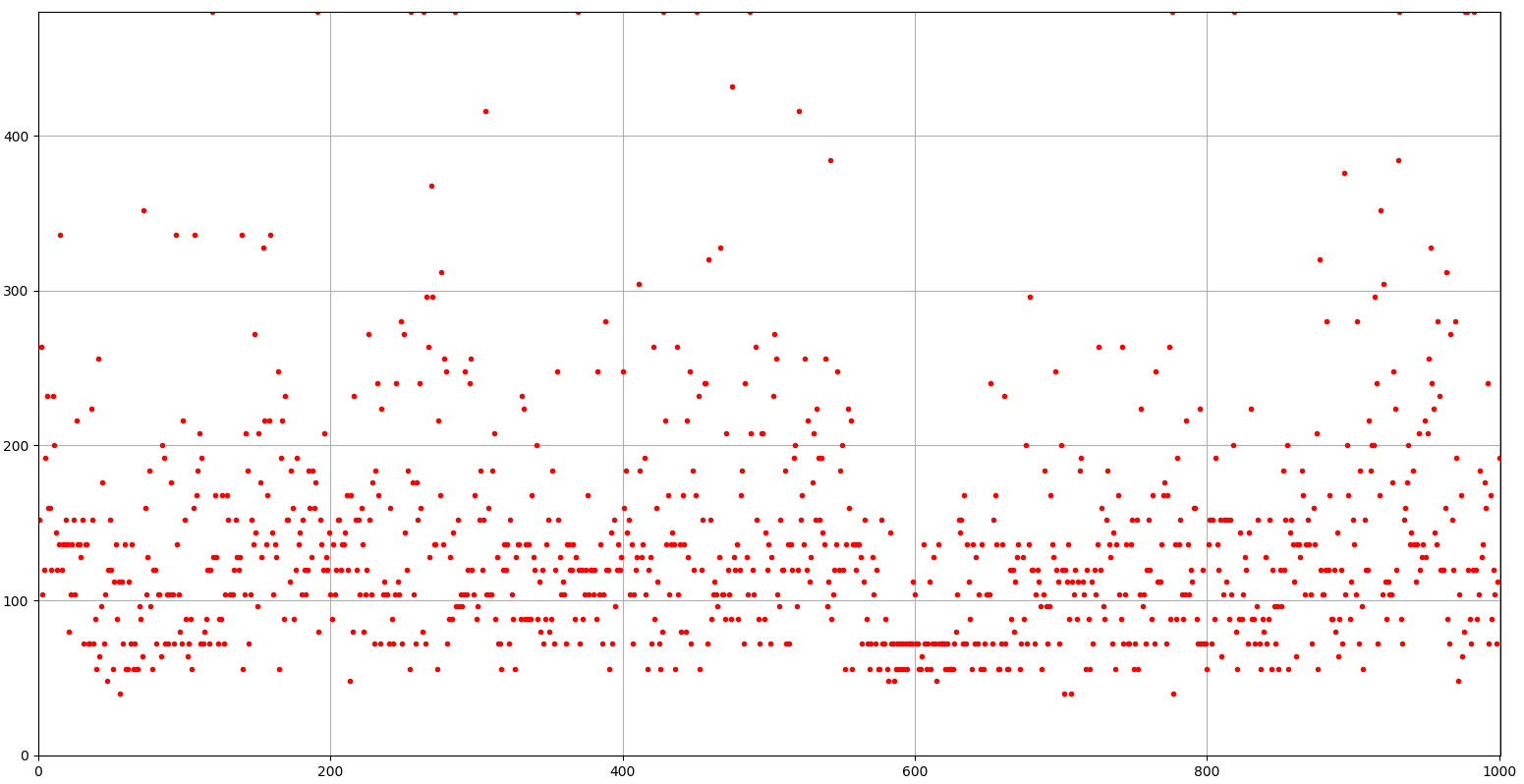
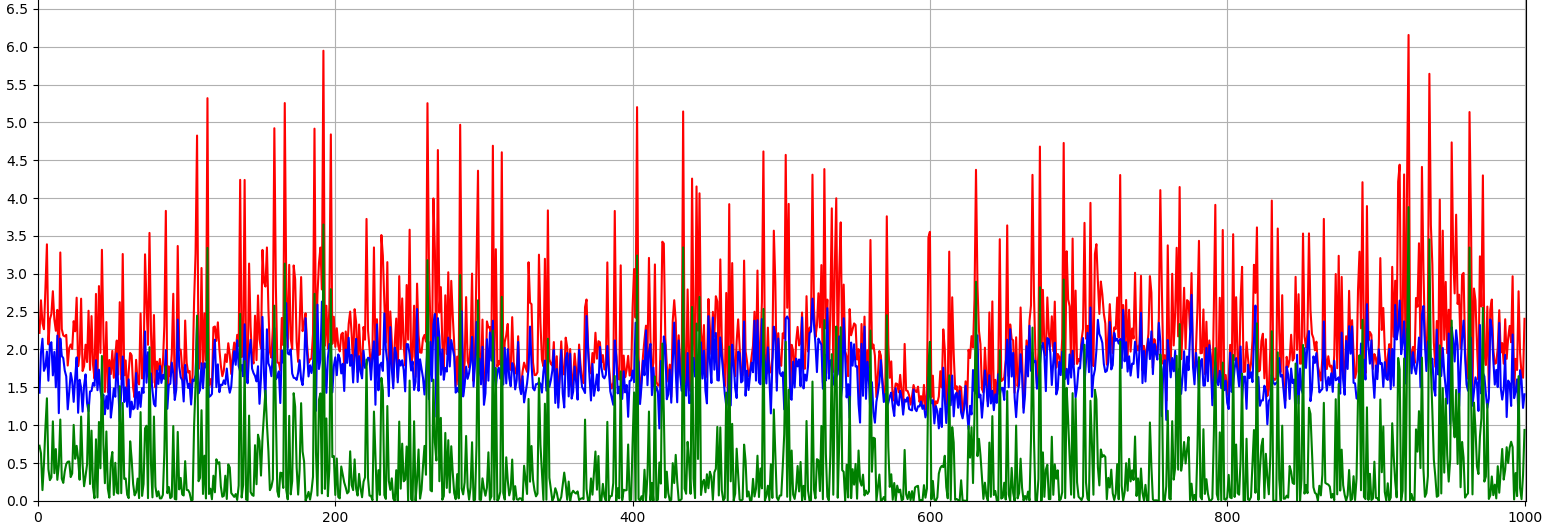
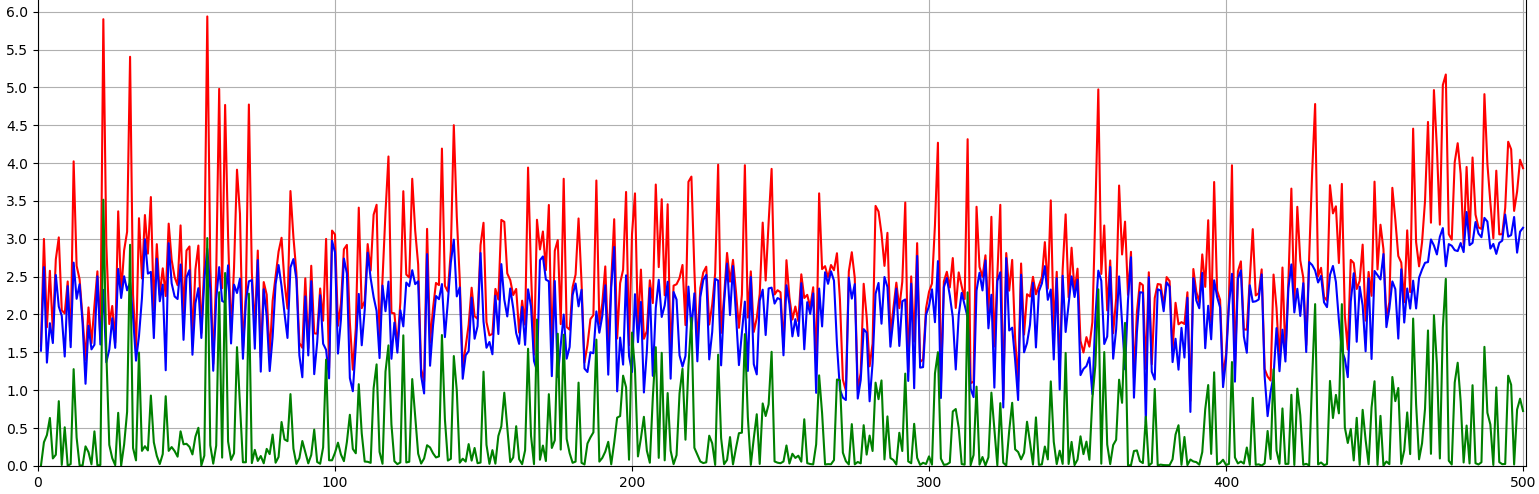
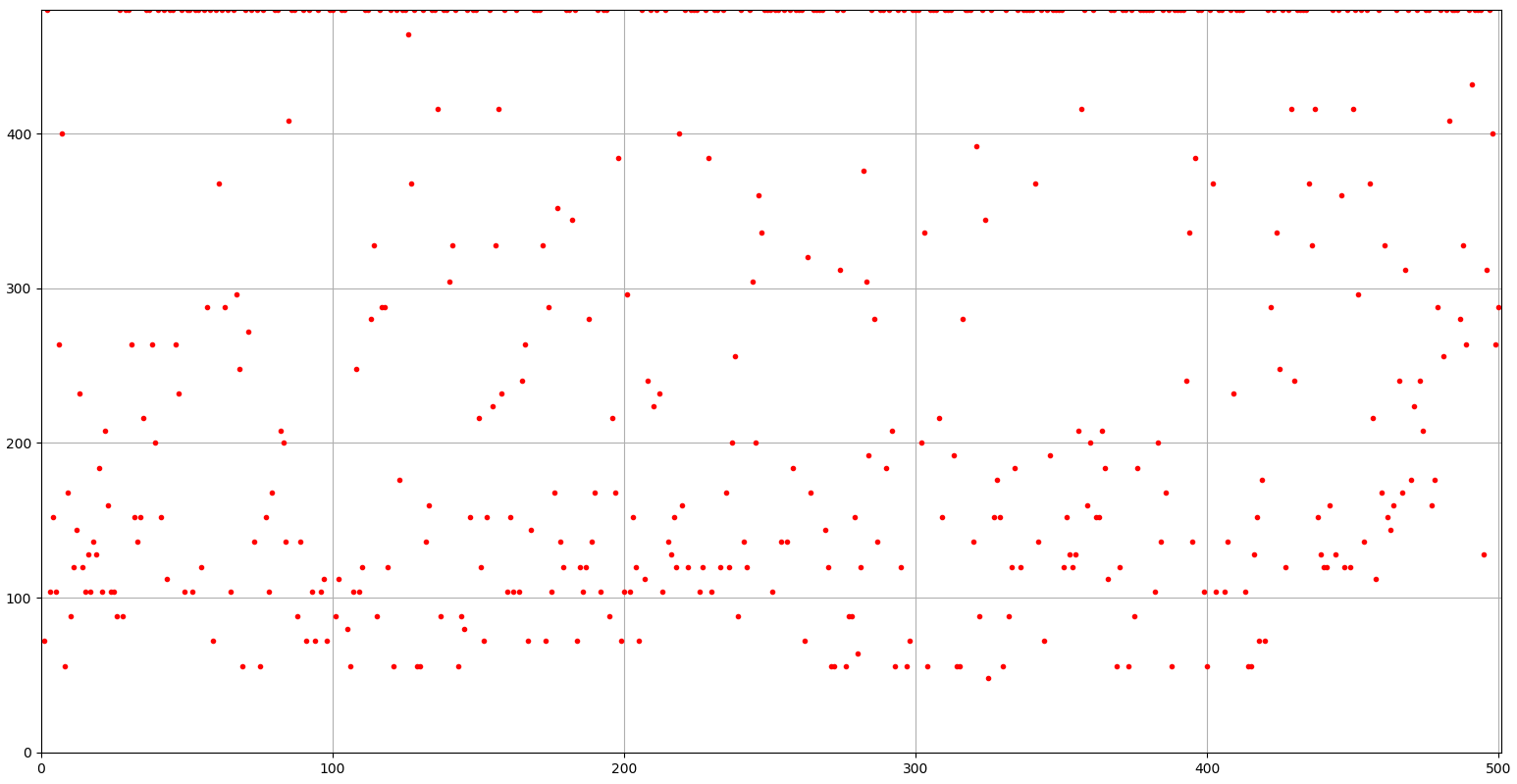
**Selfplay from checkpoint2 with only 32 data each game batch len(x) no heuristic (MCTS (512,10,1e-3); no decay)**

