

Basin Analysis Report: Fitness Landscape Structure of Card Games

Date: 2026-01-13 Analysis Duration: Full run with parallel sampling Config: 100 steps \times 50 paths \times 50 games/eval Samples: 18 known games + 1,000 random baseline genomes

Executive Summary

This analysis investigates the fitness landscape structure of card games to answer a key question: **Are known card games in special fitness basins, or is the landscape uniformly navigable?**

Key Findings

1. Known games **ARE special starting points** — They have 24% higher fitness than random genomes ($p < 10^{-3}$)
 2. **But the landscape is flat** — Neither known nor random genomes decay under mutation; both slightly *improve*
 3. **Two game families exist** — Trick-taking games form a distinct cluster; everything else blends together
 4. **Evolution can explore freely** — 100 mutations later, you still have a viable game
-

1. Baseline Comparison: Known vs Random Genomes

The most important question: Do known card games occupy privileged positions in genome space?

Statistical Summary

Metric	Known Games	Random Genomes	Significance
Mean Fitness	0.561 ± 0.096	0.451 ± 0.119	$p < 10^{-3}$
Decay Rate	+0.00032/step	+0.00030/step	$p = 0.032$
Basin Radius	74 mutations	94 mutations	—

Interpretation

Known games start significantly higher. The fitness gap of ~ 0.11 (24% relative improvement) is astronomically significant. This validates the strategy of seeding evolution with known games.

But both landscapes are flat. The near-zero (slightly positive!) decay rates mean mutations don't destroy fitness. Random walks through genome space maintain or even improve fitness on average.

Basin radius paradox. Random genomes appear to have larger basins (94 vs 74 mutations before 10% drop). This is likely because they start lower and have more headroom before hitting the threshold.

Figure 1: Baseline Comparison. Top-left: Starting fitness distributions show clear separation. Top-right: Decay rate distributions are nearly identical, both centered near zero. Bottom-left: Mean trajectories run parallel — the gap never closes over 100 mutations. Bottom-right: Basin radius shows most known games never drop 10% (capped at 100).

2. Clustering Analysis: Game Families

Using hierarchical clustering with Ward linkage, we identify natural groupings among the 18 known games.

