

Deterministic Skip-Bo Agent with Hierarchical Positional Value Play

Abstract

Skip-Bo is a commercial card game of imperfect information in which players race to clear their personal stock piles. Research on artificial agents for this domain is scarce, yet the game’s simple rules and large state space make it an attractive benchmark for sequential decision-making under uncertainty. This paper formalises a deterministic baseline agent, **Hierarchical Positional Value Play** (HPVP), that prioritises actions from different sources (stock, hand, discard) and proposes an evaluation framework that emphasises reproducibility and statistical rigour. The underlying rule set, information available to the agent and randomisation procedures are made explicit to enable fair comparison and replication. A companion web application and open-source repository provide a reference implementation and datasets for further study.

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.

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 [1]. 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) [1]. These cards form that player’s *stock pile*; the top card is flipped face-up [1]. All undealt cards become a central *draw pile* [1]. In the centre of the table up to four *building piles* will be created during play [1]. Each player also has up to four personal *discard piles* [1]. 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/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 [1]. 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 [1]. 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 [1]. There can never be more than four active building piles at once [1]. 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 [1]. 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 [1]. The *winner* of the round is the first player to play every card in their stock pile [1].

2.3 Notation and assumptions

We model the game as a two-player Markov decision process 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.

2.4 Formal model and complexity

Skip-Bo can be formalised as a Markov decision process (MDP) with hidden information. For each state s in the state space S , the available actions $A(s)$ are the legal moves described above. A transition kernel $T(s, a)$ yields a distribution over next states, capturing randomness from shuffling and drawing. A reward function $R(s, a)$ assigns +1 to terminal states in which the agent has emptied its stock pile and 0 elsewhere; intermediate rewards can be assumed zero because the objective is to maximise the probability of eventual victory. Solving this MDP optimally is intractable: the state space grows combinatorially with the number of cards. A coarse lower bound on $|S|$ 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.

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 [1]. 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 1 summarises the hierarchy. Within each priority class, actions are tie-broken deterministically.

Priority	Action class	Tie-breaking principle
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	For each playable hand card, prefer to extend an existing building pile rather than start a new one; amongst 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 <i>positional discard</i> : 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 1: Hierarchical Positional Value Play (HPVP) priorities.

Algorithm 1 (HPVP policy).

1. **Stock play:** If the top stock card can legally be played onto a building pile, play it using the tie-breaking rule for Priority 1 and return.
2. **Hand plays:** Generate the set of hand cards playable on building piles. If non-empty, select the card and target pile by the tie-breaker for Priority 2, play it and return.
3. **Discard-pile plays:** For each discard pile from left to right, if its top card can be played on a building pile, play it according to the tie-breaker in Priority 3 and return.
4. **Discard:** If none of the above apply, select a hand card to discard (preferring highest ranks) and a discard pile according to the positional discard rule in Priority 4; place the card on that pile and end the turn.

This procedure is deterministic given the current state. Because the agent never draws additional cards until its hand is empty, it may use multiple cards per turn. When all sources are exhausted, the discard step ends the turn.

3.3 Relation to existing methods

HPVP is designed as a hand-crafted heuristic baseline rather than an optimal agent. It differs from simple *greedy* play—playing the first legal card without regard to source—in its explicit hierarchy and in the positional discard rule. It also differs from *random* agents that sample uniformly from legal moves. Future work can compare HPVP to search-based agents; for example, Monte-Carlo tree search (MCTS) methods have been extensively surveyed in the literature [7] and have shown success in games with large branching factors.

4 Related work

Although Skip-Bo is widely played, scientific literature on automated Skip-Bo agents is scarce. Browne et al. survey Monte-Carlo tree search (MCTS) methods and discuss their application to card games, but Skip-Bo is not evaluated in that survey (see Ref. [7]). In affective computing research, Skip-Bo has been used as a scenario for evaluating empathic virtual agents. Becker et al. employed the game in their study of the 3D agent *Max*, describing it as a turn-taking card game in which players discard cards onto central stacks in ascending order; the same paper notes that players attempt to discard eight cards by building stacks numbered 1 through 12 (see Ref. [8]). These works treat Skip-Bo primarily as a vehicle for human-agent interaction rather than optimising gameplay. To our knowledge, no prior work has published an explicit baseline agent or evaluation framework for Skip-Bo. Our contribution is to formalise a deterministic heuristic policy, provide open implementations and data and lay out statistical methodology for comparative evaluation.

5 Evaluation methodology

5.1 Baselines

To assess HPVP we propose comparing it against several baseline strategies:

1. **Random:** At each decision point choose uniformly from all legal moves. This agent provides a lower bound on performance and helps calibrate win-rate differences.
2. **Greedy:** Always select the first available legal move in a fixed ordering of sources (stock, hand, discard) but without the positional discard heuristic.
3. **HPVP:** The deterministic policy defined above.

Additional baselines may include one-step look-ahead agents or learned policies from reinforcement learning libraries (e.g., RLCard), but these require separate implementations and are beyond this paper’s scope.

5.2 Experimental design

Because Skip-Bo is stochastic and turn order confers an advantage, experimental design must control for variance and bias. For each matchup between two agents we recommend:

- **Paired games:** Play an even number N of games with fixed seeds; in half of them Agent A starts and in the other half Agent B starts. Use the same shuffles and starting positions across pairings so differences in win rate reflect agent policy rather than card order.
- **Sample size:** Choose N via power analysis for the expected effect size. For win-rate comparisons (binary outcome), the standard error of the difference in proportions decreases with $1/\sqrt{N}$. Small effect sizes (e.g., difference of 0.05) require more games than large differences.
- **Logging:** Record for each game the shuffle seed, sequence of actions, number of turns, number of completed building piles and the winner. Publicly release raw data and code for reproducibility.

