2nd DLP Workshop:
Game AI Applications for Historical Games Research

Measuring Games: Typicality, Novelty and Quality

C. Browne

erc
European Research Council
Established by the European Commission

Maastricht University, 11 April 2022

https://ludii.games/workshop.php

Reconstructing Games

Partial rule sets

- Millions of ways to complete

Use historical context

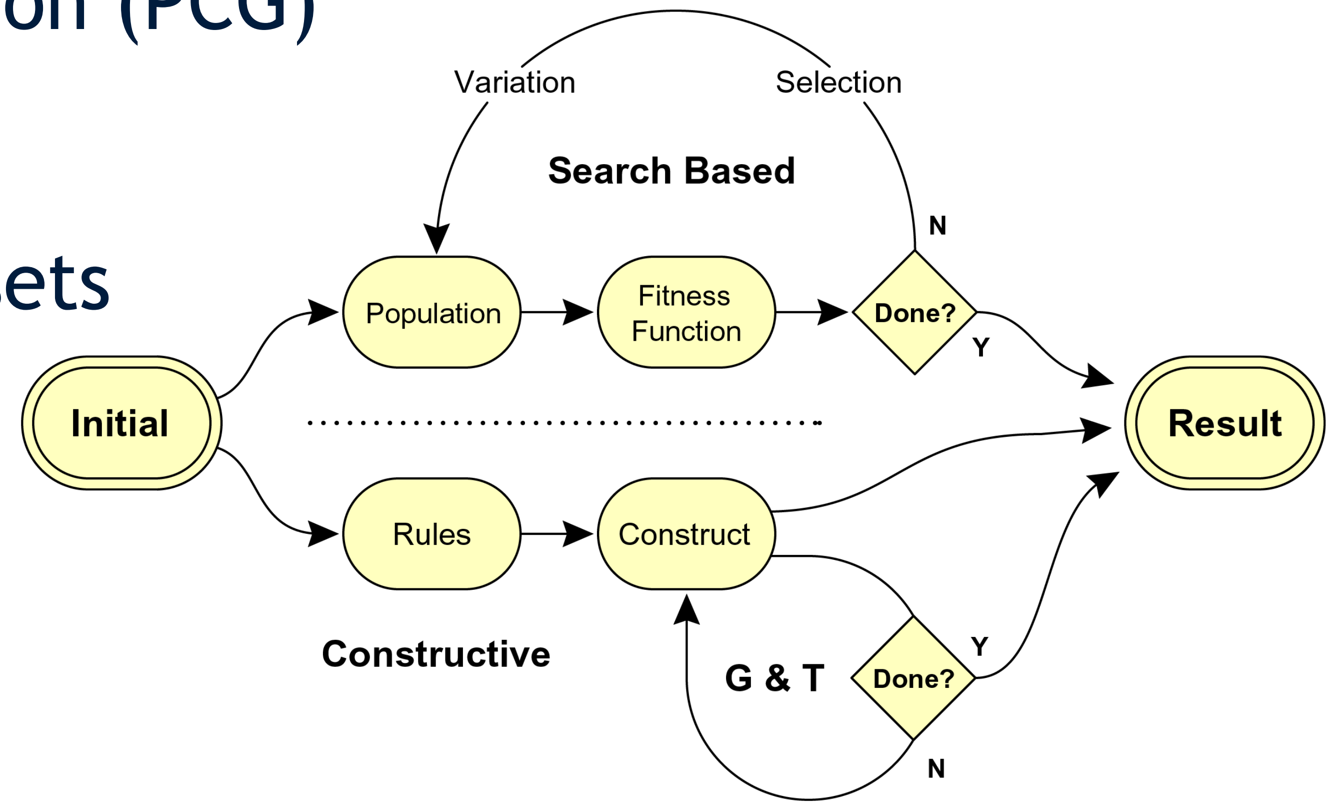
- Reduce to hundreds

How to measure these in reasonable time?

Procedural Content Generation

Procedural Content Generation (PCG)
- Generate & Test

1. Generate candidate rule sets
2. Measure them
3. Rank by score



Togelius *et al.* (2011)
"Search-Based Procedural Content Generation"
EvoApplications, Springer LNCS 2024, 141–150

Computational Creativity

Examples of ***combinatorial creativity***

Indicators of creativity:

- ***Typicality*** - Artefacts should be typical of their domain
- ***Novelty*** - Artefacts should be novel in their domain
- ***Quality*** - Artefacts should display quality in their domain

G. Ritchie (2007)
"Some Empirical Criteria for Attributing Creativity to a Computer Program"
Minds & Machines 17:67–99

Novelty

Covered by Matthew tomorrow (game distance)…

Typicality

C. Browne (2009)
"Automatic Generation and Evaluation of Recombination Games"
PhD Thesis, QUT

A typical game is:

1. **Fair**
   All players have reasonably equal chance of winning

2. **Decisive**
   Produces win/loss results more often than draws

3. **Good Length**
   Not too short (trivial) nor too long (boring)

Can be adjusted according to cultural context

Typicality

Typicality quickly filters out badly flawed rule sets

Can use random playouts

- Fast!
- Explores state space more thoroughly

Playouts are not realistic

Don’t give much insight into actual quality

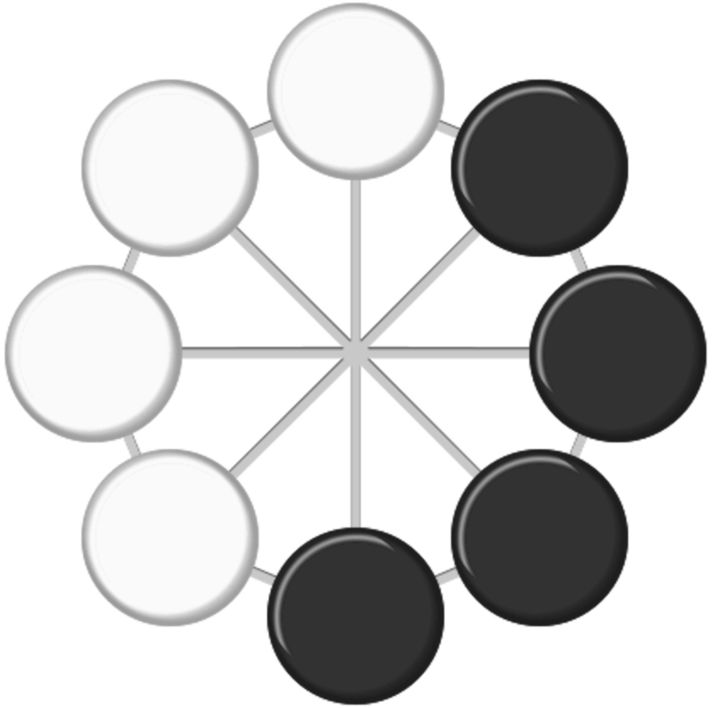
Example: Mu Torere

Traditional Maori (NZ) game

- 18th century

1. Players take turns moving a piece of their colour to the adjacent empty point

2. Player with no move loses



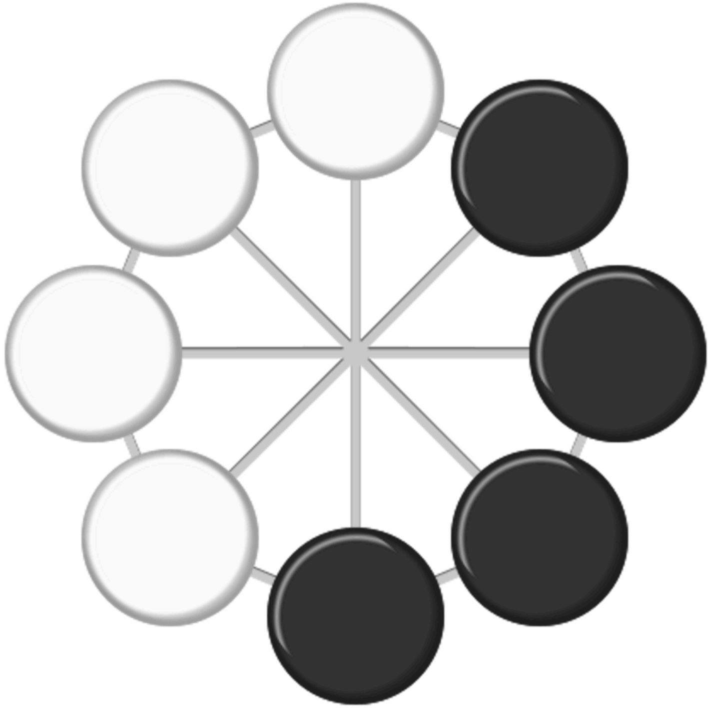
Mu Torere Rulesets

1a. Unconstrained movement

1b. Piece being moved must be adjacent to enemy piece (on first move)

1c. Piece being moved must be adjacent to enemy piece (on first two turns)

1d. Piece being moved must be adjacent to enemy piece (on all turns)



Mu Torere Rulesets

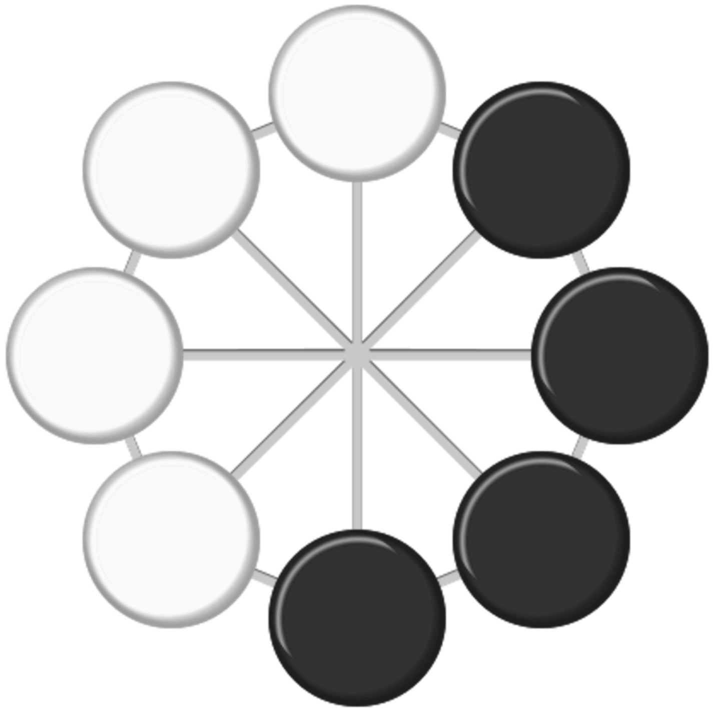
Bad length

~~1a. Unconstrained movement~~

1b. Piece being moved must be adjacent to enemy piece (on first move)

1c. Piece being moved must be adjacent to enemy piece (on first two turns)

1d. Piece being moved must be adjacent to enemy piece (on all turns)



