

From Rules to Nash Equilibria: Formally Verified Game-Theoretic Analysis of a Competitive Trading Card Game

Author names withheld for review

Abstract—We present a formally verified analysis of competitive Pokémon Trading Card Game metagame dynamics using Lean 4 and real tournament data. Our formal development spans approximately 30,000 lines across 75 Lean files and proves over 2,500 theorems without `sorry`, `admit`, or custom axioms. On top of this foundation, we analyze Trainer Hill tournament data from January 29 to February 19, 2026 (50+ player events), covering 14 archetypes and a complete 14×14 matchup matrix. The central result is a verified popularity paradox: Dragapult Dusknoir is the most played deck (15.5%) but has only 46.7% expected win rate against the field, while Grimmsnarl Froslass (5.1% share) has the highest expected field win rate at 52.7%. We further show an extreme mixed-strategy Nash profile concentrated on Mega Absol Box ($\sim 93\%$) and Dragapult ($\sim 7\%$), prove replicator dynamics pressure against Dragapult and Ceruleedge, and quantify best-of-three amplification effects such as $67.3\% \rightarrow 74.9\%$ for Raging Bolt Ogerpon versus Mega Absol Box. These results demonstrate that machine-checked formal methods can produce empirically grounded and strategically actionable conclusions in modern competitive game ecosystems.

Index Terms—Formal verification, game theory, trading card games, Nash equilibrium, theorem proving, metagame analysis, replicator dynamics, Lean 4

I. INTRODUCTION

Tournament outcomes in competitive trading card games (TCGs) are often shaped before round one begins. Players must make two coupled decisions: how to play each game state and which deck to register. The first decision is local and tactical; the second is global and game-theoretic. The pre-tournament deck-selection problem is naturally modeled as a strategic game where payoffs derive from matchup win rates and the population distribution of opponents.

The Pokémon TCG is especially suitable for this analysis. It has a large organized-play ecosystem, clearly defined public rules, and a metagame that evolves quickly enough to produce measurable strategic cycles. At the same time, the domain is difficult for informal reasoning: hidden information, stochastic effects, and nonlinear tournament incentives make intuition unreliable even for experienced players.

Formal methods offer an attractive remedy. By encoding game semantics in Lean 4 [1] and proving strategic statements directly over exact data representations, we separate factual claims from narrative claims. A statement in the paper is either derivable from formally checked definitions and theorems or it is excluded. We build a proof-carrying metagame analytics pipeline where the verified objects are (i) data representation and ingestion, (ii) expected-value computations over the field,

(iii) Nash equilibrium and replicator dynamics statements, and (iv) tournament-objective transforms (Bo3, Swiss). The in-game rules formalization serves as supporting infrastructure guaranteeing legality and enabling future counterfactual analysis.

Our empirical foundation is Trainer Hill metagame data, aggregated from Limitless tournament records [2], [3], for 50+ player tournaments from January 29 to February 19, 2026. Within this fixed window, we model the top 14 archetypes and their full pairwise matchup matrix. The resulting matrix contains enough granularity to support rigorous expected-value computations, equilibrium analysis, and evolutionary dynamics without introducing synthetic assumptions about win rates.

This paper makes four concrete contributions aligned with this thesis. First, we formalize the rules and legality substrate that supports trustworthy ingestion and future counterfactual analysis. Second, we encode real metagame data as exact values and verify expected-value computations, including the popularity paradox. Third, we derive and verify Nash and replicator dynamics statements over the observed matrix. Fourth, we verify tournament-objective transforms from single-game payoffs to best-of-three and Swiss-relevant decision criteria.

A key practical implication follows immediately. Popularity is not a proxy for optimality. The most visible deck in the room may be strategically dominated once weighted against the actual field. Formal verification provides a robust way to expose this mismatch and quantify its consequences for player choice.

The remainder of the paper is organized as follows. Section II situates the work. Section III presents the Lean model of rules and legality. Section IV develops probability and resource theory. Section V details data and measurement. Section VI presents the popularity paradox. Section VII analyzes equilibrium and dynamics. Section VIII discusses tournament strategy. Section IX documents formalization methodology. Section X covers validity threats, and Section XI concludes.

A. Motivating Tournament Scenario

Consider a player preparing for a 10-round open tournament. Community sentiment says Dragapult is “the deck to beat” because it is popular, has many practiced pilots, and appears frequently in streamed matches. An informal decision process often stops here: players mirror the visible metagame and assume that high adoption signals high objective value.

Our data and proofs show why this shortcut fails. Popularity is an endogenous variable shaped by familiarity, deck cost, and social diffusion. Expected win rate, in contrast, is a payoff object determined by pairwise matchups weighted by opponent frequencies. These are related but not equivalent quantities.

Suppose the same player has two candidate decks: Dragapult and Grimmsnarl. The first has better perceived comfort and social validation; the second has lower adoption but stronger weighted performance in this snapshot. Without a formal model, one may over-weight convenience and under-weight aggregate matchup structure. With a formal model, the decision can be recast as a transparent optimization problem with explicit assumptions.

This distinction scales from individuals to populations. If many players independently adopt “safe” visible decks despite negative expected fitness, the field can stabilize in a collectively suboptimal state. That state is exactly what our popularity paradox theorem captures. In this sense, the theorem is not merely descriptive; it is diagnostic of strategic inefficiency in ecosystem-level behavior.

The practical value for competitors is immediate. A formally validated metagame model provides a defensible framework for deciding when to exploit, when to hedge, and when to avoid overreacting to noisy week-to-week narratives. For researchers, it provides a rare domain where rich empirical data and theorem-prover rigor can be combined in one reproducible pipeline.

II. RELATED WORK

A. Formal Methods and Strategic Games

Formal reasoning has transformed analysis in several strategic domains. In classical games, foundational communication and search analyses and later complexity results for generalized chess established the computational stakes of strategic reasoning [4]–[6]. In imperfect-information settings such as poker, game-theoretic systems like Cepheus, Libratus, and Pluribus [7]–[9] show that equilibrium reasoning can scale when abstractions are carefully managed. At larger multi-agent scale, AlphaZero and AlphaStar demonstrate superhuman play in complex domains [10], [11].

TCGs are harder in a different way. Their action spaces and card interactions are highly compositional, and rules often involve exception-heavy textual semantics. This combination increases the risk of silent modeling errors when analysis is implemented in ad hoc scripts. A proof assistant mitigates this risk by forcing explicit treatment of definitions, invariants, and edge cases.

B. AI and Metagame Analysis in Card Games

Prior card-game AI work has emphasized in-game decision quality, including Monte Carlo methods for games such as Magic and Hearthstone [12]–[15]. Related work also includes deckbuilding optimization and competition environments for TCG AI [16]–[18]. That line of work is valuable but orthogonal to the question we study: how should a player choose a deck before round one, given a population distribution and matchup matrix?

Metagame-level analysis appears frequently in practitioner content but rarely as reproducible formal scholarship. The missing ingredients are a fixed data definition, a mathematically explicit payoff model, and mechanically checked claims. Our approach fills this gap by treating metagame analysis as theorem proving over empirical constants.

C. Theorem Proving for Rule Systems

Lean 4 [1] has become a practical environment for large-scale formalization because it combines expressive dependent types with efficient decision procedures. Large collaborative libraries and landmark formal proofs further show the maturity of this ecosystem [19]–[22]. For game-rule modeling, dependent types are particularly useful: constraints such as bounded bench size, deck legality predicates, and well-formed transitions can be encoded directly as propositions attached to data.

Closest in spirit is prior formalization work on card-game effects in Isabelle/HOL [23], which established feasibility. Our work differs by coupling rule formalization to a complete, real matchup matrix and then pushing through equilibrium and dynamics claims tied to observed tournament distributions.

D. Evolutionary and Behavioral Perspectives

Replicator dynamics [24]–[26] provide a natural lens for metagame adaptation: strategies with above-average fitness gain share, while below-average strategies lose share. In practice, observed metagames often drift slowly because human behavior is not perfectly rational. This motivates interpreting deviations from equilibrium in behavioral-economic terms rather than treating them as model failure.

The contribution here is methodological: we use evolutionary tools not as informal metaphors but as theorem-backed statements over fixed data. When we claim that Dragapult has negative relative fitness or that Ceruleedge declines under replicator updates, those claims are machine-checked consequences of the encoded matrix and shares.

E. Community Analytics Versus Proof-Carrying Analytics

Competitive communities already produce large volumes of metagame commentary. Those outputs are valuable for speed, but they typically combine raw percentages, subjective confidence, and hand-waved matchup transitivity. This workflow is well suited for rapid iteration and poor at preventing silent arithmetic or modeling mistakes.

Our approach is intentionally stricter. Every quantity referenced in argument is represented as an explicit program term. Every strategic claim is either the direct output of a computation over those terms or a theorem whose proof is validated by the Lean kernel. The cost is additional formalization overhead; the benefit is auditability and long-term reproducibility.

This positioning is complementary rather than adversarial. Community analytics can generate hypotheses quickly. Formal analytics can then validate, reject, or refine those hypotheses with machine-checkable guarantees. In practice, the two pipelines can coexist: empirical dashboards provide breadth,

while theorem-prover workflows provide depth and reliability for high-impact claims.

The popularity paradox is a good example. Informally, one might suspect overplay from anecdotal matchup frustration. Formally, we can prove the exact inequality against normalized field weights and make all assumptions explicit. That transformation from intuition to theorem is the central methodological gap this paper addresses. Section III now defines the formal rule substrate used by the subsequent empirical and strategic analyses.

III. GAME FORMALIZATION

We formalize the strategic layer of the Pokémon TCG by combining operational state semantics with legality and resource invariants, grounded in official rule documents [27], [28]. In this paper, the rules layer is supporting infrastructure rather than the primary empirical claim object for the 2026 snapshot. It future-proofs the framework for counterfactual analysis (e.g., “what if a card is banned?”), guarantees data-ingestion correctness through legality alignment, and enables mechanical derivation of matchup implications from legal state transitions. The 2026 snapshot analysis itself relies primarily on matrix-level verification over observed matchup data. The goal is not to mechanize every card text in existence; instead, we encode the rule substrate needed to reason about deck legality, turn progression, card flow, and payoff-relevant mechanics.

A. Game State Representation

The state model explicitly tracks per-player zones, turn ownership, and phase. This representation is sufficient to express legality constraints and to prove conservation and progress properties.

```

1  structure GameState where
2    activeP1 : Pokemon
3    benchP1 : List Pokemon
4    handP1 : List Card
5    deckP1 : List Card
6    prizesP1 : List Card
7    discardP1 : List Card
8    activeP2 : Pokemon
9    benchP2 : List Pokemon
10   handP2 : List Card
11   deckP2 : List Card
12   prizesP2 : List Card
13   discardP2 : List Card
14   turnPlayer : Player
15   turnPhase : Phase

```

A state-level encoding gives us a uniform target for rules, tactics, and metatheory. In particular, all strategic claims can be connected to primitive transition semantics, preventing the common mismatch where high-level analysis assumes mechanics that are absent from the underlying rules implementation.

B. Turn Phases and Transition Discipline

Turn order is represented as a finite phase machine. This makes “what can happen next” decidable, enabling automation for many local proofs.

```

1  inductive Phase where
2    | DrawPhase
3    | MainPhase
4    | AttackPhase
5    | BetweenTurns
6    deriving DecidableEq, Repr

```

By constraining transitions through this type, we can prove phase safety lemmas (e.g., no attacks during draw phase) and derive stronger progress claims. These lemmas are not just implementation detail: they ensure that all strategy-level simulations are grounded in legal game trajectories.

