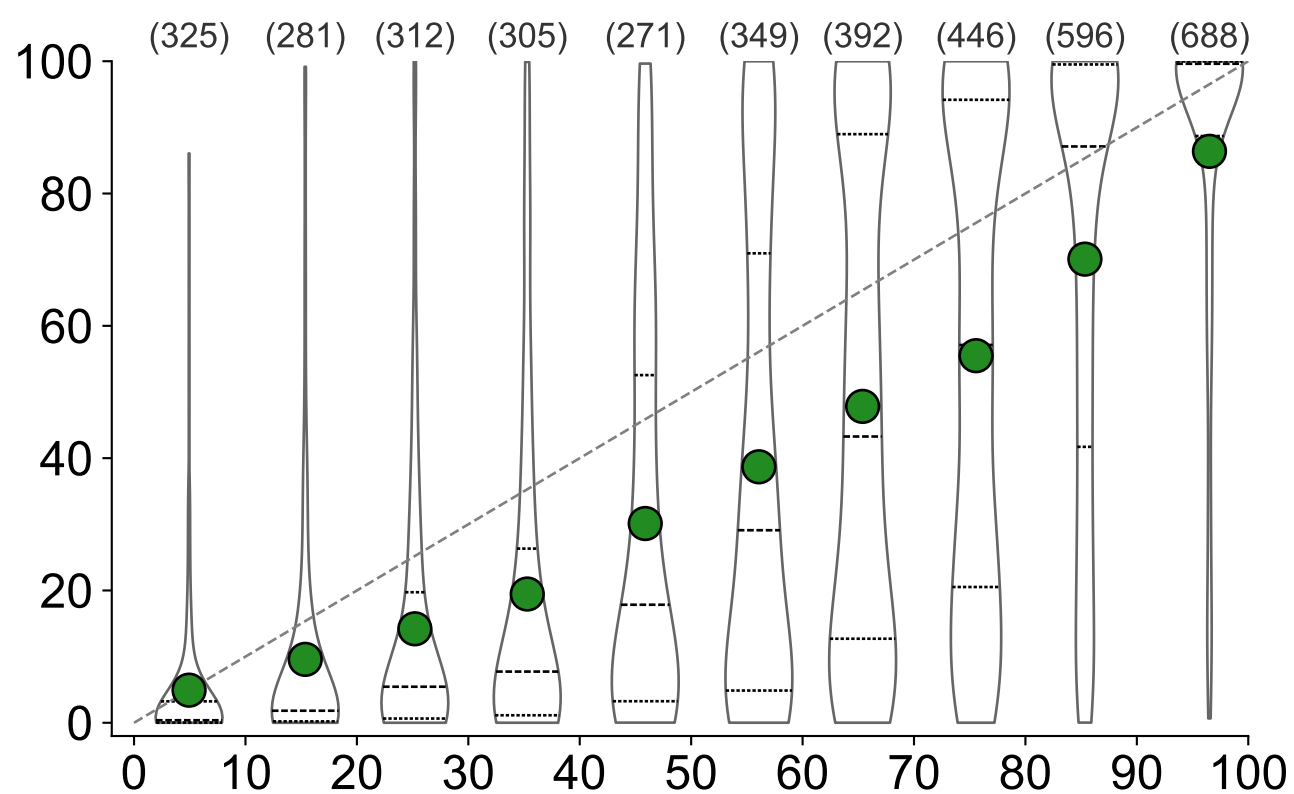
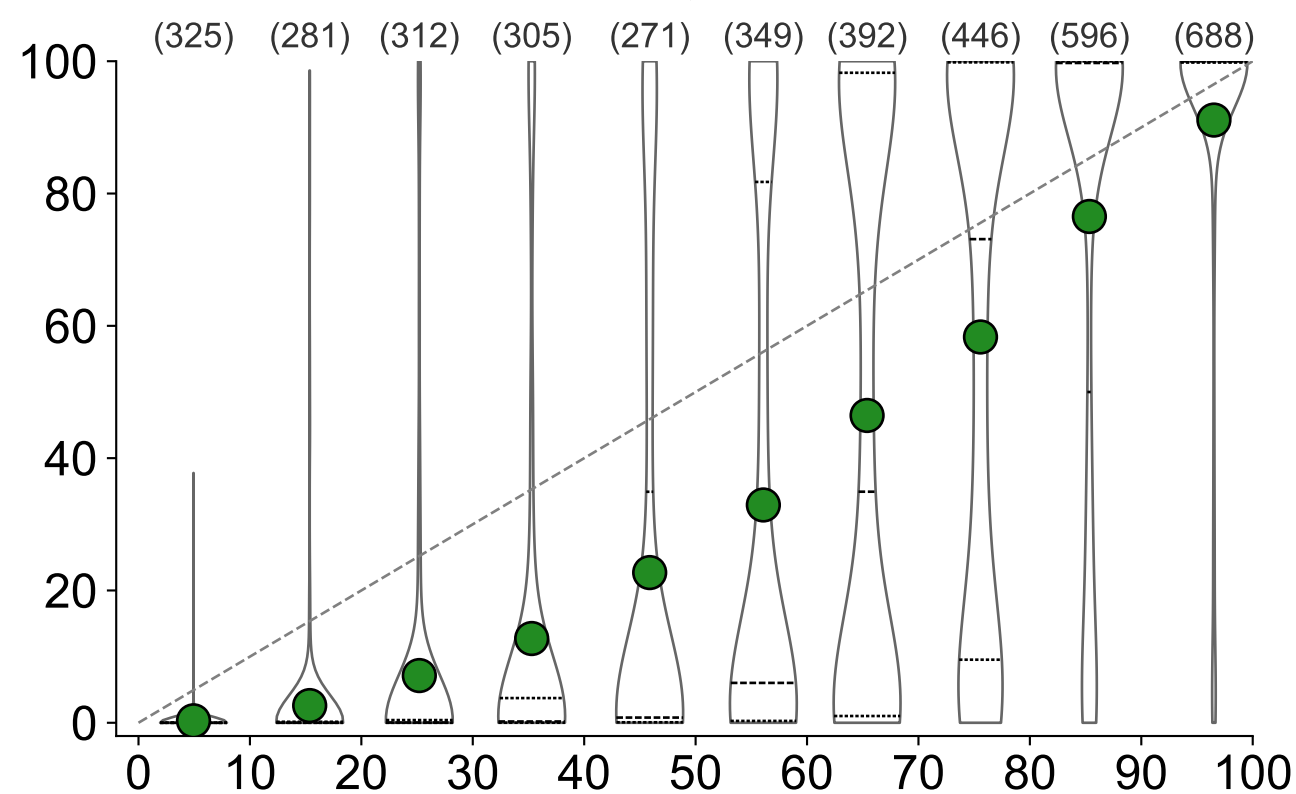


Probability model predicts