* 0-1000

****

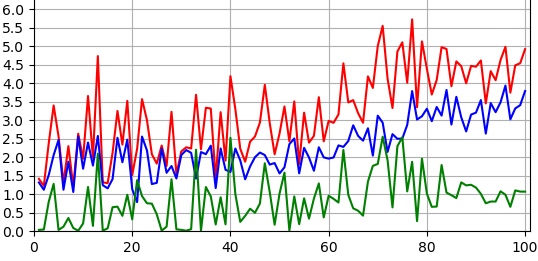
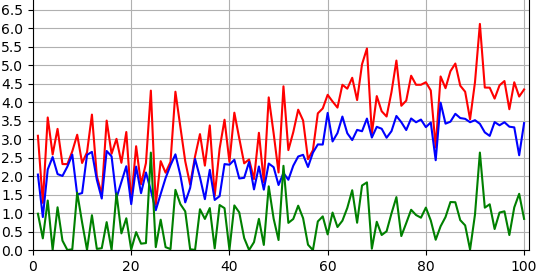
**Selfplay from checkpoint2 with only 32 data each game batch len(x) heuristic5 (MCTS (512,10,1e-3); no decay)**

* 900-1400

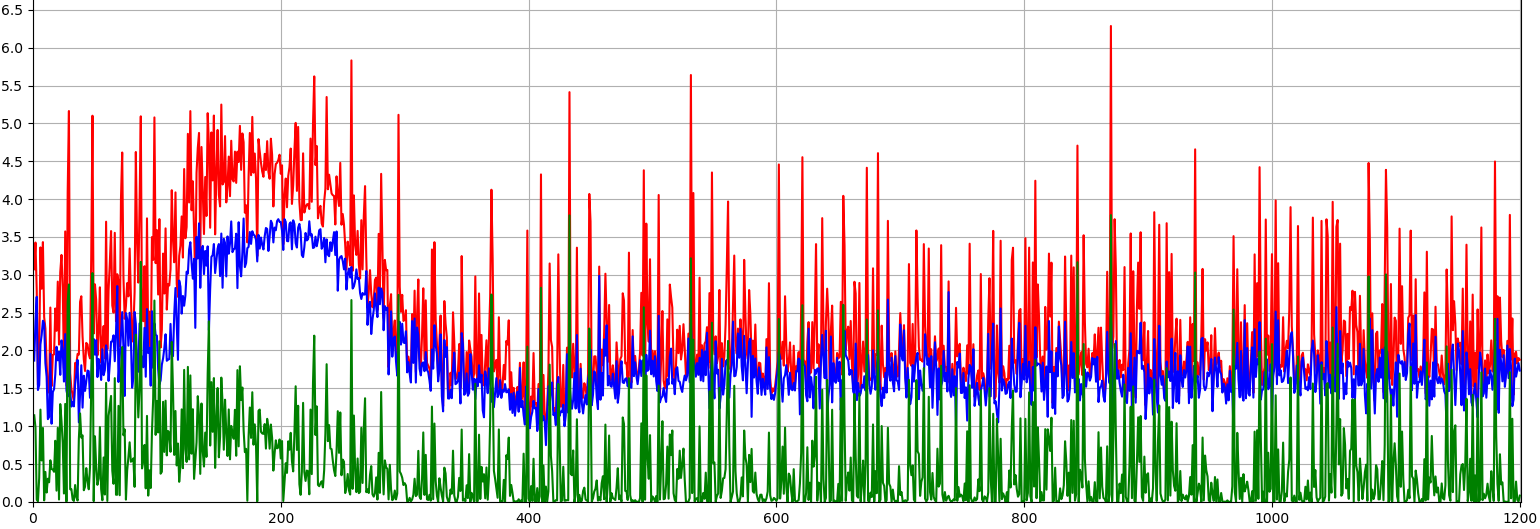


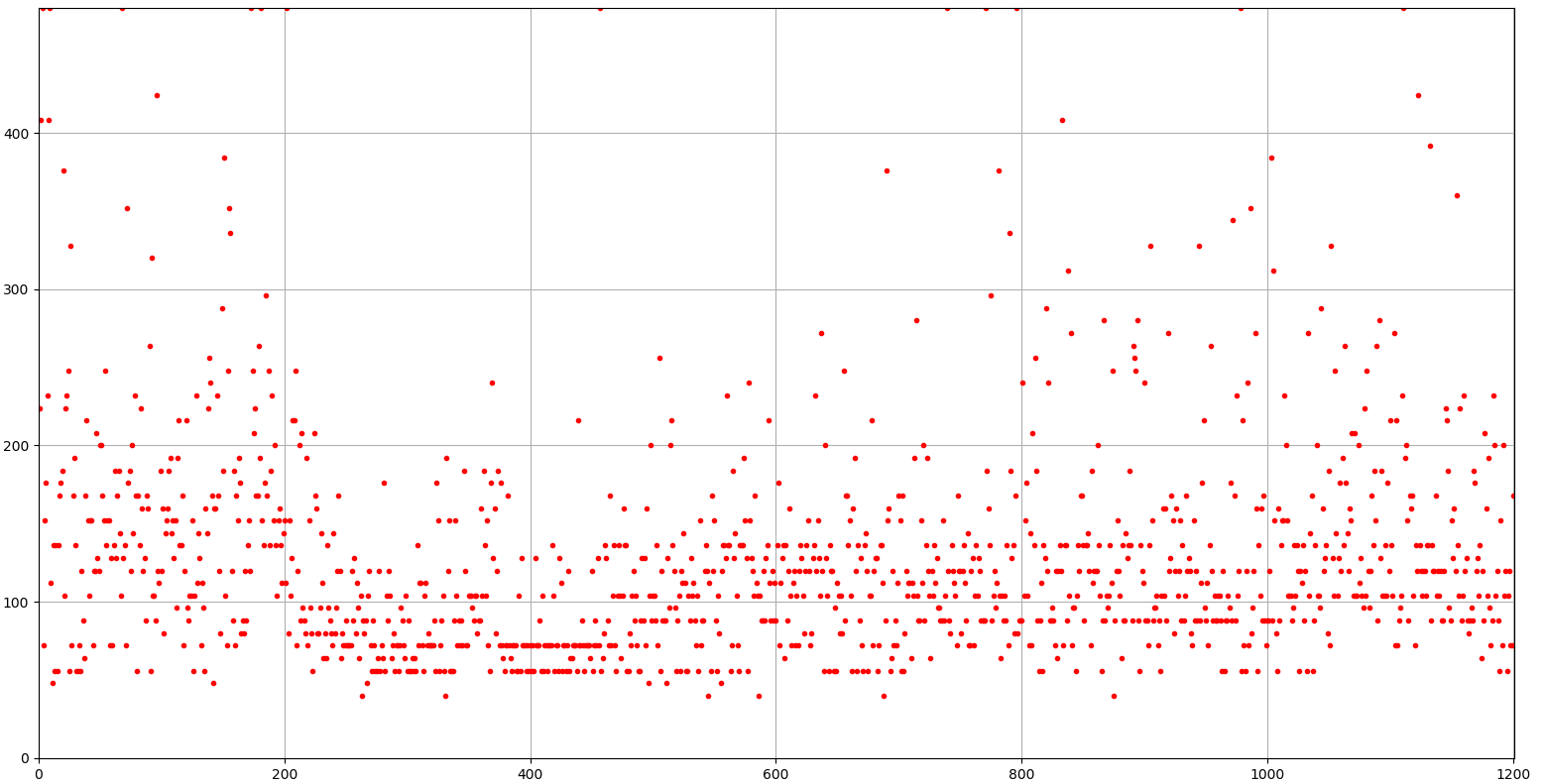
**Selfplay from checkpoint2 with only 32 data each game batch len(x) no heuristic (MCTS (512,10,1e-3); no decay)**

