強化学習エージェントによる協調行動とコミュニケーションの創発

佐藤 尚[†,☆] 内部英治[†] 銅谷賢治[†]

コミュニケーションの原型は，個体が環境や他の個体との相互作用において，報酬の獲得や適応度の向上に寄与する形で発現したと考えられる．本研究では，報酬最大化を目的とする強化学習エージェントが，余剰な行動と感覚の自由度をコミュニケーショのために使うことを学習できるための条件を，2 個体が互いに相手の縄張りに入ると報酬を得るが衝突すると罰を受けるというゲームにより検証した．このゲームでは，コミュニケーションと協調行動のそれぞれが必須ではないが，発光行動を使えるエージェント間では，互いにその光を信号として利用することで衝突を避け，報酬を獲得し合う協調行動の創発が観察された．信号の表現の仕方には多様性が見られ，また作業記憶を持つエージェント間では，信号を送る側とそれに従う側という役割分化も見られた．これは，コミュニケーションと協調行動が必須ではない状況において，意味と信号の任意の対応付けによるコミュニケーションが，コミュニケーションの達成そのものを目的としなくても一般的な行動学習の枠組みにより創発しうることを示す初めての知見である．

Emergence of Communication and Cooperative Behavior by Reinforcement Learning Agents

Takashi Sato,[†,☆] Eiji Uchibe[†] and Kenji Doya[†]

The prototype of symbolic communication would have emerged to help individuals to earn rewards and to improve fitness by using their excess degrees of freedom in action and perception. In this paper, we investigate whether and how reinforcement learning agents who aim at maximizing rewards can learn to use their redundant actions for communication in a simple game where the two agents learn to earn rewards by intruding into the other's territory on a linear track. In this task, although both cooperation and communication are not imperative, we found that the agents with lights and light sensors are able to achieve cooperative behaviors by avoiding collisions using visual communication in the middle of the track. Further analysis reveals a variety in the mapping of messages to signals. In some cases, the differentiation of roles into a sender and a receiver was observed. This is to our knowledge the first demonstration of emergence of communication by arbitrary meaning-coding mapping without an explicit objective of communication itself in a situation in which both cooperation and communication are not indispensable.

1. 序 論

我々人間を含む様々な動物のコミュニケーションの根源的な機能は，個体間の相互作用を可能に，あるいは組織化することにより，報酬の獲得や適応度の向上に寄与することであると考えられる．このようなコミュニケーションの原初的なものは，いつ，なぜ，どのように，そしてどのような行動をする個体の間で創発したのだろうか．たとえば人類の二足歩行や道具使用の起源は化石などの証拠を基に議論されるが，原初的コミュニケーションに用いられたであろう信号や言葉には，化石のような物理的証拠が存在しない．しかしながら，このような証拠の入手や実験室的再現が困難な問題に対して，理解したい対象の数理モデルを「作り」，それをコンピュータシミュレーションやロボット実験により「動かす」ことを通してその理解を試みる「構成論的手法[7),10),11)]」の有用性が，計算機技術の発展にともない広く認知されるようになった．コミュニケーションの創発を含む言語の起源と進化の問題に関しては，「進化言語学[8)]」の立場からの研究が近年興隆している．

コミュニケーション創発問題を扱う研究は，その多

† 独立行政法人沖縄科学技術研究基盤整備機構大学院大学先行研究プロジェクト
Initial Research Project (IRP), Okinawa Institute of Science and Technology (OIST) Promotion Corporation
☆ 現在，独立行政法人国立高等専門学校機構沖縄工業高等専門学校メディア情報工学科
Presently with Media Information Engineering, Okinawa National College of Technology (ONCT)

くが，いかにして形式的な記号システムに意味を内在させるかという「記号接地問題[6]」をコミュニケーション創発というテーマに置き換えて扱っている．またそれらは，アプローチの違いによって 2 つに大別される．1 つは，進化論的計算手法を採用する研究である[2),13),21)]．もう 1 つのアプローチは，コミュニケーション創発をエージェントの学習機能によって実現しようと試みるものである[9),12),14),16),17),19),☆]．

しかし，これらの研究では，信号と意味の対応付けを行いやすくするための作り込みがあらかじめ施されていたり[2),19)]，話し手や聞き手などの役割や情報伝達の手順などを含むコミュニケーション・プロトコルをあらかじめエージェントに付与していたり[2),9),12),14),17),19),21)]，あるいはコミュニケーションの成功自体を目的関数とする，またはそれを主目的とするタスクを課していたり[9),12),17)]，他にコミュニケーションの成功を主目的としていないタスクを課していたとしても，タスクの解が協調と非協調の 2 種類しか用意されておらず，前者が生じたときには利用されたコミュニケーションを強化する正の報酬（あるいは非協調が生じたときよりはるかに小さい負の報酬）が与えられるため，協調とコミュニケーションのそれぞれが共創発するパターンが現れやすくなっている[2),13),14),16),19),21)]．このため，これらの先行研究では，明示的なコミュニケーションが存在していなかった世界において，コミュニケーションがどのように創発し，どのようなコミュニケーション形態が生じうるのかという問いに答えることは難しい．また，コミュニケーションの実現に必要とされるすべての知識が生得的に備えられているとは考えにくく，そのような知識の多くは個体が世界や他個体との相互作用を通じて，様々なことを試行錯誤的に学習していくことによって漸次的に獲得されるものであると考えられる．

本研究では，コミュニケーションの成立に必要な知識やメカニズムがあらかじめ用意されていない状況下で，試行錯誤的学習機能を有する個体が環境や他個体との相互作用を通じて，意味を持たない余剰行動に特定の意味を割り当てることができるか，またそれによって他個体とのコミュニケーションが可能となるか，さらに，あらかじめ話し手と聞き手などの役割を付与しなくても自発的に役割分化したり，あるいは話者役の適切な交替（ターンテイキング）を行えるようになったりするかどうか検証する．なお，本研究で対象とするのは，Bickerton[1)] が提唱している「原型言語（Proto-language）[☆☆]」によるコミュニケーションよりもさらに単純な「原型会話（Proto-communication）」である．ここでは，原型会話を幼児の発達初期で見られるような「1 語（文/音）によるコミュニケーション」と定義する．

本稿では，まず 2 章においてコミュニケーション創発の可能性とその条件を検証するためのゲーム論的枠組みを提案し，さらにゲームのプレーヤとして用いる強化学習エージェントを説明する．3 章では，提案するゲームを用いたシミュレーション実験の結果を示す．4 章では改めて先行研究と本研究との違いについて，様々な視点から議論し，さらに得られた結果について考察する．そして最後に，5 章で本研究の結論を述べる．

2. モ デ ル

2.1 進入ゲーム

動物達の社会生活では，縄張り争い，採餌活動，そしてパートナー探しなど，実に様々な行動が見られ，その中では他個体との数多くの種類のコミュニケーションが観察されている．とりわけ，威嚇や服従を表す様々なディスプレイの示し合いは，彼らが深く傷ついたり死に至ったりするような激しい闘いを避けるうえで非常に重要な役割を果たすコミュニケーションの一種であると考えられる[5)]．

本研究では，ある個体たちが縄張り争いをしている状況を想定する．個体たちはそれぞれ利己的に他個体の縄張りへの侵入を試みようとするが，他個体との闘争にはコストがかかるため，なるべくなら闘争は避けるべきである．ここでは，上記想定状況を簡略化した「侵入ゲーム（Intrusion Game; IG）」を提案する．このゲームでは，4 つのスロットを持つ 1 次元トラック上で 2 個体のプレーヤが前後に移動することができる．中央から西側の 2 スロットは西側のプレーヤの縄張り，東側の 2 スロットは東側のプレーヤの縄張りとして割り当てられる．トラックの両端には壁があり，各プレーヤはその壁を乗り越えることができず，またプレーヤ同士は互いに擦り抜けることができない．4 スロットの空間上でとりうるプレーヤの位置パターン