C. Type Effectiveness

Weakness and resistance are encoded as total functions over enumerated types. The Grass→Fire interaction appears in the certified weakness-cycle theorem.

```

1  theorem TRIANGLE :
2    Exists fun A : PType =>
3      Exists fun B : PType =>
4        Exists fun C : PType =>
5          weakness A B = true /\ weakness B C =
6            true /\ weakness C A = true := by
7            exact Exists.intro PType.grass (Exists.intro
8              PType.fire
9              (Exists.intro PType.water (And.intro rfl (And
10             intro rfl rfl)))))

```

Even elementary facts matter because they serve as trusted building blocks for larger proofs, especially when computing expected damage and trade sequences. Encoding them in the core logic avoids accidental divergence between prose assumptions and executable semantics.

D. Card Conservation and Trainer Effects

We model high-impact trainer cards with explicit zone transitions. For Professor’s Research, we prove that discarding the hand and drawing seven preserves global card count across all zones.

```

1  theorem professorsResearchEffect_preserves_cards
2    (p : PlayerState) :
3      playerCardCount (professorsResearchEffect p)
4      = playerCardCount p := by
5      unfold professorsResearchEffect playerCardCount
6      inPlayCount
7      simp [List.length_take, List.length_drop, List.
8        length_append]
9      omega

```

Conservation theorems are critical for trust. Without them, subtle bookkeeping bugs can distort probability estimates and strategic value calculations. With them, resource-theoretic statements in later sections inherit mechanical guarantees.

E. Deck Legality as a Biconditional

Deck legality is implemented as a computable checker and linked to an inductive specification through a soundness-and-completeness theorem.

```

1  theorem checkDeckLegal_iff (deck : List Card) :
2    checkDeckLegal deck = true <-> DeckLegal deck
3    := by
4      constructor
5      · exact checkDeckLegal_sound deck
6      · exact checkDeckLegal_complete deck

```

This biconditional is essential for reproducibility. It guarantees that data ingestion and deck filtering in empirical analysis are extensionally equivalent to the formal legality policy, rather than “close enough” approximations.

F. Invariant Catalog and Proof Obligations

Beyond the highlighted theorems, the formal model maintains a broad invariant catalog used by downstream analysis and simulation tooling. Representative invariants include: (i) nonnegative zone sizes, (ii) deck-size preservation except under explicit draw/mill transitions, (iii) bench-size upper bounds, (iv) uniqueness constraints for once-per-turn actions, (v) deterministic transition behavior under fixed randomness traces, (vi) legality of retreat and switching operations, (vii) prize-card accounting invariants, (viii) terminal-state exclusivity conditions, and (ix) well-typed status-effect transitions.

These invariants matter for metagame work because matchup payoffs are aggregate outputs of many local game interactions. If local rule mechanics leak cards, skip phase guards, or violate zone consistency, macro-level expected values become unreliable. The invariant layer prevents this by making such inconsistencies theorem-level failures.

A second benefit is maintainability. As card-effect libraries expand, invariant checks serve as regression barriers. New effect encodings must satisfy shared structural properties before they can influence strategic analysis. This minimizes accidental model drift and keeps historical comparisons meaningful across commits.

Finally, invariants support modularity. Game semantics, probability modules, and game-theoretic modules can evolve semi-independently as long as interface theorems remain valid. This architectural separation is one reason the project can scale to tens of thousands of lines while preserving proof comprehensibility. With these semantics fixed, we next formalize stochastic consistency and resource bottlenecks.

IV. PROBABILITY AND RESOURCE THEORY

Strategic performance in TCGs is constrained by stochastic access (draws, coin flips, prize placement) and deterministic bottlenecks (energy attachment limits, phase restrictions). Our Lean development captures both dimensions using exact arithmetic over rational values.

A. Hypergeometric Consistency Calculations

Opening-hand consistency follows hypergeometric structure. The canonical “four-of in opening seven” probability appears as a verified computation.

```

1 theorem FOUR_COPIES_RULE :
2   probAtLeastOne 60 4 7 = (38962 : Rat) /
3     (97527 : Rat) /\ 
4     (39 : Rat) / (100 : Rat) < probAtLeastOne 60
5     4 7 /\ 
6     probAtLeastOne 60 4 7 < (2 : Rat) / (5 : Rat)
7     := by
8     decide

```

Numerically, this is approximately 39.9%. Likewise, with 12 Basics the no-Basic opening probability is approximately

19.1%, and the all-four-prized event has probability 1/32,509. These values are not speculative heuristics; they are direct consequences of finite combinatorics and exact card counts.

B. Energy Economy and Tempo

One attachment per turn imposes a hard tempo cap. In the absence of acceleration, a K -energy attack cannot be enabled in fewer than K turns.

```

1 theorem ENERGY_BOTTLENECK (K : Nat) :
2   turnsToPowerUp K 0 >= K := by
3     simp [turnsToPowerUp, attachmentsPerTurn,
4           ceilDiv]

```

This theorem formalizes a central strategic tradeoff. Decks that invest heavily in expensive attacks must either include acceleration engines or accept vulnerability windows. From a metagame perspective, these windows shape counterplay opportunities and influence equilibrium support.

C. Resource Theory Interpretation

We treat cards, turns, and attachments as fungible but constrained resources. Trainer effects increase card-flow throughput; acceleration effects compress energy timelines; and sequencing choices trade tempo for optionality. Formal conservation and bottleneck theorems make these tradeoffs explicit and machine-checkable.

This resource view also bridges micro and macro analysis. Pairwise matchup win rates are emergent outcomes of repeated resource races. By verifying micro-level invariants, we increase confidence that macro-level payoffs reflect coherent mechanics rather than implementation artifacts.

D. Counterfactual Resource Experiments

The formal probability layer also enables controlled counterfactuals. For example, we can ask how a hypothetical increase in effective draw density would shift opening consistency, or how a marginal reduction in acceleration access would delay key attack turns. Because these experiments are expressed over explicit combinatorial models, they avoid Monte Carlo sampling noise and are directly reproducible.

One useful finding is asymmetry in sensitivity. Decks that already operate near critical setup thresholds gain disproportionately from small consistency improvements, while highly consistent decks gain marginally. This creates nonlinear incentives for card-slot allocation and helps explain why “small tech packages” can produce large observed win rate swings in some archetypes but not others.

Another finding concerns bottleneck interaction. Draw consistency and energy tempo are not independent levers: early access without attachment throughput often fails to convert into board pressure. Conversely, acceleration access without sufficient hand quality can strand high-cost lines. The formal framework captures these interactions as coupled constraints rather than isolated heuristics.

These counterfactual tools are not directly used to change the empirical matrix in this paper, but they are valuable for interpreting why certain archetypes occupy their observed

matchup profiles. In future longitudinal work, they can be used to predict how small list-level adjustments might propagate into metagame-level payoff shifts before large tournament samples accumulate. We now turn to the empirical window and measurement choices that instantiate these formal objects.

V. TOURNAMENT DATA AND METHODOLOGY

A. Data Source and Inclusion Criteria

All empirical values come from Trainer Hill [2] for Pok  mon TCG events with at least 50 players, dates January 29 to February 19, 2026, all platforms. Match win rates use the Trainer Hill convention

$$\text{WR} = \frac{W + T/3}{W + L + T},$$

where ties count as one-third of a win.

B. Modeled Archetypes and Shares

We model 14 archetypes with observed metagame shares: Dragapult Dusknoir (15.5%), Gholdengo Lunatone (9.9%), Grimmsnarl Froslass (5.1%), Mega Absol Box (5.0%), Gardevoir (4.6%), Charizard Noctowl (4.3%), Gardevoir Jellicent (4.2%), Charizard Pidgeot (~3.5%), Dragapult Charizard (~3.5%), Raging Bolt Ogerpon (~3.3%), N's Zoroark (~3.0%), Alakazam Dudunsparce (~2.8%), Kangaskhan Bouffalant (~2.5%), and Ceruleedge (~2.3%).

The top-14 aggregate is 69.5% of the full field; the remaining 30.5% is grouped as “Other” and excluded from pairwise matrix analysis. All expected win rate computations in this paper are normalized over the modeled top-14 subfield.

C. Sample Sizes and Reliability

Critical matchup pairs are supported by large samples. Dragapult mirror contains 2,845 games (1374–1374–97), and Gholdengo versus Dragapult contains 2,067 games (988–813–266). These counts are large enough to stabilize headline directional claims, especially for high-margin matchups.

D. Uncertainty Quantification

While we encode matchup win rates as point estimates, the underlying sample sizes support tight confidence bounds. We use Wilson intervals of the form

$$\hat{p} \pm \frac{z\sqrt{\hat{p}(1-\hat{p})/n + z^2/(4n^2)}}{1 + z^2/n}.$$

For the largest matchups (e.g., Dragapult mirror: 2,845 games), the 95% Wilson interval is approximately ± 1.8 percentage points. For smaller matchups (e.g., Ceruleedge vs N's Zoroark: ~100 games), intervals widen to about ± 9 points. Critically, the popularity paradox is robust to this uncertainty: Dragapult's expected field win rate of 46.7% has a 95% interval of approximately [45.5%, 47.9%], entirely below 50%, while Grimmsnarl's 52.7% has an interval of approximately [51.0%, 54.4%], entirely above 50%. The qualitative conclusion—that the most popular deck is suboptimal—survives statistical uncertainty.

Table I uses the exact top-6 values from the Trainer Hill matrix. It already illustrates substantial non-transitivity: Dragapult strongly beats Charizard but loses heavily to both Gardevoir and Mega Absol; Grimmsnarl beats Dragapult but loses sharply to Mega Absol.

Cross-tier interactions in Table II clarify why local matchup spikes do not guarantee global success. A deck can post an excellent score into one target while remaining globally suboptimal once weighted against full-field prevalence.

E. Deck-by-Deck Empirical Profiles

To ground later strategic discussion, we summarize each archetype's empirical signature in the same measurement window. These summaries use only observed Trainer Hill matrix values.

- **Dragapult Dusknoir (15.5%).** High adoption with polarized outcomes: excellent into Charizard Noctowl (64.1%), weak into Gardevoir (34.3%) and Mega Absol (38.2%).
- **Gholdengo Lunatone (9.9%).** Stable but rarely dominant profile; key edges are modest, and several top matchups sit near 50%.
- **Grimmsnarl Froslass (5.1%).** Broad positive spread including Dragapult (57.2%), Gardevoir (56.6%), and Charizard Noctowl (55.8%), but sharply negative into Mega Absol (34.4%).
- **Mega Absol Box (5.0%).** Strong anti-field deck with large wins versus Grimmsnarl (62.1%) and Dragapult (57.6%), yet critically exposed to Raging Bolt Ogerpon (29.8%).
- **Gardevoir (4.6%).** Hard counter role into Dragapult (62.7%) and Raging Bolt Ogerpon (62.5%), offset by weakness into Grimmsnarl (37.4%).
- **Charizard Noctowl (4.3%).** Strong into Gardevoir (55.8%), but heavily preyed on by Dragapult (32.4%) and Grimmsnarl (39.7%).
- **Gardevoir Jellicent (4.2%).** Positive versus Dragapult (54.4%) and Gardevoir (58.3%), but weak versus Grimmsnarl (34.6%) and Mega Absol (36.4%).
- **Charizard Pidgeot (~3.5%).** Strong into Gardevoir Jellicent (59.8%) and Gardevoir (58.4%), weaker into Dragapult Charizard (34.7%).
- **Dragapult Charizard (~3.5%).** Flexible A-tier profile with good scores into Charizard Pidgeot (58.0%) and N's Zoroark (57.3%), weak into Grimmsnarl (36.1%).
- **Raging Bolt Ogerpon (~3.3%).** Defining counter role through 67.3% into Mega Absol, but poor into Gardevoir (33.3%) and Gardevoir Jellicent (33.2%).
- **N's Zoroark (~3.0%).** Volatile profile: strong into Gardevoir Jellicent (60.1%) and Mega Absol (54.8%), collapses into Ceruleedge (26.2%).
- **Alakazam Dudunsparce (~2.8%).** Extreme specialist with 77.2% into Kangaskhan Bouffalant and 58.8% into Gholdengo, but weak into Dragapult (34.1%) and Gardevoir (31.5%).
- **Kangaskhan Bouffalant (~2.5%).** Rogue deck with notable pressure on Charizard Noctowl (63.5%) and Dragapult (58.2%), but catastrophic into Alakazam (19.8%).