most people would agree with statement [question (b), %]

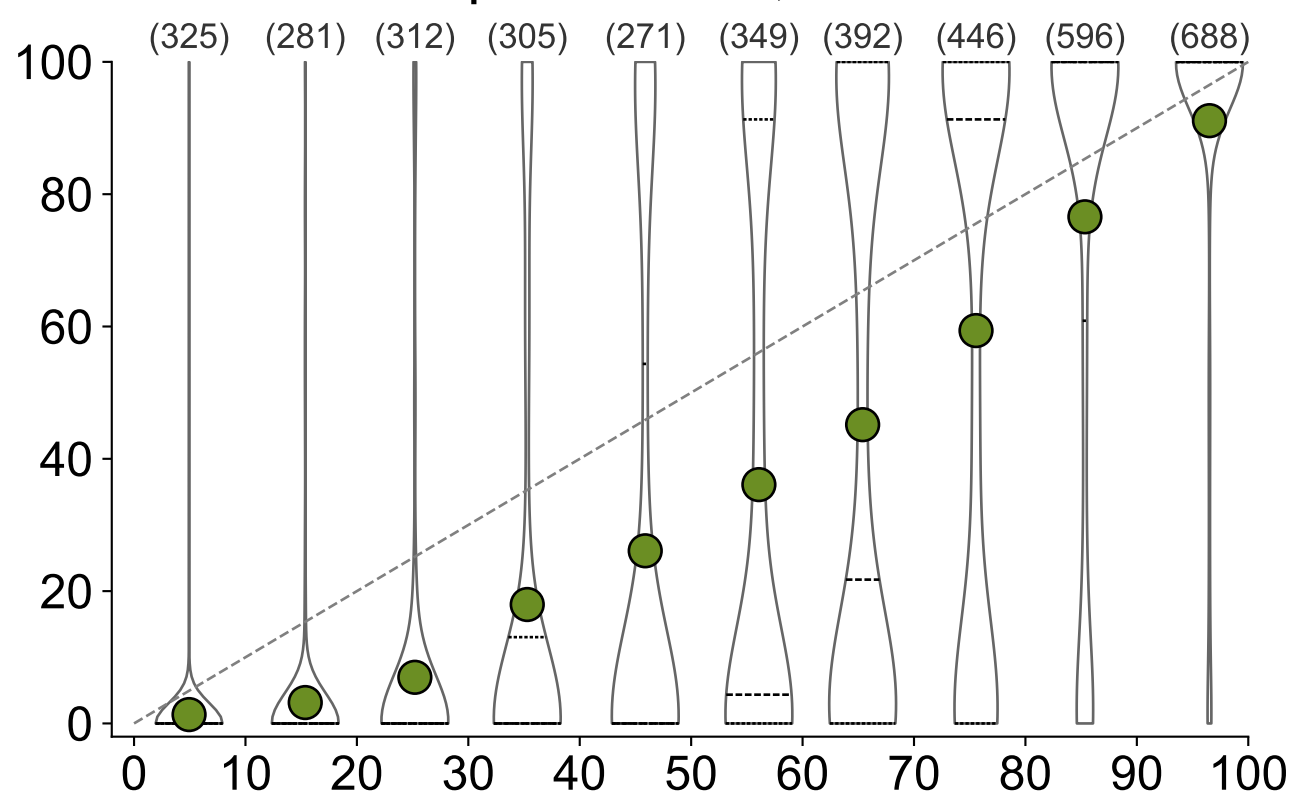