* 1300-

****

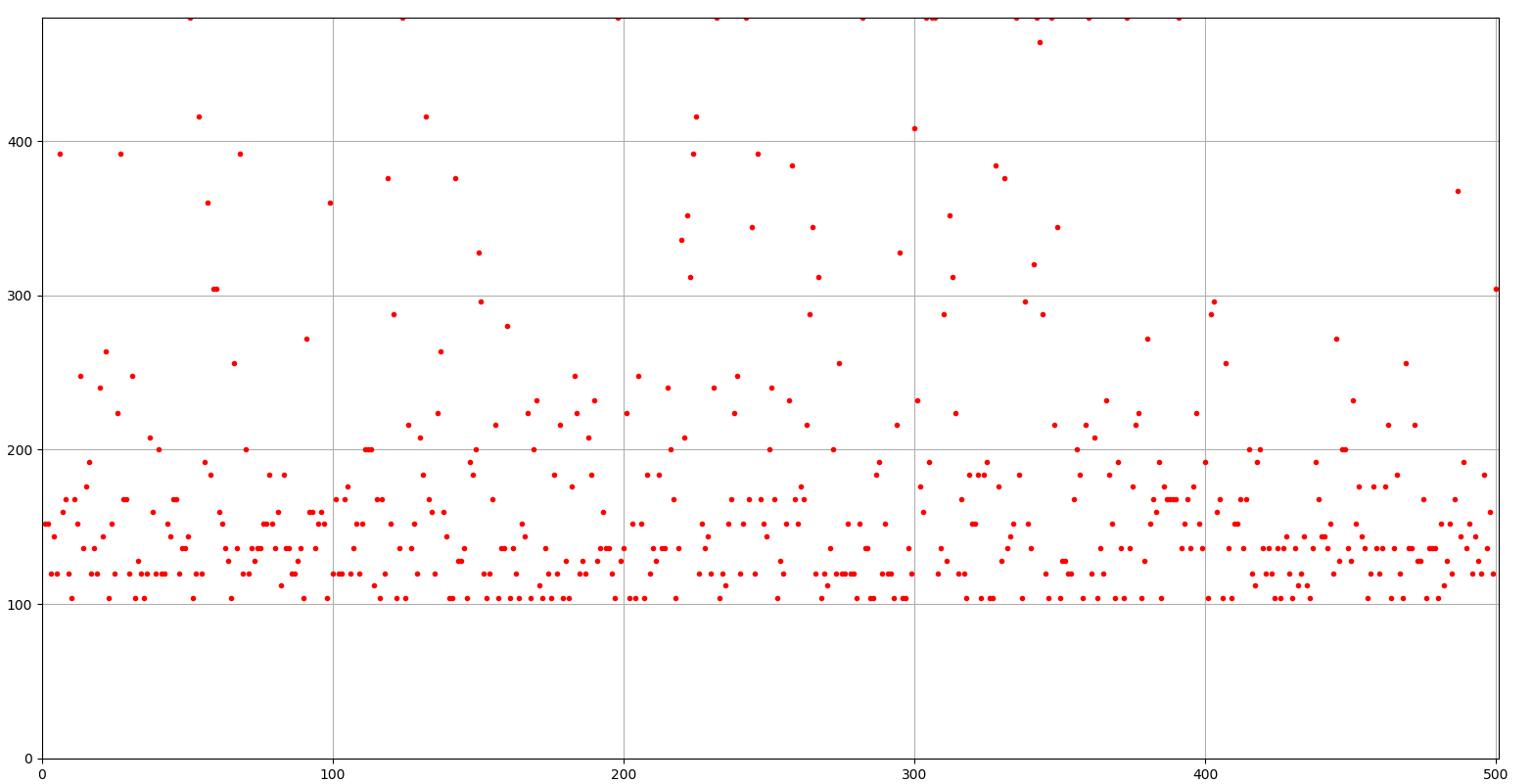
* 1300-

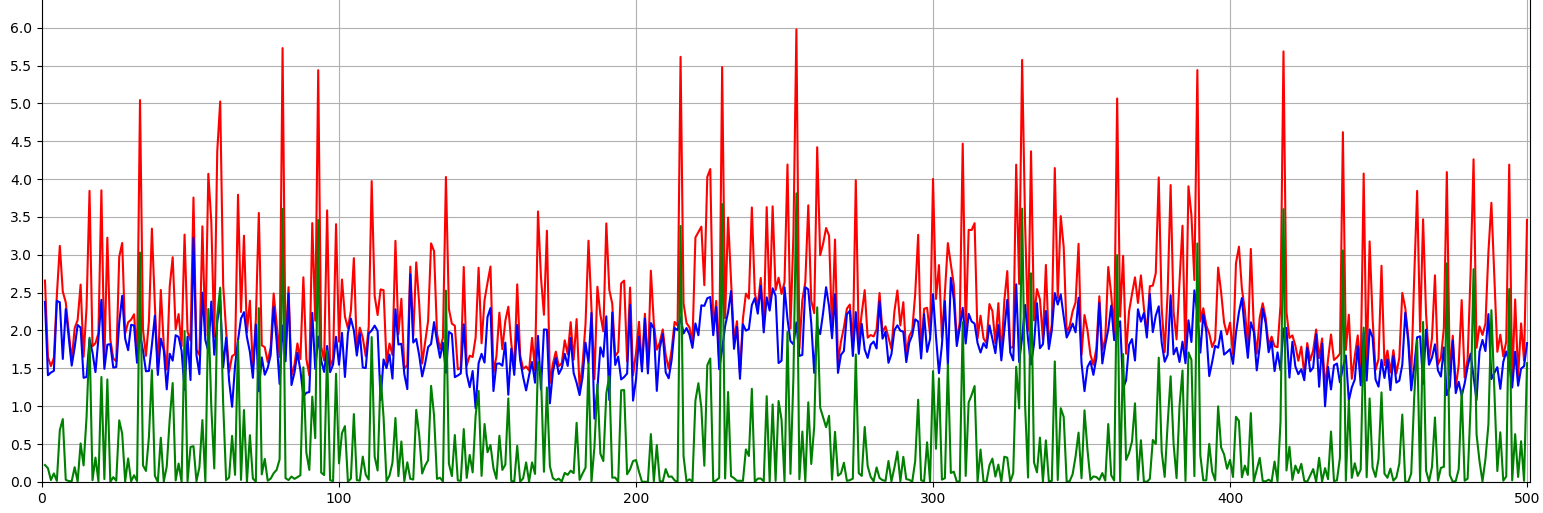




**Selfplay from checkpoint2 with only 32 data each game batch len(x) no heuristic only 100+ (MCTS (512,10,1e-3); no decay)**

* 2500

****

****

**Selfplay choose better model by competing 10 rounds every 100 rounds?**

**Fine Tuning from checkpoint2**

**Round1 policy head only; data 100 rounds; MCTS (512,10,1e-3); no decay; heuristic5; with draw**

**If better**

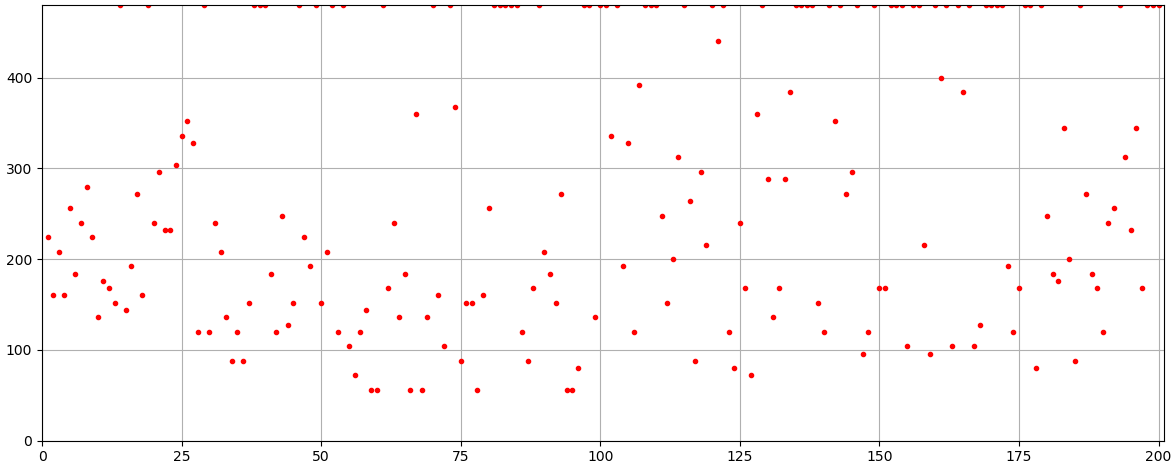
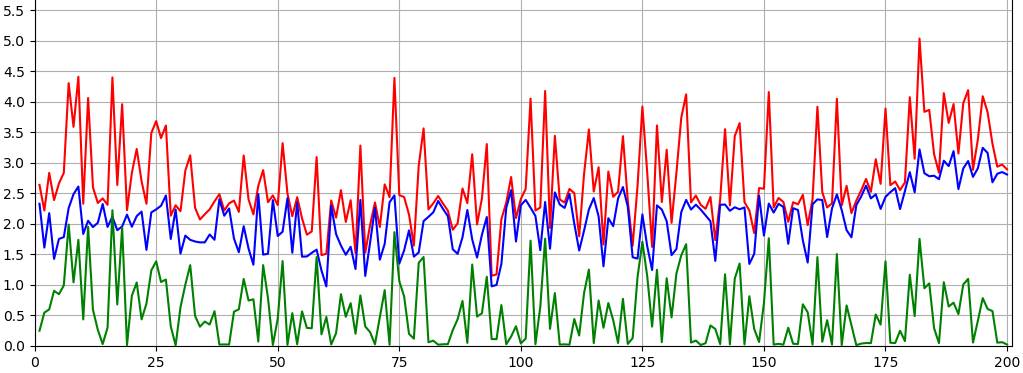
**Round1 value head only; data 200 rounds <=10 datapoints per round; MCTS (512,10,1e-3); no decay; non heuristic; with draw**

**If better**

**Round2**

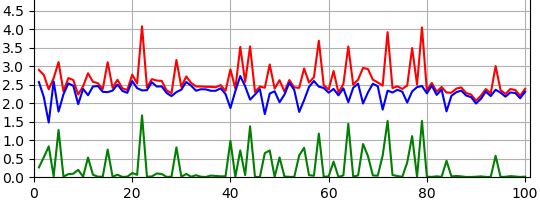
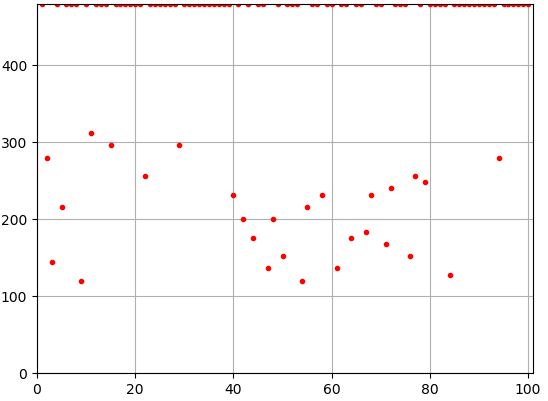
**Selfplay from checkpoint2**

**Batch Len(X); MCTS (512,10,1e-3); no decay; heuristic with draw**

* 200: doesn’t work

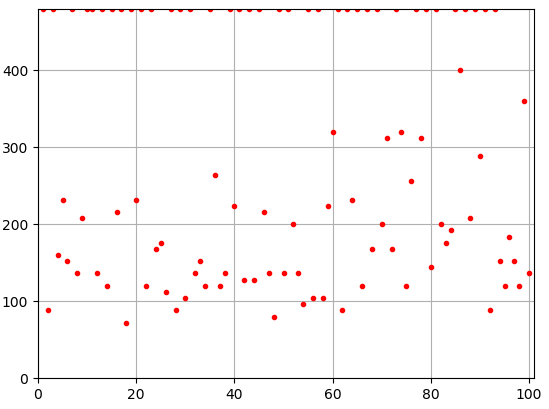
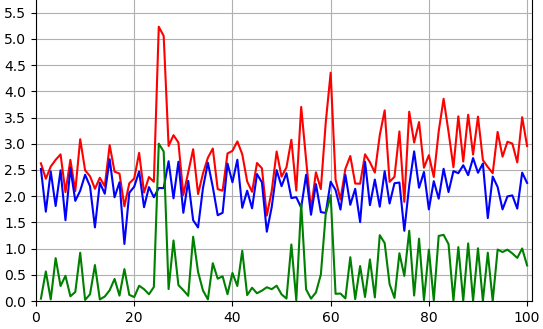
**Batch Len(X); MCTS (256,5,1e-3); no decay; heuristic2 with draw**

* 100: not good, decayed

****

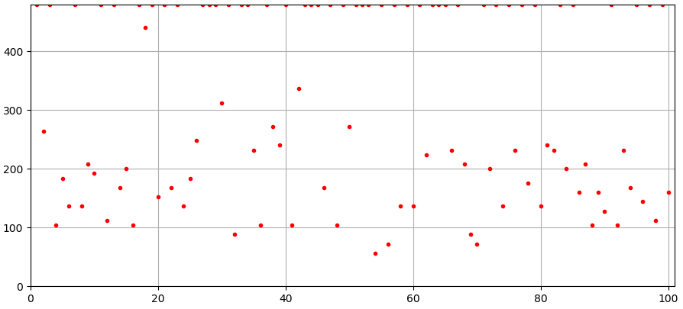
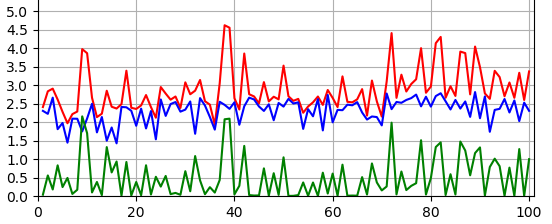
**Batch 32 for non-draw Len(X) for draw; MCTS (512,10,1e-3); no decay; 1/2heuristic3 with draw**

* 100: not good, decayed



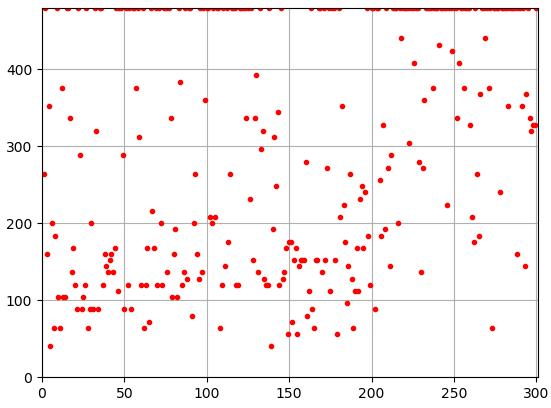
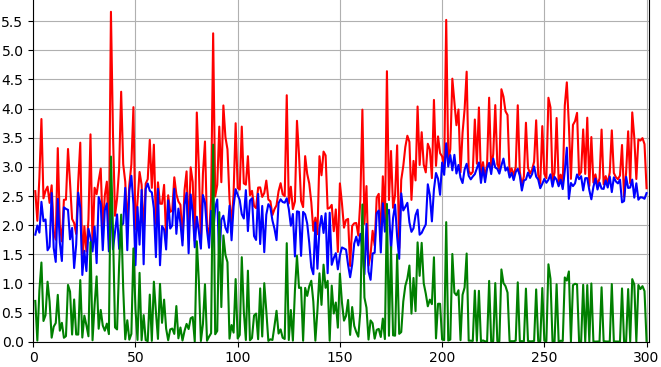
**Batch 32 for non-draw Len(X) for draw; MCTS (512,10,1e-3); no decay; 1/2heuristic4 with draw**