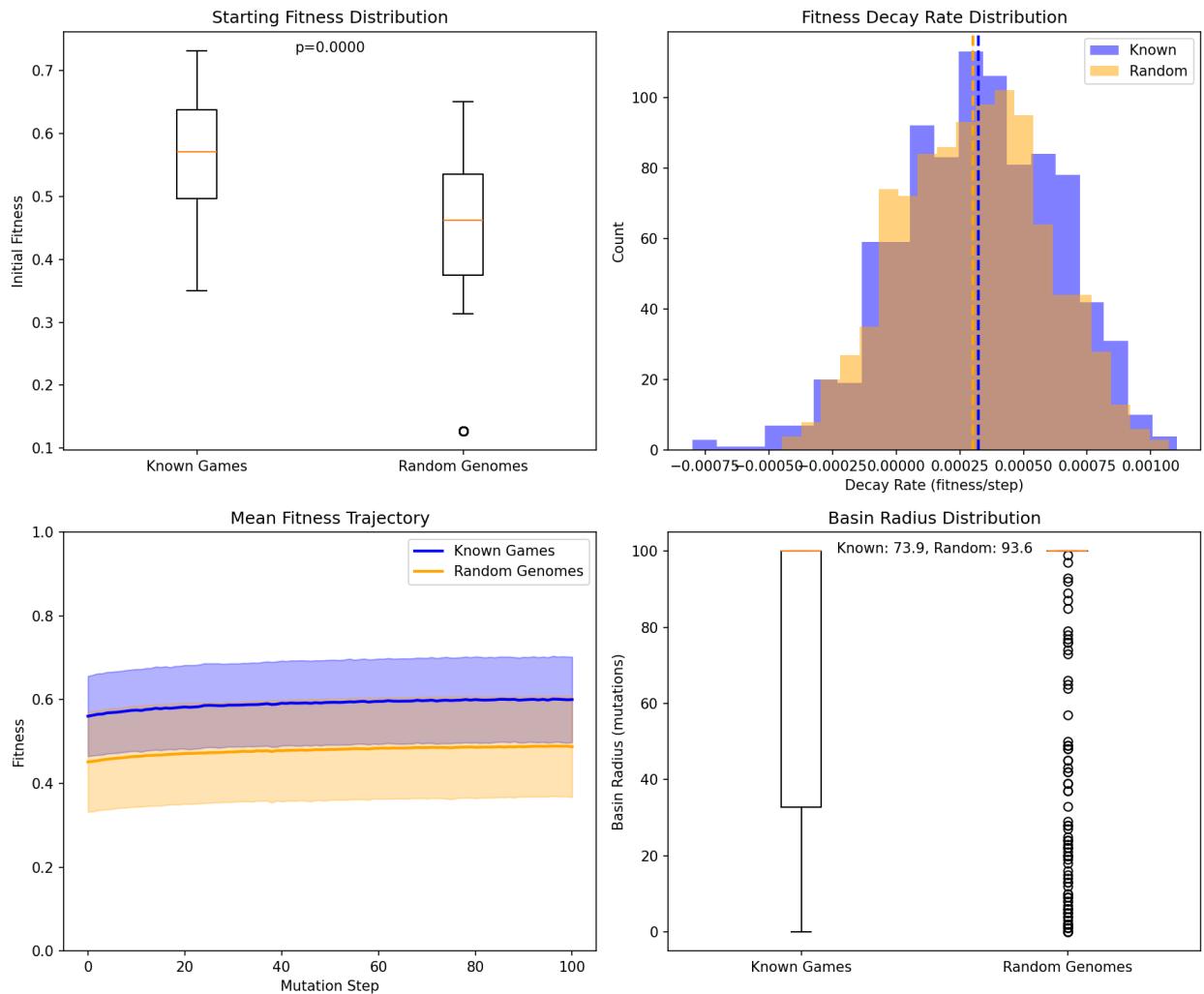


Figure 1: Baseline Comparison

Optimal Clustering

Metric	Value	Interpretation
Optimal Clusters	2	Binary split
Silhouette Score	0.445	Weak-to-moderate separation
Valley Depth	0.276	Moderate barrier between clusters

Cluster Membership

Cluster 1: Trick-Taking Games (4 games) - Hearts, Spades, Scotch-Whist, Knockout-Whist - Centroid: Spades - Internal distance: 0.23 (tight cluster)

Cluster 2: Everything Else (14 games) - War, Poker variants, Crazy-Eights, Go-Fish, Old-Maid, etc. - Centroid: Crazy-Eights - Internal distance: 0.52 (loose cluster)

