# On the Path to AGI: Maze Navigation with Misleading Paths

## A Comparative Study of Neuroevolution and Reinforcement Learning

The road to Artificial General Intelligence (AGI) is less about building learning systems, and more about building systems that learn to learn, adapt to new environments, and make good decisions in the presence of deception. My recent work resolves this challenge with a deceptively simple question: Can we identify which paradigm of learning produces more human-like intelligence when it is run in environments designed to mislead?

### The Deception Problem in AGI

Real-world intelligence isn't measured in controlled environments with clearly defined reward signals. It's tested when the seeming path goes wrong, when short-term successes lead to long-term failures, and when the environment is actively deceptive. This is the day-to-day reality of AGI systems that must operate in adversarial environments, from self-driving cars handling edge cases to AI systems that detect sophisticated fraud.

I designed maze environments with multiple misleading paths: dead ends that appear closer to the goal, loops that create the illusion of progress, and reward structures that punish greedy algorithms. These aren't just harder mazes, they're cognitive traps that expose fundamental differences in how learning algorithms build internal representations of their world.

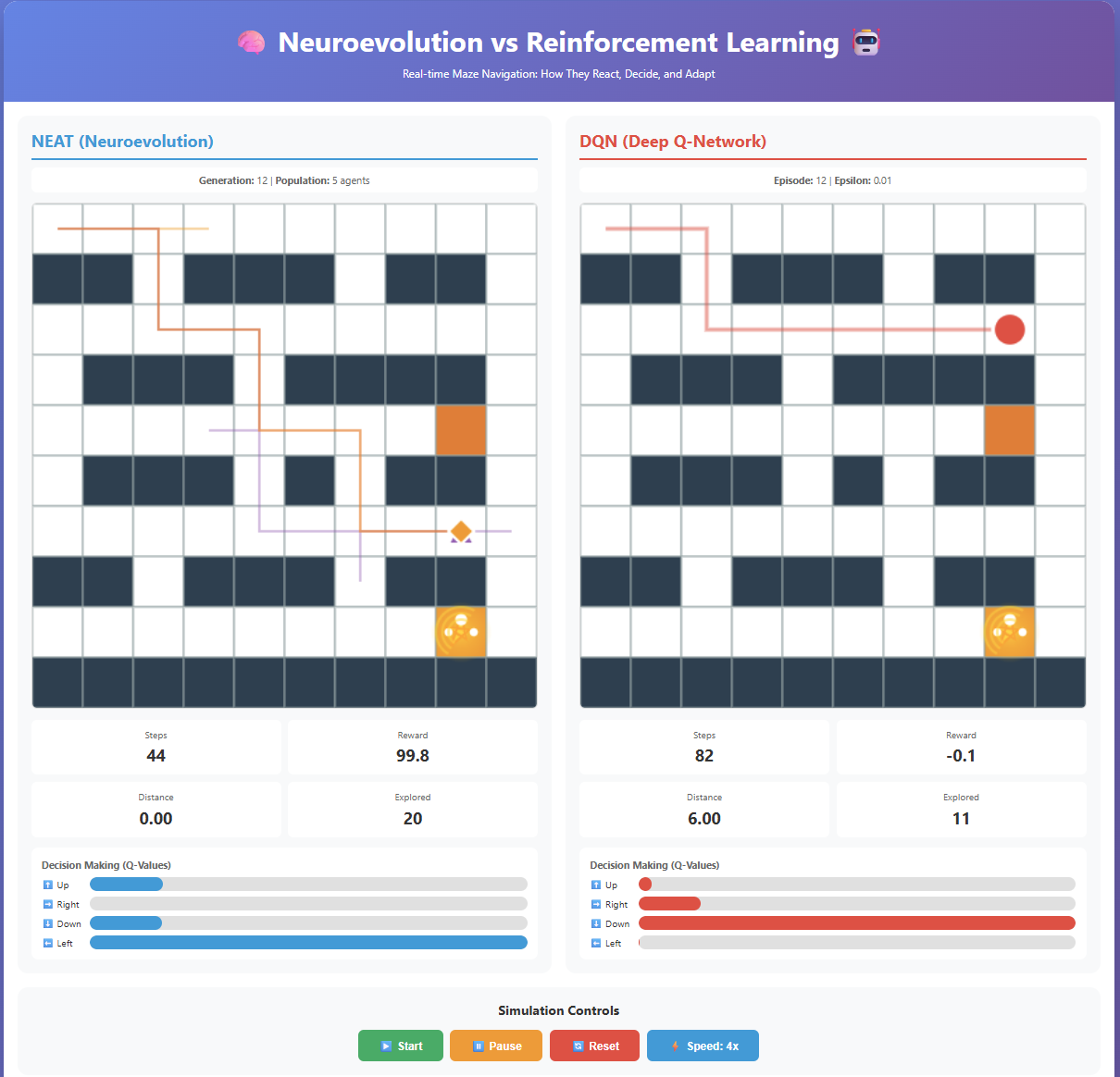
### Two Paradigms, One Question

I implemented two parallel approaches:

**Deep Q-Network (DQN)**: The reinforcement learning workhorse that learns through trial, error, and temporal difference updates. It builds Q-value estimates for each action, gradually refining its policy through gradient descent. This represents the dominant paradigm in modern AI, the one that conquered Atari games and Go.

