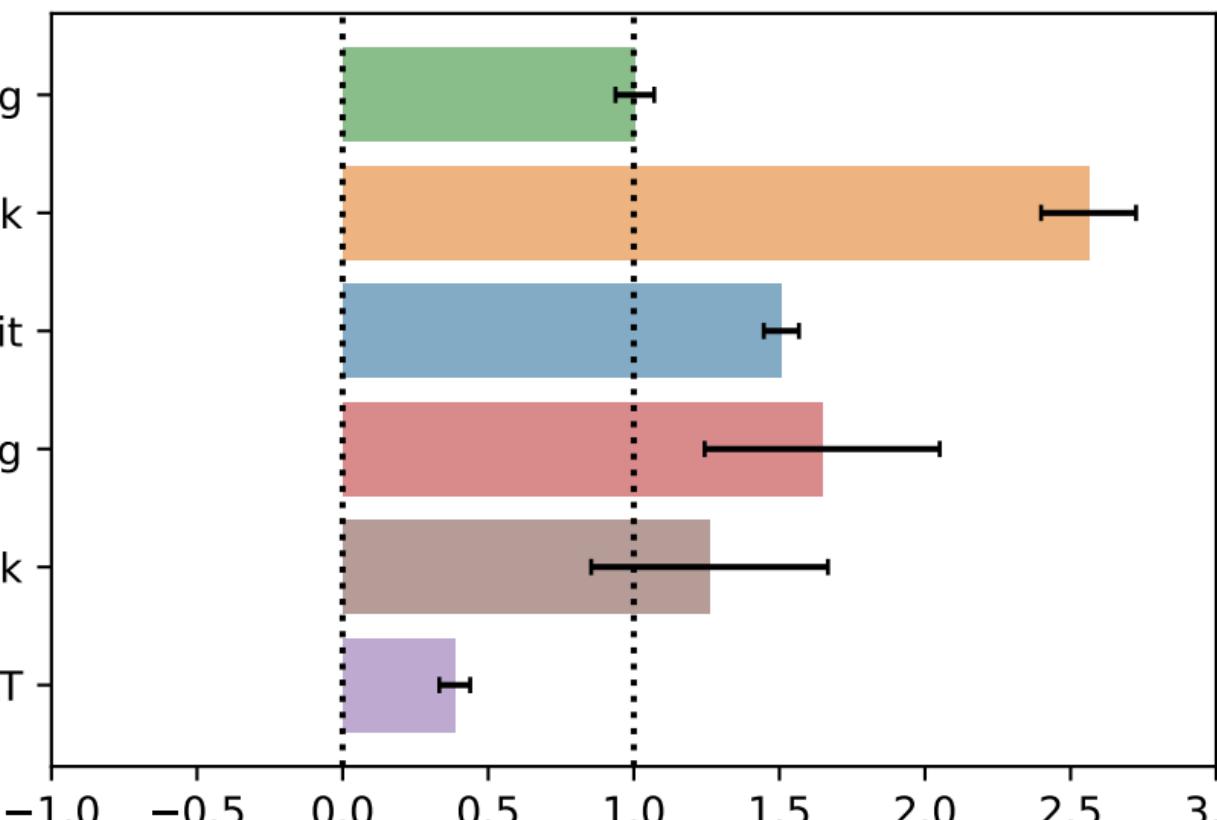


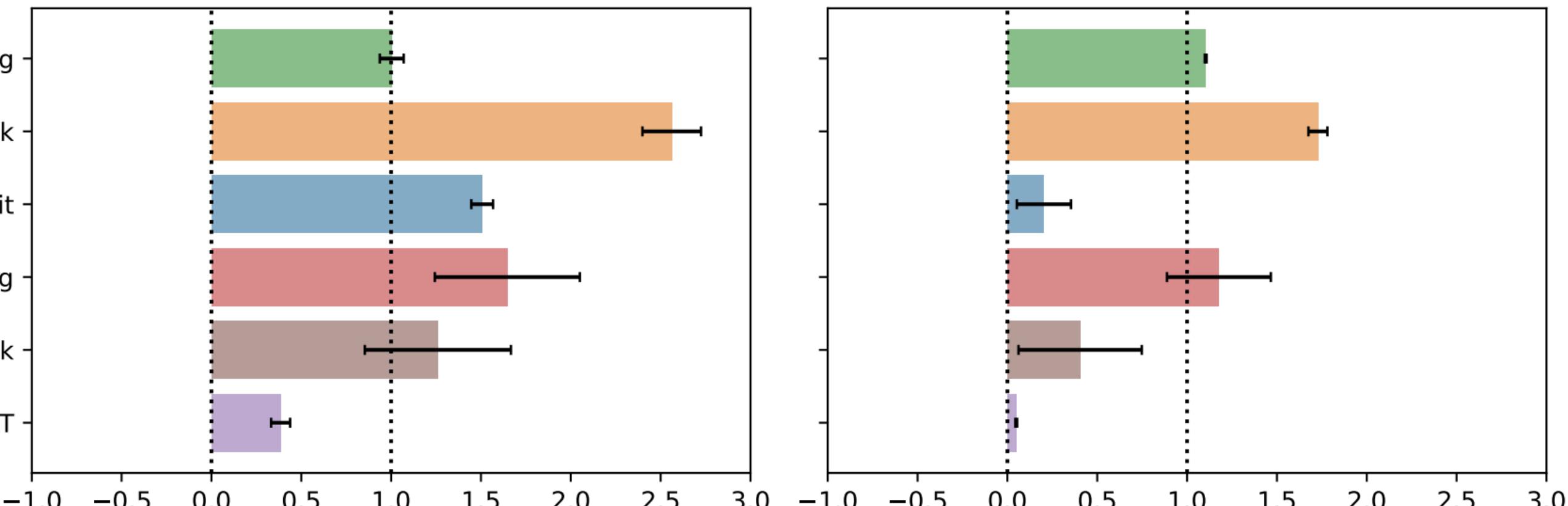
gpt-4

Probabilistic Reasoning



claude-2

Horizon Task



Values (Normalized: Random=0, Human average=1)