# ガイスターにおける自己対戦による行動価値関数の学習

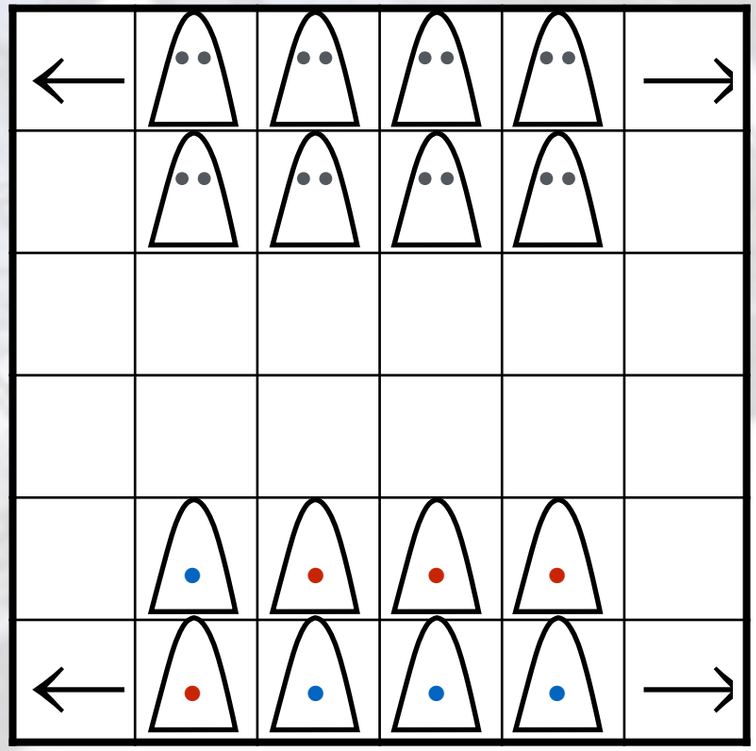
電気通信大学大学院 村松研究室 佐藤佑史

# ガイスターとは

- オバケの形をした駒を使う、２人用ボードゲーム
- 作者： Alex Randolph
- 1982年 ドイツ年間ゲーム大賞 ノミネート
- 不完全情報ゲーム

# ガイスターのルール（1）

- 駒は**青**と**赤**の２種類、各プレイヤは４つずつ持つ
- 盤のサイズは６×６
- 盤の四隅には出口のマスが存在
- 駒は前後左右１マスに移動可
- ゲームをはじめる前に自陣8マスに自由に配置



# ガイスターのルール（2）

- ゲームの勝利条件

1. 相手の**青駒**を４つ全て取る

2. 自分の**赤駒**を４つ全て取らせる

3. 自分の**青駒**を相手側の出口のマスから脱出させる

# ガイスターにおける既存研究

- Prototype-Based Learning+モンテカルロ木探索※
  - ルールを覚えたばかりの初心者程度の実力
- Ghosts Challenge 2013, 2014

---

※ Ghosts Challenge 2013 BLISS

# 強化学習

- 勝率を最大化するために、何をすべきかを学習
- 教師なし学習（棋譜データ不要）
- 自己対戦での学習
- バックギャモンが強化学習手法で成功
- 囲碁AIであるAlphaGoでも使用

# 強化学習手法：Sarsa($\lambda$)

- 勝率の見積もりを計算する行動価値関数 $Q(s,a)$

$s$：局面　　$a$：手

- $Q(s,a)$ 更新時の引数 $s,a,r,s',a'$ に由来

$r$：ゲームの結果　　$s'$：次局面　　$a'$：次局面における手

# ３層ニューラルネットワークでの関数近似

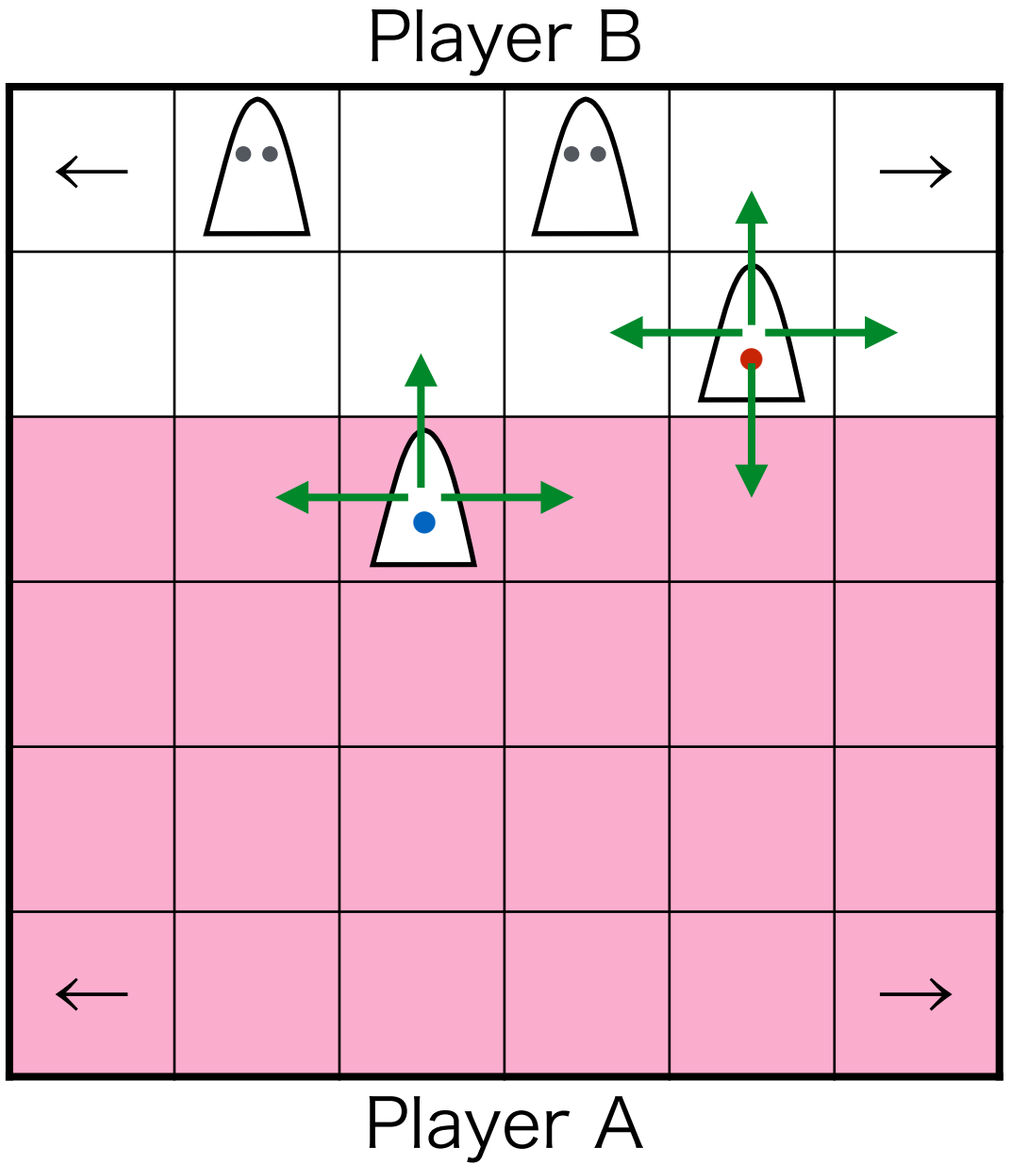
- 行動価値関数の近似に３層ニューラルネットワークを使用
- ニューラルネットワークの入力
  - 移動後の自分の青駒と赤駒の配置
  - 相手の駒の配置
  - 取った駒の数
  - 相手の駒、自分の推測
  - etc

