

Modeling 1v1 Pokemon Battles with Zero-Sum Games and Expectimax Trees

Jake Smith, Nico Salm, Kevin Wang



What is a 1v1 Pokemon Battle?

Specifically the 1v1 Smogon format

Before battling, you need to build a team. In the 1v1 format, you bring a team of 3 Pokemon.



Garchomp @ Choice Band
Ability: Rough Skin
EVs: 4 HP / 252 Atk / 252 Spe
Jolly Nature
- Outrage
- Earthquake
- Fire Fang
- Stone Edge

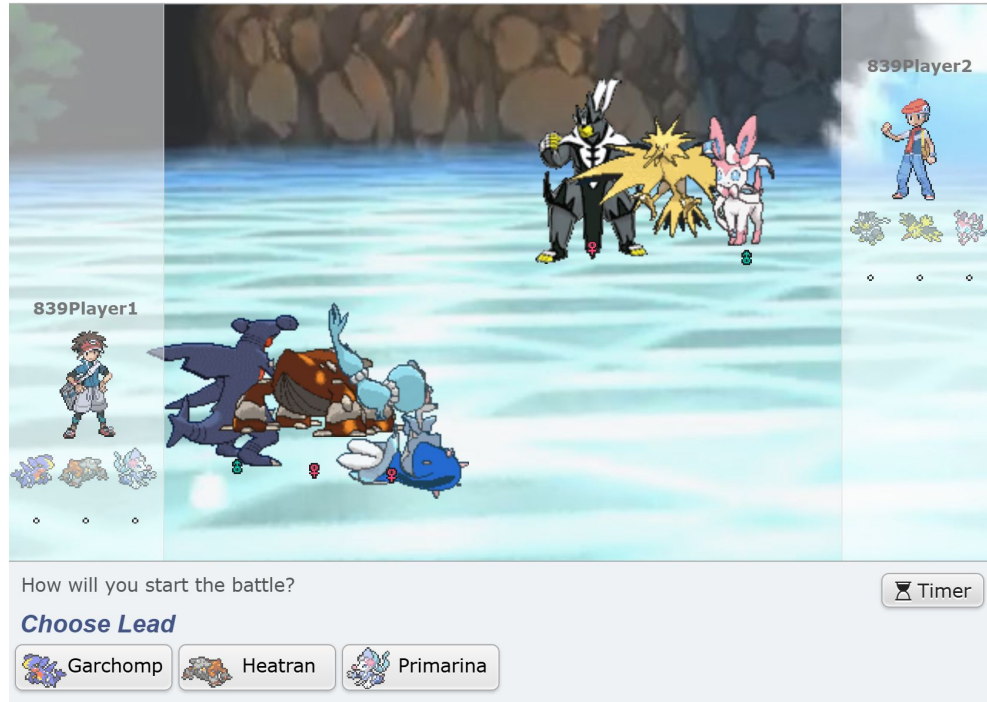


Zapdos @ Life Orb
Ability: Pressure
EVs: 16 HP / 216 SpA / 24 SpD / 252 Spe
Timid Nature
- Hurricane
- Heat Wave
- Thunderbolt
- Protect



Sylveon @ Choice Specs
Ability: Pixilate
EVs: 72 HP / 72 Def / 252 SpA / 4 SpD / 108 Spe
Modest Nature
- Hyper Beam
- Hyper Voice
- Echoed Voice
- Draining Kiss

A 1v1 Pokemon battle consists of two phases, a choosing Pokemon and a battling Pokemon phase.



