缶蹴りにおける強化学習を用いた身体的インタラクション
Physical Interaction of Kick the Can Using Reinforcement Learning

高田 亮介 †，竹内 勇剛 †
Ryosuke Takata, Yugo Takeuchi
† 静岡大学大学院総合科学技術研究科
Graduate School of Integrated Science and Technology, Shizuoka University
takata.ryosuke.18@shizuoka.ac.jp

概要

人を含む生命は，個体それぞれが決められたルールに従って行動することで，マクロから見ると協調的に振る舞っていることがある．本研究では，コンピュータシミュレーションにおいてもルールに記述されない戦略がボトムアップに獲得されるのか，という点に注目する．題材として缶蹴りを強化学習させ，集団としての社会性が形成されることを確認した．知見として，ゲームルールに記述されない，他エージェントを参照した振る舞いを習得することが明らかになった．この成果は，従来のようにトップダウンに関係を記述することで社会性が生まれるのではなく，ボトムアップに社会性が創発することを示唆している．

キーワード：缶蹴り, 強化学習, 集団, 役割

1. はじめに

缶蹴りは，世代を超えて受け継がれてきた遊びである．ルールが単純であるにもかかわらず，缶蹴りを遊んでいる人の間には多様なインタラクションが見られる．ずっと陰に隠れて様子を伺う者，息を合わせて飛び出して缶を蹴ろうとする者，オニは時々缶から離れてみせて缶に注意を向けていないふりをする．本研究では，缶蹴りに見られる身体的なインタラクションに注目する．特に，エージェントの全身的な移動動作に焦点を当てる．

従来の社会性は，トップダウンに決定されたモデルに従って振る舞うことで形成されていた．例えば，竹内 (2000) や中嶋 (2004) は予め決められた社会的応答を実行するエージェントに社会性が見出せることを示唆している [1][2]．さらに，坂本ら (2019) の接近行動分析 [3] に見られるように，身体の位置関係や向きといった身体的なパラメータに注目することでエージェントの身体的インタラクションをモデル化することが可能である．しかしながら，接近行動分析のようにトップダウンにモデル化する手法では，現実社会に適用する場合には起こり得る状況を全て記述する必要があるため，限界がある．この問題を解決するためには，ボトムアップなモデル化手法が必要である．そこで本研究では，従来のようにトップダウンに決定される社会性ではなく，集団をボトムアップにモデル化することで形成される社会性に注目する．ボトムアップに形成される社会性を確認するためには，個体レベルの振る舞いをモデル化すればよい．ここで，ルールが定められている環境であれば，機械学習によってシミュレーションが実現可能であることに注目する．本研究では，ボトムアップに振る舞いを構成するための手法として，強化学習を用いる．強化学習によって，人や他の機械に対して注意を払っていないかのように見せかける “駆け引き” や，同じ目標に向かって互いに調整しながら行動する “協調” といった高度なインタラクションを行うことができれば，インタラクションにおける戦略が広がり [4]，そこから集団としての社会性が創発されると考えられる．

佐藤ら (2007) は，侵入ゲームと呼ばれるゲーム内のエージェントに 2 値シグナルを発する仕組みを取り入れることで，強化学習によって協調行動とコミュニケーションが創発することを示唆している [5]．しかしながら，侵入ゲームは 1 次元空間のごく簡単なゲームであり，より現実空間に近い 3 次元環境での協調やコミュニケーションの創発は確かめられていない．現実社会においても協調的な振る舞いやコミュニケーションが創発することを確認するためには，より現実空間に近い環境で実験を行う必要がある．また，佐藤 (1996) は人と機械の協調を実現するためには「補完機能」つまり役割分担が可能であることが必要だとしている [6]．以上を踏まえて，本研究では侵入ゲームより複雑な缶蹴り環境で，役割分担による協調やコミュニケーションが行われるか，そして役割を持った集団によって社会性は形成されるか確認する．

本研究では，単純なルールで多様なインタラクショ

ンが見られる缶蹴りを題材にして，上述したようなエージェント間の駆け引きや協調的な振る舞いが創発する可能性をシミュレーションによって確認することを目的とする．題材に缶蹴りを選んだ理由としては，缶蹴りは伝統的な遊びで親しまれているため人に対しても実験を適用しやすいこと，オニがプレイヤを見つけた際にシグナル発信（「見つけた」という発声）を行うため，侵入ゲームの研究で確認された協調的な振る舞いやコミュニケーションがより現実空間に近い環境で確認できることが期待できることが挙げられる．本研究の成果は，缶蹴りのようなシグナルを発する状況における協調的な振る舞いや駆け引きを強化学習によって実現することにより，現実社会でも人とコミュニケーションしながら協調したり駆け引きを行うことのできるエージェントの実現への寄与が期待できる．