# 学習を行う前に：着手制限

- 着手制限
  - 手前の4行にいる駒の後退不可
  - 自駒の色を考慮しない自駒の同一配置の禁止
- 着手制限を入れないとまともに学習できない
  - ランダムに手を指すと引き分けになりやすい
  - 引き分けになる手が多くなると自己対戦であるため、ずっとゲームが引き分けになる
- **学習時のみ**でなく、**対局時の行動価値関数を利用するAIプレイヤ**に対しても同様の着手制限を課す

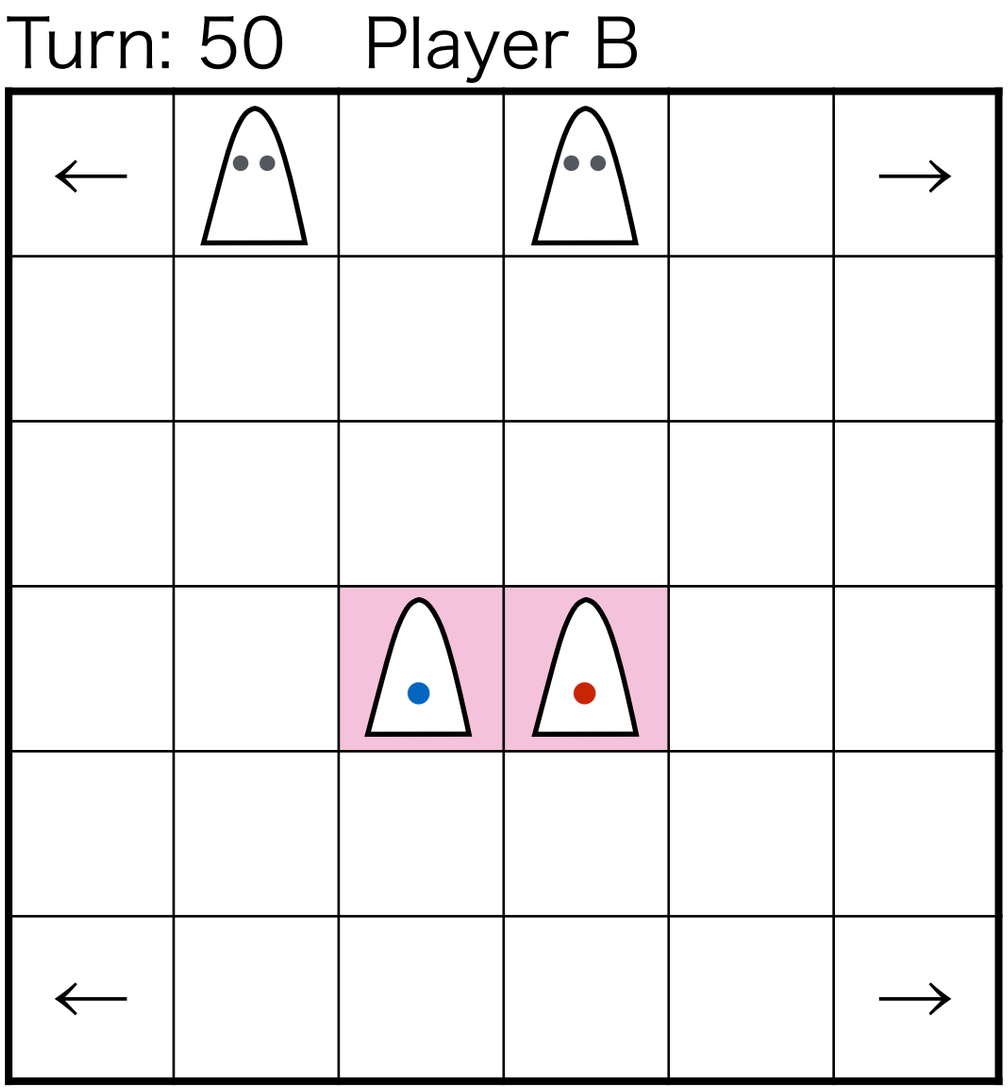
# 着手制限：後退不可

・ 手前の4行にいる駒が後退不可

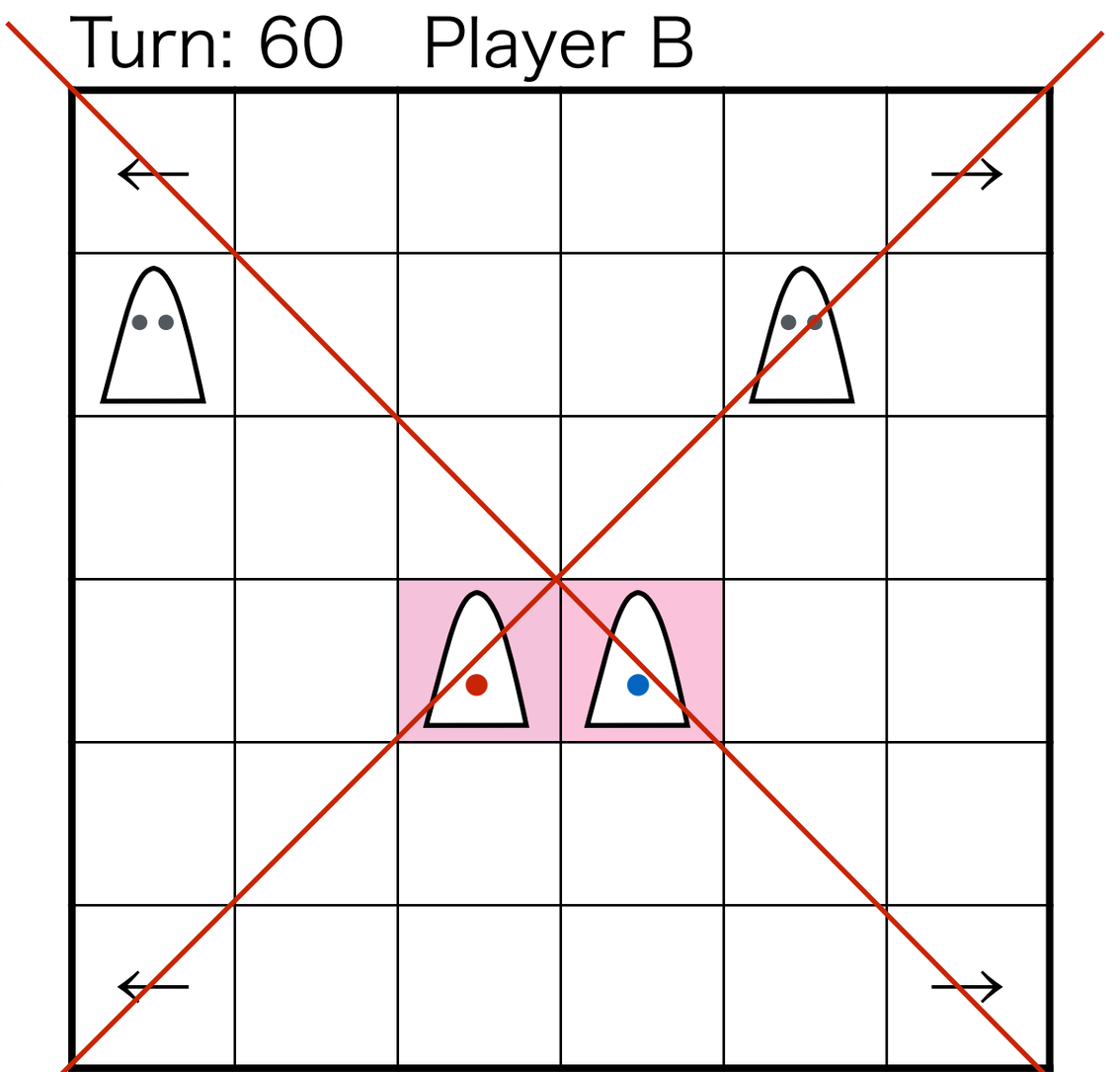


# 着手制限：同一配置禁止

・ 自駒の色を考慮しない自駒の同一配置を禁止



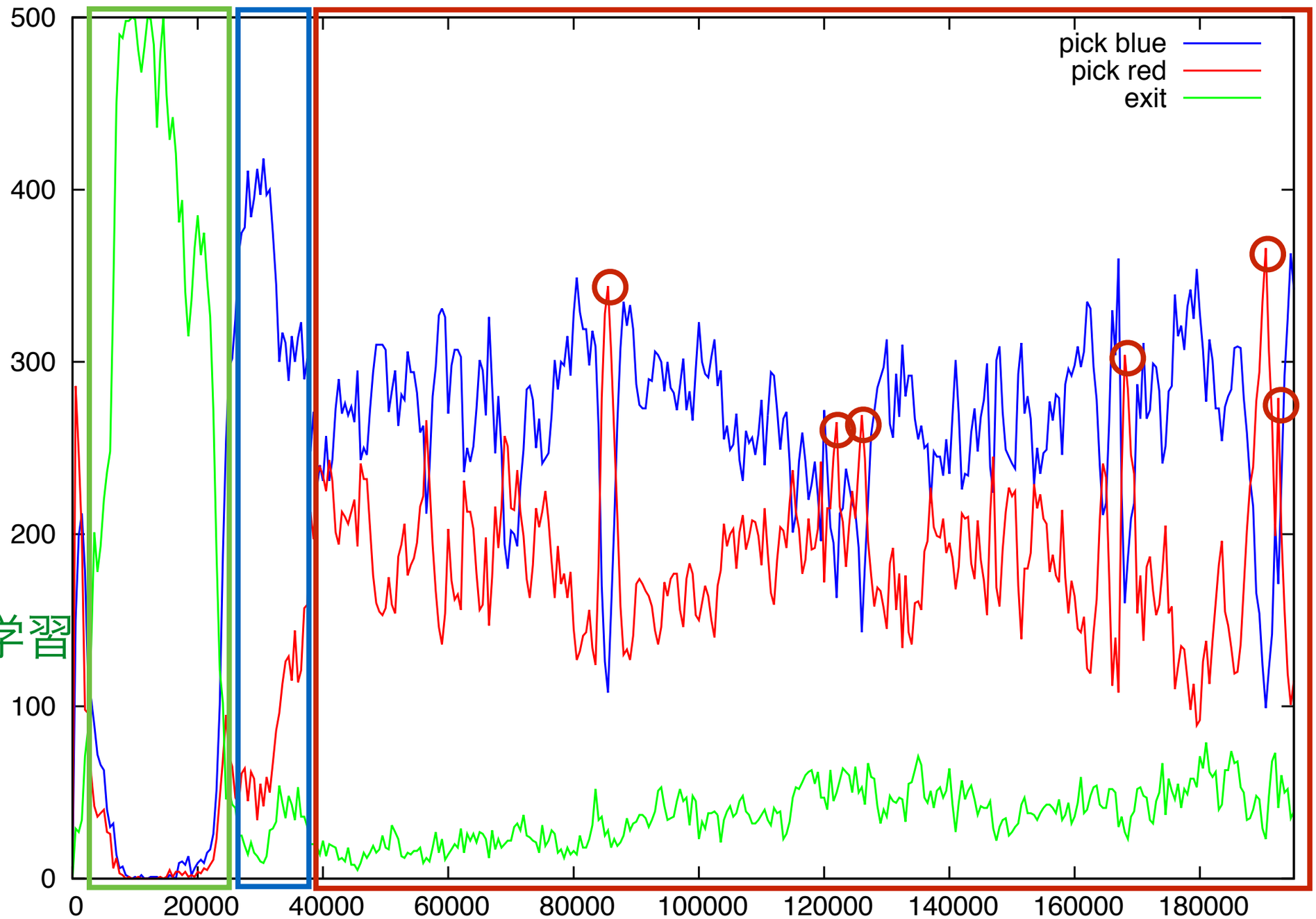

指せない



# 実験

- 150万回の自己対戦でのSarsa(λ)学習
- 500回ごとの各勝利条件を満たした回数とニューラルネットワークの重みを出力

# 実験結果（1）

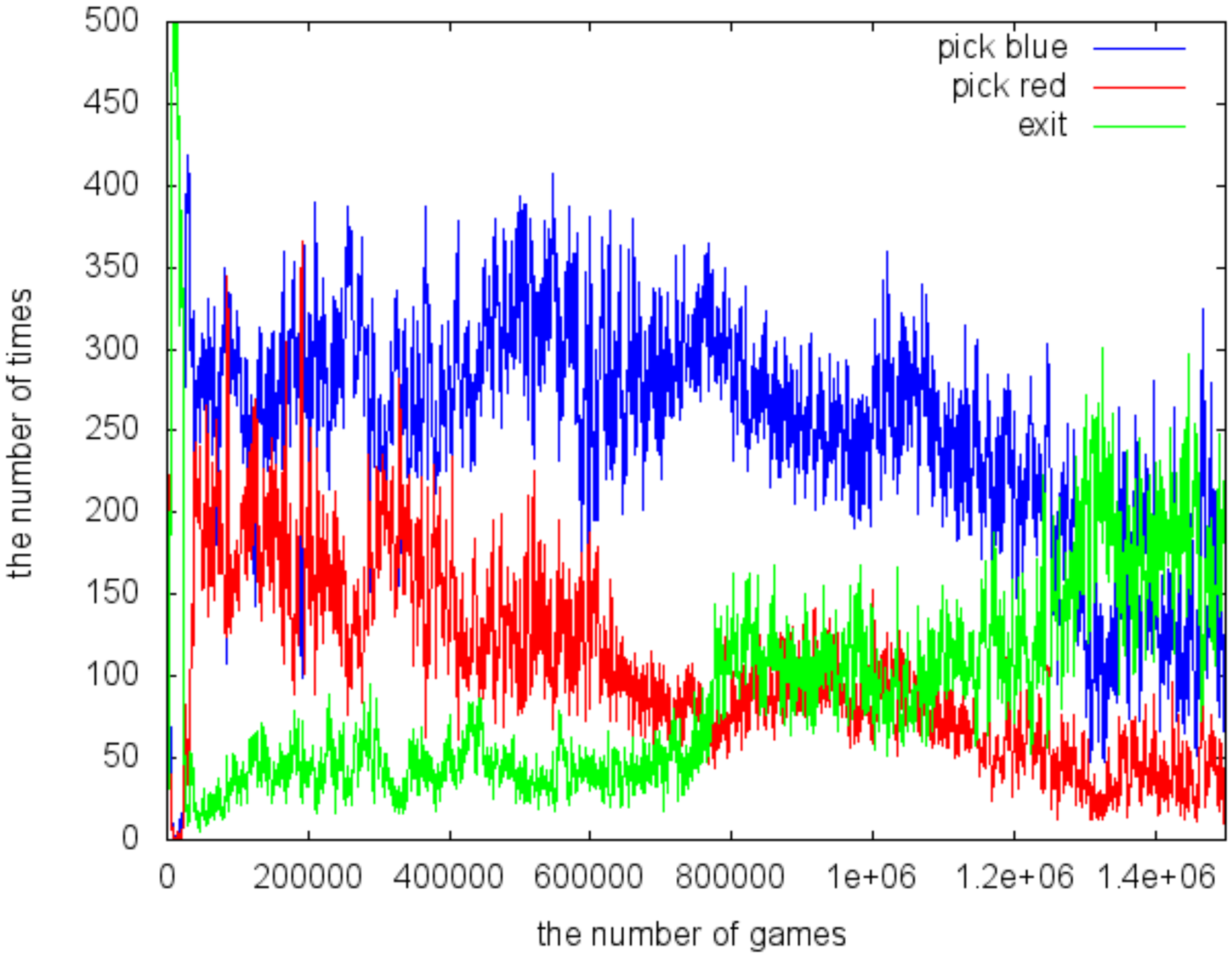



脱出手の学習

脱出する青駒取り学習

赤駒も取らせようとする

# 実験結果（2）


pick blue
pick red
exit
the number of times
the number of games
0
50
100
150
200
250
300
350
400
450
500
0
200000
400000
600000
800000
1e+06
1.2e+06
1.4e+06

# 実験結果（3）

| 学習対戦数 | 先手勝ち数 | 後手勝ち数 | 引き分け数 | 青駒取り決着数 | 赤駒取り決着数 | 脱出決着数 |
|---|---|---|---|---|---|---|
| 14000 | 257 | 223 | 30 | 0 | 1 | 479 |
| 100000 | 237 | 259 | 4 | 323 | 150 | 23 |
| 192500 | 238 | 253 | 9 | 171 | 279 | 41 |
| 500000 | 262 | 227 | 11 | 345 | 116 | 28 |

- 学習対戦数192500（Q AI）
  - 人間プレイヤが使う序盤定石をよく利用
  - ブラフとなる手を積極的に指す

# 対局実験１

- Q AIとランダムプレイヤおよびモンテカルロ木探索を用いたプレイヤ(MCTプレイヤ)との対局実験
  - ランダムプレイヤとの3000戦 （先手後手同数）
  - MCTプレイヤとの1000戦（先手後手同数）
    - プレイアウト数 10万回（約4秒）

