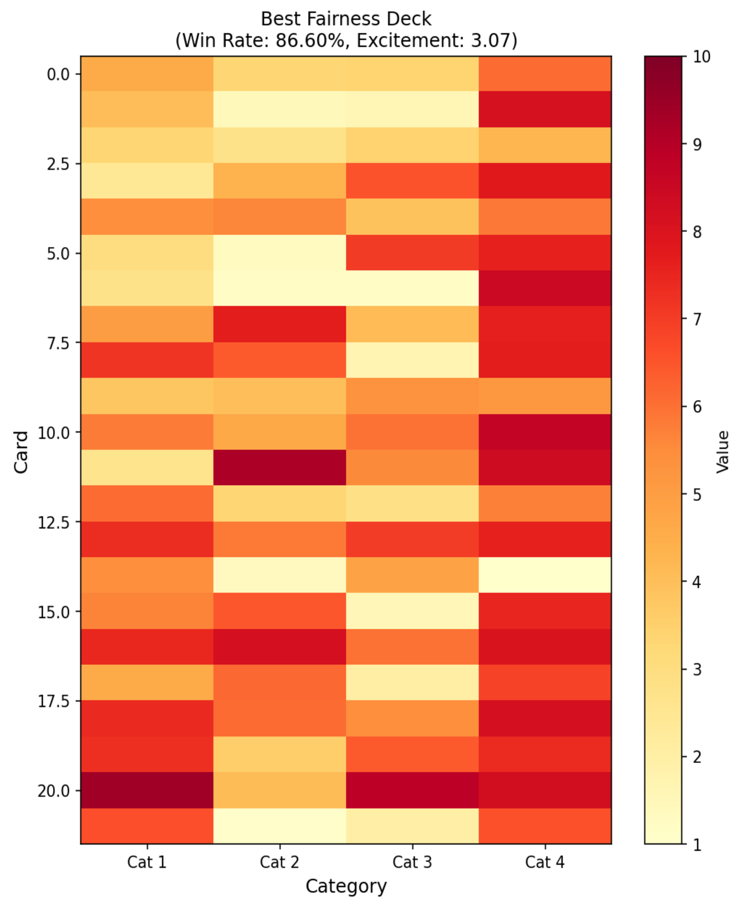
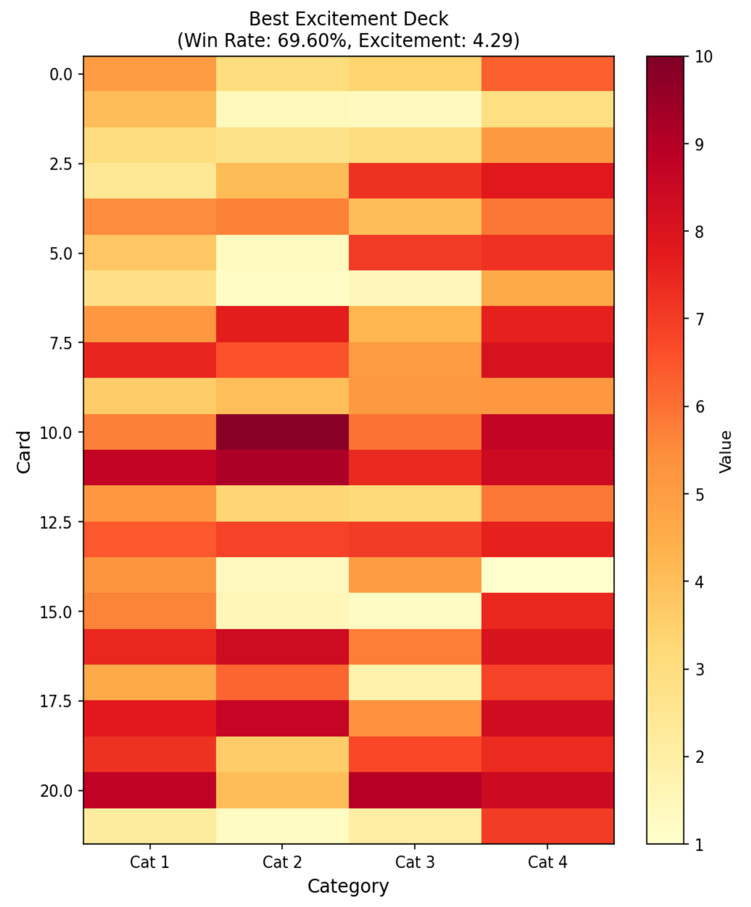
Optimization and Decision Making Case study: Automatic Game Balancing