☆ 天正ら[19)] は，進化的手法も利用している．しかし，進化の初期世代においてすでにエージェントが適切に行動選択できるようになる様子が示されていることから，強化学習が主に重要な役割を担っていると考え，ここでは天正らの研究を学習エージェントによるコミュニケーション創発の研究の 1 つとして紹介した．

☆☆ 文法構造を持たず，語順に一貫性がない 2，3 語程度の言葉．（1〜2 歳以降の）幼児言語，ピジン，人間によって言語教育を受けた類人猿の言語はすべて原型言語であると Bickerton は主張している[1)]．

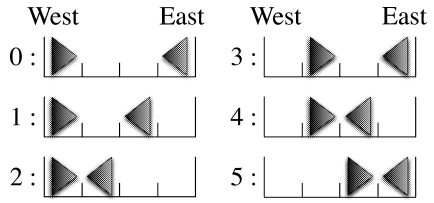

図 1 侵入ゲームにおいてプレーヤがとりうる 6 種類の位置パターン
Fig. 1 Six possible position patterns (denoted by 0 to 5) of the players in the intrusion game.

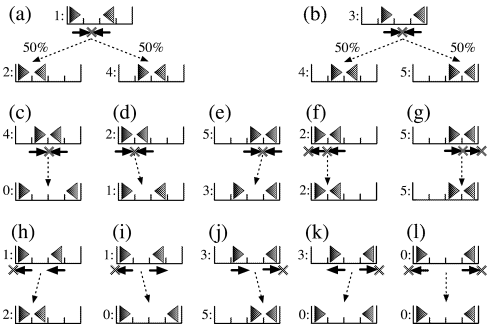

図 2 衝突時における 12 種類の状態遷移規則
Fig. 2 Twelve state transition rules of position patterns in case of collisions.

を**図 1** に示す．各位置パターンに付けられている数字は，各プレーヤによってそれぞれの位置を認識する際に用いられる位置パターン識別番号である．

IG におけるプレーヤの目的は，衝突せずに相手の縄張りに入ること（によって報酬を獲得すること）である．すなわち，図 1 における位置パターン 1 ないし 4 から 2 へ遷移したときに東側のプレーヤが，同様に位置パターン 3 ないし 4 から 5 へ遷移したときに西側のプレーヤが報酬を獲得する．一方，プレーヤが壁，他のプレーヤ，またはその両方と衝突した場合にはプレーヤに罰が与えられる．衝突には**図 2** に示すように 12 種類のケースがある．各図の上段が衝突前の状態を表し，その状態から実線矢印の方向へ各プレーヤが移動して衝突が起きた後（図中の×印は衝突が起きたところ），位置パターンがどのように遷移するかを点線矢印の先に描かれている下段が表している．図 2 において，(a)〜(e) は両プレーヤが前進したことにより生じる衝突，(f) および (g) は片方のプレーヤが相手に向かって前進し，壁際のプレーヤが壁に向かって後退することによ生じる 2 重衝突，(h)〜(l) は壁際のプレーヤが壁に向かって後退することにより生じる衝突である．なお，空きスロットに両者が進入しようとする (a) と (b) では確率的に，その他の (c)〜(l) では決定論的に状態が遷移する．

IG において最大の問題となるのは，位置パターン 4 における競合をどのように解消するかということである．プレーヤが即時的な報酬をめざして行動するとすれば，西側のプレーヤは 4 から 5 へ，そして東側のプレーヤは 4 から 2 へ，ともに前進することを選ぶはずである．しかし，位置パターン 4 において両者がともに前進を選ぶならば，2 個体は衝突し，それぞれに罰が与えられる．

もしも，位置パターン 2 ないし 5 から 4 へ遷移したときに，「1 つ前のステップで前に進んだならば，本ステップでも前に進む」と「1 つ前のステップで後ろに進んだならば，本ステップでも後ろに進む」という 2 種類のルールをプレーヤ双方が適切に利用できるならば，IG における位置パターン 4 での競合を解消することができ，各プレーヤは交互に報酬を獲得できるはずである．たとえば，プレーヤが自分の過去の行動を記憶できるならば，原理的には上記 2 種類のルールを獲得することが可能である．しかし，実際に位置パターン 4 での競合を解消するためには，各プレーヤが両方のルールを同時期に獲得していなければならない．さらに，位置パターン 0 から 4 へ遷移した後，上記ルールを適用することが逆に衝突を招くことになるため，位置パターン 4 においてどのような行動をとるべきかを規定するルールを獲得するとともに，位置パターン 2 および 5 の両方から 4 へ遷移できるようにするためのルールも獲得しておく必要がある．

2.2 強化学習エージェント

本研究では，完全なコミュニケーション・システムが生得的に各個体に備わっていると考えるのではなく，それらが環境や他個体との様々な相互作用を繰り返すことで得られた報酬を基に学習していくことによって，初めてコミュニケーションを成立させられるようになると考える．また，コミュニケーションに用いられる信号や言葉，そしてそれらの教示者が存在していなかった世界では，個体たちは各々が行うことのできる様々な行動を試しながら学習し，その過程を通して，試行錯誤的に原始的なコミュニケーション手段を獲得したのではないかと推測される．

そこで我々は，環境や他個体との相互作用によって与えられる報酬や罰に基づき，エージェントが自律的に行動を獲得するのに有効な試行錯誤的学習手法の 1 つである「強化学習[18)]」を採用する．我々の提案する IG では離散時間，離散空間，離散行動を仮定しているため，そのような離散問題に適した「Q 学習[18)]」を用いることにする．なお，Q 値の更新は次式により行う．

表 1 4 種類の強化学習エージェント
Table 1 Four types of reinforcement learning agents.

	Actions	States
(a) N (null) type	0: Go Forward 1: Go Backward	0~5: Position Pattern
(b) M (memory) type	0: Go Forward 1: Go Backward	0~5: Position Pattern 0 or 1: Past Action
(c) L (light) type	0: Go Forward with Light Off 1: Go Backward with Light Off 2: Go Forward with Light On 3: Go Backward with Light On	0~5: Position Pattern 0 or 1: Opponent's Light
(d) LM (light-memory) type	0: Go Forward with Light Off 1: Go Backward with Light Off 2: Go Forward with Light On 3: Go Backward with Light On	0~5: Position Pattern 0~3: Past Action 0 or 1: Opponent's Light

$$Q(s_t, a_t) := Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

ここで，r_{t+1} はある時間ステップ t において，エージェントが状態 s_t で行動 a_t を行ったときに得られる報酬，α は学習率（$0 < \alpha \leq 1$），γ は割引率（$0 \leq \gamma \leq 1$）である．エージェントの行動選択手法としては「ϵ グリーディ手法」を採用する．この手法は，確率 $(1-\epsilon)$ で最大 Q 値を持つ行動を選択するが，確率 ϵ でランダムに行動を選択するというものである．

本研究では，エージェントのセンサや行動の違いによって，それらの振舞いがどのように変わるのかを調べるために，**表 1** に示す 4 種類の強化学習エージェントを IG のプレーヤとして用いる．前節で示したように，IG のプレーヤの目的は，衝突せずに相手の縄張りに入ることである．また，IG における位置パターン 4 での競合は，エージェントが記憶を持つことにより，原理的には解消可能である．したがって，最低限必要となる基本行動は前後への移動行動のみであり，その他の行動はすべて余剰行動であると見なせる．