2. 缶蹴り

本研究では，日本で遊ばれている缶蹴りを題材とする．

2.1 缶蹴りのルール

缶蹴りの手順は以下の通りである．なお，本研究で注目しない手順については省略している．

1. 参加者をオニとプレイヤの役に分ける．
2. 缶をフィールドの中央に配置し，オニは 10 秒のカウントダウンを開始する．プレイヤはこの間に隠れる．
3. オニのカウントダウンが終わったらオニはプレイヤを探す．プレイヤを見つけた場合はシグナルを出した後に缶に触れることで，見つかったプレイヤは退場する．
4. オニがプレイヤを全員退場させたらオニの勝利でゲームを終了する．プレイヤのうち 1 人でも缶に触れたらプレイヤの勝利でゲームを終了する．

2.2 缶蹴りにおける振る舞い

本研究では，エージェントの振る舞いを観測することで考察を行う．ここでは，缶蹴りにおいて注目すべき振る舞いを述べる．缶蹴りでは，オニがプレイヤに対して注意を払っていないかのように振る舞うことがある．例えば，オニがプレイヤのいる方向を把握しておきながら別の方向に視線を向けることで，プレイヤは自分に注意が向けられていないと判断する可能性が高い．また，プレイヤはオニの死角から缶に向かうために，別のプレイヤが囮になることがある．本研究ではこのような高度なインタラクションを対象にしてモデル化を試みる．以上をまとめると，缶蹴りで注目するインタラクションは以下の 2 点になる．

(1) 缶に向かうプレイヤとオニとの駆け引き
(2) 缶に向かうプレイヤ同士の協調

(1) ではオニがプレイヤを探すために缶から離れる振る舞いとオニがプレイヤを捕まえるために缶の近くにいる振る舞いの切り替えに注目する．また，(2) ではプレイヤが全てオニに捕まることが無いようにタイミングを合わせて缶に向かったり，わざとオニに見つかるように振る舞う囮役のプレイヤと，オニが囮を捕まえている隙に缶に触れる攻撃役のプレイヤに役割分担が創発することに注目する．このような駆け引きや自己犠牲的な協調の振る舞いを，エージェント間の距離や身体の向きといった身体的なパラメータでのモデル化を目指す．これにより，駆け引きや協調的な振る舞いが創発する条件の記述が可能になり，将来的には人と機械の駆け引きや協調を実現するために寄与すると考えられる．

3. 学習実験

本研究で行った強化学習実験とその考察を述べる．

3.1 実験方法

缶蹴りの環境は Unity で作成した．Unity は 3 次元仮想環境の物理演算が可能で，缶蹴りのようなエージェントの相互作用がもたらす複雑系のシミュレーションに適している [7]．Unity で作成した缶蹴り環境を図 1 に示す．図 1 において，青いエージェントがプレイヤ，紫のエージェントがオニ，白い円柱は決められた位置に生成される壁である．

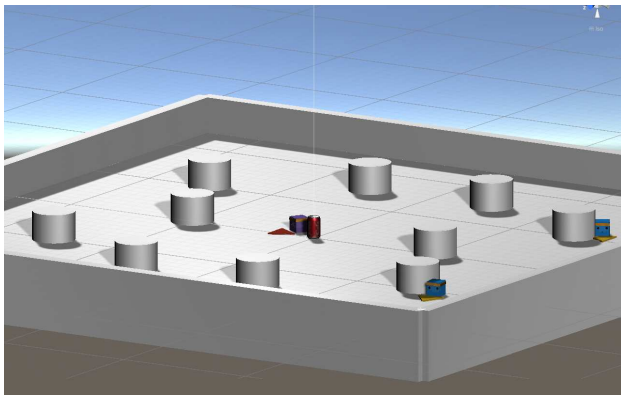

図 1 Unity で構築した缶蹴り環境

シミュレーションに用いるオニ，プレイヤは自律移動型のエージェントであり，前方 120 度の視界を有している．図 2 のように，視界内には 11 本の光線を飛ばし，エージェントは光線に当たっているオブジェクトの情報を取得する．

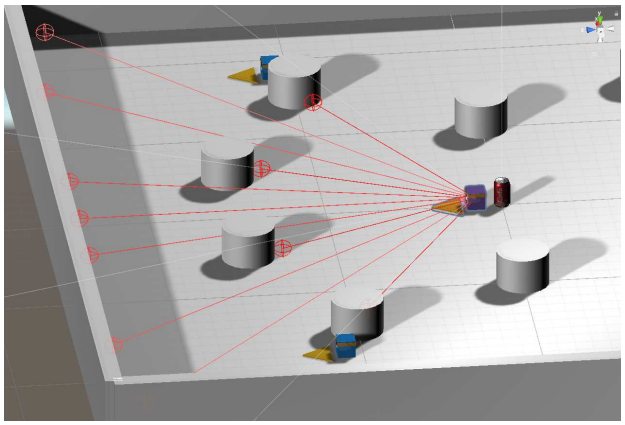

図 2　エージェントの視界と光線

フィールドの外周は壁に囲まれており，エージェントが隠れるためのオブジェクトとして円柱を 10 個配置する．缶および円柱はフィールド内の決められた位置に配置される（図 3）．