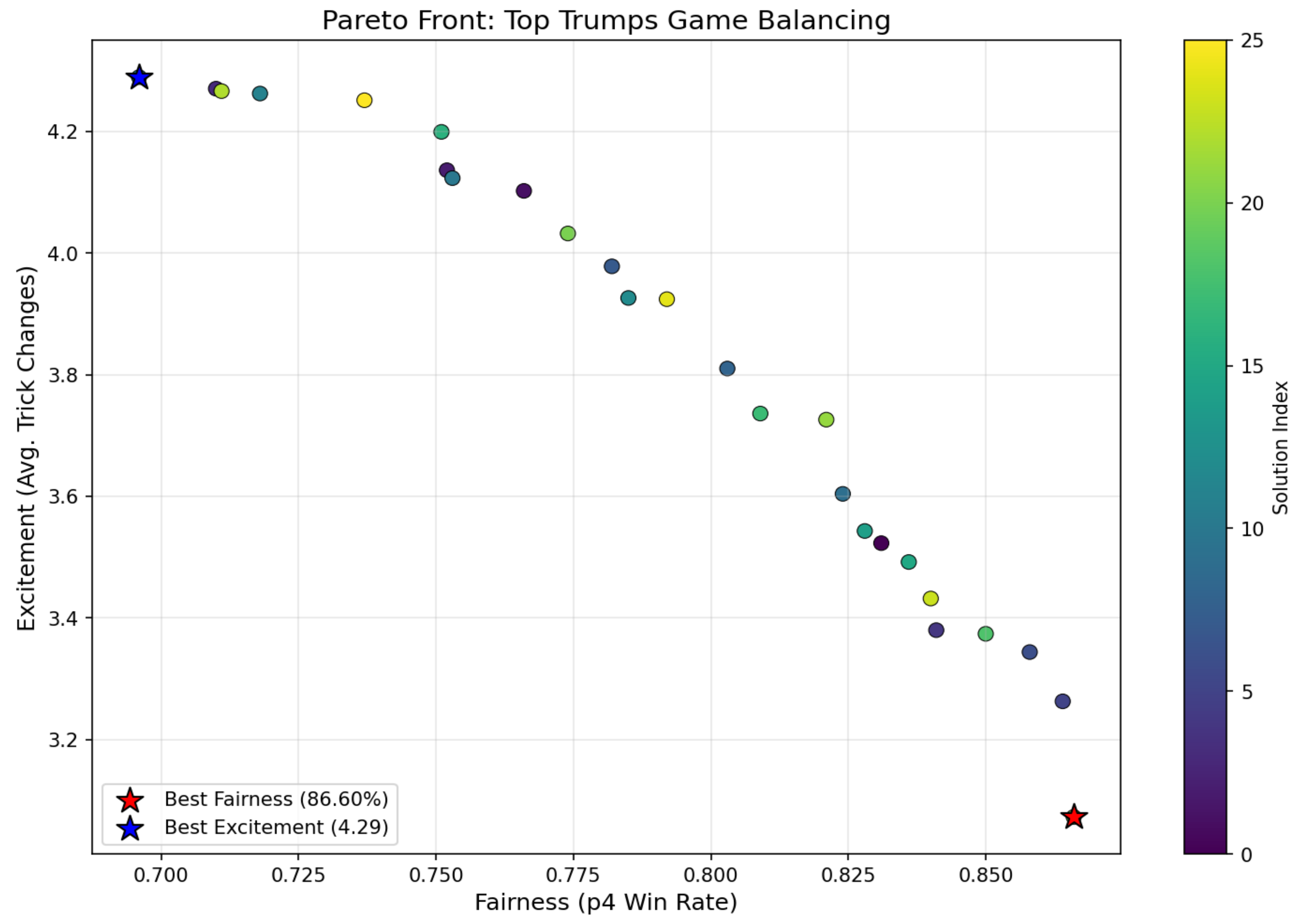


Problem

- **Objective**: Automatically balance a Top Trumps deck by optimizing **fairness** and **excitement**.
- **Problem size**: 22 cards × 4 categories → **88** continuous decision **variables**.
- **Evaluation**: Each candidate deck is assessed using **1,000** stochastic game **simulations**.
- **Challenge**: High-dimensional search space combined with **noisy** objective evaluations.
- **Goal**: Approximate the **Pareto front** capturing trade-offs between fairness and excitement.

