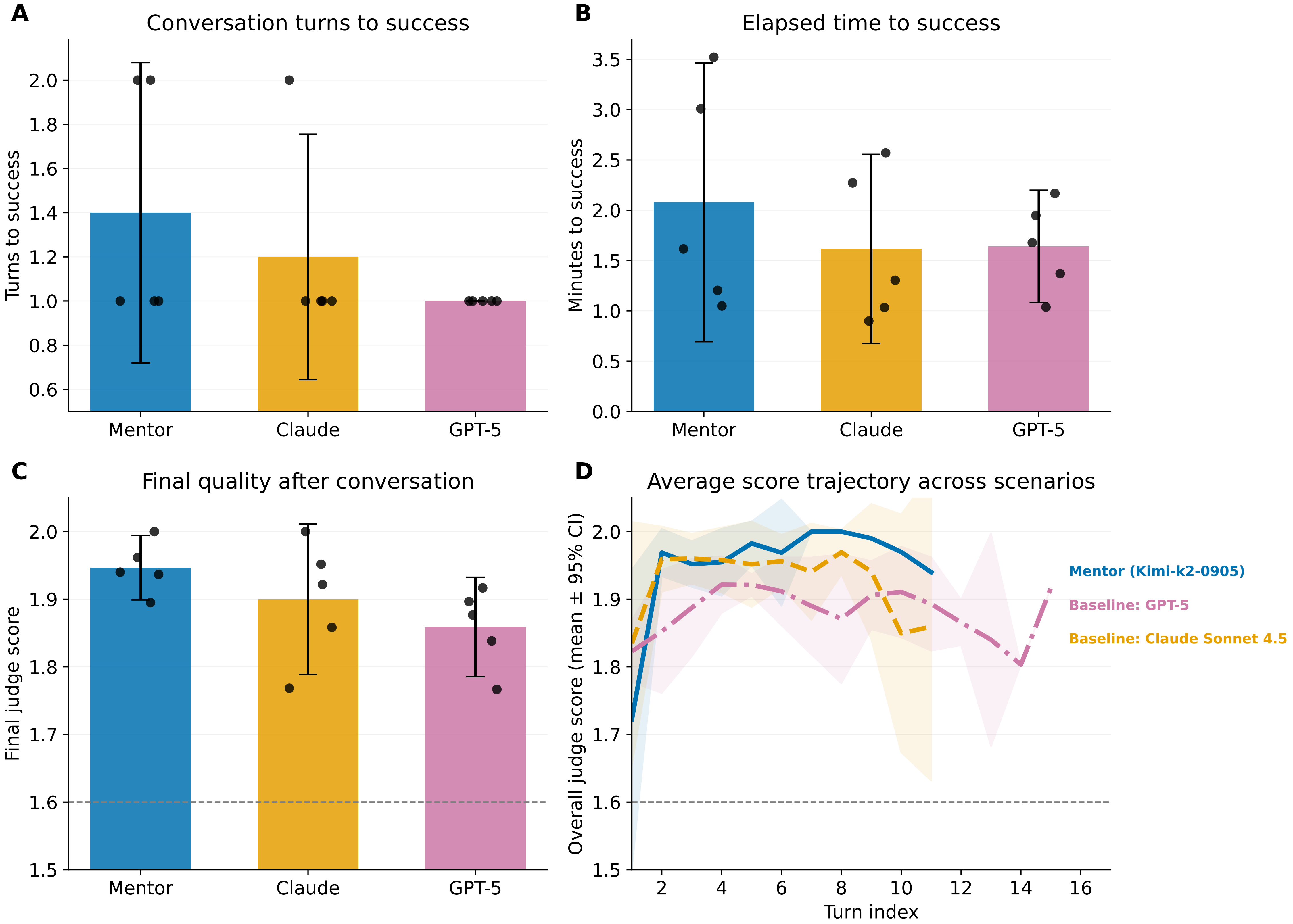
Multi-turn conversation quality and efficiency



A–C: dots mark individual scenarios (n=5 per agent), bars show group means ±95% CI; success threshold (score ≥ 1.6) denoted by dashed gray line. Panel D: mean score trajectory across sce