ここで採用するエージェントは，現在の両エージェントの位置パターンを観測可能であり，光源と光センサを持つか否か，自分の過去の行動の記憶を持つか否かによって区別される．具体的には，N（null）型のエージェントは，位置パターンセンサのみを持ち，M（memory）型のエージェントはさらに自身が 1 ステップ過去に行った行動の履歴を状態として持つ．L（light）型エージェントは光源と光センサを持ち，相手の発光の有無を観測し，そして相手に対して光を発する行動を持つ．さらに LM（light-memory）型エージェントは，光源，光センサと自分の 1 ステップ過去の行動の記憶を持つ．なお，L および LM 型エージェントは，どのスロットからでも相手の光源の状態を観測できるものとする．

3. シミュレーション実験と結果

シミュレーション実験の各種設定を以下に記す．エージェントへの（プラスの）報酬は「+1」，罰（マイナスの報酬）は壁のみ，または他個体のみと衝突した場合「−1」，そして両方と衝突した場合「−2」とする．エージェントの強化学習に関する各パラメータは，3.4 節以外では，ランダム行動選択確率 ϵ を「0.01」，学習率 α を「0.01」，そして割引率 γ を「0.9」とし，これらは変化しないものとする．また，エージェントの各 Q 値の初期値は「0.0」にセットする．3.4 節では，様々な値のパラメータセットを用いて，エージェントの振舞いに対する各パラメータの影響について調べた実験の結果を示す．

IG は，1) エージェントのセンサによる現在状態の知覚，2) 知覚結果に基づき行動を選択，3) 選択した行動を実行，4) 報酬獲得，5) Q 学習（次状態の知覚も含む）という 5 つをまとめて 1 ステップ，そして 10,000 ステップを 1 エピソードとし，「10 エピソード」まで繰り返される．エージェントの位置は，各エピソードの最初に図 1 に示される 6 種類のうちの 1 つがランダムに設定される．4 種類のエージェントの設定に対して各々異なる初期位置（6 種類）および乱数で 10 回のシミュレーションを行った．

3.1 協調行動の創発

まずはじめに，4 種類の強化学習エージェントごとの最終エピソードにおける合計報酬の多かったエージェント（勝者）と少なかったエージェント（敗者）の合計報酬のアンサンブル平均を**図 3** に示す．

N および M 型エージェントでは，ほぼすべての試行において，2 個体のうちの片方のエージェントが理論的最大値に近い合計報酬を獲得し，1 人勝ちしている．なお，エージェントが 1 人勝ちする場合，相手の縄張り前で前後に移動することによって，2 ステップに 1 回報酬を得ることができる．前述のとおり，1 エピソードが 10,000 ステップであるから，その場合での合計報酬の理論的最大値は 5,000 である．

一方，L および LM 型エージェントでは，勝者と敗者の合計報酬の平均値の差が縮まっている．実際，単純に勝者と敗者の差が縮まっているだけでなく，その差がほとんどなくなる例も見られた．これは 2 個体のエージェントが互いに報酬を獲得し合う協調行動を行っていることを示唆する．

次に，L および LM 型エージェントの位置パターンの変化を解析した結果を**図 4** に示す．ここでは，4 種

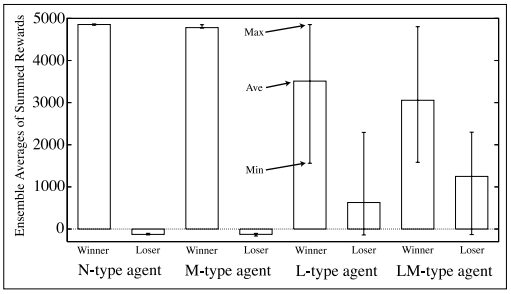

図 3 4 種類の強化学習エージェントごとの最終エピソードにおける合計報酬のアンサンブル平均
Fig. 3 Ensemble averages of summed rewards obtained at final episode for four types of agents.

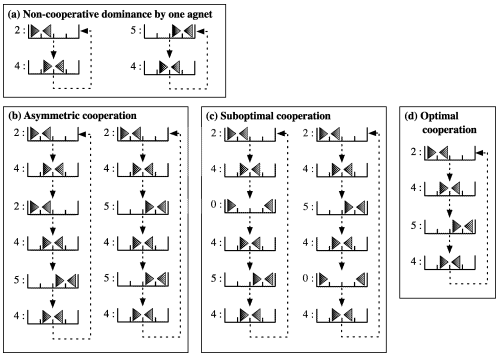

図 4 4 種類の典型的な行動パターンの例
Fig. 4 Four examples of typical behavioral patterns.

類の典型的な行動パターンを例示する．

図 4 (a) は，1 個体による非協調的支配の行動パターンを示す．この行動パターンでは，1 個体のみが 2 ステップごとにプラスの報酬を獲得できるが，他のエージェントは同じ報酬を得ることができない．敗者エージェントは，この行動パターンを変えない限り，マイナスの報酬を与えられることはない．逆に変えた場合には，勝者エージェントも行動パターンを変えない限りマイナスの報酬を得る可能性の方が高い．

図 4 (b) は，非対称的協調行動を示す．この行動パターンでは，両方のエージェントがプラスの報酬を獲得することができる．しかし，1 個体は 6 ステップ中でプラスの報酬を 2 回獲得できるが，もう一方の個体は同じステップ中でも 1 回しか報酬を得ることができない．なお，この行動パターンは，LM 型エージェントのみで見られる．

図 4 の (c) と (d) は，それぞれ準最適協調行動と最適協調行動を示す．最適協調行動とは，最少のステップ数で 2 個体のエージェントがともに最多の報酬獲得回数となる行動パターンのことであり，4 ステップおきに 1 回報酬を獲得できる．準最適協調行動では 6 ステップおきに 1 回報酬を獲得できる．

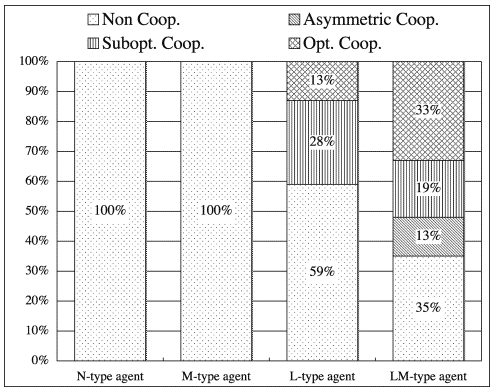

図 5 4 種類の強化学習エージェントごとの各行動パターンの発生頻度
Fig. 5 Occurrence frequency of four typical behavioral patterns for four types of agents.

図 5 は 4 種類の強化学習エージェントごとの各行動パターンの発生頻度を調べた結果である．図 3 で示唆されたように，N および M 型エージェントでは，いずれも図 4 (a) に示される非協調的支配行動パターンに 100%収束することが確かめられた．一方，L および LM 型エージェントでは，様々な協調行動を獲得することが分かった．また，LM 型エージェントの方が，L 型エージェントよりも最適協調行動を創発しやすいことが分かった．

3.2 コミュニケーションの創発

本節では，IG において見られた強化学習エージェント間でのコミュニケーション創発に焦点を当てる．IG では，2 個体のエージェントが位置パターン 4 において交互に前進行動をとるために，どのような情報を用いればよいかということが最大の問題である．各エージェントを比べた結果，N および M 型エージェントは位置パターン 4 における競合を解決することができず，1 個体による非協調的支配に 100%収束するのに対し，L および LM 型エージェントは互いに光信号を発信し合うことによって位置パターン 4 における競合を解決することができ，2 個それぞれが報酬を獲得し合う協調行動を実現することができた．

