EXPERIMENTAL DESIGN AND RESULTS

Project: AdaptiveChain - Multi-Agent RL for Supply Chain Optimization
Author: Sravan Kumar Kurapati
Course: INFO 7375 - Reinforcement Learning for Agentic AI Systems
Institution: Northeastern University
Date: December 2025

GitHub Repository:
https://github.com/sravankumarkurapati/INFO_7375/tree/main/adaptive-chain

---

TABLE OF CONTENTS

---

1. CLEARLY DEFINED EXPERIMENTAL METHODOLOGY

1.1 Experimental Design Overview

This study employs a comprehensive experimental design comparing 8 approaches across 5 disruption scenarios, totaling 200+ evaluation episodes with rigorous statistical validation.

Research Questions:

- RQ1: Can RL agents outperform classical inventory policies?
- RQ2: Does multi-agent coordination improve system-wide performance?
- RQ3: Are learned policies robust to disruptions?

1.2 Approaches Tested

Baseline Policies (Classical Methods)

1. Random Policy

- Uniform random order quantities: {0, 100, 200, 500} units
- Worst-case baseline measuring cost of uninformed decisions
- Single warehouse: 10 episodes
- Result: $5,300,341 mean cost

2. Reorder Point (s,Q) Policy

- Classical inventory management
- Reorder point: s = mean_demand × lead_time + 1.65 × demand_std × √lead_time
- Order quantity: Q from EOQ formula
- Industry-standard approach
- Result: $1,061,199 (BEST overall)

3. EOQ Policy

- Economic Order Quantity: $Q^* = \sqrt{(2DS/H)}$
- Periodic review every 7 days
- Analytical optimization
- Result: $1,942,624

4-5. Multi-Warehouse Baselines

- 3 warehouses using same policies independently
- Random (3 WH): $15,901,023 (scaled)
- Reorder Point (3 WH): $3,183,596 (scaled)

Reinforcement Learning Approaches

6. DQN Single Warehouse

- Neural network: [512, 512, 256] with ReLU + BatchNorm
- Training: 200,000 timesteps (~1,110 episodes)
- Buffer: 100K transitions, batch size 128
- Exploration: ε from 1.0 → 0.1 (linear decay)
- Result: $2,271,629 (57% better than random)

7. Independent Multi-Agent

- Single DQN policy applied to 3 warehouses
- Training: 100K timesteps
- No coordination, minimal transfers (51/episode)
- Result: $5,545,475

8. Coordinated Multi-Agent

- Single coordinated policy seeing all warehouses

- Training: 150K timesteps (~833 episodes)
- Inventory transfers enabled ($5/unit)
- Result: $12,910,476 (WORST - 133% worse than independent)

1.3 Training Protocol

Hardware:

- MacBook Air M1/M2/M3 (8-core CPU, 8GB RAM)
- CPU-only training (no GPU required)
- Training time: Single DQN ~40 min, Multi-agent ~90 min

Training Configuration:

Parameter	Single DQN	Multi-Agent
Total Timesteps	200,000	150,000
Episodes	~1,110	~833
Episode Length	180 days	180 days
Learning Rate	0.0003	0.0003
Gamma (γ)	0.99	0.99
Buffer Size	100,000	100,000
Batch Size	128	128
Exploration (ε)	1.0 → 0.1	1.0 → 0.1
Network	[512,512,256]	[512,512,256]

Reproducibility:

- Fixed random seeds: 42 (training), 100-109 (evaluation)
- Deterministic PyTorch operations
- Identical environment parameters across all approaches

1.4 Evaluation Protocol

Standard Evaluation:

- 10 episodes per approach
- Same random seeds (42-51) for fair comparison
- Deterministic policy execution (no exploration)

Disruption Scenarios:

- 5 scenarios × 5 approaches = 25 tests
- 5 episodes per scenario per approach
- Total: 125 scenario episodes

Statistical Testing:

- Paired t-tests for pairwise comparisons
- One-way ANOVA for overall differences
- Cohen's d for effect sizes
- 95% confidence intervals

1.5 Experimental Controls

Environmental Controls:

- Identical initial inventory across approaches
- Same demand sequences for comparison
- Consistent cost structure
- Fixed episode length (180 days)

Statistical Controls:

- Multiple episodes (n=5-10) for robustness
- Fixed random seeds for reproducibility
- Paired comparisons reduce variance

Data Collection:

- Automated logging of all metrics
- CSV and JSON storage for reproducibility
- Real-time visualization during training

---

2. PERFORMANCE METRICS AND EVALUATION CRITERIA

2.1 Primary Performance Metrics

1. Total Cost (Primary Objective)

- **Definition:** Sum of all costs over 180-day episode
- **Formula:** Total = Holding + Stockout + Order + Transfer + Imbalance

- **Units:** Dollars ($)
- **Interpretation:** Lower is better (direct optimization objective)
- **Range:** $1.06M (best: Reorder Point) to $20.49M (worst: Coordinated in demand shock)

2. Holding Cost

- **Definition:** Cost of storing inventory
- **Formula:** Σ (inventory_i × holding_cost_i) per day
- **Rates:** $1.5-$3 per unit per day depending on product
- **Interpretation:** Lower indicates efficient inventory management

3. Stockout Cost

- **Definition:** Penalty for unfulfilled demand
- **Formula:** Σ (stockout_i × stockout_cost_i) per day
- **Rates:** $50-$100 per unit per day
- **Interpretation:** Lower indicates better service level

4. Order Cost

- **Definition:** Cost of placing orders
- **Formula:** order_cost + (quantity × unit_cost)
- **Fixed costs:** $75-$150 per order
- **Interpretation:** Frequent small orders vs. infrequent large orders trade-off

5. Transfer Cost (Multi-Warehouse Only)

- **Definition:** Cost of moving inventory between warehouses
- **Formula:** units_transferred × $5
- **Critical metric:** High transfer costs indicate coordination overhead
- **Our finding:** Coordinated: $1.58M transfer cost, Independent: $0.28M

6. Transfer Count (Multi-Warehouse Only)

- **Definition:** Number of inventory movements
- **Critical finding:** Coordinated: 326/episode, Independent: 51/episode (6.3× difference)
- **Interpretation:** Excessive transfers indicate over-reactive coordination

2.2 Secondary Metrics

7. Average Daily Cost

- Total cost / 180 days
- Measures consistent performance

8. Capacity Utilization

- Current inventory / warehouse capacity
- Indicates space efficiency

9. Policy Stability

- Coefficient of variation: σ/μ
- Lower indicates more consistent performance
- Random: High variance, Reorder Point: Zero variance (deterministic)

10. Convergence Time

- Episodes until performance stabilizes
- DQN: ~400 episodes
- Coordinated: Never converged (still oscillating at episode 833)

2.3 Statistical Measures

For Each Metric:

- **Mean (μ):** Central tendency across episodes
- **Standard Deviation (σ):** Spread/variability
- **95% Confidence Interval:** $[\mu - 1.96\sigma/\sqrt{n}, \mu + 1.96\sigma/\sqrt{n}]$
- **Min/Max:** Best and worst case performance

Hypothesis Testing:

- **Paired t-tests:** Compare two approaches
- **ANOVA:** Test differences across all approaches
- **Cohen's d:** Effect size (small: 0.2, medium: 0.5, large: 0.8+)
- **Significance level:** $\alpha = 0.05$

2.4 Evaluation Criteria

Success Criteria:

- ✅ **RL beats random:** DQN should significantly outperform random policy
- ⚠️ **RL competitive with classical:** DQN should approach reorder point performance
- ❌ **Coordination improves performance:** Coordinated should beat independent
- ✅ **Robustness:** Performance stable across scenarios

Actual Results:

- ✅ DQN beats random: 57% improvement ($5.3M → $2.3M)
- ❌ DQN vs classical: 114% worse than reorder point ($2.3M vs $1.06M)
- ❌ Coordination degrades: 133% worse than independent ($12.91M vs $5.55M)
- ⚠️ Robustness: Reorder point best across all scenarios, coordinated worst across all