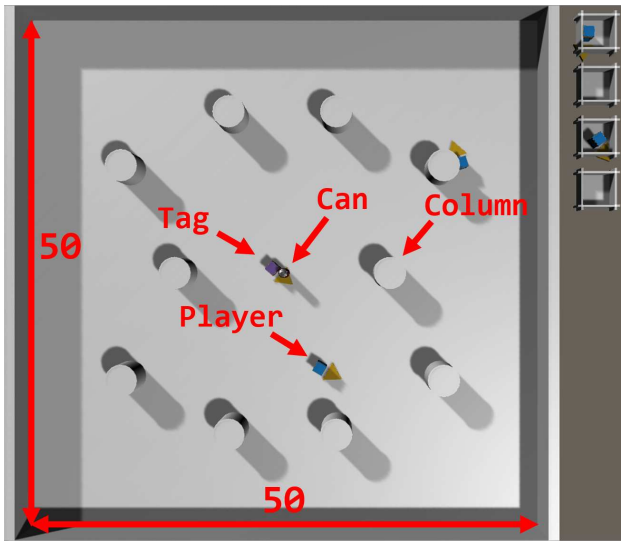

図 3　フィールドの詳細

強化学習アルゴリズムには Unity ML-Agents に搭載されている PPO (Proximal Policy Optimization)[8] を用いた．PPO は，環境からの情報取得と目的関数の最適化を交互に繰り返すアルゴリズムであり，ゲーム課題や物理演算シミュレーション等で成果を出している [8][9]．PPO の特徴は，方策関数を更新する際に，その変化量が大きくなり過ぎないようにクリッピング操作を行うことで，学習を安定化させている点である．方策の更新は式 (1) に従って行う．クリッピング操作は，式 (2) に示す方策の変化量比の値が $1-\epsilon$ より小さい場合，および $1+\epsilon$ より大きい場合に変化量を一定の値にする処理である．

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t[min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)] \tag{1}$$

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \tag{2}$$

強化学習におけるエージェントの状態空間は，図 2 のようにエージェントの視界内に飛ばされた 11 本の光線に当たっているオブジェクトの情報と，エージェントの絶対位置情報，発見情報から構成される．発見情報は 2 値をとる変数で，オニの場合は「プレイヤのうち誰かを発見しているか」，プレイヤの場合は「自分がオニに発見されているか」を表す．状態空間を表 1 にまとめる．なお，視界の光線はプレイヤを個体識別することが可能であるため，オニがプレイヤを見つけてシグナルを出す（名前を呼ぶ）という缶蹴りのルールを実装できる．

表 1　状態空間

状態	次元数
エージェントの視界内光線に当たっているオブジェクト情報	11
エージェントの絶対位置座標	3
発見情報（オニ: 誰かを発見したか，プレイヤ: 自分が発見されたか）	1

強化学習で使用する報酬に関しては，2.1 章で述べた缶蹴りのルールにおける勝利条件，敗北条件に従って設定した．オニの勝利条件かつプレイヤの敗北条件は，オニがプレイヤを発見し缶に触れることであるため，その一連の行動に対して報酬を設定した．プレイヤの勝利条件かつオニの敗北条件は，プレイヤが缶に触れることであるため，その行動に対して報酬を設定した．オニとプレイヤの利害は完全に対立しているため，缶蹴りは零和ゲームであるとみなして報酬設計を行った．以上の報酬系を表 2，および表 3 にまとめる．学習の初期状態では，学習器であるニューラルネットワークの重みがランダムに設定されるため，最初はエージェントがランダムウォークするだけの状況だが，学習を重ねることで表 2 や表 3 の報酬が最大になるような振る舞いを獲得する．

表 2　オニの報酬

内容	値
オニが（プレイヤ発見後）缶に触れる	+1/PNUM
プレイヤが缶に触れる	-1
時間経過	-0.0005

表 3　プレイヤの報酬

内容	値
プレイヤが缶に触れる	+1
オニが（プレイヤを発見し）缶に触れる	-1/PNUM
時間経過	+0.0005

PPO におけるハイパーパラメータは表 4 にように設定した．なお，今回の設定は Unity ML-Agents のデフォルト設定を用いた．

表 4　PPO のハイパーパラメータ

パラメータ名	値
バッチサイズ	128
バッファサイズ	2048
バッファに追加するステップ数	64
方策変化量の閾値 ϵ	0.2
エントロピー正規化率 β	0.005
正規化パラメータ λ	0.95
学習率 η	0.0003
割引率 γ	0.99
エポック数	3
隠れ層のニューロン数	256
隠れ層の数	2
RNN メモリサイズ	128
RNN 経験シーケンス長	64

学習実験は，オニ 1 体に対してプレイヤの数を変更した 4 条件で行った．この 4 段階のプレイヤ数の変化により，インタラクション対象の個体数が変化した場合に本研究の目的である駆け引きや協調的な振る舞いに関して変化があるのか確認した．特に，集団における個体数の変化によって協調の様相である役割がどのように変化するのかを明らかにするために今回の条件群を用意した．実験条件を表 5 に示す．表 5 において，PNUM はプレイヤの数を表す．なお，すべての条件で 5,000,000 ステップ学習させた．