ここでは，まずはじめにエージェントが獲得したコミュニケーション形態の典型例を示す．**図 6** (a) は，対称的信号伝達の例である．図中 2 および 4 段目は位置パターン 4 に到達した時点の各エージェントの光源の状態を表しており，それぞれ交互に光を発している様子が示されている．この場合，各エージェントの光信号は，位置パターン 4 から次に移行すべき位置パターンを表すことができる．一方，図 6 (b) は，非対称的

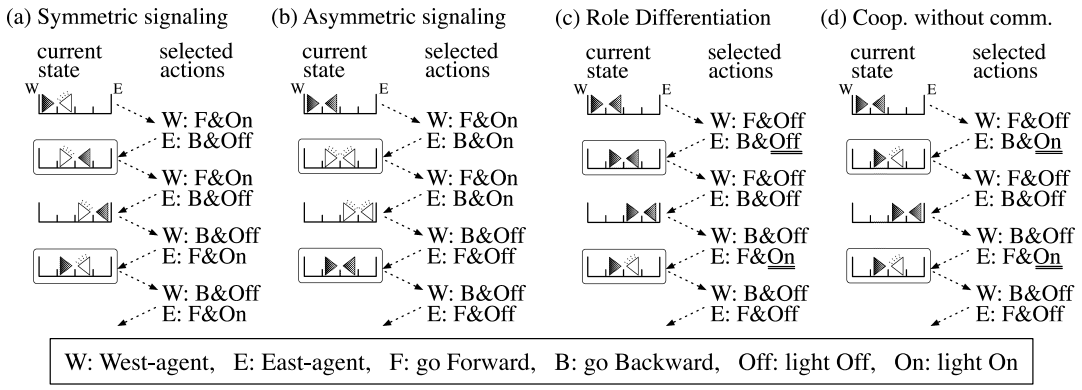

図 6 創発したコミュニケーションの典型例
Fig. 6 Typical examples of emerged communication.

信号伝達の例である．こちらは，2 個体がともに光を発する，または発しないことによって位置パターン 4 における競合を解消することで，それぞれの利益を損なわない行動をとることが可能となっている．

さらに，エージェントの行動を詳細に調べてみると，図 6 (c) および (d) に示されるようなコミュニケーションの創発例が LM 型エージェントにおいて見られた．図 6 (c) は光信号の送信者と受信者への役割分化が実現されていることを示している．すなわち，西側のエージェントは東側のエージェントが発した光信号を受信し，その情報に基づいて位置パターン 4 における競合を解決しているが，自身では光信号を発していない．一方，東側のエージェントは西側のエージェントから位置パターン 4 における競合を解決するためのヒントを与えられていないため，代わりに自身の過去の行動の履歴情報を用いてその問題を解決している．

図 6 (d) は 2 個体間での明示的なコミュニケーションに頼ることなしに実現された最適協調行動の例である．図から分かるように，位置パターン 2 または 5 から 4 へ移行する際，東側のエージェントはつねに光信号を発している．すなわち，東側のエージェントが発した光信号が位置パターン 4 における競合を解決するためには役に立たないことを意味する．また，西側のエージェントも東側のエージェントとは逆につねに光信号を発していないため，東側のエージェントは西側のエージェントの光信号を頼りにすることができない．この例では，双方が自身の過去の行動の履歴情報のみを利用して位置パターン 4 における競合を解決している．

ここで示した 4 つの典型例では，エージェントが確率 ϵ でランダムな行動をとることによって，最適協調行動パターンが崩れたとしても，ほとんどの場合，長くとも数ステップ後にはそのパターンが再生される．さらに興味深いことに，図 6 (d) のパターンが同様のランダム行動によって崩れたときには，1 度光信号の送信者と受信者という役割分化状態となり，片方向コミュニケーションによる最適協調行動を行った後に，コミュニケーションに頼らない最適協調行動の周期パターンが復活する例も見られた．これらのことから，強化学習によって，エージェントが経験可能なすべての状態・状況に対して安定した協調行動を行うことができる可能性が示唆される．

3.3 エージェントの学習プロセスと最適協調行動の安定性

続いて，図 4 に示される 4 種類の行動パターンが，どのような行動パターンを経て収束したのか，また収束するまでにエージェントがどのような行動パターンを形成可能であるのかを調べた．まず，学習中のエージェントの Q 値を 1,000 ステップごとに保存しておく．次に，各 Q 値の組を初期値としてセットしたエージェントで，ランダム行動選択確率を 0，学習係数を 0 として，すべての初期状態から改めて IG を始め，収束した行動パターンを調べた．なお，L 型エージェントがとりうる状態数は，24（6 (初期位置の種類) × 4 (2 個体の発光状態の組合せの数)）である．LM 型エージェントであれば，とりうる状態数は 384（6 (初期位置の種類) × 16 (2 個体の過去の行動の組合せの数) × 4 (2 個体の発光状態の組合せの数)）である．ここでは LM 型エージェントにおいて見られた典型例を**図 7** に示す☆．なお，グラフ中の数字列はエージェントの位置パターンの変化を表す．たとえば，「45」は位置パ

☆ グラフが 50,000 ステップまでしか描かれていない理由は，行動パターンが収束するまでにほとんど 50,000 ステップも必要としなかったからである．

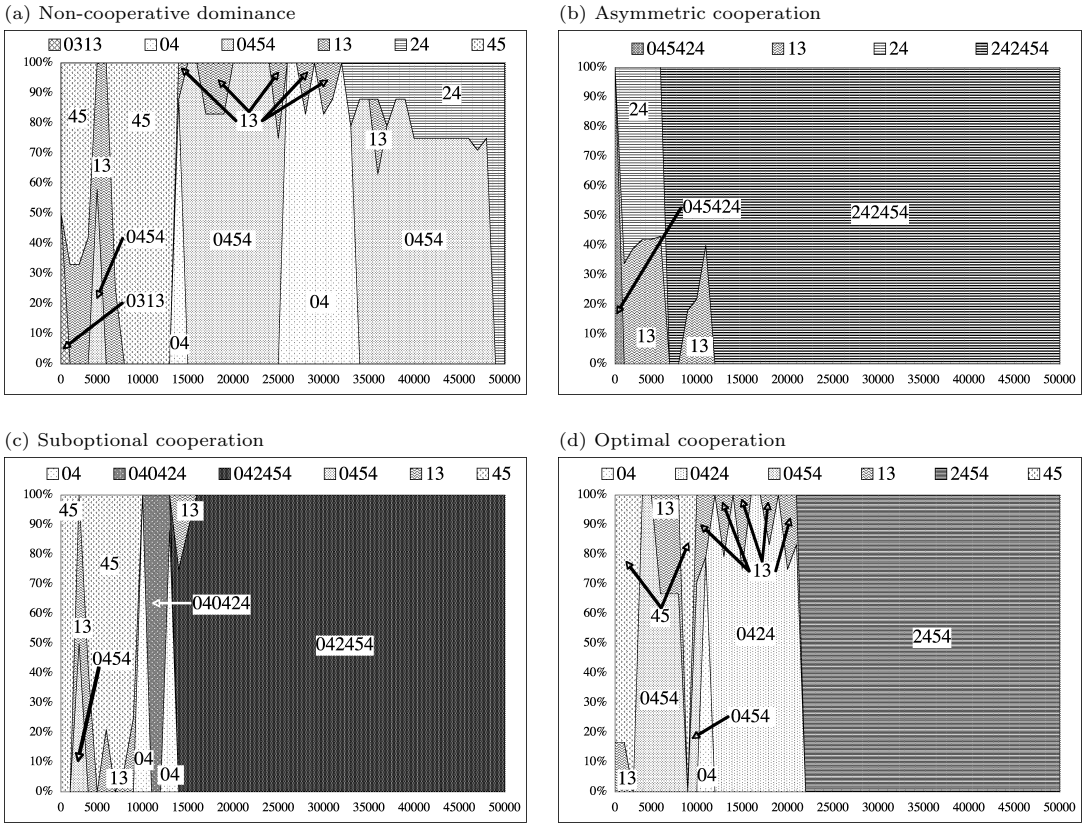

図 7 4 種類の行動パターンに収束するまでの強化学習エージェントが示すことのできる行動パターンの変化
Fig. 7 Transitions of behavioral patterns that the agents can show until converging with one of four typical behavioral patterns.