# 対局実験１の結果

## ランダムプレイヤとの3000戦の結果

| | 勝ち | 負け | 引き分け |
|---|---|---|---|
| Q AI | 1366 | 1427 | 207 |

## MCTプレイヤとの1000戦の結果

| | 勝ち | 負け | 引き分け |
|---|---|---|---|
| Q AI | 347 | 615 | 38 |

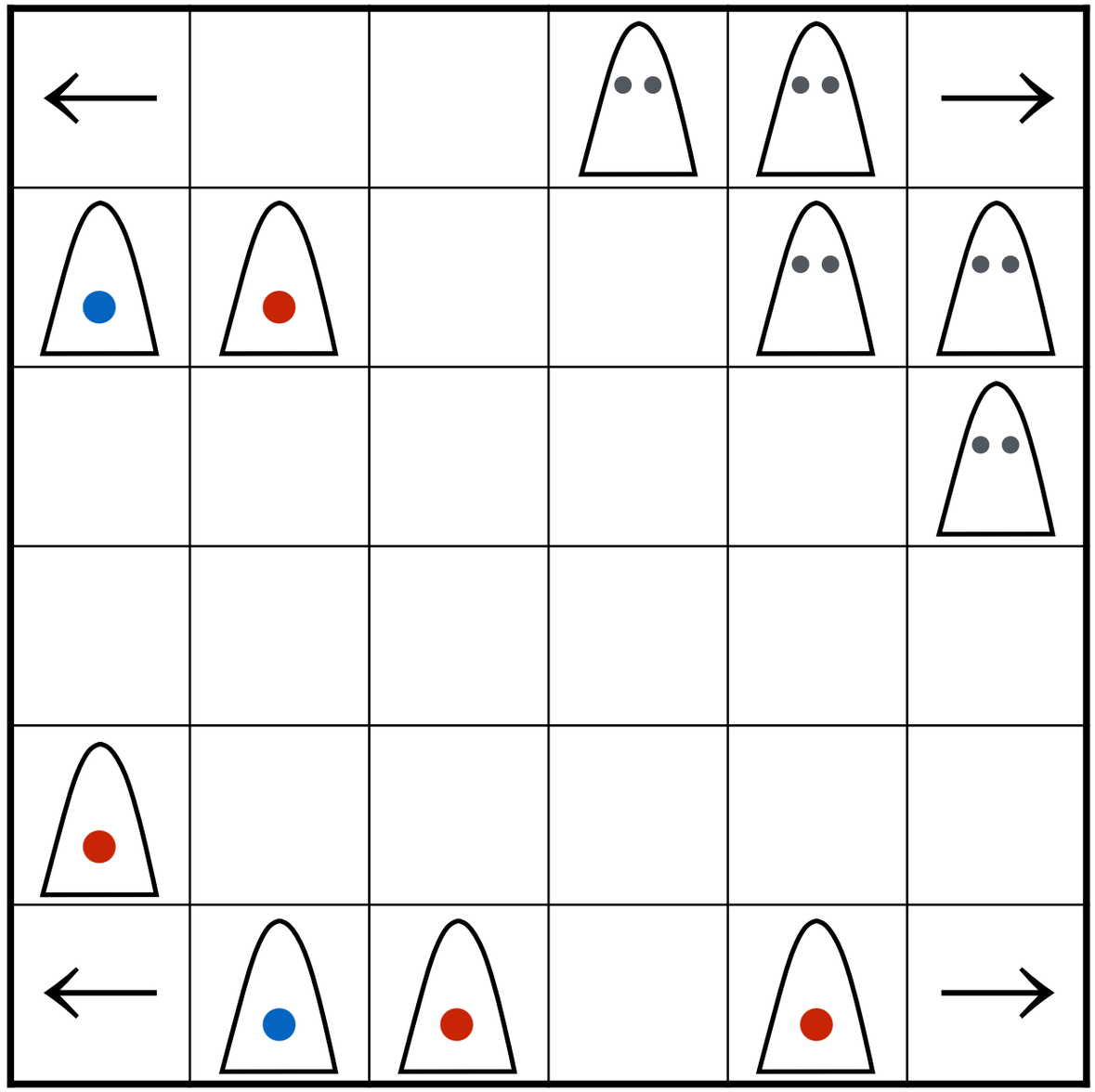
# 対局の敗因

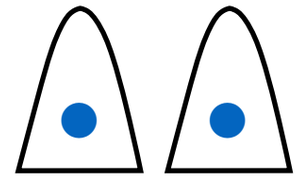
- ランダムプレイヤとの対局では推測を乱される

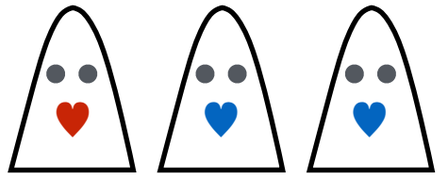