表 5　実験条件

条件名	オニの数	プレイヤの数
C-1vs1	1	1
C-1vs2	1	2
C-1vs3	1	3
C-1vs4	1	4

評価は，PPO によって獲得したオニとプレイヤの振る舞いを観察することで行う．具体的には，以下のように協調と駆け引きを定義し，評価を行う．

プレイヤ間の協調 プレイヤ間で役割分担することによってプレイヤ側が勝利する．

オニとプレイヤの駆け引き プレイヤがオニに見つかっていない状態で，オニが缶と円柱の間を往復する．

3.2　学習結果と考察

3.2.1　C-1vs1

条件 C-1vs1 における学習曲線を図 4 に示す．図 4 において，赤線がオニ，緑線がプレイヤである．学習はオニ，プレイヤのどちらも振動しており，学習が終了した 5,000,000 ステップ時点ではオニの方が優勢になっていることがわかる．学習が振動した理由としては，オニとプレイヤの報酬系が対立していることによる共進化的な作用のためだと推察される．オニは，ある時点でプレイヤが得た戦略に対応するように戦略を学習し，プレイヤはオニの戦略にさらに対応するように学習する．この学習がオニとプレイヤの双方で繰り返されることにより，学習が振動したと考えられる．

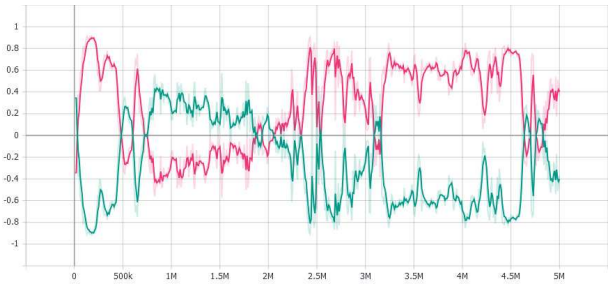

図 4　C-1vs1 の学習結果（横軸:ステップ数，縦軸:得られた報酬，赤:オニ，緑:プレイヤ）

また，条件 C-1vs1 におけるエージェントの移動経路を図 5 に示す．図 5 では，プレイヤはフィールド左上の円柱からフィールド中央の缶に向かって移動して

おり，オニは缶の位置から弧を描くように上方向に移動し，上の円柱の位置まで移動したらまた缶の位置に戻っていく様子が表されている．最終的にはオニが先に缶に触れ，オニの勝利でゲームが終了した．オニは，左上方向以外に対してプレイヤがいるか確認する動作を見せていないことから，左上にプレイヤがいることを学習していると推測される．3.1 節に示した指標に基づくと，この振る舞いは協調も駆け引きも行っていないと言える．

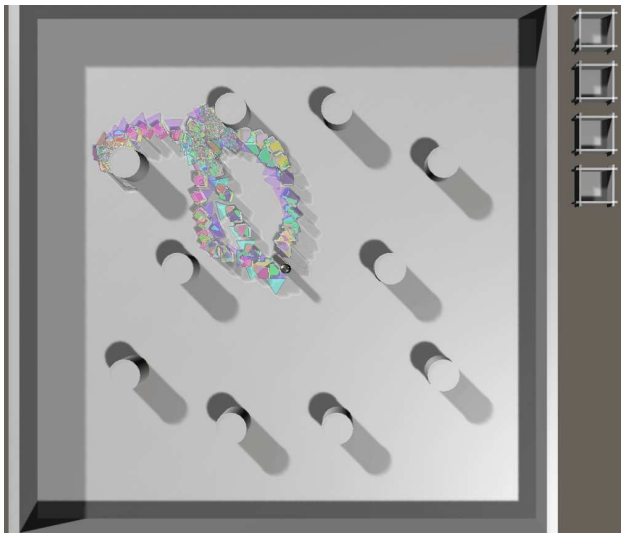

図 5　C-1vs1 におけるエージェントの移動経路

3.2.2　C-1vs2

条件 C-1vs2 における学習曲線を図 6 に示す．図 6 において，橙線がオニ，その他の線がプレイヤである．図 6 の結果から，図 4 と同様に，学習が振動していることが見てとれる．約 2,800,000 ステップまではオニとプレイヤの結果は交差しているが，約 2,800,000 ステップからはオニが優勢になり，そのまま学習が終了した．

