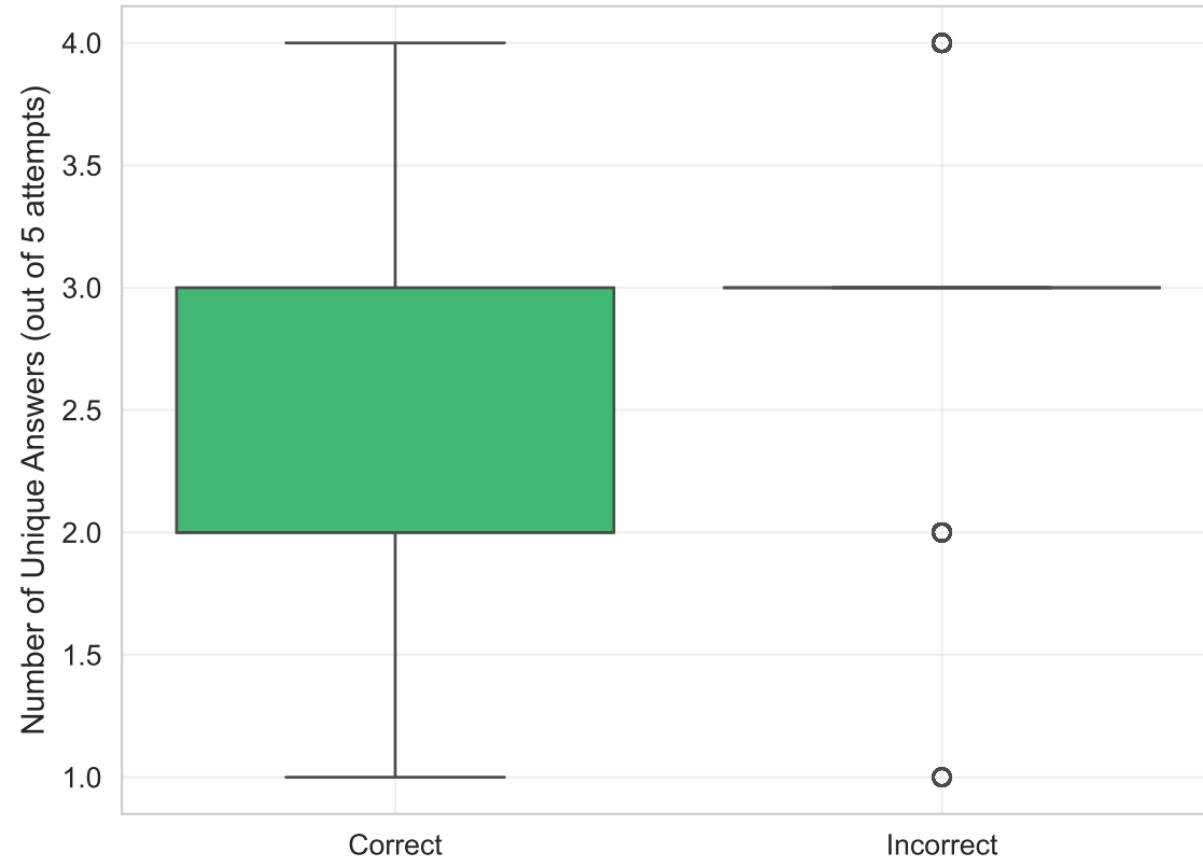


LLM Answer Uncertainty Analysis (N=10,000 questions)

Answer Uncertainty by Correctness



Mean Comparison (\pm SEM)

