修士論文

不完全情報同時手番ゲームにおいて人間のような読み合いを演出するAIプレイヤ

小西 健太郎

主指導教員 池田 心

北陸先端科学技術大学院大学
先端科学技術研究科
（情報科学）

令和5年3月

Abstract

Artificial intelligence (AI) techniques are actively researched in various fields, including image processing and natural language processing. These techniques are expected to be further developed and contribute to society. Games are also one of the actively researched fields. As a symbol of intelligence, for games such as chess, it has long been a goal to create AI players that surpass human players. Although this goal has been achieved in many games, such AlphaGo for the game of Go, there are still many issues to be addressed in research to use AI players in various ways other than being strong opponents. Examples include "entertaining" and "teaching".

In imperfect information games, where players cannot accurately grasp some information, there are Nash equilibrium strategies that are not taken advantage of by the opponents. However, it is infeasible for human players to obtain such strategies in complex games. Therefore, human players usually try to "read" their opponents' minds to predict the opponents' information and actions. The players then choose actions that may give them advantages according to the predictions. In some games, most human players choose their actions based on "mind games" where the players try to read the opponents' minds and also consider that their minds are read by the opponents. This kind of gameplay is an important element in the enjoyment of the game.

However, only a few researchers focused on creating AI players that play mind games. Existing AI players are hard to produce behaviors as they are playing mind games. This is one of the challenges for AI players in some imperfect information games.

In this study, we focus on simultaneous games (i.e., players choose actions at the same time) and try to make AI players play human-like mind games. We propose a simplified version of Pokémon battle, a game in which the main strategy among human players is to read each other's actions based on type superiority.

We propose a method to play mind games, which is to generate moves that human players find natural (called standard moves) and then strengthen moves with human-like habits or tendencies and moves that mimic humans' reading (called bias moves). The AI player for the proposed method consists of a standard move generator, a bias move generator, and an exploitation strategy. The approaches and their effectiveness are explained as follows.

The standard move generator generated standard move strategies that look natural to human players. As the first step, we created AI players based on the Monte-Carlo method. These Monte-Carlo players were strong to some extent, where the win rates against random players were higher than 97%. However, we also confirmed that some moves were unnatural due to problems such as overesti-

mation or underestimation of states.

To solve these problems, we calculated the theoretical win rates (payoffs) for all states of the simplified Pokémon battle by retrograde analysis. We then created a Nash player that follows the Nash equilibrium strategy based on the payoff matrix. The Nash player had a win rate of 72% against the Monte-Carlo player. In addition, the Nash player had fewer unnatural moves that would be considered "bad moves" by humans. However, we face two new challenges. First, the Nash player is too strong as an opponent for average human players. Second, the Nash player sometimes fails to select moves that seem promising to humans.

To adjust the strength of the Nash player, we created a δ-Nash player that adds noise of size δ to the payoff matrix (mimicking human misperception) before finding a Nash equilibrium strategy. Taking the initial state of the game as an example, the δ-Nash player uses a mixed strategy in which three or more actions, including moves that look promising to humans, are selected probabilistically. We consider that this makes the AI players more human-like in the game. In an evaluation experiment, we let the δ-Nash players with different δ settings play against the Nash player. The win rates were between 31% and 53%, which means that we could adjust the strength of the δ-Nash players by varying δ.

The bias move generator generated move strategies with human-like habits and tendencies by adding additional biases to the payoff matrix for the δ-Nash player. We created four players with different biases: favoring attack, favoring exchange, favoring effective attack, and favoring ineffective attack. We confirmed that the four players had higher probabilities of selecting the moves corresponding to the preferences we assigned. The player favoring attack selected 8% more attacks than the δ-Nash player, the one favoring exchange selected 13% more exchanges, the one favoring effective attack selected 26% more effective attacks, and the one favoring ineffective attack selected 29% more ineffective attacks.

When evaluating the strength of the biased players in terms of their win rates against the Nash player, we found that even the player with the worst win rate (the player favoring ineffective attacks whose win rate was 43.6%), the win rate was comparable to the δ-Nash player whose win rate was 45.6%. The results indicated that the biases did not force the biased players to select the biased moves in states where these moves were clearly inappropriate.

In addition, we conducted experiments to examine the effects of noise δ and bias parameter α on player strength and move probability. We found that the noise δ and the bias parameter α can be adjusted relatively independently for the strength (win rate) and the move bias, respectively. This allows us to create a variety of move strategies suited to playing mind games.

Finally, for the exploitation strategy, we proposed a method that mimicked hu-

man players' reading in games to make human players believe that they have been read. This method has not yet been evaluated with a complete implementation and will be left as future work.

概要

人工知能技術（AI）は画像処理や自然言語処理をはじめとして様々な分野で盛んに研究されており，今後も技術発展と社会への貢献が期待されている．ゲームも当該分野の 1 つであり，知性の象徴とされるチェスなどのゲームにおいて，AI プレイヤが人間を超える強さを獲得することは長らく目標とされてきた．囲碁の AlphaGo をはじめとして多くのゲームでこの目標は達成されたものの,「楽しませる」「教える」など AI プレイヤに強い対戦相手以外の多様な用途を持たせる研究については課題もまだ多い．

プレイヤが一部の情報を正確に把握できない不完全情報ゲームには,「負けない戦略・付け込まれない戦略」であるナッシュ均衡戦略が存在するが，複雑なゲームで人間プレイヤがこの戦略を求めることは事実上不可能である．そのため人間プレイヤは，相手の情報や行動を予測し，予測に対してアドバンテージを得られる行動を選択する「読み」を用いる場合がある．人間同士が読みの思考に基づいて意思決定を行う場合や，相手が読んでくることを考慮して意思決定を行う「読み合い」は一部のゲームでは主流な意思決定方法の 1 つであり，この特有の駆け引きはゲームを楽しむ上で重要な要素の 1 つといえる．

しかし,「読み合い」の駆け引きに着目した AI プレイヤは少なく，人間プレイヤの対戦相手として読み合いの駆け引きを演出することは難しい．この点は一部の不完全情報ゲームにおける AI プレイヤの課題の 1 つだと考える．

本研究では，同時手番ゲームにおける読み合いに着目し，AI プレイヤに人間のような読み合いを演出させる手法の確立に取り組む．読み合い演出の対象ゲームとして，人間プレイヤ間ではタイプ相性から相手の行動を読み合う戦略が主流であるポケモンバトルを簡易化したゲームを提案し作成した．

読み合いを演出する手法として，人間プレイヤが自然に感じる標準着手を生成した上で，人間のような癖や傾向を持つ着手や，人間の読みを模倣する着手を強調することを提案した．提案手法を実現するための，読み合いを演出する AI プレイヤは，標準着手生成部，バイアス着手生成機能，搾取戦略実施機能から成り，それぞれについて有効性やアプローチを示した．

標準着手生成部では，人間プレイヤが自然に感じる標準着手戦略の生成に取り組んだ．モンテカルロ法によって着手を決定するモンテカルロプレイヤはランダムプレイヤ相手に 97%以上の勝率を示したものの，予備実験では状態の過大・過小評価などの問題によって不自然な着手も確認された．

そのため，簡易版ポケモンバトルの全状態の真の勝率（利得）を後退解析で計算した後，その利得行列からナッシュ均衡戦略求めてそれに従うナッシュプレイヤを作成した．ナッシュプレイヤはモンテカルロプレイヤに対して，72%の勝率を示し，かつ人間が「ひどい手」と感じるような不自然な着手を減らすことができた．しかし，平均的な人間プレイヤの対戦相手としては強すぎる，人間には有望に感じる一部の行動をまったく選択しない場合があることが新たに課題となった．

ナッシュプレイヤの強さを調整するため，利得行列に大きさ δ のノイズ（人間の

誤認知を模したもの）を加えてからナッシュ均衡戦略を求める δ-ナッシュプレイヤを作成した．δ-ナッシュプレイヤは例えば初期局面において，人間には有望に見える手を含む 3 つ以上の行動を確率的に着手する混合戦略を用いる．これにより対戦していてより人間らしいと感じられる相手を作ることができたと考える．また，ナッシュプレイヤとの対戦勝率をベンチマークとした評価実験では，ノイズのサイズや性質を変化させることで勝率 53%~勝率 31%の幅で強さを調整できることを示した．

バイアス着手生成機能では，δ-ナッシュ均衡戦略を求める誤認知利得行列に更にバイアスを加えることで，人間が持つ癖や傾向を付与した着手戦略を生成した．攻撃/交換/有効攻撃/非有効攻撃を好むバイアスを付与した 4 つのプレイヤは，それぞれ攻撃選択確率が 8%，交換選択確率が 13%，有効攻撃選択確率が 26%，非有効攻撃選択確率が 29%増加するなど付与したバイアスに従った選択確率の増加が見られた．

また，バイアスを付与したプレイヤの強さについて，ナッシュプレイヤとの対戦勝率で評価したところ，通常の δ-ナッシュプレイヤが 45.6%であるのに対し，最も成績が悪い非有効攻撃を好むプレイヤであっても 43.6%の勝率を示し，明らかに適切でない状況でそれらの行動を無理やり取るようなことはしていないことが分かった．

さらに，ノイズ δ とバイアスパラメータ α がプレイヤの強さや着手確率に与える影響について評価実験を行った．結果として，ノイズ δ によって強さ (勝率) を，バイアスパラメータ α によって着手の偏りを，比較的独立に調整できることが分かった．これにより，読み合い演出に適した多様な着手戦略が作成できることが分かった．

最後に搾取戦略実施機能については，人間プレイヤに「読まれた」と思わせる着手を生成する方法として，人間の読みを模倣する手法を示した．この手法は完全な実装による評価実験で有効性を示すまでには至らなかったため，今後の課題である．

目 次

図 目 次

表 目 次

第1章　はじめに

近年，人工知能技術（AI）は急速に発展している．なかでもゲームは多くの人にとって身近な遊びであり，ルールも明確でシミュレーションが容易であること・勝率などによる定量的な評価が可能であることから，研究が盛んな分野の１つである．2016 年には囲碁において DeepMind 社が開発した AlphaGo がトップ棋士に勝利するなど [1]，完全情報ゲームでは AI プレイヤが人間以上の能力を発揮している．

ゲームを対象とした AI 研究は，囲碁や将棋などの現在の状態を完全に把握できる完全情報ゲームだけでなく，ポーカーや麻雀などの不完全情報ゲームについても盛んに行われている．不完全情報ゲームは，プレイヤが一部の情報を正確に把握することができないゲームであり，未知の情報があるなかで意思決定を行うことが求められる．このような性質は，経済や政治など人間の社会的活動にも頻繁に登場する．不完全情報ゲームにおいても，AI プレイヤは高い能力を有しており，2019 年に Facebook 社が開発した Pluribus はポーカーで最も一般的なルールである 6 人プレイでの Headsup No-limit Texas Hold'em において，複数のプロとの対戦に勝利した [2].

このように，強さに着目した AI プレイヤは様々なゲームで人間のトッププレイヤ以上の水準に到達しつつある．そのため，強い対戦相手という用途のほかに，人間らしく振る舞うことで人間を楽しませる，人間プレイヤの実力や好みに合わせることで人間に教える・練習相手になるなど多様な用途を持った AI プレイヤの研究も盛んに行われるようになっている [3, 4]．池田らは，モンテカルロ碁において着手の選択確率や勝率に基づく自然な形勢制御，統計量に基づく多様な戦略の演出のためのアプローチを提案し，接待碁のための自然な手加減や，実利派/中央派や楽観派/悲観派など人間プレイヤのような棋風を表現できることを示した [5].