* 100: decayed

****

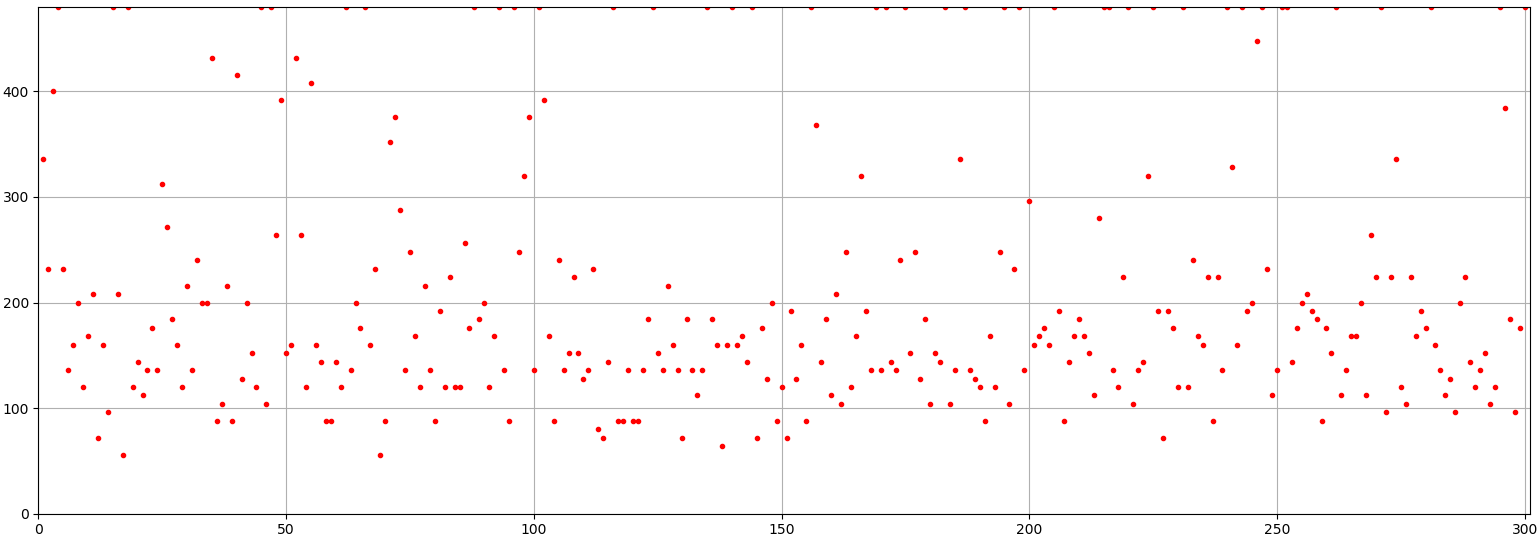
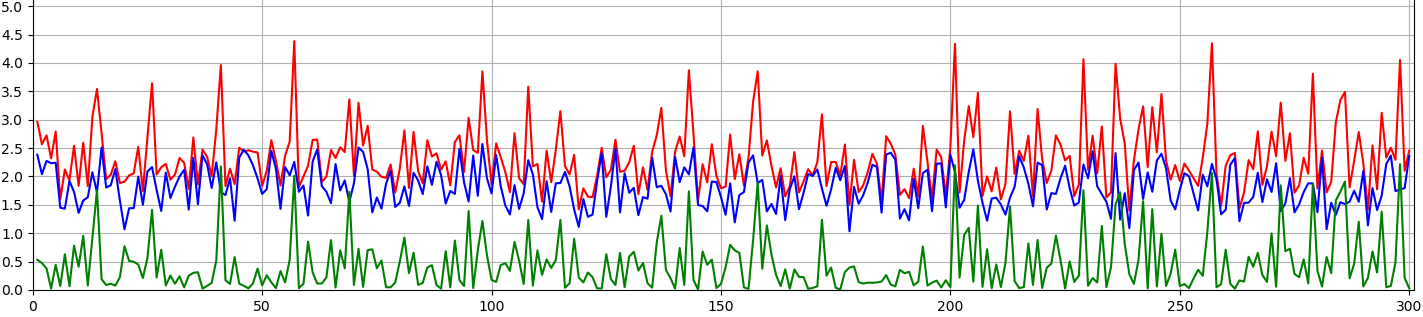
**Batch 32 for non-draw Len(X) for draw; MCTS (512,10,1e-3); no decay; heuristic5 with draw**

