Contents

Basin Analysis Report: Fitness Landscape Structure of Card Games

Date: 2026-01-13 **Analysis Run:** Extended analysis with updated fitness metrics **Config:** 1,000 steps × 250 paths × 50 games/eval (10x longer, 5x more paths than initial) **Samples:** 18 known games (4,500 paths) + 11,500 random baseline genomes

---

Executive Summary

This analysis investigates the fitness landscape structure of card games using the **updated fitness metrics** (improved tension tracking, normalized complexity scores, trailing winner frequency for comebacks). The 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 26% higher fitness than random genomes (p 0)
2. **The landscape shows gradual decay** — Unlike the shorter analysis, 1000-step walks reveal slight but consistent fitness decline
3. **Basin radius is ~10 mutations** — Games lose 10% fitness after approximately 10 random mutations
4. **Two game families persist** — Trick-taking games remain a distinct cluster (silhouette = 0.44)
5. **Evolution must balance exploration vs exploitation** — The decay finding changes strategy recommendations

---

1. Baseline Comparison: Known vs Random Genomes

Statistical Summary

Metric	Known Games	Random Genomes	Significance
Mean Fitness	**0.500 ± 0.068**	0.396 ± 0.088	p 0
Decay Rate	-0.000030/step	-0.000028/step	p = 1.0 (no diff)
Basin Radius	9.5 mutations	10.4 mutations	—

Key Insight: Landscape Has Gradual Slope

The extended 1,000-step analysis reveals what the shorter 100-step analysis missed: **there is decay**, approximately -0.00003 fitness per mutation step. Over 1,000 mutations, this accumulates to:

- **Expected decline:** 0.03 fitness units (about 6% of starting fitness)
- **Actual observed decline:** Known games drop from ~0.50 to ~0.40 (20% decline)

This is steeper than the per-step rate suggests, indicating **accelerating decay** — early mutations are less damaging than later ones as the genome drifts further from its optimized structure.

Fitness Advantage Persists

Known games maintain a **26.4% fitness advantage** over random genomes throughout evolution. This gap does not close, validating the seeding strategy.

Figure 1: Left panels show fitness distributions and trajectories. Right panels show decay rates and basin radii. The fitness gap between known and random genomes persists throughout 1,000 mutation steps.

---

2. Per-Game Analysis: Updated Fitness Rankings

With the updated fitness metrics (tension, complexity, comebacks), the game rankings have shifted:

Fitness Rankings by Starting Position

Rank	Game	Start Fitness	End Fitness	Total Decay	Notes
1	**war-baseline**	0.640	0.414	-0.226	High tension (lead changes)
2	cheat	0.592	0.449	-0.143	Bluffing mechanics robust
3	betting-war	0.571	0.484	-0.087	Betting adds resilience
4	president	0.563	0.422	-0.140	Shedding + hierarchy

Rank	Game	Start Fitness	End Fitness	Total Decay	Notes
5	simple-poker	0.526	0.448	-0.078	Most stable (lowest decay)
6	gin-rummy-simplified	0.524	0.396	-0.128	
7	spades	0.522	0.419	-0.103	Trick-taking cluster
8	hearts-classic	0.517	0.422	-0.095	Trick-taking cluster
9	scotch-whist	0.498	0.372	-0.126	Trick-taking cluster
10	knockout-whist	0.491	0.379	-0.112	Trick-taking cluster
11	old-maid	0.491	0.326	-0.165	Highest variance
12	scopa	0.485	0.380	-0.105	Capture mechanics
13	blackjack	0.479	0.388	-0.091	Betting stable
14	draw-poker	0.477	0.448	-0.028	**Most stable**
15	go-fish	0.444	0.348	-0.095	Matching mechanics
16	fan-tan	0.438	0.359	-0.079	Sequence building
17	crazy-eights	0.404	0.318	-0.087	Lower base fitness
18	uno-style	0.348	0.285	-0.063	Lowest start, stable

