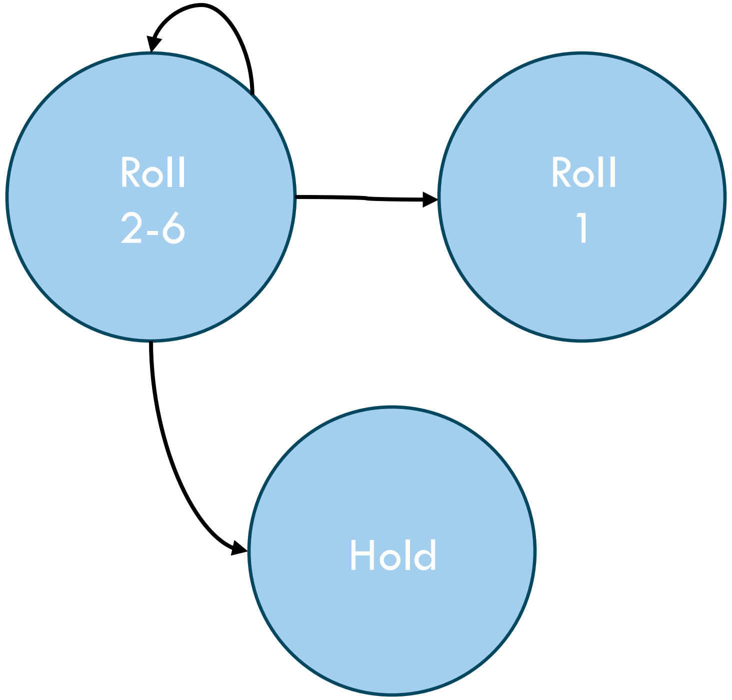
PIG: A DICE GAME OF STRATEGIC CHANCE

By Elder Veliz

12-03-2024

HOW TO PLAY

- **Set-up**: 1 fair, six-sided die and 1+ (non-)human players
- Player(s) alternate turns:
  - If a player rolls a 1, they lose all points **accumulated** in that turn and their turn ends.
  - If a player rolls a 2-6, they can choose to:
    - "hold" and bank the points accumulated in that turn,
    - or roll again.
- The first player to reach or exceed the target score of 100 wins the game.



HEURISTIC STRATEGY

Let ***X*** be the number of successful rolls until the first failure (i.e. roll a 1)

$$X \sim Geom\left(\frac{1}{6}\right),\ \mathbb{E}(X) = \frac{6}{1} - 1 = \textbf{5 rolls}$$

Let ***R*** be the value of a <u>successful</u> die roll

$$R \sim Discrete\ Uniform(2, 6),\ \mathbb{E}(R) = \frac{2+6}{2} = \mathbf{4}$$

We might be inclined to believe X, R are independent, in which case:

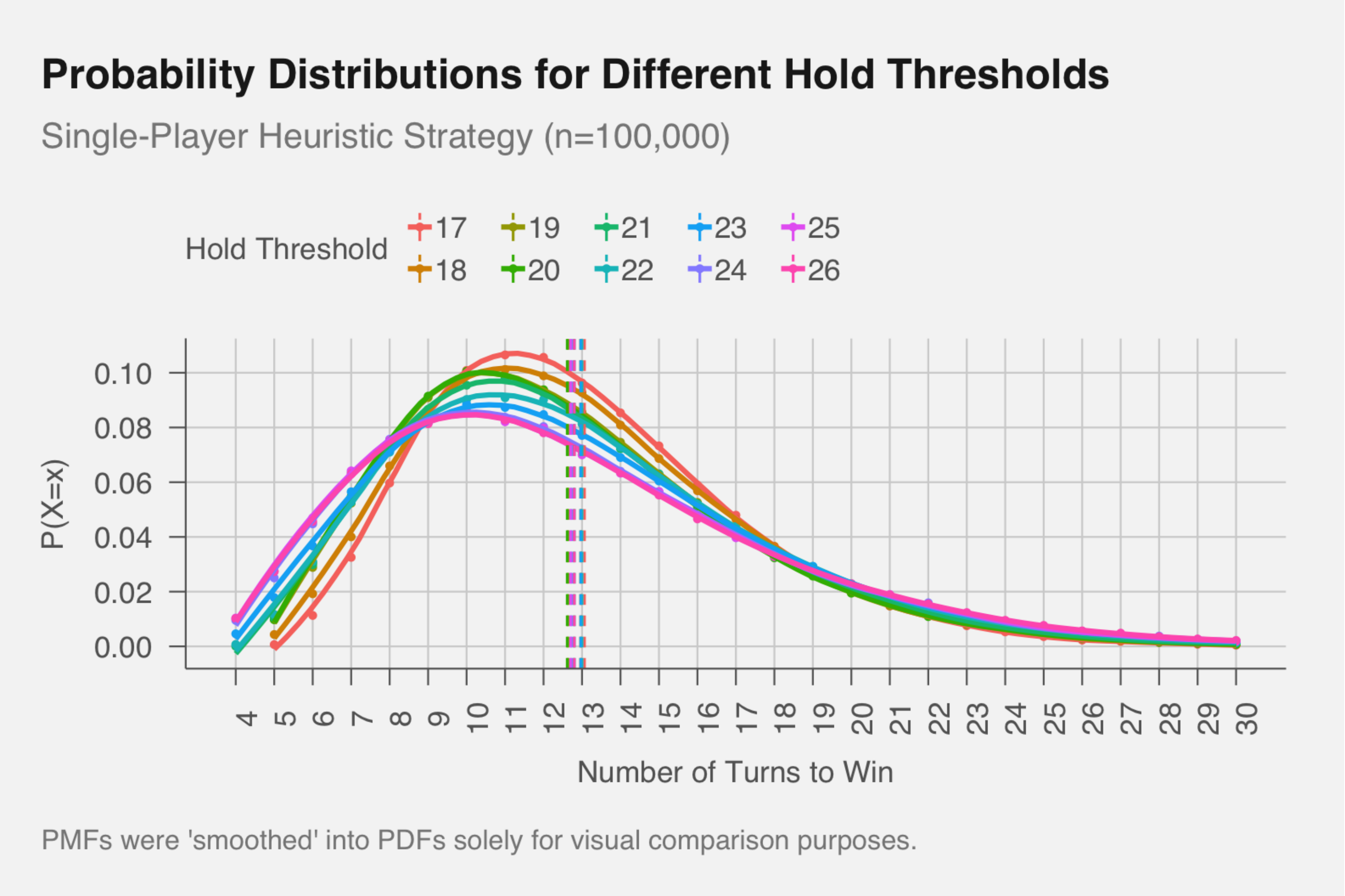
$$\mathbb{E}(X \cdot R) = \mathbb{E}(X) \cdot \mathbb{E}(R) = 5 * 4 = 20$$

