様と，強化学習手法の一つである Q 学習について説明する．

3.1 Infinite Mario Bros.

本研究では，ゲームの仕様やゲーム環境パラメータが公開されている，“Infinite Mario Bros.”[8] を学習対象とし，振る舞いの獲得と，その比較，検証を実施する．節 2.1 で述べた Robin の A*アルゴリズムエージェント [1] を 1 つの最適解として扱う．“Infinite Mario Bros” は，世界的に有名なゲームである “Super Mario Bros.” を模したアクションゲームであり，そのゲーム画面を図 1 に示す．“Infinite Mario Bros.” における仕様は以下のとおりである．

- **ステージの自動生成**
  事前に与えたシード値に従って無限にステージが生成

される．

- **エージェントの操作キャラクタ（マリオ）**
  エージェントはマリオを操作する．エージェントによるマリオの操作はキー入力 (LEFT, RIGHT, DOWN, SPEED, JUMP) により行う．フレーム毎のキーの押下状態により，マリオは対応した行動を行う．また，マリオには「でかマリオ」「ちびマリオ」が存在する．「でかマリオ」でダメージを受けた場合は「ちびマリオ」に変化し，「ちびマリオ」でダメージを受けた場合は死亡する．ダメージについては，後述の接触判定において説明する．穴に落ちた場合は「でかマリオ」「ちびマリオ」を問わず，死亡する．
- **敵キャラクタ**
  複数種類の敵キャラクタが登場し，敵キャラクタはそれぞれ独自のアルゴリズムで動作している．エージェントには，この敵キャラクタを避けて進むか，倒して進むか，どのように処理するかが求められる．マリオは敵キャラクタとの接触判定によってダメージを受ける場合がある．踏むことができる敵キャラクタは，踏む以外の行動で接触した場合ダメージを受ける．踏むことができない敵キャラクタは，接触した場合ダメージを受ける．
- **スコアの獲得**
  マリオが死亡する，または，設定された制限時間に達すると攻略は終了し，スコアを獲得する．スコアは Mario AI Competition で規定されている評価関数で計算され，敵キャラクタを倒した数，ステージを攻略した距離などに応じてスコアが上昇する．獲得スコアが高いほど優秀なエージェントとして評価される．
- **エージェントの観測情報**
  マリオの座標，マリオの状態，画面内の敵キャラクタの種類および座標，地形情報といったものを観測情報として得ることができる．エージェントの観測する地形情報は，ステージに配置されているブロックのうち，画面内にある 22 × 22 のブロックの配置情報となる．“Infinite Mario Bros.” は毎秒 24 フレームで動作しており，エージェントは毎フレーム観測情報を受け取り，マリオの行動制御を行うためのキー入力を返す必要がある．

3.2 Q 学習の概要

振る舞いの学習には大量の学習データが必要となるが，ビデオゲームにおいては，教師となるプレイデータが大量に用意できないため，学習試行により振る舞いの獲得が可能な強化学習の手法を用いる．本研究で採用した，強化学習手法の一つである Q 学習は，ゲーム内での形勢を報酬として直感的に設定できる点，ルール獲得という形で学習が進む点で，ゲームクリエイタが利用しやすいというメリットがある．Q 学習では，あるゲーム状態を s，その状態下でエージェントが可能な行動を a とした場合，状態 s と行動 a を組とし，その組に対する Q 値とよばれる評価値を算出する．あるゲーム状態での最適行動は数式 1 で決定され，Q 値が最も高い行動が最適であると出力する．また，Q 学習における Q 値の更新式は数式 2 であり，エージェントが行動するたびに Q 値を更新することで振る舞いの獲得が可能となる．

$$argmax_{a_t} Q(s_t, a_t) \tag{1}$$

$$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha((r + \gamma max_p Q(s_{t+1}, p)) \tag{2}$$