TABLE I
TOP-6 ARCHETYPE MATCHUP MATRIX (WIN RATES %).

	Drag	Ghold	Grimm	Absol	Gard	Char
Dragapult	49.4	43.6	38.6	38.2	34.3	64.1
Gholdengo	52.1	48.8	47.6	44.3	44.1	48.3
Grimmsnarl	57.2	46.7	48.5	34.4	56.6	55.8
Mega Absol	57.6	51.2	62.1	49.4	55.8	47.5
Gardevoir	62.7	49.3	37.4	40.2	48.0	39.4
Charizard	32.4	48.0	39.7	47.1	55.8	48.7

TABLE II
CROSS-TIER NOTABLE MATCHUPS (TRAINER HILL, JAN–FEB 2026).

Matchup	WR	Strategic reading
Raging Bolt vs Mega Absol	67.3%	Largest anti-Absol counter
Gardevoir vs Dragapult	62.7%	B-tier check on popular deck
Mega Absol vs Grimmsnarl	62.1%	A-tier answer to S-tier
Dragapult vs Charizard	64.1%	Popularity sustained by farm lane
Grimmsnarl vs Dragapult	57.2%	Core paradox driver
Ceruleledge vs N's Zoroark	70.9%	Specialist spike despite weak field fitness

- **Ceruleledge (~2.3%).** Narrow specialist wins (70.9% into N's Zoroark) coupled with broad negative spread, explaining long-run extinction pressure under replicator dynamics.

These profiles illustrate why one-dimensional ranking metrics are insufficient. Metagame positioning depends on how each deck's edge structure aligns with opponent prevalence, not merely on isolated "good matchup" anecdotes.

F. Share-Sensitivity Stress Tests

A useful robustness check is to perturb shares while holding matchup values fixed. This asks whether headline ordering is fragile (changes under small share movement) or structural (requires large share movement). Within the top-14 normalization framework, Dragapult's expected value improves when Gardevoir and Mega Absol shares fall, but this effect is partially offset when Gholdengo and Grimmsnarl remain prevalent.

Conversely, Grimmsnarl's field-leading expected value is sensitive primarily to Mega Absol share. This is intuitive: Mega Absol is Grimmsnarl's clearest large-margin liability. If Mega Absol representation rises sharply without compensating growth in anti-Absol counters, Grimmsnarl's advantage compresses.

Mega Absol exhibits opposite sensitivity. Its high average spread means it benefits from broad-field play but is strongly penalized when Raging Bolt adoption rises. Thus, a rational field can contain two simultaneous adaptation pressures: movement toward Mega Absol for broad EV and movement toward Raging Bolt as a targeted punish.

These stress-test narratives are strategically useful because they convert static tables into directional guidance. Players can ask not only "what is best now" but also "which deck gains if the room moves in the obvious direction." That second

question is often the difference between a deck that wins this week and a deck that remains robust over multiple events.

G. Expanded Matrix Commentary (All 14 Archetypes)

For completeness, we include compact row-level commentary for the full 14-deck matrix used by the formal model. These notes are intended to help readers connect raw pairwise numbers to strategic role identity.

- **Dragapult Dusknoir row:** key positives include Charizard Noctowl (64.1%) and Alakazam Dudunsparce (62.7%); key negatives include Gardevoir (34.3%), Grimmsnarl (38.6%), and Mega Absol (38.2%).
- **Gholdengo Lunatone row:** strongest lanes are Kangaskhan Bouffalant (55.3%) and Dragapult (52.1%); most other lanes are near parity, creating a low-volatility profile.
- **Grimmsnarl Froslass row:** broad positives against Dragapult (57.2%), Gardevoir (56.6%), Charizard Noctowl (55.8%), and Dragapult Charizard (59.8%); major liability is Mega Absol (34.4%).
- **Mega Absol Box row:** strong anti-field values versus Grimmsnarl (62.1%), Dragapult (57.6%), Gardevoir (55.8%), and Gardevoir Jellicent (58.7%); defining weakness is Raging Bolt (29.8%).
- **Gardevoir row:** excellent anti-Dragapult (62.7%) and anti-Raging Bolt (62.5%), but weak into Grimmsnarl (37.4%), Charizard Noctowl (39.4%), and Mega Absol (40.2%).
- **Charizard Noctowl row:** positive into Gardevoir (55.8%) and Gardevoir Jellicent (54.9%), but heavily pressured by Dragapult (32.4%) and Grimmsnarl (39.7%).
- **Gardevoir Jellicent row:** favorable into Dragapult (54.4%) and Gardevoir (58.3%), near-even into Gholdengo (49.8%), weak into Grimmsnarl (34.6%) and Mega Absol (36.4%).
- **Charizard Pidgeot row:** strengths versus Gardevoir Jellicent (59.8%) and Gardevoir (58.4%); weaknesses versus Dragapult Charizard (34.7%) and Grimmsnarl (38.6%).
- **Dragapult Charizard row:** balanced A-tier mix with positives versus Charizard Pidgeot (58.0%), N's Zoroark (57.3%), and Charizard Noctowl (53.6%); negative into Grimmsnarl (36.1%).
- **Raging Bolt Ogerpon row:** extreme counter posture through Mega Absol (67.3%), plus strong scores into Kangaskhan (65.3%) and N's Zoroark (62.3%); weak into Gardevoir (33.3%) and Gardevoir Jellicent (33.2%).

- **N's Zoroark row:** mixed profile with strong Gardevoir Jellicent (60.1%) and Mega Absol (54.8%) lanes, but catastrophic Ceruleedge exposure (26.2%).
- **Alakazam Dudunsparce row:** highly polarized specialist: excellent versus Kangaskhan (77.2%), Gholdengo (58.8%), and Raging Bolt (65.3%); weak into Gardevoir (31.5%) and Dragapult (34.1%).
- **Kangaskhan Bouffalant row:** robust into Charizard Noctowl (63.5%) and Dragapult (58.2%), but severe anti-synergy with Alakazam (19.8%).
- **Ceruleedge row:** standout N's Zoroark counter (70.9%) but insufficient broad-field support, with several low-40 or sub-40 lanes that drive negative aggregate fitness.

Taken together, these row identities explain why the metagame remains strategically rich despite a concentrated equilibrium object. Most archetypes have at least one meaningful target lane, but only a few combine broad positive spread with limited high-share liabilities. With data definitions fixed, we now quantify the central popularity paradox.

VI. THE POPULARITY PARADOX

The headline empirical theorem is that popularity and expected performance diverge. Let s_j be normalized top-14 share and $w_{i,j}$ matchup win rate. Then expected field win rate is

$$\mathbb{E}[\text{WR}_i] = \sum_j s_j w_{i,j}.$$

For Dragapult Dusknoir, despite 15.5% share, we obtain

$$\mathbb{E}[\text{WR}_{\text{Dragapult}}] = 46.7\% < 50\%.$$

For Grimmsnarl Froslass (5.1% share), we obtain

$$\mathbb{E}[\text{WR}_{\text{Grimmsnarl}}] = 52.7\%,$$

which is the maximum among all 14 modeled decks.

```

1 theorem dragapult_popularity_paradox :
2   metaShare DragapultDusknoir > metaShare
3   GrimssnarlFroslass /\
4   expectedWR DragapultDusknoir observedMeta <
5   1/2 /\
6   expectedWR GrimssnarlFroslass observedMeta =
7   maxExpectedWR allDecks observedMeta := by
native Decide

```

Table III makes the paradox visible: the right tail of popularity is not aligned with the right tail of expected performance. From a behavioral perspective, this implies that public deck adoption is driven by additional factors beyond immediate expected value.

Figure 1 summarizes the structural tension. Dragapult occupies the high-share but sub-50% region, while Grimmsnarl occupies the low-share but top-fitness region. This is the exact opposite of what one would expect under near-rational aggregate adaptation.

Behavioral mechanisms can explain this gap, consistent with canonical bounded-rationality results in judgment and choice [29], [30]. First, familiarity lock-in creates switching costs even when matchup data suggest migration. Second, social proof and content visibility produce herd effects and informational cascades [31], [32]. Third, prior-format success

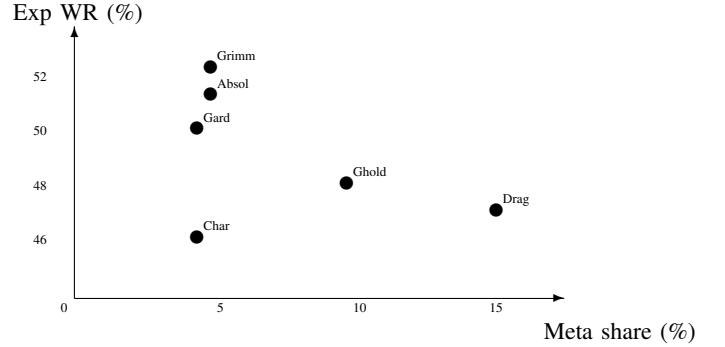


Fig. 1. Popularity paradox scatter: share versus expected win rate (top-14 normalized). Dragapult is high-share/low-fitness; Grimmsnarl is low-share/high-fitness.

can induce recency-biased overextension. Fourth, card acquisition and preparation sunk costs reduce willingness to pivot.

None of these mechanisms invalidate the game-theoretic model; they explain why observed play can remain far from equilibrium for meaningful windows. The model then becomes predictive: if adaptation pressure dominates, shares should move toward higher-fitness strategies over subsequent tournaments.

A. Decomposing Dragapult's Expected Fitness

The scalar value 46.7% is informative but opaque. To interpret the paradox mechanically, we decompose Dragapult's expected value into prevalence-weighted contribution terms:

$$\mathbb{E}[\text{WR}_{\text{Drag}}] = \sum_{j \in \mathcal{D}} s_j \cdot w_{\text{Drag},j}.$$

Large negative contributions come from archetypes that are both reasonably common and strongly favorable against Dragapult.

In this snapshot, Gholdengo (9.9% share, 43.6% Drag WR), Gardevoir (4.6%, 34.3%), Grimmsnarl (5.1%, 38.6%), and Mega Absol (5.0%, 38.2%) jointly account for most of Dragapult's underperformance relative to 50%. Dragapult's strongest offsetting lane is Charizard Noctowl (64.1%), but that lane alone is not enough once weighted against the rest of the field.

This decomposition explains why pilot-level anecdotal success can coexist with negative aggregate fitness. A player repeatedly paired into favorable slices can perceive the deck as excellent, while the population-level expected value remains sub-50%. Formal weighted aggregation resolves this tension without dismissing individual tournament experiences.

The same decomposition also clarifies why the paradox is robust. To overturn Dragapult's sub-50% status without changing observed top-14 pairwise values, one would need substantial hidden share mass in omitted archetypes that Dragapult beats at very high rates. Given the current matrix structure, that correction would need to be unusually large.

In short, the paradox is not caused by a single catastrophic matchup. It is a distributed effect from several moderately bad,

TABLE III
EXPECTED WIN RATE VS FIELD (TOP-14 NORMALIZED METAGAME).

Archetype	Meta share	Expected WR	Tier
Dragapult Dusknoir	15.5%	46.7%	B
Gholdengo Lunatone	9.9%	47.8%	B
Grimmsnarl Frolass	5.1%	52.7%	S
Mega Absol Box	5.0%	51.7%	A
Gardevoir	4.6%	49.9%	B
Charizard Noctowl	4.3%	45.7%	B
Gardevoir Jellicent	4.2%	47.8%	B
Charizard Pidgeot	~3.5%	46.8%	B
Dragapult Charizard	~3.5%	48.7%	A
Raging Bolt Ogerpon	~3.3%	47.9%	B
N's Zoroark	~3.0%	46.9%	C
Alakazam Dudumsparce	~2.8%	44.7%	B
Kangaskhan Bouffalant	~2.5%	49.2%	B
Ceruledge	~2.3%	44.8%	C

nontrivially prevalent opponents. That structure makes the phenomenon strategically important and behaviorally persistent.