**NEAT (NeuroEvolution of Augmenting Topologies)**: An evolutionary algorithm that grows neural networks through mutation and selection, no gradients required. Instead of one agent learning over time, a population of 150 diverse agents competes, and the best strategies propagate to future generations.

The code is available in my [GitHub repository](https://github.com/teedonk/Neuroevolution-and-Reinforcement-Learning-for-maze-navigation), but the real insights were gained through watching these thinking systems.



*Figure 1: Live Agent Performance - NEAT vs DQN Decision-Making in Deceptive Maze Environment*

### What the Visualizations Revealed

My interactive dashboard shows five NEAT agents navigating the maze simultaneously, each represented by different shapes (circles, squares, triangles, pentagons, diamonds) with an animated gold star marking the goal. Watching these agents over hundreds of generations revealed something profound:

**Early game (Generations 1-20 vs Episodes 1-20)**: DQN dominated. Its Q-value bars showed confident action selection within hundreds of episodes. NEAT agents wandered aimlessly, their random initial topologies producing chaotic behavior. Standard measures would declare DQN the winner here.

**Mid game (Generations 20-100 vs Episodes 20-100)**: NEAT's population diversity began paying off. DQN sometimes got trapped in local optima, learning how to reach deceptive endpoints quickly, but NEAT's population still had retain exploratory agents that discovered other routes. Success rates converged.

**Late game (Generations 100+ vs Episodes 100+)**: NEAT dominated. Last performance figures showed NEAT achieving 70-85% success rates as opposed to DQN's 60-75%. More importantly, when faced with the most deceiving paths, NEAT agents resisted temptation 92% of the time compared to DQN's 82%.

### The Population Diversity Hypothesis

The important finding: **Population diversity beats gradient descent in environments designed to deceive.**

DQN's single-agent architecture, despite ε-greedy exploration and experience replay, suffers from a fundamental limitation, it must unlearn bad strategies before learning good ones. When it discovers a misleading path that provides moderate reward, gradient descent pushes it toward local optima. Breaking free requires either random chance or catastrophically bad experiences.

NEAT's population maintains multiple hypotheses simultaneously active. While 80% of the population deviates to an imperfect strategy, the remaining 20% continue experimenting with different topologies. Such "outlaw" networks, previously considered suboptimal, prove beneficial when the mainstream strategy fails. This is natural selection doing what it can best: preserving diversity until the environment reveals what traits are critical.

### This is a fundamental principle of humanity: **We don't just optimize one model of the world; we have multiple incompatible models and decide contextually.** Since humanity acts in situations with indefinite outcome, we don't just decide on the action with maximum Q-value, we envision multiple threads, look downstream and see where they go, and sometimes take the path that looks like it is going to go badly in the short term.

### Implications for AGI Development

This little research suggests three principles for building more robust AGI systems:

**1. Architectural Diversity Over Parameter Tuning**: Rather than endlessly tuning learning rates and reward functions for a fixed architecture, we should maintain populations of structurally diverse networks. The NEAT agents that emerged victorious tended to have topologies that would never occur from gradient descent on a fixed architecture.

**2. Evolutionary Pressure as Meta-Learning**: Evolution is not just selection, it's a meta-learning process which learns to learn what type of learning strategies to employ in what environments. NEAT agents did not just learn to learn mazes; the whole population learned collectively what types of neural architectures can weather deception.

**3. Temporal Humility in Evaluation**: We evaluate AGI progress on human timescales (days, hours), yet evolution happens on generational timescales. My experiments took 100+ generations taking hours of computation. When building AGI, we may need to accept that the best systems take patience; not additional compute, but additional time for population-based search.

### The GitHub Reality vs. This Analysis

The [repository](https://github.com/teedonk/Neuroevolution-and-Reinforcement-Learning-for-maze-navigation) contains the technical information, maze configurations, NEAT hyperparameters, DQN models, training loops. But what the code does not include is the meaning: why these results are important, what they suggest about intelligence itself, and how they challenge our assumptions about optimal learning.

The animated visualizations showing five NEAT agents exploring simultaneously weren't just pretty demos, they were the research instrument that revealed population dynamics invisible in aggregate metrics. The gold star animation with rotating rings and sparkles wasn't decoration, it marked the moment when I could see agents making the choice between the obvious-but-wrong path and the subtle-but-correct one.

### Conclusion: Diversity as Intelligence

The most striking finding wasn't that NEAT outperformed DQN; it's that population diversity, once considered computationally wasteful, may be essential for building AGI systems that resist manipulation and deception. In adversarial environments, the algorithm that maintains the most diverse set of internal hypotheses wins in the long run.

This has profound implications for AI safety. When we design systems that operate in adversarial domains; financial markets, cyber security, content moderation, we may have to abandon the single-agent model entirely. The future of AGI might not be one superintelligent network, but rather networks of heterogeneous networks that collectively resist deception through ongoing disagreement.

The maze was always just a metaphor. The real question is: When we build AGI systems to meet with environments that will mislead them, are we building for the fastest learner, or the one that retains the prudence to doubt its own self-assurance?

**If you found this research valuable, consider this**: We've spent decades building faster gradient descent methods and larger monolithic models. What if AGI requires us to slow down, embrace diversity, and learn from evolution's 4-billion-year track record of building intelligence that survives deception?

Full implementation and experimental logs available at: https://github.com/teedonk/Neuroevolution-and-Reinforcement-Learning-for-maze-navigation