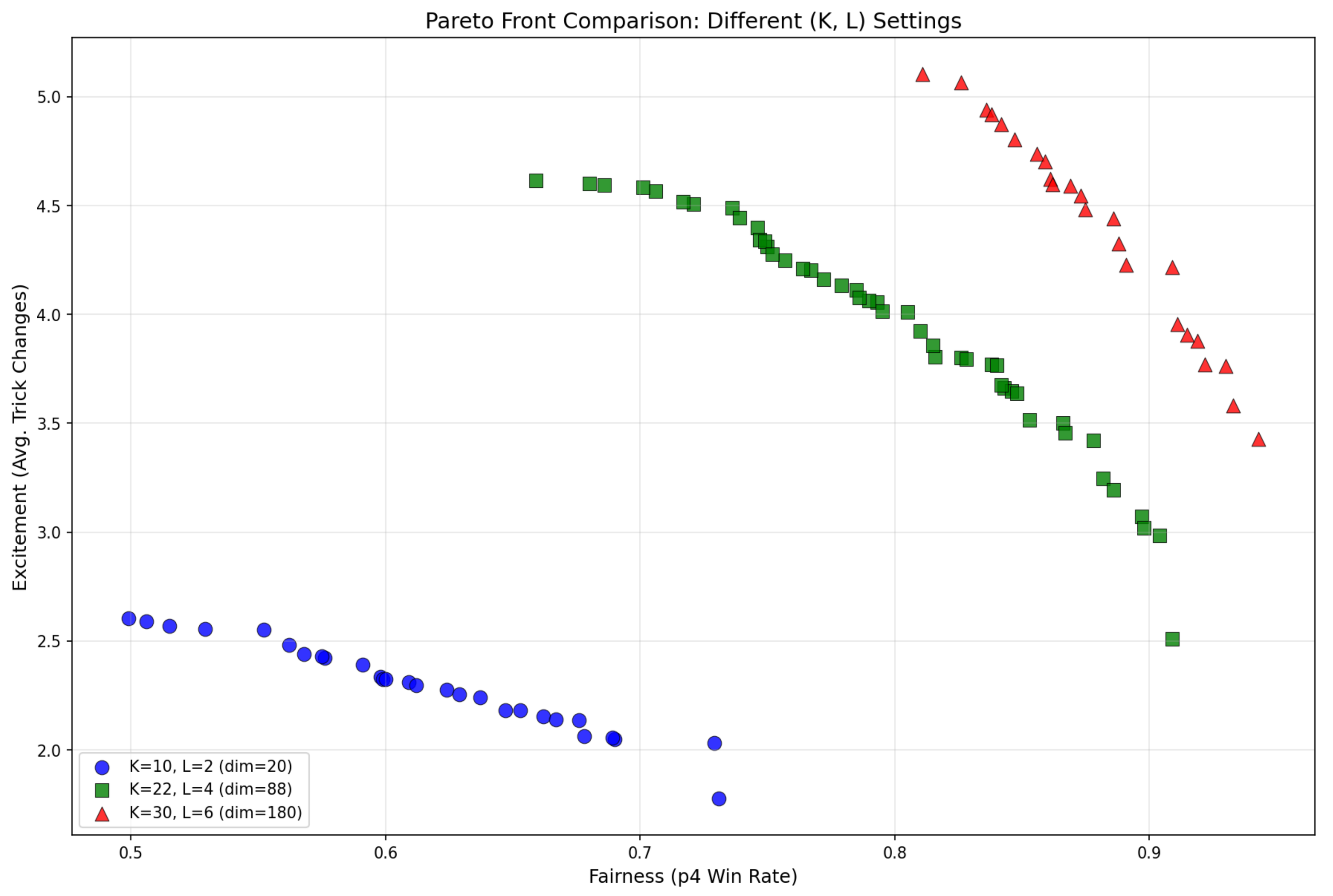
Approach

- **Optimizer**: **NSGA-II** (multi-objective evolutionary algorithm)
- **Population size**: **100** candidate decks per generation
- **Generations**: **100** evolutionary iterations
- **Search budget**: **10,000** total evaluations (100 × 100)
- **Procedure**: Evolve decks via **selection**, **crossover**, and **mutation** to approximate the fairness–excitement Pareto front