Mu Torere Rulesets

Bad length

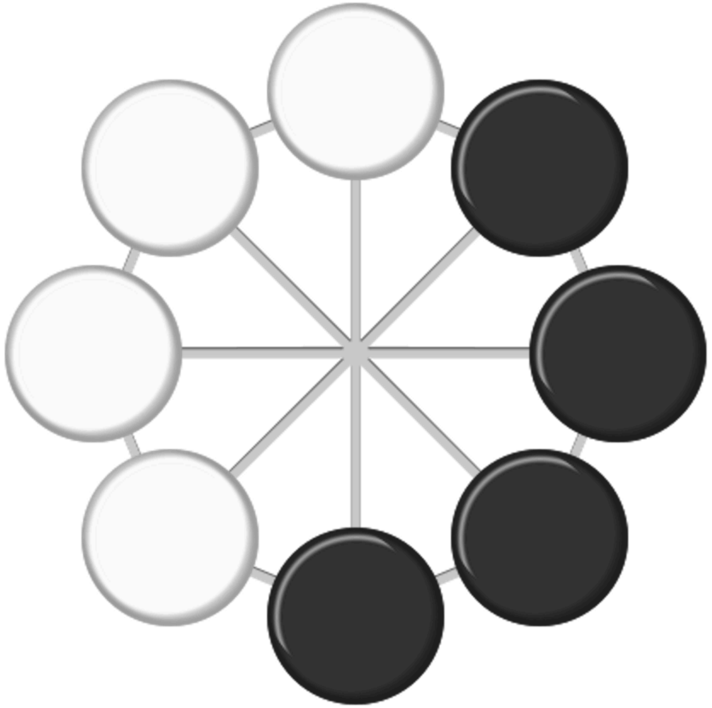
~~1a. Unconstrained movement~~

Unfair

~~1b. Piece being moved must be adjacent to enemy piece (on first move)~~

1c. Piece being moved must be adjacent to enemy piece (on first two turns)

1d. Piece being moved must be adjacent to enemy piece (on all turns)



Mu Torere Rulesets

Bad length

~~1a. Unconstrained movement~~

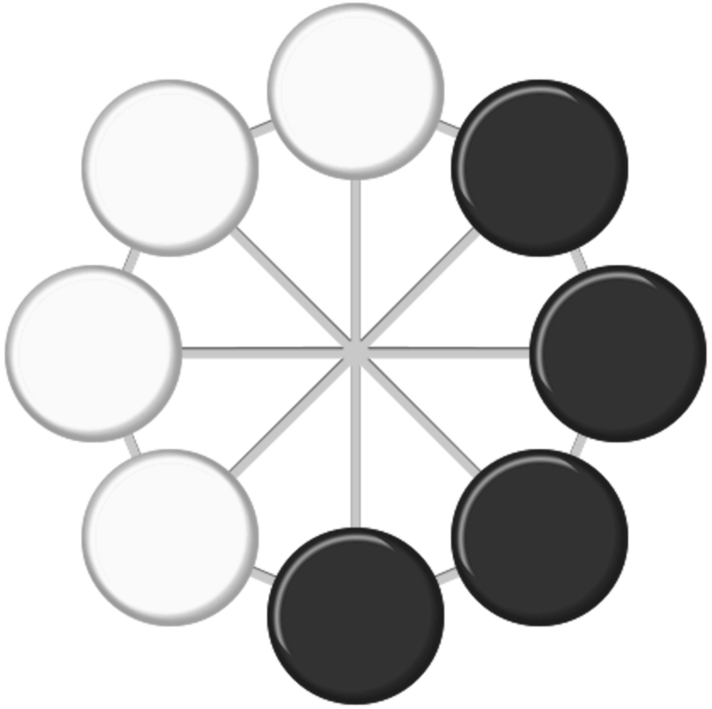
Unfair

~~1b. Piece being moved must be adjacent to enemy piece (on first move)~~

~~1c. Piece being moved must be adjacent to enemy piece (on first two turns)~~

~~1d. Piece being moved must be adjacent to enemy piece (on all turns)~~

Indecisive



Implications

Don’t want perfect play

- May not even want strong play

Need to incorporate **quality**

- Measure this across a range of games
- Measure it quickly

How?

Strategy is the Key

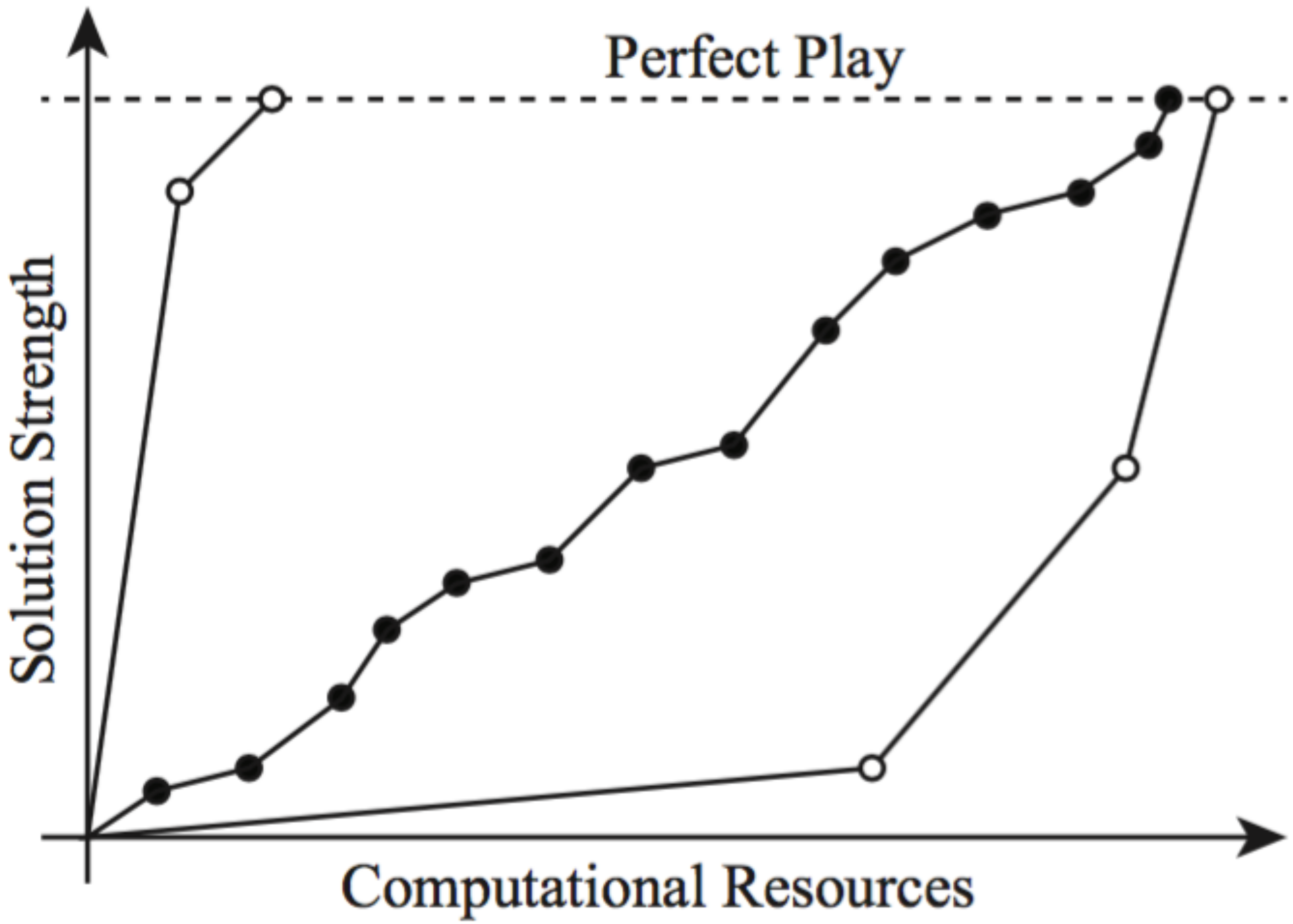
Trying to model transmission of traditional **strategy** games

Games with easily obtained strategies are:

- Easy to learn
- Easy to teach
- Interesting to play

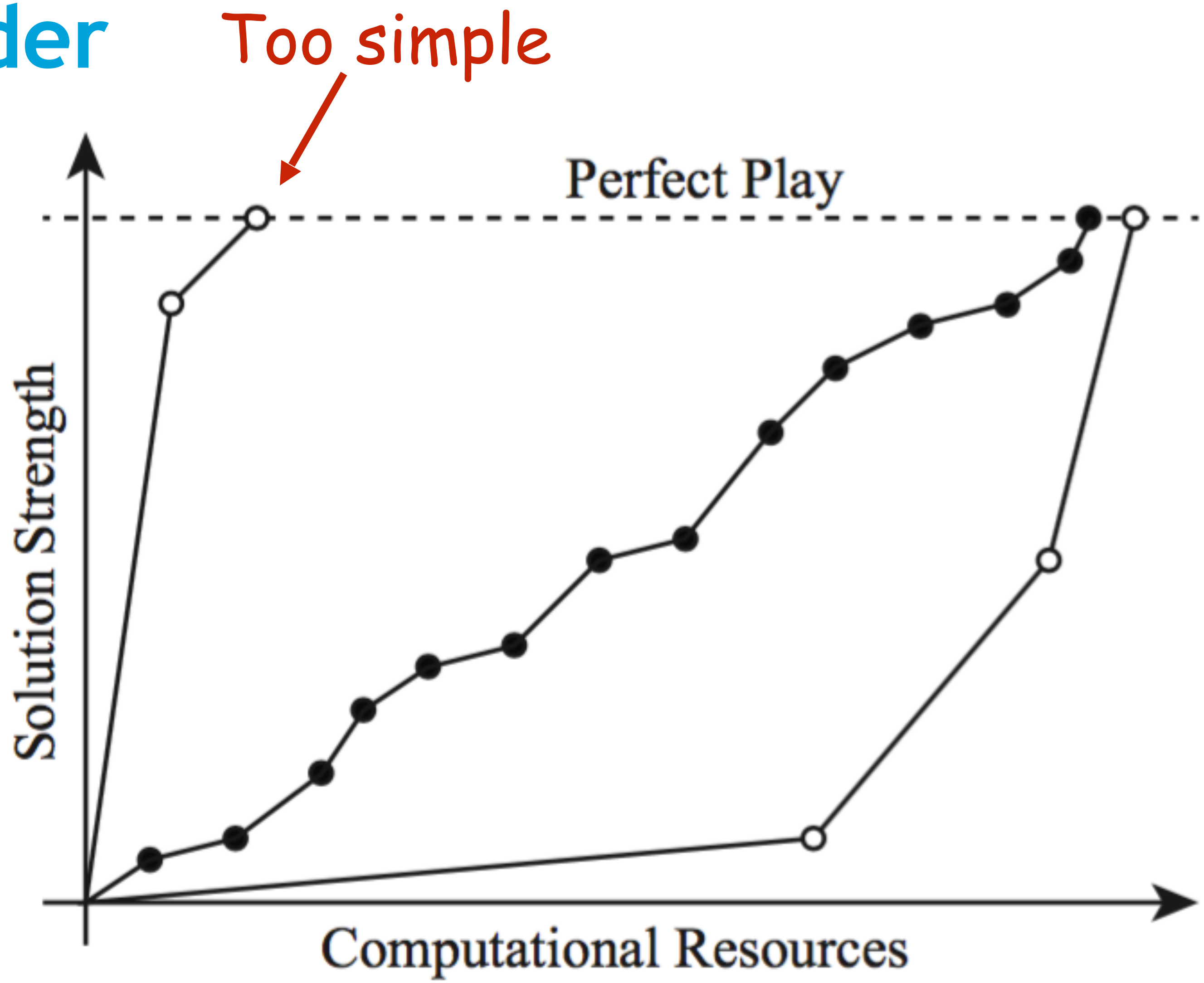
These are the games that will be passed on and survive