* 300: decayed



**Train policy head only**

**Batch 32; MCTS (512,10,1e-3); decay 0.95; heuristic with draw**

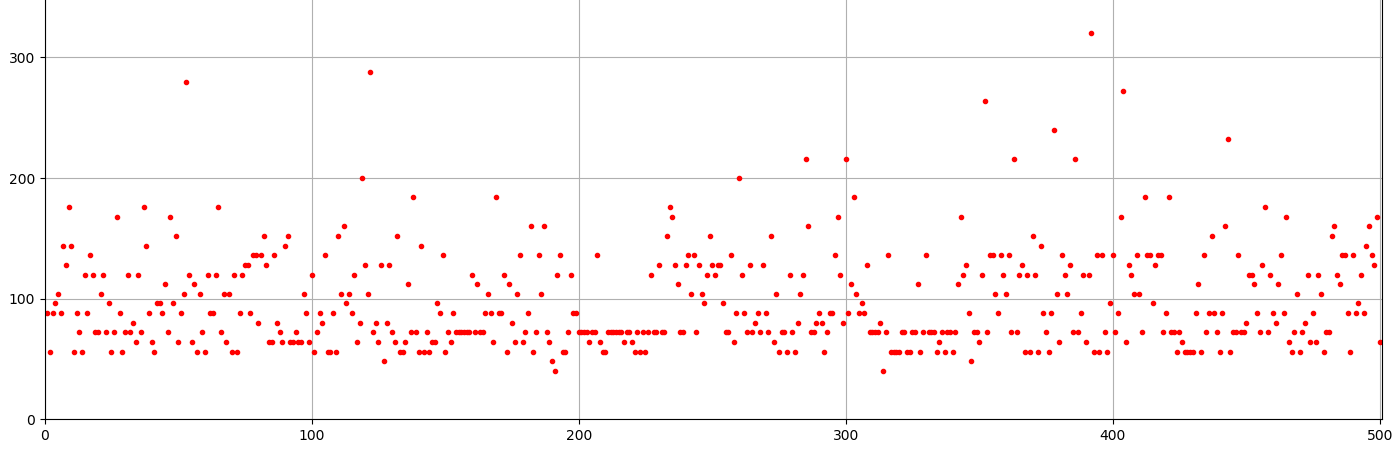
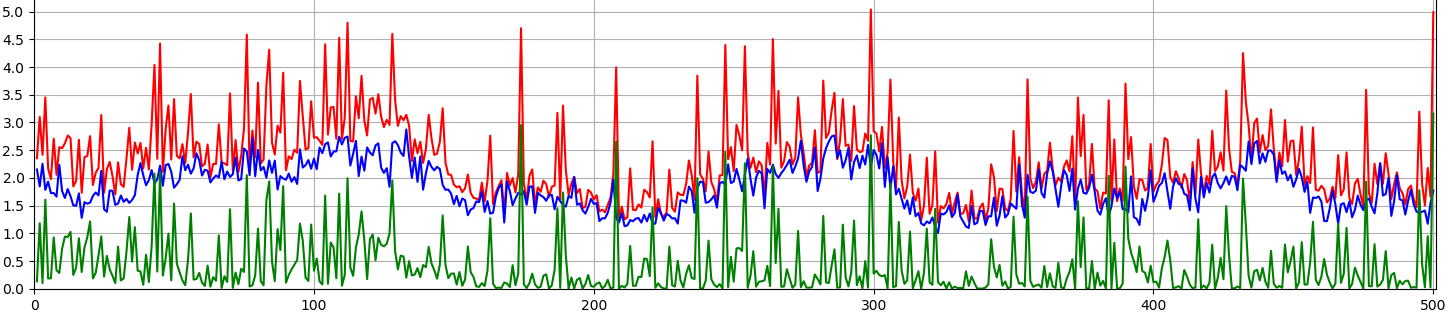
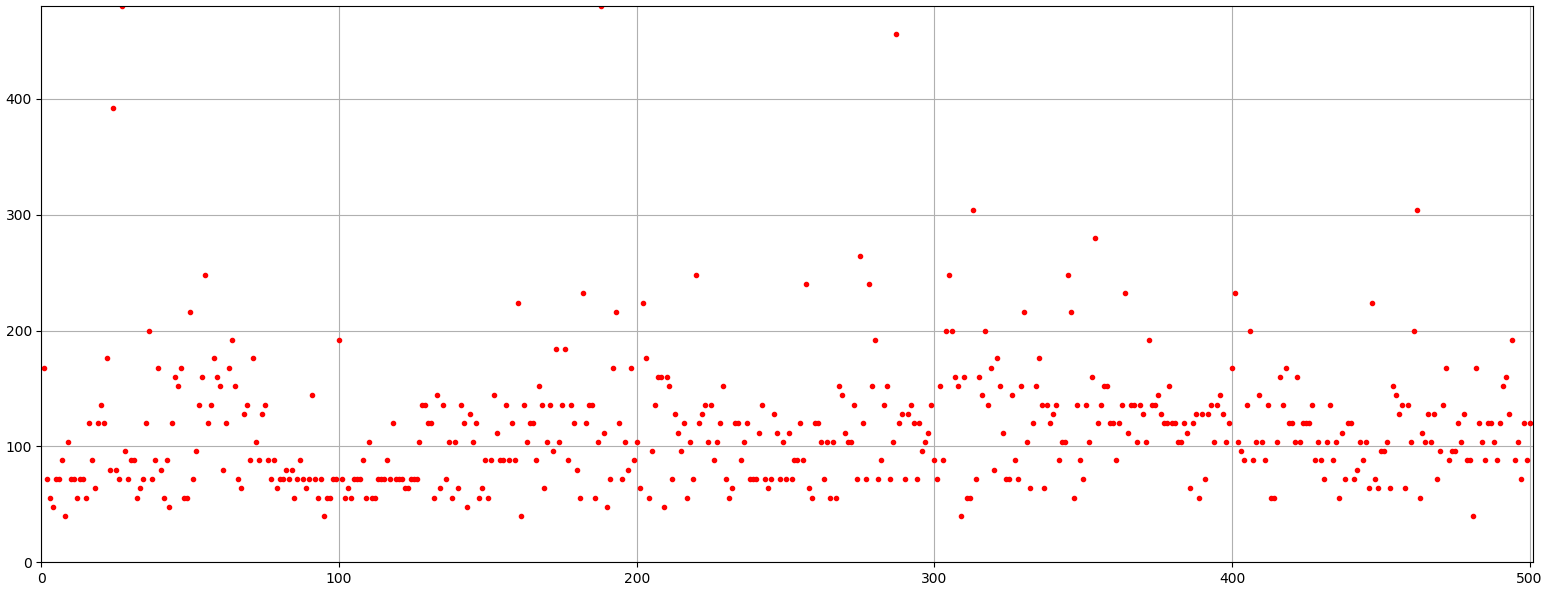
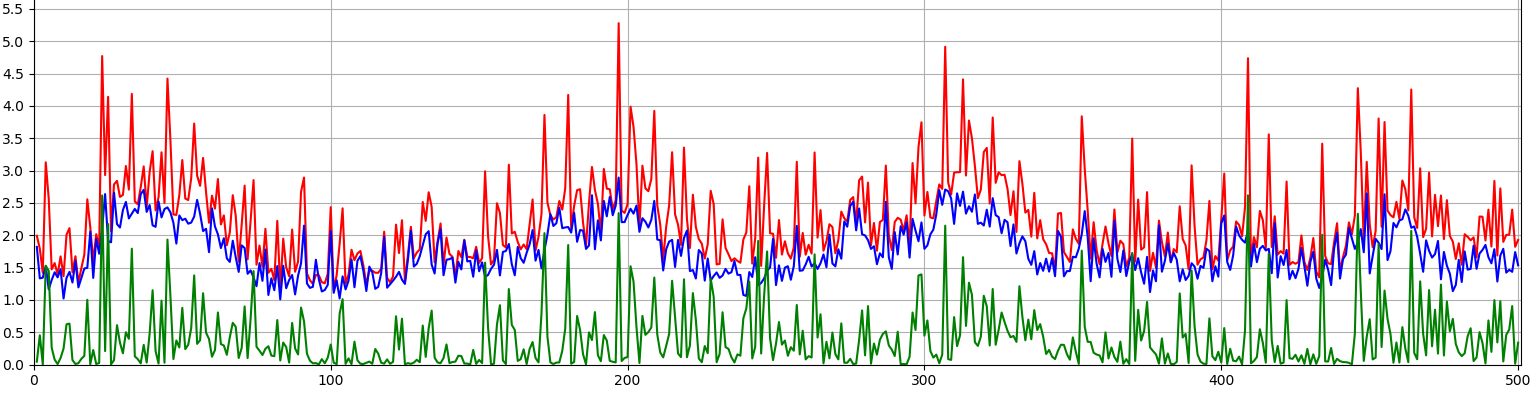
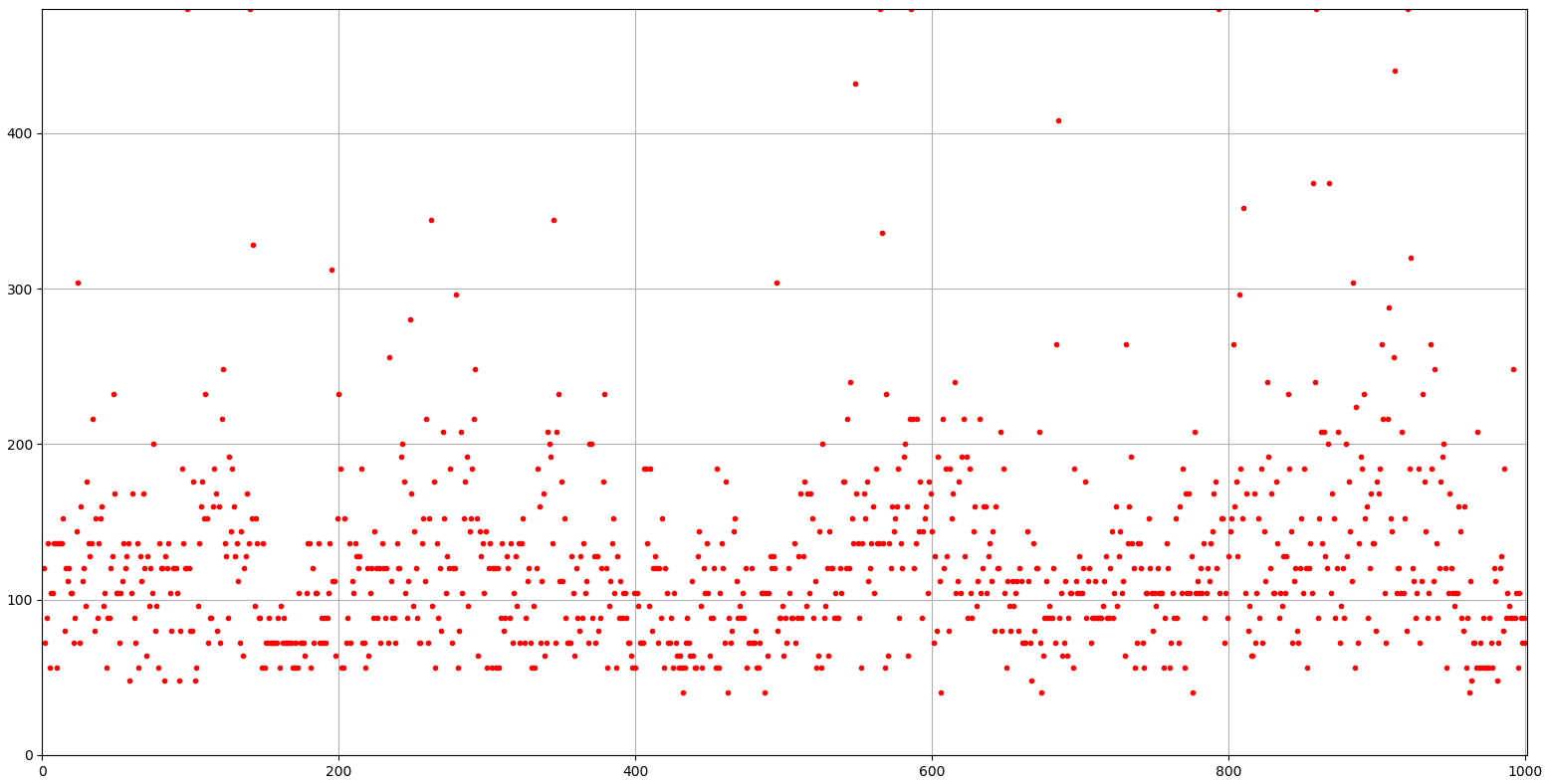
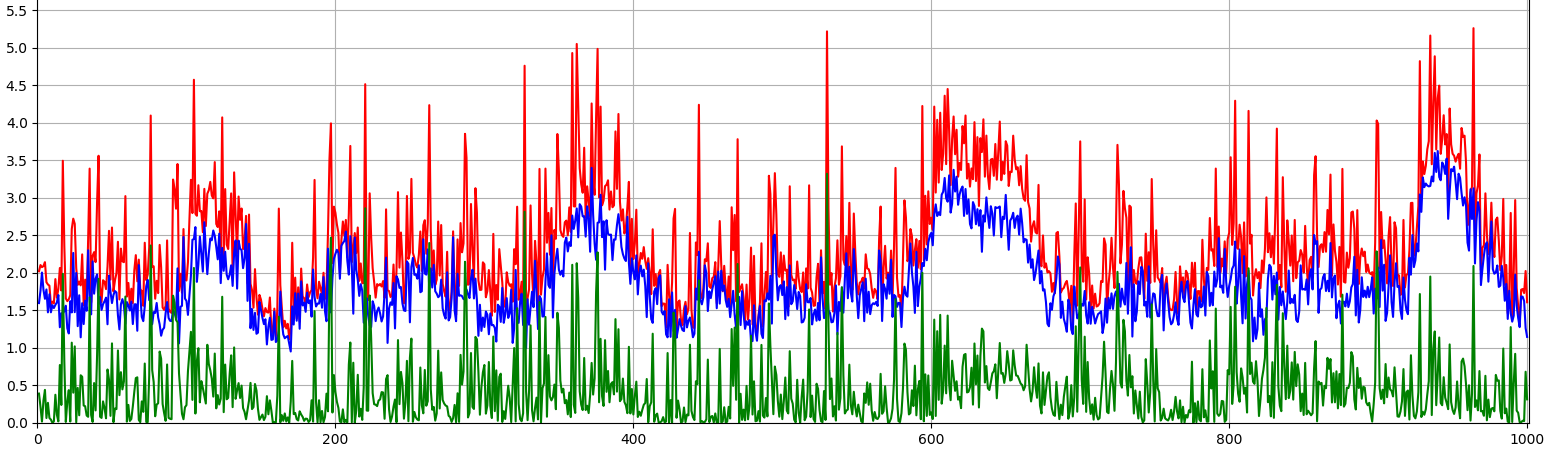
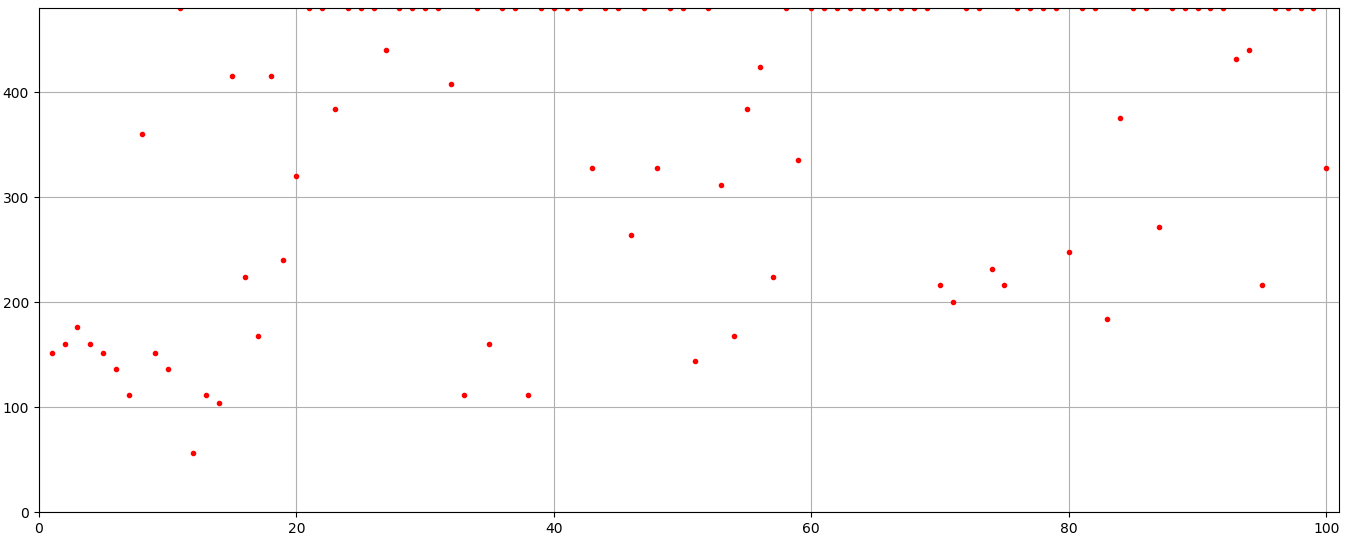
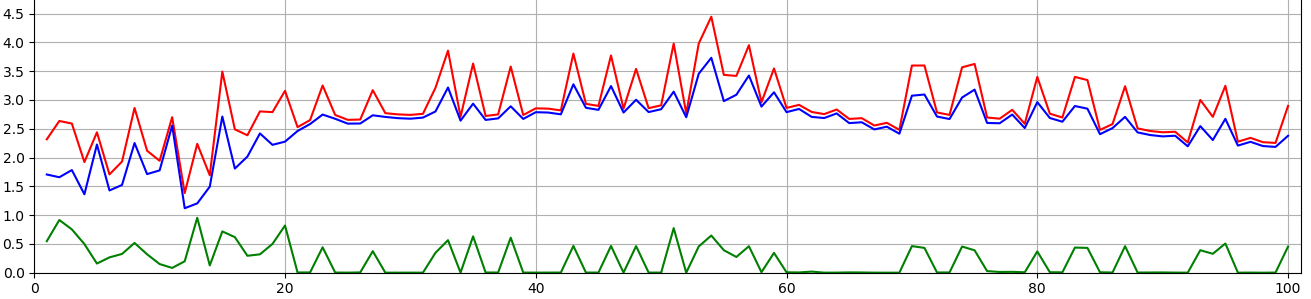
* 300: no significant improvement