数式 2 において，t はゲームのフレーム，s_t はフレーム t における状態，a_t はフレーム t においてとった行動，$Q(s_t, a_t)$ は (s_t, a_t) に対応する Q 値である．α は学習率と呼ばれ，Q 値の更新において新たな報酬をどれだけ重視するかを示す値であり，γ は割引率と呼ばれる 0 以上 1 以下の定数である．r は状態 s_t において行動 a_t をとったことによって得られる報酬である．エージェントの行動選択手法としては $\epsilon - greedy$ 法を用いる．$\epsilon - greedy$ 法は，$1 - \epsilon$ の確率で Q 値が最大となる最適行動を選択し，ϵ の確率でランダムに行動を選択する．

4. 学習フレームワークの構築

本章では，『生物の基本原則』を強化学習の制約条件として賦課した学習フレームワークについて述べる．強化学習の枠組みにおける『生物の基本原則』の扱いについて説明し，観測情報とエージェントの行動に関する状態圧縮手法について述べる．次に，Q 学習における報酬の設定方法について記述する．また，節 2.1 で述べた Robin の A*アルゴリズムエージェント (最適解)[1] に対しても，『生物の基本原則』を導入することを試みる．

4.1 『生物の基本原則』の強化学習への導入

『生物の基本原則』である「身体的な制約：ゆらぎ，遅れ，疲れ」，「生き延びるために必要な欲求 (本能)：「慣れ」親しんだものに対する希求と「新奇」なるものに対する希求」として，強化学習エージェントには以下の 4 つの制約条件を賦課する．

(a) **観測位置情報のゆらぎ**
人間プレイヤは，観測した操作キャラクタと敵キャラクタの位置（座標）を正確に認識することは難しく，誤差（ゆらぎ）が生じる．これを観測位置情報のゆらぎと定義し，操作キャラクタ，敵キャラクタの座標に対してガウスノイズを付与したものを，エージェントの観測とすることで再現する．節 3.2 で述べた数式 1 の $Q(s_t, a_t)$ の計算の際に s_t として与える操作キャラクタと敵キャラクタの位置情報に対して，さまざまな分散をもつガウスノイズを付与する．

(b) **観測から行動制御における情報認識の遅れ**
人間プレイヤは，ゲームの状態を観測し認識してから，実際に行動をするまでに遅れが発生する．これを情報認識の遅れと定義し，エージェントの観測するゲーム状態を数フレーム過去の情報にすることで再現する．節 3.2 で述べた数式 1 の $Q(s_t, a_t)$ の計算の際に s_t として与えるゲーム状態を，数フレーム前の状態とする．

(c) **キー操作変更による疲れ**
人間プレイヤは，キー操作を極めて短時間で，または，長時間連続して実施すると疲れが生じる．これをキー操作の疲れと定義し，エージェントは可能な限り少ないキー操作で攻略するよう学習することで再現する．節 3.2 で述べた数式 2 の Q 値の更新の際に，報酬 r にキー操作変更による負の報酬を与える．報酬 r については，節 4.3 で詳しく述べる．

(d) **「訓練」と「挑戦」のバランス**
人間プレイヤは，同じ行動を繰り返す事で訓練 (鍛錬) して慣れていく一方で，同じ行動に慣れすぎて飽きたり，その行動で失敗を繰り返したりすると，飽きや失敗を解消するための新奇な行動に挑戦する．これを訓練と挑戦のバランスと定義し，エージェントは失敗を繰り返した際に，新たな行動に挑戦するよう学習することで再現する．節 3.2 で述べたエージェントのランダム行動選択確率 ϵ の設定に関して，失敗を繰り返しているゲーム状態 s では大きな値を設定することで，新奇な動作に挑戦する傾向を高める．逆に，失敗をほとんどしないゲーム状態 s では小さな値を設定し，同じ動作を訓練する傾向を高める．