Strategy Ladder



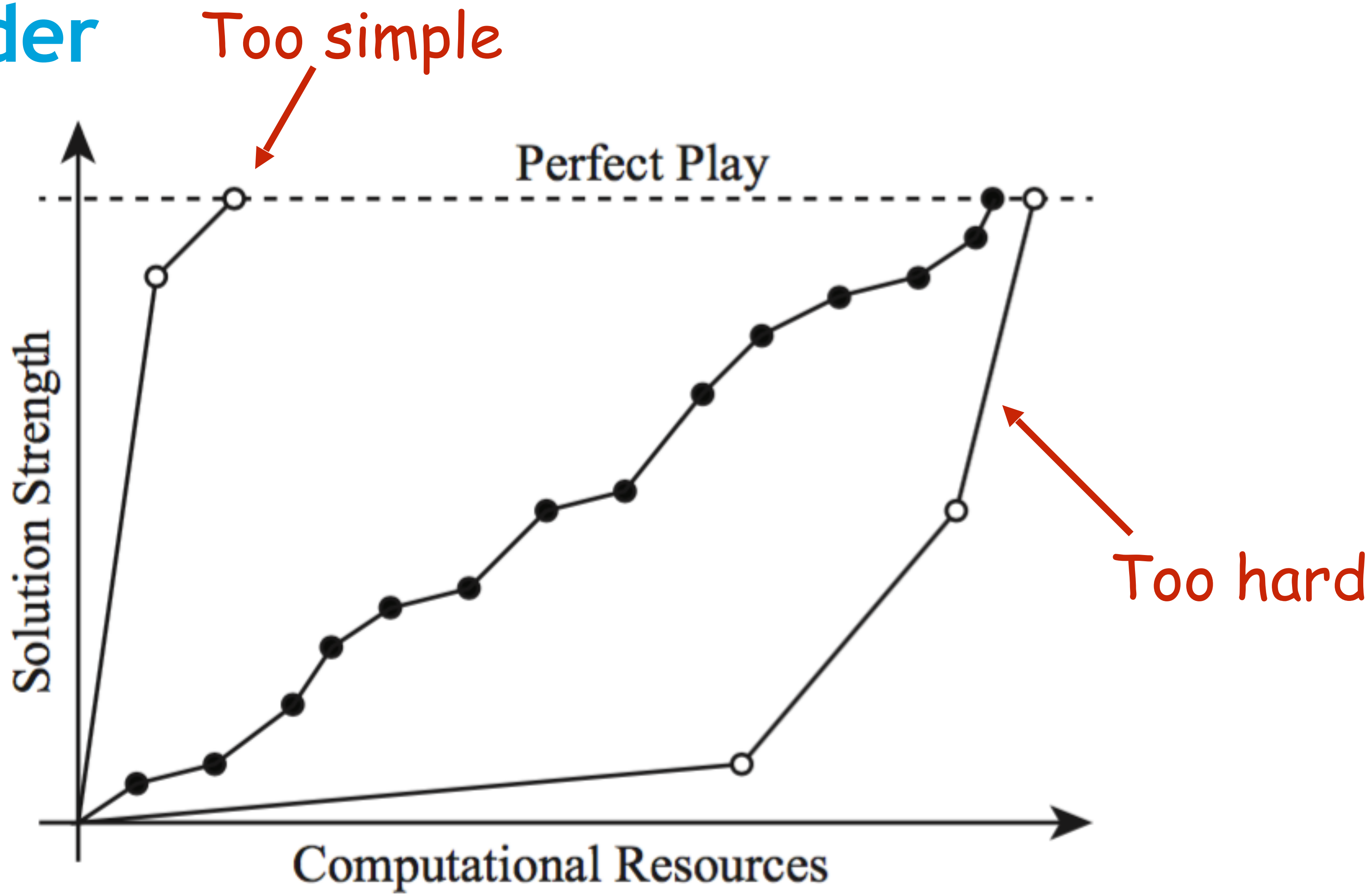
Lantz *et al*. (2007)
“Strategic Depth in Games” *AAAI’17*

Strategy Ladder



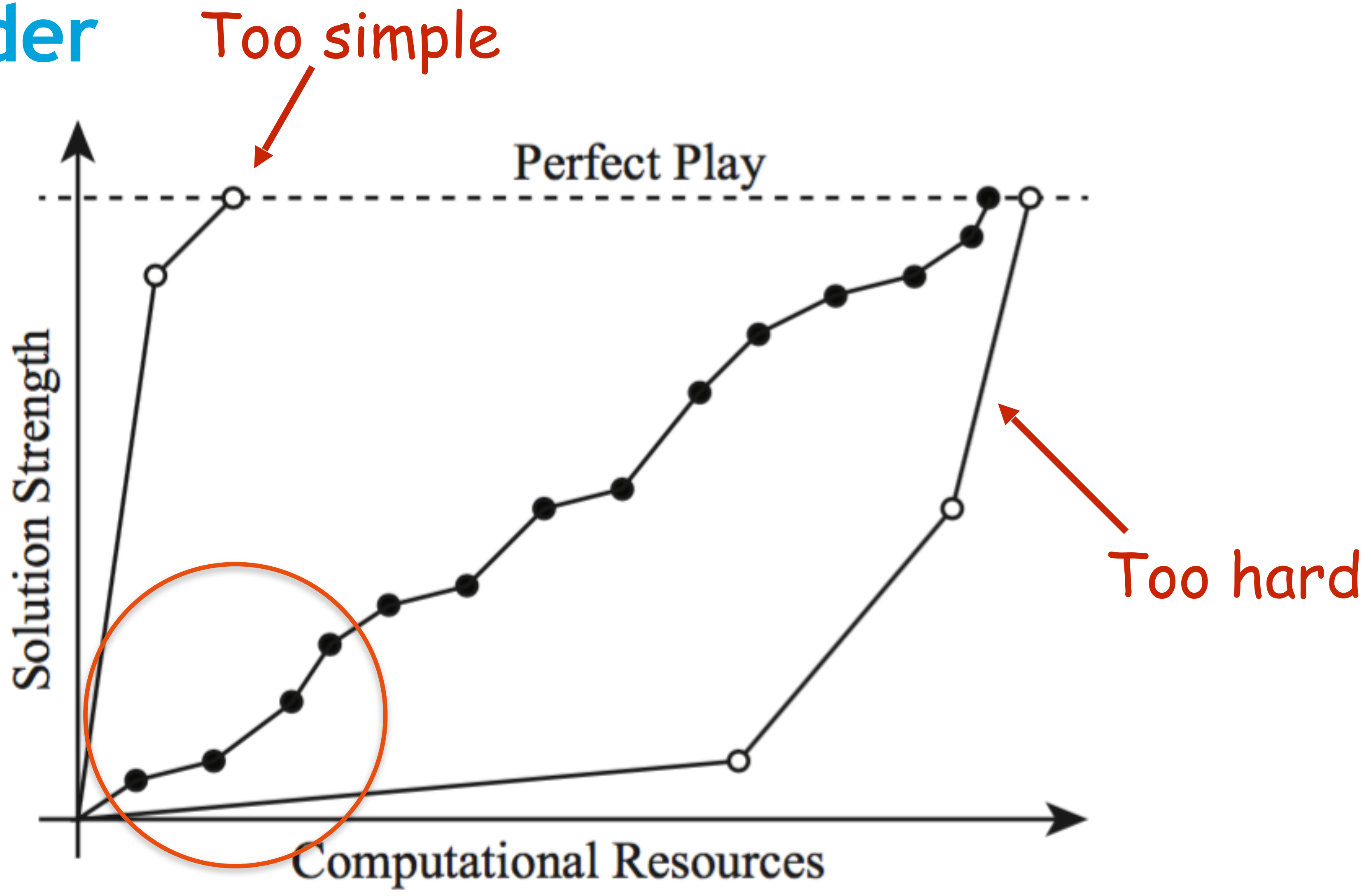
Lantz *et al.* (2007)
"Strategic Depth in Games" *AAAI'17*

Strategy Ladder



Lantz *et al.* (2007)
"Strategic Depth in Games" *AAAI'17*

Strategy Ladder



Lantz *et al.* (2007)
"Strategic Depth in Games" *AAAI'17*

Strategic Depth

Standard approach:
- Find number of distinct skill levels, e.g. ELO ratings

Requires either:
- Database of played games
  - With knowledge of players
- Generation of played games
  - Takes hours/days to compute

Problems:
- Not reliable
- We want a result in seconds/minutes

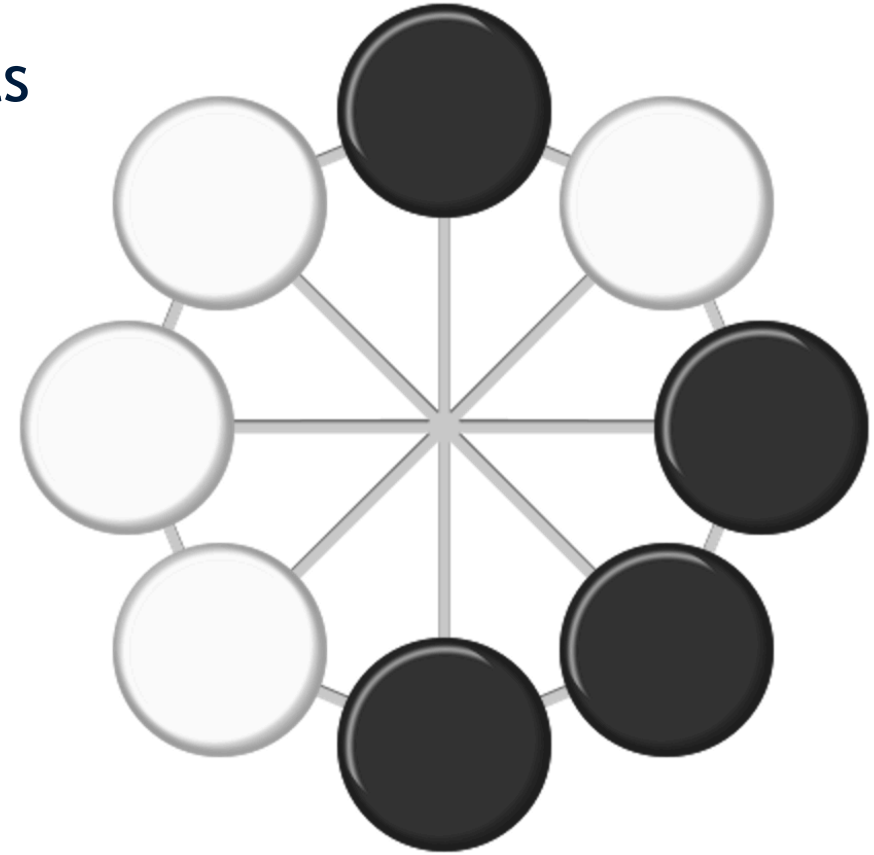
M. Thompson (2000)
"Defining the Abstract"
Abstract Games Journal

New Approach

Use MCTS (standard UCT)
- Compare weak agent vs strong agent
- Successively doubled at (low) iteration counts

For each match *m*
- Weak agent: *m* x BF playouts
- Strong agent: 2*m* x BF playouts