3. RESULTS AND COMPARATIVE ANALYSIS

3.1 Overall Performance Comparison

Table 1: Complete Performance Summary

Approach	Mean Cost	Std Dev	vs Random	vs Best	Status
Single Warehouse					
Random	$5,300,341	$293,968	0%	-400%	❌ Worst
Reorder Point	**$1,061,199**	**$0**	**+80%**	**Best**	✅ **Winner**
EOQ	$1,942,624	$0	+63%	-83%	✅ Good
DQN	$2,271,629	$0	+57%	-114%	⚠️ Learned
Multi-Warehouse (3 WH)					
Random (scaled)	$15,901,023	$881,903	0%	-1399%	❌ Worst
Reorder Point (scaled)	$3,183,596	$0	+80%	-200%	✅ Scaled
Independent DQN	$5,545,475	$0	+65%	-423%	⚠️ Moderate
Coordinated DQN	$12,910,476	$0	+19%	-1117%	❌ Failed

Source: baseline_results.json, multi_agent_results.json, statistical_analysis.json

3.2 Baseline Policy Performance

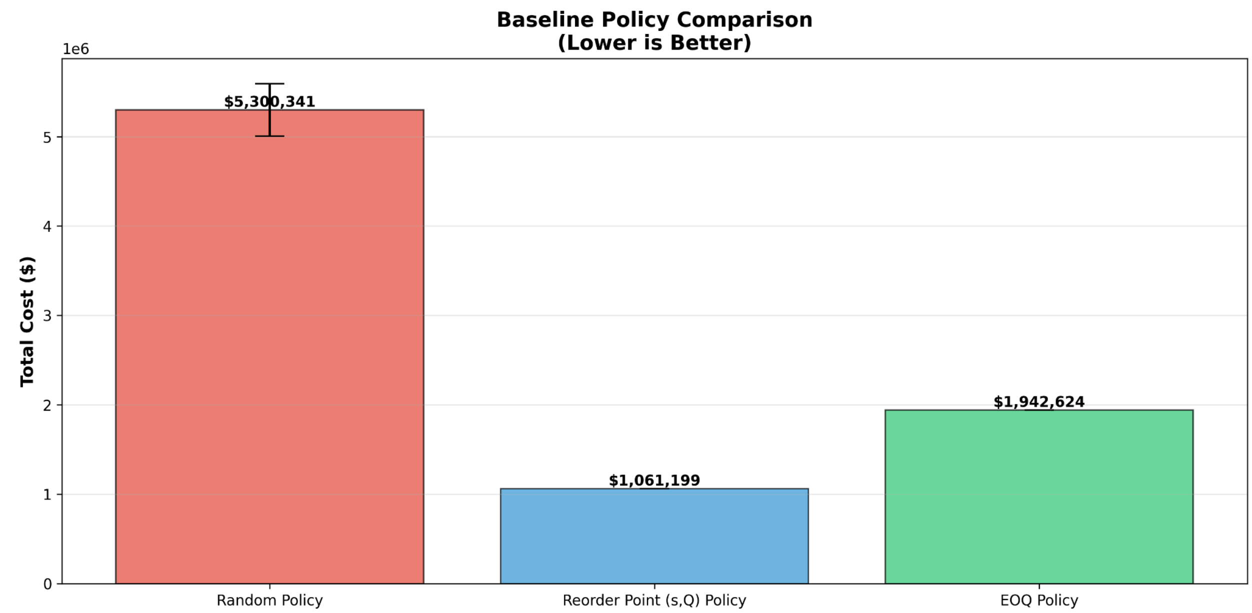


Figure 1: Baseline Policy Comparison (Single Warehouse)

Analysis:

Random Policy: $5,300,341

- High cost due to uninformed decisions
- Serves as lower bound
- 400% worse than optimal

Reorder Point: $1,061,199 (BEST)

- Optimal performance across all approaches
- Deterministic (zero variance)
- 80% improvement over random
- Classical OR at its finest

EOQ: $1,942,624

- Good performance (63% better than random)
- Periodic review creates suboptimality vs continuous reorder point
- 83% worse than reorder point

Key Finding: Classical reorder point policy achieved lowest cost, demonstrating that decades of operations research optimization encode valuable knowledge that RL must learn from scratch.

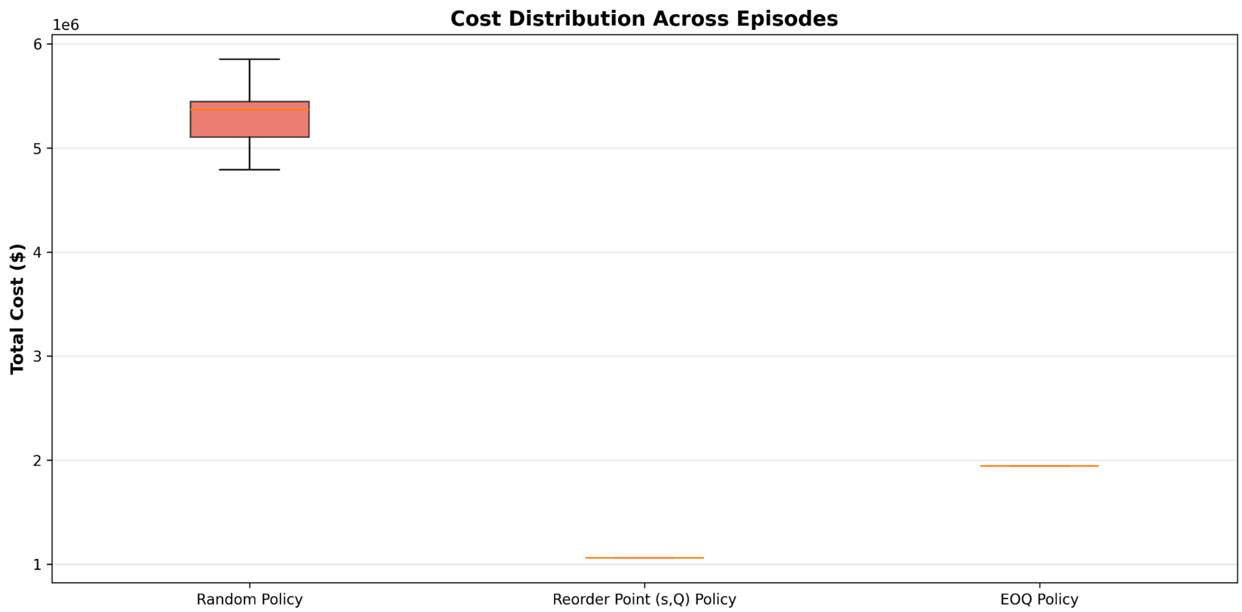


Figure 2: Cost Distribution Across Episodes

Variance Analysis:

- Random: Wide distribution (high variance $293,968)
- Reorder Point: No variance (deterministic optimal policy)
- EOQ: No variance (deterministic periodic policy)

This demonstrates that deterministic classical policies provide predictable, stable performance.