4.2 強化学習におけるゲーム状態の圧縮

節 4.1 の制約が課されたエージェントにおいて，現実的な時間で Q 学習が収束し，COM の振る舞いを獲得できるよう，ゲーム状態の次元を圧縮する必要がある．まず，エージェントの観測情報を以下のとおりに圧縮する．

- **マリオを中心に 7 × 7 ブロックの地形と敵配置**
  エージェントが観測可能な地形情報は画面を 22 × 22 ブロックに分割したものである．しかし，1 フレームあたりのマリオの移動距離は小さく，画面内全ての地形情報や敵の配置がマリオの行動に影響することはない．そこで，学習に使用する地形情報と敵の配置は，マリオを中心とした 7 × 7 ブロックとする．
- **「でかマリオ」か「ちびマリオ」か**
  「でかマリオ」か「ちびマリオ」かは，より長く攻略を進めるうえで重要な要素である．
- **マリオの進行方向**
  マリオの進行方向を 8 方向＋停止の 9 状態とする．

次に，マリオの行動の設定について述べる．マリオの制御はキー入力によって行う．このキー入力の組み合わせにおいて，行動制御に影響がある 10 パターンの組み合わせを，選択可能な行動として設定する（表 1）．

表 1 行動の種類とキー入力の組み合わせ

行動の種類	(Left,Right,Down,Jump,Speed)
右に歩く	(OFF,ON,OFF,OFF,OFF)
右に走る	(OFF,ON,OFF,OFF,ON)
右に歩きジャンプ	(OFF,ON,OFF,ON,OFF)
右に走りジャンプ	(OFF,ON,OFF,ON,ON)
左に歩く	(ON,OFF,OFF,OFF,OFF)
左に走る	(ON,OFF,OFF,OFF,ON)
左に歩きジャンプ	(ON,OFF,OFF,ON,OFF)
左に走りジャンプ	(ON,OFF,OFF,ON,ON)
しゃがむ	(OFF,OFF,ON,OFF,OFF)
静止	(OFF,OFF,OFF,OFF,OFF)

節 3.2 で述べた $Q(s_t, a_t)$ の算出において，s_t と a_t に上記の状態圧縮を施すことで，計算量を削減し現実的な学習時間で学習を収束させることが可能となる．

4.3 強化学習における報酬の設定

Mario AI Competition では「敵キャラクタを可能な限り避け，ステージをより早く，より遠くまで攻略する」ことが目標とされている．そのため，ステージを早く攻略することに対して正の報酬を与え，逆にダメージを受ける，死亡するといった攻略を阻害する要因に対して負の報酬を与えることが望ましい．また，節 4.1 で述べた，キー操作による疲れを実現するため，キー操作を変更した場合は負の報酬を与える必要がある．そこで，報酬 $reward$ を以下のとおり設定する．

$$reward = distance + damaged + death + keyPress \tag{3}$$

数式 3 において，$distance$ は行動によって進んだ距離であり，そのまま正の報酬とする．$damaged$ は行動によってダメージを受けた場合に与える負の報酬，$death$ は行動によって死亡した場合に与える負の報酬である．また，$keyPress$ は前フレームから行動を変更した場合に与える負の報酬である．

4.4 『生物の基本原則』の経路探索問題への導入

経路探索問題によるアプローチで獲得された最適解を有する Robin の A*アルゴリズムエージェントにおいて，生物の基本原則に沿った以下の 3 つの制約条件を賦課する．(d) 訓練と挑戦のバランスに関しては，学習フェーズを持たない経路探索では実現不可能であるため対象外とする．

(a) **観測位置情報のゆらぎ**
節 4.1 での強化学習への導入と同様とする．経路探索エージェントは，ノイズが付与された観測情報における最適行動を求めることになる．

Vol.2013-MUS-98 No.16
Vol.2013-EC-27 No.16
2013/3/16

(b) **観測から行動制御における情報認識の遅れ**
節 4.1 での強化学習への導入と同様とする．経路探索エージェントは，数フレーム過去における最適行動を求めることになる．