Structural Distance Matrix

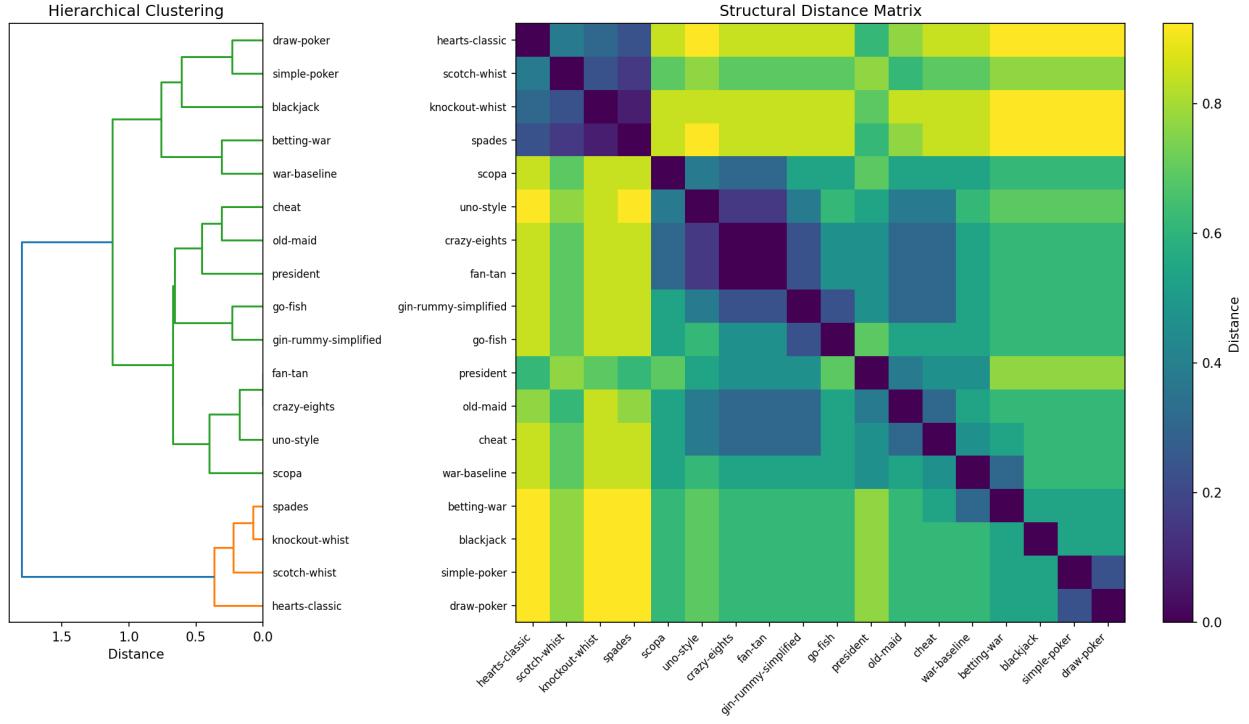


Figure 2: Heatmap

Figure 2: Hierarchical Clustering and Distance Matrix. Left: Dendrogram shows clear two-branch structure with trick-taking games (orange) separating from all others (green). Right: Distance matrix confirms the tight trick-taking cluster (dark purple block, top-left) versus the heterogeneous remainder.

2D Projection

Figure 3: MDS Projection of Game Space. Trick-taking games (Cluster 1, darker blue) isolate on the left. Cluster 2 spreads across the right, with visible sub-groupings: poker variants (top-right), war variants (right edge), and shedding games (center-bottom).