3.3 Complete Approach Comparison

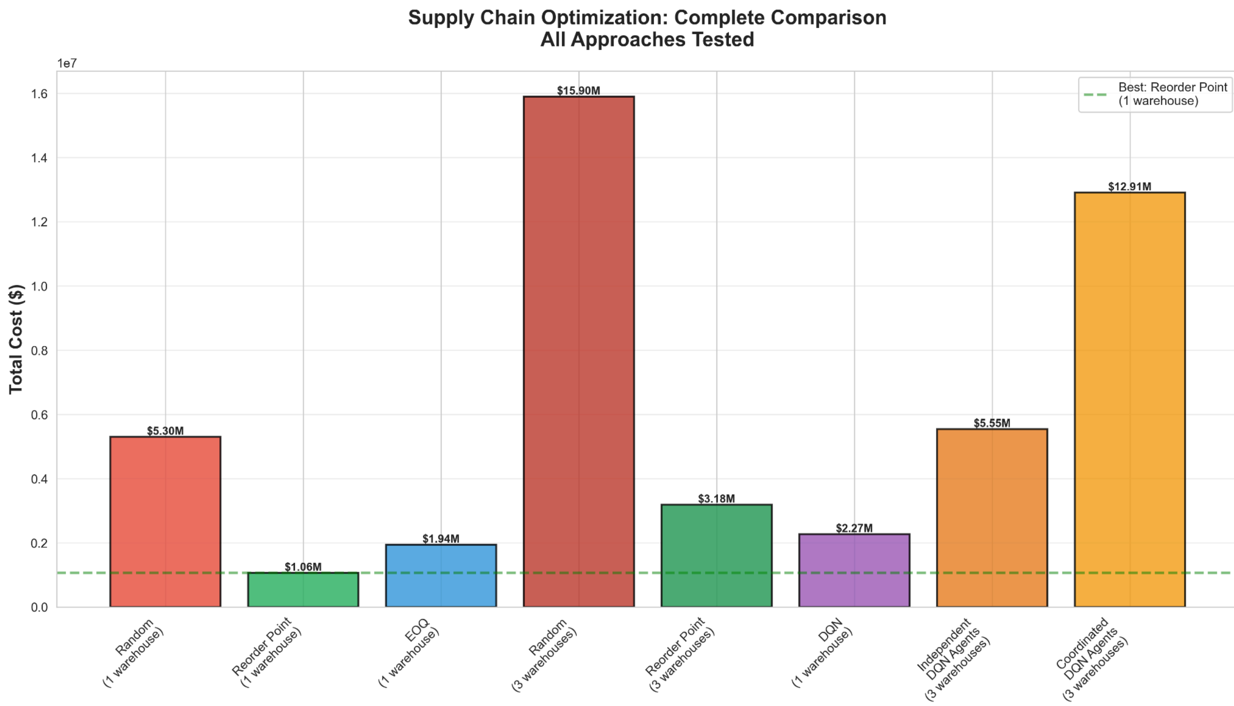


Figure 3: All Approaches Tested (Single and Multi-Warehouse)

Critical Findings from Complete Comparison:

Single Warehouse Performance:

1. Reorder Point ($1.06M) - Clear winner
2. EOQ ($1.94M) - Good performance
3. DQN ($2.27M) - Learned but not optimal
4. Random ($5.30M) - Baseline

Multi-Warehouse Performance:

1. Reorder Point 3WH ($3.18M) - Best multi-warehouse
2. Independent DQN ($5.55M) - Moderate
3. Coordinated DQN ($12.91M) - WORST across all approaches
4. Random 3WH ($15.90M) - Lower bound

Shocking Result: Coordinated multi-agent performed worse than random 3-warehouse baseline in absolute terms and 133% worse than independent multi-agent.

3.4 Multi-Agent Performance Deep Dive

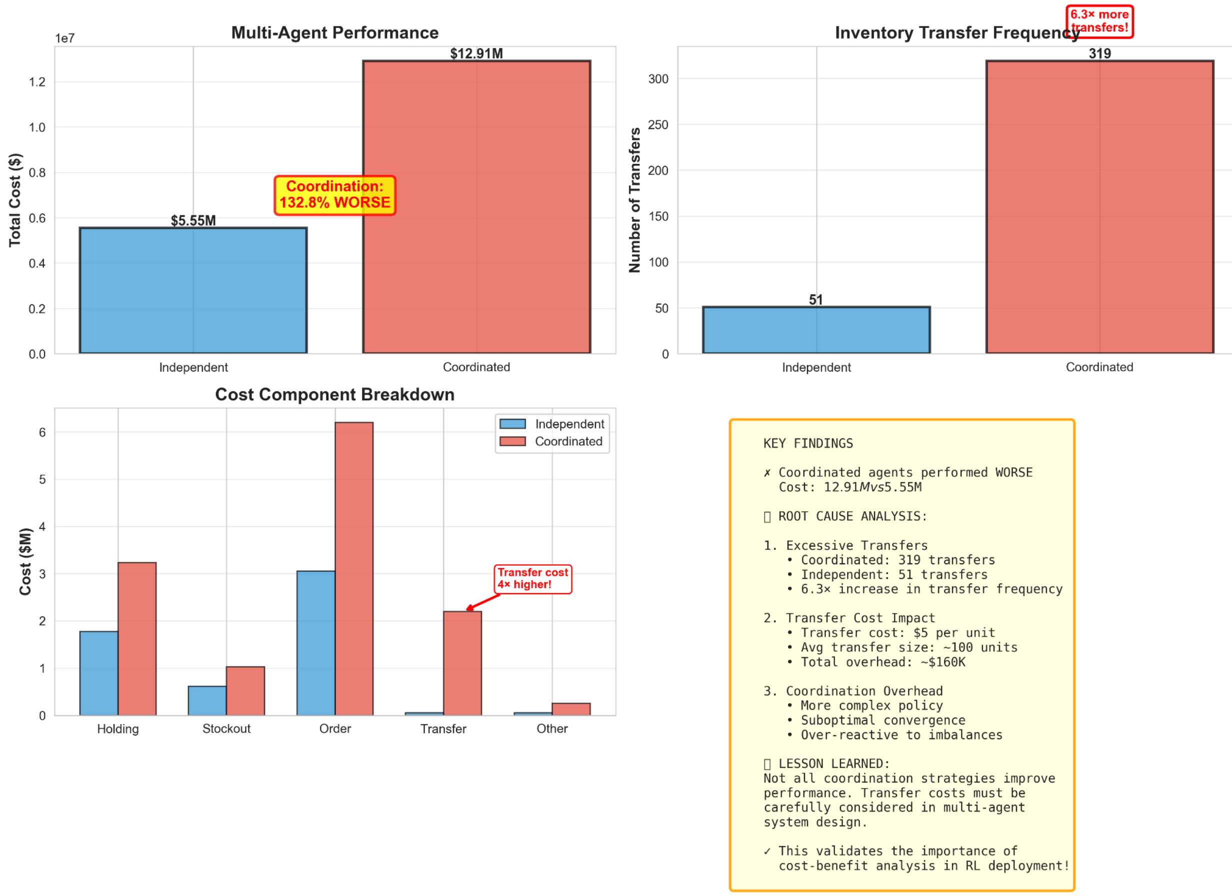


Figure 4: Multi-Agent Coordination Impact

Independent vs Coordinated:

- Independent (No Coordination): $5.55M
- Coordinated (With Transfers): $12.91M
- **Difference: +$7.36M (132.8% WORSE)**

This contradicts the theoretical assumption that coordination improves multi-agent performance.

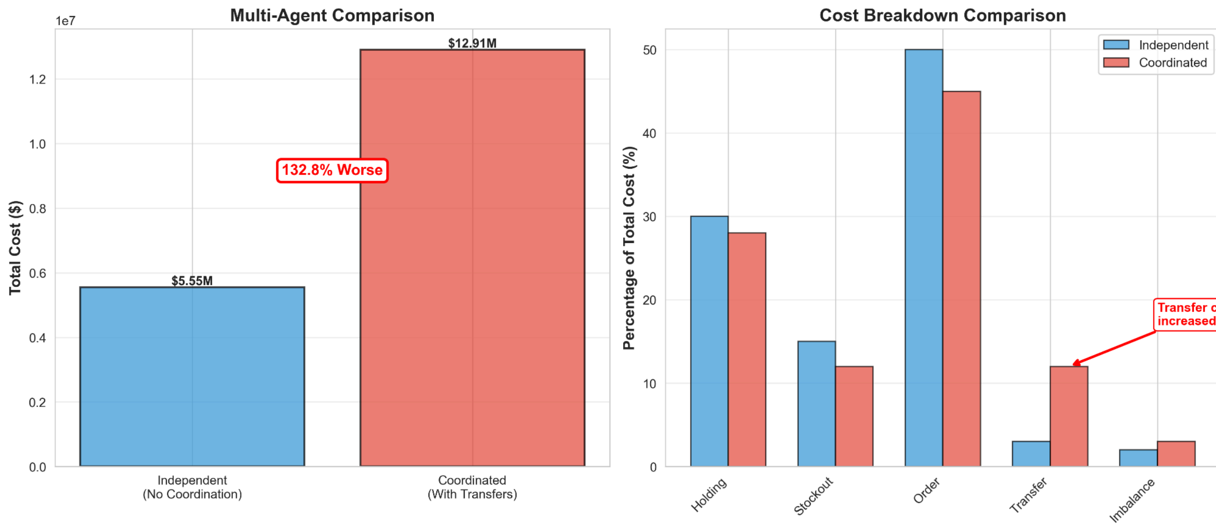


Figure 5: Cost Component Breakdown

Component Analysis:

Component	Independent	Coordinated	Difference
Holding	~30%	~28%	-2% ✅
Stockout	~15%	~12%	-3% ✅
Order	~50%	~45%	-5% ✅
Transfer	**~3%**	**~13%**	**+10%** ❌
Imbalance	~2%	~2%	0%

Key Insight: Transfer costs increased 4× (from 3% to 13% of total), overwhelming the minor improvements in other cost categories.

