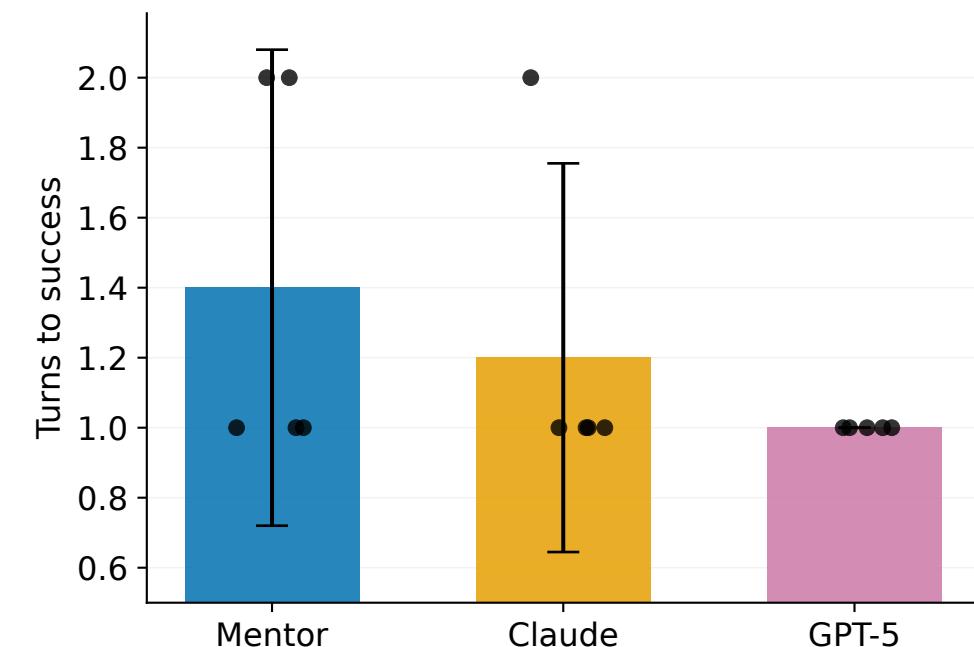


# Multi-turn conversation quality and efficiency

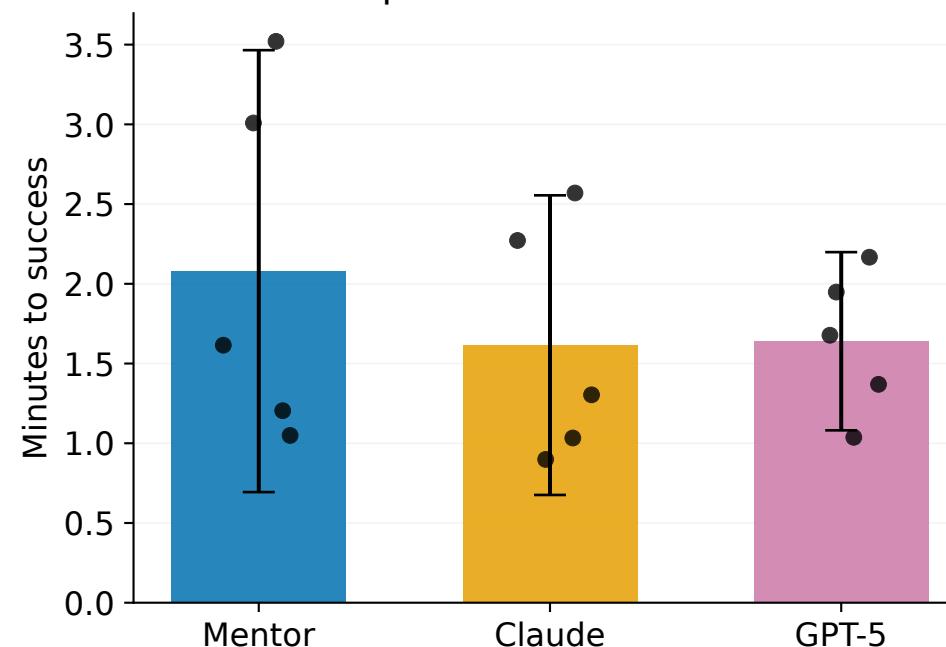
**A**

Conversation turns to success



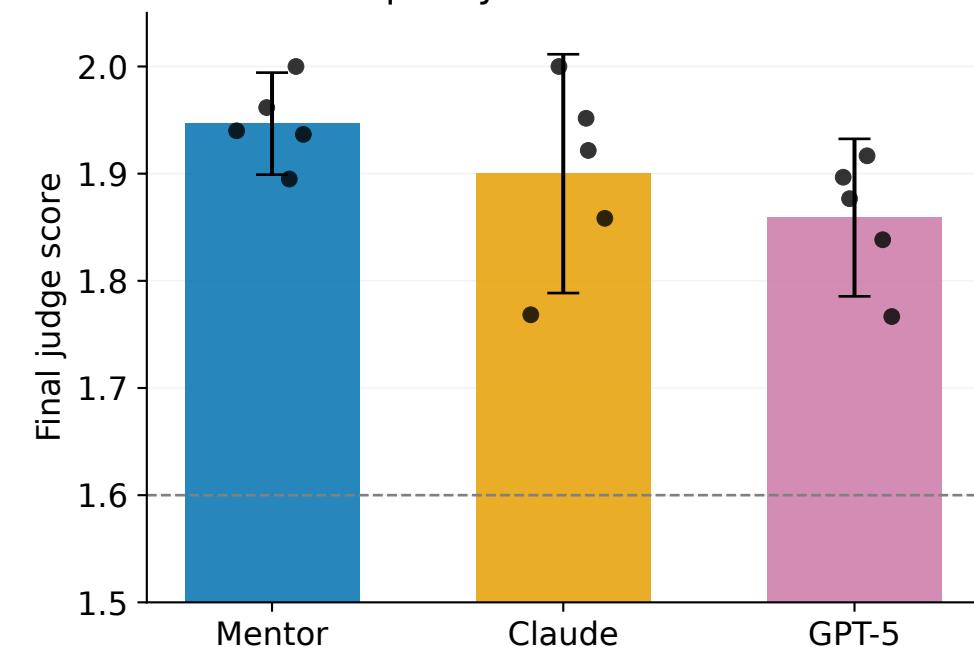
**B**

Elapsed time to success



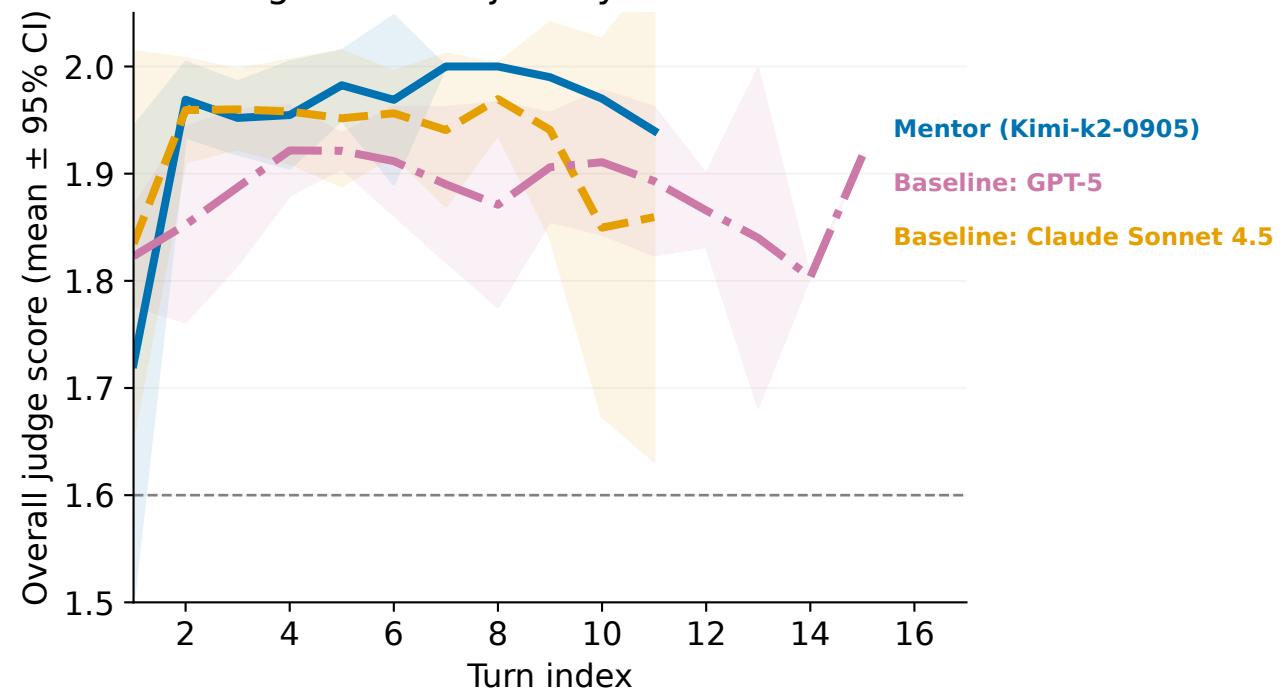
**C**

Final quality after conversation



**D**

Average score trajectory across scenarios



Panels A-C: dots mark individual scenarios (n=5 per agent), bars show group means  $\pm 95\%$  CI; success threshold (score  $\geq 1.6$ ) denoted by dashed gray line.  
Panel D: mean score trajectory across scenarios.