

# **Deterministic Skip-Bo Agent with Hierarchical Positional Value Play**

by  
Anonymous

Submitted to the Department of Computer Science  
in partial fulfillment of the requirements for the degree of  
**MASTER OF SCIENCE IN COMPUTER SCIENCE**

at the  
**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**  
February 2026

© 2026 Anonymous. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Anonymous  
Department of Computer Science  
December 22, 2025

Certified by: Supervisor Name  
Professor of Computer Science, Thesis Supervisor

Accepted by: Graduate Officer  
Professor of Computer Science  
Graduate Officer, Department of Computer Science



# Acknowledgements

This research and the accompanying simulation framework were developed with the assistance of **Gemini (AI by Google DeepMind)**. The AI contributed significantly to the formalization of the HPVP strategy, the implementation of the simulation environment, and the technical typesetting of this document. The conceptual oversight and research objectives were defined by the author.



# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Game Rules and Environment Specification</b>	<b>9</b>
2.1	Cards, Setup and Information . . . . .	9
2.2	Turn Structure . . . . .	9
2.3	Formal Model and Complexity . . . . .	10
<b>3</b>	<b>Hierarchical Positional Value Play (HPVP)</b>	<b>11</b>
3.1	Motivation . . . . .	11
3.2	Algorithm . . . . .	11
<b>4</b>	<b>Evaluation Methodology</b>	<b>13</b>
4.1	Baselines and Experimental Design . . . . .	13
4.2	Metrics and Statistical Analysis . . . . .	13
4.3	Sample Size Determination and Effect Sizes . . . . .	14
<b>5</b>	<b>Implementation and Limitations</b>	<b>15</b>
<b>6</b>	<b>Conclusion</b>	<b>17</b>



# Chapter 1

## Introduction

Classic card games such as Skip-Bo offer controlled environments for investigating heuristics, search and learning methods. Unlike Chess or Go, Skip-Bo has a hidden draw pile and private hands, placing it among the family of stochastic, imperfect-information games. The objective is straightforward—be the first to play all cards from your personal stock pile—but strategic play emerges from managing four discard piles, a limited hand and shared building piles. Despite a large player base, there is little published work on automated Skip-Bo agents. To support reproducible research, we propose a deterministic baseline agent based on simple, transparent heuristics. We also describe a simulation environment, logging format and evaluation procedures so that other agents can be compared under identical conditions.



# Chapter 2

## Game Rules and Environment Specification

### 2.1 Cards, Setup and Information

Skip-Bo is played with a deck of **162 cards**: 144 numbered cards (twelve ranks from 1 to 12, four colours without functional difference) and 18 wild “Skip-Bo” cards. At the start of a two-to-four-player game, the dealer shuffles the deck and deals 30 cards face-down to each player (20 cards for five–six players). These cards form that player’s *stock pile*; the top card is flipped face-up. All undealt cards become a central *draw pile*. In the centre of the table up to four *building piles* will be created during play. Each player also has up to four personal *discard piles*. The agent is assumed to observe its own hand, the top card of its stock pile, the top cards of its discard piles and all current building piles; it does not see the contents of the draw pile or other players’ stock or discard piles.

### 2.2 Turn Structure

Play proceeds clockwise from the dealer. At the beginning of the first turn a player draws five cards from the draw pile as their hand and on subsequent turns draws back up to five cards. A player may then play cards onto the building piles in ascending order; each building pile must start with a card of rank 1 or a Skip-Bo (wild) card. When a building pile reaches rank 12, it is removed from the table and its cards are set aside; when the draw pile is exhausted, all completed building piles are reshuffled to form a new draw pile. There can never be more than four active building piles at once. During a turn players may play cards from three sources: the face-up stock card, cards in hand and the top cards of any discard piles. They may play as many cards as legal; whenever all hand cards have been used they draw back up to five and may continue. A player’s turn ends when they cannot or choose not to make a play; they must then discard exactly one card from hand to one of their personal discard piles. The *winner* of the round is the first player to play every card in their stock pile.

