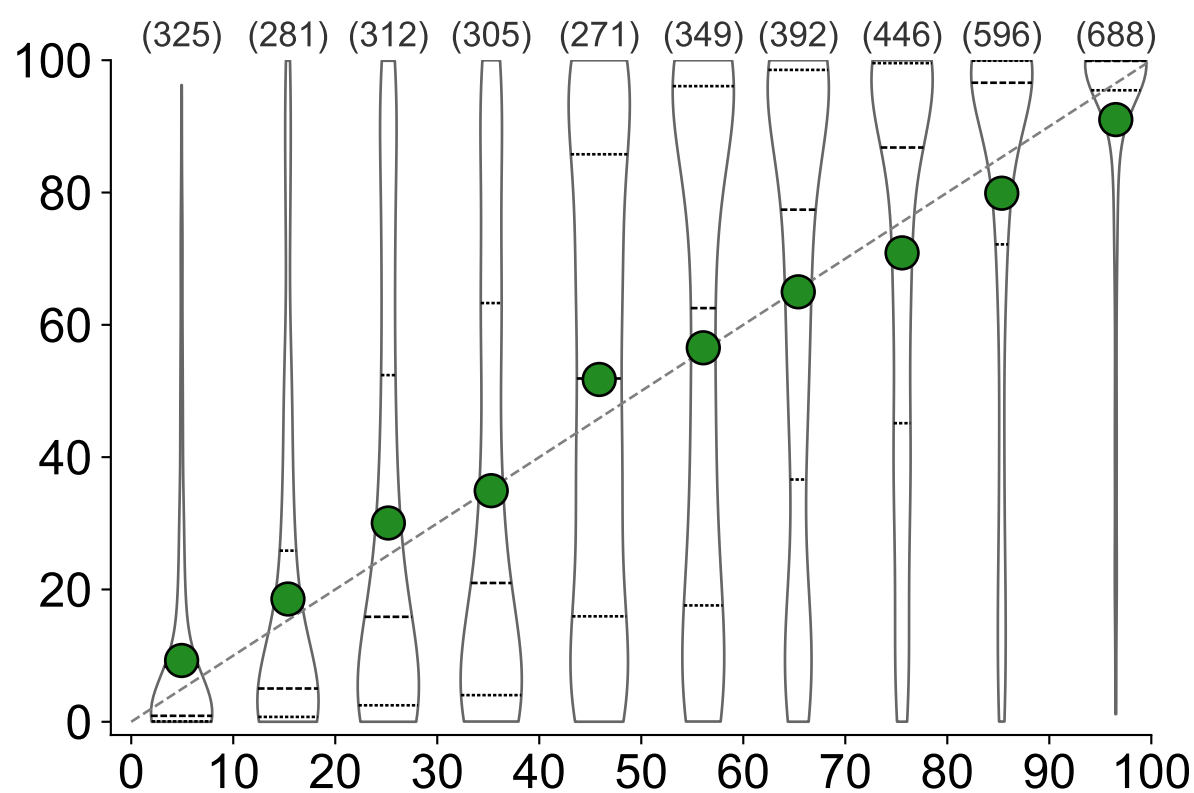
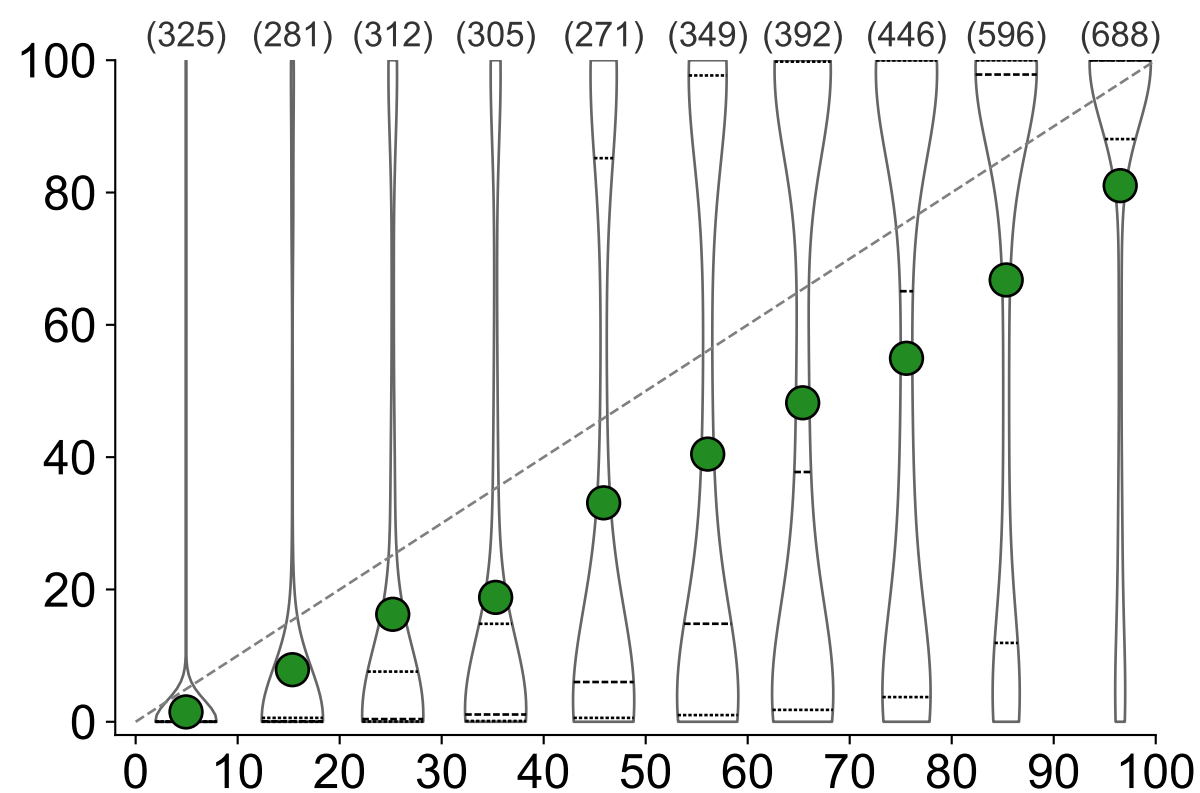


Probability model agrees with