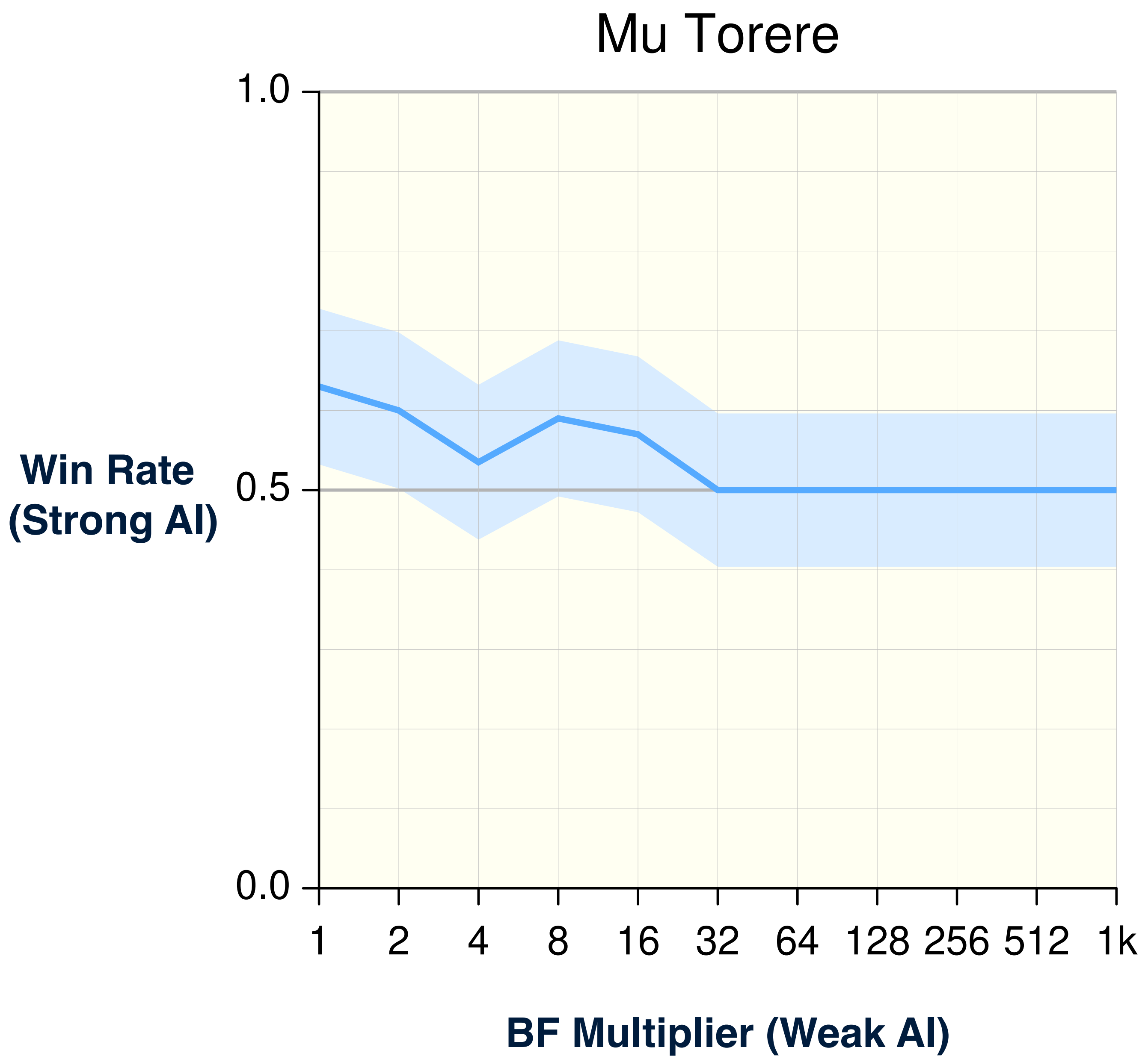
e.g. Game state *s* with BF = 4 moves
- UCT_4 vs UCT_8
- UCT_8 vs UCT_{16}
- UCT_{16} vs UCT_{32}
- UCT_{32} vs UCT_{64}
- …



Result

Plot (linear/$\log_2$ scale)

- Win rate of strong AI
- BF multiplier of weak AI



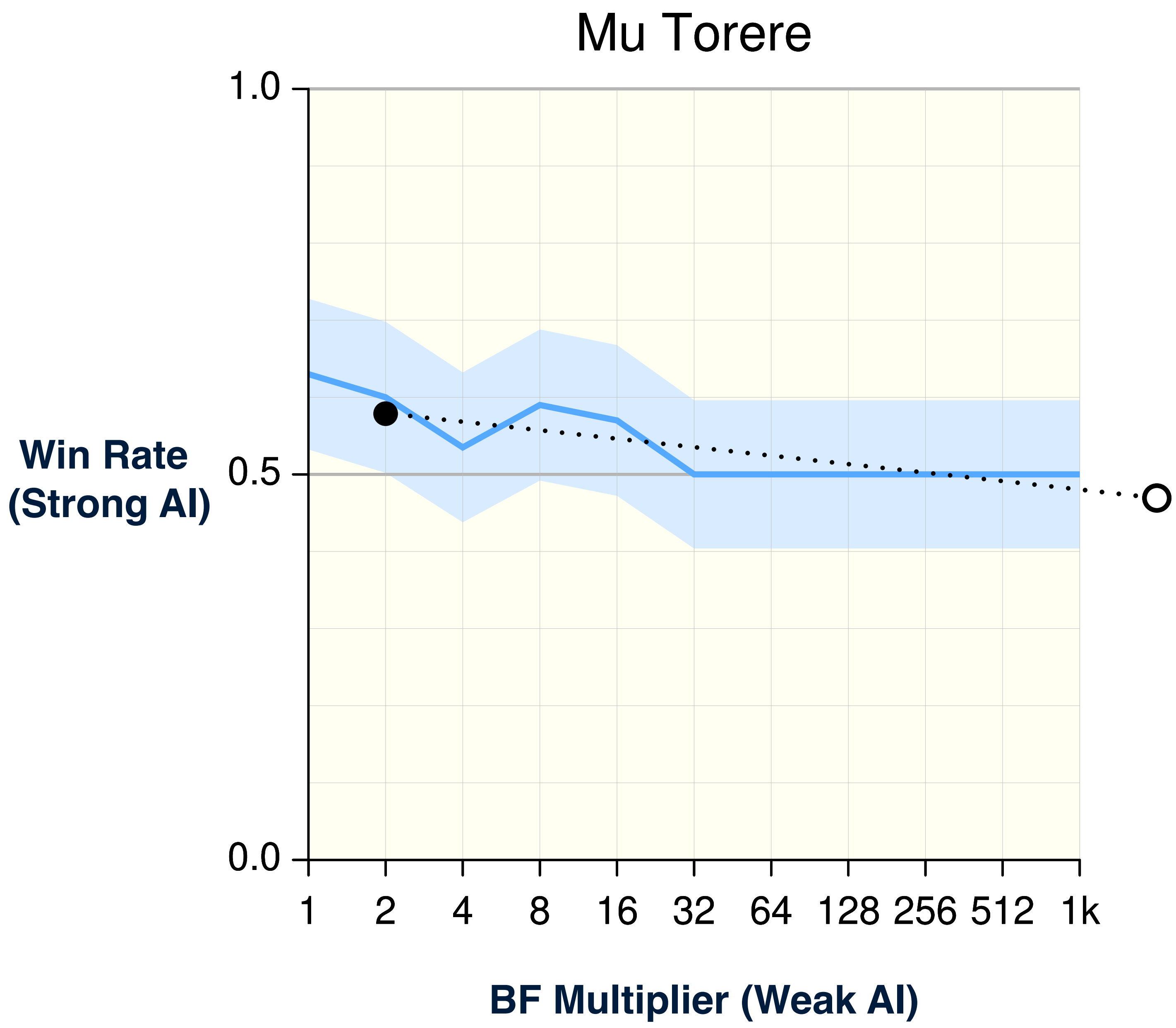
Result

Plot (linear/$\log_2$ scale)

