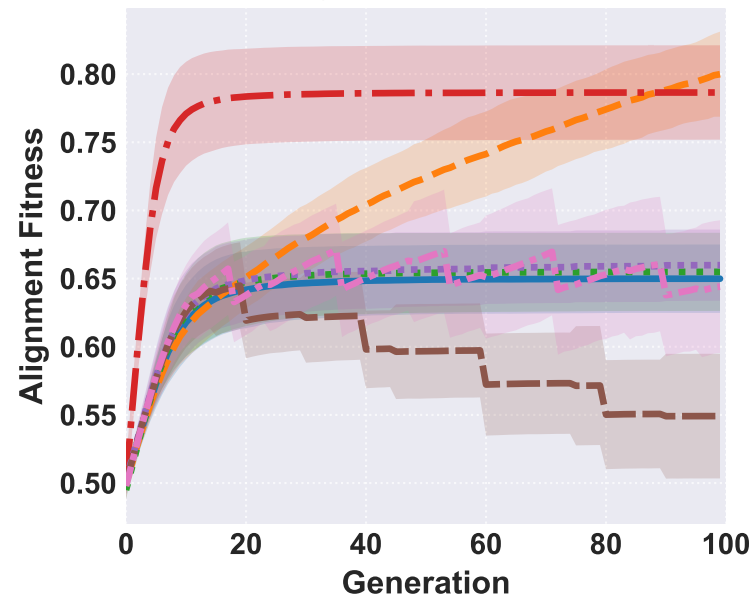
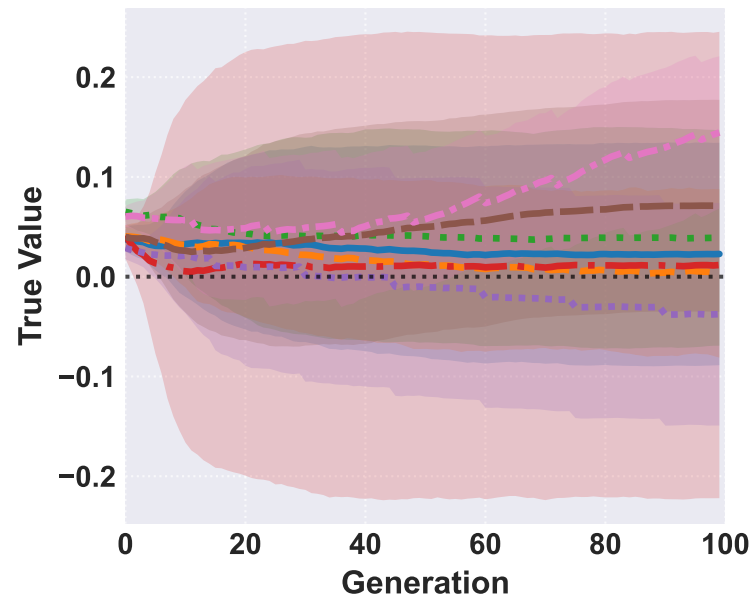


Level 4 Scenario Comparison: Metric Histories

Alignment Fitness



True Value



Deceptive Belief Ratio