- **3手以上での必勝手を理解できていない**

# Q AIがわからない３手詰めの必勝局面

Random



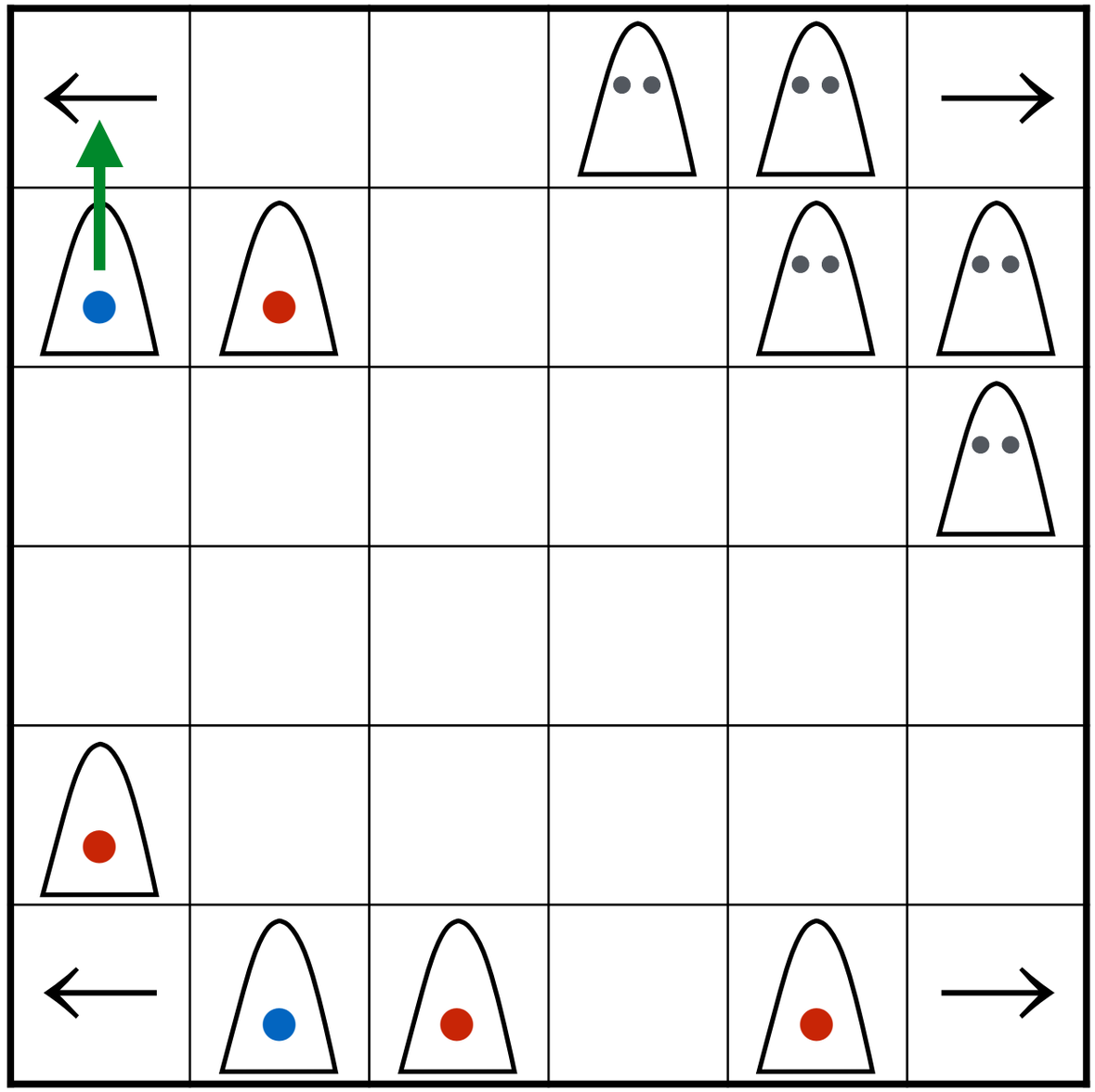


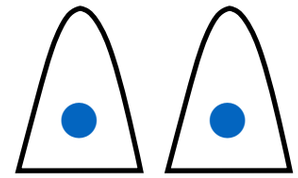


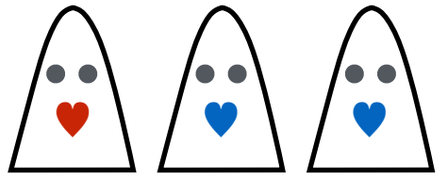
Q AI

# Q AIがわからない３手詰めの必勝局面

Random



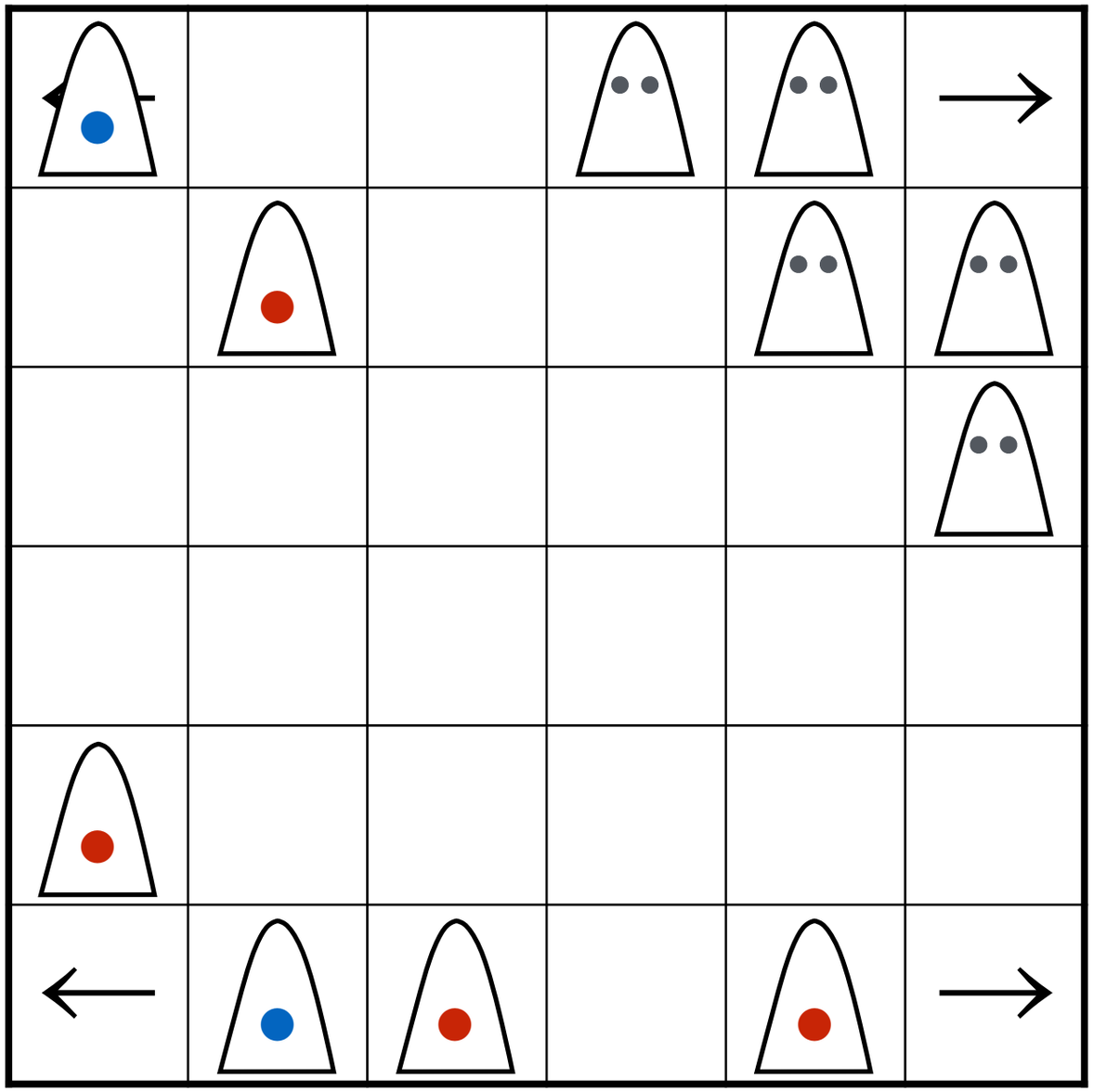


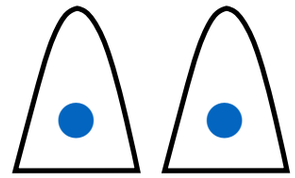


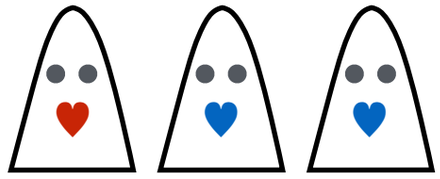