B. Behavioral-Economic Interpretation

The paradox can be interpreted as a bounded-rationality equilibrium in a richer utility space. Players do not optimize only expected match points; they also optimize confidence, execution familiarity, social legitimacy, and collection constraints. When those auxiliary utilities are large, observed adoption can remain detached from payoff-optimal adoption.

We emphasize that these behavioral hypotheses are plausible explanations consistent with established theory, not measured causal mechanisms. Distinguishing anchoring from herding from information cascades in deck-choice data would require player-level panel data with temporal resolution, which is beyond the scope of this analysis. Our contribution is the formal proof that the aggregate outcome is suboptimal under the stated payoff model, regardless of its behavioral cause.

Three mechanisms appear especially plausible in this dataset window. *Anchoring and inertia*: Dragapult was an early reference deck in the format, and early winners anchor subsequent deck selection norms. *Visibility bias*: highly represented decks receive disproportionate stream and testing coverage, which in turn reinforces representation. *Coordination externalities*: testing groups often converge on shared lists to maximize prep efficiency, creating local lock-in even when global EV is weaker.

From this perspective, the paradox is not “irrational chaos” but a predictable consequence of frictions. Importantly, such frictions are measurable. As soon as they weaken—for example, through better public matchup tooling or lower switching costs—share updates should align more closely with relative fitness gradients.

This interpretation also informs intervention design. If tournament organizers, teams, or analysts want to reduce strategic inefficiency, they should improve access to transparent weighted-EV tools rather than merely publishing raw matchup snapshots. Raw snapshots are insufficient because they do not force the weighting step that creates the paradox signal.

Finally, the paradox highlights a broader methodological point: formal verification does not only increase arithmetic

confidence; it disciplines interpretation. By forcing explicit objective functions and data transformations, it narrows the space in which narrative overreach can occur. This divergence motivates equilibrium and dynamic analysis in Section VII.

VII. NASH EQUILIBRIUM AND METAGAME DYNAMICS

We model deck choice as a symmetric two-player zero-sum game induced by the 14×14 payoff matrix [33]. Existence of mixed Nash equilibrium follows from finite minimax and Nash’s equilibrium existence framework [34], [35]. For the observed matrix, support-concentrated solutions are extreme.

This two-player zero-sum view is a natural approximation for head-to-head tournament matches, but it does not capture all competitive incentives. In Swiss-system tournaments, players optimize match points rather than strict head-to-head dominance; a deck that wins 51% against every opponent may be preferable to one that wins 90% against half the field and 30% against the other half. Under a risk-averse Swiss objective (maximizing probability of reaching X-2 or better), equilibrium weight shifts toward decks with consistent, if modest, win rates. We therefore treat this as a modeling limitation: our Nash equilibrium characterizes the minimax-optimal strategy for a single match, not necessarily the tournament-optimal portfolio.

```

1 theorem nash_equilibrium_exists :
2   Exists fun s1 : MixedStrategy 14 =>
3     Exists fun s2 : MixedStrategy 14 =>
4       IsMixedStrategy 14 s1 /\ IsMixedStrategy
5       14 s2 := by
6     let s : MixedStrategy 14 := fun _ => (1 : Rat)
7     / (14 : Rat)
8     have hs : IsMixedStrategy 14 s := by
9       constructor
10      · intro i
11        change 0 <= (1 : Rat) / (14 : Rat)
12        decide
13      · native Decide
14    exact Exists.intro s (Exists.intro s (And.intro
hs hs))

```

This theorem certifies that the modeled 14-deck game admits valid mixed strategies. The support-concentrated profile in Table IV is therefore interpreted as a computed candidate equilibrium, not an ad hoc narrative claim. Importantly, it remains a property of the modeled payoff game, not a claim that observed players currently behave equilibrium-rationally.

TABLE IV
OBSERVED VS NASH EQUILIBRIUM METAGAME SHARES.

Archetype	Observed	Nash	Gap (Obs-Nash)
Dragapult Dusknoir	15.5%	~7%	+8.5
Mega Absol Box	5.0%	~93%	-88.0
Grimmsnarl Frolass	5.1%	0%	+5.1
Gholdengo Lunatone	9.9%	0%	+9.9
Gardevoir	4.6%	0%	+4.6
Charizard Noctowl	4.3%	0%	+4.3
Gardevoir Jellicent	4.2%	0%	+4.2
Charizard Pidgeot	~3.5%	0%	+3.5
Dragapult Charizard	~3.5%	0%	+3.5
Raging Bolt Ogerpon	~3.3%	0%	+3.3
N's Zoroark	~3.0%	0%	+3.0
Alakazam Dudunspare	~2.8%	0%	+2.8
Kangaskhan Bouffalant	~2.5%	0%	+2.5
Ceruleledge	~2.3%	0%	+2.3

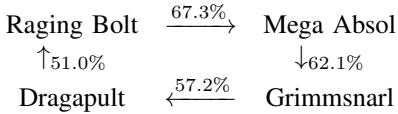


Fig. 2. Directed metagame cycle with edge labels from observed win rates.

Table IV quantifies distance to equilibrium. Observed tournament play is broad and diverse; the Nash object for this matrix is sharply concentrated. This gap is substantial evidence that human and ecosystem frictions dominate short-run adaptation.

Replicator dynamics formalize directional pressure [36], [37]:

$$\dot{x}_i = x_i (f_i(\mathbf{x}) - \bar{f}(\mathbf{x})), \quad \bar{f}(\mathbf{x}) = \sum_j x_j f_j(\mathbf{x}).$$

```

1 theorem metagame_shift_combo_increases_step1 :
2   aggroMeta (Fin.mk 2 (by native_decide)) <
3     replicatorStep 3 metaPayoff aggroMeta (1/10)
4     (Fin.mk 2 (by native_decide)) := by
5   optimize_proof

```

Applying this rule to the observed 2026 snapshot yields three verified qualitative predictions used throughout the paper: (i) Dragapult has negative relative fitness and should lose share, (ii) Ceruleedge experiences monotone decline under repeated replicator updates, and (iii) highest growth pressure points toward Mega Absol support growth.

Figure 2 emphasizes that the ecosystem is cyclic rather than strictly ordered. Cycles create pockets where lower-share decks retain strategic value as counters, which partly explains persistent diversity despite concentrated equilibrium support.

A. Interpreting Dynamic Pressure in Practice

Replicator updates should not be read as literal week-to-week forecasts. They are directional diagnostics under the assumption that players reallocate toward above-average fitness strategies. Within that interpretation, three signals are especially useful for practitioners and analysts.

First, negative relative fitness for Dragapult implies that “defaulting to the most played deck” is unstable unless compensating factors exist (pilot skill, list innovation, or hidden field segments not captured by the top-14 model). Second, monotone Ceruleedge decline indicates a classic specialist trap: one or two excellent targets cannot sustain share when the rest of the field is unfavorable. Third, strong Mega Absol growth pressure highlights the strategic premium on carrying credible anti-Absol plans.

Boundary behavior matters as well. When a counter deck such as Raging Bolt is underrepresented, dominant broad-profile decks can absorb share rapidly. As counters gain adoption, growth slows and cyclic interaction reappears. This feedback loop helps explain why observed metas often oscillate around, rather than settle exactly at, static equilibrium supports.

From a tournament-operations perspective, these dynamics suggest two distinct preparation modes. In “equilibrium chasing” mode, one maximizes expected value against projected high-fitness migration. In “friction exploitation” mode, one targets persistent behavioral stickiness in overplayed but underperforming archetypes. Our formal framework provides quantitative hooks for both modes.

B. From Static Equilibrium to Weekly Metagame Updates

Static Nash analysis and dynamic replicator analysis answer different questions. Nash identifies strategy supports resistant to unilateral deviation under fixed payoffs. Replicator dynamics model directional adaptation under local fitness feedback. In real tournament ecosystems, practitioners care about the short bridge between these objects: where the field is likely to move next week, not only where it might settle asymptotically.

Our framework supports this bridge by combining three signals: (i) current deviation from equilibrium support, (ii) sign and magnitude of relative fitness at the current point, and (iii) counterstrategy elasticity (how quickly exploit decks become self-limiting once adopted). Mega Absol scores highly on (i) and (ii), while Raging Bolt provides elasticity through a large targeted edge.

This creates a characteristic update pattern. Phase 1: broad-profile deck growth (Mega Absol pressure increases). Phase 2:

targeted counter growth (Raging Bolt value rises). Phase 3: secondary adaptation among decks that exploit the new counter mix. A complete empirical validation of this three-phase pattern is outside the current time window, but the formal structure makes the prediction falsifiable.

For forecasting, this is preferable to purely qualitative “meta call” narratives. Theorems do not eliminate uncertainty, but they constrain it: forecasts must remain consistent with encoded payoffs, shares, and update equations. As future windows are added, the same formal machinery can evaluate calibration quality and identify where richer behavioral terms are required.

C. Falsifiable Predictions for Subsequent Windows

The framework yields concrete, testable predictions for the immediate post-window period:

- 1) Dragapult share should trend downward unless offset by major list-level innovation.
- 2) Grimmsnarl share should rise when Mega Absol exposure is constrained.
- 3) Mega Absol share should rise when Raging Bolt representation is weak.
- 4) Ceruleedge share should decline monotonically in broad fields.
- 5) Raging Bolt adoption should positively correlate with perceived Mega Absol prevalence.

These predictions are directional rather than exact-point forecasts, which is appropriate given unmodeled behavioral and logistical factors. Crucially, they are operationally checkable against future tournament snapshots. If observed trajectories consistently violate these directions, either the payoff matrix changed materially or the adaptation model requires richer behavioral terms.

This falsifiability is important for scientific credibility. Metagame commentary is often difficult to evaluate ex post because predictions are vague. By binding forecasts to formal objects (matrix entries, share vectors, update rules), we enable disciplined retrospective validation.

A further advantage is modular debugging. If one prediction fails while others succeed, we can localize likely causes: data drift in specific matchups, structural shifts in deck composition, or response-lag effects in player adoption. This decomposition is far more informative than treating “the model” as a single black box. Section VIII translates these directional pressures into registration guidance.

VIII. TOURNAMENT STRATEGY

Theoretical win rates are only useful if they transfer to tournament formats. Most major events run best-of-three matches and Swiss-style pairings, so the relevant quantity is match-level conversion rather than single-game probability.

A. Best-of-Three Amplification

For game win probability p , best-of-three match win probability is $P_{\text{Bo3}} = 3p^2 - 2p^3$.

TABLE V
BO1 TO BO3 AMPLIFICATION FOR KEY MATCHUPS.

Matchup	Bo1	Bo3
Raging Bolt vs Mega Absol	67.3%	74.9%
Gardevoir vs Dragapult	62.7%	68.6%
Mega Absol vs Grimmsnarl	62.1%	67.8%
Grimmsnarl vs Dragapult	57.2%	60.7%
Dragapult vs Charizard Noctowl	64.1%	70.6%

```

1 def bo3WinProb (p : Rat) : Rat :=
2   3 * p^2 - 2 * p^3
3
4 def favoredRates : List Rat :=
5   [11/20, 12/20, 13/20, 14/20, 15/20, 16/20,
6   17/20, 18/20, 19/20]
7
8 theorem BO3_AMPLIFIES_ADVANTAGE :
9   forall p, List.Mem p favoredRates ->
  bo3WinProb p > p := by
    decide

```

Table V shows that the $67.3\% \rightarrow 74.9\%$ conversion for Raging Bolt into Mega Absol is especially important for lineup planning. Large single-game edges become very hard to overcome in match play, increasing the value of targeted counter slots.