statement [question (a), %]

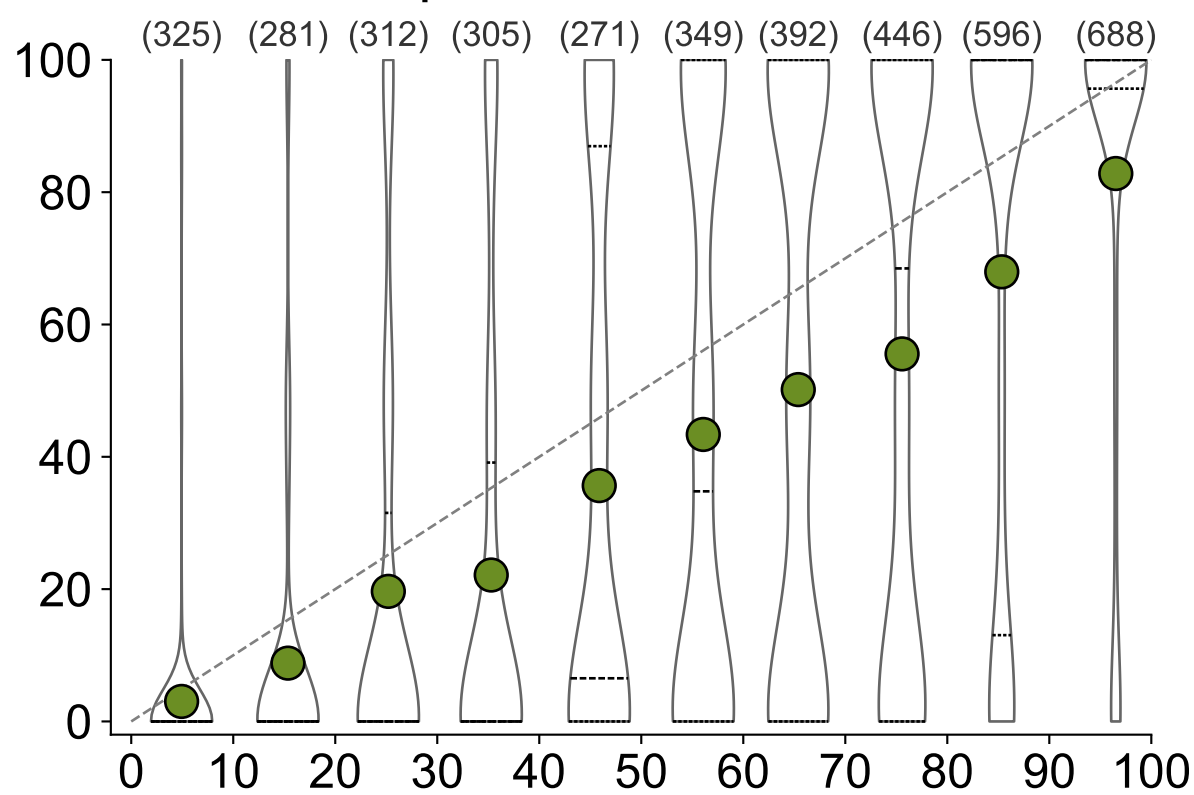
GPT-3.5: $r = .68$, RMSE = 29.36