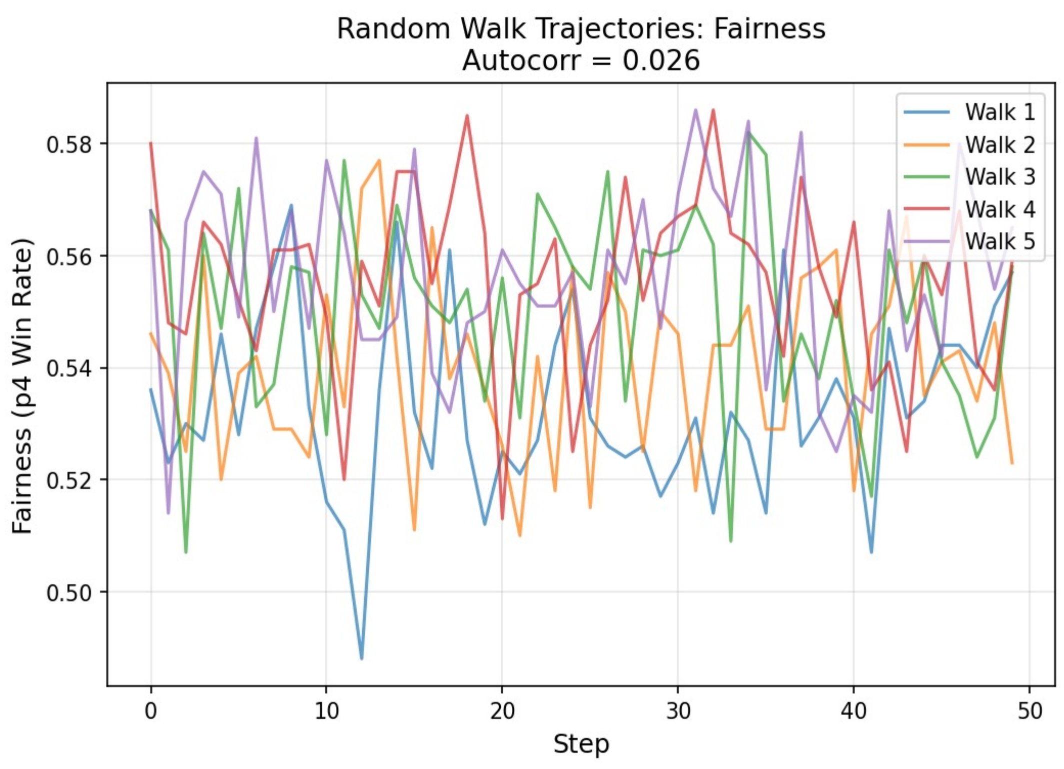
Results

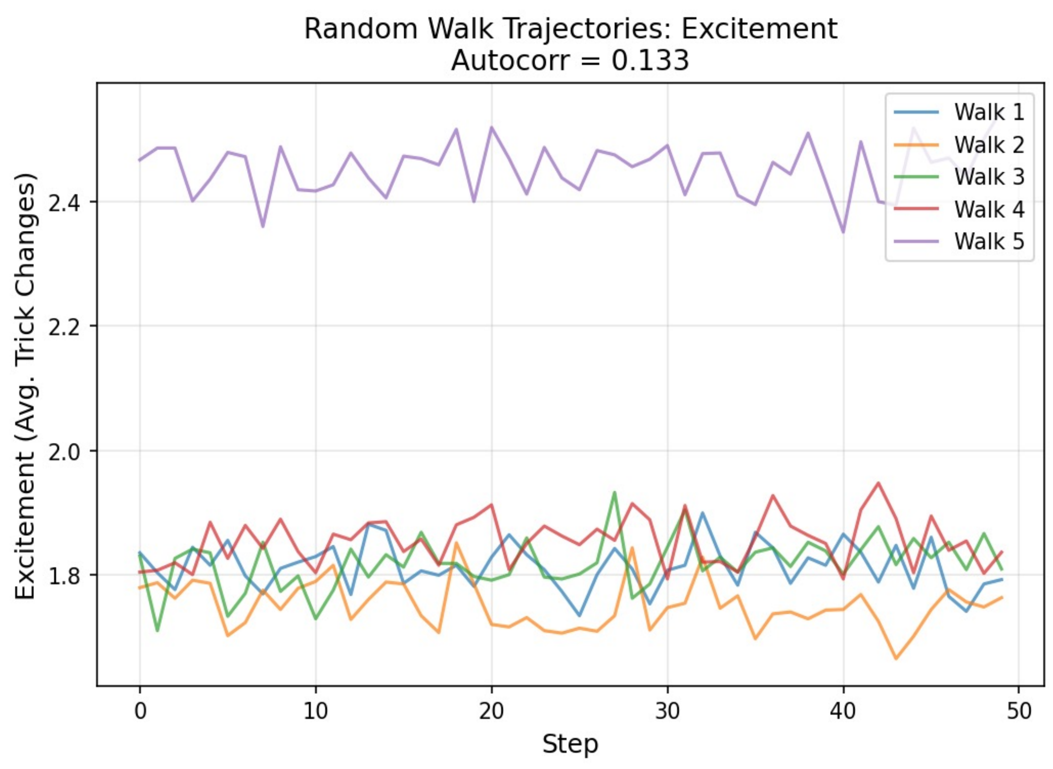
- **Solution Quality:** Hypervolume = 4.170591 (reference point [0.1, 0.1]).
- **Pareto Front:** 26 non-dominated solutions discovered.
- **Objective Ranges:** * **Fairness (p4 win rate):** 0.696 – 0.866
  - **Excitement (trick changes):** 3.07
- **Extreme Trade-offs:**
  - **Best Fairness**: 86.60% win rate, 3.07 excitement.
  - **Best Excitement**: 69.60% win rate, 4.29 excitement.
- **Runtime Performance**: Total runtime of 4.2 minutes (average 0.025s per evaluation).