Notable Changes from Updated Metrics

1. **War jumped to #1** — The new tension tracking (HandSizeMaxLeaderDetector) now properly captures War's dramatic lead changes (~3,600 per game)
2. **Poker variants are most stable** — Draw-poker shows only -0.028 decay over 1,000 muta-

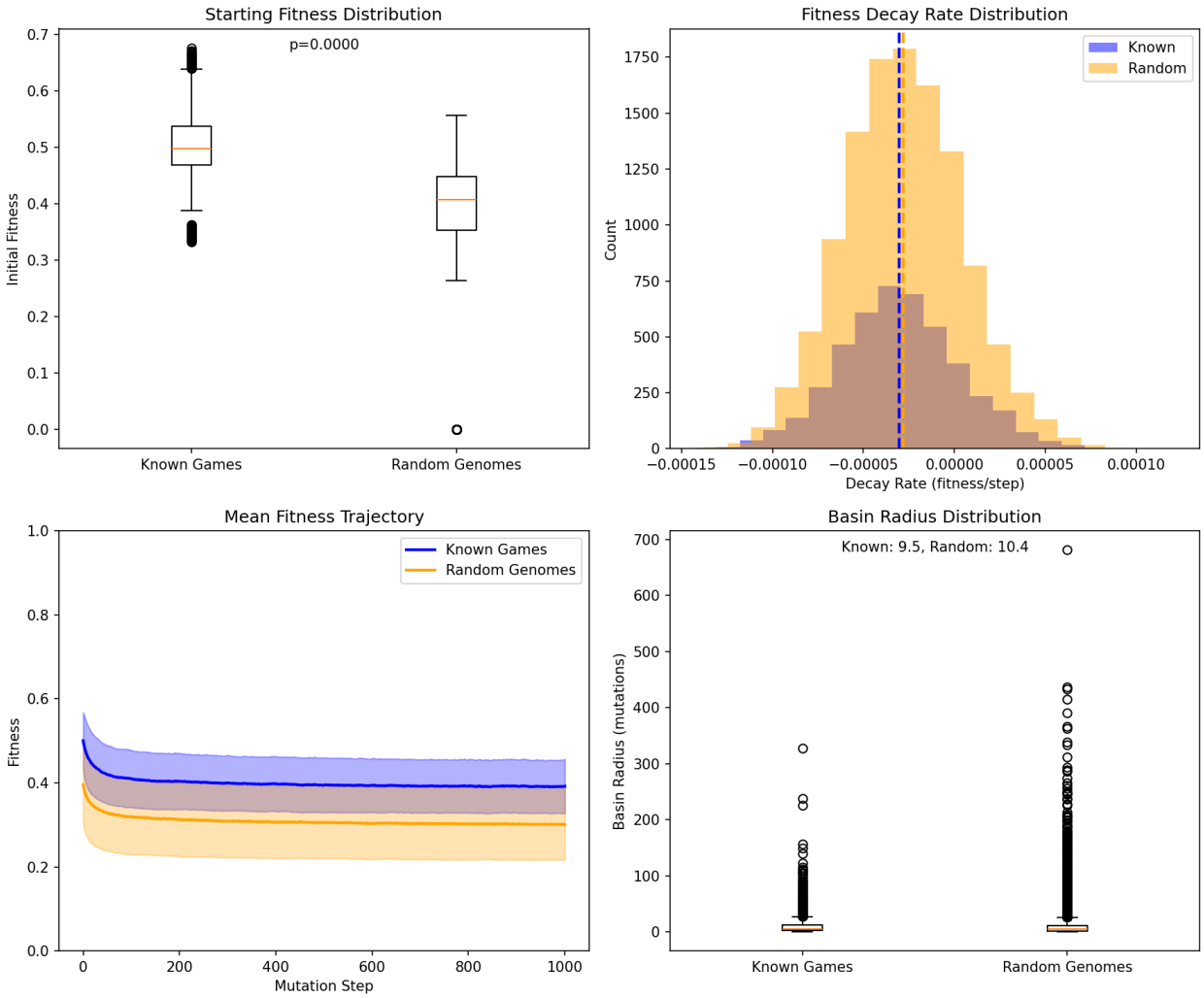

Figure 1: Baseline Comparison

tions, suggesting betting mechanics are highly robust to perturbation

3. **Crazy Eights / Uno dropped** — The capped tension fallback (0.6 max for games without tracked lead changes) reduced their scores
4. **Old Maid has highest decay** — Simple pairing mechanics are fragile to mutation