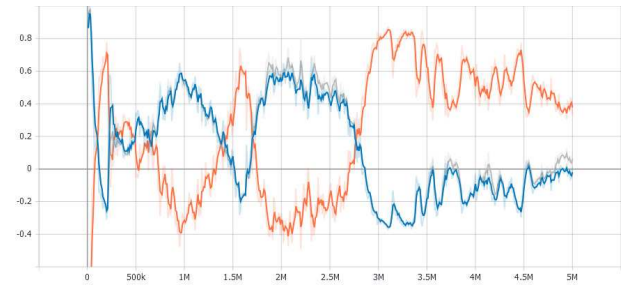

図 6　C-1vs2 の学習結果（横軸:ステップ数，縦軸:得られた報酬，橙:オニ，青・灰:プレイヤ）

また，条件 C-1vs2 におけるエージェントの移動経路を図 7 に示す．図 7 では，プレイヤは 2 体存在する（左上がプレイヤ 1，右上がプレイヤ 2）．プレイヤ 1 は，左上方向に円柱から離れるように移動している．プレイヤ 2 は，円柱の後ろから缶に向かって移動している．オニは，上方向に向かって移動し，プレイヤ 1 を発見した後に缶に触れてプレイヤ 1 を捕まえている．プレイヤ 2 は，オニがプレイヤ 1 を捕まえるために缶に向かって移動するタイミングと同時に缶に向かって移動を始めていた．最終的にはプレイヤ 2 が缶に触れて，プレイヤの勝利でゲームが終了した．プレイヤ 1 はプレイヤ 2 の攻撃をオニの視界に入れないために，オニに見つかることでオニを缶の方向に移動させる必要がある．そのため，プレイヤ 1 はオニに見つかりやすい位置に移動していると推察される．そして，オニがプレイヤ 1 を捕まえるために行動している途中でプレイヤ 2 が缶に向かって移動し，缶に触れることに成功している．つまり，プレイヤ 1 はプレイヤ 2 が缶に触れるための “囮” の役割を担っていると言える．この結果は，3.1 節の指標に基づくと，役割の創発によって勝利するための協調的な振る舞いである．また，プレイヤ 2 はオニに見つかっておらず，オニは缶と円柱の間を往復していることから，駆け引きの振る舞いも見せていると言える．この囮の役割は，攻撃を行うプレイヤ 2 がいなければ成立しないため，囮の役割を持つプレイヤ 1 と攻撃の役割を持つプレイヤ 2 のそれぞれの役割は，2 体のプレイヤが同じ環境で同時に学習したから形成されたと推測される．ここで，例えば別の環境で別のプレイヤ 1 や別のオニと共に学習したプレイヤ 2 を混ぜても，プレイヤ 2 の囮は有効に機能しないと考えられる．

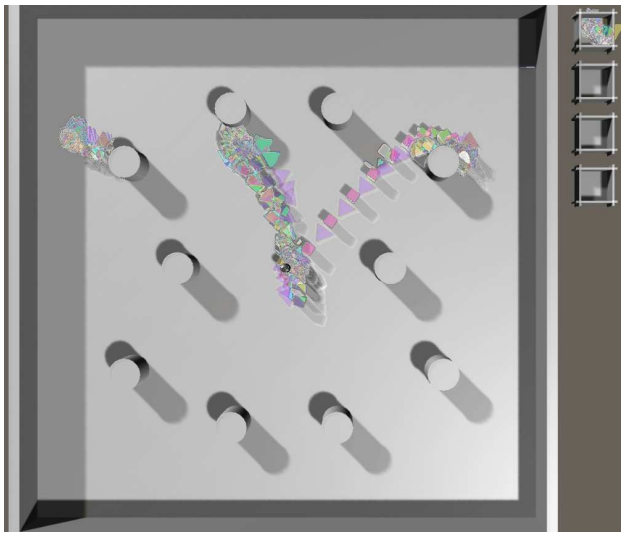

図 7　C-1vs2 におけるエージェントの移動経路

3.2.3 C-1vs3

条件 C-1vs3 における学習曲線を図 8 に示す．図 8 において，赤線がオニ，その他の線がプレイヤである．約 1,200,000 ステップまでは図 4 や図 6 と同様に，オニとプレイヤの得られた報酬が交差を繰り返しているが，約 1,200,000 ステップ以降は，振動はあるが全体的にオニが劣勢になっている様子がわかる．

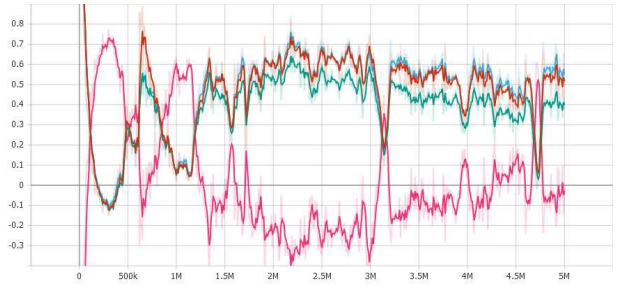

図 8　C-1vs3 の学習結果（横軸:ステップ数，縦軸:得られた報酬，赤:オニ，茶・水・緑:プレイヤ）

また，条件 C-1vs3 におけるエージェントの移動経路を図 9 に示す．図 9 では，プレイヤは 3 体存在する（左上がプレイヤ 1，右上がプレイヤ 2，左下がプレイヤ 3）．プレイヤ 1 は，ゲーム終了まで左上の円柱に隠れている．プレイヤ 2 は，円柱から離れて右上の外周の方に移動して，オニに見つかり捕まっている．プレイヤ 3 は，缶に向かって移動している．オニは，図 5 や図 7 のようにプレイヤを探しに行かず，缶に密接するように缶の周囲を回るように移動している．最終的にオニが右上方向を向いている隙にプレイヤ 3 が缶に触れて，プレイヤの勝利でゲームが終了した．プレイヤ 2 は，C-1vs2 の場合と同様に，オニの気を引き付けてわざと捕まる “囮” の役割を担っていると推察される．この結果は，3.1 節の指標に基づくと，役割による協調的な振る舞いであると言える．ここで，“囮” の役割を持つプレイヤ 2 と “攻撃” の役割を持つプレイヤ 3 が互いの役割を有効に機能させるためには，同じ環境で学習する必要があると考えられる．