- Win rate of strong AI
- BF multiplier of weak AI

Find regression line

- Ignore BF = 1
- Extrapolate to next BF



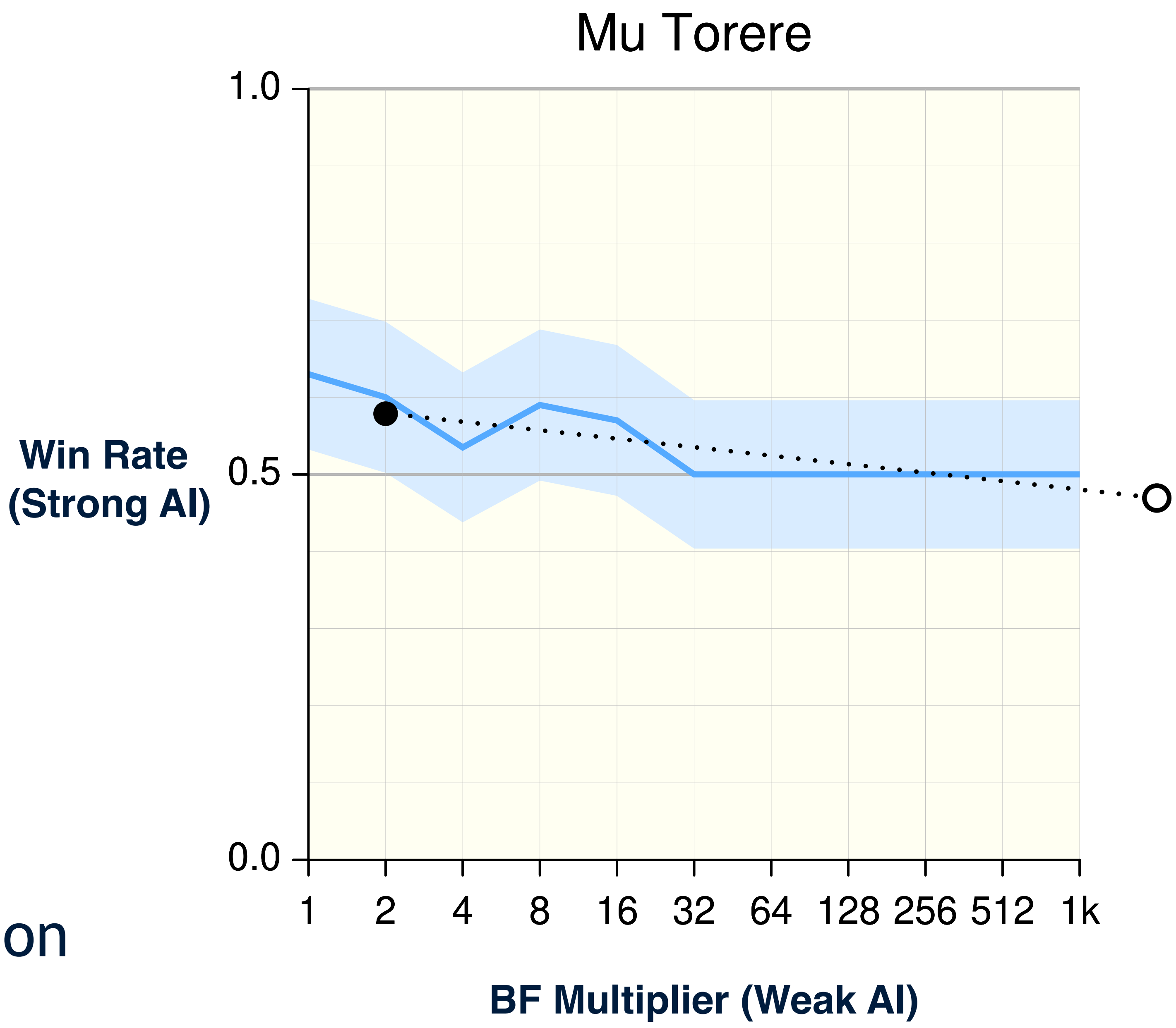
Result

Plot (linear/$\log_2$ scale)
- Win rate of strong AI
- BF multiplier of weak AI

Find regression line
- Ignore BF = 1
- Extrapolate to next BF

Smooths out fluctuations

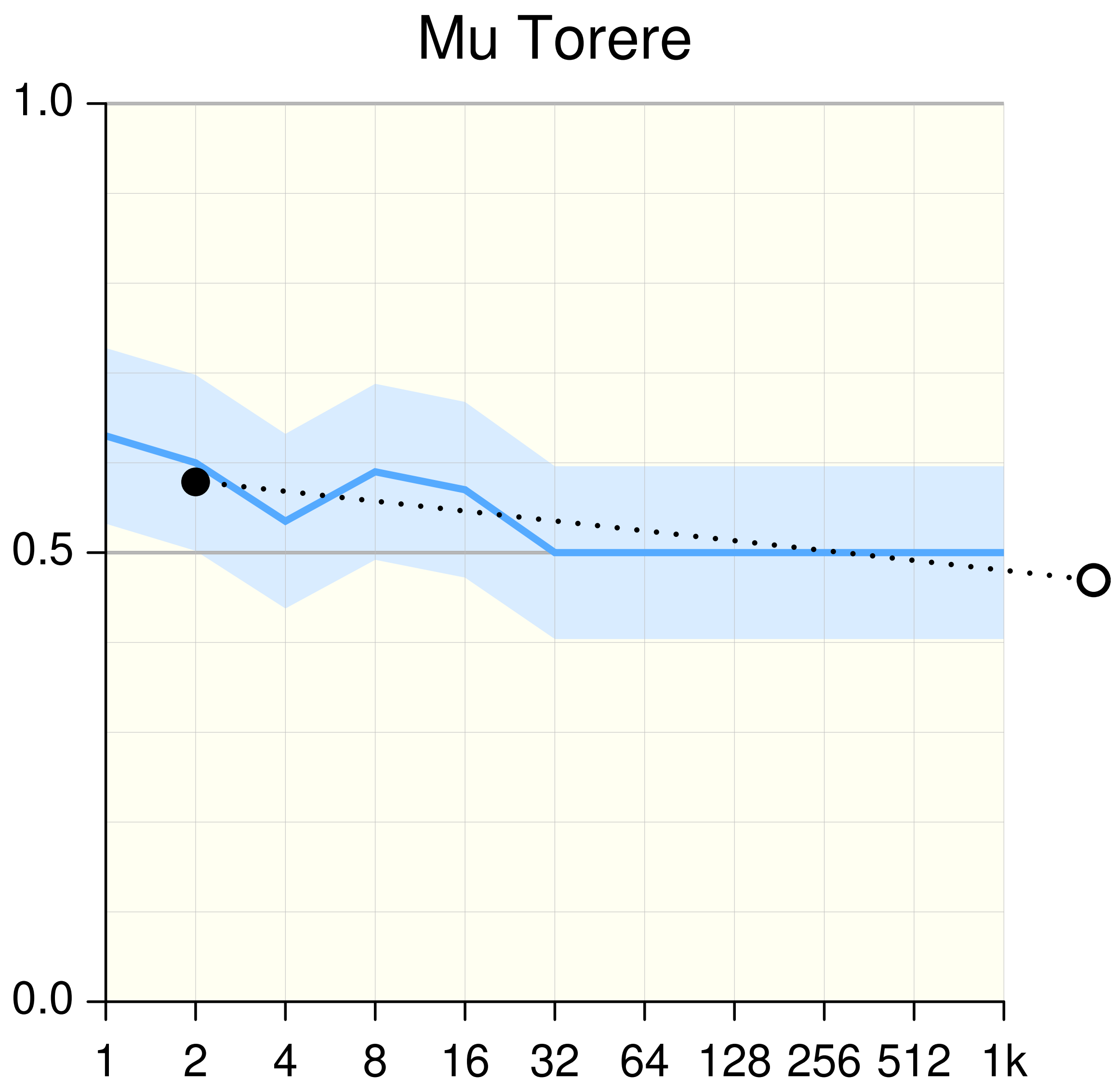
Prediction based on trend
- Haussdorf/fractal dimension



Odd UCT Behaviour

Ignore result of

- UCT_{0BF} vs UCT_{1BF}
- UCT_{1BF} vs UCT_{2BF}



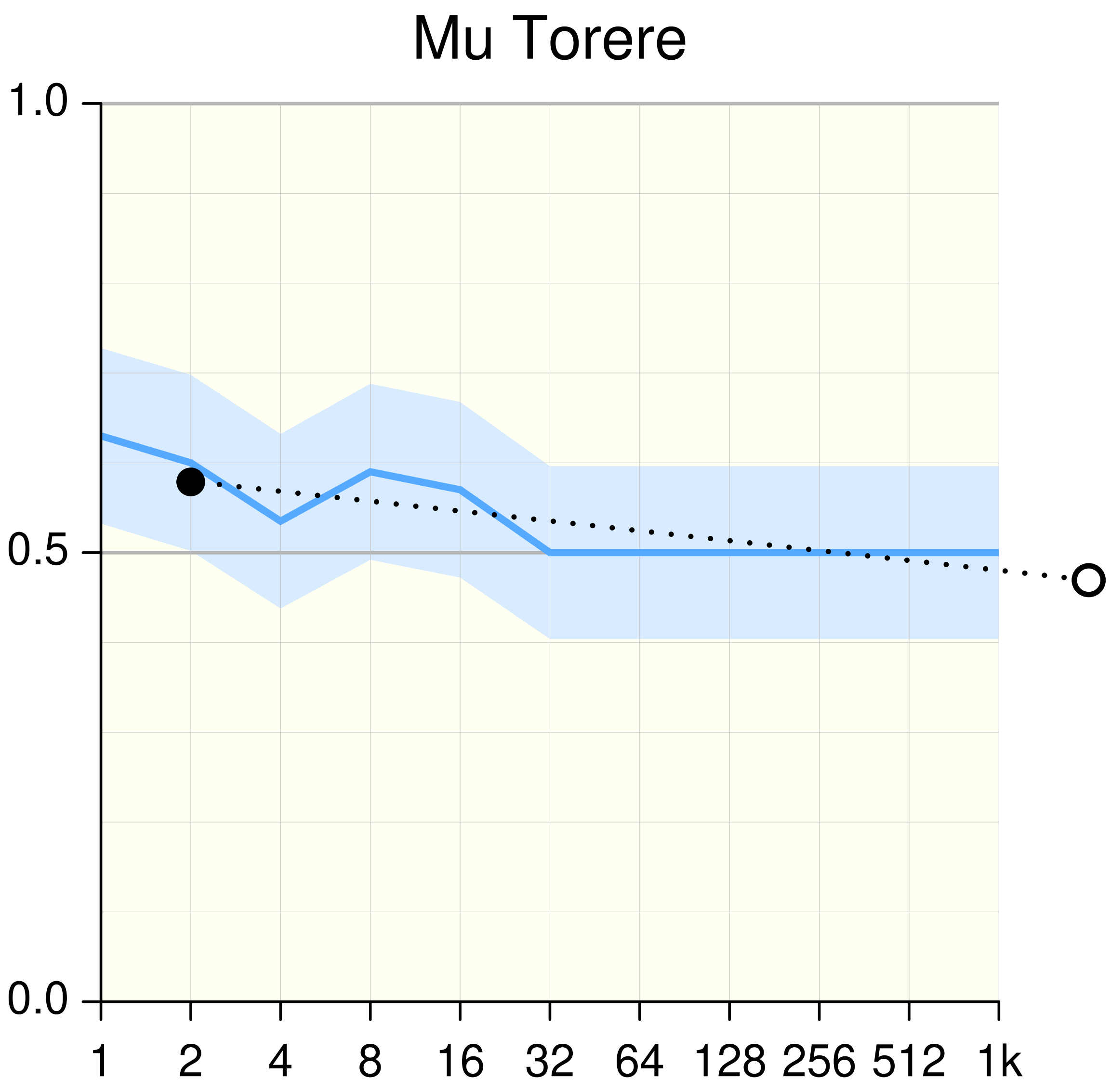
Odd UCT Behaviour

Ignore result of

- UCT_{0BF} vs UCT_{1BF}
- UCT_{1BF} vs UCT_{2BF}

UCT with 0 .. BF iterations

- Is random selection



Odd UCT Behaviour

Ignore result of

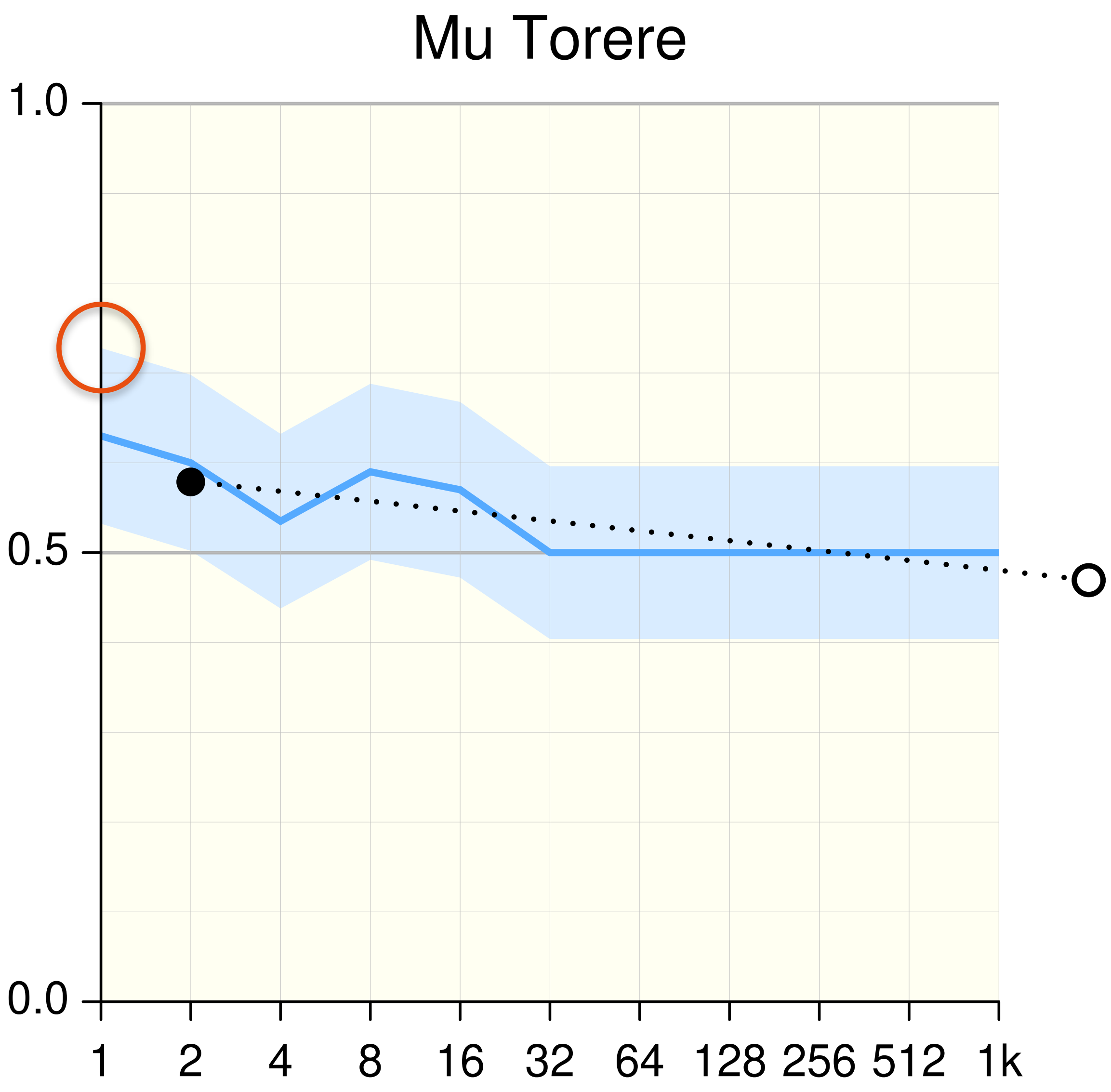
- UCT_{0BF} vs UCT_{1BF}
- UCT_{1BF} vs UCT_{2BF}