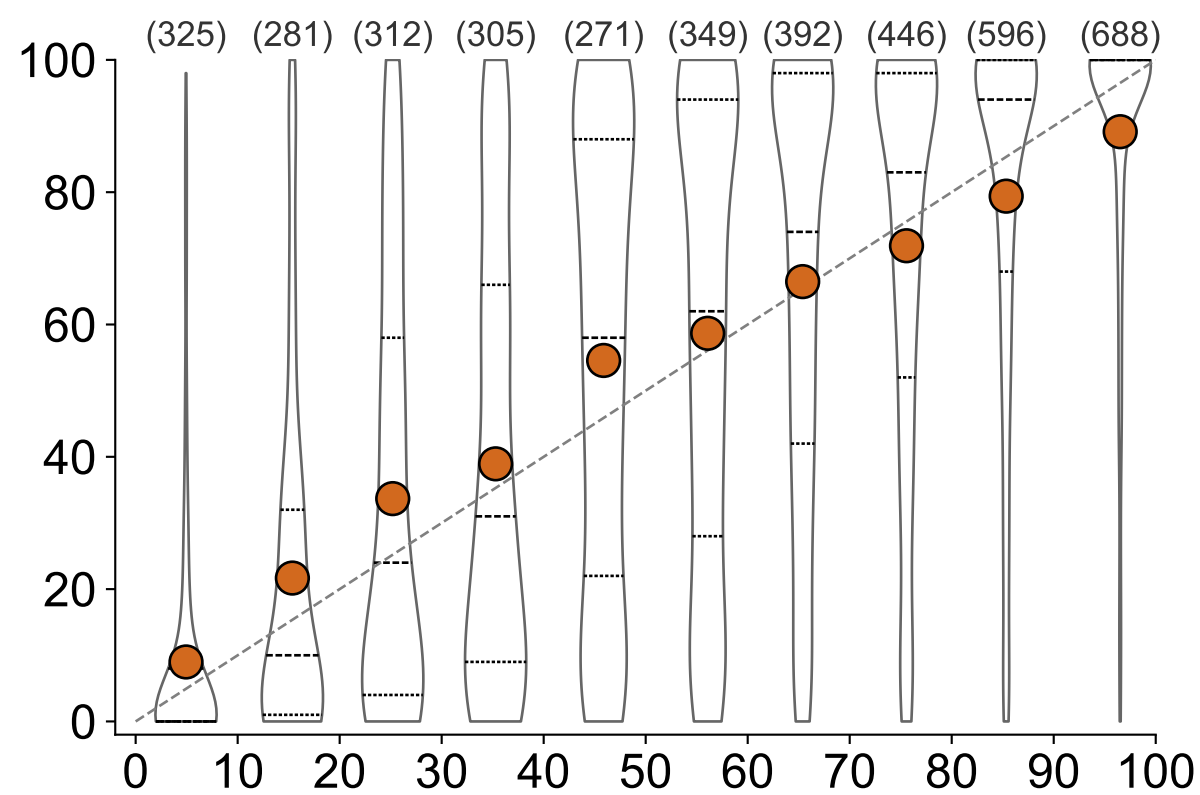
GPT-4o: $r = .61$, RMSE = 38.66



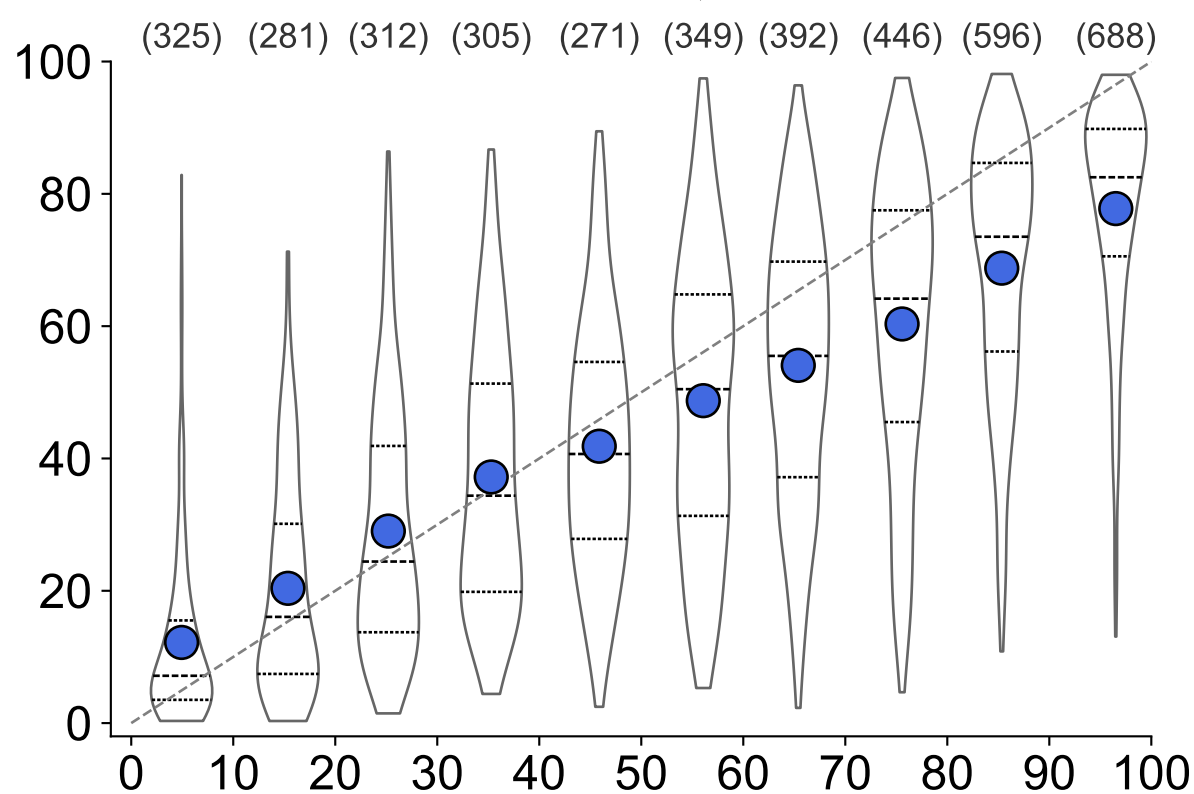