This has spurred the following heuristic:

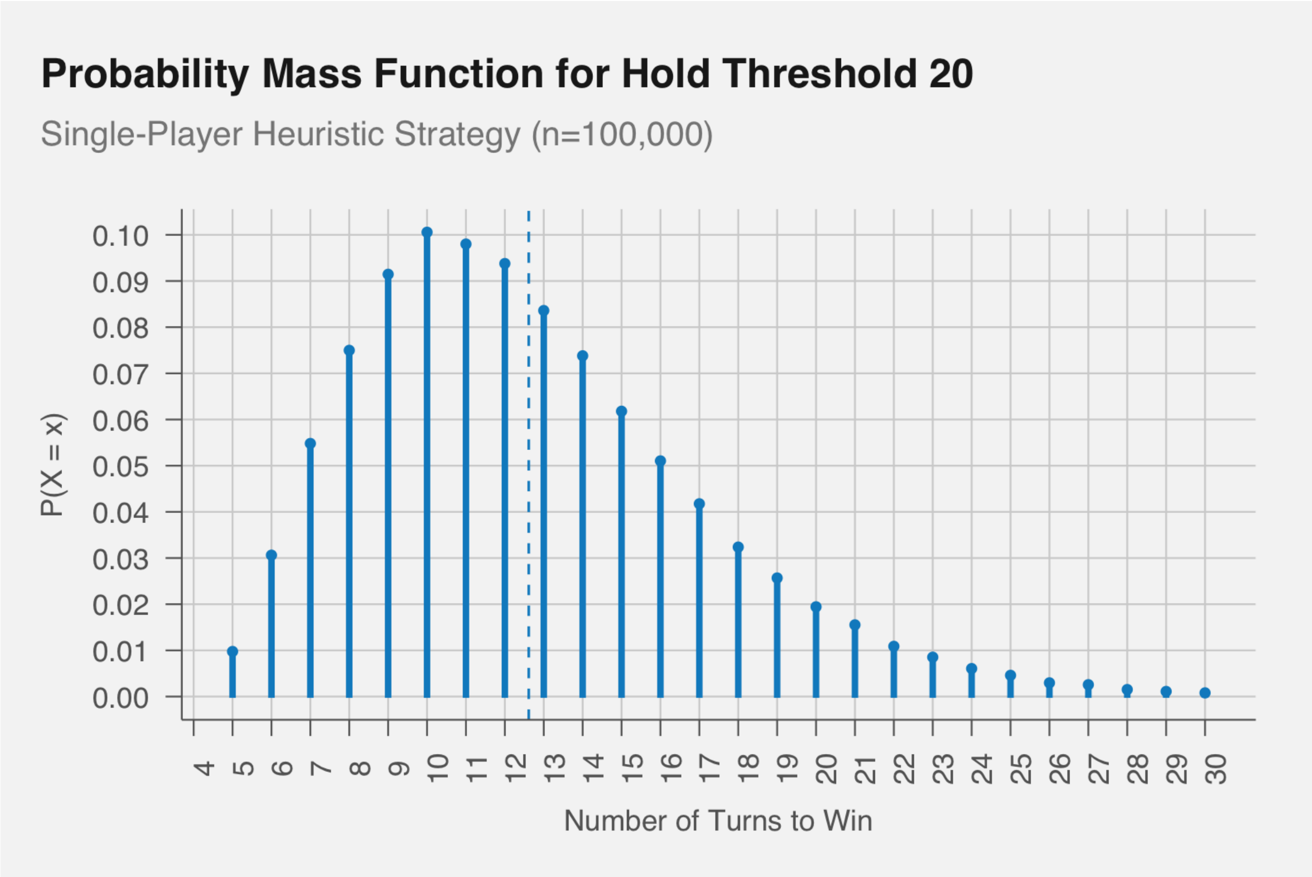
Let T denote the cumulative turn total.