3.2.4 C-1vs4

条件 C-1vs4 における学習曲線を図 10 に示す．図 10 において，青線がオニ，その他の線がプレイヤである．約 1,800,000 ステップまでは図 4，図 6，図 8 と同様に，オニとプレイヤの得られた報酬が交差を繰り返しているが，約 1,800,000 ステップ以降は，オニが劣勢になっている様子がわかる．プレイヤの数が増えると，表 2 に示したオニの負の報酬を得やすくなるため，条件 C-1vs4 は他の条件と比べてオニとプレイヤの得られた報酬に差が出たと考えられる．

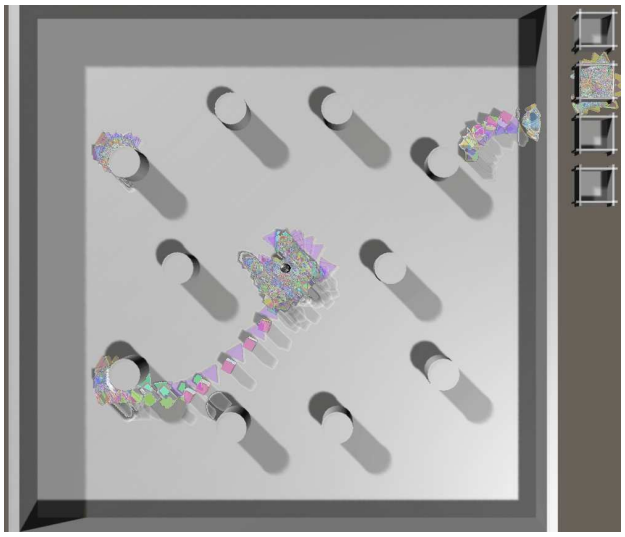

図 9　C-1vs3 におけるエージェントの移動経路

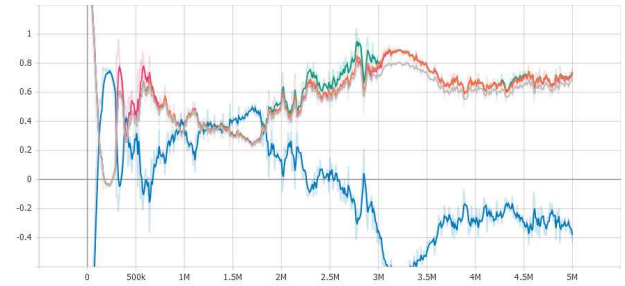

図 10　C-1vs4 の学習結果（横軸:ステップ数，縦軸:得られた報酬，青:オニ，橙・灰・緑・赤:プレイヤ）

また，条件 C-1vs4 におけるエージェントの移動経路を図 11 に示す．図 11 では，プレイヤは 4 体存在する（左上がプレイヤ 1，右上がプレイヤ 2，左下がプレイヤ 3，右下がプレイヤ 4）．プレイヤ 1，プレイヤ 2，プレイヤ 3 は，ゲーム終了まで円柱の後ろに隠れている．プレイヤ 4 は他のプレイヤと異なり，缶に向かって移動している．オニは，条件 C-1vs3 の振る舞いと同様に，缶に密接するように缶の周囲を回るように移動している．最終的にオニが左上方向を向いている隙にプレイヤ 4 が缶に触れて，プレイヤの勝利でゲームが終了した．この振る舞いは，3.1 節の指標に基づくと協調的であると言える．他の条件の傾向から，囮の役割を持つプレイヤが出現すると予想されたが，攻撃を行うプレイヤ 4 以外のプレイヤは全て円柱の後ろに隠れる振る舞いを見せた．この振る舞いには，表 3 に示したプレイヤの報酬系における時間経過による微少な報酬が関わっていると考えられる．プレイヤはゲームが長引くだけで報酬が得られるため，“円柱の後ろに隠れる” という振る舞いは，ゲームを長引かせて報酬を得る目的で行っていると推察される．オニは報酬

を得る（あるいは負の報酬を得ない）ために，隠れたプレイヤ 1・プレイヤ 2・プレイヤ 3 を探すために缶から離れるか，プレイヤ 4 の攻撃に対処できるように缶の近くにいるかという選択を行う必要がある．その結果，オニは缶の周囲を移動する振る舞いと缶から離れる振る舞いを学習したと推察される．このような，缶から離れた後に近づく行動を繰り返すような振る舞いは，3.1 節の指標に基づくと，駆け引きを行っていると言える．