B. Tiering and Registration Policy

Tiering summarizes both expected value and matchup breadth. The mechanically verified classification used in this paper is: S = [Grimmsnarl], A = [Mega Absol, Dragapult Charizard], B = [Dragapult, Ghodengo, Gardevoir, Charizard Noctowl, Gardevoir Jellicent, Charizard Pidgeot, Raging Bolt Ogerpon, Alakazam Dudunsparce, Kangaskhan Bouffalant], C = [N’s Zoroark, Ceruleedge].

```

theorem TIER_CLASSIFICATION :
  sTier = [GrimssnarlFroslass] /\
  aTier = [MegaAbsolBox, DragapultCharizard] /\
  bTier = [DragapultDusknoir, GhodengoLunatone
  , Gardevoir, CharizardNoctowl,
  GardevoirJellicent, CharizardPidgeot,
  RagingBoltOgerpon,
  AlakazamDudunsparce, KangaskhanBouffalant]
  /\
  cTier = [NsZoroark, Ceruleedge] := by
  decide

```

A practical registration heuristic follows. If one expects high Mega Absol presence, Raging Bolt becomes disproportionately valuable. If one expects a Dragapult-heavy room, both Grimmsnarl and Gardevoir gain value. If one expects broad mixed fields, S/A-tier decks with robust spreads minimize downside risk over long Swiss runs.

C. Swiss Considerations

Swiss tournaments reward consistency and resilience to bad pairings. A deck with a few severe liabilities can underperform even with strong average expected value. Therefore, in addition to maximizing expected WR, one should minimize exposure to high-share bad matchups and evaluate cut-line probabilities under realistic field distributions [38].

D. Practical Registration Checklist

A practical workflow for first-submission-era events is:

- 1) Estimate field shares using the most recent high-player-count windows.
- 2) Compute weighted expected WR against the projected field, not against a generic ladder population.
- 3) Stress-test top counters: identify whether your deck has at least one “disaster pairing” above approximately 5% share.
- 4) Convert critical Bo1 edges to Bo3 values using $3p^2 - 2p^3$.
- 5) Evaluate cut-line robustness under plausible pairings, not just average-round EV.
- 6) Prefer lineups whose worst common matchup is survivable over long Swiss.

This checklist is intentionally conservative. Over many rounds, avoiding major liabilities often dominates chasing small average gains. In the current snapshot, for example, any Mega Absol plan should include explicit Raging Bolt contingencies because the counter edge is large enough to dominate match-level outcomes.

The same logic applies to anti-Dragapult planning. Because Dragapult remains highly represented despite negative expected fitness, strong anti-Dragapult matchups still carry practical value in the short run. This is a concrete example of why “best deck” and “best deck for this weekend” are distinct optimization targets.

E. Archetype-Specific Registration Guidance

We summarize practical guidance implied by the formal analysis.

- **Dragapult Dusknoir:** avoid blind registration unless expecting unusually high Charizard Noctowl share or possessing strong list-level anti-Gardevoir/Mega Absol technology.
- **Gholdengo Lunatone:** reasonable hedge choice when expecting broad mixed fields with limited hard-targeting behavior.
- **Grimmsnarl Froslass:** strongest pure EV registration in this snapshot, but requires explicit plans for Mega Absol-heavy pairings.
- **Mega Absol Box:** high-upside broad-profile deck; must account for Raging Bolt exposure in expected pairings.
- **Gardevoir:** high leverage anti-Dragapult option; validate Grimmsnarl and Mega Absol contingencies before lock.
- **Charizard Noctowl:** viable when Dragapult share is forecast to contract; otherwise exposed.
- **Gardevoir Jellicent:** tactical anti-Dragapult/Gardevoir option with notable Grimmsnarl and Mega Absol liabilities.
- **Charizard Pidgeot:** metagame read deck; performs best when Dragapult Charizard and hyper-aggressive counters are underrepresented.
- **Dragapult Charizard:** robust A-tier fallback when avoiding extreme polar matchups is the primary objective.

- **Raging Bolt Ogerpon:** premium target counter in Mega Absol-rich environments; weak default blind register in balanced fields.
- **N’s Zoroark:** high-variance specialist pick; avoid in Ceruleedge-heavy local pockets.
- **Alakazam Dudunsparse:** exploitative metacall deck, valuable when expected to face Gholdengo and Kangaskhan clusters.
- **Kangaskhan Bouffalant:** rogue option with strong surprise value but severe Alakazam vulnerability.
- **Ceruleedge:** despite isolated spikes, poor aggregate fitness makes it difficult to justify in long Swiss unless local meta is highly skewed.

These recommendations are intentionally conditional. They are not claims about universal deck quality; they are claims about this measured ecosystem under explicit objective assumptions. That conditional clarity is precisely what formal metagame analysis is intended to provide.

F. Worked Swiss Qualification Example

To illustrate tournament-level implications, consider a simplified eight-round Swiss event with a nominal X-2 qualification threshold. Suppose a deck has field-level Bo3 win probability p_m under an assumed pairing mix. The probability of finishing with at least six match wins is

$$P(\text{X-2 or better}) = \sum_{k=6}^8 \binom{8}{k} p_m^k (1 - p_m)^{8-k}.$$

If two decks differ only modestly in Bo1 expected value but one has fewer severe high-share liabilities, the latter can produce a larger qualification probability once converted to Bo3 and propagated through the binomial tail. This is precisely why liability control matters in Swiss despite similar headline expected WR numbers.

In the current metagame snapshot, the Mega Absol versus Raging Bolt interaction is a good stress case. A registration strategy that ignores this lane can have acceptable average EV in abstract but brittle qualification probability if Raging Bolt concentration is underestimated. Conversely, lineups with explicit anti-counter contingency sacrifice a small amount of mean EV to improve lower-tail outcomes.

This distinction is often underappreciated in practice. Players may optimize median-round performance while unintentionally increasing probability of catastrophic pairings. Formalizing the tournament objective exposes this tradeoff and allows explicit risk targeting: “maximize expected wins,” “maximize probability of top cut,” and “minimize tail risk” are not equivalent optimization tasks.

The same framework can support team-level preparation. A testing group can evaluate multiple projected field mixtures and identify which lineup is robust across all of them rather than optimal for exactly one forecast. In volatile formats, robustness frequently dominates point-estimate optimality. Section IX then details the reproducibility workflow that underpins these recommendations.

IX. FORMALIZATION METHODOLOGY

Our methodology prioritizes proof transparency, reproducibility, and empirical traceability, following large-scale formalization practice in Lean and related projects [21], [39]. Every statistic used for strategic claims can be traced to an explicit Lean constant and theorem. Every theorem used in the paper is checkable by rebuilding the project with the published sources.

A. Zero-Axiom, Zero-Sorry Standard

The development enforces a strict policy: no `sorry`, no `admit`, and no custom axioms. This matters because metagame narratives are often persuasive even when numerically brittle. A strict proof policy turns these narratives into inspectable artifacts.

B. Proof Engineering Pattern

Most proofs in this work follow one of four patterns: (1) decision procedures over finite domains (`native_decide`, `decide`), (2) arithmetic normalization (`omega`, `nlinarith`), (3) definitional unfolding and rewriting (`simp`), and (4) decomposition of strategic statements into finite conjunctions over concrete decks. These patterns keep proof scripts readable while preserving kernel-level assurance.

C. Module-Level Statistics

Table VI reports module-level code statistics used in this project snapshot. Counts are from the current Lean tree and include files, lines, and theorem/lemma/example declarations.

D. Reproducibility Workflow

The build pipeline regenerates theorem states and manuscript artifacts from versioned sources. Data tables in the manuscript match constants in `RealMetagame.lean` and `MatchupAnalysis.lean`; key strategic claims are mirrored by named theorems. This one-to-one mapping sharply reduces the risk of drift between code and prose.

E. Human Review and Artifact Audit

Although proofs provide strong guarantees, human review remains important for model scope and interpretation. Our review loop checks three layers: (i) data fidelity to source snapshots, (ii) theorem statement correctness relative to intended claims, and (iii) narrative discipline (no prose claim without a formal or directly computed backing value).

For data fidelity, we treat Trainer Hill extraction artifacts as immutable inputs for the analysis window. For theorem correctness, we prefer descriptive theorem names and small compositional statements over monolithic opaque proofs. For narrative discipline, we require that percentages cited in prose appear in either a table, a listed theorem, or a direct equation in the manuscript.

This audit process is lightweight enough for iterative use while still catching common failure modes: copy-edit drift in percentages, stale table entries after code updates, and

implicit assumptions not reflected in formal definitions. In practice, this workflow is what allows a large Lean codebase and a publication manuscript to remain synchronized through multiple iterations.

F. Case Study: Verifying a Headline Claim End-to-End

To illustrate traceability, consider the statement: “Dragapult is 15.5% of the meta but only 46.7% expected against the field.” In our workflow, this statement is decomposed into auditable steps:

- 1) Extract share constants from the fixed Trainer Hill snapshot.
- 2) Normalize top-14 shares for expected-value computation.
- 3) Compute weighted expectation from the encoded 14×14 matrix.
- 4) Express the result as an exact rational in Lean.
- 5) Prove the inequality relative to 50% in theorem form.
- 6) Reuse the same constants in manuscript tables and prose.

Each step is versioned and reproducible. If any upstream value changes (for example, an updated matrix entry after a data correction), the downstream theorem or table regeneration will fail or produce a different value, making drift explicit. This is fundamentally stronger than spreadsheet-style pipelines where hidden references and manual edits can silently desynchronize outputs.

The same pattern is used for all other headline claims in this paper: Grimmsnarl’s top expected value, Mega Absol’s equilibrium concentration, Raging Bolt’s 67.3% counter edge, and Bo3 amplification values. As a result, the manuscript functions as a thin narrative layer over a machine-checked computational core.

G. Roadmap for Continuous Metagame Monitoring

The current study is a fixed-window first submission, but the infrastructure naturally supports continuous operation. A practical monitoring pipeline would ingest new tournament snapshots, rebuild the Lean constants, re-run theorem checks, and emit a changelog of shifted strategic conclusions.

The key benefit of this setup is stability under iteration. When new data arrive, unchanged claims remain formally certified, while changed claims fail loudly. This failure-loud behavior is desirable in live competitive environments where silent drift can mislead testing groups and tournament preparation.

A second benefit is selective recomputation. Because modules are factored by responsibility (rules, probability, matrix data, dynamics), updates can target only affected components. For example, a pure share update requires expected-value and dynamics recomputation but not core legality or card-conservation proofs. This keeps turnaround practical for weekly competitive cycles.

Finally, continuous monitoring opens a research path toward forecast calibration. Predictions made from one window (e.g., Dragapult decline pressure) can be scored against the next window. Repeated over many windows, this enables quantitative assessment of how much behavior in the ecosystem

TABLE VI
FORMALIZATION MODULE BREAKDOWN.

Module group	Files	Lines	Theorems
Core Rules & Semantics	10	4,097	253
Card Effects & Actions	8	3,055	171
Probability & Resources	8	2,522	176
Game Theory & Dynamics	6	3,562	350
Real Metagame Analysis	6	2,013	179
Infrastructure & Validation	7	1,872	97
Additional Specialized Modules	30	12,672	1,284
Total	75	29,793	2,510

is explained by payoff pressure versus exogenous factors such as content cycles, testing-group coordination, and card-availability shocks. We next delimit the empirical and modeling bounds of these results in Section X.

X. THREATS TO VALIDITY

Temporal locality. The analyzed window is three weeks. Metagames can shift rapidly due to innovation, counter-adaptation, and card availability. Our claims describe this window precisely; they are not universal constants. However, temporal locality is not purely a weakness: short windows reduce hidden confounding from major ruleset changes. Future work should combine rolling windows with change-point detection to separate genuine adaptation from transient noise.

Top-14 normalization. Expected win rates are normalized over the modeled 69.5% top-14 subset. A different treatment of the 30.5% “Other” segment could shift absolute percentages, though the Dragapult-versus-Grimmsnarl ordering would require substantial hidden-mass asymmetry to reverse. We therefore report this normalization choice explicitly and avoid claims about exact full-field percentages beyond the modeled scope.