"Roll" while $\boldsymbol{T < 20}$; "Hold" when $\boldsymbol{T \geq 20}$.

RESULTS OF DIFFERENT "HOLD" THRESHOLDS



RESULTS OF DIFFERENT "HOLD" THRESHOLDS



Threshold	Mean	Median	Variance
17	13.06	12	4.03
18	12.96	12	4.25
19	12.64	12	4.38
20	**12.61**	**12**	**4.43**
21	12.80	12	4.58
22	13.00	12	4.81
23	12.96	12	5.06
24	12.73	12	5.23
25	12.70	12	5.28
26	12.80	12	5.39

MARKOV DECISION PROCESS (MDP)

Used to model decision-making in situations where outcomes are jointly under the control of a **decision-maker** and partly **random.**

Assumptions:

Markov Property: $P(s_{t+1}|s_t, a_t, \cdots, s_0, a_0) = P(s_{t+1}|s_t, a_t)$

i.e. the future state depends **only** on the current state and action

Finite State and Action Spaces

Stationary Transition Probabilities and Rewards

Full Observability

Bounded Reward

MARKOV DECISION PROCESS (MDP)

An **agent** will cycle through a series of states, actions, and rewards

Framework:

Set of Possible States

Set of Possible Actions

Reward Model: assigns rewards to given state and action

Value Function: *total* reward expected to accumulate from given state

Transition Model: PMF over next state given current state and action

Policy (π): maps an action to a given state

TOWARD AN "OPTIMAL" STRATEGY

In a 2-player game, observe that Pig is a sequence of transitions between states. Our goal is maximize expected gains while managing risks.

Let the triplet $(\boldsymbol{T}, \boldsymbol{S_{me}}, \boldsymbol{S_{op}})$ define a **game state (S):**

Our current turn total, $\mathbf{T} \in [0,105]$

Our current score, $\boldsymbol{S_{me}} \in [0,105]$

Our opponent's score, $\boldsymbol{S_{op}} \in [0,105]$

At each game state, we can take two **actions (A):**

Roll the die and

a) roll a 2:6, and add its outcome R to the turn total T

b) roll a 1, ending our turn

Hold the turn total T, add it to our current score S_{me}, and end our turn

TOWARD AN "OPTIMAL" STRATEGY

The outcome of the die determines our **transition probabilities** between states:

Roll 1: Our turn ends; state transitions to $(0, S_{me}, S_{op})$ with the opponent's turn.

Roll 2-6: Turn total **T** increases; state transitions to $(T + R, S_{me}, S_{op})$, still our turn.

Define a simple **reward model**:

+1 if we win $(S_{me} + T \geq 100)$

-1 if we lose $(S_{op} \geq 100)$

0 otherwise

SOLVING FOR THE "OPTIMAL" STRATEGY

We use a **zero-sum** framework where maximizing our pay minimizes opponent's payoff

Utility of Holding:

If $(S_{me} + \mathrm{T} \geq 100)$: $V\left(T, S_{me}, S_{op}\right) = 1$

Otherwise: $V_{Hold}\left(T, S_{me}, S_{op}\right) = -V(0, \boldsymbol{S_{op}}, \boldsymbol{S_{me}} + T)$

Utility of Rolling:

$$V_{Roll}\left(T, S_{me}, S_{op}\right) = \frac{1}{6}\left(-V\left(0, \boldsymbol{S_{op}}, \boldsymbol{S_{me}}\right) + \sum_{R=2}^{6} V\left(T + R, S_{me}, S_{op}\right)\right)$$

For given state $s \in \mathrm{S}$, **value function:**

$$V(s) = \max_{a \in \{Roll, Hold\}} Expected\ Utility\ of\ Action\ a$$

Bellman Equation: $V(s) = \max_{a \in A} \sum_{s'} P(s'|s, a)[R(s, a, s') + \gamma V(s')]$