Q AI

# Q AIがわからない３手詰めの必勝局面

Random



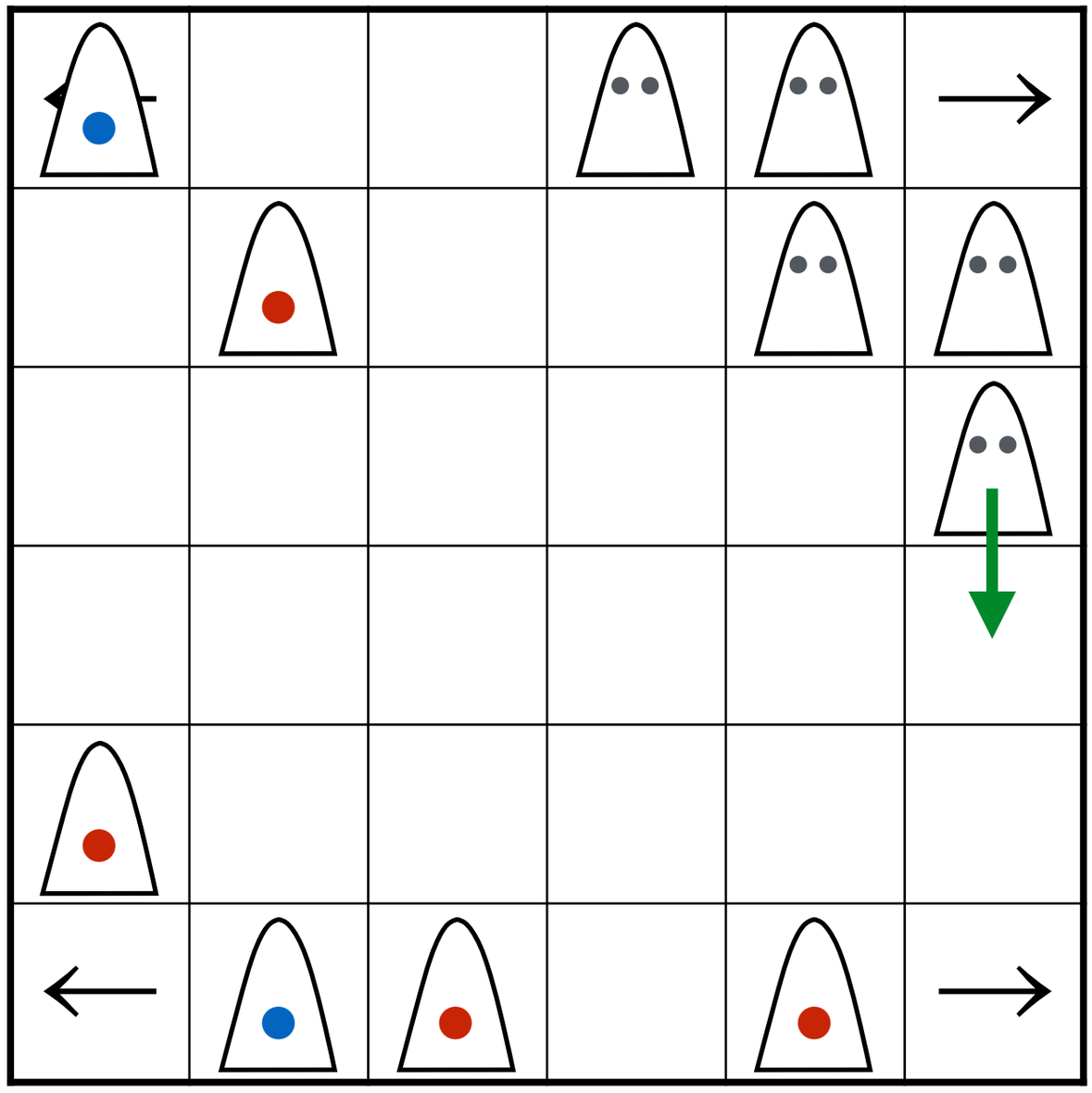


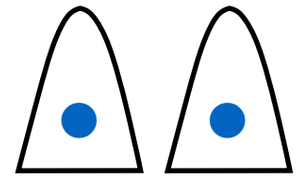


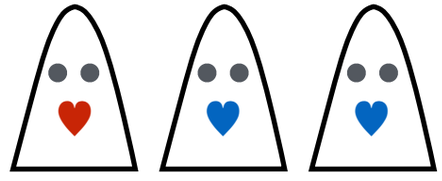
Q AI

# Q AIがわからない３手詰めの必勝局面

Random



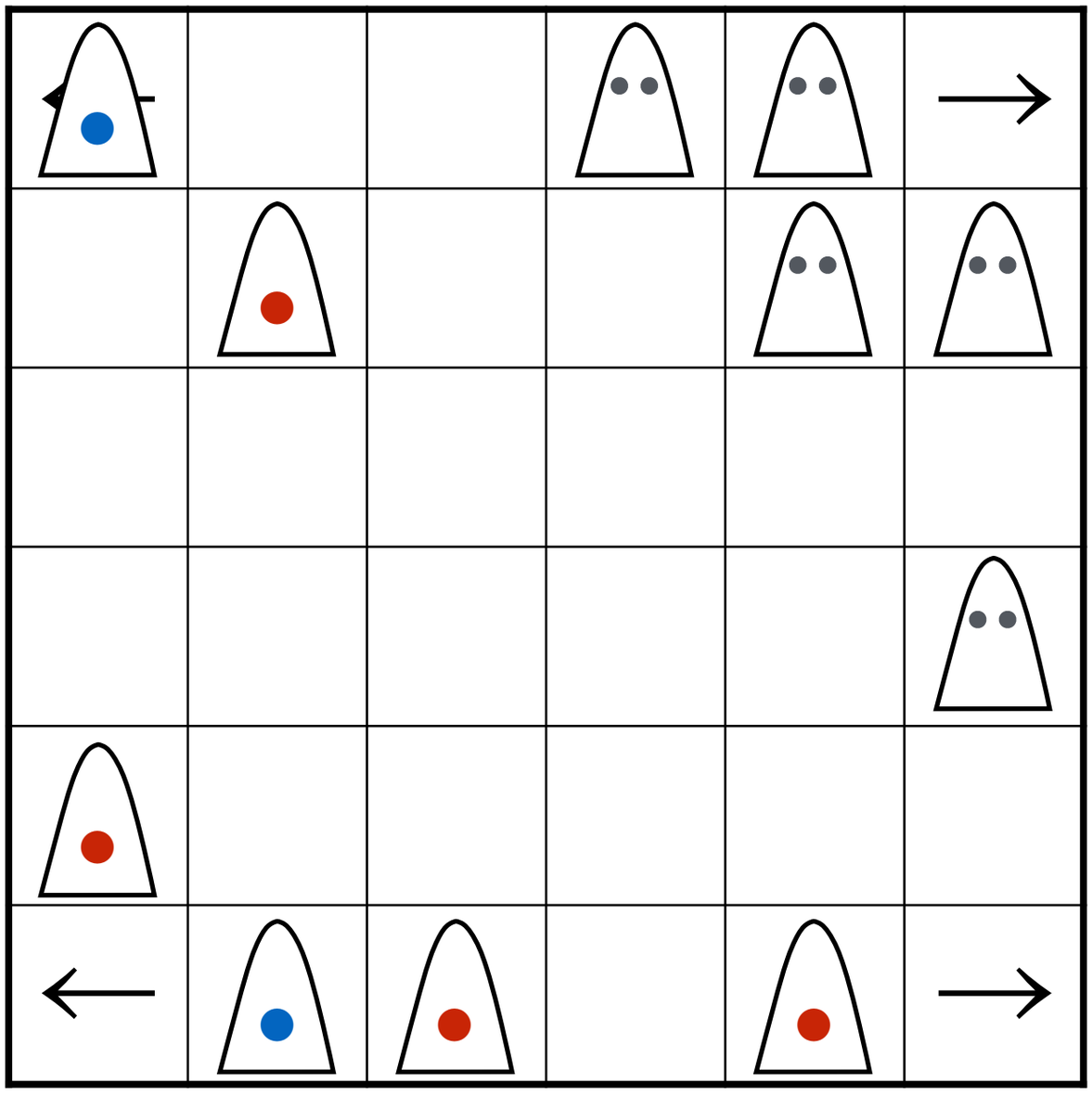


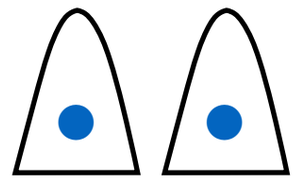


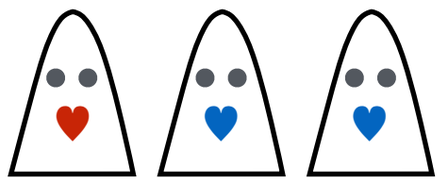