UCT with 0 .. BF iterations

- Is random selection

UCT with BF .. 2BF iterations

- Usually better than expected



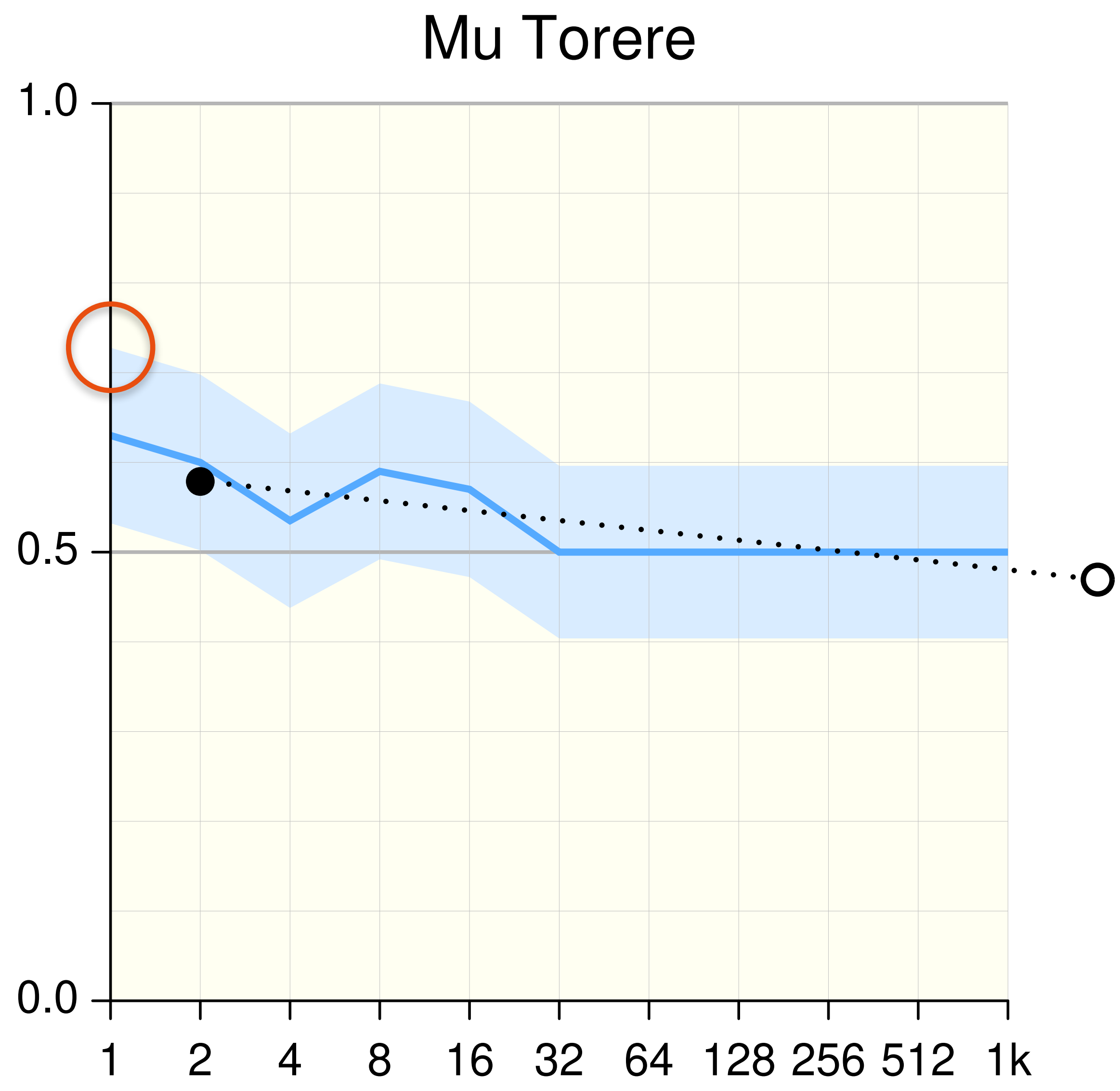
Odd UCT Behaviour

Ignore result of
- UCT_{0BF} vs UCT_{1BF}
- UCT_{1BF} vs UCT_{2BF}

UCT with 0 .. BF iterations
- Is random selection

UCT with BF .. 2BF iterations
- Usually better than expected

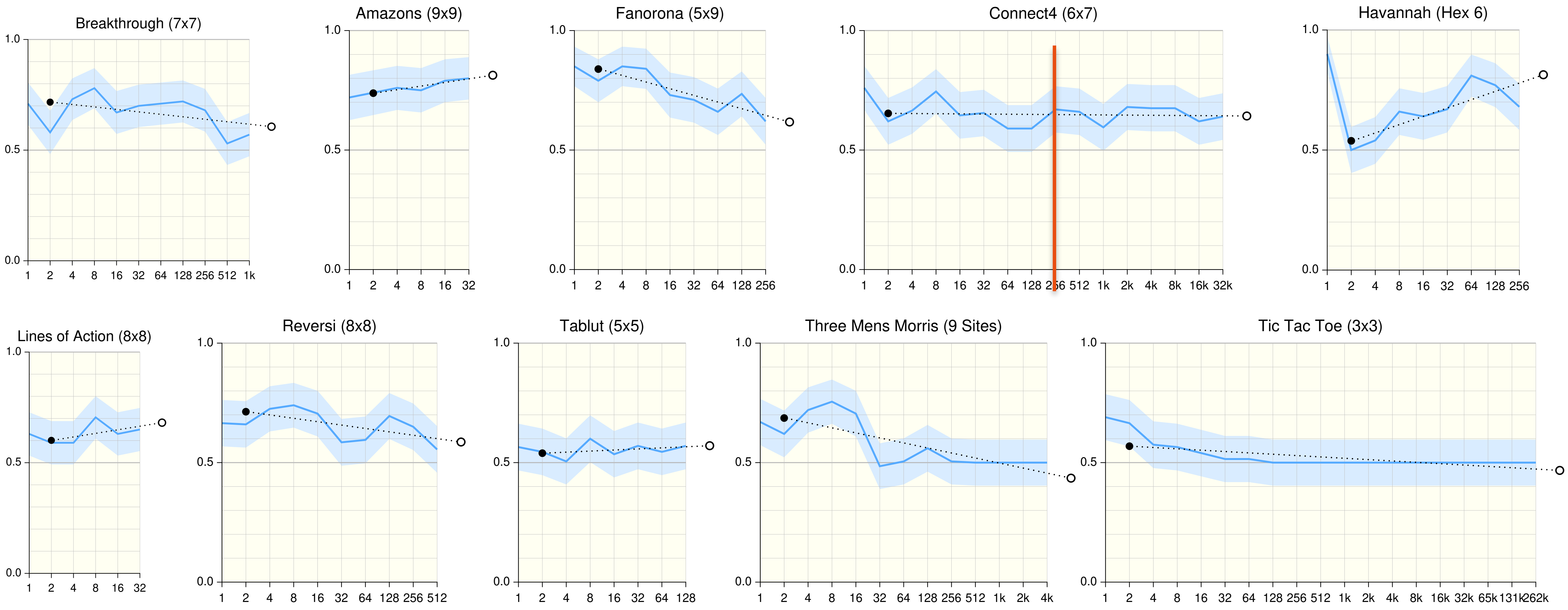
Calculation time: 20s



More Results

20s run time (Stephen Tavener's *AI AI*)

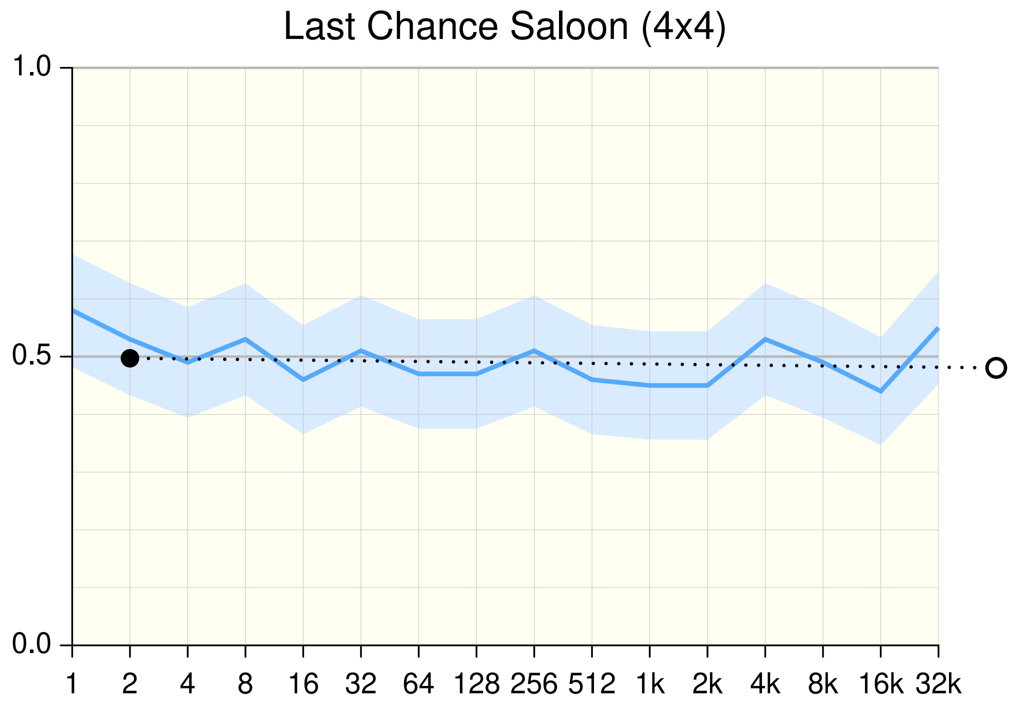
- 100 trials per match or 95% confidence (whichever comes first)

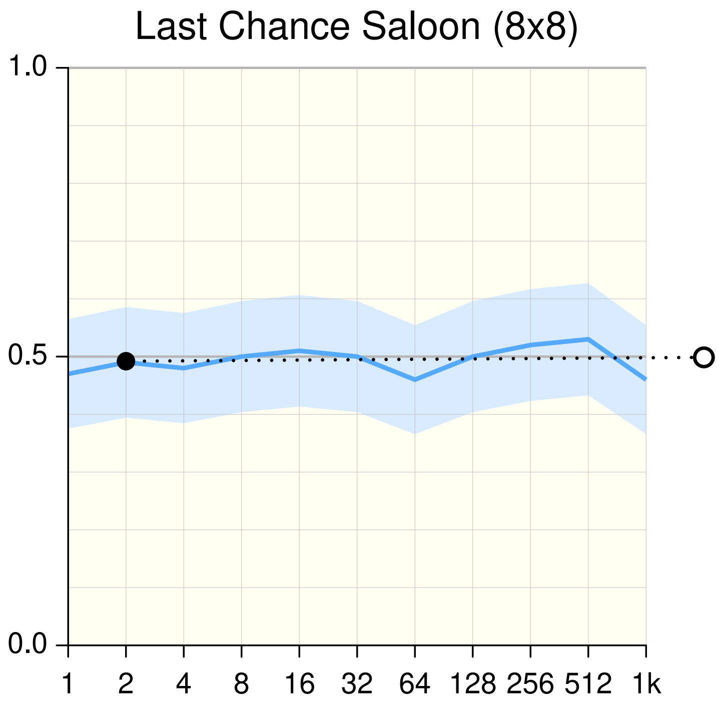


Pure Chance Games

Last Chance Saloon
- Add piece to empty cell
- When board is full,
  roll d6 to determine winner