一方で不完全情報ゲームにおいて人間らしさを表現するためには，不完全情報ゲームに固有の人間らしさを考慮しなければならない．不完全情報ゲームの解として，ナッシュ均衡となる混合戦略（確率的戦略）があり，各プレイヤは他のどのような戦略を選んでも利得を高くできず相手につけ込むことができない．したがって，可能な限り自分の目的を達成するように行動する合理的なプレイヤ同士の対戦においてはナッシュ均衡解に従うことが有力である．しかし，人間プレイヤがこれらの戦略を求めるのは計算量が膨大で容易ではないほか，仮に正しい確率的戦略が求まったとしても，正確にその確率分布通りに着手することは難しい．そのため，このような性質を持つゲームにおいて，人間プレイヤは相手の情報や行動

を予測する「読み」を用いて意思決定を行うことがある．予測した情報や行動に対して期待利得が高い行動を選択し，実際に予測が正しかった場合には大きなアドバンテージを得られる．人間同士がこの思考に基づいて意思決定を行う場合や，相手が読んでくることを考慮して意思決定を行う「読み合い」では，現在の局面の情報のみで相手の行動を予測するわけではなく，相手プレイヤの傾向や心理的な状況を予測に用いる場合がある．このとき，人間プレイヤは仮にナッシュ均衡の意味で適切でない戦略に従い着手を行ったとしても，結果としてアドバンテージを得ることができれば駆け引きに勝った自分の行動を肯定することができる．

ポケモンなどの不完全情報ゲームの一部では，人間プレイヤ同士の対戦において読み合いが主流な戦略として用いられている．前述の通りゲームの解であるナッシュ均衡となる混合戦略を用いることは事実上不可能であり，また対戦相手の人間プレイヤには多くの癖や偏りがあるために，読みを用いることが勝つための現実的な近道となるためである．さらに，この特有の駆け引きは不完全情報ゲームを楽しむ上で重要な要素の１つといえる．しかし，市販ゲームなどで用いられているAIプレイヤには，ナッシュ均衡に近い戦略に従う強力なAIプレイヤや，ルールベースのAIプレイヤなど様々なアルゴリズムがあるなかで,「読み合い」という駆け引きに着目しているものは少なく，人間プレイヤのように相手の傾向や心理面を考慮した行動選択をしていないことが多いと推測する．そのため，人間プレイヤが対戦しても読み合いの駆け引きになっていると実感することが少なく，ゲームを十分に楽しめないことや，読み合いの駆け引きを伴ったプレイングスキルの向上に繋がらない点が問題として挙げられる．

本研究では不完全情報ゲームの一種である同時手番ゲームにおいて，AIプレイヤに人間プレイヤのような読み合いを演出させる手法の確立に取り組む．人間プレイヤのような読み合いを演出する手法として，人間の読みを模倣した着手や，人間のように読まれる可能性がある癖を持った着手を強調することで実現する．具体的に，提案AIプレイヤは標準着手生成部，人間的な癖を付与するバイアス着手生成機能，搾取戦略実施機能で構成される．標準着手生成部は人間が自然に感じるような着手戦略の生成を行い，バイアス着手生成機能において読まれる可能性がある特定の行動を好むようなバイアスを付与した着手を生成する．また，搾取戦略実施機能では相手プレイヤの癖を見抜き，予測される戦略に対してアドバンテージを得られるような着手戦略を生成する．これらを統合したモデルによって読み合いを演出することで，人間プレイヤが読み合いが起きていることを実感でき，より人間同士の対戦に近い体験ができるようになることを目指す．

本論文は以下の形で構成される．2章では，本研究と対象となる不完全情報ゲームおよび読み合いについて概要を述べる．3章では，対象ゲームや，軸となる手法についての関連研究を紹介する．4章では，本研究において作成した同時手番ゲームおよび，提案手法について説明する．5章では，AIプレイヤの着手生成部，6章では，人間のような癖を与えるバイアス部の実装について述べる．7章では，搾取戦略実施機能のアプローチについて述べる．最後に8章では，本研究の成果や今

後の展望をまとめる.

第2章　対象ゲーム

本研究では，不完全情報ゲームの一種である同時手番ゲームを対象として，AIプレイヤに人間のような読み合いを演出させる手法の確立を目指している．そのため，本章ではナッシュ均衡などの不完全情報ゲームの基礎的な理論について説明する．また，不完全情報ゲームにおいて，人間プレイヤ同士の対戦で生じる「読み」や「読み合い」についても紹介する．そして，4 章で後述する提案ゲームのもとになるポケモン対戦についても概要を説明する．

2.1　不完全情報ゲームについて

2.1.1　展開形ゲーム

将棋やチェスなど多くのボードゲームでは，ゲームは複数の手番で構成されており，プレイヤは各手番で何らかの行動選択を行う．このようなプレイヤの手番と行動をゲーム木によって記述するモデルとして展開形ゲーム [6] がある．将棋や囲碁などの相互手番で完全情報なゲームだけでなく，不完全情報性を持つゲーム，例えばポケモンなどのプレイヤが同時に行動するゲームや，自分より前に行動したプレイヤの選択が観察できないプレイヤが存在するゲームも展開形ゲームに含まれる．展開形ゲームは式 (2.1) に示す以下の 5 つの要素の組で定義される．以降，[6] の記述を再構成して記す．

$$\Gamma = (K, P, p, U, h) \tag{2.1}$$

1. ゲームの木 K
   ゲームの木 K は初期点 0 をもつ有限な有向木で点 (node) と枝 (edge) から構成される．点 x と初期点 0 を結ぶ木の点と枝の系列を，点 x へのパス (path) という．異なる 2 つの点 x と y に対して，点 y が初期点 0 から点 x のパス上にあるとき，x は y の後にあるといい，$x > y$ と表す．点 w が $x > w$ なる点をもたないとき，w を木の頂点といい，その全体を W で表す．頂点以外の点を手番といい，その全体を X で表す．また，ゲームの木 K の 1 つの頂点 w に対して，初期点 0 と頂点を結ぶパスをプレイ w という．手番 x に対して x と x の直後の点を結ぶ 1 本の枝を手番 x における選択肢といい，その全体を $A(x)$ で表す．ゲームのプレイとは，初期点 0 から始まって各手番でプレ

イヤが 1 つの選択肢を選択することによってゲームが進行し，最後に 1 つの頂点に到達してゲームが終了することである．

2. プレイヤ分割 P
プレイヤ分割 $P = [P_0, P_1, \cdots, P_n]$ は，ゲームの木 K の手番の全体 X の 1 つの分割である．0 以外の添字 $i = 1, \cdots, n$ はゲームに参加するプレイヤを表し，

(a) $X = P_0 \cup P_1 \cup \cdots P_n$

(b) 任意の 2 つの P_i と $P_j (i \neq j)$ に対して $P_i \cap P_j = \emptyset$

(c) $P_i \neq \emptyset, i = 1, \cdots, n$

が成り立つ．集合 $P_i (i = 1, \cdots, n)$ はプレイヤ i の手番の全体を表す．プレイヤの集合を $N = \{1, \cdots, n\}$ とする．集合 P_0 はゲームにおける偶然手番の全体である．偶然手番では，プレイヤの意思とは無関係にある偶然メカニズムによって枝の選択が行われる．

3. 偶然手番の確率分布族 p
ゲームのすべての偶然手番 $x \in P_0$ に対して，x での選択肢の集合 $A(x)$ 上の 1 つの確率分布 p_x が定められている．確率分布 p_x が各選択肢 $e \in A(x)$ に付与する確率を $p_x(e)$ とするとき，$\sum_{e \in A(x)} p_x(e) = 1$ および $0 \leq p_x(e) \leq 1$ が成り立つ．偶然手番 $x \in P_0$ の確率分布 p_x の族 $\{p_x | x \in P_0\}$ を p とおく．

4. 情報分割 U
情報分割 $U = [U_0, \cdots, U_n]$ は，プレイヤ分割 $P = [P_0, \cdots, P_n]$ の 1 つの再分割である．すなわち，すべての $i = 0, \cdots, n$ に対して，U_i は集合 P_i の部分集合 u の族であり，

(a) $P_i = \bigcup_{u \in U_i} u$

(b) U_i の任意の異なる 2 つの集合 u と v に対して $u \cap v = \emptyset$

が成り立つ．$U_i (i = 0, \cdots, n)$ をプレイヤ i の情報分割といい，U_i に属する集合をプレイヤ i の情報集合という．プレイヤの情報集合 u は次のような意味をもつ．いま，ゲームのプレイによって u の手番 x に到達したとする．このとき，プレイヤ i は，

(c) 情報集合 u のある手番に到達したことを知る．

(d) しかし，情報集合 u のどの手番に実際に到達したかは知らない．

さらに，次の 2 つの性質が情報集合のもつ意味から必要とされる．

(e) 情報集合 u は同じプレイと 2 回以上交わってはならない．

(f) 情報集合 u に含まれるすべての手番は同じ数の枝をもつ.

情報集合 u の各手番はゲームの初期点 0 からそれぞれ異なるパスをもつが, それらを除いてはプレイヤは実質的に同じ意思決定問題に直面する．プレイヤの情報集合 u において，u の手番がもつすべての枝のうちで実質的に同じ意思決定問題を表す複数の枝を同一の選択肢とみなし，これを情報集合 u における 1 つの選択肢という．情報集合 u における選択肢の全体を $A(u)$ で表す.

5. 利得関数 h
利得関数 h はゲームの木 K の各頂点 $w \in W$ に対して利得ベクトル $h(w) = (h_1(w), \cdots h_n(w))$ を対応させる．ここで，第 i 成分はプレイヤ i の利得を表す.

式 (2.1) で定義した展開形ゲーム Γ においてすべてのプレイヤ $i(1, \cdots, n)$ のすべての情報集合 $u \in U_i$ がただ 1 つの手番からなるとき，ゲーム Γ は完全情報ゲームであり，この条件を満たさないゲームは不完全情報ゲームに含まれる [6]．本章で例示するジャンケンゲームは隠された情報がないように見えるが，展開形ゲームでモデル化した場合には,「片方のプレイヤが着手を決定し，それが何かわからない状態でもう片方のプレイヤが着手を決定する」と捉えることになる．そのため, ジャンケンゲームのような同時手番ゲームも不完全情報ゲームとして扱う.

2.1.2 戦略と期待利得

展開形ゲームにおいて，プレイヤが各手番でどのような行動を選択するかという計画を戦略とよぶ．情報集合 u における選択肢の全体 $A(u)$ 上の 1 つの確率分布 s_{iu} を情報集合 $u \in U_i$ におけるプレイヤ i の局所戦略といい，プレイヤ i の行動戦略 s_i はプレイヤの各情報集合 $u \in U_i$ に対して u における 1 つの局所戦略 s_{iu} を対応させる関数で表現できる．プレイヤ i の行動戦略の全体を S_i で表す．また，実現確率やプレイヤの期待利得について以下のように定義する [6].

- $s = (s_1, \cdots, s_n)$: プレイヤの行動戦略の組.
- $s_{iu}(e)$: プレイヤ i が情報集合 $u \in U_i$ において枝 $e \in A(u)$ を選択する確率.
- $E(w)$: ゲームの木の頂点 $w \in W$ に対して初期点 0 から w へのパス上におけるすべての枝．集合 $E(w)$ は偶然手番の枝の集合 $E_0(w)$ とプレイヤ $i(1, \cdots, n)$ の枝の集合 $E_i(w)$ からなる.
- $c(w)$: 集合 $E_0(w)$ に含まれる偶然手番のすべての枝 e が選択される確率の積. ただし，プレイ w が偶然手番を含まないときは $c(w) = 1$ とおく.
- $p(w \mid s)$: 行動戦略の組 s に従ってゲームがプレイされるとき，頂点 $w \in W$ が到達される確率（頂点 w の実現確率）.
すなわち，$p(w \mid s) = c(w) \prod_{i=i}^{n} \prod_{e \in E_i(w)} s_{iu}(e)$.

- $H_i(s_1, \cdots, s_n)$: $s = (s_1, \cdots, s_n)$ をプレイヤの行動戦略の組とするときの，戦略の組 $s = (s_1, \cdots, s_n)$ に対するプレイヤ i の期待利得.
  すなわち，$H_i(s_1, \cdots, s_n) = \sum_{w \in W} p(w \mid s) h_i(w)$.