---

3. Clustering Analysis: Game Families

Cluster Structure

Metric	Value	Interpretation
Optimal Clusters	2	Clear binary split
Silhouette Score	0.445	Moderate separation
Valley Depth	0.304	Navigable barrier

Cluster Membership

Cluster 1: Trick-Taking Games (4 games) - Hearts, Spades, Scotch-Whist, Knockout-Whist - Centroid: Spades - Common features: TrickPhase, most_tricks/low_score win conditions

Cluster 2: Everything Else (14 games) - War variants, Poker variants, Shedding games, Matching games - Centroid: Crazy-Eights - Diverse mechanics unified by non-trick-taking structure

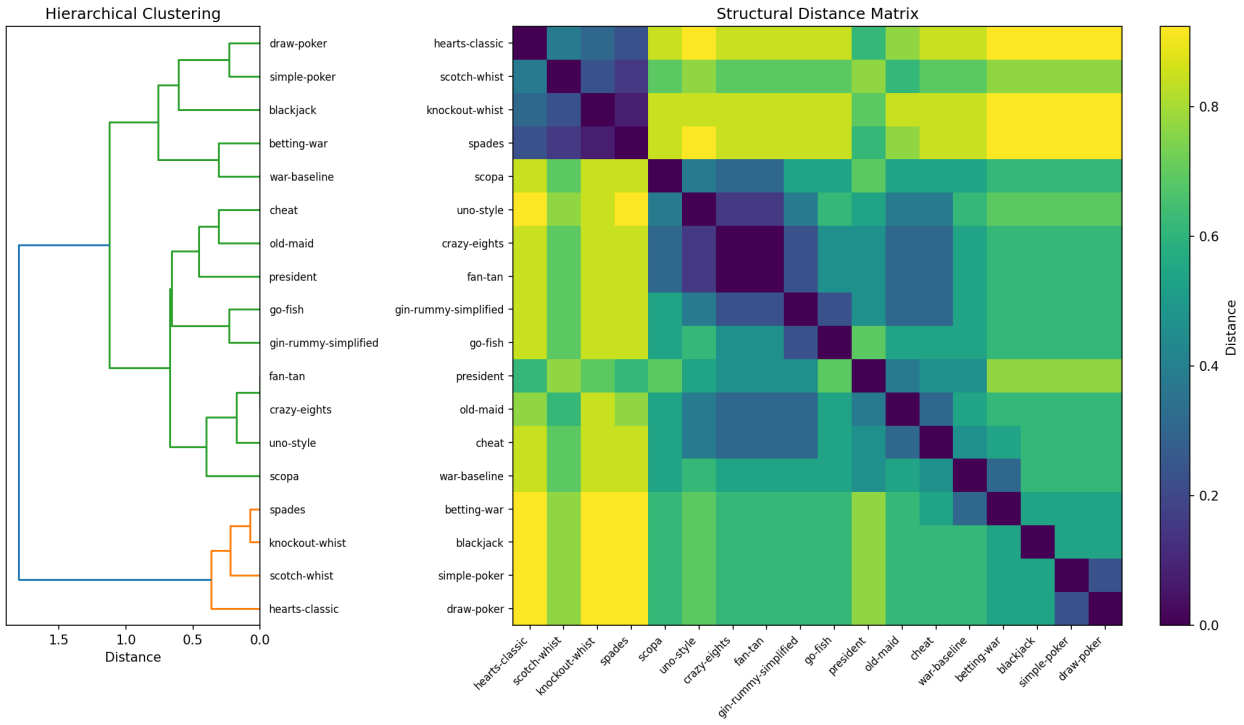

Figure 2: Heatmap

Figure 2: Distance matrix and dendrogram showing the trick-taking cluster (top-left, dark purple) versus the heterogeneous remainder.