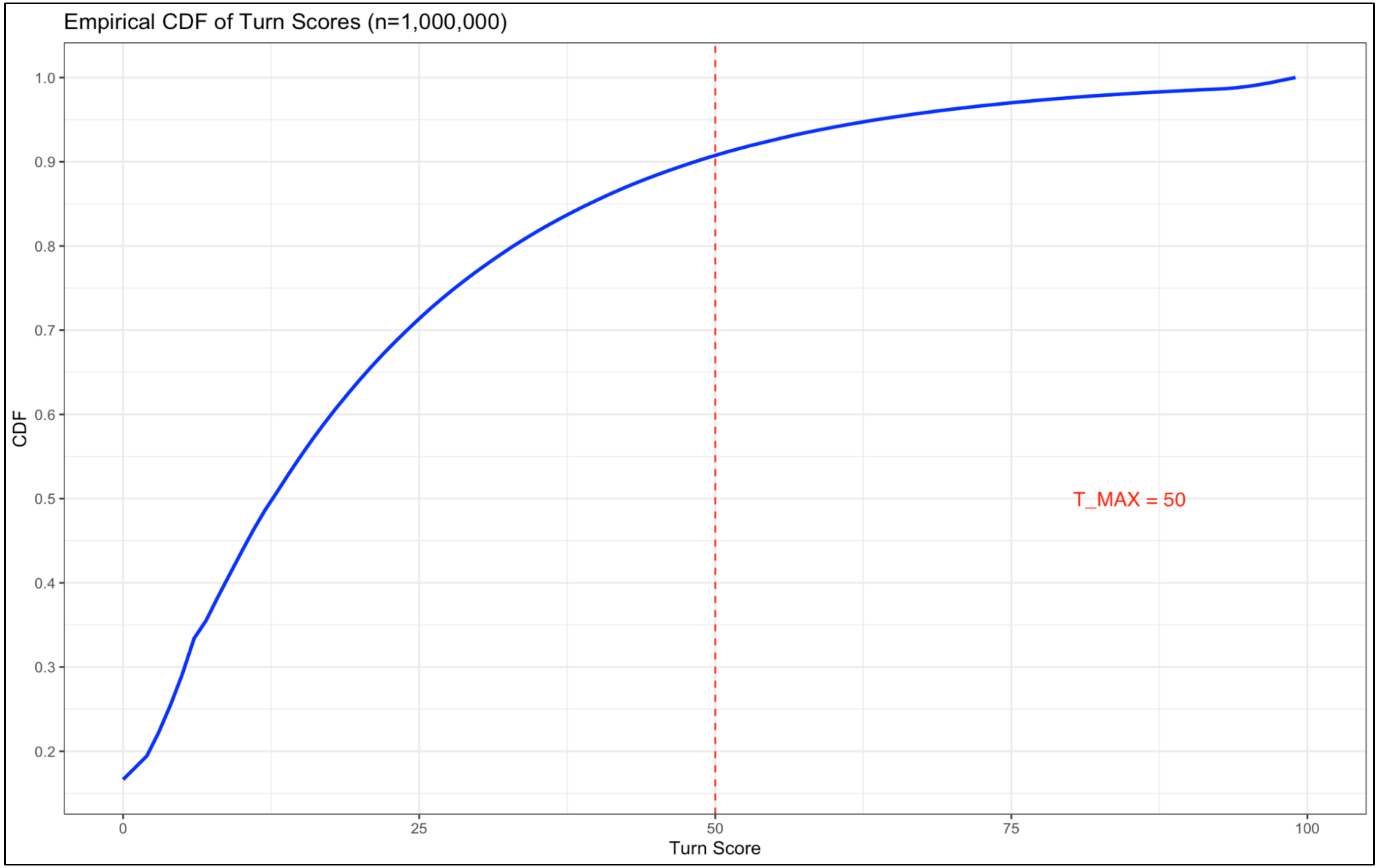
Policy: At each state, action with highest utility recorded as optimal action.