ターン 4 と 5 の繰返し，すなわち，図 4 (a) に示されるような 1 個体による非協調的支配の行動パターンを意味する．

図 7 (a) は，1 個体による非協調的支配のパターンに収束するまでの変遷を示す．この例では，「45」や「0454」という西側のエージェントにとって有利な行動パターンが支配的となっている．しかし，49,000 ステップより逆転し，どの状態から IG を始めても，東側のエージェントが 1 人勝ちするパターンに収束するようになっている．

図 7 のそのほかの図は，それぞれ (b) が非対称的協調行動，(c) が準最適協調行動，そして (d) が最適協調行動に収束するまでの変遷の例を示す．これらと図 7 (a) の違いは，早い段階で位置パターン 4 から 2 へのルートと，4 から 5 へのルートを別々に，あるいは同時に選べるようになっていることである．すなわち，これらの行動パターンを示すエージェントは，1 つ前の状態（自身の行動，相手の光信号，またはそれら両方）に従って，位置パターン 4 における競合を適切に解消できるようになっているのである．また，図から分かるように，エージェントはとりうるすべての状態から IG を始めたとしても，最終的に最適協調行動パターンを示すようになっている．これによって，確率 ϵ でランダムな行動をしたがために，最適協調行動パターンが一時的に崩れたとしても，エージェントがすぐにそのパターンを回復できる方策を獲得できることが確かめられた．

さらに，エージェントのコミュニケーション形態の違いが最適協調行動の安定性にどのような影響を及ぼすのかを調べた．最適協調行動の安定性は，エージェントの行動が最適協調行動に収束するまでに必要としたステップ数で評価した．このステップ数が少なければ少ないほど，確率 ϵ で生じるランダムな行動によって乱れにくい安定な最適協調行動であるといえる．

ここでは，最終エピソードにおいて最適協調行動を示すことができたすべてのエージェント，すなわち，対称的および非対称的コミュニケーションを行い最適協調行動を示すことができた L 型エージェント，同様

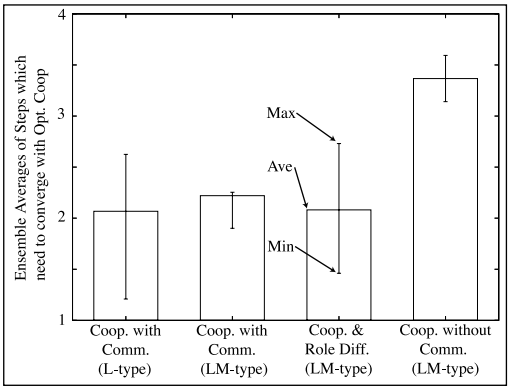

図 8 最終エピソードにおいて最適協調行動に収束するまでに必要とするステップ数のアンサンブル平均
Fig. 8 Ensemble averages of steps which need to converge with an optimal cooperation at the final episode.

のコミュニケーションを行い最適協調行動を示すことができた LM 型エージェント，役割分化状態で最適協調行動を示すことができた LM 型エージェント，そしてコミュニケーションに頼らずに最適協調行動を示すことができた LM 型エージェントを対象とした．

具体的な計算手順は図 7 のときとほぼ同様であるが，今回は最終エピソードでの上記エージェントの Q 値のみを用いて，エージェントのとりうるすべての状態から IG を始めて，最適協調行動に収束するまでに要したステップ数のアンサンブル平均を調べた．その結果を**図 8** に示す．一見して分かることは，コミュニケーションに頼らずに最適協調行動を示すことができたエージェントの行動パターンが収束するまでに必要としたステップ数が，他のケースに比べて多いということである．換言するならば，コミュニケーションに頼らない最適協調行動は，他の最適協調行動よりも不安定であるといえる．これは，コミュニケーションに頼らない最適協調行動を行うエージェントの方策において，状態と行動との間の 1 対 1 の関係があまり重複なく形成されていることが原因である．つまり，いくつかの状態から 1 つの行動というような多対 1 の関係が他のエージェントに比べて少なく，冗長性が低いために，余計な位置パターンのルートを通ることになっているのである．

3.4 最適協調行動の創発に対する各パラメータの影響

最後に，強化学習に関する各パラメータが，エージェントの振舞いに対してどのような影響を及ぼすのかを調べた．その実験結果を**図 9** に示す．ここでは，各パラメータの影響を調べるために，エージェントに 100 エピソードまで IG をプレイさせることを，1 組のパラメータセットあたり，初期位置パターンおよび乱数シードを変えて 60 試行行い，最終エピソードにおいて安定した最適協調行動の創発が見られたゲームをカウントし，その合計値を指標とした（よって，最高値は 60 である）．図 9 の各行は上段から順に，M 型，L 型，そして LM 型エージェントを用いて実験した結果を示している．また，各列は，左から学習率 α が 0.01，0.05，そして 0.1 のときの結果である．N 型エージェントではすべてのパラメータセットで最適協調行動が見られなかったため結果を省略した．

まず，M 型エージェントの結果に注目しよう．先の実験において，図 3 および 5 に示されるように，M 型エージェントは N 型エージェントと同様，1 個体による非協調的支配の行動パターンしか示すことができなかった．ところが，割引率 γ を非常に大きい値（たとえば，0.999）に設定した場合，すなわち，即時報酬よりも将来の報酬に関心を持つようにした場合，学習率 α の値が小さければ，興味深いことに，M 型エージェントによる IG でも最適協調行動が示された．しかしながら，M 型エージェントは，1 つ前の自身の行動を記憶できるだけであるため，位置パターン 4 における競合を解消することができる行動ルールの組合せを 1 種類（「位置パターン 4 に前進して遷移したならば次も前進」と「位置パターン 4 に後退して遷移したならば次も後退」という組合せ）しか形成できない．したがって，確率 ϵ でその行動ルールとは関係なくランダムに行動した場合には，最適協調行動のパターンが簡単に崩れてしまうため，安定して交互に報酬を獲得し続けることはできないのである．また，学習率 α の値を大きくした場合，学習 1 回あたりの Q 値の変化量が大きくなるため，最適協調行動を維持できる代わりの行動ルールを形成できない M 型エージェントでは不安定化する．

次に，L 型エージェントの結果を見てみると，M 型エージェントが IG をプレイしたときよりもさらに最適協調行動の発生頻度が高くなっていることが分かる．これは，L 型エージェントが位置パターン 4 での競合を解消できる行動ルールの組合せを M 型エージェントよりも多い 4 種類形成することができるからである（図 6 (a) と (b) のほかに移動行動が入れ替わった 2 種類）．共通していることは，割引率 γ の値が小さいときよりも大きいときの方が，そしてランダム行動選択確率 ϵ の値が大きいときよりも小さいときの方が最適協調行動を創発しやすいということである．ただし，学習率 α の値を大きくするにつれて最適協調行動の発生頻度は低くなり，α が 0.1 のときでは，ど

