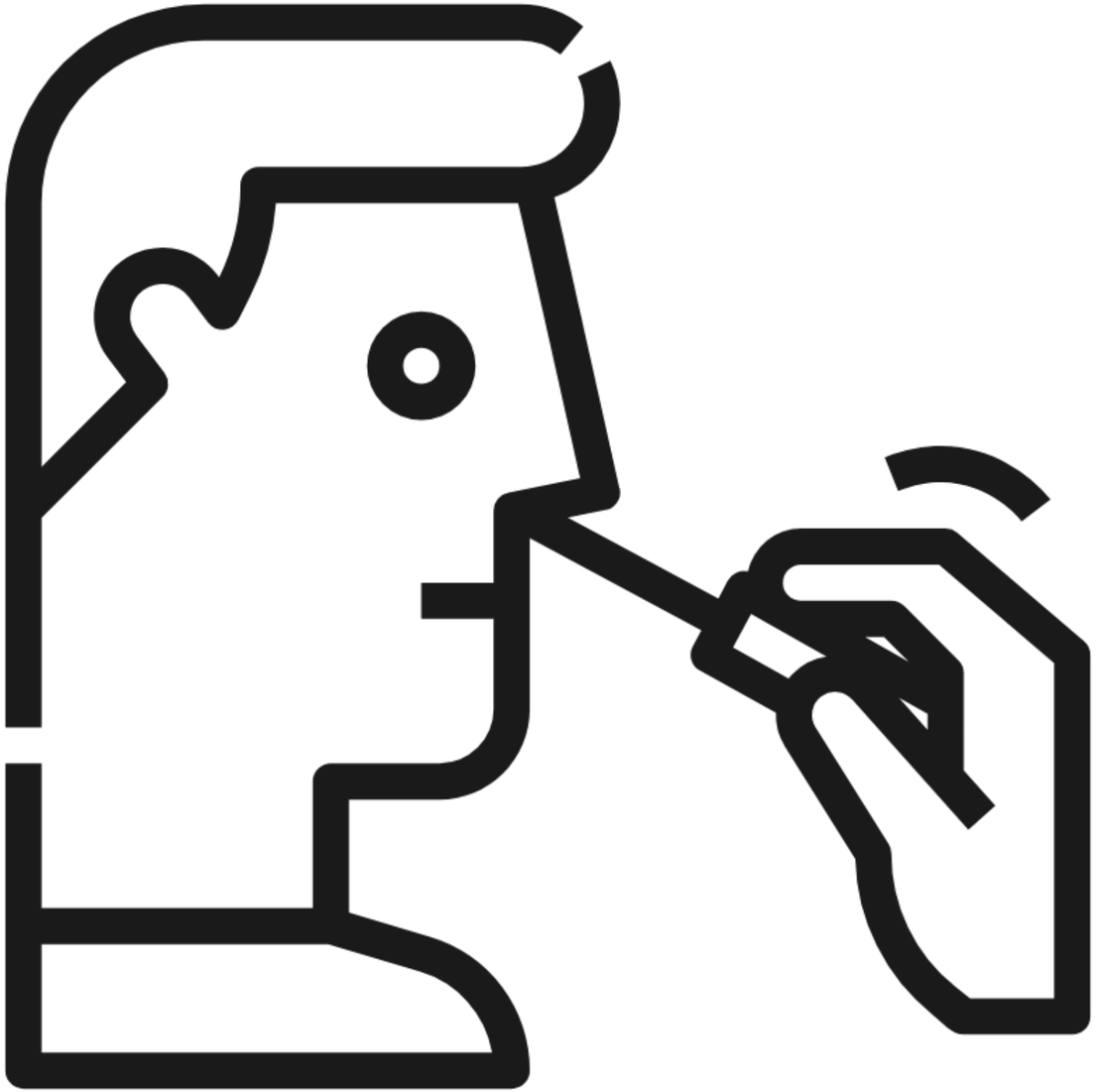
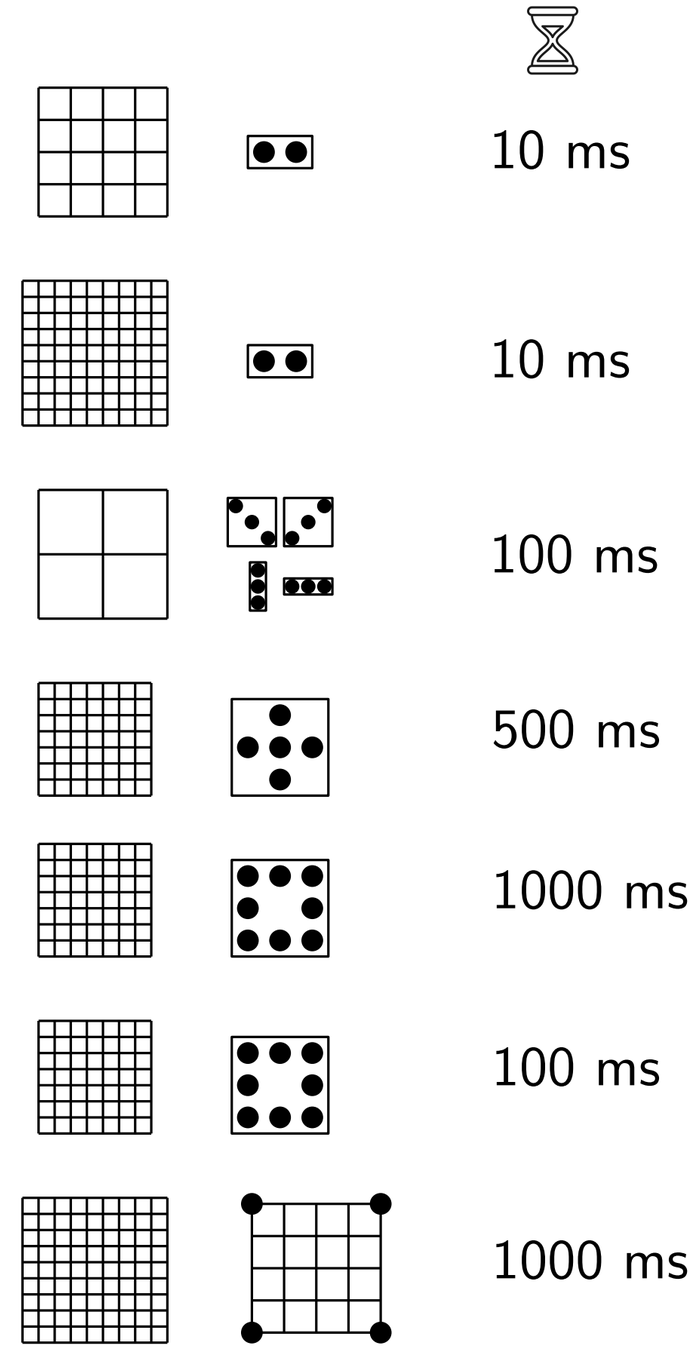
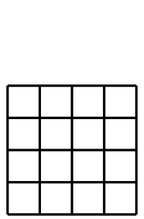
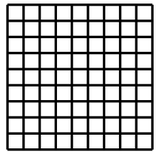
Bandit Grading