ADJUSTMENTS FOR CONVERGENCE

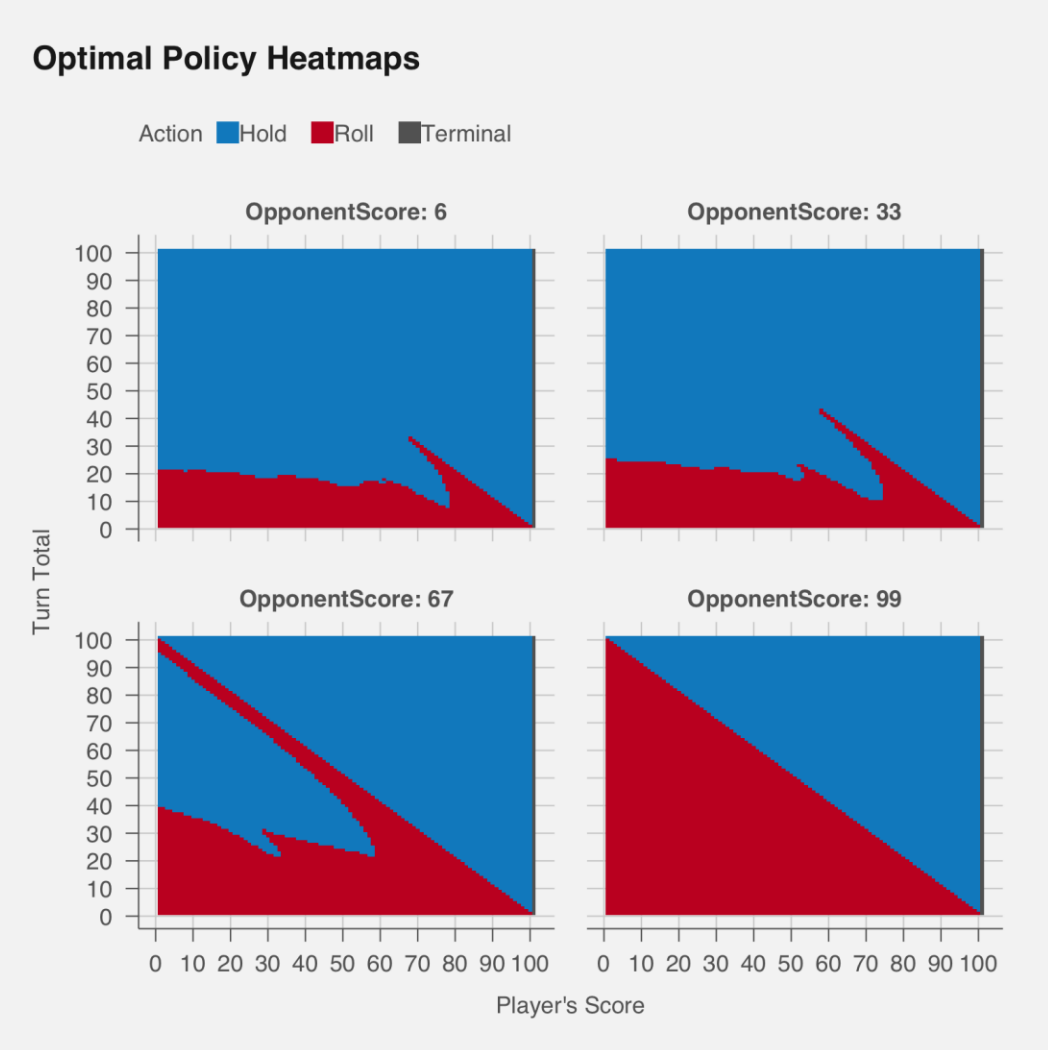
$\boldsymbol{\gamma = 1}$, discount factor chosen emphasize *future* over immediate rewards

$\boldsymbol{T \in [0, 50]}$, turn total truncated to reduce computation

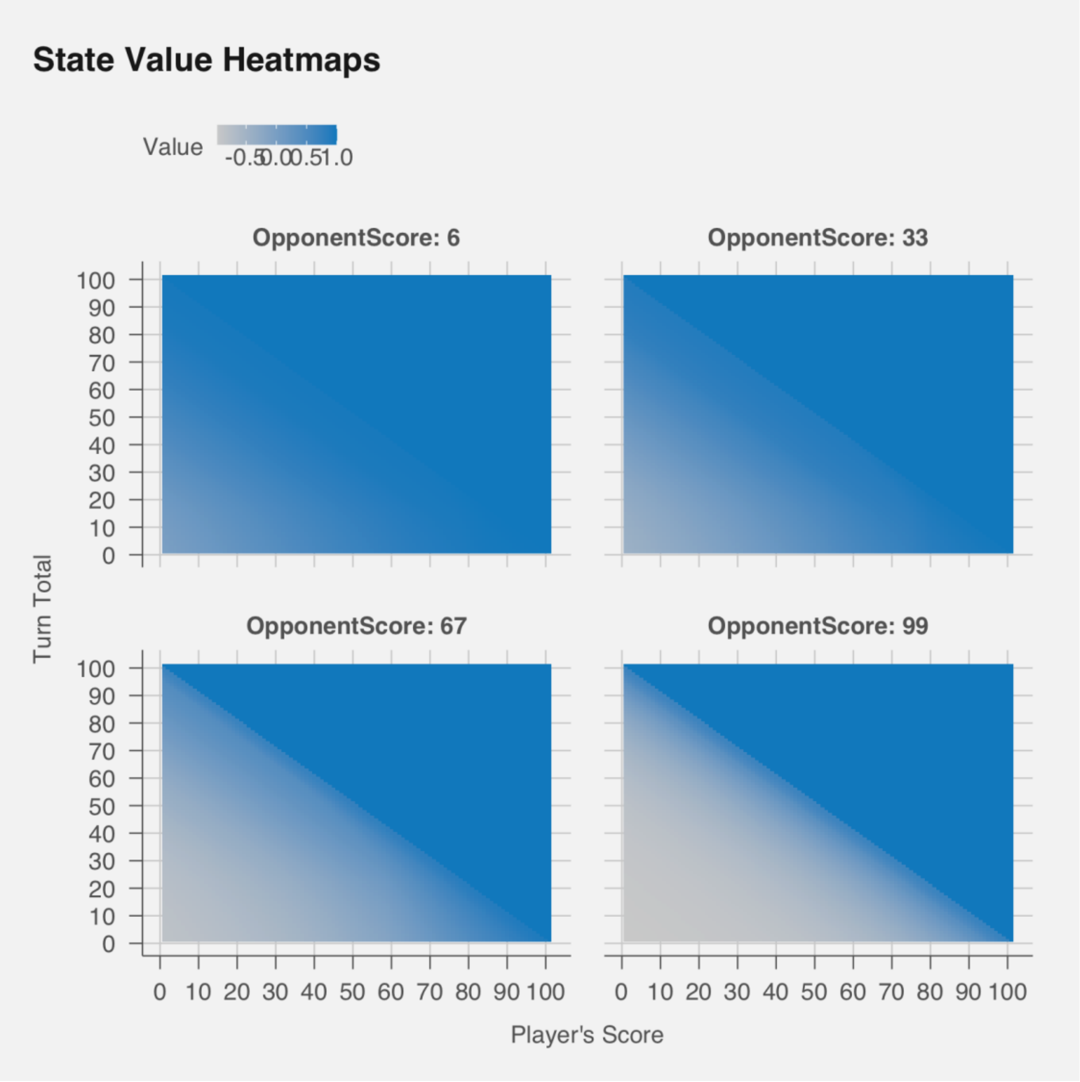
Zero-Sum framework assumes *both* players play optimally