(c) **キー操作変更による疲れ**
人間プレイヤは，極めて短時間でのキー操作の変更には限界がある．そこで，経路探索エージェントに対して，極めて短時間でのキー操作の変更を禁止することで，キー操作の疲れを再現する．

5. 計算機実験と獲得された振る舞いの検証

本章では，本研究の自律的行動獲得機構による強化学習エージェントについて，学習試行により正常に学習が進んでいるかどうか確認する．また，『生物の基本原則』を導入した強化学習エージェント，経路探索エージェントの振る舞いについて，『生物の基本原則』を導入していない場合の振る舞いと比較して，『人間にとって自然』と映るような振る舞いが表出しているかどうか検証する．

5.1 強化学習エージェントの学習性能の検証

本研究で提案した自律的行動獲得機構が正常に動作していることを示すため，(a) 観測位置情報の「ゆらぎ」，(b) 観測から行動制御における情報認識の「遅れ」，のパラメータセットを変更し，獲得スコアの推移を調べる．強化学習エージェントに与える「ゆらぎ」と「遅れ」のパラメータセットとして，以下の 3 セットを設定し，3 つの強化学習エージェントを生成した．

- ゆらぎ 0.0 ブロック・遅れ 0 フレーム (0.0 秒)
- ゆらぎ 0.5 ブロック・遅れ 6 フレーム (0,25 秒)
- ゆらぎ 1.0 ブロック・遅れ 12 フレーム (0.5 秒)

毎試行ランダム生成されるステージを対象として学習試行を行い，学習試行回数は 10 万ゲーム，200 ゲームごとの獲得スコアの平均をとる．Q 学習に関連するパラメータ設定として，学習率 α を 0.2，割引率 γ を 0.9，$\epsilon - greedy$ 法におけるランダム選択確率を 0.05 と設定した．また，報酬 r における $distance$ は進んだ距離 ×2.0，$damaged$ は-50.0，$death$ は-100.0，$keyPress$ は-5.0 とした．獲得スコアの比較対象として，最適解である Robin の A*アルゴリズムエージェント [1](節 2.1 参照) を用いた．学習試行実験の結果を図 2 に示す．図 2 から，どのパラメータセットにおいても正常に学習が進んでいることが確認できた．最適解に相当する A*アルゴリズムエージェントの獲得スコアほどではないものの，自律的行動獲得機構により，比較的良いパフォーマンスの振る舞いが獲得できていることが示された．

5.2 獲得された振る舞いの検証

『生物の基本原則』を導入した強化学習エージェント，経路探索エージェントの振る舞いが，『生物の基本原則』を導入していない振る舞いと比べ，どのように変化したかを検証する．