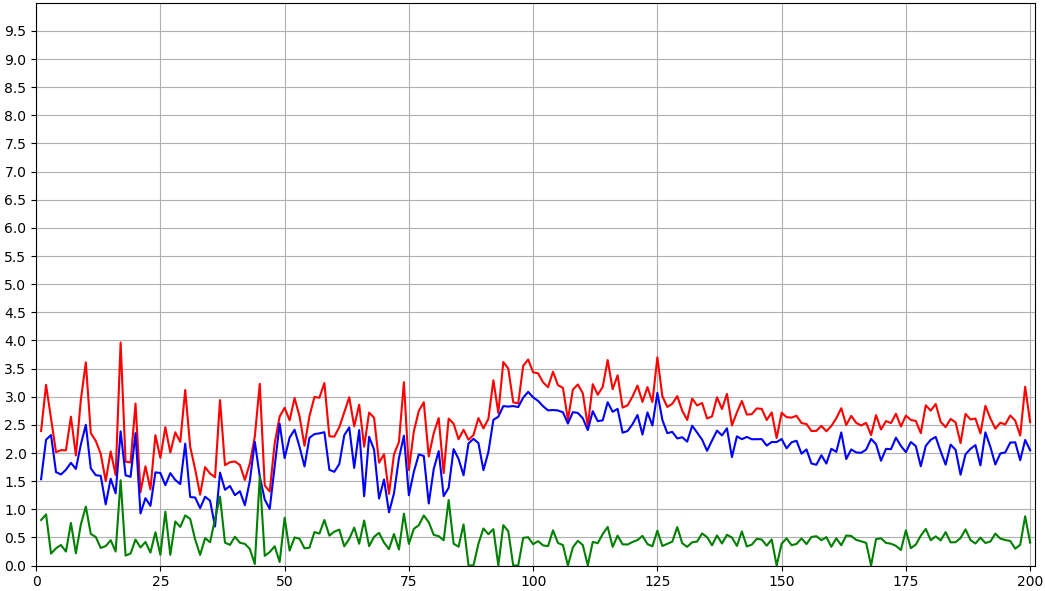
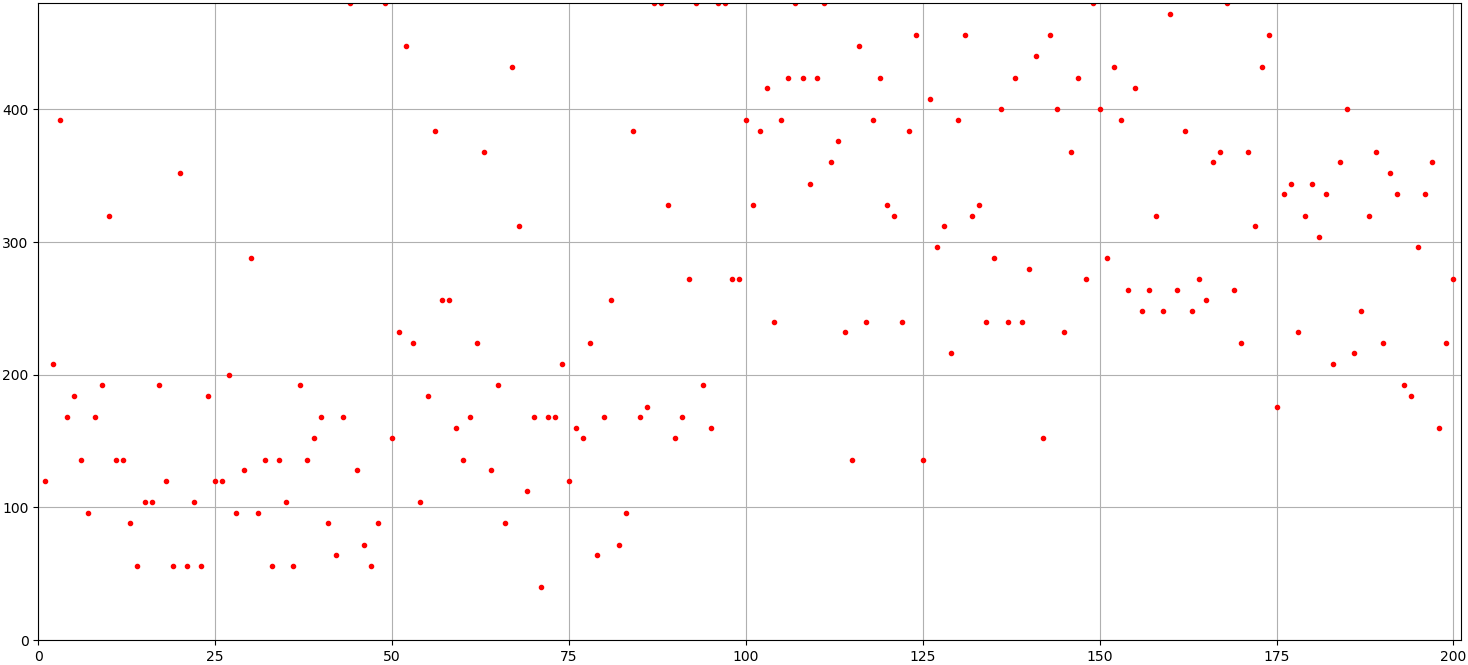
**Train value head only**

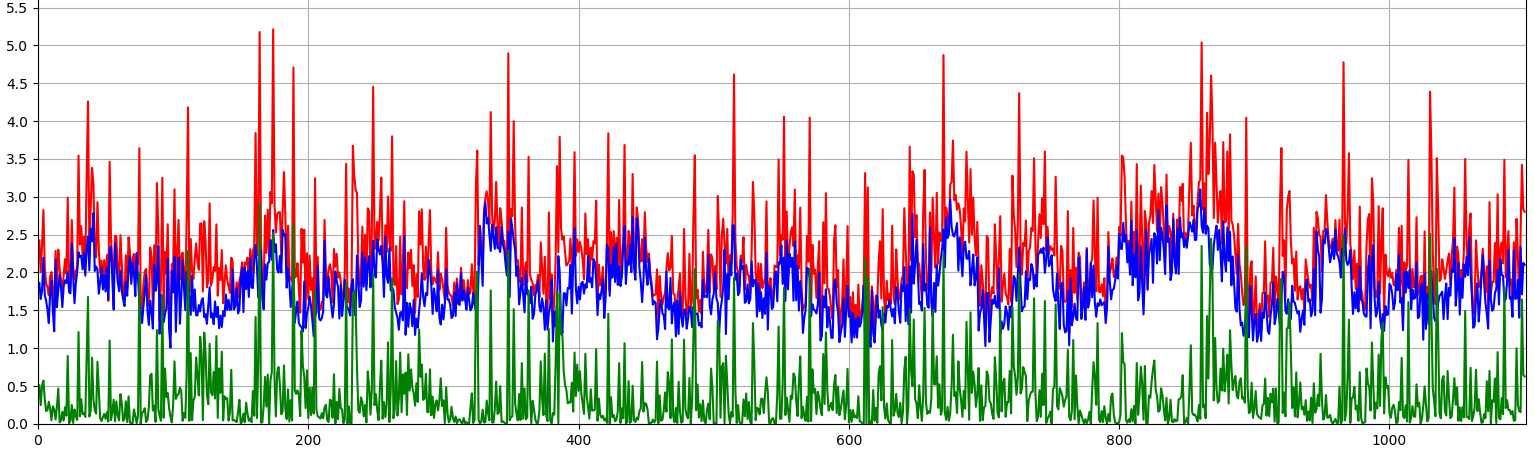
**Batch 32; MCTS (512,10,1e-3); decay 0.95; non-heuristic with draw**

**Selfplay from checkpoint1**

**Batch 32; MCTS (512,10,1e-3); decay 0.95; with draw**

* 0-500: ****
* 500-1000:
* 1000-2000:
* Start from 1900 with heuristic function: too many draws, did not perform well – 10:-10
* Start again from 1900 with heuristic, take 1/10 of draw:

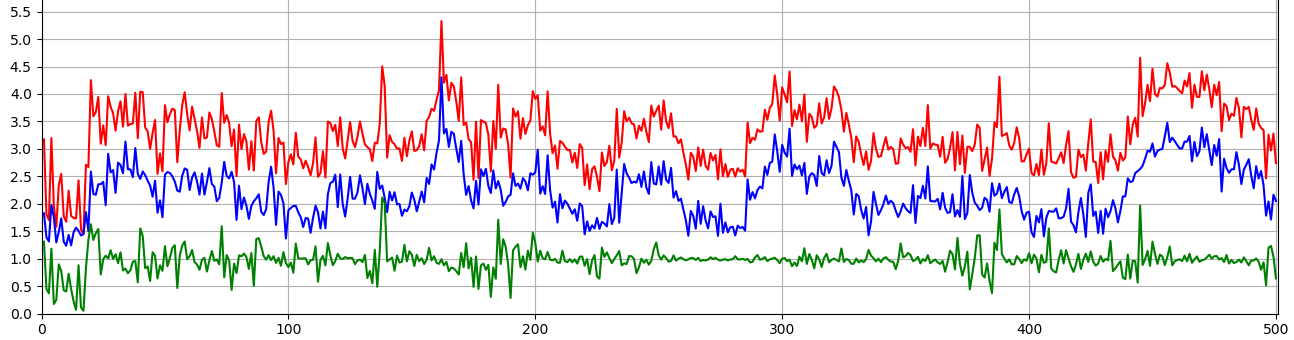
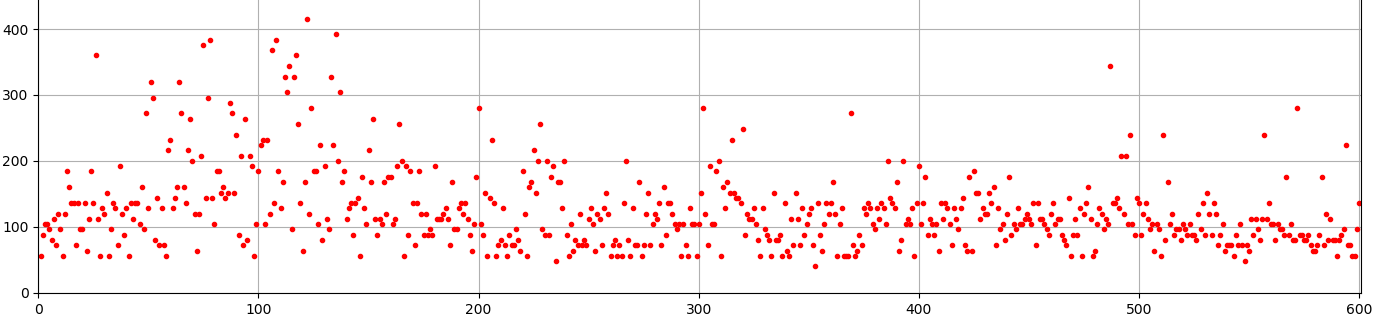
****

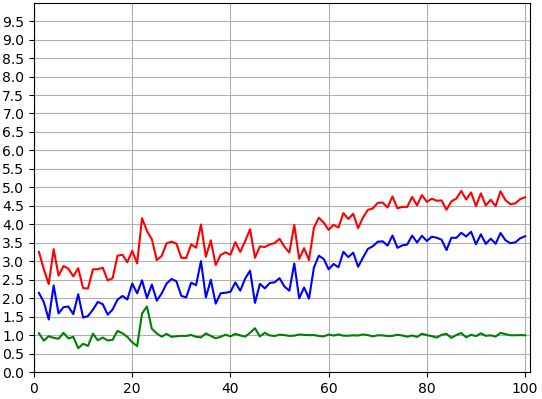
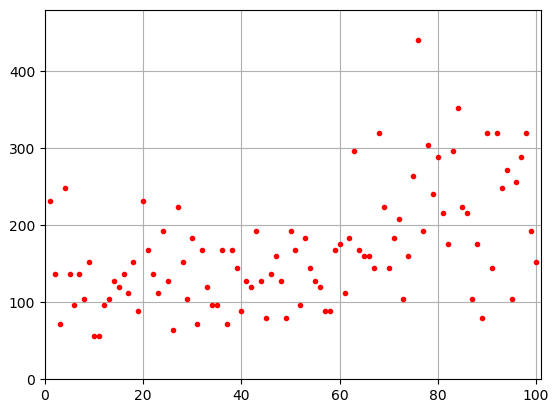
* Rid of heuristic 1900-3000: 