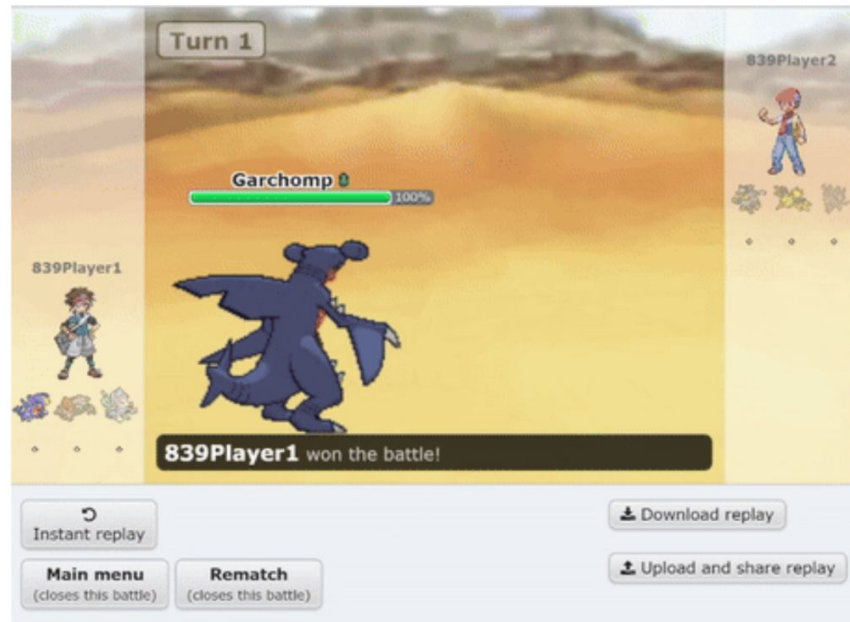
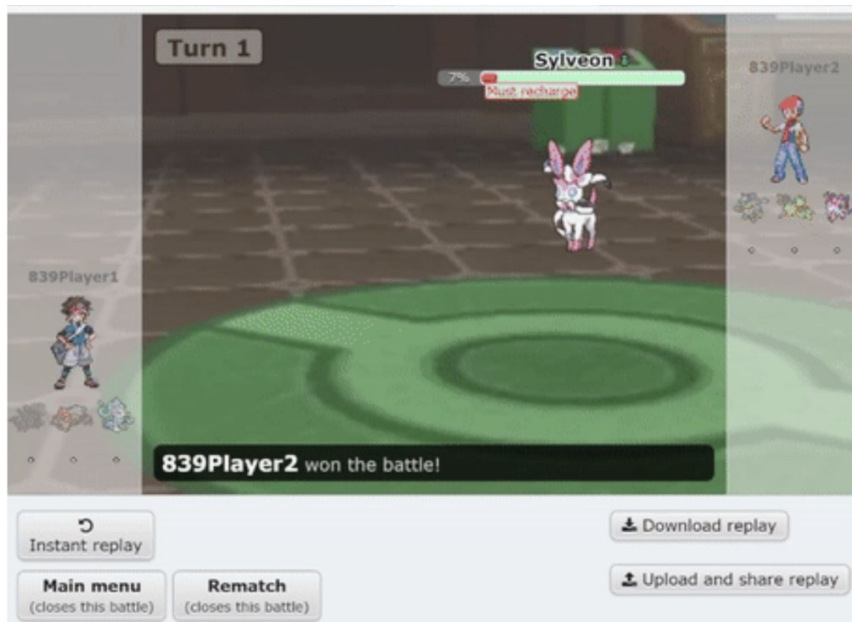
A 1v1 Pokemon battle consists of two phases, a choosing Pokemon and a battling Pokemon phase.



When a player's Pokemon's HP reaches 0, the other player wins



Pokemon has lots of non-determinism (critical hits, accuracy, damage rolls, etc.), causing even identical matchups and moves to produce different outcomes.



Project Goals:

1. Find optimal strategies which account for battle phase randomness
2. Calculate win rates for each Pokémon matchup
3. Use results to model the selection phase and find an equilibria



Which move choices
are optimal throughout
a battle?



Win-rate: X%



Win-rate: (1-X)%

Zapdos

How do Pokémon fare
against one another?



Which Pokémon should
you lead with?