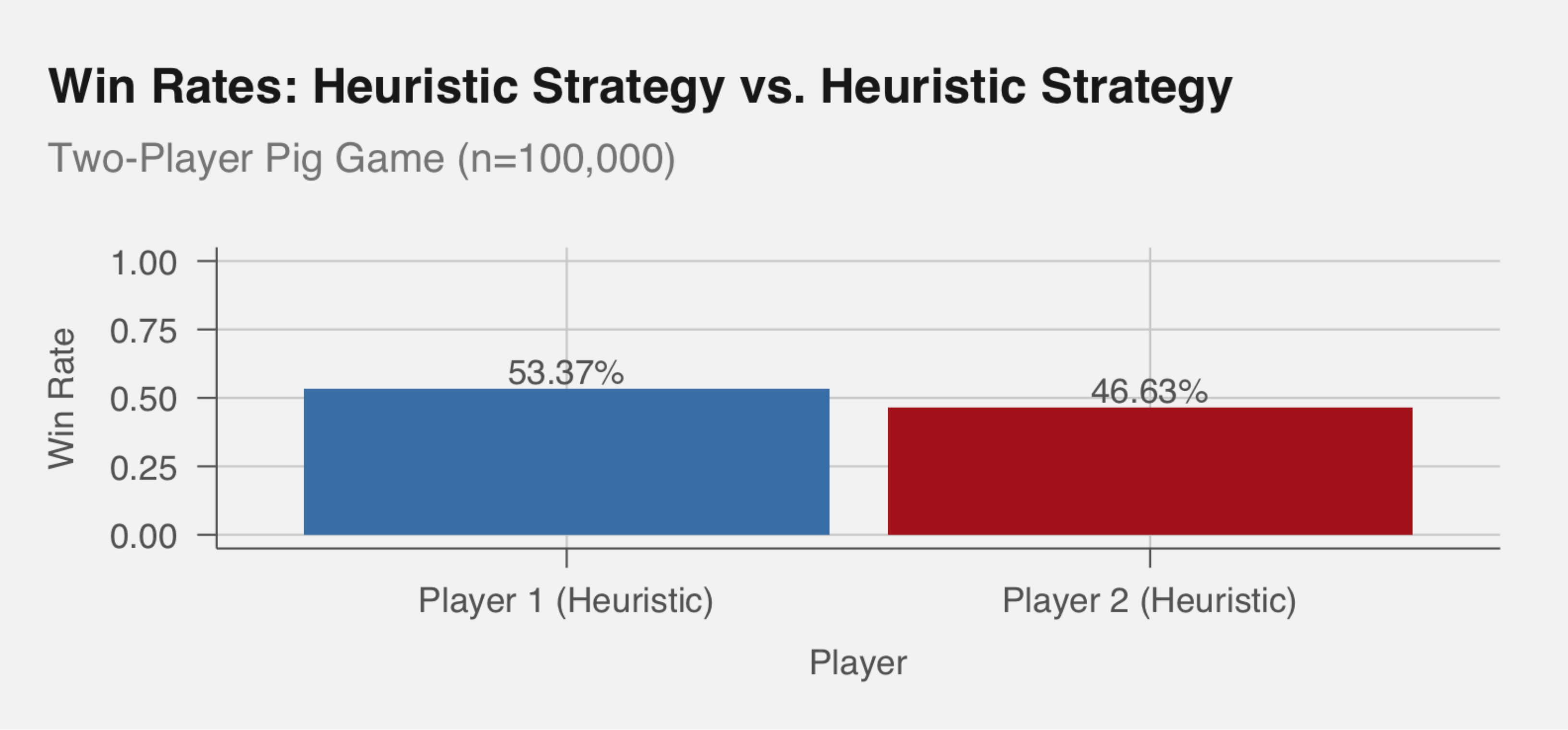
EVIDENCE OF SENSIBLE OPTIMAL POLICY



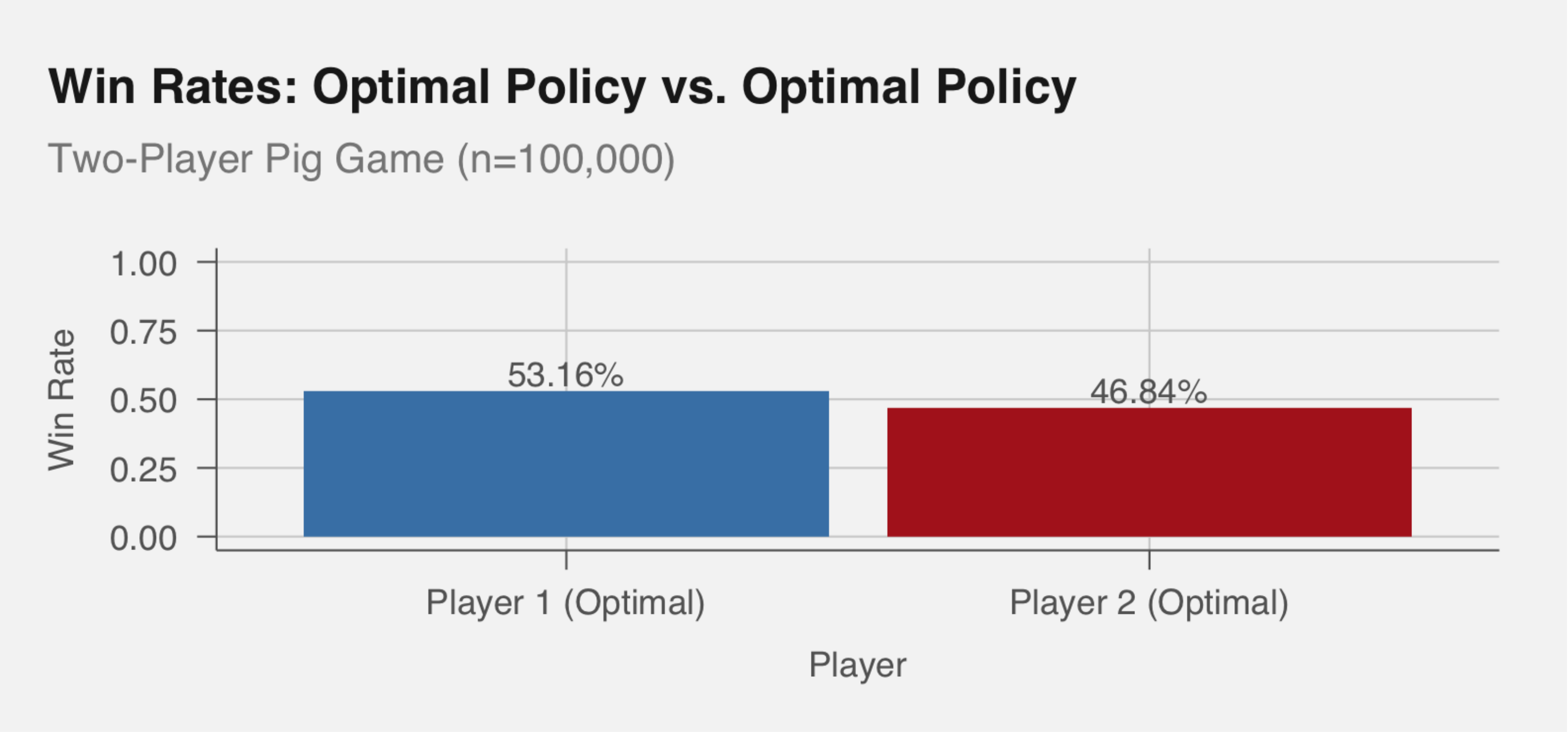
EVIDENCE OF SENSIBLE STATE VALUES



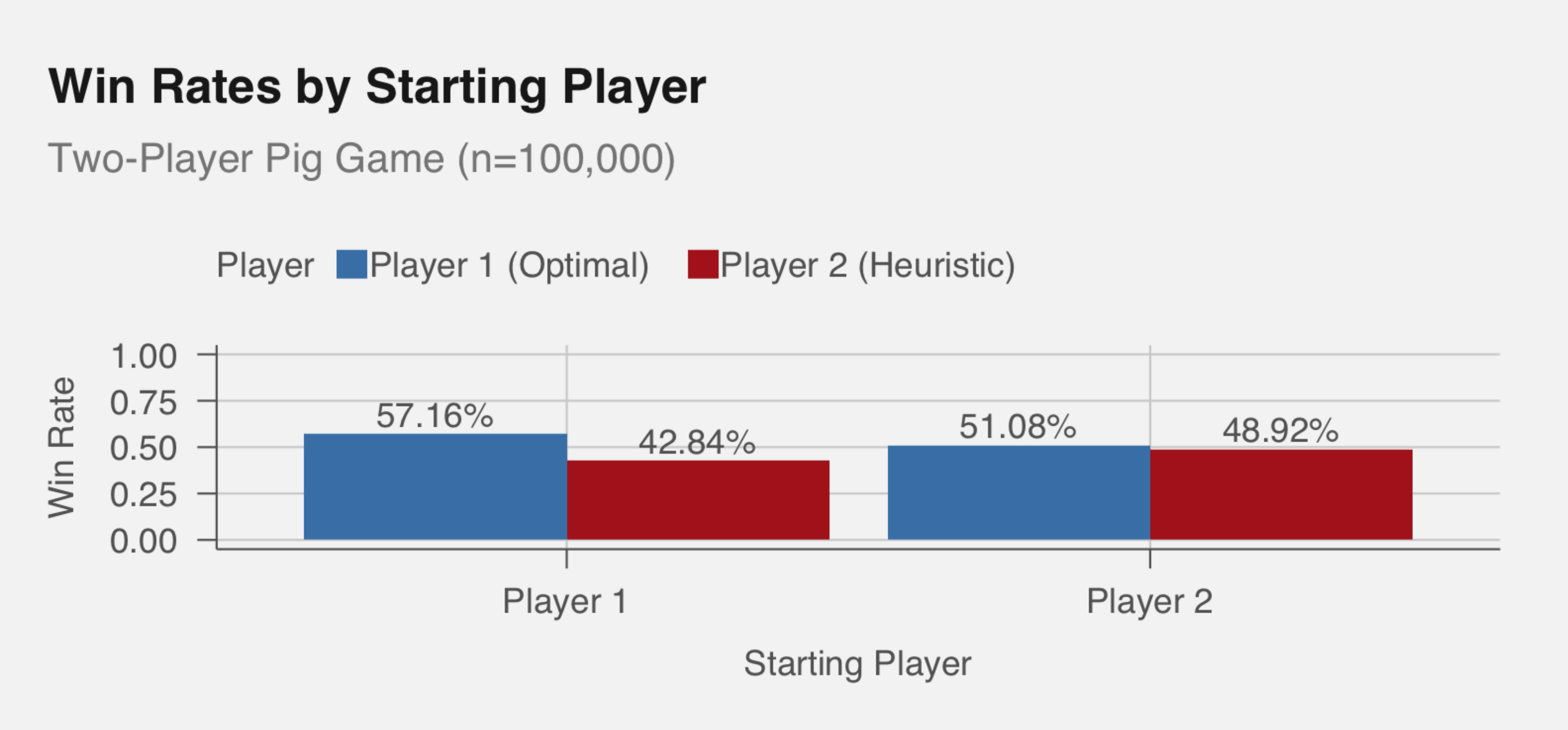
TWO HEURISTIC AGENTS



TWO OPTIMAL AGENTS



OPTIMAL VS. HEURISTIC AGENT



FURTHER INVESTIGATIONS

Starting player advantage

- Determine viable method to balance the game.

Finetuning of γ (balance between immediate and future rewards)

- Will propensity to "hold" with smaller γ yield benefits?

Influence of target score on strategies

- How does decreasing or increasing target score affect heuristic and optimal policy strategies?

Incorporate slight randomness in policy choices

- Will a non-deterministic decision-making process more realistically model human behavior?

REFERENCES

Adams, R. P. (2019). *Markov Decision Processes*. COS324 Elements of Machine Learning, Princeton University. Retrieved from https://www.cs.princeton.edu/courses/archive/spring19/cos324/files/expectimax.pdf

Game On Family, "How to Play the Pig Dice Game," https://gameonfamily.com/pig-dice/.

Lafferty, J. (2024). *Reinforcement Learning: Policy Methods*. S&DS 365 Intermediate Machine Learning, Yale University. Retrieved from https://github.com/YData123/sds365-fa24/raw/main/lectures/lecture-oct-30.pdf