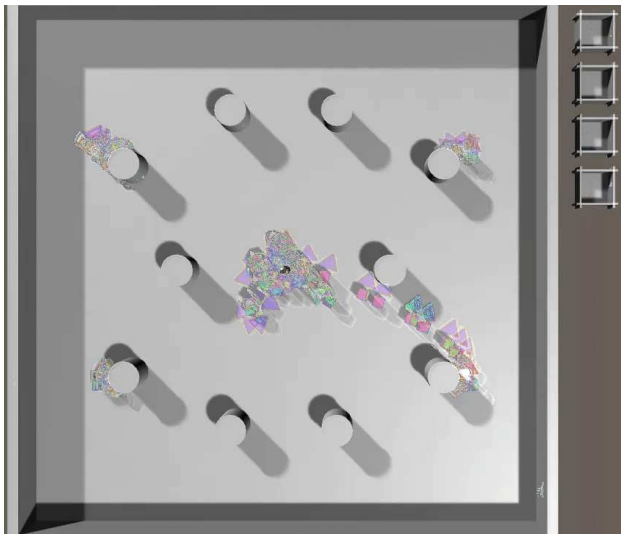

図 11　C-1vs4 におけるエージェントの移動経路

3.2.5　集団の形成と駆け引き・協調の創発

条件 C-1vs1 のように，プレイヤが 1 体の場合はオニに見つかるまで円柱に隠れる振る舞いを見せるが，条件 C-1vs2 や C-1vs3 でのプレイヤは，C-1vs1 のように振る舞うプレイヤが複数体獲得されることは無く，それぞれのプレイヤが別々の振る舞いを見せた．今回の結果を表 6 にまとめる．

表 6　実験で得られた役割

条件名	プレイヤ 1	プレイヤ 2	プレイヤ 3	プレイヤ 4
C-1vs1	隠遁	-	-	-
C-1vs2	囮	攻撃	-	-
C-1vs3	隠遁	囮	攻撃	-
C-1vs4	隠遁	隠遁	隠遁	攻撃

これらの振る舞いは同じ環境下では互いに有効的に機能し，結果的に報酬を得るような振る舞いへと繋がっていた．条件 C-1vs4 においては，4 体のプレイヤのうち 3 体が C-1vs1 と同じ振る舞いだったが，残り 1 体のプレイヤが攻撃を行うため，その攻撃に対処しなければならない．つまり攻撃を行うプレイヤは，円柱の後ろに隠れる 3 体のプレイヤをオニが探しに行かないように牽制する役割も持っていると言える．強化学習によって，オニがプレイヤを探しに行くか，缶の近くにいるか，という駆け引きを獲得したと考えられる．以上より，条件 C-1vs1，C-1vs2，C-1vs3，C-1vs4 とプレイヤ数を増やすに従って，プレイヤは個々的な振る舞いの集まりでは無く，それぞれが役割を持つ集団が形成されていると考えられる．集団の中にいるプレイヤの振る舞いは，個別に見ると報酬に繋がらないが，他のプレイヤの振る舞いと合わさることで報酬に繋がる振る舞いになるため，今回の実験で得られたプレイヤは協調的であると言える．

本研究の成果は，侵入ゲームで見られた協調的な振る舞いやコミュニケーション [5] を，より現実空間に近い環境へ拡張した結果であると考えられる．橋本 (2015) は，コミュニケーションを言語的コミュニケーション・記号非言語コミュニケーション・非記号コミュニケーションの 3 つのレベルに分けている [10]．今回の実験によって創発した缶蹴りのコミュニケーションは，オニによるプレイヤを発見した際のシグナルによる記号非言語コミュニケーション，およびオニとプレイヤ・プレイヤ同士の身体位置関係による非記号コミュニケーションであると考えられる．現実社会に応用してエージェントと人との駆け引きや協調を展望すると，相手の気を引くことが可能であり，かつプレイヤ間で「どのように缶に向かって移動していくか」といった作戦が伝達可能である言語による言語的コミュニケーションを可能にすることが今後の課題である．

3.2.6　エージェントの視点から見た振る舞い

ここまで，環境を俯瞰的に見た第三者的な視点から考察したが，実際の学習はエージェントの主観的な視点から得られる情報を基に行われる．エージェントが得られる環境の情報は，自身の位置と，自身の視界にあるオブジェクトの情報である．そのため，例えば円柱の後ろに隠れたプレイヤは実際には全く観測されていない状態と同じであり，オニが左上に隠れたプレイヤを見つける振る舞いは，学習によって網羅的にフィールドを探索した結果「左上にプレイヤが隠れている」という経験を積んだことにより獲得したと考えられる．さらに，図 7 のように，プレイヤからすると他のプレイヤが視界に入っていなくても，オニを観測して缶を蹴りに行くことを決めることがある．これは，オニを観測することで間接的に味方プレイヤの状態を同定していることになる．学習でプレイヤが行うことは表 3 に示した報酬を最大化することであり，表 6 に示した役割はそのサイドエフェクトとして創発さ