## 2.3 Formal Model and Complexity

We model the game as a two-player Markov decision process (MDP) with hidden state. A game state  $s$  includes the ordered draw pile, each player’s stock pile, four discard piles per player (top cards visible), up to four building piles (top cards visible) and the current hand of the acting player. The *action space*  $A(s)$  comprises all legal moves: playing the top stock card, playing any hand card, playing the top card of any discard pile or discarding a hand card to one of the four discard piles. To guarantee reproducibility we fix a pseudo-random number generator seed for shuffling at the beginning of each match and alternate starting player roles across games.

Solving Skip-Bo optimally is intractable. A coarse lower bound on the size of the state space  $|S|$  is the number of ways to distribute 162 distinct cards into two stock piles, two hands, eight discard piles and up to four building piles, ignoring within-pile order, i.e.,  $\binom{162}{30,30,5,5,4,4,84}$ . This combinatorial explosion motivates the use of heuristic policies and simulation-based evaluation.

# Chapter 3

## Hierarchical Positional Value Play (HPVP)

### 3.1 Motivation

Casual Skip-Bo strategy guides emphasise playing from the stock pile whenever possible and organising discard piles to permit future sequences. HPVP formalises this advice into a deterministic policy. The goal is not to approximate an optimal agent but to provide a transparent baseline against which more sophisticated agents (e.g., Monte-Carlo tree search or reinforcement learning) can be evaluated.

### 3.2 Algorithm

HPVP considers the set of legal actions in a fixed priority order. Table 3.1 summarises the hierarchy. Within each priority class, actions are tie-broken deterministically.

**HPVP Policy.** At each turn the agent attempts to play the top stock card if possible (Priority 1). If that is not possible, it plays the most promising card from its hand (Priority 2). Failing that, it plays the top card of its discard piles (Priority 3) in left-to-right order. If no play is available, it discards a hand card according to the positional discard rule (Priority 4) and ends the turn.

<b>Priority</b>	<b>Action class</b>	<b>Tie-breaking principle</b>
1	Play the face-up stock card onto any building pile	If multiple building piles can accept the card, choose the pile whose top rank is highest ( $< 12$ ) to maximise immediate continuation; if no building piles accept the card, defer to lower priorities.
2	Play hand cards onto building piles	Prefer to extend an existing building pile rather than start a new one; among several options choose the play that leaves the smallest gap between the new top rank and 12.
3	Play the top card of a discard pile onto building piles	Discard piles are processed from left to right. For each pile, if the top card is playable, apply the same tie-breaking as in Priority 2.
4	Discard a hand card	Select a card to discard based on positional discard: choose the discard pile whose current top card is maximally less than 12; if all four piles have top rank 12 or are empty, use the leftmost pile. When selecting a hand card to discard, prefer high-rank cards to preserve low cards for sequences.

Table 3.1: Hierarchical Positional Value Play (HPVP) priorities.

# Chapter 4

## Evaluation Methodology

### 4.1 Baselines and Experimental Design

To evaluate HPVP one must compare it against alternative strategies. Obvious baselines include a random agent (uniformly choosing among legal moves), a myopic greedy agent (playing any legal move that releases the top stock card without deeper planning) and learning-based agents such as Monte-Carlo tree search (MCTS). Each match should consist of a large number of games in which starting roles and random seeds are alternated to mitigate first-player advantage. Logging should record the random seed, the sequence of actions, the number of turns and the identity of the winner for each game. Public release of raw data and code enables replication.

### 4.2 Metrics and Statistical Analysis

**Win Rate ( $\rho$ ).** The fraction of games won by the agent when alternately starting. Confidence intervals can be computed using binomial methods (e.g., Wilson or Clopper–Pearson). To detect small differences in win rate, large sample sizes may be necessary.

**Turn Length ( $\bar{L}$ ).** The average number of turns taken by the winner. Because turn counts can be skewed or heavy-tailed, non-parametric tests (e.g., the Mann–Whitney  $U$  test) are appropriate when normality cannot be assumed. Non-parametric tests make fewer assumptions about the data distribution and compare the ranks of observations from two independent samples.

**Rounds to Win ( $R_{\text{win}}$ ).** The number of complete rounds played before the stock pile is depleted. Medians and inter-quartile ranges should be reported and compared across agents using non-parametric methods.