2.1.3 ナッシュ均衡

展開形ゲーム $\Gamma = (K, P, p, U, h)$ において，行動戦略の組 $s^* = (s_1^*, \cdots, s_n^*)$ がナッシュ均衡点であるとは，すべてのプレイヤ $i = 1, \cdots, n$ のすべて行動戦略 $s_i \in S_i$ に対して，

$$H_i(s^*) \geq H_i(s_i, s^*_{-i}) \tag{2.2}$$

が成立することである [6]．ただし，s^*_{-i} は，行動戦略の組 $s^* = (s_1^*, \cdots, s_n^*)$ から第 i 成分 s_i^* を除いた行動戦略の組を表す．つまり，他のプレイヤの戦略を固定したうえで，ある一人が期待利得を増やそうとしても，そのような戦略の組がない均衡状態のことをいう．

ナッシュ均衡戦略について具体的に考えるため，勝ちのとき 1，負けたとき-1，引き分けは 0 の利得を得られる零和 2 人ジャンケンゲームを取り上げる．このゲームにおいて，ナッシュ均衡戦略はお互いにグー，チョキ，パーをそれぞれ 1/3 の確率で出す戦略となる．お互いがこの戦略に従った場合，期待利得は 0 となり，片方のプレイヤがナッシュ均衡戦略以外の戦略，例えばグーを 1/2, チョキとパーを 1/4 ずつ出すような戦略に変更しても，戦略を変更したプレイヤの期待利得は 0 より高くならない．このように，もし戦略の組 s^* がプレイされるならば，どのプレイヤも自分だけ戦略を変更する動機をもたず，ゲームのプレイは s^* で均衡することになる．

前述の例のような単純なゲームではナッシュ均衡戦略を求めることは難しくない．ただし，より複雑な不完全情報ゲームでナッシュ均衡戦略を求めるのは容易でなく，後悔 (regret) とよばれる値を最小化させることでナッシュ均衡に近い解を得る手法 [7] などが知られている．人間プレイヤには複雑なゲームにおいてナッシュ均衡戦略を求めることは事実上不可能であるため，人間プレイヤ同士の対戦では着手戦略に偏りが生じることでナッシュ均衡戦略以上に大きな利得を得ることができる戦略が存在する．そのため，人間同士の対戦において，相手プレイヤに対する最適応答を求めるために人間プレイヤは「読み」や「読み合い」による意思決定を行う場合がある．

2.2 不完全情報ゲームにおける読み合い

一部の不完全情報ゲームでは，人間プレイヤは相手の情報や行動を予測する「読み」を用いて意思決定を行う場合がある．例えば，ポーカーにおいてブラフを高い割合で用いてくる相手プレイヤがいる場合，今回もブラフなのではないかと予

測して，ベットなどの強気な選択を行うケースは，行動の「読み」として捉えることができる．また，麻雀において相手プレイヤの捨て牌の情報から，手牌を推測したうえで意思決定を行う場合は，情報の「読み」といえる．

人間同士がこのような「読み」の思考に基づいて意思決定を行う場合や，相手が読んでくることを考慮して意思決定を行う状況を本研究では「読み合い」と呼ぶ．この読み合いについて繰り返しゲーム，例えば連続してプレイする２人ジャンケンゲームの例で考える．これまでの試行から，相手プレイヤがグーを多く選択するプレイヤであると予測される場合，次の手もグーである可能性が高いと考えチョキ以外の手を選択する意思決定は「読み」の思考からも人間的には自然である．対して相手プレイヤも，自分がグーを多く選択する傾向をあえて利用し，相手の読みを逆手に取るような行動を行うことも読み合いによる自然な意思決定といえる．

前述の繰り返しジャンケンゲームに見られるような読み合いの駆け引きは，人間同士の対戦では頻繁に体験するものであり，不完全情報ゲームを楽しむ上で重要な要素の１つといえる．

2.3　ポケモンバトル

図 2.1: ポケモンバトルの対戦画面（出典：「ポケットモンスター ソード・シールド」公式サイト [8]）

「ポケットモンスター」は任天堂株式会社，株式会社クリーチャーズ，株式会社ゲームフリークによって開発・発売されているゲームシリーズであり，ターン制のバトルシステムを基本としたRPGである．ゲームの中心となっているターン制バトルシステム（以下，ポケモンバトル）は，ストーリー中だけでなくプレイヤ同士のオンライン対戦でも人気が高いコンテンツである．シリーズごとにゲー

ムルールやバトルに使用できるポケットモンスター（以下，ポケモン）が異なるが，本稿では最も一般的なルールであるシングルバトルについて説明する．

シングルバトルルールは 2 人のプレイヤが 1 匹ずつ場にポケモンを出していくルールであり，主にパーティーの作成，ポケモンの選出，コマンドバトルの 3 段階で構成されている．まず，パーティーの作成では各プレイヤは事前に数百ある種族のポケモン中から任意の 6 匹で構成されたパーティーを作成する．ポケモンは HP，こうげき，ぼうぎょ，とくこう，とくぼう，すばやさの計 6 つのステータスがあり，ポケモンの種族ごとに大きくステータスの特徴が異なるほか，同じ種族であっても個体ごとにバラつきがある．また，各プレイヤは個体ごとに一定の能力値を任意のステータスに自由に振り分けることができる．ポケモンは持つことができる 4 つのワザ，持ち物と呼ばれる特殊効果を持つアイテム，特殊な能力である特性などもプレイヤが任意に選択することができる．

次に，ポケモンの選出では，開始と共に相手プレイヤのパーティーを構成する 6 匹のポケモンの種族が一定時間だけ公開される．各プレイヤは公開された相手のポケモンの種族を踏まえて，自分のパーティーから 3 匹のポケモンを選択する．

お互いにポケモンの選出が終了すると，実際に選出したポケモンを用いたコマンドバトルが開始する．各プレイヤは 1 ターンに一度，場に出ているポケモンのワザ 4 種，および控えのポケモンと場のポケモンを入れ替える交換 2 種の計 6 種の選択肢から同時に行動を選択する．選択されたコマンドは交換，すばやさの高いポケモン，すばやさの低いポケモンの順に実行され，これらのコマンド選択を繰り返して最終的に相手のポケモンの HP をすべて 0 にしたプレイヤの勝利となる．これらのコマンド選択時，相手ポケモンの種族以外の情報は隠されているため，同時手番以外にも不完全情報性を持つゲームといえる．また，ポケモンバトルにはワザのダメージが増加する急所や追加効果などの確率的な要素があり，不確実性を持つゲームであるともいえる．

ポケモンバトルは前述の通り，隠された情報や相手の選択したコマンドを予測する必要があるため，お互いに相手の情報や行動を予測し，意思決定を行う「読み合い」が発生しやすいゲームと考える．また，ポケモンごとにタイプとよばれる相性があり，場に出ているポケモンの相性関係から相手の行動を読み合うことは主流な戦略の 1 つでもある．しかし，ポケモンバトルはパーティーの作成やポケモンの選出など多段階の要素で構成されていることや，ポケモンの種類やワザ，ステータスなど高すぎる自由度や不確実性の問題から，ゲーム AI 研究において非常に難易度の高い課題の 1 つといえる．そのため，ポケモンバトルを学術研究の対象として扱うことには手法が複雑になることや，実装や評価の難化が懸念される．以上の理由から，人間のような読み合いの演出の第一歩としてポケモンバトルを取り扱うことは不適当と考える．そこで，本研究ではポケモンバトルをもとにして，読み合いが生じるような本質は残しつつも高度または些末な要素は取り除いた同時手番ゲームを作成し，提案ゲーム上で人間のような読み合いを演出する手法の確立に取り組む．提案するポケモンライクゲームの詳細については 4 章

で説明する.

第3章　関連研究

本章では，対戦相手をモデリングする手法，および AI プレイヤに人間らしい振る舞いを実現させる手法について紹介する．また，前章で説明したポケモンバトルに関する AI プレイヤの研究についても紹介する．

3.1　相手モデル

前章で説明した通り，不完全情報ゲームにおいてナッシュ均衡戦略はゲームの解であるものの，複雑なゲームで均衡戦略を求めることは容易ではない．本節では，複雑な不完全情報ゲームや一部のリアルタイムゲームにおいて，高い利得が得られる戦略を獲得するために相手プレイヤをモデリングする手法である相手モデル (opponent model) について紹介する．

Van der Kleij は，ポーカーの 1 種である Texas Hold'em においてプレイスタイルに基づいてプレイヤをクラスタリングし，それぞれのクラスタに対応するモデルを用いることでプレイヤの行動などを予測する手法を提案している [9]．

Sato らは，格闘ゲームにおいて複数のルールベースプレイヤを組み合わせ，相手に応じてプレイヤを切り替える手法を提案している [10]．この手法ではプレイヤの切り替えを，与えたダメージと受けたダメージの差分を報酬とする多腕バンディット問題と捉えることで，相手に有利な行動を取れるよう制御している．

3.2　人間らしい AI プレイヤ

AI プレイヤに強い対戦相手以外に楽しませる，教える，練習相手になるなどの多様な用途を持たせるためには，多様な展開や戦略を実現する，人間プレイヤの実力を模倣するなど，人間プレイヤに近い性能や人間プレイヤのように振る舞うことが求められる．本研究においても AI プレイヤに人間らしい振る舞いを演出させることを目的としており，特に人間同士が対戦する場合に生じる「読み合い」に着目している．AI プレイヤに人間らしい振る舞いを実現させる手法には様々な着目点や実現手法がある．

人間らしい振る舞いを獲得する方法として，代表的な手法に教師あり学習がある．人間プレイヤの着手情報などを教師データとして学習したモデルを用いることで，人間プレイヤに近い振る舞いを行う研究が報告されている．杵渕らは，将

棋において人間が思考する際に考慮する指し手の時系列情報である「流れ」に着目し，教師あり学習による流れを考慮した着手予測器を提案している [11].

また，統計量やルールを用いることで人間らしい振る舞いを表現する方法も有力である．池田らは，囲碁においてモンテカルロ木探索を用いる場合に，プレイアウトの統計量などに基づいて人間プレイヤのような多様な戦略の演出が可能であることを示した [5]．モンテカルロ碁に共通する手順である，地合いを数える，コミを加える，勝敗を定めるという各手順に補正を加えることで，実利派/中央派，悲観派/楽観派，好戦派/厭戦派などの多様な棋風の演出を実現している．

近年では人間らしい振る舞いの自動獲得を目指す研究も行われている．藤井らは，アクションゲームを対象として，人間が生得的に持つ性質から生じる制約や欲求である生物学的制約を導入することで，強化学習の 1 種である Q 学習によって人間らしい振る舞いを持つ NPC を自動獲得できることを示した [12]．具体的には，座標認識の”ゆらぎ”，操作の”遅れ”，”疲れ”や，失敗を繰り返す局面では新奇な行動を，失敗しづらい局面では安定的な行動を行う”訓練と挑戦のバランス”を生物学的制約として，Q 学習の報酬やゲームの観測情報に用いることで，NPC は人間のように不慣れな操作やムラのある動き，安全を重視した行動などを獲得した．

3.3 ポケモンバトルにおける AI プレイヤ

前章で説明したポケモンバトルは現在のゲーム AI 研究においても非常に挑戦的な課題の 1 つであり，ゲーム本来のルールで人間プレイヤ以上の実力を示したという報告は知る限り無い．また，ゲーム本来のルールで取り組むこと自体が難しいため，現在では簡略化したルール上でコマンドバトルを行う AI プレイヤの研究が主に取り組まれている．

Ihara らは，ポケモンの種族を限定したルールにおいて ISMCTS(Information set Monte Carlo tree search) と呼ばれる手法を適用し，有効性を示した [13]．ISMCTS は，通常は完全情報の決定論的ゲームに用いられるモンテカルロ木探索を，情報集合を用いたゲーム木によって隠れた情報や不確実性を扱えるようにすることで不完全情報ゲームに適用する手法である [14].