Q AI

# Q AIがわからない３手詰めの必勝局面

Random



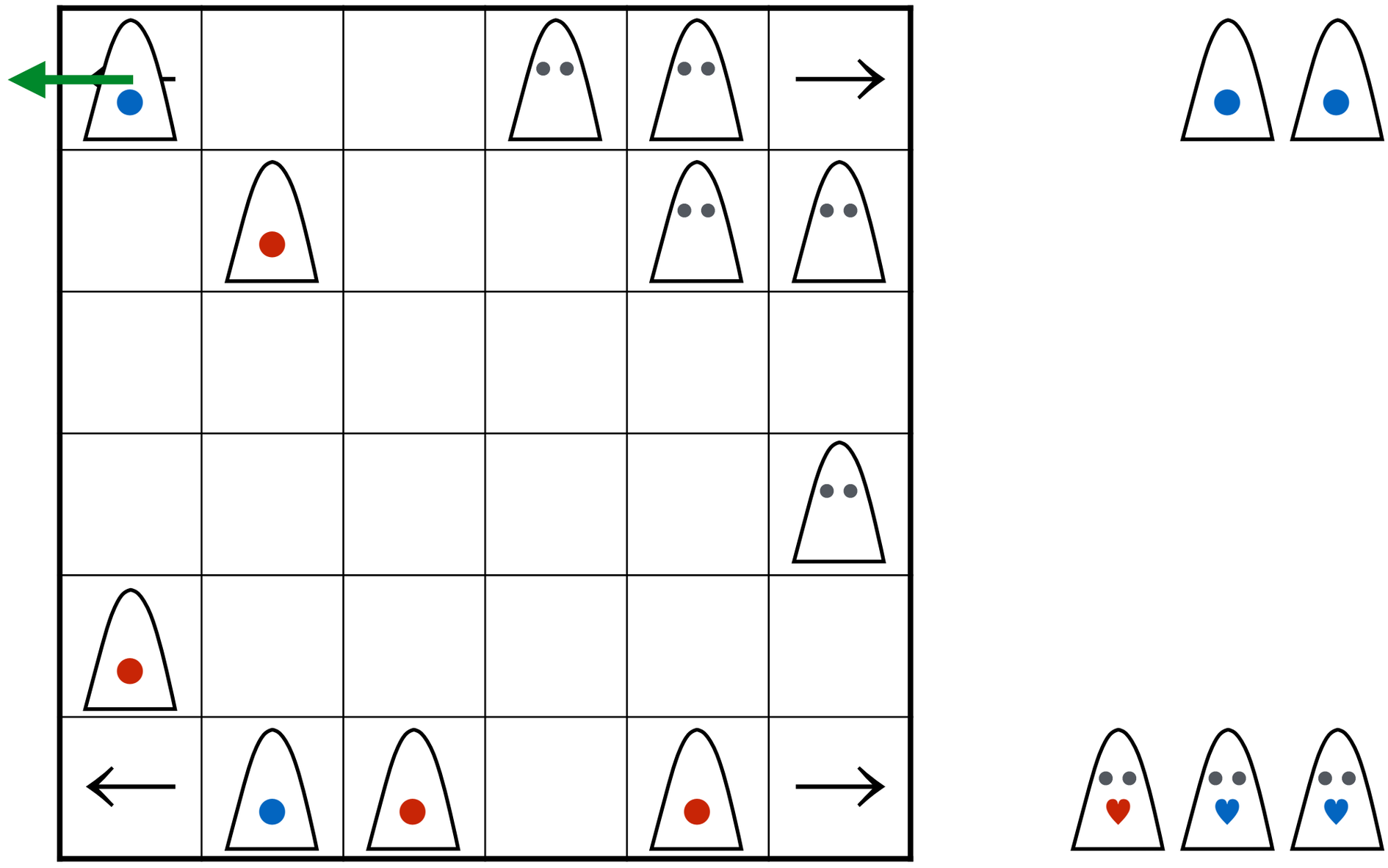




Q AI

# Q AIがわからない３手詰めの必勝局面

Random



Q AI

# Df-pn*アルゴリズム

- AND/OR木を探索する手法
- 詰将棋などで利用されるアルゴリズム
- **ただし、不完全情報ゲームでは利用不可能**

＊ Depth first proof number

# ガイスターの完全情報ゲーム化

- 相手の駒を**紫色**とし、完全情報ゲームとする
- **紫色**の駒は**青駒**として脱出することができる
- **紫色**の駒は取ると**赤駒**になる