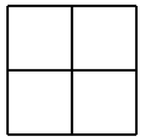
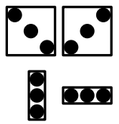

10 ms
10 ms
100 ms
500 ms
1000 ms
100 ms
1000 ms



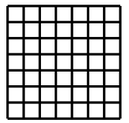

10 ms



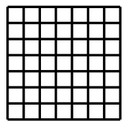

10 ms

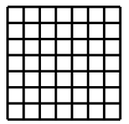
100 ms



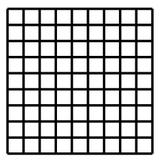
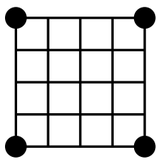

500 ms




1000 ms



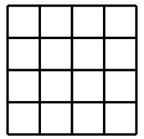

100 ms

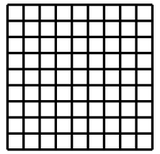
1000 ms

reference e-greedy	919/81/0	521/479/0	945/15/40	100/0/0	74/0/26	31/0/69	100/0/0
reference ucb	940/60/0	964/36/0	936/21/43	100/0/0	95/0/5	64/0/36	100/0/0

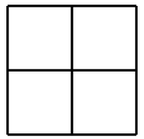
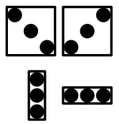
Win / Loose / Tie



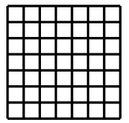

10 ms



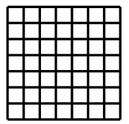

10 ms

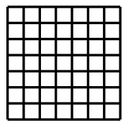
100 ms



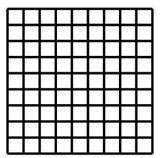
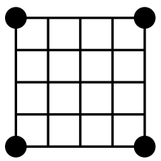