また，ポケモンバトルには公式のシミュレータが存在しないため，API が提供されている Pokemon showdown[15] に存在するルール上での成績がベンチマークになっていることも多く，過去には showdown 上での AI コンペティションも開催されている [16]．Huang らは，showdown 上の 7thGen RandomBattle とよばれる対戦に用いられるポケモンがすべてランダムに決まるルールにおいて，自己対戦によってニューラルネットワークの学習を行う Actor-critic RL 法で構築されたエージェントが Glicko-1 レーティングで 1677 に到達し，ランダムで選ばれる対戦相手に 72%の確率で勝利できる実力を示した [17].

第4章　提案手法

2.1.3 項で説明したように，不完全情報ゲームにはナッシュ均衡戦略と呼ばれるゲームの解があるが，複雑なゲームでこの戦略を求めることは容易ではなく，人間プレイヤには事実上不可能である．そこで 2.2 節でも説明したように人間プレイヤは一部の不完全情報ゲームにおいて，お互いに相手の情報や行動などを予測して意思決定を行う「読み合い」を用いる場合がある．本研究では同時手番ゲームを対象として，AI プレイヤに人間のような読み合いを演出させる手法の確立に取り組む．

人間プレイヤのような読み合いを演出する手法として，本研究では行動の読みに特に着目し，人間の読みを模倣した着手や，人間のように読まれる可能性がある癖を持った着手を強調することで実現する．

本章では，本研究における提案手法として，対象とする同時手番ゲームおよび読み合いを演出する提案プレイヤの詳細について述べる．

4.1　簡易版ポケモンバトル

簡易版ポケモンバトルとは，2.3 で紹介したポケモンバトルをもとにして，作成した同時手番ゲームである．簡易版ポケモンバトルは，タイプによる相性関係や同時着手性などの読み合いが生じるような本質は残しつつ，高度または些末な要素を取り除いたゲームである．本節では，ポケモンバトルとの相違点や，簡易版ポケモンバトルの詳細なルールについて説明する．

4.1.1　ポケモンバトルとの相違点

ポケモンバトルは 2.3 節でも紹介したように，隠された情報や相手の選択したコマンドを予測する必要がある．そのため，お互いに相手の情報や行動を予測し，意思決定を行う読み合いが発生しやすいゲームといえる．また，ポケモンの相性関係から相手の行動を読み合う戦略は，人間同士の対戦においては主流な戦略の 1 つであり，本研究でもこの点に特に着目している．しかし，ポケモンバトルには本研究で行う読み合いの演出に不適当な要素が多く存在していることも確かである．そのため，本研究で提案する簡易版ポケモンバトルでは，同時手番によるコマンド選択と，タイプ相性によって行動を読む戦略という本質を残すことで，読

み合いが発生しやすいゲーム性を維持しつつ，ゲームを難化させている様々な要素を取り除く，または制限するといった変更を行った．表 4.1 にゲームの相違点を示す．

まず，本研究で着目している行動の読み合いはコマンドバトルで生じるため，簡易版ポケモンバトルではパーティーの作成，ポケモンの選出は行わないものとする．つまり，3 匹のポケモンがお互いに選出された状態からゲームが開始する．

ポケモンの種族やワザの総数，アイテムや特性，ステータスの種類などの自由度が高い要素はゲームの戦略を多様にする側面があるが，行動の読みを行う上では関係性が薄いため，自由度を下げる，要素を取り除くなどの変更を行った．

また，ポケモンバトルがもつ不完全情報性には，相手の選択した行動が分からない同時手番性と，相手のステータスやワザなどの情報が分からないという 2 つの性質がある．後者の性質は情報の読みを行う上で重要な要素であるが，本研究では読み合い演出の第一歩として，行動の読みに対象を限定するため，同時手番性のみを残し，隠れた情報が存在しないゲーム性に変更した．ダメージが増加する急所や，追加効果などポケモンバトルがもつ確率的な要素についても，行動の読みへの影響が薄いこと，シミュレーションの再現性などの理由から，後述する初期局面の決定以外に不確実性がないゲーム性に変更した．不完全情報性，不確実性がないゲーム性に調整したことで，簡易版ポケモンバトルにおいて読み合いが発生するかの懸念もあったが，実際に試験し読み合いが起こりうることを確認している．

要素	ポケモンバトル	簡易版ポケモンバトル
パーティーの作成	あり	なし
ポケモンの選出	あり	なし
ポケモンの種族	1000 種以上	6 種
ポケモンが使えるワザ	約 100 種から 4 種選択	3 種固定
ステータスの種類	6 種	2 種（HP，すばやさ）
特性，アイテム	あり	なし
隠された情報（不完全情報性）	あり	なし（同時手番のみ）
確率的な要素（不確実性）	あり	なし

表 4.1: ポケモンバトルとの主な相違点

4.1.2 ルール

以下にポケモンライクゲームの詳細なルールについて示す．

1. 概要
   簡易版ポケモンバトルは，お互いのプレイヤが場に 1 匹ずつポケモンを出し，1 ターンに 1 度コマンドを選択するシングルバトル形式である．

2. 用いることができるポケモン
   簡易版ポケモンバトルで用いるポケモンは以下の要素で構成されている．

   (a) ステータス：
   ポケモンは HP，すばやさの 2 種類のステータスを持つ．HP の最大値は 5, すばやさの最大値は 10 となっている．ポケモンライクゲームに慣れていない被験者でも先読みがしやすいこと，次節以降で後述する AI プレイヤがナッシュ均衡戦略を計算できるようにするなどの理由から，簡易版ポケモンバトルでは HP の最大値を 5 に設定している．

   (b) タイプ：
   ポケモンはほのお，みず，くさ，でんきの 4 種類のタイプのいずれかをもつ．各タイプの相性関係は表 4.2 を参照．

   (c) ワザ：
   各ポケモンは 3 つのワザを持つ．ワザにはそれぞれタイプと威力が存在する．

3. ゲーム内で可能な行動
   ゲーム内で選択可能なコマンドは以下の 2 種類である．

   (a) ワザ (攻撃)：
   ワザは相手ポケモンに攻撃する行動であり，タイプと威力の 2 種類のパラメータがある．威力は基本 2 ダメージだが，攻撃を受ける（ぼうぎょ）ポケモンのタイプと，ワザのタイプの相性関係によって威力が変化する．基本ダメージに表 4.2 の倍率をかけた値が，最終的に相手に与えるダメージになる．また，補正倍率が 1.5 のとき，さらに攻撃するポケモンとワザのタイプが一致した場合には倍率が 2.0 となる．

   (b) 交換：
   交換は場に出ているポケモンと，控えの HP が 0 でないポケモンを入れ替える行動である．交換を行ったターンは，新たに場に出たポケモンはそれ以上行動選択ができない．

4. ゲームの進行
   ゲームは以下に示す手順で進行される．

   (a) パーティーの決定：
   ゲームが開始するとプレイヤには表 4.3, 表 4.4 に示す 2 種類のパーティーのうち，どちらかのパーティーがランダムに与えられる．パーティーは 3 体のポケモンで構成されている．プレイヤは相手のポケモンの種族と情報（HP，すばやさ，ワザの種類）を確認することができる．

(b) 場に出るポケモンの決定：
ゲーム開始時にお互いのパーティーから，それぞれ 1 匹ずつランダムに選ばれたポケモンが場に出される．ポケモンの選出を行わないため，このような仕様になっており，ポケモンの組み合わせによっては初期局面の時点で多少の形勢の偏りがある．

(c) コマンド選択
各プレイヤは 1 ターンに一度，自分の場に出ているポケモンの持つワザ 3 種と，控えのポケモンへの交換 2 種の最大 5 種類の選択肢から 1 つを同時に選択する．選択したコマンドの実行は，交換，すばやさの高いポケモンのワザ，すばやさの低いポケモンのワザの順で処理される[1]．また，選択したコマンドが実行される前に自分の場に出ているポケモンの HP が 0 になった場合には，選択していた行動はキャンセルされ，まだ HP が 0 でないポケモンがいる場合，それらのポケモンからプレイヤが選択して場に出す．

(d) 勝敗の判定
相手プレイヤのパーティーのポケモンすべての HP を 0 にしたプレイヤの勝利となる．また，20 ターン目をむかえた時点でゲームの勝敗がついていない場合は，以下に示す順で勝敗を判定する．

i. 20 ターン目時点で，残りポケモン数（HP が 0 でないポケモン）が多いプレイヤの勝利となる．
ii. 残りポケモン数が同じ場合，残りポケモンの HP 総数が大きいプレイヤの勝利となる．
iii. HP 総数が同じ場合引き分けとなる．

こうげき ＼ ぼうぎょ	ほのお	くさ	でんき	みず
ほのお	0.5	1.5	1.0	0.5
くさ	0.5	0.5	1.5	1.0
でんき	1.0	0.5	0.5	1.5
みず	1.5	1.0	0.5	0.5

表 4.2: 各タイプの相性関係と，ダメージの補正倍率

[1]簡易版ポケモンバトルではすばやさが同じポケモンがいないよう設定した．

名称	ポケモン A	ポケモン B	ポケモン C
タイプ	くさ	みず	でんき
HP	5	5	5
すばやさ	5	4	10
覚えるワザ	くさ，ほのお，でんき	みず，ほのお，でんき	でんき，くさ，みず

表 4.3: パーティー 1 の各パラメータ

名称	ポケモン D	ポケモン E	ポケモン F
タイプ	くさ	ほのお	でんき
HP	5	4	4
すばやさ	6	7	8
覚えるワザ	くさ，ほのお，でんき	みず，ほのお，でんき	でんき，くさ，みず

表 4.4: パーティー 2 の各パラメータ

4.2 読み合い演出の全体像

不完全情報ゲームにおいて人間プレイヤが読み合いが起きていることを感じ取る場面は実に様々ではあるが，代表的な状況としては「相手の情報や行動を読んでアドバンテージを得ることができたとき」,「相手に自身の情報や行動を読まれたと思ったとき」の 2 つが挙げられる．そのため，対戦中にこのような体験をすることができれば人間プレイヤは読み合いが起きていることを実感することができると考え，人間のような読み合いを演出する手法を検討する．

本研究では対戦中に前述のような人間が読み合いを実感できる状況をつくることで読み合いを演出する手法を考案する．まず，読み合い演出を行う前提として，人間プレイヤが自然に感じる着手の生成を提案する．そのうえで，対戦中に読み合いを実感できる状況を作り出す方法として，人間の読みを模倣した着手や人間のように読まれる可能性がある癖を持った着手を強調することを提案する．以下，個別具体的に説明する．

1 つ目の提案は，人間のような読み合いを演出する上で前提となる要素として，人間が自然に感じる着手戦略を生成することである．多くの人間プレイヤはゲームを楽しむという観点においては相手に自分が意図を理解できない手や，悪手に感じるような手を選択されることを嫌う．また，着手の一貫性がなかったり，意図が理解できない戦略に対して人間プレイヤは「読み」による意思決定を円滑に行うことができず，人間のような読み合いを演出するうえで弊害となる．そのため，本研究では後述する 2 つの提案手法，特に癖や傾向をバイアスとして付与する元となる戦略として，人間が不自然に感じない着手戦略の生成について取り組む．

2つ目の提案は，AIプレイヤが人間の「読み」を模倣することで，相手に付け入るような戦略を用いることである．人間プレイヤは殆どの場合，着手に付け込まれる癖や傾向がある．そこで，相手プレイヤに付け入るような戦略を用いることで，人間プレイヤに「相手に自分の行動を読まれた」と思わせることができると考える．このようにAIプレイヤ側が読みを模倣することで，擬似的にお互いが「読み」の思考に基づいて意思決定を行う状態を表現する．

読みの模倣には，3.1節で紹介した相手モデルを用いる．相手モデルは複雑な不完全情報ゲームなどで，相手をモデリングする手法である．相手モデリングによって相手の行動を予測することで，予測される行動に対してアドバンテージを得る搾取戦略を用いて「読み」を表現する．詳細なアプローチは4.2.1にて述べる．