Modeling Battles

For our project, we looked at 7 of the most used Pokemon in the 1/2021 leaderboard, all with their most popular movesets.



Primarina



Sylveon



Heatran



Urshifu



Rillaboom



Garchomp



Zapdos

To model battles, we construct a 4 by 4 payoff matrix representing move choices every turn. For damaging moves, the value is the HP done to the opponent. Setup and defensive moves impact future turn payoff matrices.

Garchomp's Moves (select one to show detailed results)

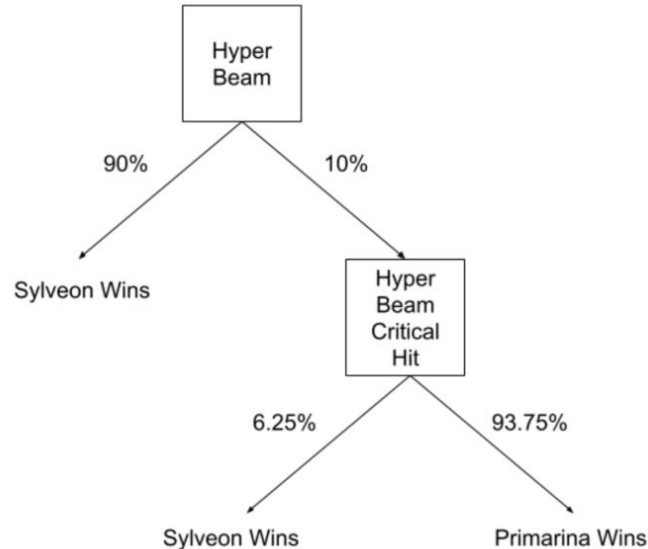
Outrage	0 - 0%
Earthquake	89.6 - 106%
Fire Fang	38.9 - 46.1%
Stone Edge	59.8 - 70.7%

Sylveon's Moves (select one to show detailed results)

Hyper Beam	274.8 - 324%
Hyper Voice	165.9 - 194.9%
Echoed Voice	73.7 - 87.1%
Draining Kiss	77 - 91% (59.3 - 70.2% recovered)

252 Atk Choice Band Garchomp Earthquake vs. 72 HP / 72 Def Sylveon: 313-370 (89.6 - 106%) -- 37.5% chance to OHKO

To account for randomness, we construct a tree of possible game outcomes. Players will then choose moves that maximize their expected win rate across all possible scenarios.



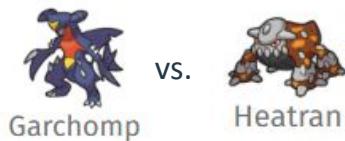
Along with randomness, we also need to account for mind games. A move like Sucker Punch can create complex scenarios where the optimal play fully depends on what move your opponent chooses

Sucker Punch

This move enables the user to attack first.
This move fails if the target is not readying an attack.

POWER: 70 ACC: 100 PP: 5

To account for all of these factors, we developed Python simulations with 1 million iterations to simulate randomness and used RL to determine optimal move choices.



Outrage	36.2 - 42.7%
Earthquake	0 - 0%
Fire Fang	0 - 0%
Stone Edge	40.4 - 47.6%

Outrage has lower damage but higher accuracy while Stone Edge crits more, hard to tell which is better

Simulation Output:

Stone Edge Win Rate: 0.1831, Heatran Win Rate: 0.8169
Outrage Win Rate: 0.1637, Heatran Win Rate: 0.8363
Earthquake Win Rate: 0, Heatran Win Rate: 1
Fire Fang Win Rate: 0, Heatran Win Rate: 1

Stone Edge is the optimal strategy

(Note: Garchomp is holding the Choice Band item meaning it's locked into the first move it uses)

We then ran the simulation with both players' optimal move choices to determine win-rates.

Optimal Strategies



Heatran

Turn 1: Will-O-Wisp
Turns 2+: Dragon Pulse



Garchomp

Turns 1+: Stone Edge



Win Rates



Heatran

81.7%

vs.