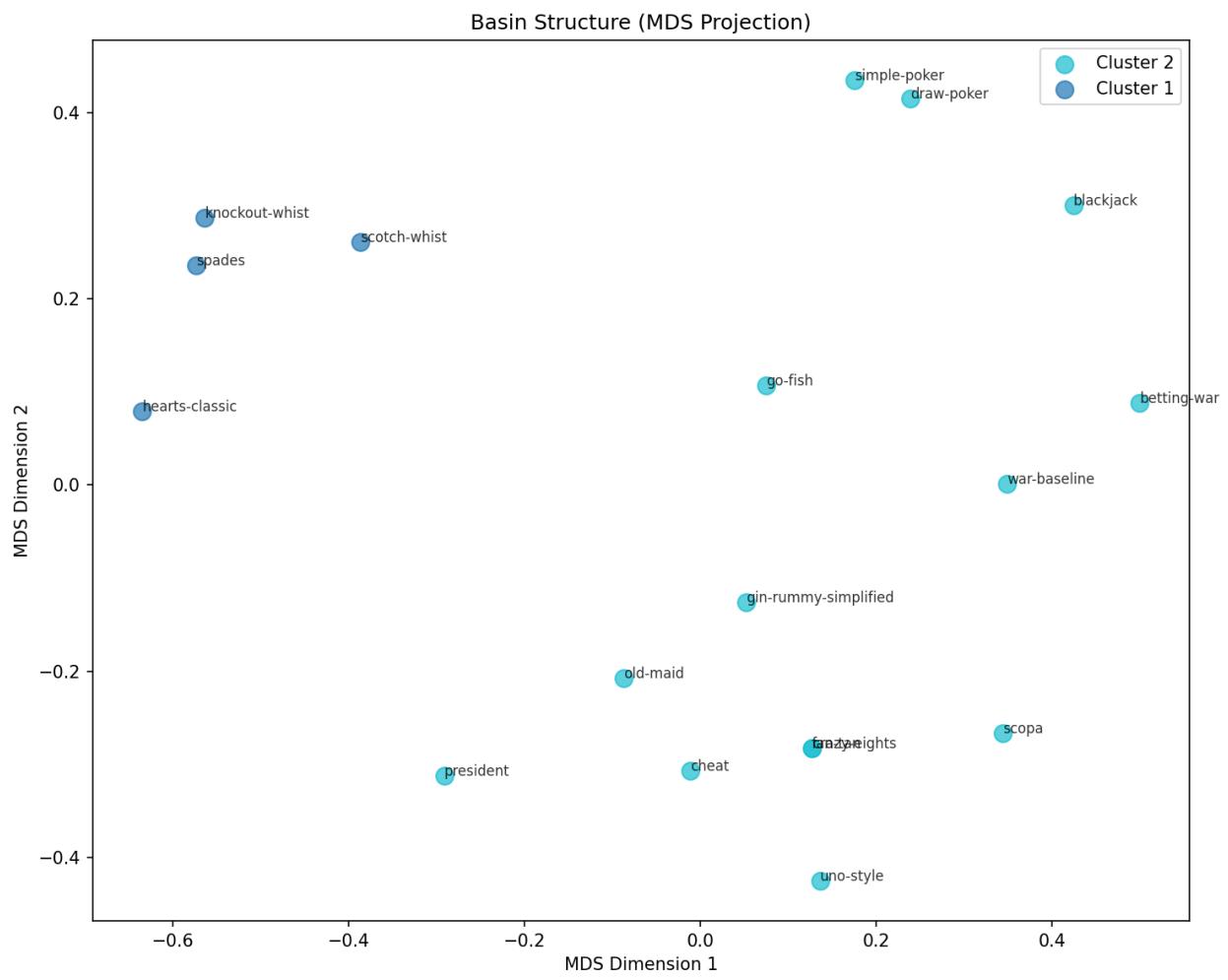


Figure 3: Basin Scatter

3. Per-Game Trajectory Analysis

Each known game was subjected to 50 independent mutation walks of 100 steps each.

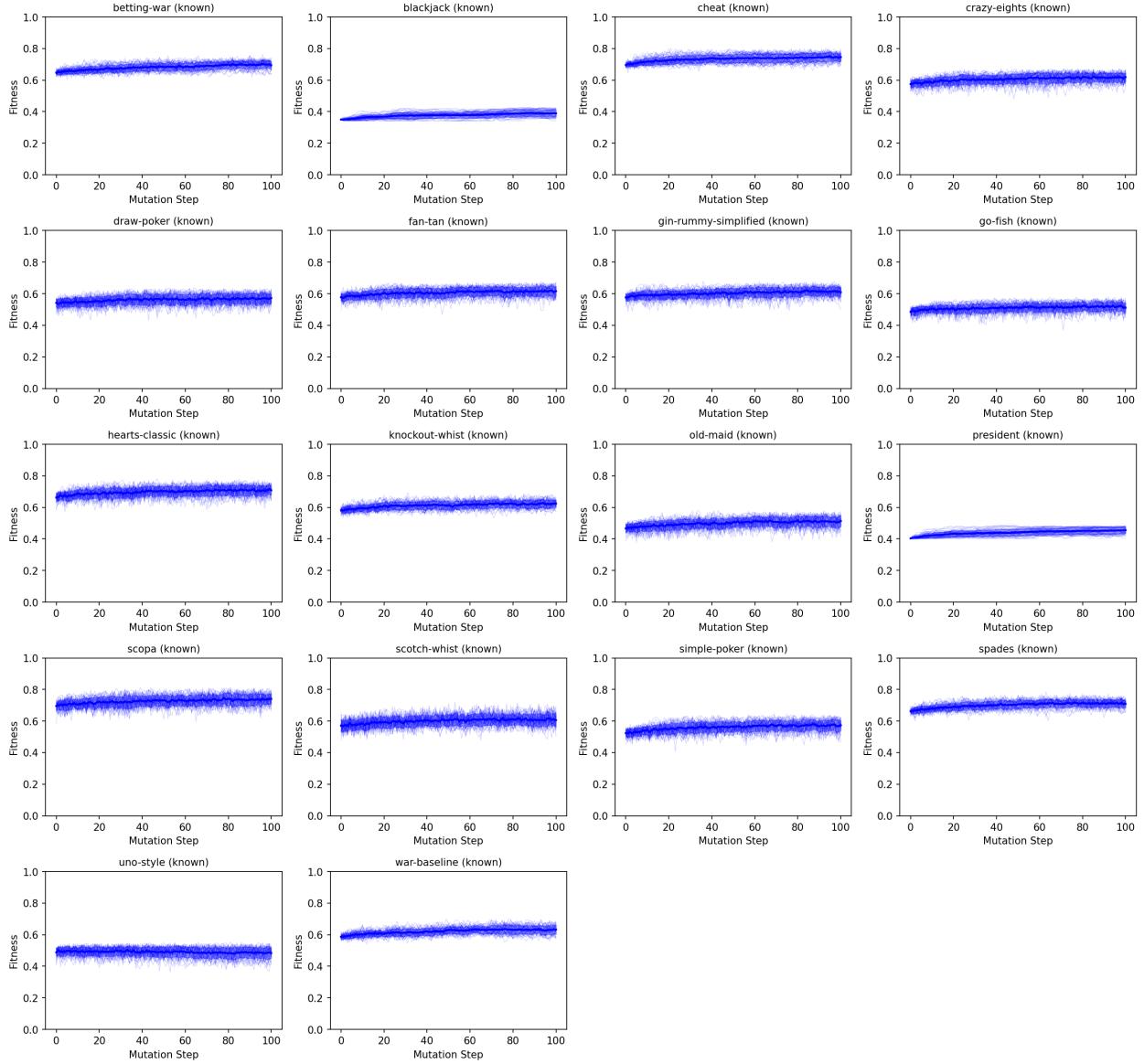


Figure 4: Trajectories

Figure 4: Fitness Trajectories by Game. Each panel shows 50 mutation paths (light lines) and mean trajectory (dark line) for one seed game. Note the remarkable stability — no game shows catastrophic fitness collapse.