3つ目の提案としては，AIプレイヤが人間のように付け込まれる癖や傾向を持った戦略を用いることである．人間のように読まれる可能性がある傾向や癖を持った戦略を用いて，隙をつくることで人間プレイヤにあえて付け入らせる．このような体験によって「相手の情報や行動を読んでアドバンテージを得ることができた」と思わせることができると考える．

このような手法を実現するために，付け込まれる癖や傾向を持った戦略がどのようなものかについて説明する．人間プレイヤは，ゲームにおいて何らかの基本とする考えや，好みを持っており，これらの傾向はゲーム中に一貫して現れる場合が多い．また，人間がある意図を持って選択した行動にはナッシュ均衡戦略と比較して何らかの偏りがあり付け込まれる隙が生じる．人間同士の対戦では，お互いにこの偏りから生じる隙を狙って読み合いになることも多い．そのため，本研究では前述のような着手の癖や傾向をいくつかのバイアスとして定義し，ゲーム中に一貫してバイアスを付与した着手を行う[2]．

4.2.1 提案AIプレイヤの概要

前述した提案手法を実際に実現するAIプレイヤの構成を図4.1に示す．提案AIプレイヤは，標準着手生成部，人間的な癖を付与するバイアス着手生成機能，搾取戦略実施機能で構成される．以下，それぞれの機能について概要を説明する．

1. 標準着手生成部

   標準着手生成部では，4.1節で説明した簡易版ポケモンバトルを対象として，人間プレイヤが不自然に感じないような合理性のある戦略を生成する．この戦略は，ナッシュ均衡戦略をベースにした混合戦略（確率分布）であるが，人間プレイヤが不自然に感じない程度に実力を弱めに調整している．標準着手生成部の詳細な実装内容については5章で述べる．

[2]人間プレイヤは相手に読まれたと感じ戦略を変更する場合もあるが，読み合いの応用例であるため，読み合い演出手法の確立を目指す本研究では扱わないこととする．

2. バイアス着手生成機能

   バイアス着手生成機能では，標準着手生成部で生成した戦略をもとにして，人間がもつような癖や傾向をバイアスとして付与した戦略を生成する．このバイアスは，AI プレイヤの選択時に決められ，数試合一貫して用いる．バイアス着手生成機能の詳細な実装内容については 6 章で述べる．

3. 搾取戦略実施機能

   搾取戦略実施機能は，現在または過去のゲーム中における相手プレイヤの行動から，相手プレイヤをモデリングし，次の行動を予測する相手モデルを持つ．相手モデルは，「交換を嫌う」，「相性の良い攻撃を好む」など既定の着手モデルを複数個持っており，相手のこれまでの行動に最も近い着手モデルが示す戦略を次の相手の着手とみなす．そして，予測される着手に対してアドバンテージを得られるような搾取戦略を生成する．ただし，搾取戦略は毎ターン行うわけではなく，モデルの予測に用いる尤度の閾値などに応じて実施する．搾取戦略実施機能の詳細なアプローチについては 7 章で述べる．

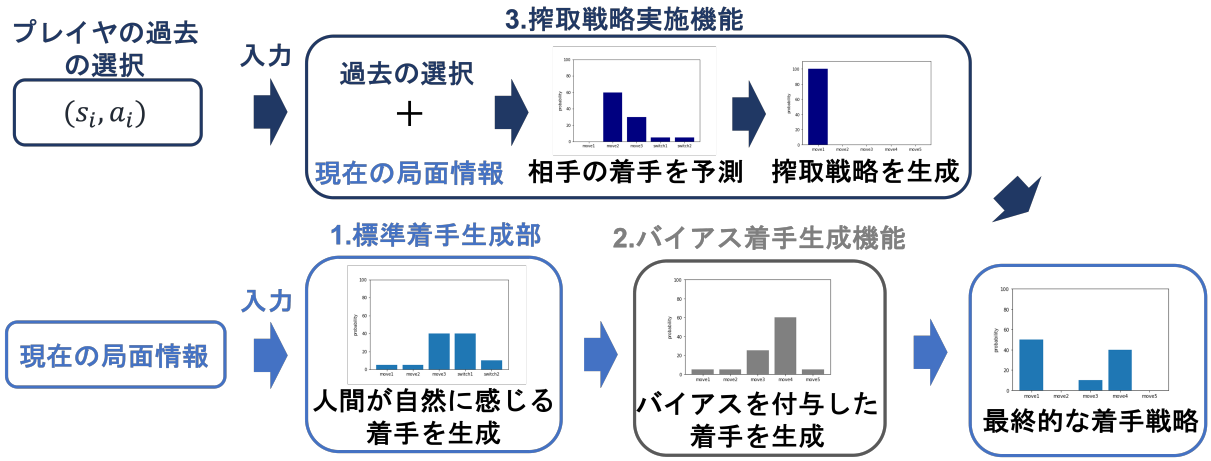

図 4.1: 読み合いを演出する AI プレイヤの構成

第5章　標準着手生成手法

本章では簡易版ポケモンバトルを対象として，人間プレイヤが自然に感じるような標準的な着手戦略を生成する手法について述べる．5.1 節では完全情報ゲームなど他のゲームでよく使われる手法であるモンテカルロ法によるプレイヤを紹介する．この手法は予備実験によって不十分であると感じたため，5.2 節で述べる後退解析によるナッシュ均衡戦略をベースとして，5.3 節で述べる δ-ナッシュプレイヤを提案する．

5.1　原始モンテカルロプレイヤ

不完全情報ゲームは，プレイヤが情報の一部を正確に把握できないゲームであるため，完全情報ゲームで用いられるような $\alpha\beta$ 探索などの従来のゲーム木探索で着手を決定する手法は効果的ではない．そのため，人間プレイヤが自然に感じるよう，比較的合理性の高い着手を生成するため，モンテカルロ法をまず検討した．モンテカルロ法は，ランダムシミュレーションによるプレイアウトを行うことで状態の良さを推定し，それに基づいて着手を決定する手法である．以下に本研究で行ったモンテカルロ法の手順について示す．

- ある状態 s_i において次の行動を決める場合，自身の行動の選択肢 5 通り $(a_1, \cdots, a_5)$ と，相手の行動の選択肢 5 通り $(b_1, \cdots, b_5)$ の組み合わせ計 25 通りについてランダムシミュレーションを行う．

- 25 通りの組み合わせについて，それぞれ 20 回ずつ勝敗がつくまでシミュレーションを行う．現在の状態 s で行動の組 (a_1, b_1) についてプレイアウトを行う場合，状態 s では (a_1, b_1) を着手し，次状態以降ではそれぞれの行動をランダムで決定しシミュレーションを行う．

- それぞれの組み合わせについてプレイアウトの勝ち数を用いて評価する．(a_i, b_j) について規定回数のプレイアウトを行ったときの勝ち数を e_{ij} とする．各着手の評価には，行動の組み合わせに対する評価 $(e_{11}, e_{12}, e_{13}, \cdots, e_{55})$ までの値を用いる．モンテカルロ法の一般的な作法に従い，着手 a_i の評価は $\sum_j e_{ij}$ とする．

- 各着手の評価をもとに，現在の状態での行動を決定する．また，行動を決定する際に，以下の 2 種類の方策を用いた．

  1. greedy: 各着手の中で勝率最大となる行動を選択．
  2. soft-max: 各着手の評価値に応じて確率的に行動を選択．

作成した原始モンテカルロプレイヤと，ランダムプレイヤを 500 回対戦させ，その勝率によって性能を評価した．対戦成績を表 5.1 に示す．原始モンテカルロプレイヤはランダムプレイヤに対して十分な勝率を示した．

続いて，人間が対戦した場合に着手が自然かどうかについて調べた．複数回のテストプレイによって調査を行った結果，以下のような問題点が判明した．次節ではこれらの課題点を修正する手法について述べる．

- ランダムなプレイアウト結果によって，一部のノードが過剰に評価されてしまうことで，人間から見ると悪手にみえる行動選択を行ってしまう場合がある．
- 仮にすべてのノードについて正確な勝率が求まったとしても，5 つの行動の勝ち数を加算しているため，相手が取らないであろう行動のプレイアウト結果を高く評価してしまい誤った行動選択をしてしまう．

方策	ランダムプレイヤに対する勝率
greedy	0.976
soft-max	0.972

表 5.1: 原始モンテカルロプレイヤのランダムプレイヤに対する勝率

5.2 後退解析によるナッシュ均衡戦略

前節で用いた原始モンテカルロプレイヤは，ランダムプレイヤに高い勝率を示すなど評価できる点もあったものの，着手の自然さという意味では多くの課題が残った．本研究ではこれらの問題点を修正するため，ナッシュ均衡戦略を求めることにする．ナッシュ均衡戦略はゲームの解であり，不完全情報ゲームにおいては最も合理的な戦略であるため，人間が「ひどい手だ」という意味で不自然に感じる着手を行う可能性を減らすことが期待できる．以下，本節では簡易版ポケモンバトルにおいてナッシュ均衡戦略を求める手法について述べる．

5×5 の利得行列で表現できるゲームのナッシュ均衡戦略は，各状態の勝率を求めることができれば線形計画法によって求めることができる．本研究ではナッシュ均衡戦略を求めるにあたり，簡易版ポケモンバトルの状態数がそれほど多くない

ことや，同時手番以外の不完全情報性や不確実性がないことを利用する．ゲームの全状態を数え，それらの状態について最終ターンから後退解析によって状態を評価することでナッシュ均衡戦略を計算した．以下に後退解析の手順を示す．

1. 状態数の列挙

   まず，ゲームを解析する上で，各ターンにおけるゲームの全状態を列挙する．簡易版ポケモンバトルの全状態数は実際には到達しない状態なども含めると $6^6 \times 3^2 \times 20 = 8,398,080$ である．

2. 最終ターンにおける状態の評価

   簡易版ポケモンバトルではゲームが 20 ターン目を迎えると，必ず勝敗がつく仕様になっている．そのため，20 ターン目時点での状態について，勝ち：勝率 1.0，負け：勝率 0，引き分け：勝率 0.5 として，評価を行う．

3. 最終ターン以外の状態の評価

   状態 s で行動 (a_i, b_j) を着手した場合の勝率については，遷移先の状態の評価を確認すれば確定的に求めることができる．そのため，状態 s における各行動の組に対する勝率を最大 5×5 の利得行列で表すことができる．この利得行列から，線形計画法によってナッシュ均衡戦略を求める．この計算には gambit[18] を用いた．
   状態 s におけるナッシュ均衡戦略を $(p_{a1}, \cdots, p_{a5}), (p_{b1}, \cdots, p_{b5})$，行動 (a_i, b_j) によって遷移した次状態の勝率を w_{ij} とすると状態 s における真の勝率は $\sum_{i,j}(w_{ij} \cdot p_{ai} \cdot p_{bj})$ となる．

4. 各ターンにおける全状態の評価

   後退解析では，前述の評価方法を用いて最終 20 ターンから 19 ターン目と順に各ターンにおけるすべての状態の評価を行う．双方がナッシュ均衡戦略を用いた場合を真の勝率として，全状態の勝率を計算することができる．

5. ナッシュ均衡戦略の計算

   ゲームの全状態の勝率を計算することができれば，任意ターンの任意の状態において，利得行列を作成し，ナッシュ均衡戦略を計算できる[1]．

後退解析によってゲームの全状態について勝率による評価を行うことで，ナッシュ均衡戦略に従う AI プレイヤ（以下，ナッシュプレイヤ）を作成した．

ナッシュプレイヤの強さを確認する上で，前節で作成したモンテカルロプレイヤ（soft-max）およびランダムプレイヤと 500 回の対戦を行った．対戦成績を表 5.2 に示す．ナッシュプレイヤは，モンテカルロプレイヤにも大きく勝ち越してい

[1]実際には step3,4 ですべての状態のナッシュ均衡戦略は求まっているが，本研究ではコンピュータのメモリ (RAM) を節約するためこのような実装とした．

る．本質的にはナッシュ均衡戦略は「負けない戦略・付け込まれない戦略」であるが，一方でモンテカルロプレイヤには状態の過大評価・過小評価による不適切な着手があるため，このような結果になったと考える．