Garchomp

18.3%





Example

Rillaboom vs. Urshifu was one of the more complicated battles we had to model. Not only does this matrix have no clear best option for either player, but the chance of Sucker Punch dealing enough damage also varies significantly by damage rolls.

		 Urshifu	
		Sucker Punch DARK 	Wicked Blow DARK 
 Rillaboom	Swords Dance NORMAL 	Draw (This Turn)	Loss, Win
	Grassy Glide GRASS 	Loss, Win	Win, Loss

With our simulation, we were able to determine the optimal move choices for both players given all the randomness, and then reran the simulation to determine each player's win rate.

	<div>Sucker Punch</div> <div></div> <div>Wicked Blow</div>				
<div>Swords Dance</div> <div></div> <div>Grassy Glide</div>	<table> <tr> <td> <div>Draw (This Turn)</div> </td><td> <div>Loss, Win</div> </td></tr> <tr> <td> <div>Loss, Win</div> </td><td> <div>Win, Loss</div> </td></tr> </table>	<div>Draw (This Turn)</div>	<div>Loss, Win</div>	<div>Loss, Win</div>	<div>Win, Loss</div>
<div>Draw (This Turn)</div>	<div>Loss, Win</div>				
<div>Loss, Win</div>	<div>Win, Loss</div>				

Optimal Strategies



Turn 1: Fake Out

Turn 2: Grassy Glide

Turn 3+:

$P(\text{Swords Dance}) = P(\text{Sucker Punch deals enough damage})$

$P(\text{Grassy Glide}) = 1 - P(\text{Swords Dance})$

Win-Rate: **69.5%**



Turn 1: Wicked Blow

Turn 2: Wicked Blow

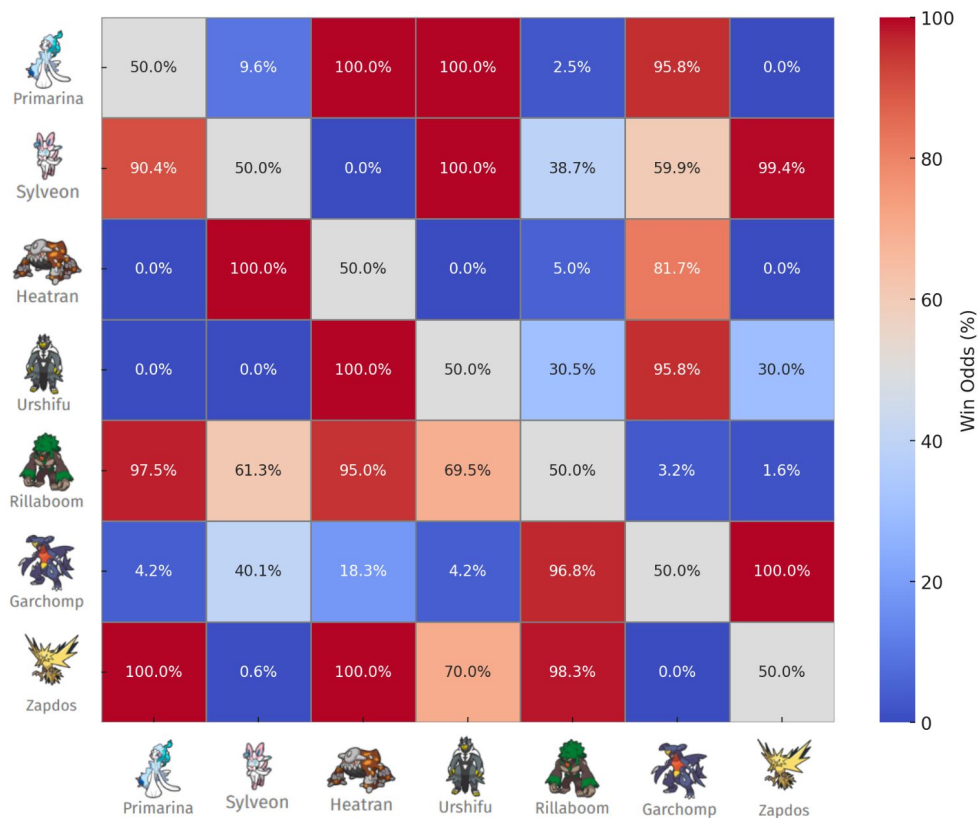
Turns 3+:

$P(\text{Sucker Punch}) = 50\%$

$P(\text{Wicked Blow}) = 50\%$

Win-Rate: **30.5%**

We repeated this process for every possible matchup, leading to the following payoff matrix:

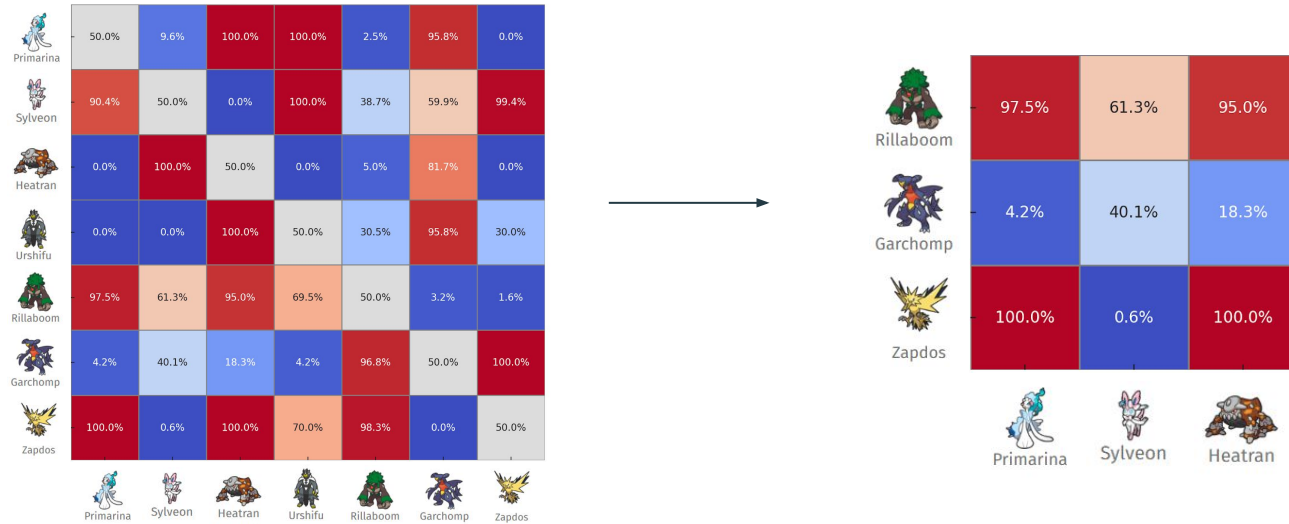










Modeling the Choosing Pokemon Phase









To model the choosing Pokemon phase, we first construct 3x3 subsets of our payoff matrix to represent battles, where players want to choose Pokemon that give them the best possible win rate.



After we have one of these matrices, we compute the Nash equilibrium to derive each players' optimal strategy and their win-rate.

 Rillaboom	97.5%	61.3%	95.0%
 Garchomp	4.2%	40.1%	18.3%
 Zapdos	100.0%	0.6%	100.0%
	 Primarina	 Sylveon	 Heatran









 Rillaboom	97.5%	61.3%	95.0%
 Garchomp	4.2%	40.1%	18.3%
 Zapdos	100.0%	0.6%	100.0%
	 Primarina	 Sylveon	 Heatran

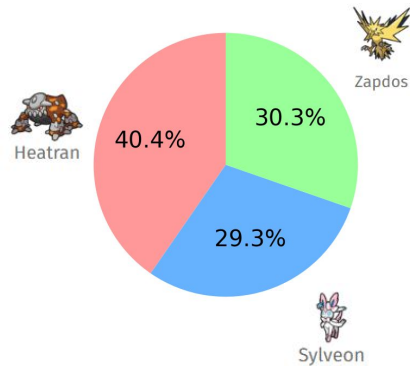
Pure Nash Equilibrium.
Row player chooses Rillaboom,
Column player chooses Sylveon.