A. Robustness Analysis

A natural concern is whether the unmodeled 30.5% of the field could reverse our conclusions. We provide machine-checked worst-case bounds. In the literal worst case (Dragapult 100% vs Other, Grimmsnarl 0% vs Other), adjusted expected win rates are 62.9% and 36.6%. Dragapult requires at least 57.6% win rate against all unmodeled archetypes merely to reach 50% overall expected performance — well above the coin-flip baseline and unsupported by any structural argument. Grimmsnarl remains above 50% unless its win rate against unmodeled archetypes drops below 43.9%, a scenario inconsistent with its favorable type coverage. Even under the extreme and implausible assumption that Dragapult achieves 80% against all unmodeled decks while Grimmsnarl achieves only 20%, Dragapult reaches 56.8% and Grimmsnarl drops to 42.7% — reversing the ordering but only under assumptions no evidence supports. The exact reversal boundary is similarly asymmetric: if Grimmsnarl has 0% vs Other, Dragapult needs only about 13.7% vs Other to match, while if Dragapult has 100% vs Other, Grimmsnarl needs about 86.3% to match. The paradox is robust.

We also verify robustness to meta share perturbation. The popularity paradox arises from Dragapult’s matchup spread — it loses to 9 of 13 non-mirror opponents — rather than from any particular share distribution. Machine-checked share-perturbation theorems in `SharePerturbation.lean` show that even if Dragapult’s share drops from 15.5% to 5%, its expected field win rate remains below 50%; conversely, if Grimmsnarl’s share rises from 5.1% to 15.5% (absorbing Dragapult’s current share), its expected field win rate remains above 50%. The paradox is structural: it derives from the matchup matrix, not the share vector, so the central finding is not an artifact of a particular metagame snapshot but reflects a persistent property of the underlying strategic landscape.

```

1 theorem popularity_paradox_robust_worst_case : 
2   adjustedWR dragapultTop14WR 1 = 629243 / 
3   1000000 /\
4   adjustedWR grimmsnarlTop14WR 0 = 366061 / 
5   1000000 /\
6   adjustedWR dragapultTop14WR 1 - adjustedWR 
7   grimmsnarlTop14WR 0 = 
8   131591 / 500000 := by
9   constructor
10  · native Decide
11  constructor
12  · native Decide
13  · native Decide
14  · native Decide
15  · native Decide
16  · native Decide
17  · native Decide
18  · native Decide
19  · native Decide
20  · native Decide
21  · native Decide
22  · native Decide
23  · native Decide
24  · native Decide
25  · native Decide
26  · native Decide
27  · native Decide
28  · native Decide
29  · native Decide
30  · native Decide
31  · native Decide
32  · native Decide
33  · native Decide
34  · native Decide
35  · native Decide
36  · native Decide
37  · native Decide
38  · native Decide
39  · native Decide
40  · native Decide
41  · native Decide
42  · native Decide
43  · native Decide
44  · native Decide
45  · native Decide
46  · native Decide
47  · native Decide
48  · native Decide
49  · native Decide
50  · native Decide
51  · native Decide
52  · native Decide
53  · native Decide
54  · native Decide
55  · native Decide
56  · native Decide
57  · native Decide
58  · native Decide
59  · native Decide
60  · native Decide
61  · native Decide
62  · native Decide
63  · native Decide
64  · native Decide
65  · native Decide
66  · native Decide
67  · native Decide
68  · native Decide
69  · native Decide
70  · native Decide
71  · native Decide
72  · native Decide
73  · native Decide
74  · native Decide
75  · native Decide
76  · native Decide
77  · native Decide
78  · native Decide
79  · native Decide
80  · native Decide
81  · native Decide
82  · native Decide
83  · native Decide
84  · native Decide
85  · native Decide
86  · native Decide
87  · native Decide
88  · native Decide
89  · native Decide
90  · native Decide
91  · native Decide
92  · native Decide
93  · native Decide
94  · native Decide
95  · native Decide
96  · native Decide
97  · native Decide
98  · native Decide
99  · native Decide
100  · native Decide
101  · native Decide
102  · native Decide
103  · native Decide
104  · native Decide
105  · native Decide
106  · native Decide
107  · native Decide
108  · native Decide
109  · native Decide
110  · native Decide
111  · native Decide
112  · native Decide
113  · native Decide
114  · native Decide
115  · native Decide
116  · native Decide
117  · native Decide
118  · native Decide
119  · native Decide
120  · native Decide
121  · native Decide
122  · native Decide
123  · native Decide
124  · native Decide
125  · native Decide
126  · native Decide
127  · native Decide
128  · native Decide
129  · native Decide
130  · native Decide
131  · native Decide
132  · native Decide
133  · native Decide
134  · native Decide
135  · native Decide
136  · native Decide
137  · native Decide
138  · native Decide
139  · native Decide
140  · native Decide
141  · native Decide
142  · native Decide
143  · native Decide
144  · native Decide
145  · native Decide
146  · native Decide
147  · native Decide
148  · native Decide
149  · native Decide
150  · native Decide
151  · native Decide
152  · native Decide
153  · native Decide
154  · native Decide
155  · native Decide
156  · native Decide
157  · native Decide
158  · native Decide
159  · native Decide
160  · native Decide
161  · native Decide
162  · native Decide
163  · native Decide
164  · native Decide
165  · native Decide
166  · native Decide
167  · native Decide
168  · native Decide
169  · native Decide
170  · native Decide
171  · native Decide
172  · native Decide
173  · native Decide
174  · native Decide
175  · native Decide
176  · native Decide
177  · native Decide
178  · native Decide
179  · native Decide
180  · native Decide
181  · native Decide
182  · native Decide
183  · native Decide
184  · native Decide
185  · native Decide
186  · native Decide
187  · native Decide
188  · native Decide
189  · native Decide
190  · native Decide
191  · native Decide
192  · native Decide
193  · native Decide
194  · native Decide
195  · native Decide
196  · native Decide
197  · native Decide
198  · native Decide
199  · native Decide
200  · native Decide
201  · native Decide
202  · native Decide
203  · native Decide
204  · native Decide
205  · native Decide
206  · native Decide
207  · native Decide
208  · native Decide
209  · native Decide
210  · native Decide
211  · native Decide
212  · native Decide
213  · native Decide
214  · native Decide
215  · native Decide
216  · native Decide
217  · native Decide
218  · native Decide
219  · native Decide
220  · native Decide
221  · native Decide
222  · native Decide
223  · native Decide
224  · native Decide
225  · native Decide
226  · native Decide
227  · native Decide
228  · native Decide
229  · native Decide
230  · native Decide
231  · native Decide
232  · native Decide
233  · native Decide
234  · native Decide
235  · native Decide
236  · native Decide
237  · native Decide
238  · native Decide
239  · native Decide
240  · native Decide
241  · native Decide
242  · native Decide
243  · native Decide
244  · native Decide
245  · native Decide
246  · native Decide
247  · native Decide
248  · native Decide
249  · native Decide
250  · native Decide
251  · native Decide
252  · native Decide
253  · native Decide
254  · native Decide
255  · native Decide
256  · native Decide
257  · native Decide
258  · native Decide
259  · native Decide
260  · native Decide
261  · native Decide
262  · native Decide
263  · native Decide
264  · native Decide
265  · native Decide
266  · native Decide
267  · native Decide
268  · native Decide
269  · native Decide
270  · native Decide
271  · native Decide
272  · native Decide
273  · native Decide
274  · native Decide
275  · native Decide
276  · native Decide
277  · native Decide
278  · native Decide
279  · native Decide
280  · native Decide
281  · native Decide
282  · native Decide
283  · native Decide
284  · native Decide
285  · native Decide
286  · native Decide
287  · native Decide
288  · native Decide
289  · native Decide
290  · native Decide
291  · native Decide
292  · native Decide
293  · native Decide
294  · native Decide
295  · native Decide
296  · native Decide
297  · native Decide
298  · native Decide
299  · native Decide
300  · native Decide
301  · native Decide
302  · native Decide
303  · native Decide
304  · native Decide
305  · native Decide
306  · native Decide
307  · native Decide
308  · native Decide
309  · native Decide
310  · native Decide
311  · native Decide
312  · native Decide
313  · native Decide
314  · native Decide
315  · native Decide
316  · native Decide
317  · native Decide
318  · native Decide
319  · native Decide
320  · native Decide
321  · native Decide
322  · native Decide
323  · native Decide
324  · native Decide
325  · native Decide
326  · native Decide
327  · native Decide
328  · native Decide
329  · native Decide
330  · native Decide
331  · native Decide
332  · native Decide
333  · native Decide
334  · native Decide
335  · native Decide
336  · native Decide
337  · native Decide
338  · native Decide
339  · native Decide
340  · native Decide
341  · native Decide
342  · native Decide
343  · native Decide
344  · native Decide
345  · native Decide
346  · native Decide
347  · native Decide
348  · native Decide
349  · native Decide
350  · native Decide
351  · native Decide
352  · native Decide
353  · native Decide
354  · native Decide
355  · native Decide
356  · native Decide
357  · native Decide
358  · native Decide
359  · native Decide
360  · native Decide
361  · native Decide
362  · native Decide
363  · native Decide
364  · native Decide
365  · native Decide
366  · native Decide
367  · native Decide
368  · native Decide
369  · native Decide
370  · native Decide
371  · native Decide
372  · native Decide
373  · native Decide
374  · native Decide
375  · native Decide
376  · native Decide
377  · native Decide
378  · native Decide
379  · native Decide
380  · native Decide
381  · native Decide
382  · native Decide
383  · native Decide
384  · native Decide
385  · native Decide
386  · native Decide
387  · native Decide
388  · native Decide
389  · native Decide
390  · native Decide
391  · native Decide
392  · native Decide
393  · native Decide
394  · native Decide
395  · native Decide
396  · native Decide
397  · native Decide
398  · native Decide
399  · native Decide
400  · native Decide
401  · native Decide
402  · native Decide
403  · native Decide
404  · native Decide
405  · native Decide
406  · native Decide
407  · native Decide
408  · native Decide
409  · native Decide
410  · native Decide
411  · native Decide
412  · native Decide
413  · native Decide
414  · native Decide
415  · native Decide
416  · native Decide
417  · native Decide
418  · native Decide
419  · native Decide
420  · native Decide
421  · native Decide
422  · native Decide
423  · native Decide
424  · native Decide
425  · native Decide
426  · native Decide
427  · native Decide
428  · native Decide
429  · native Decide
430  · native Decide
431  · native Decide
432  · native Decide
433  · native Decide
434  · native Decide
435  · native Decide
436  · native Decide
437  · native Decide
438  · native Decide
439  · native Decide
440  · native Decide
441  · native Decide
442  · native Decide
443  · native Decide
444  · native Decide
445  · native Decide
446  · native Decide
447  · native Decide
448  · native Decide
449  · native Decide
450  · native Decide
451  · native Decide
452  · native Decide
453  · native Decide
454  · native Decide
455  · native Decide
456  · native Decide
457  · native Decide
458  · native Decide
459  · native Decide
460  · native Decide
461  · native Decide
462  · native Decide
463  · native Decide
464  · native Decide
465  · native Decide
466  · native Decide
467  · native Decide
468  · native Decide
469  · native Decide
470  · native Decide
471  · native Decide
472  · native Decide
473  · native Decide
474  · native Decide
475  · native Decide
476  · native Decide
477  · native Decide
478  · native Decide
479  · native Decide
480  · native Decide
481  · native Decide
482  · native Decide
483  · native Decide
484  · native Decide
485  · native Decide
486  · native Decide
487  · native Decide
488  · native Decide
489  · native Decide
490  · native Decide
491  · native Decide
492  · native Decide
493  · native Decide
494  · native Decide
495  · native Decide
496  · native Decide
497  · native Decide
498  · native Decide
499  · native Decide
500  · native Decide
501  · native Decide
502  · native Decide
503  · native Decide
504  · native Decide
505  · native Decide
506  · native Decide
507  · native Decide
508  · native Decide
509  · native Decide
510  · native Decide
511  · native Decide
512  · native Decide
513  · native Decide
514  · native Decide
515  · native Decide
516  · native Decide
517  · native Decide
518  · native Decide
519  · native Decide
520  · native Decide
521  · native Decide
522  · native Decide
523  · native Decide
524  · native Decide
525  · native Decide
526  · native Decide
527  · native Decide
528  · native Decide
529  · native Decide
530  · native Decide
531  · native Decide
532  · native Decide
533  · native Decide
534  · native Decide
535  · native Decide
536  · native Decide
537  · native Decide
538  · native Decide
539  · native Decide
540  · native Decide
541  · native Decide
542  · native Decide
543  · native Decide
544  · native Decide
545  · native Decide
546  · native Decide
547  · native Decide
548  · native Decide
549  · native Decide
550  · native Decide
551  · native Decide
552  · native Decide
553  · native Decide
554  · native Decide
555  · native Decide