さらに，ナッシュプレイヤの着手について人間プレイヤが自然に感じるかどうかについて評価を行った．著者が複数回のテストプレイを行った結果，ナッシュプレイヤの着手に不自然に感じる手は少なく，自然な着手の作成を行うことが可能になった．また，ナッシュプレイヤの行う確率的な着手によって，一定程度相手プレイヤに読まれたと感じるような体験を得ることができた．しかし，ナッシュプレイヤは 2.1.3 で紹介したようにゲームの解の 1 つであり，相手に付け入られない戦略であることから人間プレイヤの対戦相手としては強すぎるという問題が生じる．一方的に相手の癖や傾向を読むことができず，負ける展開は読み合い演出の前提からも好ましくない．そこで次節において，ナッシュプレイヤの着手を調整する手法について述べる．

対ランダムプレイヤ	対モンテカルロプレイヤ
0.968	0.720

表 5.2: ナッシュプレイヤの各プレイヤに対する勝率

5.3 δ-ナッシュプレイヤ

ナッシュ均衡戦略に従うナッシュプレイヤは明らかに悪い着手や付け込まれるような着手をしないため，人間プレイヤにも自然な相手であると感じさせることができる．しかし平均的な人間プレイヤの対戦相手としては強すぎるという問題点がある．そこで本節では標準的な強さの着手を生成するため，ナッシュプレイヤの強さを調整する手法について述べる．

ナッシュ均衡戦略を自然な着手に感じやすい理由として，相手に付け入られない性質が挙げられる．しかし，当然ながら相手に付け入られない戦略に対しては人間プレイヤは読みによるアドバンテージを得る戦略を考えることができないため，読み合いの演出に適しているとは言い難い．そこで，ナッシュ均衡戦略の性質をある程度維持しながら，強さを人間に近づける方法にとしてナッシュ均衡戦略を求める利得行列にノイズを付与する手法を提案する．人間プレイヤは，5×5 通りの行動に対して，正確に利得を評価することは難しく，真の勝率が例えば 60% などの局面を人間は 50%などに間違えてしまうことは多々ある．このような人間の不正確さを導入した誤認知利得行列からナッシュ均衡戦略を求めることで，ナッシュ均衡戦略の性質をある程度維持しながら，強さを調整することができる．

本研究では，ナッシュプレイヤの強さを調整する手法として，以下に示す δ-ナッシュ均衡戦略を提案する．δ-ナッシュ均衡戦略は，ノイズを付与した誤認知利得行

列からナッシュ均衡戦略を求める手法である．δ-ナッシュ均衡戦略を求める手順について以下に示す．

- i 行 j 列の利得行列の要素 p_{ij} について，それぞれ一様分布 $(-\delta, \delta)$ などに従うノイズを付与する．ノイズを付与した利得行列から 5.2 節の手順 3 で用いた線形計画法による方法でナッシュ均衡戦略 N を計算する．
- 上記の試行を n 回行う．k 試行目のナッシュ均衡戦略を N_k として，δ-ナッシュ均衡戦略 $N' = \sum_k N_k/n$ を求める．

δ-ナッシュ均衡戦略は，付与するノイズのサイズや，ノイズの分布などの性質を変更することで，強さの調整や着手確率を変更することが可能である．次節では，δ-ナッシュ均衡戦略に従うエージェントである δ-ナッシュプレイヤについて評価実験を行う．

5.4 着手生成手法の評価実験

5.4.1 確率的な着手に関する評価

後退解析によって求めた勝率をそのまま用いたナッシュ均衡戦略は，人間ならば選択してもおかしくない行動をまったくとらないことがある．例えば，自分 2 行動，相手 1 行動に単純化した例で，真の勝率が 50%の手と 45%の手があった場合，後者は絶対に選択されない．しかし，人間にとっては真の勝率を正確に計算することは難しいことが多いため，実際には人間プレイヤが後者を選択してもおかしくない．

一方で δ-ナッシュプレイヤは，より人間に近い勝率評価を行うため，このような人間であれば有力に見える着手の確率を増加させることが期待できる．本項では，このような着手確率のバラつきについて検証を行った．

確率的な着手戦略を評価するため，簡易版ポケモンバトルの初期局面を用いる．簡易版ポケモンバトルは，プレイヤ間のパーティー入れ替えを加味しなければ，9 つの初期局面が存在する．これらの初期局面[2]においてノイズ δ を変化させた場合の δ-ナッシュ均衡戦略を確認することで，着手戦略のバラつきを評価した．

$$\begin{bmatrix} 0.42 & 0.00 & 0.42 & 0.08 & 0.54 \\ 1.00 & 0.59 & 1.00 & 0.08 & 0.38 \\ 0.42 & 0.00 & 0.42 & 0.22 & 0.25 \\ 0.00 & 0.09 & 0.00 & 0.16 & 0.02 \\ 0.00 & 0.00 & 0.01 & 0.16 & 0.17 \end{bmatrix} \tag{5.1}$$

[2] この実験ではパーティー 2 の設定は表 4.4 のポケモン E, ポケモン F の HP を 5 にした設定を用いた．

式 (5.1) は初期局面 1 における真の勝率による利得行列である．この利得行列から求めたナッシュプレイヤ，一様分布 $(-\delta, \delta)$ でノイズを付与した $n = 10$ の δ-ナッシュプレイヤの着手戦略の組 $(a_1, \cdots, a_5), (b_1, \cdots, b_5)$ を表 5.3 に示す．

δ	着手戦略の組み合わせ
0（ナッシュ）	(0.00, 0.30, 0.70, 0.00, 0.00), (0.00, 0.19, 0.00, 0.81, 0.00)
0.05	(0.00, 0.27, 0.52, 0.11, 0.09), (0.00, 0.18, 0.00, 0.80, 0.02)
0.1	(0.01, 0.25, 0.39, 0.13, 0.22), (0.00, 0.18, 0.00, 0.76, 0.06)
0.2	(0.08, 0.43, 0.35, 0.08, 0.06), (0.00, 0.19, 0.00, 0.74, 0.07)

表 5.3: 初期局面における各プレイヤの着手戦略

表 5.3 のナッシュ均衡戦略は，a, b の戦略ともに 2 つの行動のみ確率的に着手する戦略になっている．一方で δ-ナッシュ均衡の着手戦略は $\delta = 0.05$ でも 3 つ以上の行動を選択する可能性がある．人間的には有望にみえる行動は δ が低い場合には選択確率が低いものの，δ が増加するにつれ，選択確率も増加している．しかし，δ が大きくなることで人間から悪手にみえる行動についても低確率で着手する可能性があるため，この点は課題である．検証では例示した局面以外のすべての初期局面についても同様の傾向を確認している．

以上の検証結果から δ-ナッシュプレイヤは，人間プレイヤの誤認知を模倣することで，少し損する手を含んだ人間らしい多様な着手を行うことができると考える．今後はこの δ-ナッシュプレイヤを 4.2.1 項で述べた標準着手生成部として用いる．

5.4.2 プレイヤの強さと着手の自然さに関する評価

δ-ナッシュ均衡戦略に従う δ-ナッシュプレイヤについて，ノイズのサイズを変化させることで勝率を調整できるかについて評価実験を行った．また，付与するノイズの性質などを変更させることで勝率や着手戦略にどのような影響があるかについても検証を行った．

以下に示す 4 種類のノイズを付与した $n = 10$ の δ-ナッシュプレイヤについて，それぞれナッシュプレイヤと 500 回対戦させ勝率を確認した．表 5.4 に対戦成績を示す[3]．

- 一様分布 $(-\delta, \delta)$
- [4]正規分布 $(0, \sqrt{\frac{1}{12}((-\delta) - \delta)^2})$
- 一様分布 $(\pm\delta(1 + 4p(1 - p)))$

[3]比較実験に用いている seed 値の影響で短期的にナッシュプレイヤに 50%以上の勝率を示すことがある．

[4]正規分布の分散は δ を固定した際の比較のために一様分布と同様の値を用いた．

- 正規分布 $(0, \delta(1+4p(1-p)))$

ノイズδ	0	0.05	0.1	0.2
ナッシュ$(\delta=0)$	0.528	-	-	-
一様分布 $(-\delta, \delta)$	-	0.526	0.500	0.458
一様分布 $(\pm\delta(1+4p(1-p)))$	-	0.516	0.460	0.378
正規分布 $(0, \sqrt{\frac{1}{12}((-\delta)-\delta)^2})$	-	0.534	0.512	0.434
正規分布 $(0, \delta(1+4p(1-p)))$	-	0.456	0.400	0.316

表 5.4: ノイズの分布とパラメータごとの δ-ナッシュプレイヤの勝率比較

ノイズサイズによる強さの調整

ナッシュプレイヤとの対戦成績では，一様分布，正規分布ともにノイズが大きくなるにつれ，勝率が下がっていることが確認できる．δ-ナッシュプレイヤは，ナッシュプレイヤと比較して強さを抑えることに成功しており，プレイヤの強さという観点では人間が自然に感じる着手の条件を満たしている．

ノイズを付与する分布の比較

一様分布によってノイズを与えたプレイヤと，正規分布によってノイズを与えたプレイヤの勝率をそれぞれ比較すると，一部例外はあるものの正規分布によってノイズを与えたプレイヤの勝率の方が低いことがわかる．また，分散を揃えた分布同士を比較しても，正規分布の方が低い勝率を示している．

ノイズ δ を変動させた際の比較

一様分布 $(\pm\delta(1+4p(1-p)))$ および，正規分布 $(0, \delta(1+4p(1-p)))$ は i 行 j 列の利得行列の要素（勝率）p によって，ノイズのサイズが変動するモデル（以下，δ 可変モデル）である．

δ 可変モデルは，人間プレイヤの場合，真の勝率が極端なときよりも，互角に近いほうが，真の勝率を見誤る可能性が高いことを反映させたモデルである．例えば，真の勝率が 0% のときに，勝率が 10% だと誤認することよりも，真の勝率が 50% のときに勝率 60% と誤認することの方が多いのは人間プレイヤであれば自然といえる．

これらのモデルについては，δ 固定モデルよりも勝率が下がる傾向が確認できる．とくに δ が大きい場合には，分散も大きくなるため正規分布では傾向が特に顕著である．δ 可変モデルはノイズが大きい場合には，弱くなりすぎる問題があるため，ノイズが小さい場合には，より人間らしい勝率を求める上で有効だと考える．

着手の自然さ

複数回のテストプレイを行うことでプレイヤの着手の自然さについて確認を行い，以下のように評価した．

- ノイズ δ が大きいとき，人間が悪手に感じるような行動を一定程度評価し，低確率で着手することがある．この場合，δ 可変モデルのほうがノイズが大きくなり不自然に感じやすいため，δ 固定モデルの方が適している．
- ノイズ δ が小さい場合には，悪手に対する補正が低いため自然に感じる場合が多い．この場合には，より人間らしい勝率である δ 可変モデルが適している．

有望な設定

強さと着手の自然さに関する評価から有望な設定について考察する．まず，より人間に近い誤認知を表現できる点で δ 可変モデルは有望なモデルといえる．強さについては，$\delta = 0.05$ のモデルはナッシュプレイヤとの差が小さいため，ノイズ δ は 0.1 から 0.2 が適切な範囲だと考える．しかし，δ 可変モデルは，ノイズが大きいと分散が大きくなるため，この影響を一定程度抑えるため一様分布を用いることが望ましい．

評価をまとめると，一様分布 $(\pm\delta(1+4p(1-p)))$，δ =0.1 から 0.2 が読み合いを演出する前提となる標準着手として有望だと考える．ただし，人間プレイヤの実力に応じて，δ-ナッシュプレイヤの強さを調整する必要性もあるため，有望な設定はこの限りではない．

第6章　バイアス着手生成機能

本章では，AI プレイヤに人間のように読まれる可能性がある癖や傾向を持った戦略を着手させる手法について説明する．6.1 節ではバイアスを付与する手法について説明する．6.2 節では付与する 2 種類のバイアスと，その結果について述べる．6.3 節では，バイアスの補正式やノイズのパラメータが与える影響について検証を行う．