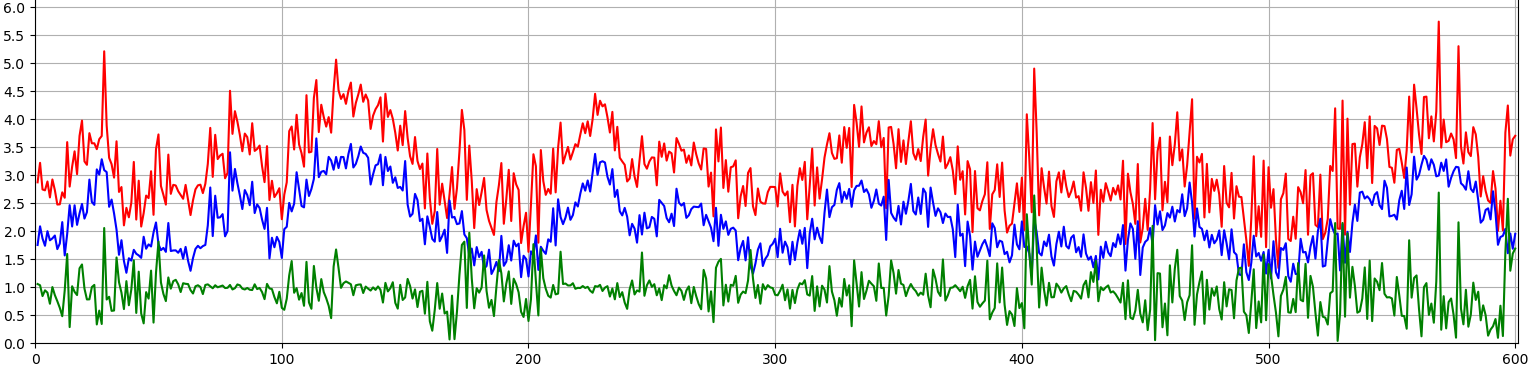
**ROUND3 selfplay**

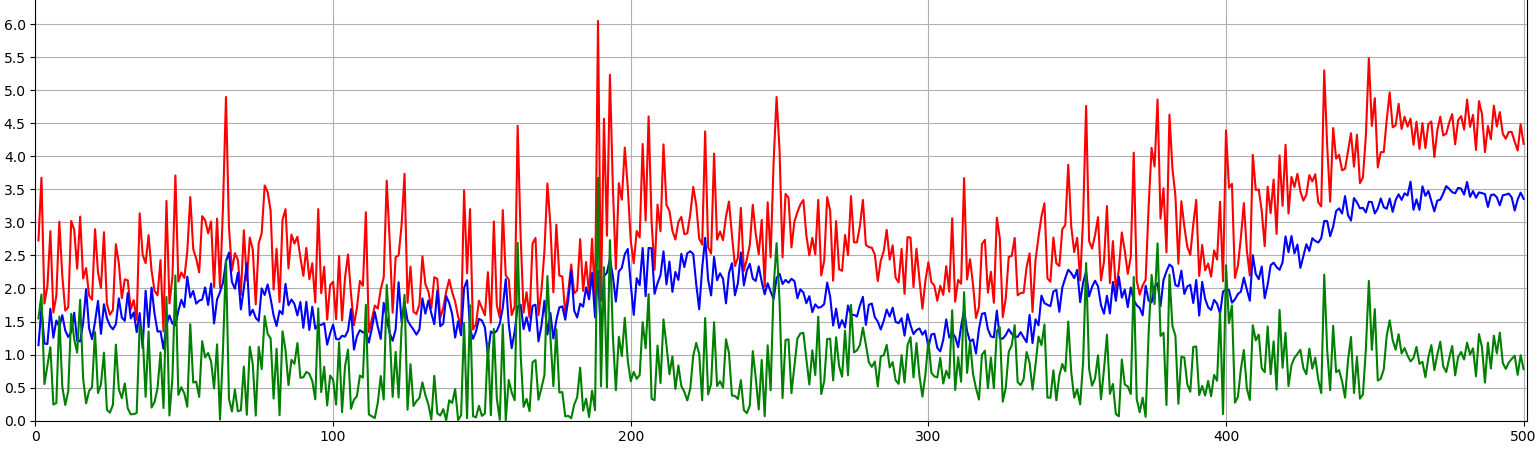
**batch 32 selfplay without draw**

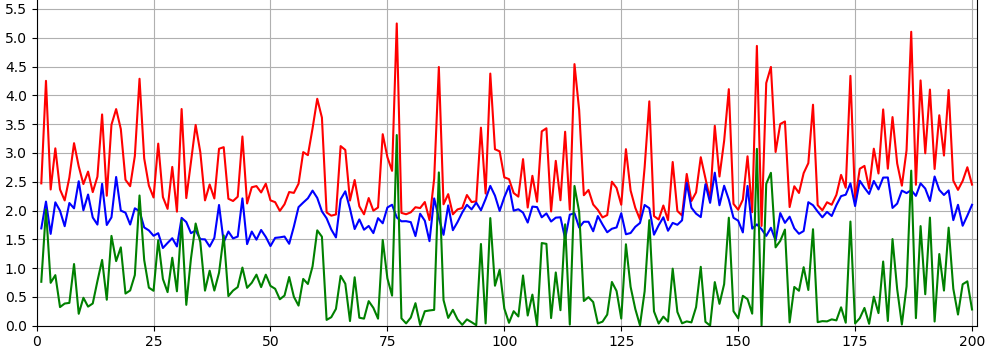
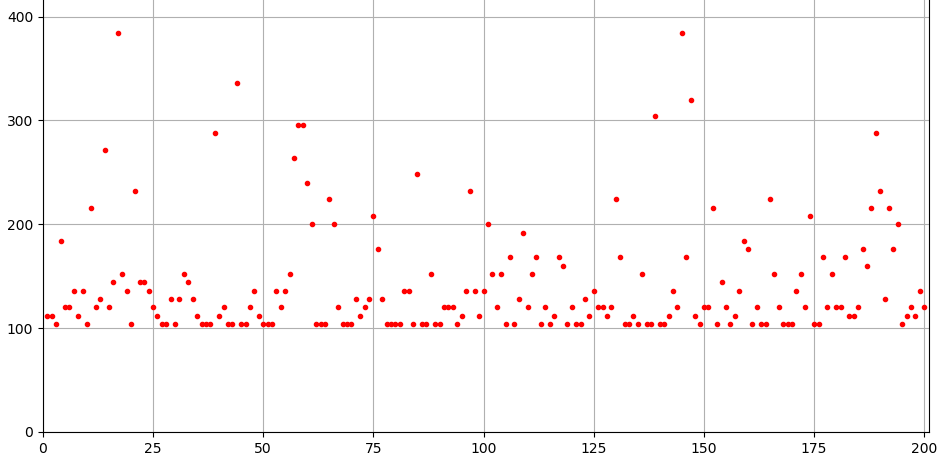
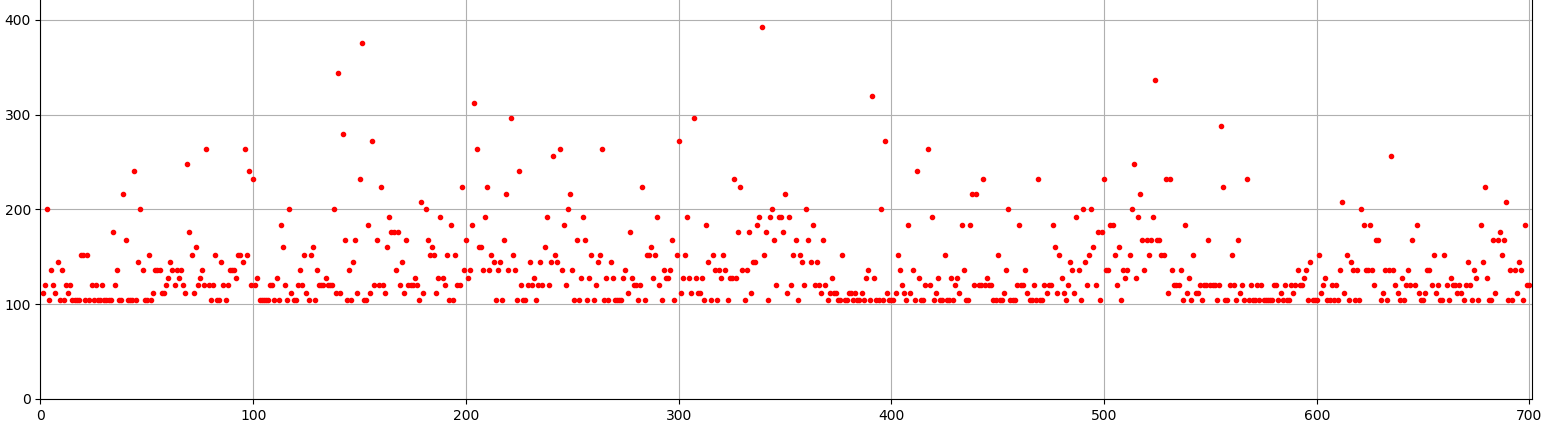
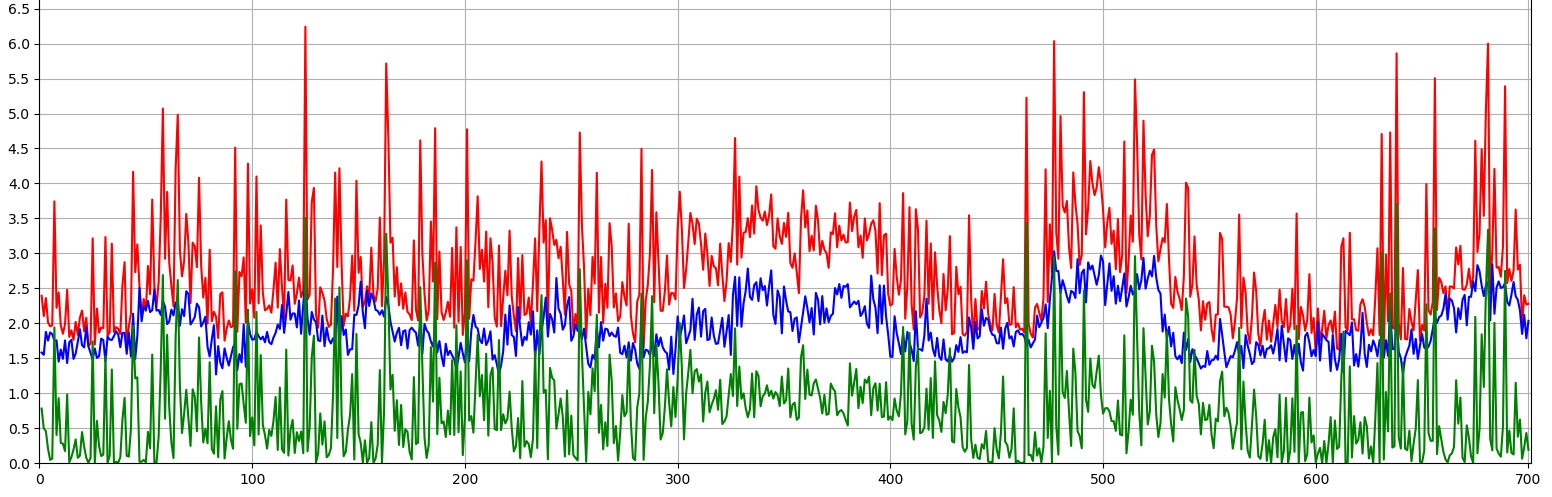
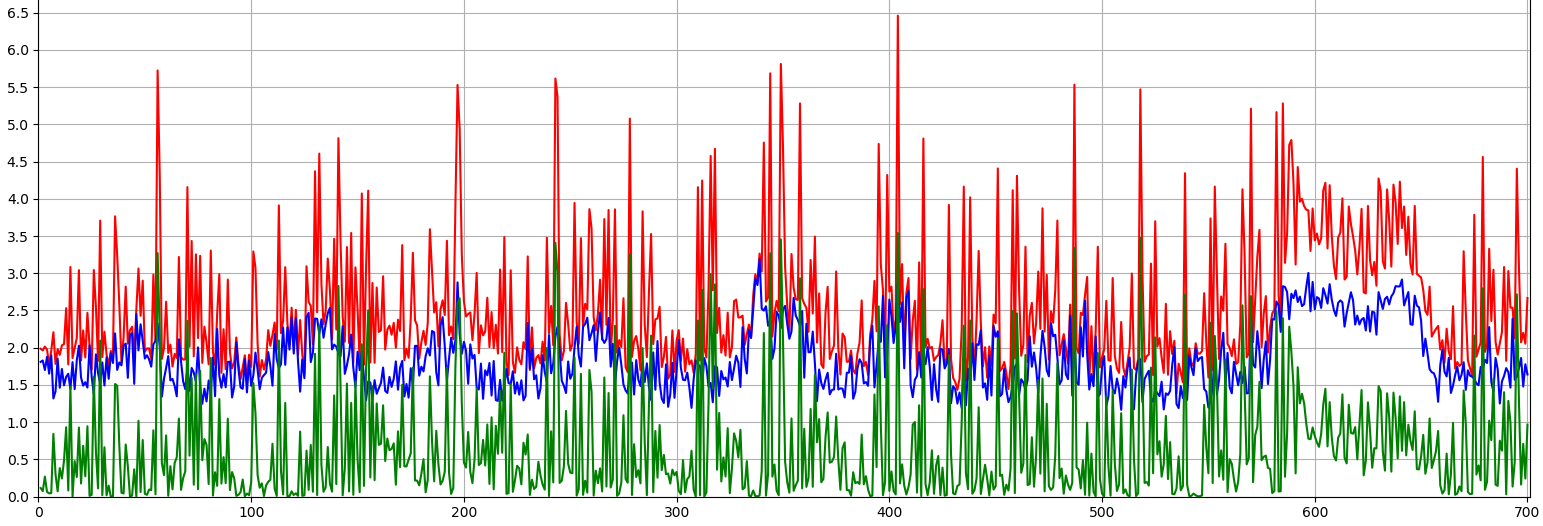
**bn: 1000**