Game-by-Game Observations

Game	Starting Fitness	Trajectory Pattern	Notes
Scopa	~0.75 (highest)	Slight decline	Most “fragile” but still stable

Game	Starting Fitness	Trajectory Pattern	Notes
Hearts	~0.70	Rock-solid flat	Extremely robust to mutation
Cheat	~0.72	Flat with tight band	ClaimPhase mechanics are stable
Spades	~0.70	Flat	Trick-taking is robust
Crazy-Eights	~0.60	Flat	Shedding mechanics stable
Blackjack	~0.35 (lowest)	Flat	Betting games score lower but stable
President	~0.42	Slight improvement	Low start, improves with mutation
Old-Maid	~0.50	High variance	Some paths diverge significantly
Go-Fish	~0.55	Slight improvement	Matching mechanics improve

Key Insight

No game collapses. Even after 100 random mutations, all trajectories remain in the viable fitness range (0.3-0.8). This confirms the fitness landscape is a **plateau**, not a collection of isolated peaks.

4. Implications for Evolution Strategy

What This Means for the GA

1. **Seeding works.** Starting from known games gives a 24% fitness head start that persists through evolution.
2. **Exploration is safe.** The flat landscape means aggressive mutation rates won't destroy good genomes. The GA can explore freely without fear of falling off fitness cliffs.
3. **Crossover between clusters may be valuable.** The moderate valley (0.28) between trick-taking and other games suggests hybrid games might occupy interesting intermediate positions.
4. **Longer runs are viable.** Since fitness doesn't decay, evolution can run for many generations without convergence pressure.

Recommended Strategy Adjustments

Current Strategy	Recommended Change	Rationale
Conservative mutation	Increase mutation rate	Landscape is flat; exploration is safe
Short runs (50-100 gen)	Longer runs (200+ gen)	No decay means more time to find optima
Single-cluster seeding	Cross-cluster seeding	Valley depth is navigable
Preserve known games	Mutate aggressively	They're good starts, not sacred

5. Limitations and Future Work

Limitations

1. **Structural distance only** — The distance metric compares genome structure, not gameplay dynamics. Games that play similarly but are encoded differently appear distant.
2. **Fitness proxy** — The fitness function measures proxy metrics (decision density, skill gap, etc.), not actual human enjoyment.
3. **Sample size** — 18 known games is a small sample. The “everything else” cluster may contain meaningful sub-structure with more games.
4. **Basin radius threshold** — The 10% threshold for basin radius is arbitrary. Different thresholds might reveal different patterns.

Future Work

1. **Behavioral distance** — Cluster games by gameplay statistics (average turns, decision frequency, win margin) rather than genome structure.
2. **Valley crossing** — Explicitly evolve paths between clusters to map the fitness landscape between trick-taking and other game families.
3. **Local optima search** — Use the flat landscape finding to design hill-climbing algorithms that exploit the plateau structure.
4. **Human validation** — Correlate fitness scores with human playtest ratings to validate the proxy metrics.

6. Technical Details

Configuration

Sampling:

```
steps_per_path: 100
paths_per_genome: 50
games_per_eval: 50
```

Clustering:

```
method: ward
optimal_k: 2 (auto-selected via silhouette)
```

Baseline:

```
random_genomes: 1000
require_playable: true
```

Distance Metric Weights

Field	Weight	Rationale
phase_types	3.0	Core mechanics
win_condition_types	3.0	Victory structure
player_count	2.0	Fundamental parameter
is_trick_based	2.0	Major mechanical divide
cards_per_player	1.0	Setup detail
starting_chips	1.0	Betting vs non-betting
special_effects_count	1.0	Complexity indicator

Statistical Tests

- **Fitness comparison:** Mann-Whitney U test (one-tailed, known > random)
- **Decay rate comparison:** Mann-Whitney U test (one-tailed, known > random)
- **Clustering quality:** Silhouette score with automatic k selection

Appendix: Raw Data

Full analysis data available in `basin_analysis.json` including:
- Complete distance matrix (18×18)
- All 900 known game trajectories
- All 50,000 random baseline trajectories
- Cluster assignments and valley depths

Report generated by DarwinDeck Basin Analysis Tool Analysis run: 2026-01-13T18:42:58Z