Zero signal
- As expected





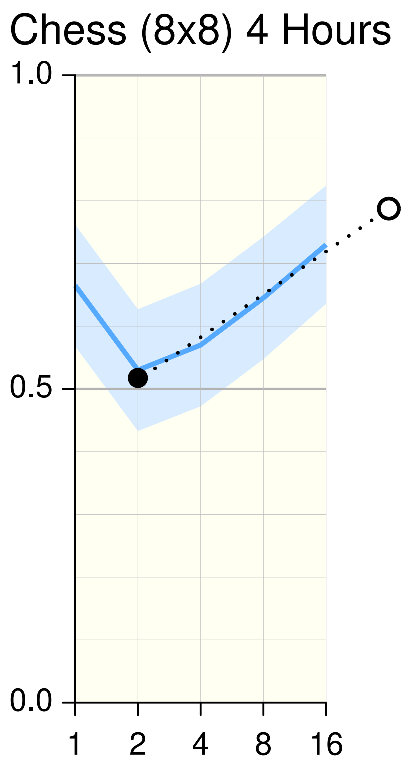
Complex Games

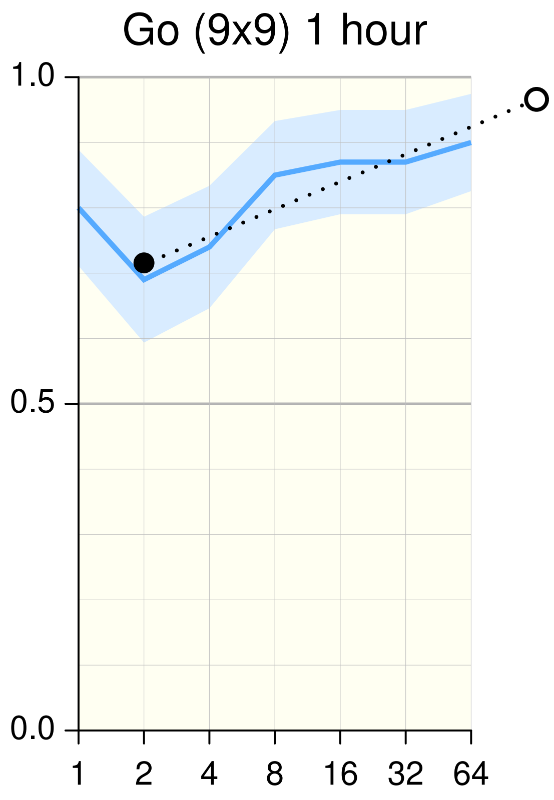
Chess

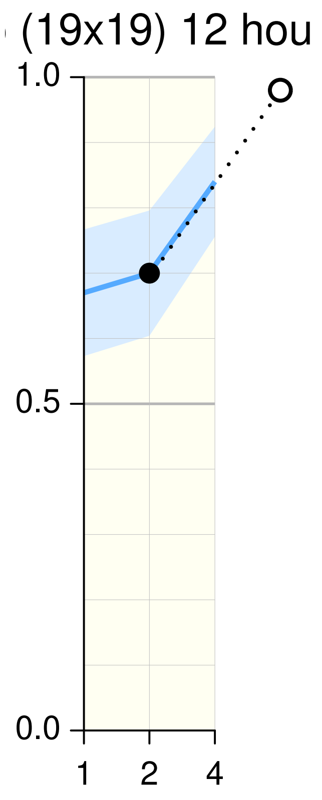
- 4 hours (BF = ~40)

Go

- 9x9 (BF = ~50)
- 19x19 (BF = ~200)







What Is This Measuring?

Advantage of deeper search over shallower search
- Capacity for states to be misleading
- Resolved with deeper search

Humans do this by learning strategies
- “Chunk” knowledge about the game

Plots show potential for strategies to exist
- Not the strategies themselves!

Simple Games

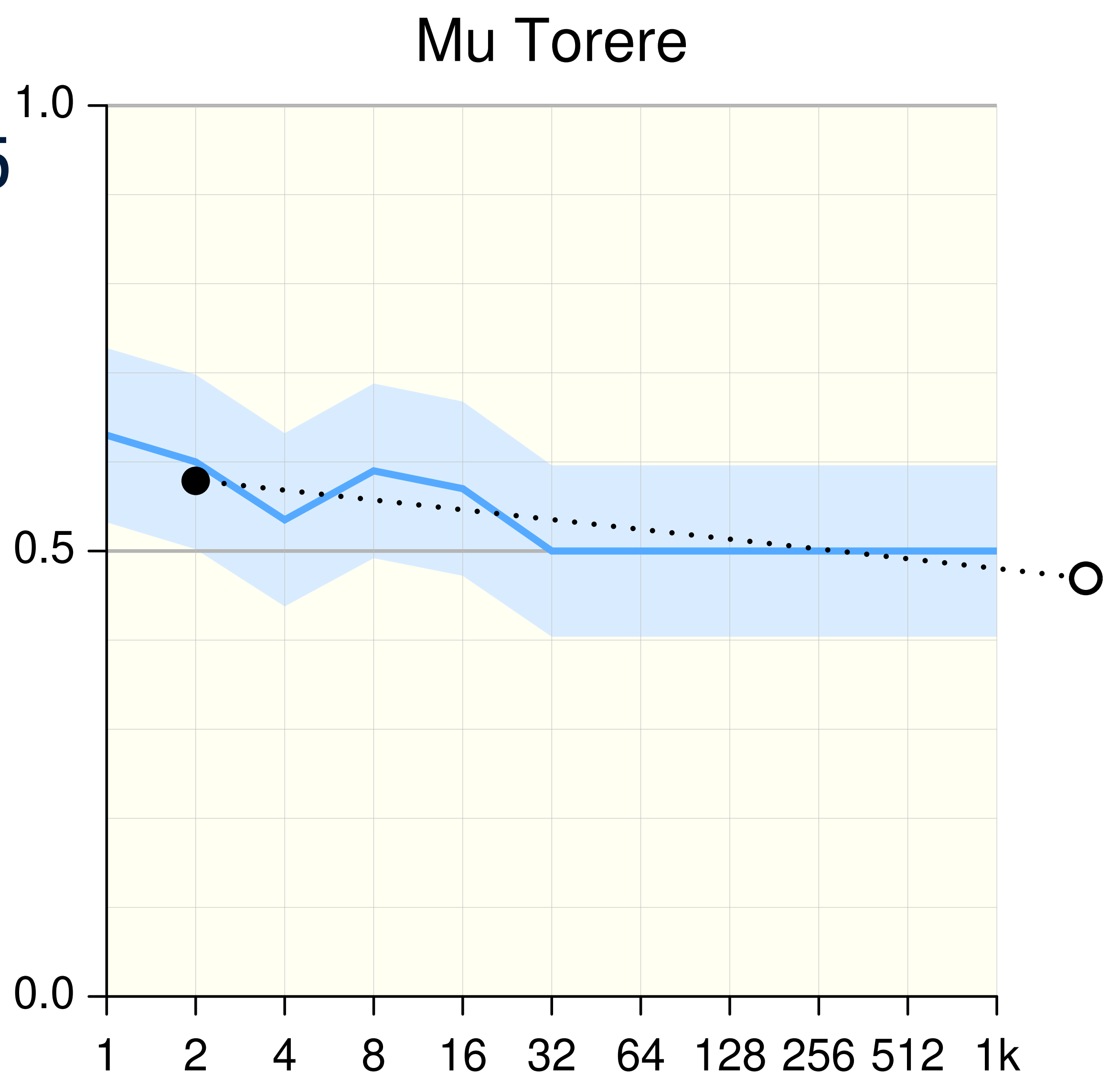
Simple games quickly converge to 0.5

Strategies will leave non-zero signal at lower iterations

Better measure:

$$SP = y + (1 - y)\ |A|$$

Catches simple games

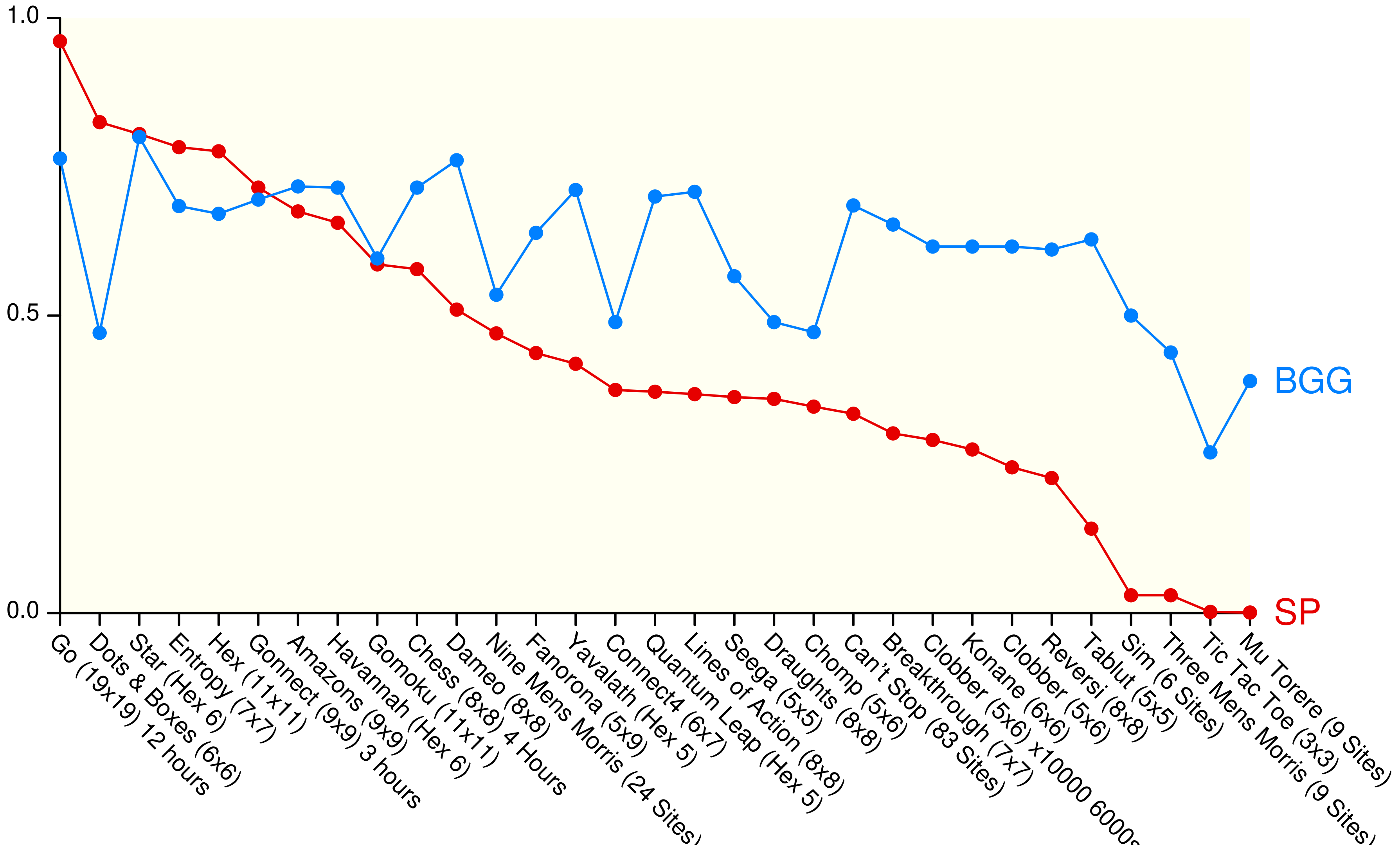


Comparison with BGG Ratings

BoardGameGeek
- Geek rating

Some correlation

Can distinguish
- No strategy
- Some strategy
- Signs of deeper strategy



Summary

Detects indicators of strategy
- Skill/chance balance?