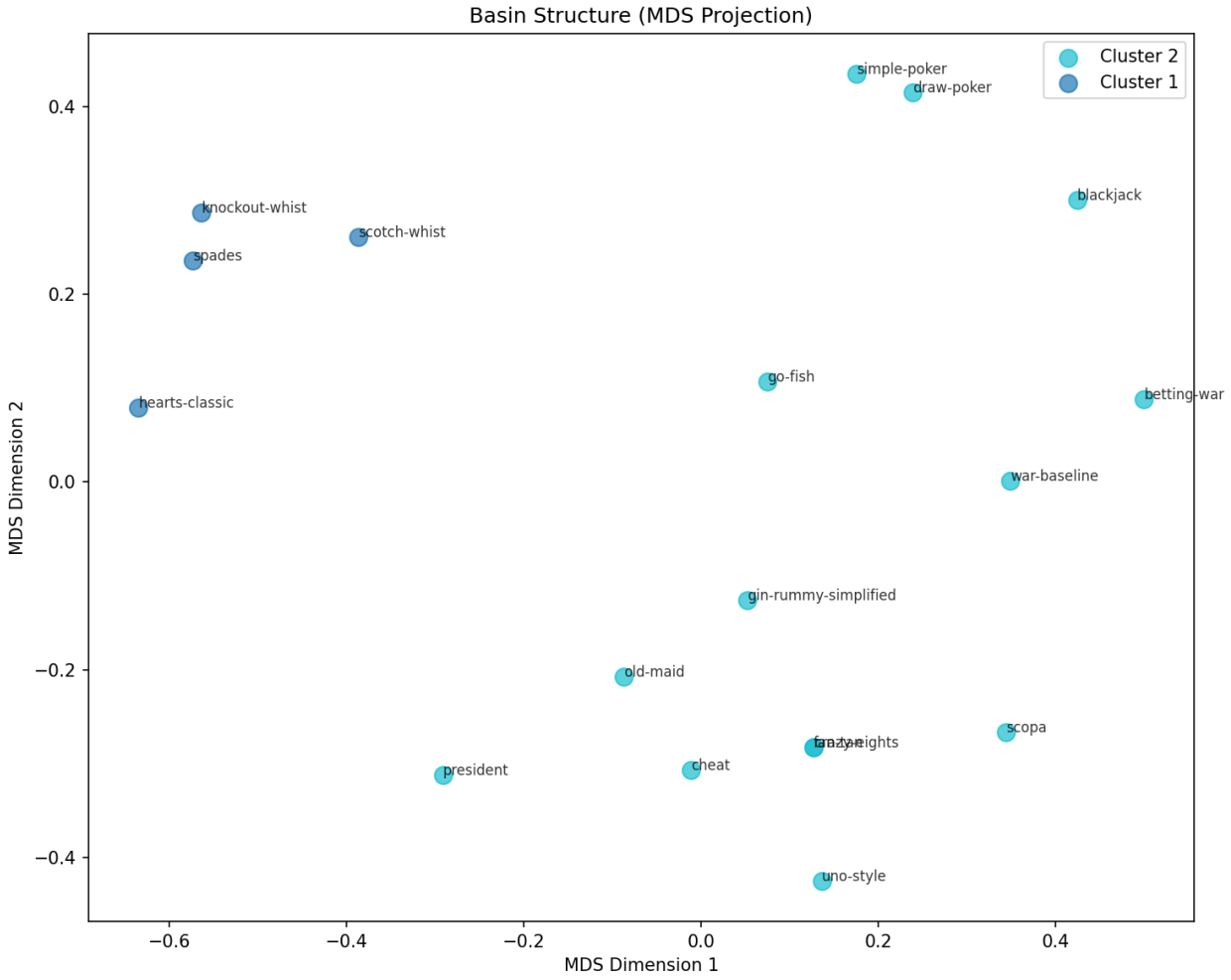

Figure 3: Basin Scatter

Figure 3: MDS projection showing spatial relationships. Trick-taking games cluster on the left; poker variants group on the right; shedding games spread across the center.