Row player has a 61.3% chance of winning.

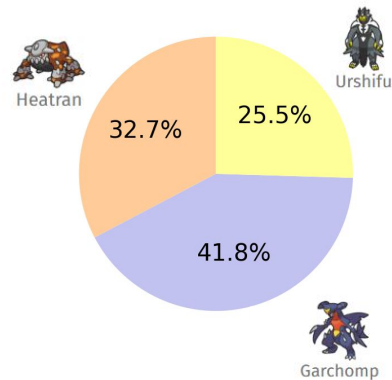
Unlike the last game, some games ended up with Nash equilibria with complex mixed strategies.

	50.0%	81.7%	0.0%
	0.0%	59.9%	100.0%
	100.0%	0.0%	70.0%
			
	Heatran	Garchomp	Urshifu

Mixed Nash Equilibrium.
Row player chooses:



Column player chooses:



Row player has a 50.5% chance of winning.

Findings

After computing the Nash equilibrium for every possible matchup, we discovered the best and worst performing teams for our subset of Pokemon:

Best Performing (62.4% Win Rate):



Worst Performing (29.4% Win Rate):



We were also able to determine the average win rate for teams containing each Pokemon.



Sylveon

- 55.3%



Rillaboom

- 50.39%



Urshifu

- 46.07%



Zapdos

- 52.6%



Primarina

- 47.66%



Garchomp

- 50.55%



Heatran

- 47.44%

Further Work

To better reflect Pokemon's nature as an imperfect information game, a natural extension would be to implement different moveset variations of the same Pokemon.



Garchomp @ Choice Band
Ability: Rough Skin
EVs: 4 HP / 252 Atk / 252 Spe
Jolly Nature
- Outrage
- Earthquake
- Fire Fang
- Stone Edge

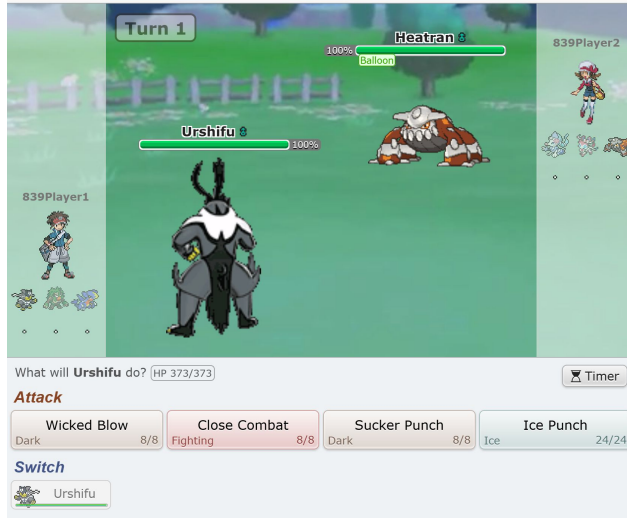
Win-Rate against Air Balloon Heatran: 18.1%



Garchomp @ Life Orb
Ability: Rough Skin
EVs: 4 HP / 252 Atk / 252 Spe
Jolly Nature
- Dragon Claw
- Earthquake
- Fire Fang
- Stone Edge

Win-Rate against Air Balloon Heatran: 100%

Another possible extension is to try to find optimal strategies and Nash equilibria in full Pokemon battles outside of the 1v1 format, with switching included.



4 choices per turn



9 choices per turn



Questions?

PROFESSOR