↓

**Df-pnによる必勝手探索が可能**

# Q AI-Dfpn

- Df-pnによる150msecの必勝手探索
- 必勝手が見つからなければ、Q AIを用いる
- Q AI-Dfpnを用いて対局実験

# Q AI-Dfpnでの対局結果

## ランダムプレイヤとの3000戦の結果

| | 勝ち | 負け | 引き分け |
|---|---|---|---|
| Q AI-Dfpn | 1971 | 916 | 113 |

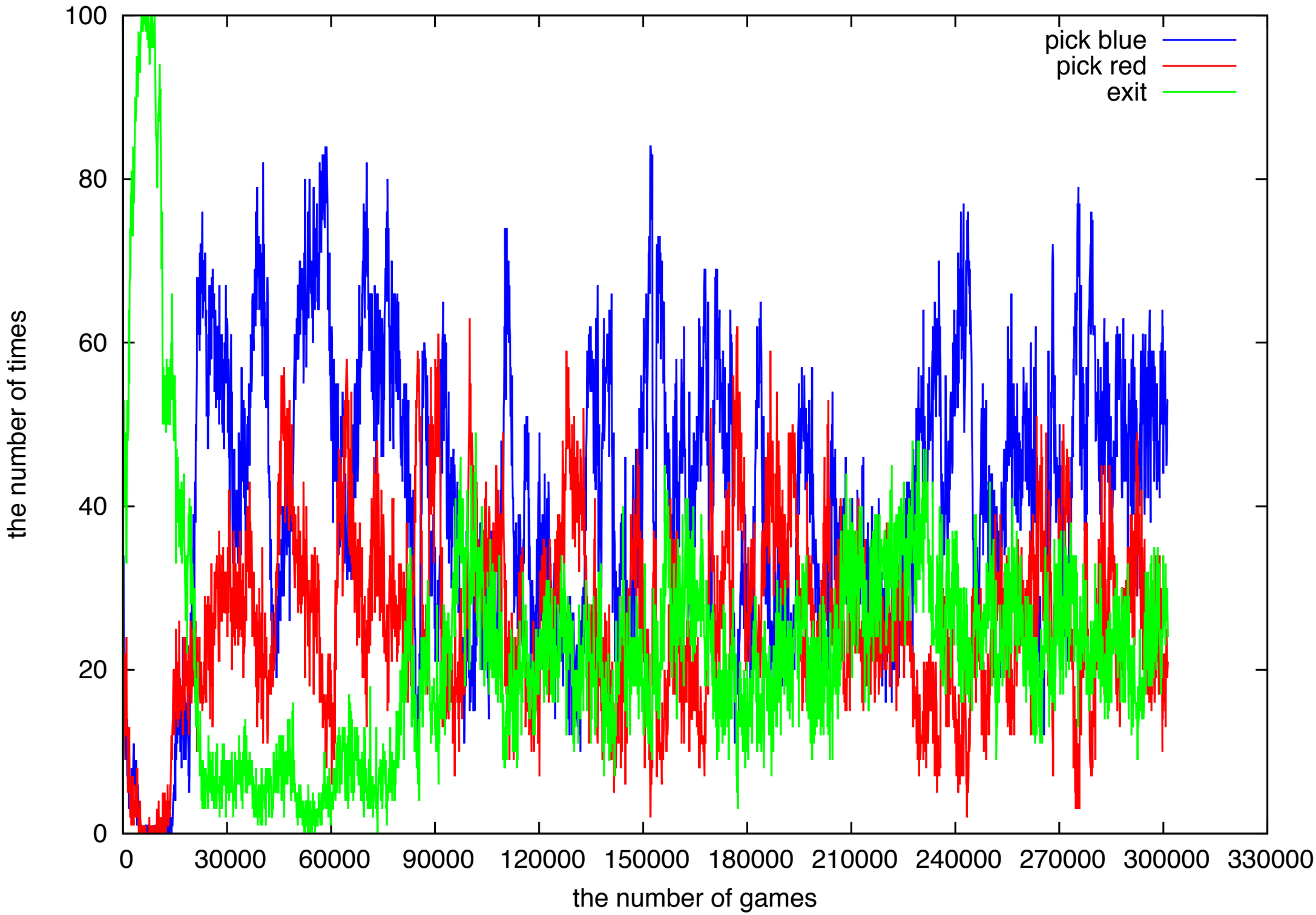
## MCTプレイヤとの2000戦の結果

| | 勝ち | 負け | 引き分け |
|---|---|---|---|
| Q AI-Dfpn | 1178 | 769 | 53 |

# 実験２

- 自己対戦による必勝手探索を組み込んだSarsa(λ)学習（Q AI2）
- 100戦ごとの勝利条件を満たした回数を出力
- NNの入力に２つの特徴を追加
  - 移動後の相手の駒の出口までの最短距離
  - 移動後に相手の脱出手を防ぐことができるか
- 必勝手探索は150msec

# 実験結果（1）


pick blue
pick red
exit
the number of times
the number of games
0
20
40
60
80
100
0
30000
60000
90000
120000
150000
180000
210000
240000
270000
300000
330000

# Q AI2-Dfpnでの対局

- Q AI2-Dfpnで対局実験を行なう

| 学習対戦数 | 先手勝ち数 | 後手勝ち数 | 引き分け数 | 青駒取り決着数 | 赤駒取り決着数 | 脱出決着数 |
|---|---|---|---|---|---|---|
| 130100 | 39 | 38 | 23 | 16 | 51 | 10 |

# Q AI2-Dfpnでの対局結果

## ランダムプレイヤとの1000戦の結果

| | 勝ち | 負け | 引き分け |
|---|---|---|---|
| Q AI2-Dfpn | 516 | 239 | 245 |

## MCTプレイヤとの1000戦の結果

| | 勝ち | 負け | 引き分け |
|---|---|---|---|
| Q AI2-Dfpn | 510 | 447 | 47 |

# まとめ

- 不完全情報ゲームを完全情報ゲームとしてモデル化し、必勝手探索を行なう手法の提案
- Q AI-DfpnでMCTプレイヤに勝ち越した
- どの学習段階が最も強いのかわからない

# 今後の課題

- 学習において、自己対戦をどこで止めればいいのかわからない

- 自己対戦の場合、同一戦略での対戦のみでしか学習できない

- 必勝手探索を用いた敵赤駒の特定