L1: Alignment Fitness vs True Value 0.35 14 0.30 12 Swept Parameter (Beta) Mean Final True Value 0.25 15 10 0.20 0.15 0.10 4 0.05 0.00 0.0 0.1 0.2 0.3 0.4 0.5 **Mean Final Alignment Fitness**