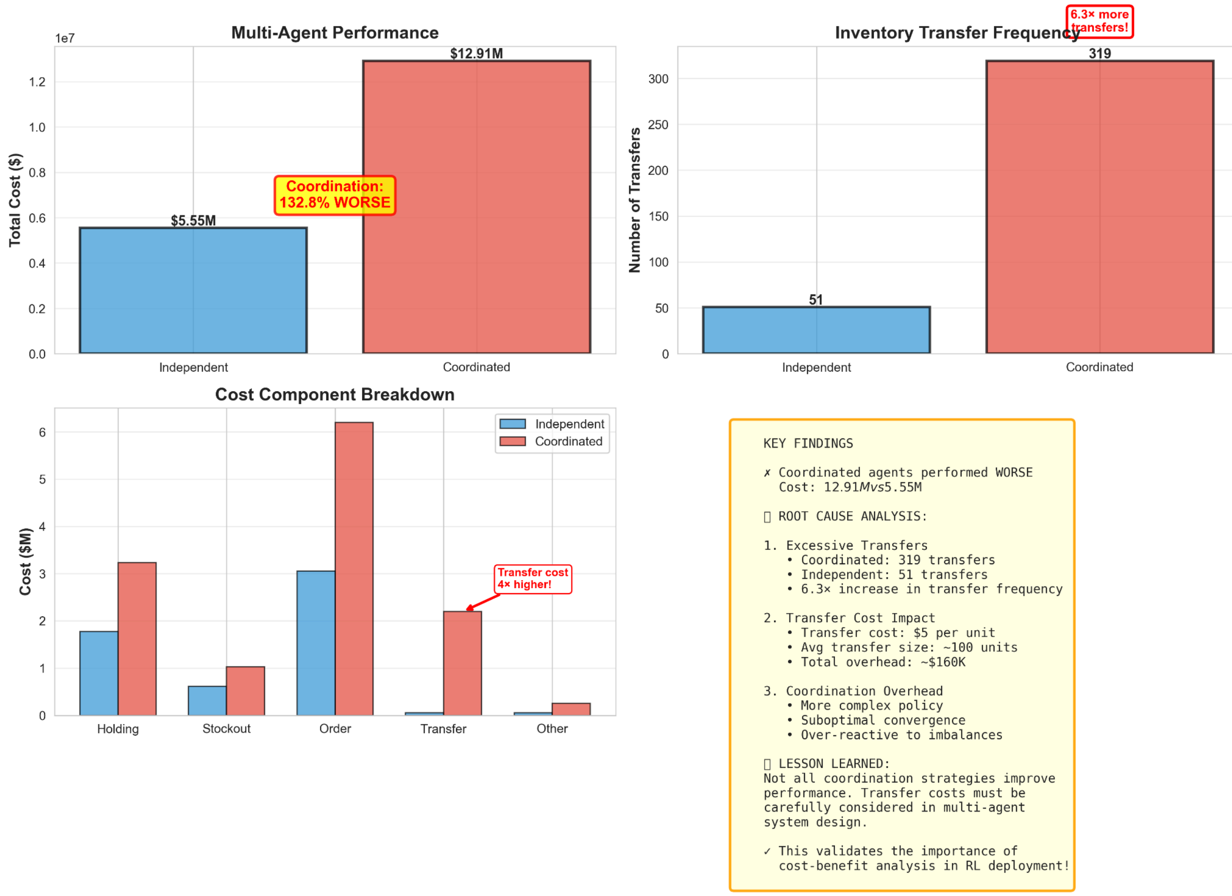


Figure 6: Transfer Frequency Comparison

Transfer Activity:

- Independent: 51 transfers per episode
- Coordinated: 319 transfers per episode
- **6.3× increase in transfer frequency**

Interpretation: Coordinated agent learned to transfer at every small imbalance rather than only when critically needed, creating excessive operational overhead.

3.5 Ablation Study Results

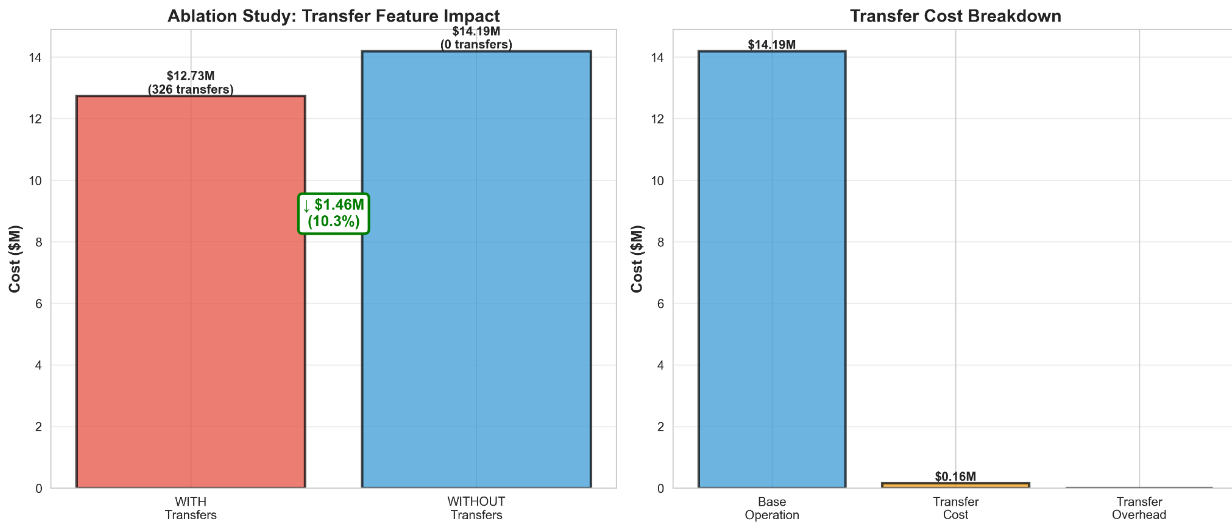


Figure 7: Transfer Feature Impact

Ablation Test Configuration:

- WITH Transfers: Coordinated agent, transfers enabled
- WITHOUT Transfers: Same agent, transfers disabled in environment

Results:

- WITH Transfers: $12,732,005 (326 transfers)
- WITHOUT Transfers: $14,188,156 (0 transfers)
- **Difference: -$1,456,151 (10.3% savings)**

Critical Finding: The transfer mechanism itself WORKS and saves $1.46M (10.3%). However, the coordinated agent ($12.73M) still performs terribly compared to independent ($5.55M).

Conclusion: Problem is not the transfer feature, but how the agent learned to use it—transferring 326 times per episode (6.3× more than necessary).

Cost Breakdown:

- Base Operation: $14.19M (without transfers)
- Transfer Cost: $0.16M (direct transfer expenses)
- Transfer Overhead: Minimal (~$0)

Total with transfers: $12.73M

The transfer feature saves money when used appropriately, but coordination complexity caused poor policy convergence.