6.1　バイアスの付与方法

4.2.1 項で説明したように，バイアス着手生成機能では，標準着手生成部で生成した戦略と同様の考え方を行いつつも，人間がもつような癖や傾向をバイアスとして付与した戦略を生成する．癖や傾向となるバイアスを付与する方法として，標準着手生成部が生成した着手確率を直接操作する方法がまず考えられる．しかし，直接着手確率を操作すると，やり方や程度によってはナッシュ均衡戦略の性質から大きく外れることになり，人間プレイヤが違和感を感じる可能性がある．

そこで，バイアス着手生成機能では，5.3 節で述べた δ-ナッシュ均衡戦略を求める際の利得行列の要素にバイアスを付与することを提案する．ノイズを付与した利得 p_δ に，さらにバイアスによる補正をかけた誤認知利得行列を作成する．この利得行列を用いて δ-ナッシュ均衡戦略を求めることで，人間のような癖や傾向を持った戦略を生成する．もとの勝率 p_δ に対して，式 (6.1) に示す関数の出力 p' を補正後の勝率とした．なお，式中における α はその行動の組み合わせを選好するパラメータで，負ならば避けようとすること，正ならば好むことを表す．

$$p' = p_\delta + \alpha(1 + 20p_\delta(1 - p_\delta)) \tag{6.1}$$

この式では，p_δ が 0 または 1 に近い場合は，p_δ には小さい変異（α 程度）しか与えられない．一方で，p_δ が 0.5 に近い場合には，p_δ には大きい変異（6α 程度）が与えられる．はっきり負けや勝ちとなるような局面の評価はバイアスであまり変わらないと考えるため，このような変異の差を与える式とした．

6.2　付与する2種類のバイアスとその結果

前節で提案した手法を用いて実際に人間プレイヤのような傾向や癖を付与し，その結果を検証する．AI プレイヤに実際に与えるバイアスを以下に示す．

攻撃/交換を好むバイアス

4.1 節で説明したように，簡易版ポケモンバトルの行動は攻撃と交換に分類できる．人間プレイヤの好みや性格，プレイスタイルによって攻撃/交換の選択比率が偏ることは十分に考えられるため，これらの選択肢すべてにバイアスを付与する．具体的には，攻撃を好むバイアスの場合，行動 a_i が攻撃行動だった場合，$p_{i1}, p_{i2} \cdots, p_{i5}$ に (6.1) 式で正の α を用いて利得の補正を行う．

有効/非有効な攻撃を好むバイアス

簡易版ポケモンバトルの攻撃行動はダメージの補正倍率 1.5 倍以上の攻撃（以下，有効攻撃），補正倍率 1.0 倍の攻撃（以下，普通攻撃），補正倍率 0.5 倍の攻撃（以下，非有効攻撃）に分類できる．比較的素直なプレイヤや，安定思考のプレイヤは有効攻撃を好み，素直でないプレイヤや先読みが得意なプレイヤは非有効攻撃を好む傾向が想定される．このバイアスでは，相手プレイヤが場に出しているポケモンに対して，有効/非有効な攻撃行動に対してバイアスを付与する．

6.2.1　手法の有効性に関する評価

提案したバイアスが有効であるかについて，評価実験を行った．誤認知利得に，更にバイアスによる補正をかけた利得行列からナッシュ均衡戦略を求めるプレイヤの 500 戦分の対戦ログを分析することで，手法の有効性を検証した．なお，検証時の各パラメータはノイズ $\delta = 0.05$，$\alpha = 0.03$ とした．

図 6.1 に通常の δ-ナッシュプレイヤ (delta nash)，攻撃を好むバイアスを付与した δ-ナッシュプレイヤ (attack)，交換を好むバイアス (switch) を付与した δ-ナッシュプレイヤの対戦ログの分析結果を示す．図中 (a) は攻撃行動と交換行動の選択確率を示しており，図中 (b) は攻撃行動のうち，意思決定の段階で相手の場のポケモンに有効/普通/非有効な攻撃を選択した確率を示している．また，図中 (c) は交換行動のうち，意思決定の段階で相手の場のポケモンにタイプ相性が有利/普通・不利な交換を選択した確率を示す．

図中 delta nash と attack をそれぞれ比較すると，攻撃を好むバイアスを与えたプレイヤの攻撃選択確率が 8%以上増加していることが確認できる．

また，図中 switch の交換を好むバイアスを付与したプレイヤの着手分析結果と通常の δ-ナッシュプレイヤを比較すると，交換選択確率が 13%以上増加している．

次に，攻撃行動のうち有効な攻撃を好むバイアスを与えたプレイヤ (effective attack)，非有効な攻撃を好むバイアスを与えたプレイヤ (ineffective attack) の着手分析結果を図 6.2 に示す．

通常の δ-ナッシュプレイヤと，有効な攻撃を好むバイアスを付与したプレイヤを比較すると，有効な攻撃を選択する確率は 26%以上と大幅に増加している．有効攻撃を選択する確率が増加したことで，攻撃行動の選択確率も 2%程度の増加が確認できる．非有効な攻撃を好むバイアスを付与したプレイヤについても，非有効な攻撃の選択確率が約 26%と大幅に増加している．

以上の結果から，定義したバイアスについて，それぞれ意図した行動の着手確率を増加させることができており，人間プレイヤが持つような癖や傾向を表現することができたといえる．

さらに，人間のような癖や傾向を付与した δ-ナッシュプレイヤの強さについても検証を行った．5.4.2 項の実験と同様に，ナッシュプレイヤと 500 回対戦を行い，プレイヤの勝率について確認した結果を表 6.1 に示す．なお，検証時のパラメータは，正規分布 $(0, \delta(1+4p(1-p)))$，ノイズ $\delta = 0.05$，$\alpha = 0.03$ とした．

各バイアスを付与したプレイヤは，通常の δ-ナッシュプレイヤ (delta nash) と比較しても，勝率の差は比較的小さいことがわかった．また，攻撃を好むバイアス (attack) と有効な攻撃を好むバイアス (effective attack) に関しては，delta nash よりも高い勝率であることが確認できる．以上の結果から，人間的な癖や傾向を付与した δ-ナッシュプレイヤは，δ-ナッシュ均衡戦略の性質を一定程度保持していると考える．

また，著者がテストプレイとして各バイアスを与えたプレイヤと複数回対戦することで挙動を観察した．テストプレイでは，攻撃を好むバイアスを与えたプレイヤであっても，例えば初期局面でタイプ相性が不利な場合は交換を行うなど，バイアスを与えた着手以外も状況に応じて用いることを確認した．

delta nash	attack	switch	effective attack	ineffective attack
0.456	0.470	0.456	0.484	0.436

表 6.1: 各バイアスを付与した δ-ナッシュプレイヤの勝率

6.3 各パラメータが与える影響

本節ではバイアスのパラメータ α やノイズのパラメータ δ が，人間的な癖や傾向を付与した δ-ナッシュプレイヤの強さや着手確率にどのような影響を与えるかについて検証を行う．

実験は (6.1) 式のパラメータ α とノイズ δ を変化させることで，このプレイヤのナッシュプレイヤに対する勝率および着手確率にどのような変化が起こるかについて検証する．なお，ノイズの分布には正規分布 $(0, \delta(1+4p(1-p)))$ を用いた．

以下の表 6.2 に攻撃を好むバイアスを付与した δ-ナッシュプレイヤについて，各パラメータを変化させた場合の勝率および，攻撃選択確率について示す．なお，ナッシュプレイヤとの対戦数は各 500 回である．プレイヤの勝率とノイズ δ の関係性は 5.4.2 項で述べた通りだが，バイアスを付与したプレイヤでも同様に勝率が下がる傾向がある．α が勝率に与える影響としては，ノイズが小さい場合には $\alpha = 0.03$ 付近がピークになっているものの，ノイズが大きくなると α が高い方が勝率が良いという結果になった．これは，攻撃行動が必ず相手ポケモンに 1 ダメージ以上与えられる損が少ない行動であることが理由として考えられる．交換を好むバイアスや，非有効な攻撃を好むバイアスなど，ハイリスクハイリターンなバイアスについても同様の検証を行ったところ，α が大きくなるにつれ，勝率が下がる結果となった．

バイアスを付与した着手確率については，α が高い方が着手確率が上昇している．また，δ が高いほうが，やや攻撃選択確率が高いものの，それほど大きな差ではないといえる．

このようにノイズのパラメータ δ では強さを，バイアスパラメータ α では各着手の偏りを，おおむね自由に調整して多様な戦略を表現できることが明らかになった．7 章では，多様な着手戦略が得られることを利用し，プレイヤの着手を予測することで，相手プレイヤから更にアドバンテージを得る手法についてアプローチを示す．

α	0			0.03			0.06		
δ	0.05	0.1	0.2	0.05	0.1	0.2	0.05	0.1	0.2
勝率	0.456	0.400	0.316	0.470	0.424	0.326	0.446	0.404	0.382
攻撃選択確率	0.768	0.777	0.799	0.851	0.848	0.849	0.875	0.885	0.889