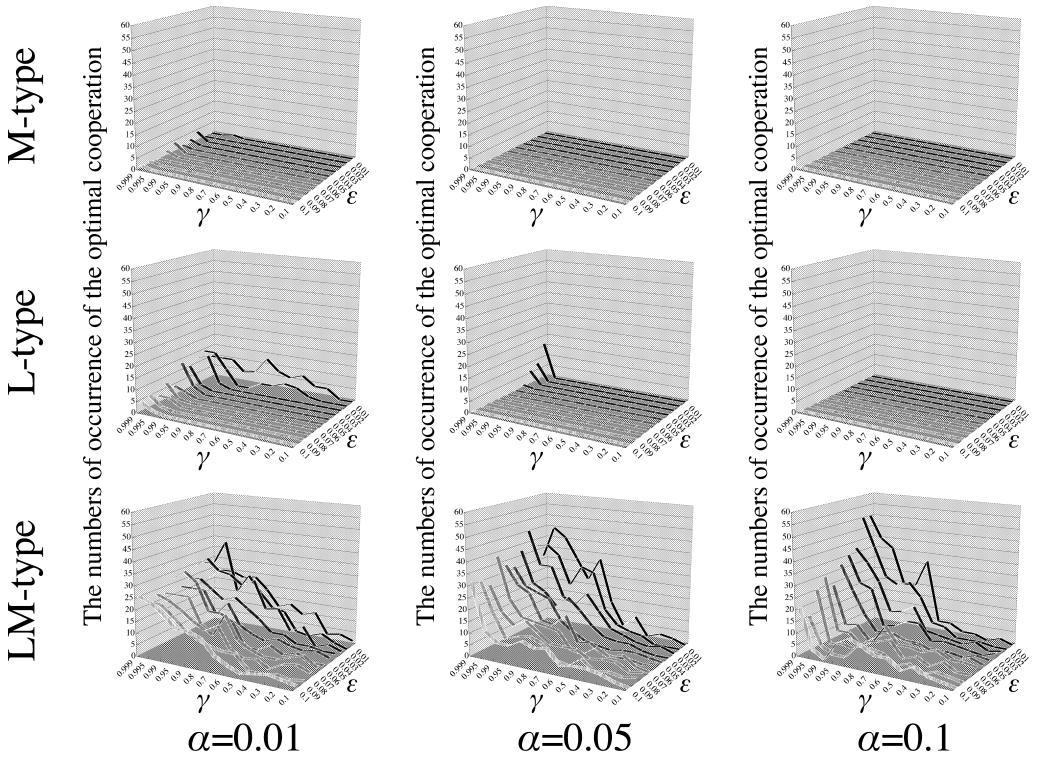

図 9 3 種類の強化学習エージェントごとの最適協調行動の創発に対する各パラメータの影響
Fig. 9 Effects of parameters on the optimal cooperation for three types of agents.

のような γ と ϵ の組合せにおいても最適協調行動が見られなかった．L 型エージェントは，4 種類の行動ルールの組合せのうちのいずれかによって最適協調行動を形成可能であるが，ランダム行動選択によって一対の行動ルールのどちらか一方でも壊れてしまった場合，図 6 (a) および (b) から分かるように，最適協調行動は維持できなくなる．α が大きい場合には，1 回のランダム行動選択によって行動ルールが壊れやすくなるため，最適協調行動の発生頻度が低下したと考えられる．

LM 型エージェントでは，割引率 γ の値が大きく，ランダム行動選択確率 ϵ の値が小さい場合には，最適協調行動の発生頻度が非常に高くなることが分かった．また，γ の値が十分大きい場合には，ϵ の値が大きくても L 型エージェントより最適協調行動の発生頻度を高く保てることが確かめられた．さらに興味深いことに，他のケースとは異なり，学習率 α の値が大きくなるにつれて，最適協調行動の発生頻度は低くならず，逆に高くなるということが分かった．これらは第 1 に，LM 型エージェントが形成可能な位置パターン 4 での競合を解消できる行動ルールの組合せが 16 種類（一部を図 6 に示している）と L 型エージェントよりも倍以上多く，適切な行動ルールの組合せをエージェント間で獲得できる確率が高いために見られ，そして第 2 に，ランダム行動選択によって一対の行動ルールのどちらか一方が壊れてしまったとしても，もう一方の行動ルールと対になって位置パターン 4 での競合を解消できる行動ルールがさらに 3 種類ある（つまり，1 種類の行動ルールに対して，4 種類の行動ルールを組み合わせることができ，そのようなセットが 4 組あるため合計 16 種類となる）ために，L 型エージェントよりも早く挽回できるということが理由であると考えられる．

これらの結果から示唆されることは，将来の報酬を十分考慮に入れられるようにすることが，互いに報酬を獲得できる協調行動の創発に重要な役割を果たしているということである．さらに，コミュニケーションに用いられる 1 つの信号または記号に対して，複数の意味や行動を割り当てられることと，信号または記号とその内容との結びつきにあらかじめ必然性がない（様々な組み合わせ方が可能）ということがコミュニケーションの創発には重要であると考えられる．

4. 議　　論

4.1 先行研究との相違点

本研究は，動物や人間の行動学習のモデルとしても支持されている強化学習のアルゴリズム[3),4)]を用いており，コミュニケーション創発をエージェントの学習機能によって実現しようと試みる研究に属する．我々と同様に学習エージェントを用いた林ら[9)]，小島ら[12)]，そして Steels ら[17)] 等の研究では，コミュニケーションの成功自体を目的関数としている，あるいはその成功を主目的とするタスクをエージェントに課しており，我々がエージェントに課している IG とは問題設定自体が大きく異なっている．

コミュニケーションの成功自体を主目的とせず，採餌や衝突回避などの行動の実現を課題とした研究[2),13),14),16),19),21)]でも，適切なコミュニケーションなしに最適な協調行動が実現不可能な課題を扱っている．しかも，協調する場合としない場合とでエージェントに与えられる報酬の値の大きさに偏りがあるため，報酬の最大化にはコミュニケーションと協調行動がともに必須な課題となっている．なお，Werner らの研究[21)]では，雄エージェントが雌エージェントのもとにたどり着き子を得るという最適行動に対して，数値的な報酬が与えられるわけではないが，それによって，似たようなコミュニケーション信号の出し方や解釈の仕方を持つエージェントが集団内に増えるため，コミュニケーションと最適解の双方がともに成立しやすくなるという点は他の先行研究と同じであると考えられる．

一方，我々が提案した IG では，図 6 (d) や割引率 γ に大きな値を設定した M 型エージェントによる IG でも不安定ながら最適解が得られることが示された図 9 から明らかなように，コミュニケーションに用いられる信号自体が最適協調行動を行ううえで役に立たなくなった場合，または完全にコミュニケーションを行えない場合でも最適協調行動が得られることが確かめられており，コミュニケーションと最適協調行動の共創発がおおむね必然となっている他の研究とは大きく異なっている．

そのうえ，コミュニケーションの成功自体を主目的としていない先行研究[2),13),14),16),19),21)]の課題では，協調するか否か，あるいは得をするか損をするかの 2 種類しか用意されておらず，前述のとおり，協調行動を行った場合には正の報酬が与えられたり適応度が高くなったりして，それを導いたコミュニケーションも同時に強化される形となっていることから，あらかじめ非協調行動に収束しにくい設定になっている．しかし，我々の IG とそのプレーヤとして採用した強化学習エージェントでは，様々な協調行動が見られただけでなく，図 4 (a) に示されているように，片方のエージェントのみが得をし，もう片方のエージェントは得しないけれども損にはならない（報酬 0）という行動パターンも見られていることから明らかなように，協調行動のみに収束しやすくなるような偏りが生じる課題設定にはなっていない．