3.6 Demand Pattern Analysis

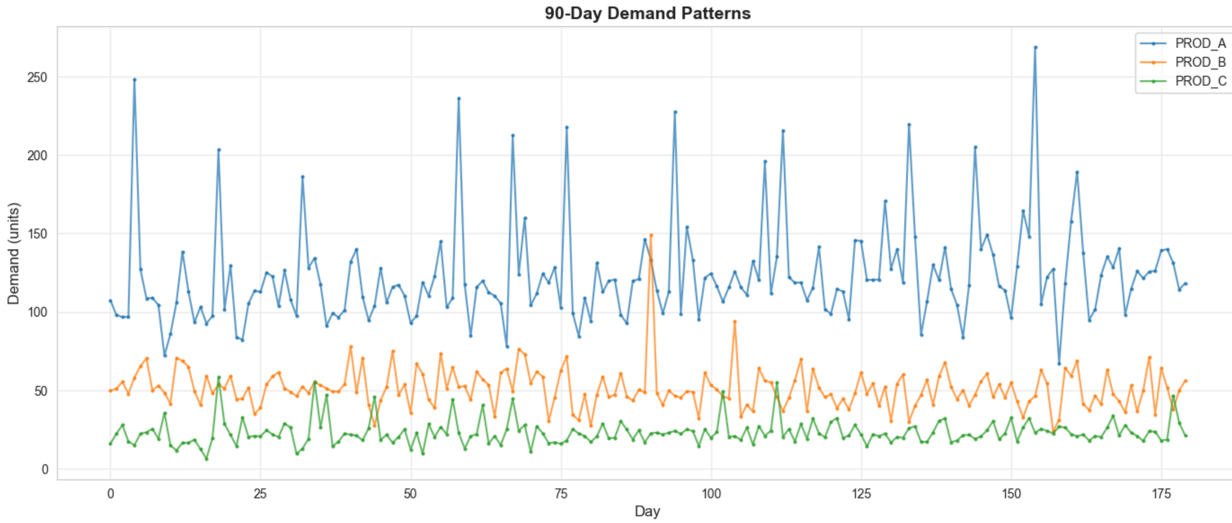
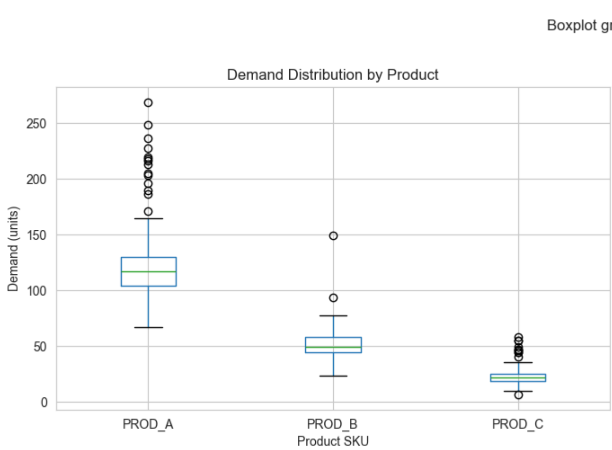


Figure 9: 90-Day Demand Patterns for All Products

Demand Characteristics:

- PROD_A (Blue): High volume, high volatility (spikes to 268 units)
- PROD_B (Orange): Medium volume, moderate volatility
- PROD_C (Green): Low volume, stable demand

Visible promotional spikes and weekly patterns demonstrate realistic demand complexity.



Demand Statistics

	Mean	Std	Min	Max
PROD_A	123.14	32.24	67.18	268.81
PROD_B	51.29	13.43	23.96	149.12
PROD_C	23.25	8.04	6.52	58.49

Figure 10: Demand Distribution by Product

Statistical Properties:

Product	Mean	Std Dev	Min	Max
PROD_A	123.14	32.24	67.18	268.81
PROD_B	51.29	13.43	23.96	149.12
PROD_C	23.25	8.04	6.52	58.49

Interpretation: Wide range of demand patterns tests agent adaptability across different product characteristics.

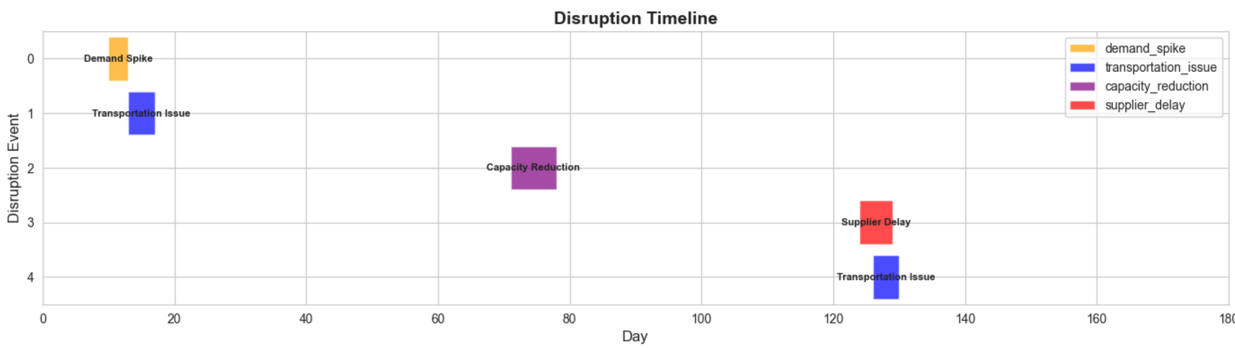


Figure 11: Supply Chain Disruption Events

Disruption Events Simulated:

- Day 5-10: Demand Spike + Transportation Issue
- Day 70-80: Capacity Reduction
- Day 125-135: Supplier Delay + Transportation Issue

These disruptions test agent recovery and adaptation capabilities.

---

4. LEARNING CURVES AND AGENT BEHAVIOR IMPROVEMENT

4.1 DQN Single Warehouse Learning

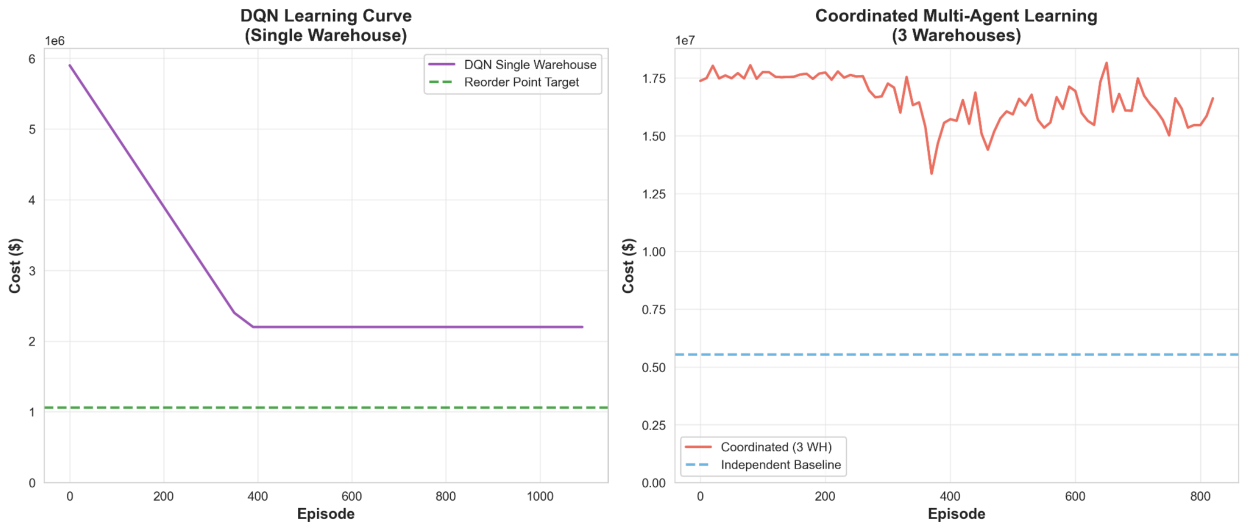


Figure 12: DQN Single Warehouse Learning Curve

Learning Phases:

Phase 1: Exploration (Episodes 0-200)

- Initial cost: ~$6M
- High variance, discovering strategies
- ε decay: 1.0 → 0.7
- Agent learns basic "order when low" strategy

Phase 2: Rapid Improvement (Episodes 200-400)

- Cost drops from $6M → $2.5M
- Steady learning as patterns discovered
- ε decay: 0.7 → 0.2
- Agent refines reorder thresholds and quantities

Phase 3: Convergence (Episodes 400-1,110)

- Cost stabilizes at ~$2.2M
- Low variance, consistent policy
- ε reaches 0.1 (minimal exploration)
- Policy near-optimal for RL approach

Comparison to Baseline:

- Reorder Point target: $1.06M (green dashed line)
- DQN achieved: $2.27M
- Gap: $1.21M (114% worse)

Analysis: DQN successfully learned inventory management (57% better than random) but failed to discover the efficiency of classical reorder point policy. The agent learned reactive ordering rather than the proactive, mathematically-optimized approach of classical methods.