表 6.2: 攻撃を好むバイアスを付与した δ-ナッシュプレイヤのパラメータごとの勝率および攻撃選択確率の変化

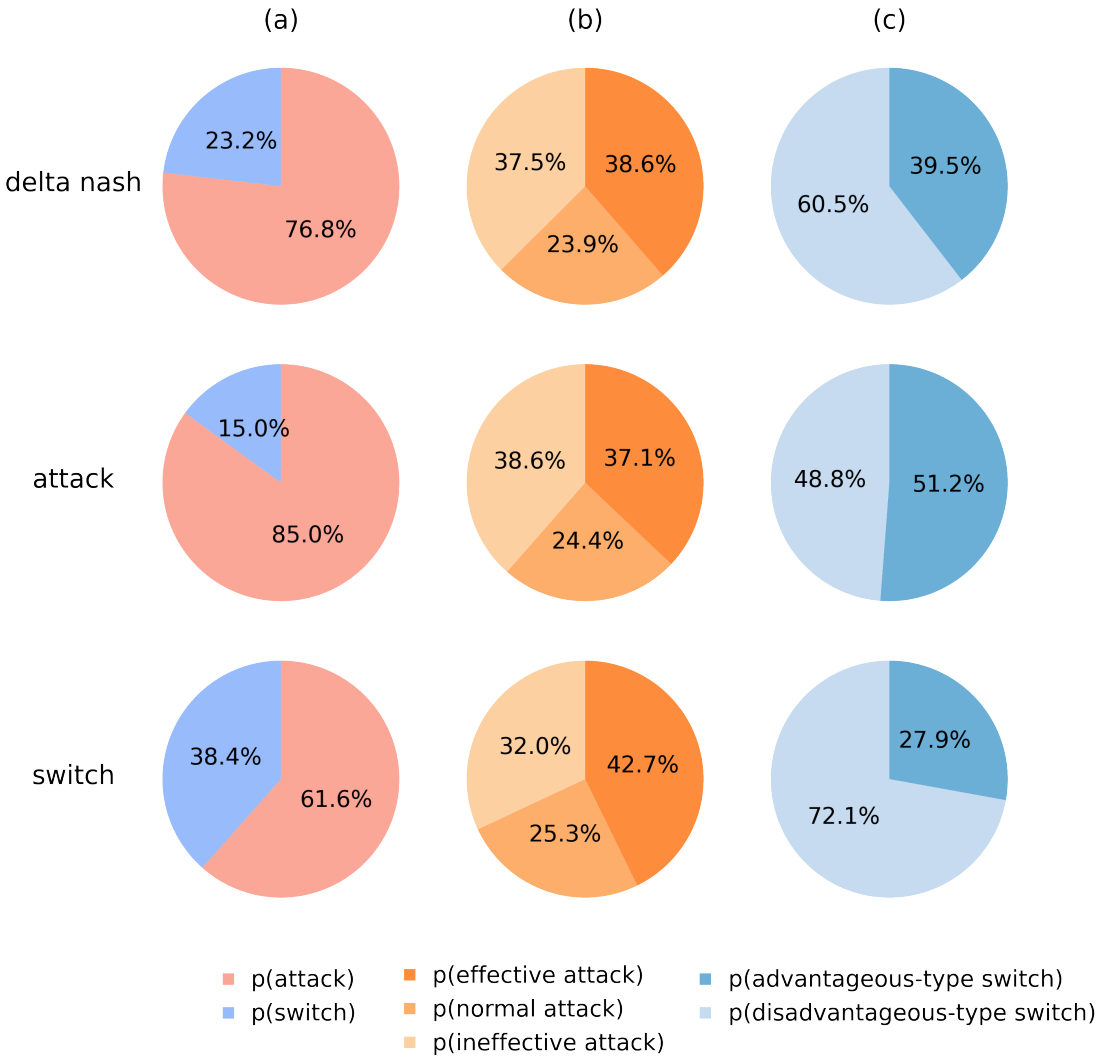

図 6.1: 通常の/攻撃/交換を好むバイアスを付与した δ-ナッシュプレイヤの着手分析結果

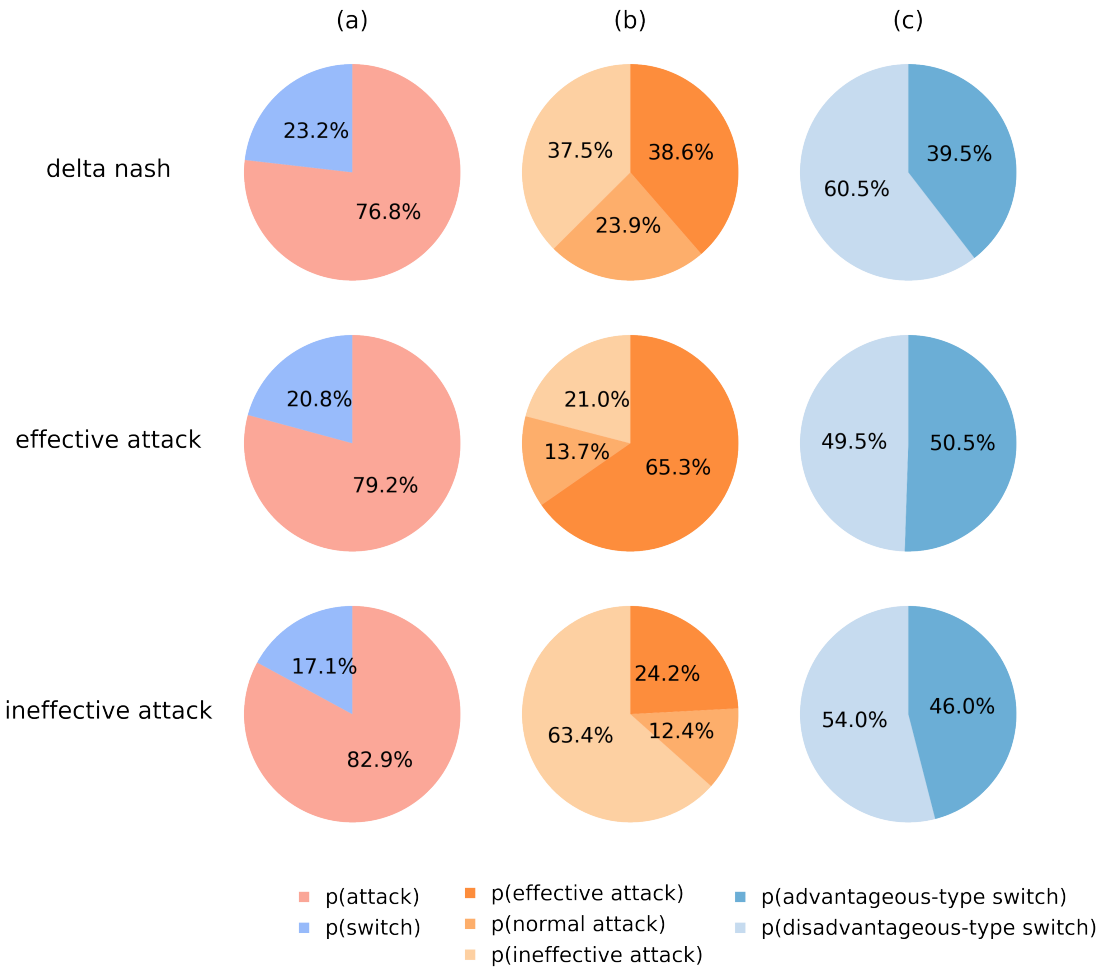

図 6.2: 通常の/有効攻撃/非有効攻撃を好むバイアスを付与した δ-ナッシュプレイヤの着手分析結果

第7章　搾取戦略実施機能のアプローチ

5章ではδ-ナッシュ均衡戦略は，ナッシュ均衡戦略の性質を一定程度維持しており，人間的には有望な手を含めた多様な着手戦略によって人間プレイヤに「読まれた」と思わせるような行動ができることを説明した．

本章では読み合いの演出において，さらに人間プレイヤに「読まれた」と思わせるような着手を強調するため，人間のような読みを模倣した着手を生成する手法についてアプローチを示す．

人間プレイヤが行う行動の読みは大きく分けて，次の相手プレイヤの行動予測と，予測される相手の行動に対してアドバンテージを得る戦略を考えることの2つから成る．本研究では前者を相手モデリングによる着手予測，後者については予測された行動に対して搾取戦略を用いることを提案する．

まず，相手モデリングによる着手予測に関するアプローチについて説明する．相手モデリングには，6.3節の実験で確認した各バイアスを付与したプレイヤについて，パラメータを調整することで多様な着手戦略を得られることを利用する．以下に相手モデリングによる着手予測の具体的な手順を示す．

- 現在または過去のゲームにおいて，各ターンごとに相手プレイヤの着手と，攻撃/交換/有効攻撃/非有効攻撃の各バイアス，バイアス定数α, ノイズδをそれぞれパラメータとしたδ-ナッシュ均衡戦略（の組）を保持する．
- 相手の行動履歴$a_i(i = 1, \cdots, n)$について，k番目のδ-ナッシュ均衡戦略N_kの尤度$L_k = \prod_i^n p(N_k|a_i)$をそれぞれ求める．$k^* = \arg\max_k L_k$として最も尤度が高くなるモデル$N_{k^*}$を求める．
- 現在の状態に対して，N_{k^*}を用いて相手の行動確率分布を得る．

次に，予測される行動に対する搾取戦略の生成方法についても説明する．相手の戦略が固定されている場合，相手の戦略に対する最適応答反応は多くの場合，ある行動を確率1で選択する純粋戦略になることが知られている．そのため，真の勝率による利得行列をもとに，自身の着手の中で最も勝率が高い行動を確率1で着手することで，相手の行動に対してアドバンテージを得ることができると考える．

このアプローチによって人間プレイヤのような「読み」を模倣する搾取戦略を強調することが可能だと考えたものの，評価実験によって有効性を示すまでには

至らなかったため，搾取戦略実施機能の完全な実装とその評価については今後の展望とする．

第8章　おわりに

本研究では，ポケモンバトルを簡略化した同時手番ゲームを対象として，AI プレイヤに人間のような読み合いを演出させる手法の確立を目指した．読み合いを演出する手法として，人間が自然に感じる標準着手を生成した上で，人間のような癖や傾向を持つ着手や，人間の読みを模倣した着手を織り交ぜることを提案した．

まず，提案したゲーム上で読み合いを演出する前提として，人間プレイヤが自然に感じる着手の生成に取り組んだ．後退解析によってゲームの全状態の真の勝率を求め，ナッシュ均衡戦略に従うナッシュプレイヤを作成した．さらに，ナッシュプレイヤの強さや着手確率を調整する手法として，ノイズを加えた誤認知利得行列からナッシュ均衡戦略を求める δ-ナッシュプレイヤを提案した．評価実験から，δ-ナッシュプレイヤは強さの調整が可能であることや，ナッシュプレイヤには選ばれないが人間には有望に見える着手を確率的に選ぶことによって，より人間らしいと感じられる対戦相手であることを示した．

次に，人間のような癖や傾向を持つ着手として，δ-ナッシュ均衡戦略を求める誤認知利得行列を更にバイアスによって補正し，人間が持つ癖を付与した着手戦略を生成した．攻撃/交換/有効攻撃/非有効攻撃を好むバイアスを付与した 4 つのプレイヤは，それぞれ付与したバイアスに従った行動の選択確率の増加が見られた．さらに，強さと着手の偏りを比較的独立に調整できることで，読み合い演出に適した多様な着手戦略が作成できることを示した．

人間の読みを模倣した着手に関してはアプローチを示したものの，有効性を示すまでには至らなかった．今後の展望としては，人間の読みを模倣する搾取戦略実施機能の完全な実装と，各機能を統合した統合モデルを用いた被験者実験により，実際に人間プレイヤ同士の対戦に近い読み合いを演出することができるかについて有効性を検証することなどがある．

また，提案した簡易版ポケモンバトル以外の同時手番ゲームに対しても，本研究の手法を応用し読み合い演出に適した多様な着手戦略を作成に貢献できると考えている．さらに，将来的にはオリジナルのポケモンバトルなどの隠れた情報がある不完全情報ゲームを対象とした読み合い演出も試みたい．

参考文献

[1] Silver, David, et al. ”Mastering the game of Go with deep neural networks and tree search.” nature 529.7587 (2016): 484-489.

[2] Brown, Noam, and Tuomas Sandholm. ”Superhuman AI for multiplayer poker.” Science 365.6456 (2019): 885-890.

[3] 池田心. ”楽しませる囲碁・将棋プログラミング.” オペレーションズ・リサーチ: 経営の科学 (2013).

[4] 仲道隆史, and 伊藤毅志. ”プレイヤの技能に動的に合わせるシステムの提案と評価.” 情報処理学会論文誌 57.11 (2016): 2426-2435.

[5] 池田心. ”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.” (2012).

[6] 岡田章. ”ゲーム理論［新版］”，有斐閣（2011）

[7] Zinkevich, Martin, et al. ”Regret minimization in games with incomplete information.” Advances in neural information processing systems 20 (2007).

[8] 「ポケットモンスター　ソード・シールド」公式サイト. https://www.pokemon.co.jp/ex/sword_shield/ (アクセス:2023/01/21).

[9] Van der Kleij, A. A. J. ”Monte Carlo tree search and opponent modeling through player clustering in no-limit Texas hold ’ em poker.” University of Groningen, The Netherlands (2010).

[10] Sato, Naoyuki, et al. ”Adaptive fighting game computer player by switching multiple rule-based controllers.” 2015 3rd international conference on applied computing and information technology/2nd international conference on computational science and intelligence. IEEE, 2015.

[11] 杵渕哲彦, and 伊藤毅志. ”流れを考慮した将棋における人間の指し手との一致率向上手法.” 情報処理学会論文誌 58.9 (2017): 1549-1554.

[12] 藤井叙人, et al. ”生物学的制約の導入によるビデオゲームエージェントの「人間らしい」振舞いの自動獲得.” 情報処理学会論文誌 55.7 (2014): 1655-1664.

[13] Ihara, Hiroyuki, et al. "Implementation and evaluation of information set Monte Carlo Tree Search for Pokémon." 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2018.

[14] Cowling, Peter I., Edward J. Powley, and Daniel Whitehouse. "Information set monte carlo tree search." IEEE Transactions on Computational Intelligence and AI in Games 4.2 (2012): 120-143.

[15] pokemon-showdown. https://github.com/smogon/pokemon-showdown (アクセス：2023/01/19).

[16] Lee, Scott, and Julian Togelius. "Showdown ai competition." 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE, 2017.

[17] Huang, Dan, and Scott Lee. "A self-play policy optimization approach to battling pokémon." 2019 IEEE Conference on Games (CoG). IEEE, 2019.

[18] Gambit. https://gambitproject.readthedocs.io/en/latest/index.html（アクセス：2023/01/25）.