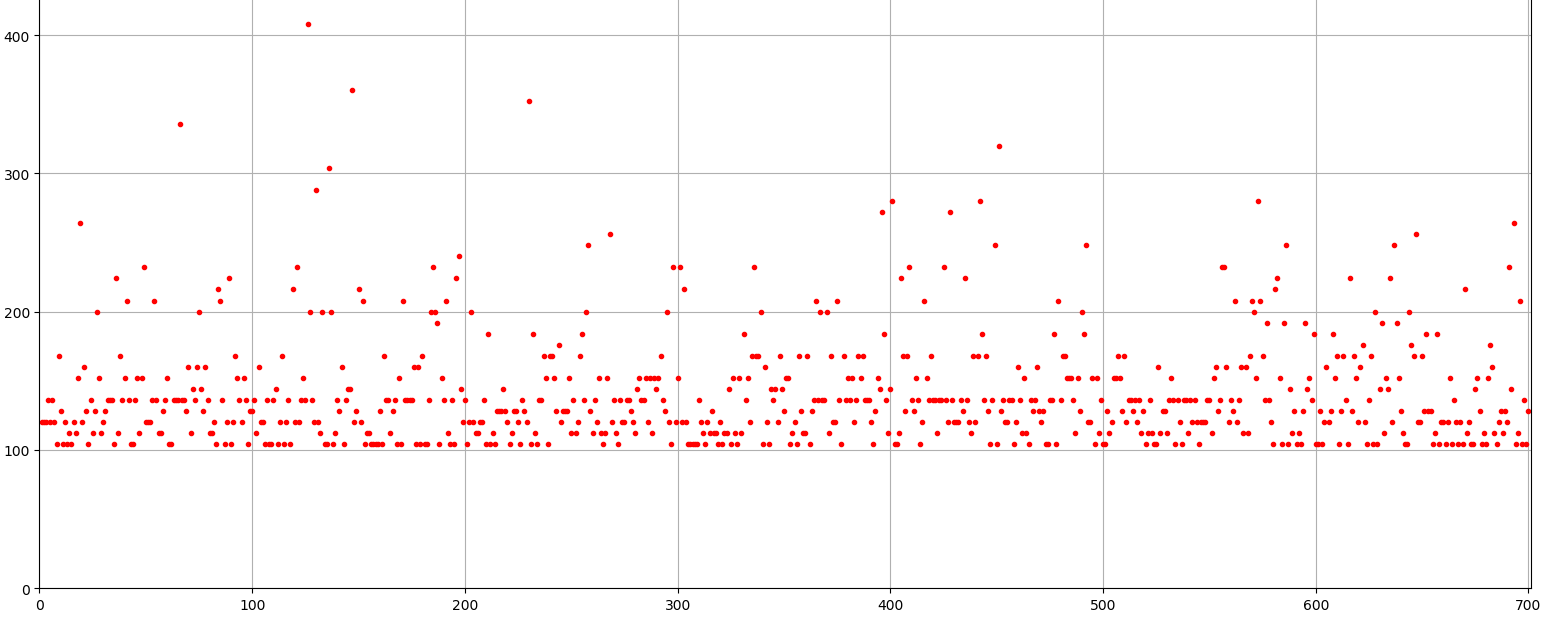
* policy parabola shaped, decreased from 4~ to 1~ and stayed at ~1-2 for 250-550 and goes back up to 2~; value ok up till 500, decayed to ~0 for 600-700
* 500 vs 700: -2 -6
* At 500: knows attack, cannot defense at all (few exceptions)
* restart from 500: value ok, best at 900, know little defense
* restart from 900: 
* restart from 900 again:



* again: know little more defense
* start from 1500 with Adam set back to default (alpha=0.001 instead of 0.002) and 1024,20,1e-2: best at 1900 knows slightly more defense



* continue from 1900 with 512, 10; round only 100 up:
* Continue from 2100:
* From 2800:



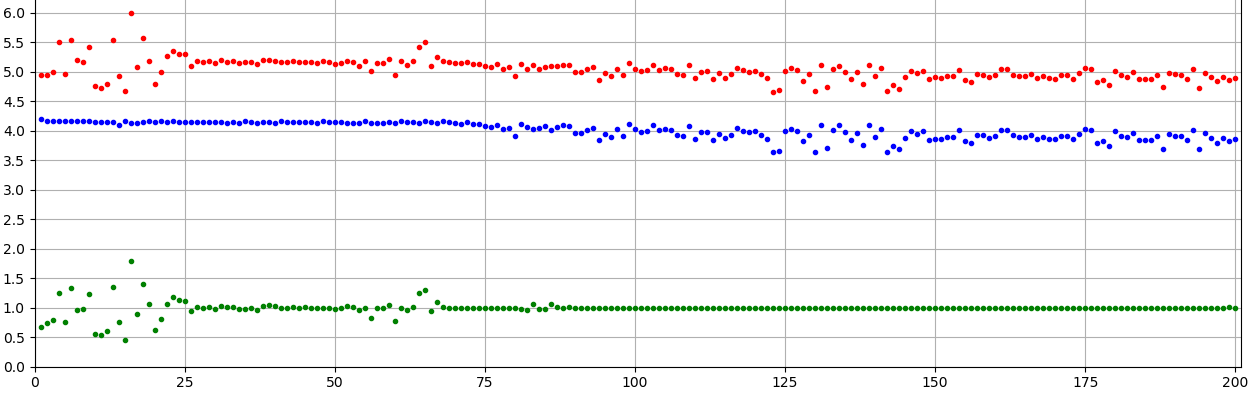
* **Make checkpoint1 at 3000**
  + **S1 (400 games)**: MCTS (512,10,1e-3); decay 0.98 without draw: improved but round length still ~100, still not learning defense well
  + **S2 (500 Games)**: MCTS (1024,10,1e-3); decay 0.95 without draw: improved but round length still ~100, knows defense a little better
  + **S3 (500 Games)**: MCTS (512,10,1e-3); decay 0.95 with draw: similar to s2
  + **S4 ()**: MCTS (1024,5,1e-3); decay 0.95 with draw: ?
  + #decided on S3
* **?**

**bn mimic: 300**

* Attempt1: value soon decays to 0 (~50 plays) and policy stuck ~4.15
* Attempt2: policy parabola shaped, decreased from ~4 to ~3 for 100-200 and goes back up to ~4 and goes back down slightly with value diminished but not stuck for 200-300
  + At 300: knows attack but not good, almost no defense

**batch 32 selfplay without value without draw**

**bn:** value decays

****

**batch 32 selfplay without value with draw**

**bn:** value decays

****

**ROUND 2 score mentor**

**Mentor:**

* loss: 4.1763 - dense\_loss: 3.7460 - dense\_2\_loss: 0.3793
* loss: 4.3153 - dense\_loss: 3.8421 - dense\_2\_loss: 0.4251

**New:**

* loss: 4.0198 - dense\_loss: 3.5770 - dense\_2\_loss: 0.3757
* loss: 4.1403 - dense\_loss: 3.6718 - dense\_2\_loss: 0.4186
* selfplay: does not converge

**batch 32 train**

M

loss: 4.3259 - dense\_loss: 3.8341 - dense\_2\_loss: 0.4628

loss: 4.3321 - dense\_loss: 3.8512 - dense\_2\_loss: 0.4546

n

loss: 4.3923 - dense\_loss: 3.8796 - dense\_2\_loss: 0.4893

loss: 4.3751 - dense\_loss: 3.8711 - dense\_2\_loss: 0.4803

b (bad)

loss: 4.3392 - dense\_loss: 3.8012 - dense\_2\_loss: 0.5148

loss: 4.3617 - dense\_loss: 3.7929 - dense\_2\_loss: 0.5451

b2

loss: 4.3072 - dense\_loss: 3.7996 - dense\_2\_loss: 0.4817

loss: 4.3025 - dense\_loss: 3.7972 - dense\_2\_loss: 0.4790

**ROUND1 random**

**basic-mimic-1-3-100-1 vs basic-mimic-1-3-500-1**

[-1, 1, 1, -1, -1, -1, 1, -1, -1, -1]

[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

-4 10

0.4974849224090576 0.5786485195159912 0.7571446895599365

0.3484046459197998 0.4915498733520508 0.5887539386749268

**basic-mimic-1-3-500-1 vs basic-mimic-1-3-500-1 (value=0)**

[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

[-1, 1, -1, -1, -1, 1, 1, 1, -1, 1]

10 0

0.3990347385406494 0.5426368713378906 0.6478967666625977

0.4284639358520508 0.5364730596542359 0.7768850326538086

[1, 1, -1, 1, -1, 1, 1, 1, 1, 1]

[-1, -1, 0, -1, -1, 1, 1, -1, -1, 0]

6 -4

0.334014892578125 0.5132874965667724 0.7174293994903564

0.35282421112060547 0.5326704978942871 0.7350823879241943

**basic-mimic-1-3-500-1 vs basic-mimic-mentor-500-1**

[-1, 1, 1, 1, -1, 1, -1, 1, -1, 1]

[1, 1, 1, 1, -1, 1, 1, 1, -1, 1]

2 6

0.41414880752563477 0.5172166109085083 0.6286647319793701

0.35588836669921875 0.4641659021377563 0.6372437477111816

[1, 1, 1, 1, 0, -1, -1, -1, -1, -1]

[1, 1, 1, 1, 1, -1, -1, -1, -1, -1]

-1 0

0.391340970993042 0.4798299789428711 0.5555174350738525

0.3417356014251709 0.4872976541519165 0.5994999408721924

\* 1-3 without shortcut !!!!

**Train**

1-3 with shortcut 500: loss: 4.9195 - dense\_loss: 2.3776 - dense\_2\_loss: 0.7961

New (1-1 with new shortcut) 500: loss: 3.2302 - dense\_loss: 2.3219 - dense\_2\_loss: 0.8057

New 200: loss: 3.5362 - dense\_loss: 2.5079 - dense\_2\_loss: 0.9499

**basic-mimic-mentor-500-1 vs basic-mimic-1-3-500-1**

[-1, 1, 0, 1, -1, 1, 1, 1, -1, 1]

[1, -1, 1, 1, -1, 1, 1, 1, -1, -1]

3 2

0.40769410133361816 0.5160478115081787 0.5848267078399658

0.40119218826293945 0.573302435874939 0.706871747970581

[-1, -1, 1, -1, 1, -1, 1, 1, 1, 1]

[1, -1, -1, -1, 1, 1, -1, -1, 1, -1]

2 -2

0.5266532897949219 0.5938489198684692 0.7024517059326172

0.5769269466400146 0.6386082410812378 0.7322661876678467

**basic-mimic-1-3-500-1 vs oldbest**

[1, 1, 1, 1, 1, 1, 1, -1, 1, -1]

[1, -1, 1, -1, -1, -1, -1, 1, -1, -1]

6 -4

0.5630660057067871 0.6226654529571534 0.757455587387085

0.5238938331604004 0.6544863224029541 0.7910943031311035

**Both value=0**

[1, -1, -1, 1, -1, 1, 1, -1, 1, -1]

[1, -1, 1, 1, -1, 1, 1, -1, 1, -1]

0 2

0.3986935615539551 0.5673752307891846 0.8307795524597168

0.3450655937194824 0.5424115657806396 0.9594321250915527

**basic-mimic-mentor-500-1 vs oldbest**

[1, 1, 1, 1, 1, 1, 1, 1, 1, -1]

[-1, 1, -1, -1, -1, 1, 1, -1, 1, -1]

8 -2

0.5141105651855469 0.6367408752441406 0.7145662307739258

0.587655782699585 0.6849261283874511 0.8193473815917969

**New vs mentor**

[1, 1, -1, 1, -1, 1, 1, -1, 1, -1]

[-1, 1, 1, -1, -1, 1, 1, 0, 1, 1]

2 3

0.5342652797698975 0.661589241027832 0.7334170341491699

0.5144686698913574 0.6067826509475708 0.7041065692901611