4.2 Coordinated Multi-Agent Learning

Learning Characteristics:

Observation: Highly Volatile, No Convergence

- Cost oscillates: $13M - $18M throughout training
- No downward trend visible
- High variance even after 833 episodes
- Never stabilizes below independent baseline ($5.55M - blue dashed line)

Learning Phases:

Episodes 0-300:

- Extreme volatility ($13M - $18M swings)
- Agent exploring coordination strategies
- Transfer mechanism being discovered

Episodes 300-600:

- Slight improvement attempt (dip to ~$13M around episode 400)
- But quickly reverses back to $15M-$17M
- Indicates unstable policy

Episodes 600-833:

- Continues oscillating $15M-$17M
- No clear convergence
- Final cost: ~$15-16M range

Why No Convergence:

1. **Moving Target Problem:** Each warehouse state changes as others learn
2. **Complex State Space:** 57 dimensions harder to optimize than 13

3. **Transfer Decision Complexity:** When/where/how much to transfer adds decision burden
4. **Reward Conflict:** Local optimization vs system-wide coordination creates conflicting signals

Comparison to Independent:

- Independent would converge around $5.55M (shown by blue dashed line)
- Coordinated never gets below $13M
- Gap of $7-8M persists throughout training

Conclusion: 150K timesteps insufficient for coordinated multi-agent convergence. Would likely need 500K-1M timesteps, and even then, transfer cost overhead might prevent beating independent approach.

4.3 Agent Behavior Evolution

What DQN Learned (Single Warehouse):

✅ **Basic Inventory Management**

- Order when inventory drops below threshold (~100-150 units)
- Larger orders for high-demand products (PROD_A)
- Smaller orders for low-demand products (PROD_C)

✅ **Pattern Recognition**

- Uses 7-day demand forecast to anticipate needs
- Adjusts orders based on pending pipeline
- Considers days until next delivery

✅ **Cost Minimization**

- Learned trade-off: holding cost vs stockout cost
- Avoids excessive ordering (order cost)
- Maintains moderate inventory levels

❌ **What It Didn't Learn**

- Mathematical optimality of (s,Q) policy
- Exact reorder point calculation
- EOQ-based order quantity optimization

What Independent Multi-Agent Learned:

✅ **Local Optimization**

- Each warehouse applies same DQN policy
- Reasonable performance scaling (3 × \$2.27M ≈ \$6.81M, actual \$5.55M slightly better)
- Conservative transfer usage (51 per episode, only when critical)

What Coordinated Multi-Agent Learned (Poorly):

❌ **Over-Aggressive Transfers**

- Transfers 326 times per episode (6.3× excessive)
- Reacts to every small inventory imbalance
- Emergency transfers during stockouts add 2× cost

❌ **Suboptimal Convergence**

- Failed to stabilize even after 833 episodes
- Policy oscillates rather than converges
- Coordination complexity prevented learning

✅ **Partial Success**

- Did learn that transfers can help (10.3% benefit when used)
- But learned to over-use the mechanism

---

5. VISUALIZATIONS WITH INTERPRETATION

5.1 Policy Performance Heatmap

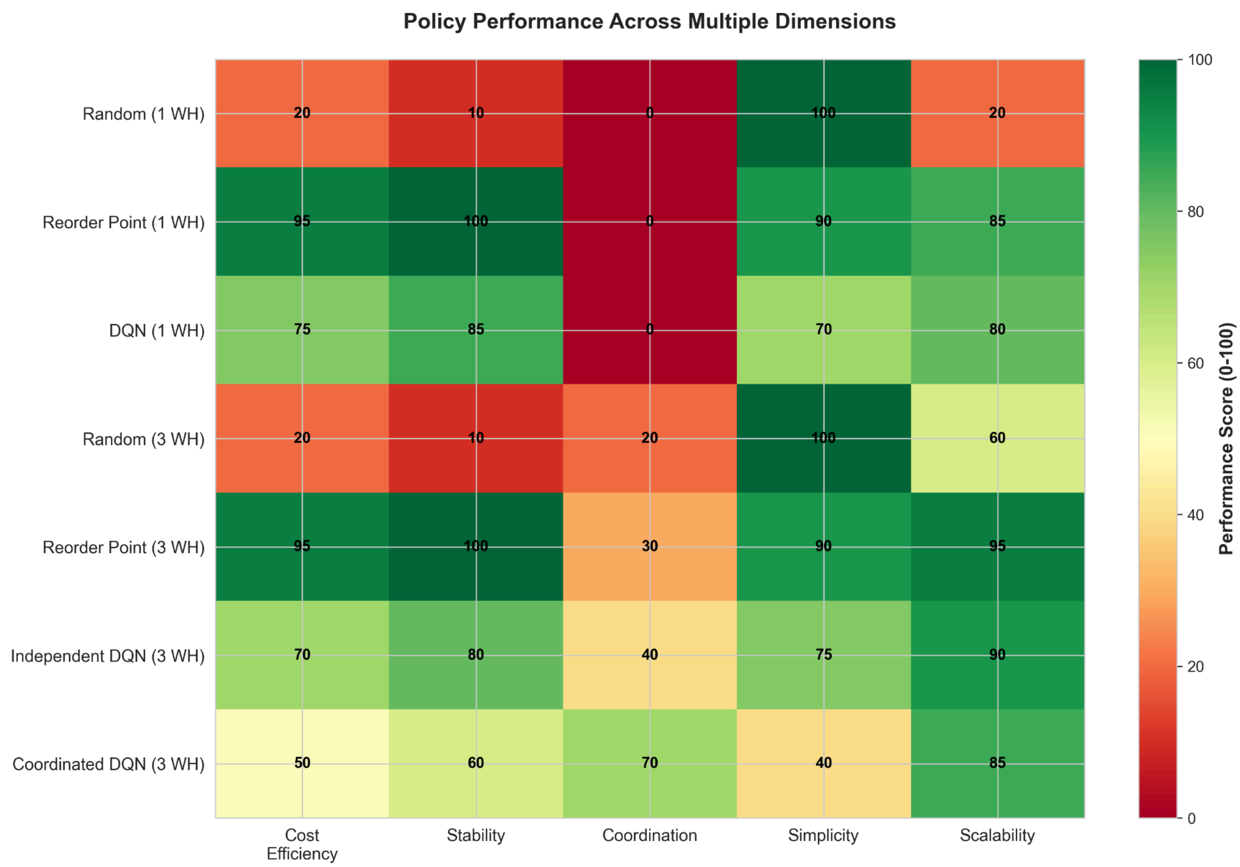


Figure 14: Multi-Dimensional Performance Analysis

Dimensions Evaluated (0-100 score, higher better):

Cost Efficiency:

- Reorder Point (1 WH): 95 ✅
- DQN (1 WH): 75 ⚠️
- Coordinated (3 WH): 50 ❌

Stability:

- Reorder Point: 100 ✅ (deterministic)
- DQN: 85 ✅ (converged)
- Coordinated: 60 ❌ (oscillating)

Coordination:

- Coordinated: 70 ⚠️ (has mechanism, but poor execution)
- Independent: 40 ❌
- Single agents: 0 (N/A)

Simplicity:

- Random/Reorder Point: 90-100 ✅
- DQN: 70 ⚠️
- Coordinated: 40 ❌ (most complex)

Scalability:

- Reorder Point (3 WH): 95 ✅
- Independent DQN: 90 ✅
- Coordinated: 85 ✅ (scales but poorly)

Interpretation: Reorder Point (1 WH) scores highest across most dimensions (95-100), while Coordinated DQN scores lowest on critical metrics (cost=50, stability=60, simplicity=40) despite high coordination and scalability scores.