500 ms




1000 ms



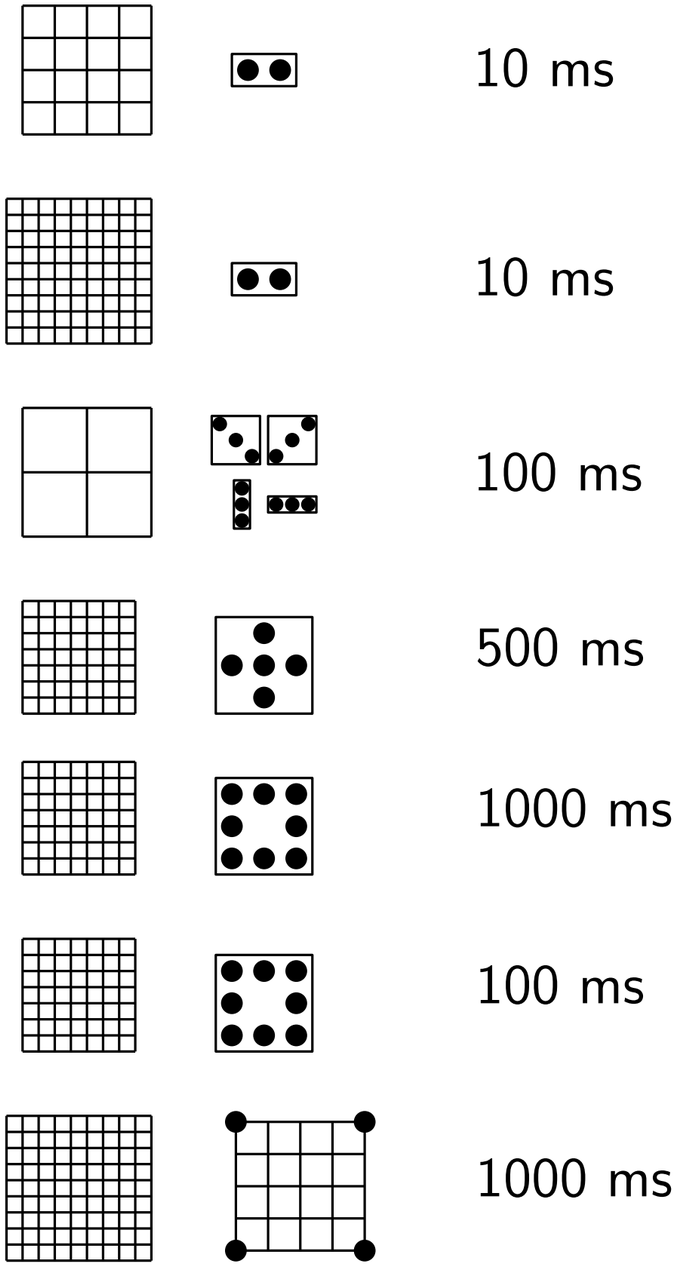

100 ms

1000 ms

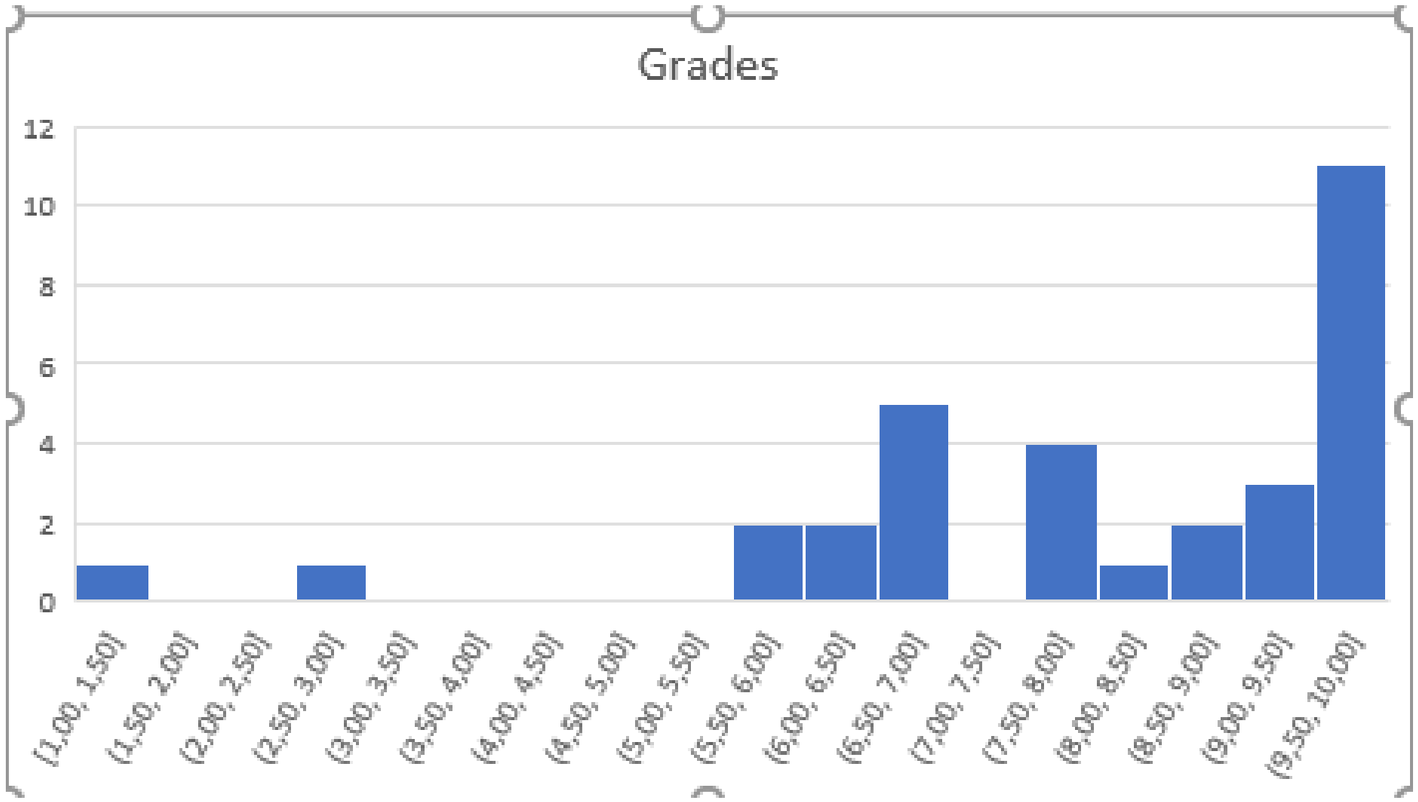
reference e-greedy	919/81/0	521/479/0	945/15/40	100/0/0	74/0/26	31/0/69	100/0/0
reference ucb	940/60/0	964/36/0	936/21/43	100/0/0	95/0/5	64/0/36	100/0/0

Win / Loose / Tie



reference e-greedy	919/81/0	521/479/0	945/15/40	100/0/0	74/0/26	31/0/69	100/0/0
reference ucb	940/60/0	964/36/0	936/21/43	100/0/0	95/0/5	64/0/36	100/0/0