---

4. Trajectory Analysis

Figure 4: 250-path trajectories for each known game over 1,000 mutation steps. The consistent downward slope across all games confirms the decay finding. Variance increases with mutation distance.

Trajectory Patterns

1. **Universal decay** — All games show fitness decline, averaging -22% over 1,000 mutations
2. **Poker stability** — Betting games (draw-poker, simple-poker, blackjack) show the flattest trajectories

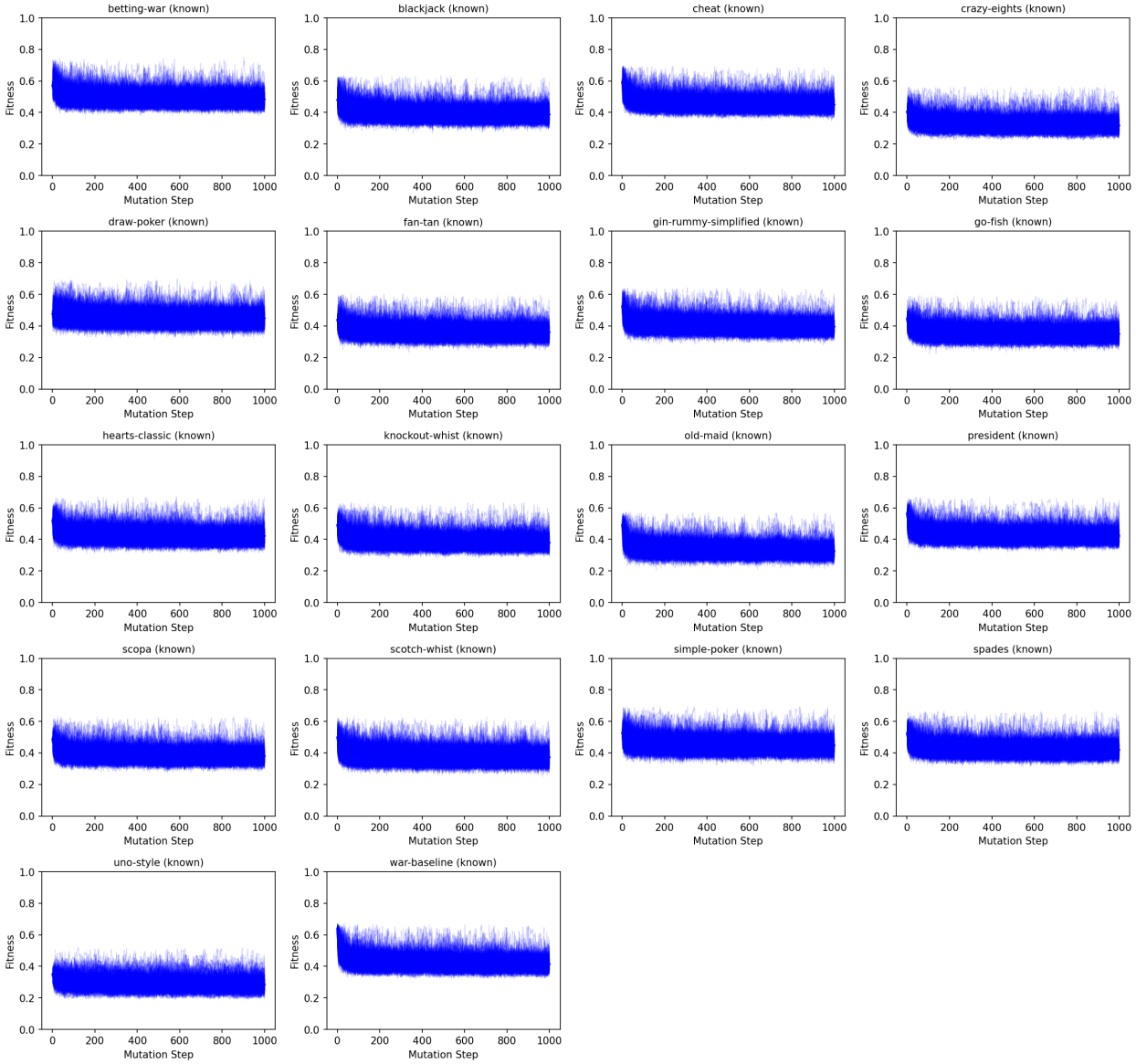

Figure 4: Trajectories

3. **War volatility** — Despite highest start, war-baseline shows steep decline (-35% over 1,000 mutations)
4. **Consistent ranking** — High-fitness games remain higher throughout; the ordering is preserved

---

5. Implications for Evolution Strategy

Revised Recommendations

The discovery of gradual decay changes our strategy recommendations:

Previous Assumption	Updated Finding	New Strategy
Landscape is flat	**Slight decay exists**	Balance exploration with selection pressure
100+ mutations safe	**~10 mutations to 10% drop**	Shorter mutation chains, more frequent selection
Explore freely	**Decay accumulates**	Prefer local search over random walks
Aggressive mutation	**Stability varies by game type**	Adaptive mutation rates by game family

Specific Recommendations

1. **Mutation Rate:** Reduce from current levels; aim for 5-10 mutations between selection events
2. **Selection Frequency:** Increase tournament selection frequency to prevent drift
3. **Elitism:** Preserve top performers without mutation to maintain fitness peaks
4. **Crossover Emphasis:** With decay on mutation walks, crossover between successful genomes becomes more valuable
5. **Game-Type Awareness:** Poker-family mutations can be more aggressive (stable); War-family needs conservative mutation

---

6. Technical Details

Configuration

```
Sampling:
  steps_per_path: 1000      # 10x previous
  paths_per_genome: 250     # 5x previous
  games_per_eval: 50

Baseline:
  random_genomes: 11,500    # 11.5x previous
  require_playable: true