Good
- Fast
- Easy to implement and calculate
- Doesn’t need existing AI

Bad
- Result depends on search time

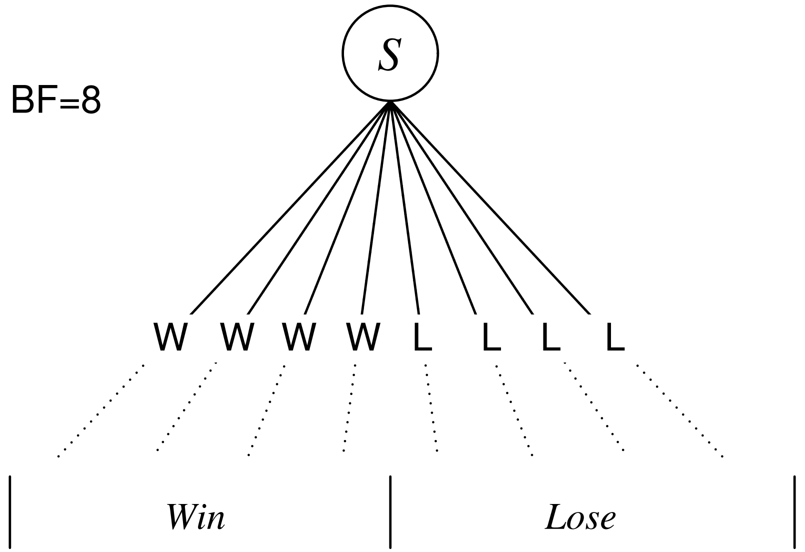
Information from nothing

Feedback?

Odd UCT Behaviour

Imagine a state *s* with BF=8 move choices

- Four always lead to wins (W)
- Four always lead to losses (L)

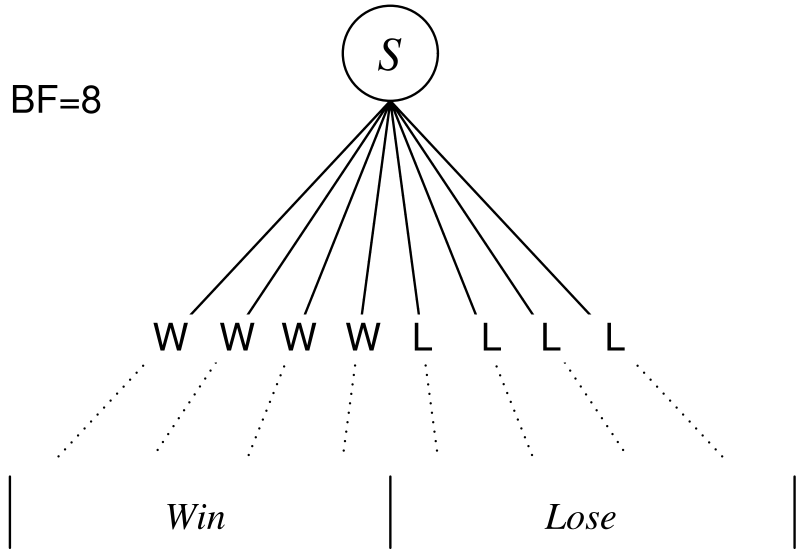


Odd UCT Behaviour

Imagine a state *s* with BF=8 move choices
- Four always lead to wins (W)
- Four always lead to losses (L)

Visits

```
  |WLWWLLWLW|WWWWWWWW|LLLLWWWW|...
0xBF        1xBF     2xBF      3xBF
```



Exploitation

$$UCB_j = \bar{X}_j + C\sqrt{\frac{ln\ N}{n_j}}$$

Exploration

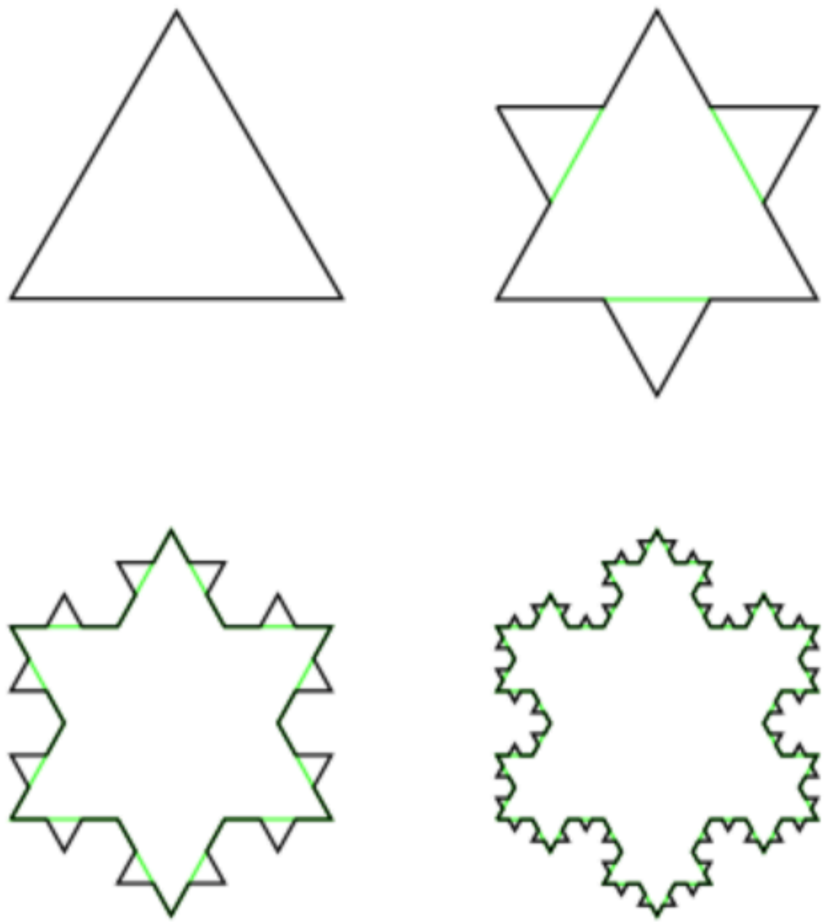
Haussdorf Dimension

Felix Haussdorf (1918)

Log/log plot
- Length
- Precision

$$\log_{\varepsilon} N = -D = \frac{\log N}{\log \varepsilon}$$

e.g. Koch snowflake:



That is, for a fractal described by $N = 4$ when $\varepsilon = \frac{1}{3}$, $D = 1.2619$

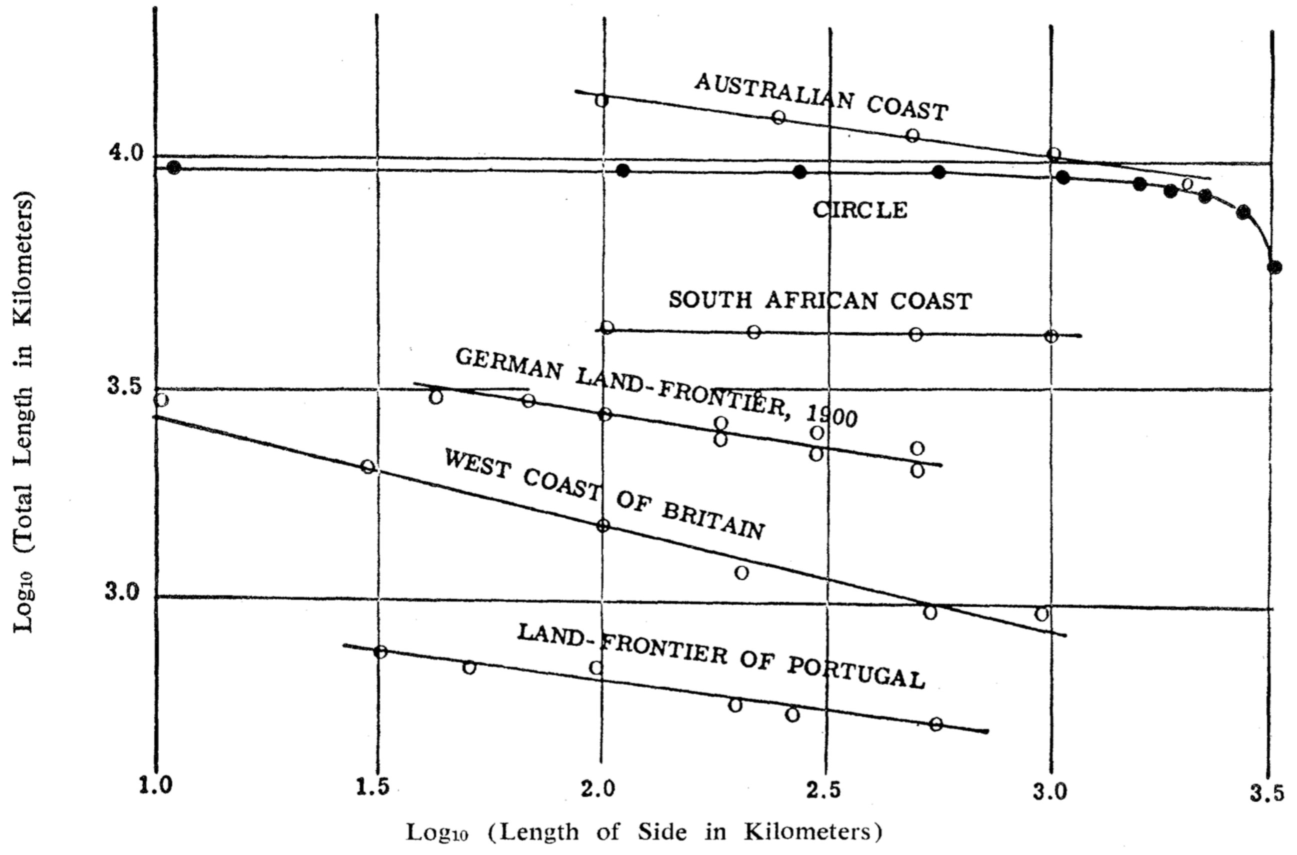
Richardson’s Coastlines

Lewis Fry Richardson (1961)

Measured coastlines at successively halved scales

Log/log plot
- Total length
- Step length

Slope of regression line



B. Mandelbrot (1961)
“How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension”
Science 156:3775, 636–638

Fractal Dimension

Mandelbrot formalised this as **fractal dimension** (1967)

S=100km L=3,400km

S=50km L=2,800km