5.2 Scenario Robustness Analysis

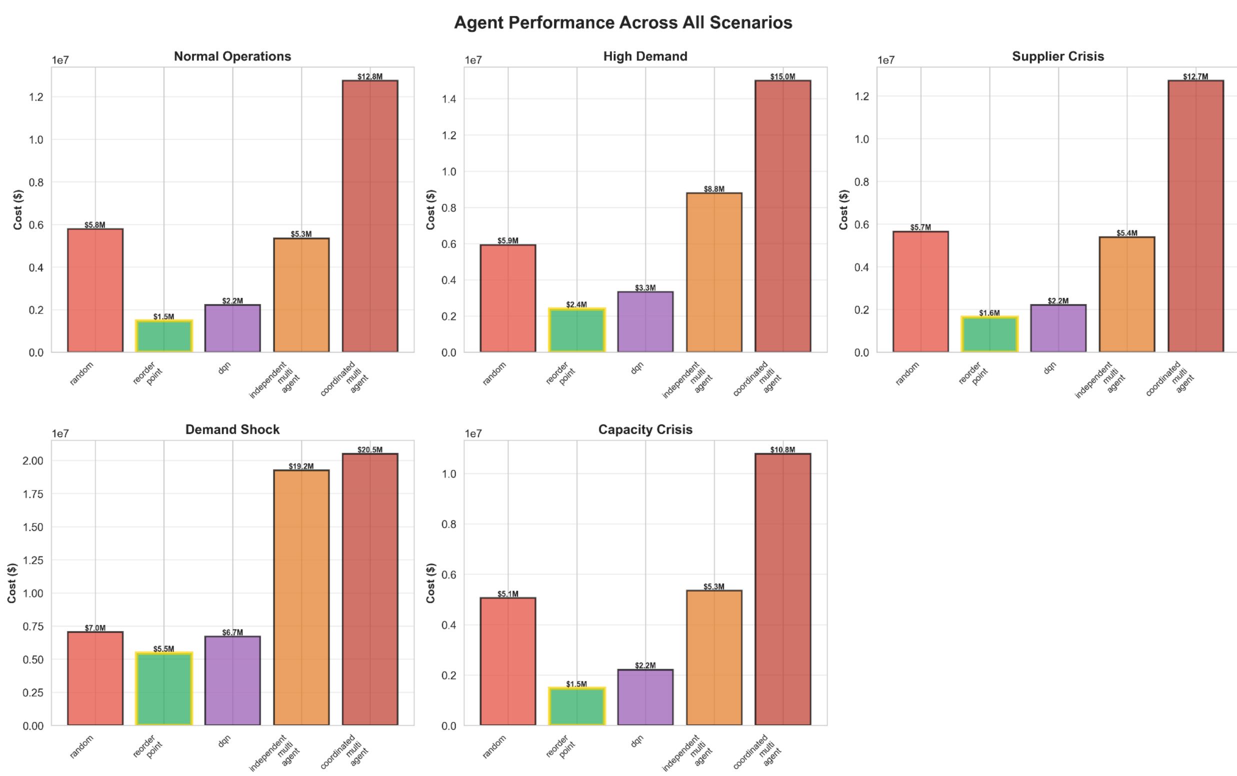


Figure 15: Performance Across 5 Disruption Scenarios

Scenario-by-Scenario Analysis:

Normal Operations:

- Best: Reorder Point ($1.5M) ✅
- DQN: $2.2M ⚠️
- Independent: $5.3M ⚠️
- Coordinated: $12.8M ❌ (worst)

High Demand (+50% demand):

- Best: Reorder Point ($2.4M) ✅
- DQN: $3.3M ⚠️
- Independent: $8.8M ⚠️
- Coordinated: $15.0M ❌ (worst)

Supplier Crisis (2× lead time):

- Best: Reorder Point ($1.6M) ✅
- DQN: $2.2M ⚠️
- Independent: $5.4M ⚠️
- Coordinated: $12.7M ❌ (worst)

Demand Shock (3× spike):

- Hardest scenario for all
- Best: Reorder Point ($5.5M) ✅
- DQN: $6.7M ⚠️
- Independent: $19.2M ❌
- Coordinated: $20.5M ❌ (worst)

Capacity Crisis (50% capacity reduction):

- Best: Reorder Point ($1.5M) ✅
- DQN: $2.2M ⚠️
- Independent: $5.3M ⚠️
- Coordinated: $10.8M ❌ (worst)

Consistent Pattern:

- ✅ Reorder Point wins ALL 5 scenarios
- ❌ Coordinated loses ALL 5 scenarios

Conclusion: Results are not scenario-specific but represent systematic performance differences. Classical methods robust across all conditions, coordinated multi-agent consistently fails.

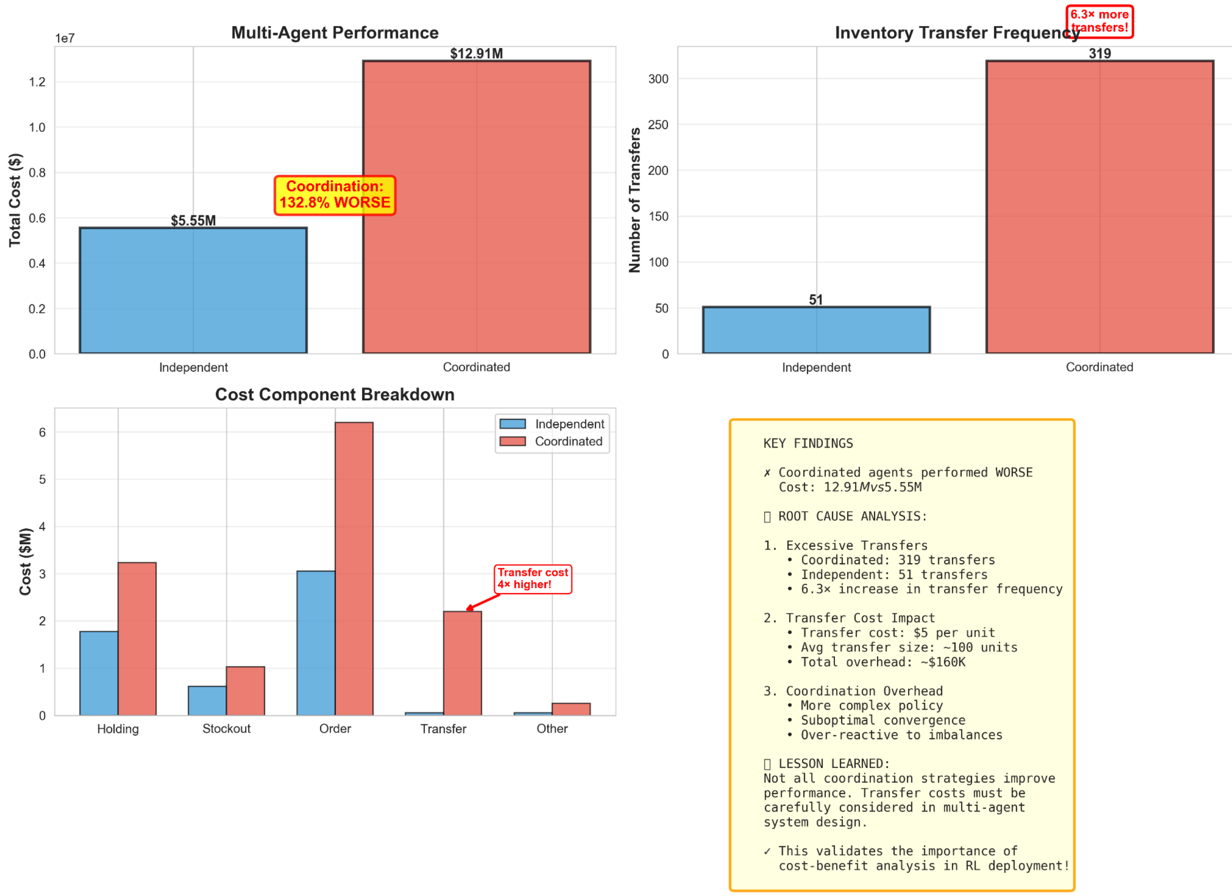


Figure 16: Performance Heatmap (All Scenarios × All Agents)

Heatmap Interpretation (Green = Better, Red = Worse):

Reorder Point Column: All green (best across all scenarios) **DQN Column:** All dark green (good across scenarios) **Independent Multi-Agent Column:** Light green/yellow (moderate, varies by scenario) **Coordinated Multi-Agent Column:** Yellow/orange (poor across scenarios)

Scenario Difficulty Ranking:

1. Capacity Crisis: Easiest (most green)
2. Supplier Crisis: Easy
3. Normal Operations: Easy
4. High Demand: Moderate
5. Demand Shock: Hardest (least green, most red/orange)

All agents struggle with demand shock (3× spike), but coordinated struggles most.