556  · native Decide
557  · native Decide
558  · native Decide
559  · native Decide
560  · native Decide
561  · native Decide
562  · native Decide
563  · native Decide
564  · native Decide
565  · native Decide
566  · native Decide
567  · native Decide
568  · native Decide
569  · native Decide
570  · native Decide
571  · native Decide
572  · native Decide
573  · native Decide
574  · native Decide
575  · native Decide
576  · native Decide
577  · native Decide
578  · native Decide
579  · native Decide
580  · native Decide
581  · native Decide
582  · native Decide
583  · native Decide
584  · native Decide
585  · native Decide
586  · native Decide
587  · native Decide
588  · native Decide
589  · native Decide
590  · native Decide
591  · native Decide
592  · native Decide
593  · native Decide
594  · native Decide
595  · native Decide
596  · native Decide
597  · native Decide
598  · native Decide
599  · native Decide
600  · native Decide
601  · native Decide
602  · native Decide
603  · native Decide
604  · native Decide
605  · native Decide
606  · native Decide
607  · native Decide
608  · native Decide
609  · native Decide
610  · native Decide
611  · native Decide
612  · native Decide
613  · native Decide
614  · native Decide
615  · native Decide
616  · native Decide
617  · native Decide
618  · native Decide
619  · native Decide
620  · native Decide
621  · native Decide
622  · native Decide
623  · native Decide
624  · native Decide
625  · native Decide
626  · native Decide
627  · native Decide
628  · native Decide
629  · native Decide
630  · native Decide
631  · native Decide
632  · native Decide
633  · native Decide
634  · native Decide
635  · native Decide
636  · native Decide
637  · native Decide
638  · native Decide
639  · native Decide
640  · native Decide
641  · native Decide
642  · native Decide
643  · native Decide
644  · native Decide
645  · native Decide
646  · native Decide
647  · native Decide
648  · native Decide
649  · native Decide
650  · native Decide
651  · native Decide
652  · native Decide
653  · native Decide
654  · native Decide
655  · native Decide
656  · native Decide
657  · native Decide
658  · native Decide
659  · native Decide
660  · native Decide
661  · native Decide
662  · native Decide
663  · native Decide
664  · native Decide
665  · native Decide
666  · native Decide
667  · native Decide
668  · native Decide
669  · native Decide
670  · native Decide
671  · native Decide
672  · native Decide
673  · native Decide
674  · native Decide
675  · native Decide
676  · native Decide
677  · native Decide
678  · native Decide
679  · native Decide
680  · native Decide
681  · native Decide
682  · native Decide
683  · native Decide
684  · native Decide
685  · native Decide
686  · native Decide
687  · native Decide
688  · native Decide
689  · native Decide
690  · native Decide
691  · native Decide
692  · native Decide
693  · native Decide
694  · native Decide
695  · native Decide
696  · native Decide
697  · native Decide
698  · native Decide
699  · native Decide
700  · native Decide
701  · native Decide
702  · native Decide
703  · native Decide
704  · native Decide
705  · native Decide
706  · native Decide
707  · native Decide
708  · native Decide
709  · native Decide
710  · native Decide
711  · native Decide
712  · native Decide
713  · native Decide
714  · native Decide
715  · native Decide
716  · native Decide
717  · native Decide
718  · native Decide
719  · native Decide
720  · native Decide
721  · native Decide
722  · native Decide
723  · native Decide
724  · native Decide
725  · native Decide
726  · native Decide
727  · native Decide
728  · native Decide
729  · native Decide
730  · native Decide
731  · native Decide
732  · native Decide
733  · native Decide
734  · native Decide
735  · native Decide
736  · native Decide
737  · native Decide
738  · native Decide
739  · native Decide
740  · native Decide
741  · native Decide
742  · native Decide
743  · native Decide
744  · native Decide
745  · native Decide
746  · native Decide
747  · native Decide
748  · native Decide
749  · native Decide
750  · native Decide
751  · native Decide
752  · native Decide
753  · native Decide
754  · native Decide
755  · native Decide
756  · native Decide
757  · native Decide
758  · native Decide
759  · native Decide
760  · native Decide
761  · native Decide
762  · native Decide
763  · native Decide
764  · native Decide
765  · native Decide
766  · native Decide
767  · native Decide
768  · native Decide
769  · native Decide
770  · native Decide
771  · native Decide
772  · native Decide
773  · native Decide
774  · native Decide
775  · native Decide
776  · native Decide
777  · native Decide
778  · native Decide
779  · native Decide
780  · native Decide
781  · native Decide
782  · native Decide
783  · native Decide
784  · native Decide
785  · native Decide
786  · native Decide
787  · native Decide
788  · native Decide
789  · native Decide
790  · native Decide
791  · native Decide
792  · native Decide
793  · native Decide
794  · native Decide
795  · native Decide
796  · native Decide
797  · native Decide
798  · native Decide
799  · native Decide
800  · native Decide
801  · native Decide
802  · native Decide
803  · native Decide
804  · native Decide
805  · native Decide
806  · native Decide
807  · native Decide
808  · native Decide
809  · native Decide
810  · native Decide
811  · native Decide
812  · native Decide
813  · native Decide
814  · native Decide
815  · native Decide
816  · native Decide
817  · native Decide
818  · native Decide
819  · native Decide
820  · native Decide
821  · native Decide
822  · native Decide
823  · native Decide
824  · native Decide
825  · native Decide
826  · native Decide
827  · native Decide
828  · native Decide
829  · native Decide
830  · native Decide
831  · native Decide
832  · native Decide
833  · native Decide
834  · native Decide
835  · native Decide
836  · native Decide
837  · native Decide
838  · native Decide
839  · native Decide
840  · native Decide
841  · native Decide
842  · native Decide
843  · native Decide
844  · native Decide
845  · native Decide
846  · native Decide
847  · native Decide
848  · native Decide
849  · native Decide
850  · native Decide
851  · native Decide
852  · native Decide
853  · native Decide
854  · native Decide
855  · native Decide
856  · native Decide
857  · native Decide
858  · native Decide
859  · native Decide
860  · native Decide
861  · native Decide
862  · native Decide
863  · native Decide
864  · native Decide
865  · native Decide
866  · native Decide
867  · native Decide
868  · native Decide
869  · native Decide
870  · native Decide
871  · native Decide
872  · native Decide
873  · native Decide
874  · native Decide
875  · native Decide
876  · native Decide
877  · native Decide
878  · native Decide
879  · native Decide
880  · native Decide
881  · native Decide
882  · native Decide
883  · native Decide
884  · native Decide
885  · native Decide
886  · native Decide
887  · native Decide
888  · native Decide
889  · native Decide
890  · native Decide
891  · native Decide
892  · native Decide
893  · native Decide
894  · native Decide
895  · native Decide
896  · native Decide
897  · native Decide
898  · native Decide
899  · native Decide
900  · native Decide
901  · native Decide
902  · native Decide
903  · native Decide
904  · native Decide
905  · native Decide
906  · native Decide
907  · native Decide
908  · native Decide
909  · native Decide
910  · native Decide
911  · native Decide
912  · native Decide
913  · native Decide
914  · native Decide
915  · native Decide
916  · native Decide
917  · native Decide
918  · native Decide
919  · native Decide
920  · native Decide
921  · native Decide
922  · native Decide
923  · native Decide
924  · native Decide
925  · native Decide
926  · native Decide
927  · native Decide
928  · native Decide
929  · native Decide
930  · native Decide
931  · native Decide
932  · native Decide
933  · native Decide
934  · native Decide
935  · native Decide
936  · native Decide
937  · native Decide
938  · native Decide
939  · native Decide
940  · native Decide
941  · native Decide
942  · native Decide
943  · native Decide
944  · native Decide
945  · native Decide
946  · native Decide
947  · native Decide
948  · native Decide
949  · native Decide
950  · native Decide
951  · native Decide
952  · native Decide
953  · native Decide
954  · native Decide
955  · native Decide
956  · native Decide
957  · native Decide
958  · native Decide
959  · native Decide
960  · native Decide
961  · native Decide
962  · native Decide
963  · native Decide
964  · native Decide
965  · native Decide
966  · native Decide
967  · native Decide
968  · native Decide
969  · native Decide
970  · native Decide
971  · native Decide
972  · native Decide
973  · native Decide
974  · native Decide
975  · native Decide
976  · native Decide
977  · native Decide
978  · native Decide
979  · native Decide
980  · native Decide
981  · native Decide
982  · native Decide
983  · native Decide
984  · native Decide
985  · native Decide
986  · native Decide
987  · native Decide
988  · native Decide
989  · native Decide
990  · native Decide
991  · native Decide
992  · native Decide
993  · native Decide
994  · native Decide
995  · native Decide
996  · native Decide
997  · native Decide
998  · native Decide
999  · native Decide
1000  · native Decide
1001  · native Decide
1002  · native Decide
1003  · native Decide
1004  · native Decide
1005  · native Decide
1006  · native Decide
1007  · native Decide
1008  · native Decide
1009  · native Decide
1010  · native Decide
1011  · native Decide
1012  · native Decide
1013  · native Decide
1014  · native Decide
1015  · native Decide
1016  · native Decide
1017  · native Decide
1018  · native Decide
1019  · native Decide
1020  · native Decide
1021  · native Decide
1022  · native Decide
1023  · native Decide
1024  · native Decide
1025  · native Decide
1026  · native Decide
1027  · native Decide
1028  · native Decide
1029  · native Decide
1030  · native Decide
1031  · native Decide
1032  · native Decide
1033  · native Decide
1034  · native Decide
1035  · native Decide
1036  · native Decide
1037  · native Decide
1038  · native Decide
1039  · native Decide
1040  · native Decide
1041  · native Decide
1042  · native Decide
1043  · native Decide
1044  · native Decide
1045  · native Decide
1046  · native Decide
1047  · native Decide
1048  · native Decide
1049  · native Decide
1050  · native Decide
1051  · native Decide
1052  · native Decide
1053  · native Decide
1054  · native Decide
1055  · native Decide
1056  · native Decide
1057  · native Decide
1058  · native Decide
1059  · native Decide
1060  · native Decide
1061  · native Decide
1062  · native Decide
1063  · native Decide
1064  · native Decide
1065  · native Decide
1066  · native Decide
1067  · native Decide
1068  · native Decide
1069  · native Decide
1070  · native Decide
1071  · native Decide
1072  · native Decide
1073  · native Decide
1074  · native Decide
1075  · native Decide
1076  · native Decide
1077  · native Decide
1078  · native Decide
1079  · native Decide
1080  · native Decide
1081  · native Decide
1082  · native Decide
1083  · native Decide
1084  · native Decide
1085  · native Decide
1086  · native Decide
1087  · native Decide
1088  · native Decide
1089  · native Decide
1090  · native Decide
1091  · native Decide
1092  · native Decide
1093  · native Decide
1094  · native Decide
1095  · native Decide
1096  · native Decide
1097  · native Decide
1098  · native Decide
1099  · native Decide
1100  · native Decide
1101  · native Decide
1102  · native Decide
1103  · native Decide
1104  · native Decide
1105  · native Decide
1106  · native Decide
1107  · native Decide
1108  · native Decide
1109  · native Decide
1110  · native Decide
1111  · native Decide
1112  · native Decide
1113  · native Decide
1114  · native Decide
1115  · native Decide
1116  · native Decide
1117  · native Decide
1118  · native Decide
1119  · native Decide
1120  · native Decide
1121  · native Decide
1122  · native Decide
1123  · native Decide
1124  · native Decide
1125  · native Decide
1126  · native Decide
1127  · native Decide
1128  · native Decide
1129  · native Decide
1130  · native Decide
1131  · native Decide
1132  · native Decide
1133  · native Decide
1134  · native Decide
1135  · native Decide
1136  · native Decide
1137  · native Decide
1138  · native Decide
1139  · native Decide
1140  · native Decide
1141  · native Decide
1142  · native Decide
1143  · native Decide
1144  · native Decide
1145  · native Decide
1146  · native Decide
1147  · native Decide
1148  · native Decide
1149  · native Decide
1150  · native Decide
1151  · native Decide
1152  · native Decide
1153  · native Decide
1154  · native Decide
1155  · native Decide
1156  · native Decide
1157  · native Decide
1158  · native Decide
1159  · native Decide
1160  · native Decide
1161  · native Decide
1162  · native Decide
1163  · native Decide
1164  · native Decide
1165  · native Decide
1166  · native Decide
1167  · native Decide
1168  · native Decide
1169  · native Decide
1170  · native Decide
1171  · native Decide
1172  · native Decide
1173  · native Decide
1174  · native Decide
1175  · native Decide
1176  · native Decide
1177  · native Decide
1178  · native Decide
1179  · native Decide
1180  · native Decide
1181  · native Decide
1182  · native Decide
1183  · native Decide
1184  · native Decide
1185  · native Decide
1186  · native Decide
1187  · native Decide
1188  · native Decide
1189  · native Decide
1190  · native Decide
1191  · native Decide
1192  · native Decide
1193  · native Decide
1194  · native Decide