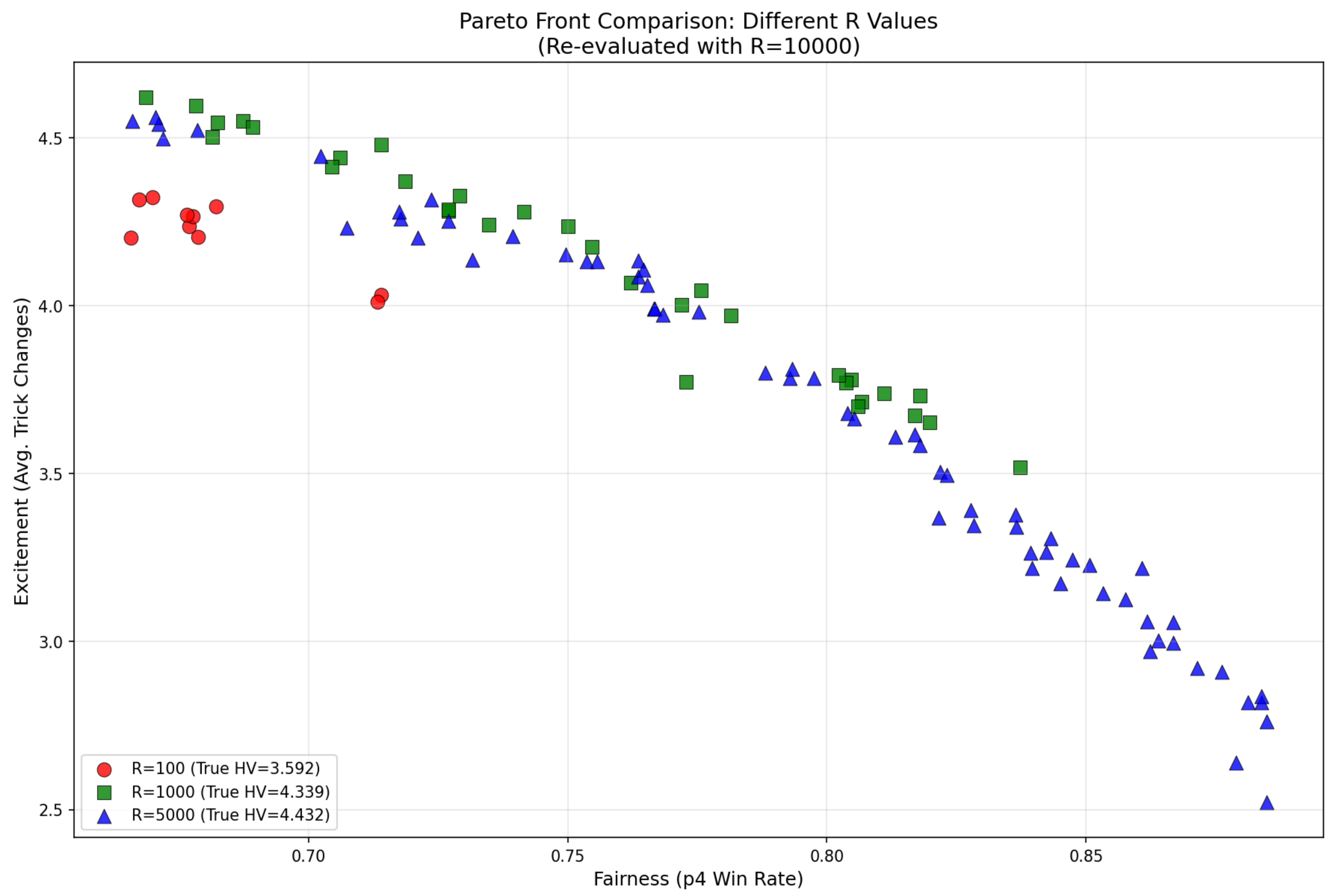


Impact of Card (K) and Category (L) Scaling

- **Search Space Complexity:**
  - Increasing K or L expands the decision variables **linearly**, but the search space grows **exponentially**. A standard 52-card deck with 6 categories would involve **312 variables**, requiring a larger population size (>100) to maintain diversity.
- **Pareto Front Density:**
  - Higher K allows for more **granular "balancing"**. Smaller decks (like your K=22) result in more **distinct "steps"** on the Pareto front, whereas larger decks would produce a **smoother**, more **continuous curve**.







Problem Landscape Properties

- **High-Dimensional Epistasis:**
  - The fitness of a single card is not independent; it depends entirely on the values of the other 21 cards in the deck. This makes the landscape **highly interdependent.**
- **Ruggedness:**
  - Small adjustments to a single category value can cause a "cascade" effect in trick-taking logic, creating a rugged landscape where **local search algorithms would likely fail**, justifying the use of **NSGA-II**.
- **Objective Correlation:**
  - Fairness and Excitement are generally **negatively correlated**. Maximizing excitement (trick changes) often requires "swingy" cards that inherently disrupt perfectly equal win rates.

Influence of Simulation Repetitions (R)

- **The "Noise" Factor:**
  - Because Top Trumps is stochastic (due to the shuffle), R acts as a **signal-to-noise filter**.
- **Solution Reliability:**
  - With low R (e.g., R=100), a deck might appear on the Pareto front simply because it got a **"lucky"** series of **shuffles**.
- **Convergence Trade-off:**
  - Increasing R to 1,000 significantly improves the **robustness** of the results but increases the computational runtime from seconds to minutes.

Niklas Bormann | Tristan Heitger | Yannick Wenger