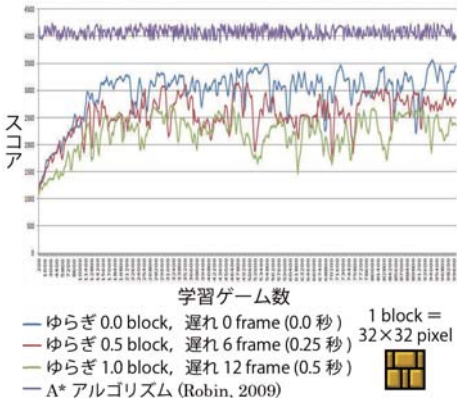

図 2 異なるパラメータセットで学習した際の獲得スコアの推移
どのパラメータセットにおいても正常に学習が進んでいる．

強化学習エージェントについては，毎試行同じステージが生成されるようにした状態で 5 万ゲームの学習試行を行い，比較的高いスコアを獲得したプレイにおける振る舞いを比較する．節 5.1 の実験環境からの変更として，ランダム選択確率 ϵ を 0.0125 と設定した．『生物の基本原則』を導入した強化学習エージェントは，ゆらぎ 0.25 ブロック，遅れ 3 フレーム (0.125 秒)，報酬 r の $keyPress$ は-5.0，失敗を繰り返しているゲーム状態での ϵ は 0.2125 とした．導入していないものは，ゆらぎ 0.0 ブロック，遅れ 0 フレーム (0.0 秒)，報酬 r の $keyPress$ は 0.0，失敗の有無に関わらず ϵ は 0.0125 のままとした．また，経路探索エージェントについては，『生物の基本原則』を導入したものは，ゆらぎ 1/12 ブロック，遅れ 2 フレーム (1/12 秒)，3 フレーム (1/8 秒) 未満でのキー操作変更を禁止した．導入していないものは，ゆらぎ 0.0 ブロック，遅れ 0 フレーム (0.0 秒)，キー操作変更の制限は無しとし，A*アルゴリズムエージェントと同等である．

比較の結果，強化学習エージェント，経路探索エージェント共に，『生物の基本原則』の導入の有無によって，表出した振る舞いの傾向に差異があった（図 3〜図 5）．敵を回避する場面，大量の敵が存在する場面，穴を飛び越える場面において，導入無しでは，最小限のジャンプ，かつ，ノンストップで攻略しているのに対し，導入有りでは，敵の前で一瞬止まる，安全になるまで待つ，余裕を持ってジャンプする，といった振る舞いが表出した．導入無しでは，パフォーマンスのみを重視しているが，導入有りでは，安全性も考慮した振る舞いを獲得しているといえる．

次に，『生物の基本原則』を導入した強化学習エージェントについて，訓練をせず，失敗に対する挑戦のみを実施するよう，ランダム選択確率 ϵ を 0.0，失敗を繰り返しているゲーム状態での ϵ を 0.2 に変更する．挑戦のみ実施するエージェントで獲得された振る舞いは，コントローラの操作やゲームのルールに慣れていない，あたかもゲームの初

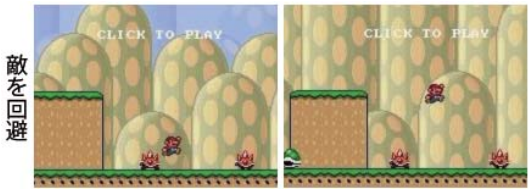

図 3 導入無し（左）と導入有り（右）での比較 (敵を回避)
導入無しでは最小限のジャンプで走り抜けるのに対し，導入有りでは大きくジャンプする．

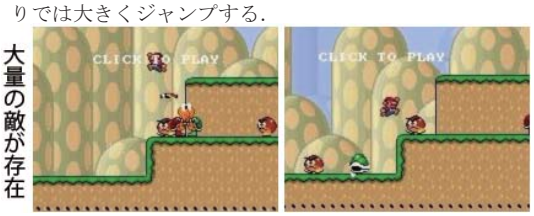

図 4 導入無し（左）と導入有り（右）での比較 (大量の敵が存在)
導入無しではためらわずに攻略しているのに対し，導入有りでは安全になるまで待つ．

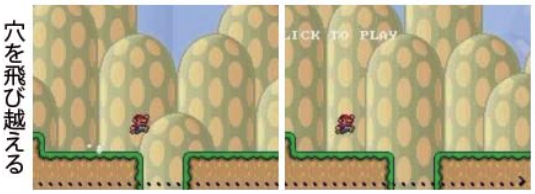

図 5 導入無し（左）と導入有り（右）での比較 (穴を飛び越える)
導入無しでは穴のギリギリからジャンプしているのに対し，導入有りでは余裕を持ってジャンプする．

級者であるかのような，たどたどしい振る舞いであった．この結果から，訓練と挑戦のバランシングにより，人間の熟達過程をシミュレートできる可能性が示唆された．

5.3 様々なゲームジャンルへの適用に向けての考察

本章の計算機実験の結果から，強化学習，経路探索の枠組みに『生物の基本原則』を導入することで，『人間にとって自然』と映る「敵を前にしてためらう」や「余裕を持ってジャンプする」といった，情緒的と解される振る舞いが獲得された．