When comparing multiple agents across several metrics, adjust  $p$ -values for multiple testing (e.g., Holm correction) and report effect sizes. For example, effect sizes could include differences in median turn counts or the probability that a randomly selected game from one agent takes longer than from another.

### 4.3 Sample Size Determination and Effect Sizes

When comparing two agents' win rates ( $p_1, p_2$ ), researchers should compute the number of paired games  $N$  needed to detect a difference  $\delta = p_1 - p_2$  with desired significance  $\alpha$  and power  $1 - \beta$ . Under the asymptotic normal approximation for a difference in proportions, the approximate sample size is

$$N \approx \frac{(z_{\alpha/2} + z_\beta)^2 [p_1(1 - p_1) + p_2(1 - p_2)]}{\delta^2}, \quad (4.1)$$

where  $z_q$  denotes the  $q$ -quantile of the standard normal distribution. For instance, distinguishing a 5% difference in win rate at  $\alpha = 0.05$  with 80% power requires on the order of thousands of games. Exact confidence intervals for a win rate can be computed by inverting the binomial distribution; the Clopper–Pearson interval provides a conservative exact interval.

To compare distributions of turn lengths or rounds-to-win between agents, non-parametric tests such as the Mann–Whitney  $U$  test are appropriate because they do not assume normality. The test ranks all observations from both samples and assesses whether one agent tends to produce longer or shorter games. Parametric tests (e.g., the two-sample  $t$ -test) can be more powerful when sample sizes are large and distributions are approximately symmetric; however, their assumptions should be checked. Regardless of the test used, confidence intervals and effect sizes should accompany  $p$ -values to convey uncertainty.

# Chapter 5

## Implementation and Limitations

The HPVP agent and evaluation framework are implemented in TypeScript and are available under an open-source licence at the repository [stritti/skip-bo-strategy-evaluation](https://github.com/stritti/skip-bo-strategy-evaluation). The project includes a web interface that allows users to simulate games, view replay data and export results. To ensure reproducibility, we specify the game rules and randomisation procedures explicitly and release the code with version tags and commit hashes. Future agents can be integrated by adhering to the same state and action interfaces.

Several factors limit the generality of the current study. First, the analysis assumes standard Skip-Bo rules; variations in stock size, number of players or draw-pile replenishment may change optimal strategies. Second, HPVP is deterministic and does not consider hidden information or opponent modelling; sophisticated strategies could exploit its predictability. Third, this paper focuses on formalising the environment and the HPVP policy rather than presenting exhaustive experimental results. Finally, while non-parametric tests are recommended when distributions are skewed or non-normal, parametric tests can remain reliable with sufficiently large sample sizes due to the central limit theorem.



# **Chapter 6**

## **Conclusion**

This thesis introduces a reproducible framework for evaluating Skip-Bo agents and formalises the Hierarchical Positional Value Play heuristic. By making the rules, information assumptions and randomisation explicit and by providing open-source code and data, we aim to lower the barrier to entry for researchers interested in imperfect-information card games. Future work will integrate search-based and learning-based agents and perform comprehensive empirical comparisons.



# References

1. **Skip-Bo Rules.** “Skip Bo Rules: How to play Skip Bo, PDF and big guide,” Skip-BoRules.com, accessed December 19 2025.
2. **Clopper–Pearson Method.** Statistics How To, “Clopper–Pearson Exact Method,” accessed December 19 2025.
3. **Comparing Two Proportions – Sample Size.** Select Statistical Consultants, “Comparing Two Proportions – Sample Size,” accessed December 19 2025.
4. **Mann–Whitney U Test Explained.** Jim Frost, “Mann–Whitney U Test Explained,” Statistics By Jim, accessed December 19 2025.
5. **Survey of Monte Carlo Tree Search Methods.** Cameron Browne et al., “A Survey of Monte Carlo Tree Search Methods,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, 2012, pp. 1–43.
6. **Evaluating Affective Feedback of the 3D Agent Max.** Christian Becker et al., “Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game,” in *Affective Computing and Intelligent Interaction*, 2005.