さらに，話し手と聞き手という役割とそれらの間での情報伝達の手順などをあらかじめエージェントに付与し，能力的に異質なエージェントに設定した状態から実験を行っている，あるいは，各エージェントの初期値を同じ値に設定していないために，厳密には同質なエージェントとして設定しているとはいえない研究[2),9),12)~14),17),19),21)]とは異なり，本研究では，各エージェントの Q 値をゼロクリアした状態，すなわち，各エージェントが能力的にまったく同質な状態からすべてのシミュレーション実験を行っている．そのような設定にもかかわらず，エージェント同士が確率的な行動選択と状態遷移を通じた相互作用を繰り返しながら，各々が独立に行動学習していくことによって，ここで見られた役割分化が自発的に生じることが確かめられた．しかも，同時に信号を発することによって情報伝達するというコミュニケーション形態（2 個体それぞれが話し手であり，なおかつ聞き手でもある）が創発することも確認できた（図 6 (b)）．これらの結果は，話し手と聞き手の役割をあらかじめ固定している，あるいは話し手が聞き手に対して情報伝達した後にその役割を入れ替えるとあらかじめ決めている研究では見られていない新たな結果であると考えられる．

4.2 創発したコミュニケーションにおける信号の意味解釈

本稿で提案された IG では，まず，強化学習エージェントが他個体との相互作用を繰り返す中で，「光を発する/発しない」というそれ自体には何の意味も持たない行動の利用法を試行錯誤的に探索した結果，その光を効果的に用いて互いに衝突を回避できるようになり，さらに交替で報酬を獲得し合うという協調行動を発現させることができた．

対称的信号伝達法（図 6 (a)）を獲得したエージェントたちの振舞いを客観的にとらえるならば，エージェントが位置パターン 4 において次に自分が進もうとしている方向（行おうとしている行動）を光信号により相手に知らせているかのように見える（宣言的信号）．または，光を発しているエージェントがその光によっ

て相手に自分が進もうとしている方向（行おうとしている行動）とは逆の方向へ移動（逆の行動を選択）することを求めているようにも解釈可能である（命令的信号）．

非対称的信号伝達法（図 6 (b)）を獲得したエージェントたちの例では，位置パターン 4 において各個体の発光行動が同じ意味（たとえば，発光=前進）を持つのではなく，2 個体が同時に信号を発するまたは発しないこと自体が 1 つの意味を帯びていると解釈することが可能であろう．換言するならば，エージェント間でコンセンサスのとれた一意的な合図（たとえば，位置パターン 4 において 2 個体がともに発光した状態 = 2 個体がともに次のステップで東側へ移動）のようなものであると考えられる．

4.3 原型会話の創発要件

本研究で見られたコミュニケーション創発は，「信号が必ずしも受信者を何らかの成功へと導くわけではなく，またそれ自体が直接何かの役に立つわけではなかったとしても，平均すれば送信者が受信者の反応によって利益を得るような信号の伝達」という Halliday らの動物コミュニケーションの定義[5] に添うものであり，それが強化学習によって「利益を得る」行動，つまり，協調行動を実現する信号の伝達行動が繰り返し強化されることで発現することが示された．

Tomasello は，伝達の働きを持つ合図は，社会的な相互作用の反復を通して，2 者間で互いの行動を形成していくことによって作り出されると述べている[20]．我々のシミュレーション実験の結果はこの Tomasello の主張を支持するものである．

さらに Tomasello は，言語記号の習慣的な使い方を習得するためには，1) 他者が意図を持つ個体であると理解すること，2) 共同注意場面☆に参加すること，3) 共同注意場面で何かに注意を向けさせようとする誰かの伝達意図を理解すること，そして，4) 模倣学習のプロセスを通して，大人と役割を交替し，大人が自分に向かって使った記号を大人に向かって使うことが必要であると主張している[20]．これらの中でも彼は特に 3) の伝達意図理解の重要性を説いており，さらに他者の伝達意図を理解するためには，「相手は ‘私が X に対する注意を共有することを’ 意図している」ということを理解しなければならない，と述べている[20]．一方，本研究における強化学習エージェントたちは，1) 他のエージェントが意図を持つ個体であると理解していない，2) あくまで 2 者間で 2 項関係を形成しているのみであるので共同注意場面に参加していない，3) Tomasello のいう意味での相手の「伝達意図」を理解していない，しかし，4) 模倣学習機能を持っていないにもかかわらず，相手が使った信号と同じ意味を持つ信号を使うことはできるようになった．我々の想定している原型会話以上の複雑なコミュニケーションの創発には，Tomasello が重要視している「伝達意図理解」が必要であるのかもしれない．しかし，原型言語によるコミュニケーションよりもさらに単純な原型会話ならば，模倣学習機能を必要とせず，他者との相互作用と試行錯誤的学習による余剰行動の役割探索の繰り返しによって発現させられる可能性があることを本研究の結果は示している．

☆ 2 者が第 3 の何かに，または第 3 の何かに向けられている相手の注意に，ある程度の時間にわたって注意を向けるという社会的なやりとりのこと[20]．

5. 結　　論

本研究では，コミュニケーションが成立する以前の世界を想定し，信号に対する特定の意味とその教示者をあらかじめ用意しないという状態からコミュニケーション創発の可能性を議論するための侵入ゲームを提案した．このゲームのプレーヤとしては，強化学習を生得的機能とするエージェントを採用した．信号を送信できるエージェントでは，信号の意味と使い方のそれぞれを自発的に獲得できること，そしてそれらが互いに報酬を獲得し合える協調行動をコミュニケーションとともに創発することを示した．さらに，作業記憶を持つエージェントでは，能力的格差がない状態からゲームを始めたにもかかわらず，相互作用と行動学習の結果として，信号の送信者と受信者への役割分化が見られ，そのうえ，コミュニケーションに頼る協調行動に比べて若干不安定ではあるが，もはや余剰行動は余剰行動のままとし，コミュニケーションには頼らなくても行動パターンを学習した各個体がそれぞれの作業記憶を利用することで協調行動を実現できることを示した．そして最後に，強化学習に関する各パラメータが協調行動に対してどのような影響を及ぼすのかを調べ，その結果，割引率 γ に大きい値を設定する，すなわち，将来の報酬に対する価値を非常に高いものとすることが協調行動の創発に対して重要な役割を果たすことを示した．また，コミュニケーションの創発には，信号とその内容との結びつきにあらかじめ必然性がなく，様々な組み合わせ方が可能であること，そして，1 つの信号に対して複数の意味や行動の割当てが可能であることが重要である可能性があることを示した．これらの結果は，協調行動やコミュニケーションが必須ではない状況において，意味と信号の任意の対

応付けによる様々なコミュニケーション形態が，コミュニケーションの達成そのものを目的としなくても一般的な行動学習の枠組みにより創発しうることを示す初めての知見である．

本稿で提案した侵入ゲームは，非常に簡素で抽象的なゲームであるにもかかわらず，示唆に富む様々な結果を示すことができた．しかしながら，その単純さゆえに，議論できることは限られており，ここで見てきた以上の複雑なコミュニケーションの創発現象を今回の設定のままで目にすることは叶わないと思われる．コミュニケーションに用いられる信号や言葉は，それを使う個体が置かれた環境を写し取る鏡としてとらえることが可能である．したがって，より複雑なコミュニケーションの創発には，より複雑な環境とそれに見合う個体の自由度を用意する必要があると思われる．個体にとって，他個体を環境の一部と見なせるならば，その数を増やすことによって環境の複雑さを増すことはできる．その効果としては，多様な言語グループへのクラスター化のダイナミクスや共通言語体系がいかにして生まれるか，などについて議論できることがあげられる．

また，本研究で採用したエージェントは，端的にいうならば，単純な刺激–反応系（Stimulus-Response System）にすぎない．これに対して，力学系を内部モデルとして採用する場合，そのエージェントは自律的に内部状態を変化させること，すなわち，「内部ダイナミクス[15]」を内包することができ，それによって外部刺激と行動との間の「1 対多の関係」を形成することができる．このような内部ダイナミクスを持つエージェントで構成されるマルチエージェント・システムは，エージェントが置かれた環境だけでなくその内部状態の変化にも依存するコミュニケーションというダイナミックな現象を解析するための有用なツールになりうると考えられる．これらの研究を進めることで，コミュニケーション，そして言語の起源に関するさらなる知見が得られるものと期待される．