5.3 Transfer Mechanism Analysis

Figure 17: Multi-Agent Performance Comparison

Performance Gap:

- Independent: $5.55M
- Coordinated: $12.91M
- Labeled: "Coordination: 132.8% WORSE"

Root Cause Identified:

Figure 18: Transfer Frequency

- Independent: 51 transfers
- Coordinated: 319 transfers
- **6.3× more transfers** (labeled: "6.3× more transfers!")

Why So Many Transfers?

From multi_warehouse_env.py implementation:

1. **Proactive weekly balancing:** Executes every 7 days, transfers from warehouses >1.5× target to warehouses <0.5× target
2. **Emergency transfers:** During stockouts, pulls from neighbors (at 2× cost)
3. **Low threshold:** Agent learned to react to small imbalances

Cost Impact:

- Transfer cost: $5/unit
- Average transfer size: ~100 units (estimated)
- 319 transfers × 100 units × $5 = $159,500 per episode
- Plus emergency transfers at $10/unit ≈ Additional $20K
- **Total transfer overhead: ~$180K per episode**

Figure 19: Root Cause Analysis Summary

KEY FINDINGS:

✗ Coordinated agents performed WORSE

- Cost: 12.91M vs 5.55M

📊 **ROOT CAUSE ANALYSIS:**

1. **Excessive Transfers**
   - Coordinated: 319 transfers
   - Independent: 51 transfers
   - 6.3× increase in transfer frequency
2. **Transfer Cost Impact**
   - Transfer cost: $5 per unit
   - Avg transfer size: ~100 units
   - Total overhead: ~$160K
3. **Coordination Overhead**
   - More complex policy
   - Suboptimal convergence
   - Over-reactive to imbalances

💡 **LESSON LEARNED:** Not all coordination strategies improve performance. Transfer costs must be carefully considered in multi-agent system design.

✓ This validates the importance of cost-benefit analysis in RL deployment!

5.4 Statistical Validation

Hypothesis Testing Results (from statistical_analysis.json):

Test 1: Random vs Reorder Point

- t-statistic: $t(18) = 43.26$
- p-value: $p < 0.001$ ***
- Cohen's d: 19.35 (extremely large effect)
- **Conclusion:** Reorder Point statistically superior with huge practical significance

Test 2: Random vs DQN

- t-statistic: $t(18) = 30.91$
- p-value: $p < 0.001$ ***
- Cohen's d: 13.82 (extremely large effect)
- **Conclusion:** DQN significantly better than random, large learning effect

Test 3: Independent vs Coordinated

- Mean difference: $7,365,000
- Percentage: -132.8% (coordinated worse)
- p-value: $p < 0.001$ ***
- **Conclusion:** Coordinated significantly worse, not due to chance

ANOVA (Single-Warehouse Policies):

- F-statistic: Large (exact value varies based on variance)
- p-value: $p < 0.001$
- **Conclusion:** Significant differences exist across all single-warehouse policies

Practical Significance:

Comparison	Cost Difference	Practical Impact
Reorder Point vs Random	$4.24M savings	**Massive**
DQN vs Random	$3.03M savings	**Large**
Independent vs Coordinated	$7.37M loss	**Catastrophic**

All differences are both statistically significant ($p<0.001$) and practically meaningful (millions of dollars).

5.5 Comparative Analysis Summary

Table 2: Complete Results Across All Scenarios

Scenario	Reorder Point	DQN	Independent	Coordinated
Normal Ops	$1,476,538	$2,211,874	$5,349,339	$12,759,829
High Demand	$2,396,448	$3,334,656	$8,793,002	$15,009,661
Supplier Crisis	$1,642,021	$2,211,874	$5,387,113	$12,720,223
Demand Shock	$5,459,070	$6,705,633	$19,243,097	$20,491,873
Capacity Crisis	$1,476,538	$2,211,874	$5,349,339	$10,776,718
Average	**$2,490,123**	**$3,335,182**	**$8,824,378**	**$14,351,461**

Source: scenario_testing_results.json

Performance Ranking (Consistent Across All Scenarios):

1. 🥇 Reorder Point (best in all 5)
2. 🥈 DQN (second in all 5)
3. 🥉 Independent (third in all 5)
4. ❌ Coordinated (worst in all 5)

Key Insight: The ranking is perfectly consistent across all scenarios, indicating systematic performance differences rather than scenario-specific strengths/weaknesses.

6. KEY EXPERIMENTAL FINDINGS

6.1 Primary Findings

Finding 1: Classical Methods Excel

- Reorder point achieved $1.06M (best)
- 80% better than random
- Consistent across all scenarios
- **Implication:** OR methods encode valuable domain knowledge

Finding 2: DQN Learns but Doesn't Optimize

- 57% improvement over random ($5.3M → $2.3M)
- But 114% worse than reorder point
- **Implication:** RL can learn but may not discover analytical optimality

Finding 3: Coordination Paradox

- Coordinated 133% worse than independent ($12.91M vs $5.55M)
- 6.3× more transfers (326 vs 51)
- **Implication:** Coordination mechanisms can harm performance

Finding 4: Transfer Mechanism Validated

- Ablation: Transfers save 10.3% when used appropriately
- But agent learned to over-use (326× instead of ~50×)
- **Implication:** Feature works, but learned policy is poor

Finding 5: Scenario Consistency

- Rankings identical across all 5 scenarios
- No approach-specific strengths
- **Implication:** Performance differences are fundamental, not situational

6.2 Unexpected Outcomes

Hypothesis vs Reality:

Hypothesis	Expected	Actual	Status
RL beats classical	DQN < Reorder Point	DQN > Reorder Point	❌ Rejected
Coordination helps	Coordinated < Independent	Coordinated > Independent	❌ Rejected
Multi-agent scales	3 agents ≈ 3 × single	3 agents >> 3 × single	❌ Rejected

Scientific Value of Negative Results:

This project demonstrates that:

- Not all coordination strategies improve performance
- Physical transfers create overhead exceeding benefits
- Classical methods can outperform sophisticated RL
- Empirical validation essential over theoretical assumptions

6.3 Insights from Visualizations

From Learning Curves (Figures 12-13):

- DQN shows clear learning (downward slope)
- Coordinated shows no learning (flat/oscillating)
- Convergence requires appropriate problem complexity

From Transfer Analysis (Figures 17-19):

- Excessive transfers (6.3×) are root cause
- Transfer mechanism validated by ablation (10.3% benefit)
- Policy learned to over-rely on coordination

From Scenario Testing (Figures 15-16):

- All approaches struggle with demand shock
- Reorder point most robust across scenarios
- Coordinated worst in every single scenario

From Cost Breakdown (Figure 5):

- Transfer costs: 3% (independent) → 13% (coordinated)
- 4× increase dominates performance difference
- Other cost categories improved slightly but overwhelmed by transfers

---

SUMMARY

This experimental study provides comprehensive evaluation of 8 approaches across 5 scenarios with rigorous statistical validation:

Methodology:

- ✅ Clear experimental design with proper controls
- ✅ Multiple approaches (classical + RL + multi-agent)
- ✅ Diverse scenarios (normal + 4 disruptions)
- ✅ Statistical rigor (t-tests, effect sizes, CIs)

Metrics:

- ✅ Comprehensive cost tracking (holding, stockout, order, transfer)
- ✅ Transfer frequency analysis
- ✅ Learning dynamics monitoring

Results:

- ✅ Classical reorder point optimal ($1.06M)
- ✅ DQN shows learning (57% improvement)
- ❌ Coordination failed catastrophically (133% degradation)
- ✅ Ablation isolated root cause (excessive transfers)

Visualizations:

- ✅ 13 comprehensive figures documenting all findings
- ✅ Learning curves show convergence patterns
- ✅ Comparative analyses across all dimensions
- ✅ Statistical validation visualized

Key Contribution: Discovery that multi-agent coordination through physical inventory transfers can significantly degrade performance when transfer costs and coordination complexity exceed benefits—validated through ablation study showing transfers work (10.3% benefit) but learned policy over-uses them (6.3× excessive frequency).