ビデオゲームエージェントにおける「人間らしさ」は，ゲームジャンルごと，ゲームタイトルごとに違った要素が必要となる．それゆえ，ビデオゲームに人間らしいエージェントを組み込む場合，本来はゲームタイトルに合った人間らしさの解析が重要であった．しかし，本研究では，『生物の基本原則』のみを組み込んだ学習フレームワークにより，「人間プレイヤがゲームをしている」かのような振る舞いを表出できることを示した．これは，開発者のヒューリスティックスや，ゲームタイトルごとの人間らしさの解析に頼ることなく，さらに学習手法やゲームジャンルを問わず，『人間らしい』COM の振る舞いを獲得できる可能性が示唆されたといえる．

6. おわりに

本稿では，ビデオゲームを対象とし，『生物の基本原則』を強化学習，ないし，経路探索に組み込むことで，**『人間にとって自然』と映る COM** の振る舞いを，自律的に獲得できるフレームワークを提案した．計算機実験では，『生物の基本原則』を強化学習，ないし，経路探索に導入したエージェントにおいて，ためらいや恐怖，余裕といったエモーションを想起させるような振る舞いが獲得された．また，初級者っぽい，中級者っぽいといった，ゲームの熟達過程を思わせる振る舞いも見られた．『生物の基本原則』という，開発者のヒューリスティックスに依らない要素の導入により，学習手法に限定されることなく，『人間にとって自然』と映る COM の振る舞いを獲得できる可能性が示唆された．

今後の課題として，人間的と解される振る舞いの要因を調査するため，人間プレイヤのプレイログ集積による熟達過程の記録と分析や，人間プレイヤの振る舞いと獲得された COM の振る舞いとの比較を実施する必要がある．また，本研究の学習フレームワークが他のゲームジャンルにも適用可能かどうかも検証していく．

参考文献

[1] Togelius, J., Karakovskiy, S. and Baumgarten, R.: The 2009 Mario AI Competition, *Evolutionary Computation (CEC) 2010 IEEE*, pp. 1–8 (2010).
[2] 保木邦仁：局面評価の学習を目指した探索結果の最適制御，*GPW2006*, pp. 78–83 (2006).
[3] Fujita, H. and Ishii, S.: Model-based reinforcement learning for partially observable games with sampling-based state estimation, *Neural Computation*, Vol. 19, pp. 3051–3087 (2007).
[4] Schrum, J., Karpov, I. V. and Miikkulainen, R.: Human-like Behavior via Neuroevolution of Combat Behavior and Replay of Human Traces, *2011 IEEE Conference CIG' 11*, pp. 329–336 (2011).
[5] Soni, B. and Hingston, P.: Bots Trained to Play Like a Human are More Fun, *2008 IEEE International Joint Conference on Neural Networks*, pp. 363–369 (2008).
[6] 池田心，Viennot, S.：モンテカルロ碁における多様な戦略の演出と形勢の制御〜接待碁 AI に向けて〜, *GPW2012*, pp. 47–54 (2012).
[7] 藤井叙人，片寄晴弘：戦略型トレーディングカードゲームのための戦略獲得手法，情処論，Vol. 50, No. 12, pp. 2796–2806 (2009).
[8] Persson, M.: Infinite Mario Bros. , http://www.mojang.com/notch/mario/.
[9] Togelius, J.: Mario AI Championship , http://www.marioai.org/.
[10] J.Togelius, S.Karakovskiy, J.Koutnik and J.Schmidhuber: Super Mario Evolution, *2009 IEEE Conference CIG'09*, pp. 156–161 (2009).
[11] 2K Australia: The 2K BotPrize , http://botprize.org/.