謝辞 本研究を進めるにあたり，沖縄科学技術研究基盤整備機構大学院大学先行研究プロジェクト神経計算ユニットの各氏には有益な議論をしていただいた．とりわけ，Thomas Strösslin 氏，Viktor Zhumatiy 氏，吉本潤一郎氏，伊藤真氏には数多くの重要なご助言をいただいた．これらの方々にお礼申し上げる．

参 考 文 献

1) Bickerton, D.: *Language and Species*, University of Chicago Press (1990). 筧 壽雄（監訳），岸本秀樹，西村秀夫，吉村公宏（訳）：ことばの進化論，勁草書房 (1998).
2) Cangelosi, A. and Parisi, D.: The emergence of a “language” in an evolving population of neural networks, *Connection Science*, Vol.10, No.2, pp.83–97 (1998).
3) Doya, K.: What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex, *Neural Networks*, Vol.12, pp.961–974 (1999).
4) Doya, K.: Complementary roles of basal ganglia and cerebellum in learning and motor control, *Current Opinion in Neurobiology*, Vol.10, pp.732–739 (2000).
5) Halliday, T.R. and Slater, P.J.B. (Eds.): *Animal Behaviour*, Blackwell Scientific Publications (1983). 浅野俊夫，長谷川芳典，藤田和生（訳）：動物コミュニケーション—行動の仕組みから学習の遺伝子まで，西村書店 (1998).
6) Harnad, S.: The symbol grounding problem, *Physica D*, Vol.42, pp.335–346 (1990).
7) 橋本 敬：構成論的手法，ナレッジサイエンス—知を再編する 64 のキーワード，杉山公造ほか（編），pp.132–135, 紀伊國屋書店 (2002).
8) 橋本 敬：言語進化とはどのような問題か?—構成論的な立場から，第 18 回日本人工知能学会予稿集（CD-ROM）(2004).
9) 林 兼大，有田隆也：明示的な意味共有の不要な自己組織型言語発生モデル，第 31 回知能システムシンポジウム資料，pp.13–18 (2004).
10) 金子邦彦，津田一郎：複雑系のカオス的シナリオ，朝倉書店 (1997).
11) 金子邦彦，池上高志：複雑系の進化的シナリオ，朝倉書店 (1998).
12) 小島英生，村尾 元，玉置 久，北村新三：強化学習エージェント間におけるコミュニケーションの創発に関する研究，第 45 回システム制御情報学会研究発表講演会（SCI’01）論文集，pp.1–2 (2001).
13) Marocco, D. and Nolfi, S.: Emergence of communication in embodied agents: Co-adapting communicative and non-communicative behaviours, *Modeling language, cognition and action: Proc. 9th Neural Computation and Psychology Workshop*, Cangelosi, A., et al. (Eds.) (2004).
14) Mazurowski, M.A. and Zurada, J.M.: Emergence of communication in multi-agent systems using reinforcement learning, *Proc. 4th IEEE International Conference on Computational Cybernetics* (*ICCC 2006*), pp.281–286 (2006).
15) Sato, T. and Hashimoto, T.: Dynamic social simulation with multi-agents having internal

dynamics, *New Frontiers in Artificial Intelligence: JSAI 2003 and JSAI 2004 Conferences and Workshops Niigata, Japan, June 2003 and Kanazawa, Japan, May/June 2004 Revised Selected Papers*, Sakurai, A., et al. (Eds.), LNAI Vol.3609, pp.237–251, Springer-Verlag (2007).
16) 柴田克成，伊藤宏司：利害の衝突回避のための交渉コミュニケーションの学習，計測自動制御学会論文集，Vol.35, No.11, pp.1346–1354 (1999).
17) Steels, L. and Vogt, P.: Grounding adaptive language games in robotic agents, *Proc. 4th European Conference on Artificial Life*, Husbands, C. and Harvey, I. (Eds.), MIT Press (1997).
18) Sutton, R.S. and Barto, A.G.: *Reinforcement Learning*, MIT Press (1998). 三上貞芳，皆川雅章（訳）：強化学習，森北出版 (2000).
19) 天正新二郎，前川 聡，吉本潤一郎，柴田智広，石井 信：マルチエージェント環境におけるコミュニケーションの段階的創発，電子情報通信学会「人工知能と知識処理」・情報処理学会「知能と複雑系」合同研究会，AI2004-83, Vol.104, No.727, pp.19–24 (2005).
20) Tomasello, M.: *The Cultural Origins of Human Cognition*, Harvard Univ. Press (1999). 大堀壽夫，中澤恒子，西村義樹，本多 啓（訳）：心とことばの起源を探る—文化と認知，勁草書房 (2006).
21) Werner, G.M. and Dyer, M.G.: Evolution of communication in artificial organisms, *Artificial Life II*, Langton, C., et al. (Eds.), pp.659–687, Addison-Wesley (1991).

(平成 18 年 12 月 21 日受付)
(平成 19 年 6 月 8 日再受付)
(平成 19 年 10 月 1 日再々受付)
(平成 19 年 10 月 11 日採録)

佐藤 尚

1974 年生．2005 年 3 月北陸先端科学技術大学院大学知識科学研究科知識システム基礎学専攻博士後期課程修了．同年 4 月沖縄科学技術大学院大学先行研究プロジェクト神経計算ユニット研究員．2007 年 10 月より沖縄工業高等専門学校メディア情報工学科助教．複雑系，人工生命，マルチエージェント・システム，ミクロマクロ・ループ，進化言語学等の研究に従事．博士（知識科学）．人工知能学会，進化経済学会各会員．2004 年度（第 18 回）人工知能学会全国大会優秀賞．

内部 英治

1972 年生．1999 年大阪大学大学院電子制御機械工学専攻博士後期課程修了．同年日本学術振興会未来開拓推進事業「分散協調視覚による動的 3 次元状況理解」プロジェクト，2001 年 4 月科学技術振興事業団 ERATO 川人動態脳学習プロジェクト，同年 10 月（株）ATR 人間情報科学研究所，2003 年 5 月（株）ATR 脳情報研究所の研究員．2004 年より沖縄科学技術大学院大学先行研究プロジェクト神経計算ユニット研究員．強化学習，ロボティクスの研究に従事．博士（工学）．1998 年度人工知能学会研究奨励賞，2004 年 SAB best philosophical paper award.

銅谷 賢治

1961 年生．1986 年 3 月東京大学大学院工学系研究科計数工学専攻修士課程修了．同年 4 月東京大学工学部計数工学科助手．1991 年 7 月カリフォルニア大学サンディエゴ校生物学科客員研究員．同年 9 月東京大学大学院工学系研究科博士（工学）．1993 年 10 月ハワードヒューズ医学研究所およびソーク生物学研究所研究員．1994 年 10 月 ATR 人間情報通信研究所主任研究員．1996 年 10 月科学技術振興事業団 ERATO グループリーダー．1999 年 11 月科学技術振興事業団 CREST 研究代表者．2003 年 5 月 ATR 脳情報研究所計算神経生物学研究室室長．2004 年 3 月沖縄科学技術大学院大学先行研究プロジェクト神経計算ユニット代表研究者．Society for Neuroscience，日本神経科学学会，日本神経回路学会，電子情報通信学会各会員．2000，2003，2005，2006 年日本神経回路学会論文賞，2007 年日本学術振興会賞，塚原仲晃記念賞．