Win / Loose / Tie



epsilon-greedy

Easy strategy good enough to win consistently

epsilon-greedy

Easy strategy good enough to win consistently

epsilon-decay

Improving upon epsilon greedy.

epsilon-greedy

Easy strategy good enough to win consistently

epsilon-decay

Improving upon epsilon greedy.

upper confidence bound

Maybe harder to fine tune.

epsilon-greedy

Easy strategy good enough to win consistently

epsilon-decay

Improving upon epsilon greedy.

upper confidence bound

Maybe harder to fine tune.

uniformly roll-outs

Surprising that this works!

epsilon-greedy

Easy strategy good enough to win consistently

epsilon-decay

Improving upon epsilon greedy.

upper confidence bound

Maybe harder to fine tune.

uniformly roll-outs

Surprising that this works!

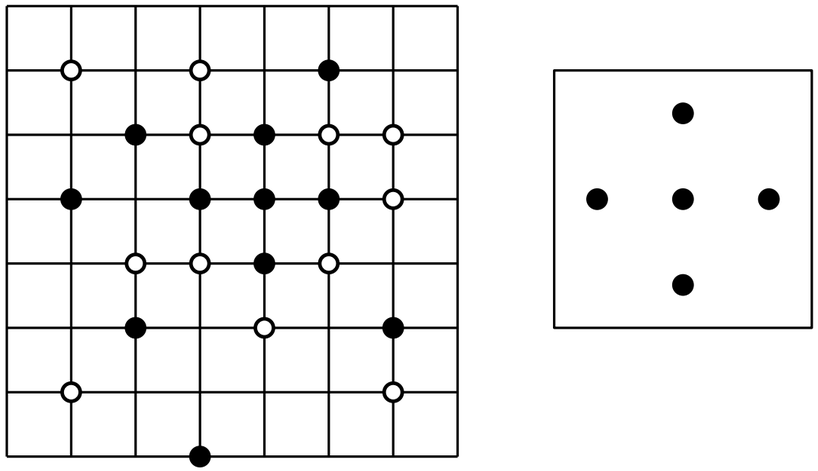
domain knowledge vs clever?

Sometimes hard to draw a line

100 roll-outs

if Win $\rightarrow +1$ for each position where stone is placed

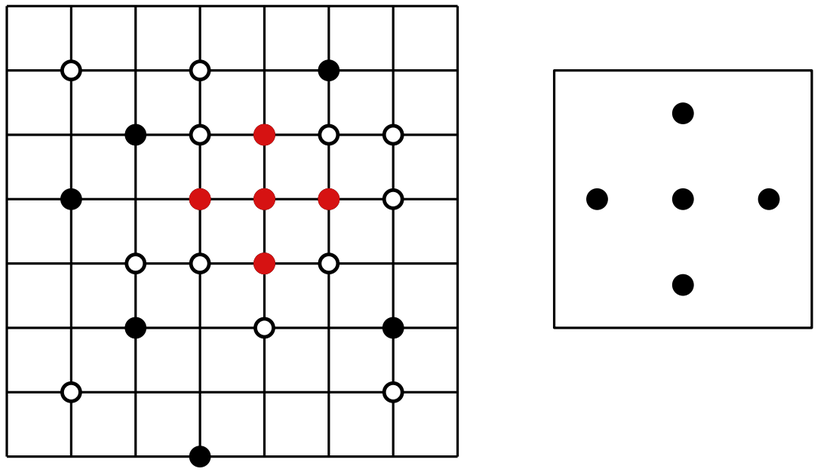
implicitly traces out the winning shape



100 roll-outs

if Win $\rightarrow +1$ for each position where stone is placed

implicitly traces out the winning shape



Alignment

learning goal

- Ability to use bandit strategies in a simple game setting.
- Ability to use roll-outs.
- Understanding the power of simple roll-outs.
- Able to structurally test and fine tune.
- comfortable to implement (affective goal)

grading

- Tested performance in different settings.
- checked for following instructions

Why better?

short anonymous survey

https://forms.gle/DJdyXGE42KN1usNV8