5.3 Metrics and statistical analysis

- **Win rate (ρ):** The fraction of games won by the agent when alternately starting. Use confidence intervals based on the binomial distribution (e.g., Wilson or Clopper–Pearson) to quantify uncertainty.
- **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., Mann–Whitney U test) are appropriate when normality cannot be assumed. Non-parametric tests make fewer assumptions about the data distribution [4, 6].
- **Rounds to win (R_{win}):** The number of complete rounds played before the stock pile is depleted. Analyse using median and inter-quartile range and compare agents with non-parametric methods.

When comparing multiple agents across several metrics, adjust p -values for multiple testing (e.g., Holm correction) and report effect sizes.

5.4 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 α and power $1 - \beta$. Under the asymptotic normal approximation for a difference in proportions, the approximate sample size is given by

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

where z_q denotes the q -quantile of the standard normal distribution. For instance, distinguishing a 5% difference in win rate (e.g., 0.55 vs. 0.50) at $\alpha = 0.05$ with 80% power requires roughly 6 500 games. Investigators should adjust N based on expected effect size and practical constraints [3]. For a given observed win rate \hat{p} over n games, exact confidence intervals can be computed by inverting

the binomial distribution. The Clopper–Pearson interval, which inverts the cumulative binomial distribution and uses Beta distribution quantiles, provides a conservative exact interval for p [2].

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 require no assumption of normality. The U test compares the ranks of observations from two independent samples and assesses whether one agent tends to produce longer or shorter games. It is important to note that the U statistic reflects stochastic dominance rather than a direct difference of medians. When the sample sizes are sufficiently large and the distributions are approximately symmetric, parametric tests like the two-sample t -test can be more powerful. Regardless of the test used, researchers should report effect sizes alongside p -values; for example, the difference in median turn counts or the common language effect size (probability that a randomly selected game from one agent exceeds one from another). All reported statistics should include confidence intervals to convey uncertainty. [4, 5]

6 Implementation and companion artefacts

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`. 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 in Section 2 and release the code with version tags and commit hashes. Future agents can be integrated by adhering to the same state and action interfaces.

7 Limitations and threats to validity

Several factors limit the generality of the current study:

- **Rule variants:** The analysis assumes standard Skip-Bo rules; variations in stock size, number of players or draw-pile replenishment may change optimal strategies. Researchers should document any deviations from the rules described in Section 2 to ensure comparability.
- **Deterministic policy:** HPVP is deterministic and does not consider hidden information or opponent modelling. Sophisticated strategies could exploit its predictability; however, as a baseline this transparency is valuable.
- **Absence of empirical results:** This paper focuses on formalising the environment and the HPVP policy rather than presenting exhaustive experimental results. The companion web project collects data for thousands of games; analysing these results and comparing to learning-based agents is future work.
- **Statistical assumptions:** While we advocate non-parametric tests for skewed metrics, some parametric tests remain robust under mild deviations from normality [5]. Researchers should inspect data distributions before choosing tests.

8 Conclusion

This paper 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,” SkipBoRules.com, accessed 19 December 2025. This online article provides the official Skip-Bo rules, including deck composition (144 numbered cards and 18 wild cards), stock sizes for different player counts, the formation of draw, building and discard piles, and strategic advice such as playing from the stock pile whenever possible.
2. **Clopper–Pearson Exact Method.** Statistics How To. “Clopper–Pearson Exact Method,” StatisticsHowTo.com, accessed 19 December 2025. Describes the Clopper–Pearson exact binomial confidence interval as an alternative to the normal approximation; the method inverts a binomial test and uses quantiles of the Beta distribution.
3. **Comparing Two Proportions – Sample Size.** Select Statistical Consultants. “Comparing Two Proportions – Sample Size,” Select-Statistics.co.uk, accessed 19 December 2025. Presents a formula for calculating the sample size required to detect a difference between two proportions using critical values of the normal distribution.
4. **Mann–Whitney U Test Explained.** Jim Frost. “Mann–Whitney U Test Explained,” StatisticsByJim.com, accessed 19 December 2025. Explains the Mann–Whitney U test as a non-parametric test comparing two independent groups using ranks and notes that the test determines whether one population tends to produce higher values than another.
5. **Nonparametric Tests vs. Parametric Tests.** Jim Frost. “Nonparametric Tests vs. Parametric Tests,” StatisticsByJim.com, accessed 19 December 2025. Discusses when parametric tests can be used with non-normal data and explains that parametric tests can remain robust with sufficiently large sample sizes due to the central limit theorem.
6. **Nonparametric Statistical Tests for Continuous Data.** Francis S. Nahm, “Nonparametric statistical tests for the continuous data: the basic concept and the practical use,” *Korean Journal of Anesthesiology*, 2016. Available at: [pmc.ncbi.nlm.nih.gov/articles/PMC4754273/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC4754273/). Notes that when the normality assumption is violated, non-parametric tests—based on ranks rather than raw values—are appropriate alternatives to parametric tests.
7. **Survey of Monte Carlo Tree Search Methods.** Cameron Browne, Edward Powley, Daniel Whitehouse, 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. DOI: 10.1109/TCI-AIG.2012.2186810. Reviews Monte Carlo tree search algorithms and their applications to games.

8. **Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game.** Christian Becker, Helmut Prendinger, Mitsuru Ishizuka and Ipke Wachsmuth, “Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game,” in *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, 2005. Available at: www.becker-asano.de/210.pdf. Uses Skip-Bo as a card-game scenario for evaluating empathic agent feedback; describes players discarding cards onto central stacks in ascending order and attempting to build stacks numbered 1–12.