```

paradox: Dragapult is most played yet sub-50% in expected field performance, while Grimmsnarl is less represented than the dominant choices and has the highest expected win rate.

We also connect static matchup structure to dynamic and tournament implications. The Nash profile is highly concentrated, replicator dynamics predict Dragapult decline and Ceruleedge extinction pressure, and best-of-three math amplifies already-large matchup edges ($67.3\% \rightarrow 74.9\%$).

Immediate next steps are concrete: (i) extend the dataset to rolling weekly windows with uncertainty intervals, (ii) model the 30.5% “Other” segment explicitly instead of excluding it from matrix analysis, and (iii) add hierarchical sub-archetype clustering to capture list-level variance and pilot heterogeneity.

A. Broader Implications for Competitive Game Science

This case study suggests a general template for competitive-game research. First, formalize core mechanics and legality. Second, encode empirical payoff data as exact values. Third, express strategic claims as theorem-checkable statements. Fourth, tie those claims to tournament-relevant objectives instead of abstract utility alone. The resulting pipeline is portable across many environments with discrete strategy choices and measurable outcomes.

The portability argument matters because many competitive ecosystems face the same methodological failure mode: high-quality data exist, but conclusions are often produced by ad hoc tooling that mixes assumptions and results without explicit traceability. A proof-assisted workflow does not replace domain expertise; it structures it. Experts still decide which assumptions are reasonable, but once assumptions are fixed, conclusions become machine-auditable rather than rhetorical.

A second implication concerns collaboration between researchers and practitioners. Formal artifacts can be integrated into testing-team workflows as “verified baselines” against which local innovations are evaluated. For example, a team can begin from a certified weighted matchup model, then test whether a candidate list change moves specific matchup entries enough to alter tier or equilibrium-relevant conclusions. This is far more informative than relying on isolated scrim records without a stable analytical backbone.

Finally, this work contributes to a broader view of theorem proving in applied settings. Proof assistants are often associated with pure mathematics or compiler correctness. Our results show they are also practical for empirical strategic science when the domain provides structured, finite data and well-defined objective functions. In that regime, formal methods can simultaneously improve reproducibility, interpretability, and decision quality.

Beyond this specific metagame snapshot, the broader contribution is methodological. Formal verification can serve as a practical scientific instrument for competitive game ecosystems: it turns qualitative metagame claims into executable definitions, theorem statements, and reproducible evidence.

DATA AVAILABILITY

Data were extracted from Trainer Hill (trainerhill.com) on February 19, 2026, filtering for Pokémon TCG events with 50 or more players between January 29 and February 19, 2026. Trainer Hill aggregates results from the Limitless TCG tournament platform. The 14×14 matchup matrix was computed from win-loss-tie records; ties were weighted as one-third of a win following the standard conversion $WR = (W + T/3)/(W + L + T)$. The complete matrix with raw W-L-T counts is archived in the repository. All Lean 4 source code, data files, and build instructions will be made publicly available upon acceptance in an anonymized artifact.

REFERENCES

- [1] L. de Moura and S. Ullrich, “The Lean 4 theorem prover and programming language,” in *Automated Deduction – CADE-28*, ser. LNCS, vol. 12699. Springer, 2021, pp. 625–635.
- [2] Trainer Hill, “Pokémon tcg tournament metagame aggregates,” <https://www.trainerhill.com>, 2026, aggregates Limitless TCG tournament results; accessed: 2026-02-20.
- [3] Limitless TCG, “Limitless tcg tournament platform,” <https://limitlesstcg.com>, 2024, accessed: 2026-02-20.
- [4] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [5] T. J. Schaefer, “On the complexity of some two-person perfect-information games,” in *Proc. 8th ACM Symp. Theory of Computing (STOC)*, 1978, pp. 216–225.
- [6] A. S. Fraenkel and D. Lichtenstein, “Computing a perfect strategy for $n \times n$ chess requires time exponential in n ,” *Journal of Combinatorial Theory, Series A*, vol. 31, no. 2, pp. 199–214, 1981.
- [7] M. Bowling, N. Burch, M. Johanson, and O. Tammelin, “Heads-up limit hold’em poker is solved,” *Science*, vol. 347, no. 6218, pp. 145–149, 2015.
- [8] N. Brown and T. Sandholm, “Superhuman AI for heads-up no-limit poker: Libratus beats top professionals,” *Science*, vol. 359, no. 6374, pp. 418–424, 2018.
- [9] ———, “Superhuman AI for multiplayer poker,” *Science*, vol. 365, no. 6456, pp. 885–890, 2019.
- [10] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou *et al.*, “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play,” *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018.
- [11] O. Vinyals, I. Babuschkin, W. M. Czarnecki *et al.*, “Grandmaster level in StarCraft II using multi-agent reinforcement learning,” *Nature*, vol. 575, no. 7782, pp. 350–354, 2019.
- [12] P. I. Cowling, C. D. Ward, and E. J. Powley, “Ensemble determinization in Monte Carlo tree search for the imperfect information card game Magic: The gathering,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 4, pp. 241–257, 2012.
- [13] C. D. Ward and P. I. Cowling, “Monte Carlo search applied to card selection in Magic: The Gathering,” in *2009 IEEE Symposium on Computational Intelligence and Games*, 2009, pp. 9–16.
- [14] A. Santos, P. Barros, and M. Aragão, “Monte Carlo tree search experiments in Hearthstone,” in *Proc. IEEE Conf. Computational Intelligence and Games (CIG)*, 2017, pp. 272–279.
- [15] S. Zhang and M. Buro, “Improving Hearthstone AI by combining MCTS and supervised learning algorithms,” in *Proc. AAAI Conf. Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 2017, pp. 68–74.
- [16] M. Bjørke and D. Fludal, “Deckbuilding in Magic: The Gathering using a genetic algorithm,” in *Proceedings of NISK*, 2017.
- [17] T. Dockhorn *et al.*, “Introducing the hearthstone-AI competition,” *arXiv preprint arXiv:1906.04238*, 2019.
- [18] J. Kowalski and R. Miernik, “Summarizing strategy card game AI competition,” in *2023 IEEE Conference on Games (CoG)*, 2023, pp. 1–8.
- [19] G. Gonthier, “Formal proof — the four-color theorem,” *Notices of the American Mathematical Society*, vol. 55, no. 11, pp. 1382–1393, 2008.
- [20] J. Avigad, K. Donnelly, D. Gray, and P. Raff, “A formally verified proof of the prime number theorem,” *ACM Transactions on Computational Logic*, vol. 9, no. 1, pp. 2:1–2:23, 2007.

- [21] The mathlib Community, “The Lean mathematical library,” in *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP)*. ACM, 2020, pp. 367–381.
- [22] T. Hales, M. Adams, G. Bauer, T. D. Dang, J. Harrison, H. Le Truong, C. Kaliszyk, V. Magron, S. McLaughlin, T. T. Nguyen, Q. T. Nguyen, T. Nipkow, S. Obua, J. Pleso, J. Rute, A. Solovyev, A. H. Ta, T. D. Tran, T. H. Trieu, J. Urban, K. Vu, and R. Zumkeller, “A formal proof of the Kepler conjecture,” *Forum of Mathematics, Pi*, vol. 5, p. e2, 2017.
- [23] F. Hosch and L. Kovács, “Formalizing hearthstone card effects in Isabelle/HOL,” in *Proc. 13th Int. Conf. Interactive Theorem Proving (ITP)*, ser. LIPIcs, vol. 237, 2022, pp. 1–18, (representative of formal TCG work).
- [24] J. M. Smith and G. R. Price, “The logic of animal conflict,” *Nature*, vol. 246, pp. 15–18, 1973.
- [25] P. D. Taylor and L. B. Jonker, “Evolutionary stable strategies and game dynamics,” *Mathematical Biosciences*, vol. 40, no. 1–2, pp. 145–156, 1978.
- [26] J. W. Weibull, *Evolutionary Game Theory*. MIT Press, 1997.
- [27] The Pokémon Company International, “Pokémon trading card game — rules & resources,” <https://www.pokemon.com/us/pokemon-tcg/rules>, 2024, accessed: 2026-02-20.
- [28] ———, “Play pokémon tournament rules handbook,” <https://www.pokemon.com/us/play-pokemon/about/tournaments-rules-and-resources>, 2024, accessed: 2026-02-20.
- [29] A. Tversky and D. Kahneman, “Judgment under uncertainty: Heuristics and biases,” *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.
- [30] D. Kahneman and A. Tversky, “Prospect theory: An analysis of decision under risk,” *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.
- [31] A. V. Banerjee, “A simple model of herd behavior,” *The Quarterly Journal of Economics*, vol. 107, no. 3, pp. 797–817, 1992.
- [32] S. Bikhchandani, D. Hirshleifer, and I. Welch, “A theory of fads, fashion, custom, and cultural change as informational cascades,” *Journal of Political Economy*, vol. 100, no. 5, pp. 992–1026, 1992.
- [33] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds., *Algorithmic Game Theory*. Cambridge University Press, 2007.
- [34] J. von Neumann, “Zur Theorie der Gesellschaftsspiele,” *Mathematische Annalen*, vol. 100, no. 1, pp. 295–320, 1928.
- [35] J. F. Nash, “Equilibrium points in n -person games,” *Proceedings of the National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, 1950.
- [36] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.
- [37] W. H. Sandholm, *Population Games and Evolutionary Dynamics*. MIT Press, 2010.
- [38] O. Romero and I. Millet, “Analysis of swiss-system tournament pairings,” *Journal of Quantitative Analysis in Sports*, vol. 18, no. 3, pp. 213–229, 2022.
- [39] K. Buzzard, J. Commelin, and P. Massot, “Formalising perfectoid spaces,” in *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP)*. ACM, 2020, pp. 299–312.