れたものである．以上のように，ボトムアップなアプローチによって目的を達成するために社会性を形成し得ることが示唆された．

3.2.7 集団による社会性

今回の結果のように，組織化された集団としての役割が創発することは，星野 (1993) や上田 (1995)，池上 (2003) が述べるように，生命としての特性を人工的に創り出すことに繋がると考えられる [11][12][13]．星野 (1993) は，ボトムアップに，かつエージェントに個々の動作を行わせる局所的制御による集団の並列行動生成によって，人工生命が成立すると述べている [11]．本研究で実現した缶蹴りエージェントは上記のアプローチに従っており，缶蹴りに勝利するという大目的のもとで自律的に集団的な振る舞いを獲得した人工生命であると言える．

缶蹴りというゲームルールの下で学習することにより，異なる役割を持つ個体から成る集団が創発された．この集団をミクロなレベルで見ると，オニに見つかりやすい位置に移動するプレイヤといったように，個体レベルとしては最適化されていないが，マクロなレベルでは “囮” という役割によってプレイヤ集団の勝利に貢献していることになる．つまり，個体レベルの振る舞いをボトムアップに学習した結果，集団としての社会性が創発されたと言える．

4. おわりに

本研究では，缶蹴り遊びを複数のエージェントで強化学習させることにより，駆け引きや協調といった高次のインタラクションを創発する可能性を確認する目的で実験を行った．実験の結果，プレイヤを 1 体から 4 体まで増加させることで，それぞれ以下に示すように異なる役割と振る舞いを見せる集団が形成された．

(1) プレイヤが 1 体の場合は基本的に円柱の後ろに隠れて，オニに見つかると缶に向かって移動するようになった．
(2) プレイヤが 2 体の場合はオニの気を引くように振る舞う “囮” の役割と缶に向かって移動する “攻撃” の役割に分かれた．
(3) プレイヤが 3 体の場合はゲーム終了まで円柱の後ろに隠れるプレイヤ，オニの気を引くように振る舞う “囮” のプレイヤ，缶に向かって移動する “攻撃” のプレイヤに分かれた．
(4) プレイヤが 4 体の場合は 3 体がゲーム終了まで円柱の後ろに隠れ，1 体が缶に向かって移動した．攻撃を行うプレイヤがいるためにオニは缶から離れることができず，隠れたプレイヤによって時間経過の報酬を得るようになった．

以上の振る舞いはボトムアップに構成されたモデルに従っているため，トップダウンにモデル化するよりも現実社会への適用が容易である．さらに，侵入ゲームのような 1 次元空間のごく簡単なゲームではなく，3 次元空間上でのシミュレーションであるため，現実社会へのインタラクションに応用できる可能性がある．今後はこれらの振る舞いから身体的なパラメータを抽出して，オニがプレイヤを探しに行くかどうかの駆け引きやプレイヤ同士の役割による協調を数理的に分析し，ボトムアップに集団による社会性が創発していることを裏付けるためのモデル化を図りたい．

文献

[1] 竹内勇剛, 片桐恭弘: ユーザの社会性に基づくエージェントに対する同調反応の誘発, 情報処理学会論文誌, Vol.41, No.5, pp.1257-1266 (2000).
[2] 中嶋宏, 森島泰則, 山田亮太, 川路茂保: 人間-機械協調システムにおける社会的知性 -心のモデルとパーソナリティによるエージェントの社会的応答について-, 人工知能学会論文誌, Vol.19, No.3, pp.184-196 (2004).
[3] 坂本孝丈, 吉岡源太, 竹内勇剛: 話しかけ場面における相手の受容度に応じた接近行動のモデルに基づく分析, 日本知能情報ファジィ学会誌, Vol.31, No.5, pp.842-851 (2019).
[4] 伊藤昭, 大橋資紀, 寺田和恵: 非零和ゲームの強化学習–相手の行動を読むプログラム. 情報処理学会研究報告知能と複雑系（ICS）, Vol.109(2005-ICS-141), pp.53–60 (2005).
[5] 佐藤尚, 内部英治, 銅谷賢治: 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, 情報処理学会論文誌数理モデル化と応用 (TOM), Vol.48, No.19, pp.55-67 (2007).
[6] 佐藤知正: 人間機械協調システム, 計測自動制御学会, Vol.35, No.4, pp.262-267 (1996).
[7] Juliani, A., Berges, V. P., Vckay, E., Gao, Y., Henry, H., Mattar, M., Lange, D.: Unity: A general platform for intelligent agents, arXiv preprint arXiv:1809.02627 (2018).
[8] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347 (2017).
[9] Bøhn, E., Coates, E. M., Moe, S., Johansen, T. A.: Deep reinforcement learning attitude control of fixed-wing uavs using proximal policy optimization, In 2019 International Conference on Unmanned Aircraft Systems (ICUAS) ,pp.523-533 (2019).
[10] 橋本敬: 言語とコミュニケーションの創発に対する複雑系アプローチとはなにか, 計測と制御, Vol.53, No.9, pp.789-793 (2016).
[11] 星野力: 人工生命の現状と将来への期待, 計測と制御, Vol.32, No.8, pp.677-683 (1993).
[12] 上田完次: 人工生命と生物指向人工物, 科学基礎論研究, Vol.23, No.1, pp.1-6 (1995).
[13] 池上高志: 人工生命から見た集合知, 人工知能学会誌, Vol.18, No.6, pp.690-696 (2003).