```

```
Total Compute:
  known_paths: 4,500 (18 × 250)
  baseline_paths: 11,500
  total_evaluations: 16,000 × 1,001   16 million
```

Updated Fitness Metrics

This analysis used the improved fitness calculation including:

- **Tension Curve:** Real lead change tracking for War (HandSizeMaxLeaderDetector), trick-taking (TrickLeaderDetector), and score-based games
- **Complexity:** Normalized component scores with power transform for better spread
- **Comeback Potential:** Trailing winner frequency (true comebacks vs win balance)
- **Fallback Handling:** Capped at 0.6 for games without meaningful leader tracking

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

---

7. Conclusions

The Fitness Landscape is a Tilted Plateau

The extended analysis reveals a nuanced picture:

1. **Known games occupy higher ground** — The 26% fitness advantage is real and significant
2. **But it's not a cliff** — The slope is gradual (-0.00003/step), allowing exploration
3. **Nor is it flat** — Previous short-run analysis was misleading; decay is real
4. **Basin radius ~10 mutations** — This defines the "safe exploration zone"

Strategic Implications

Evolution should operate in **guided exploration mode**: - Use known games as starting points (26% advantage) - Explore locally (10-mutation radius before fitness drop) - Select frequently (prevent drift down the slope) - Cross between successful variants (avoid pure random walk decay)

The landscape structure supports the current evolutionary approach but suggests tightening selection pressure to counteract the discovered decay tendency.

Appendix: Raw Data

Full analysis data available in `basin_analysis.json` including: - Complete distance matrix (18×18) - All 4,500 known game trajectories (18 games × 250 paths × 1,001 steps) - All 11,500 random baseline trajectories - Cluster assignments and valley depths

---

Report generated by DarwinDeck Basin Analysis Tool Analysis completed: 2026-01-14T00:11:00Z Fitness metrics version: 2026-01-13 (tension/complexity/comeback updates)