Claude 3 Opus: $r = .60$, RMSE = 38.67



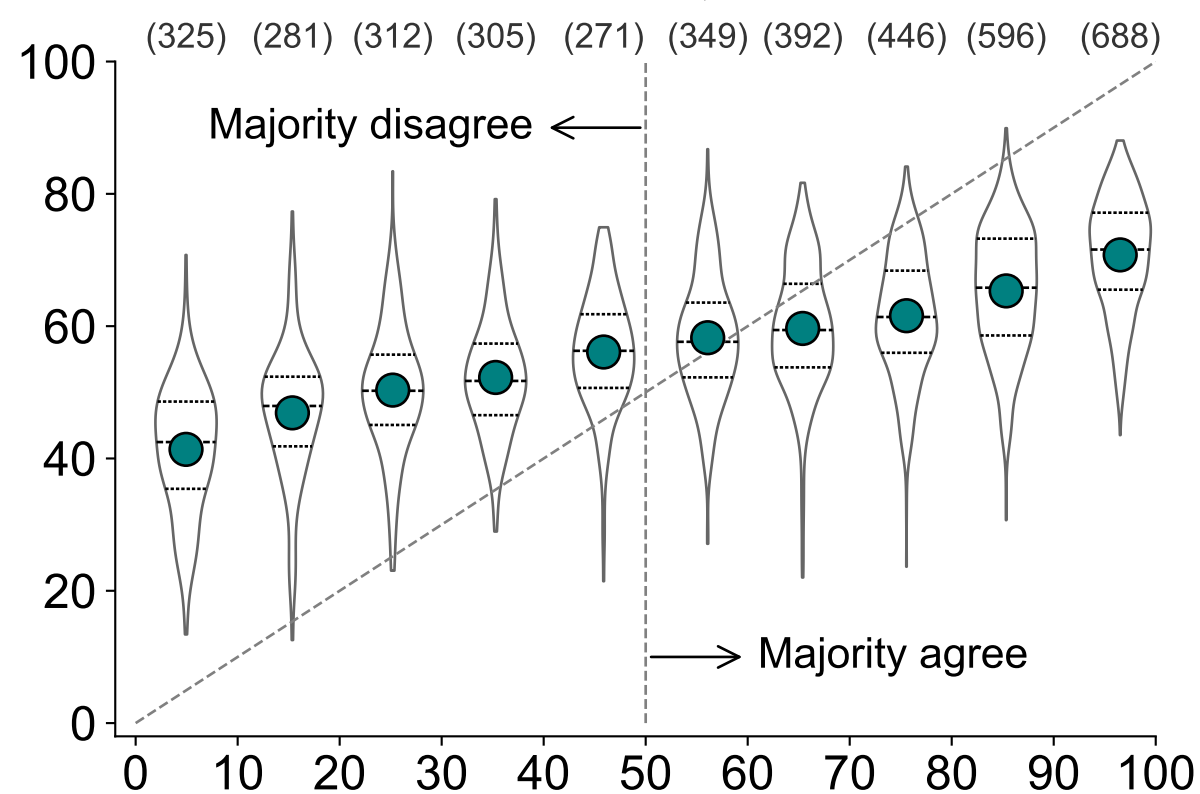
Gemini Pro 1.0: $r = .68$, RMSE = 28.59



Flan-T5-XXL: $r = .77$, RMSE = 21.83



Falcon-180B: $r = .71$, RMSE = 23.56



Percent of humans who agree with statement