GPT-3.5:  $r = .71$ , RMSE = 31.29



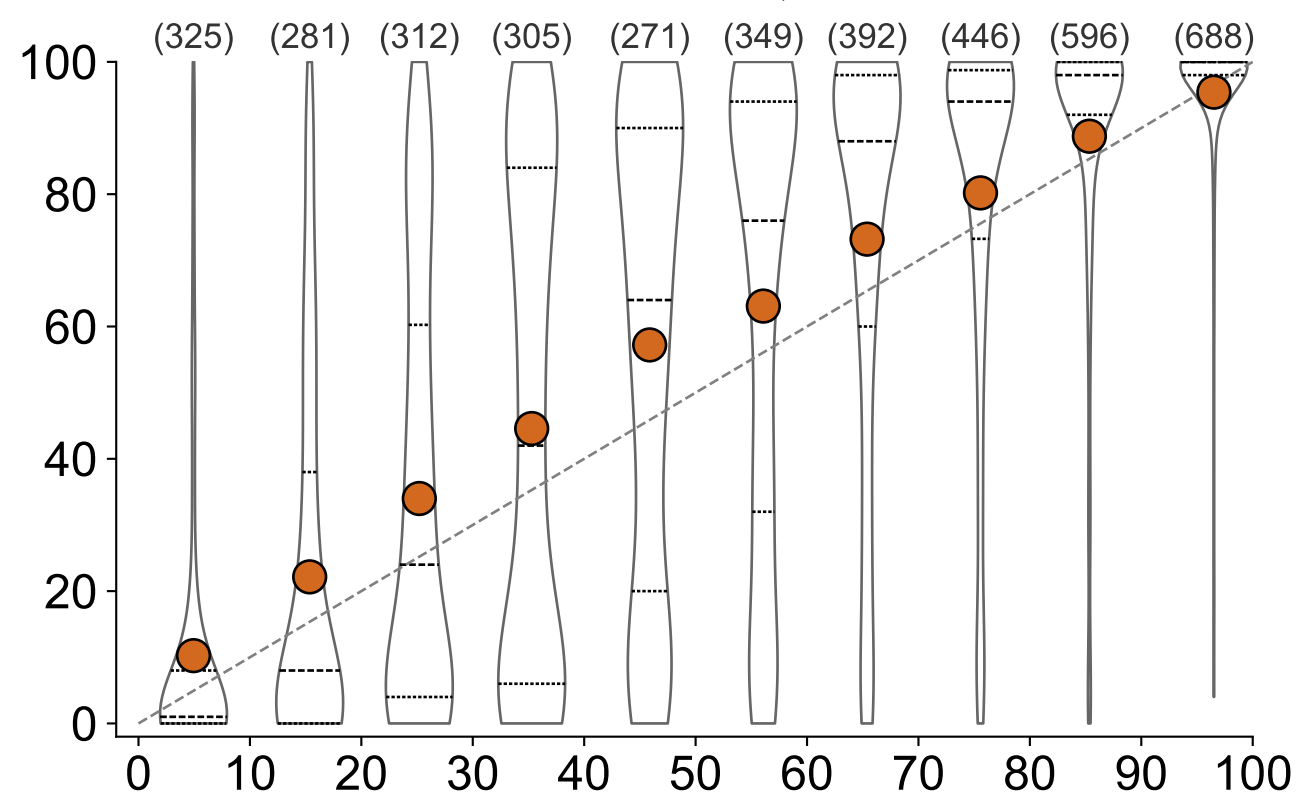
GPT-4o:  $r = .73$ , RMSE = 34.14



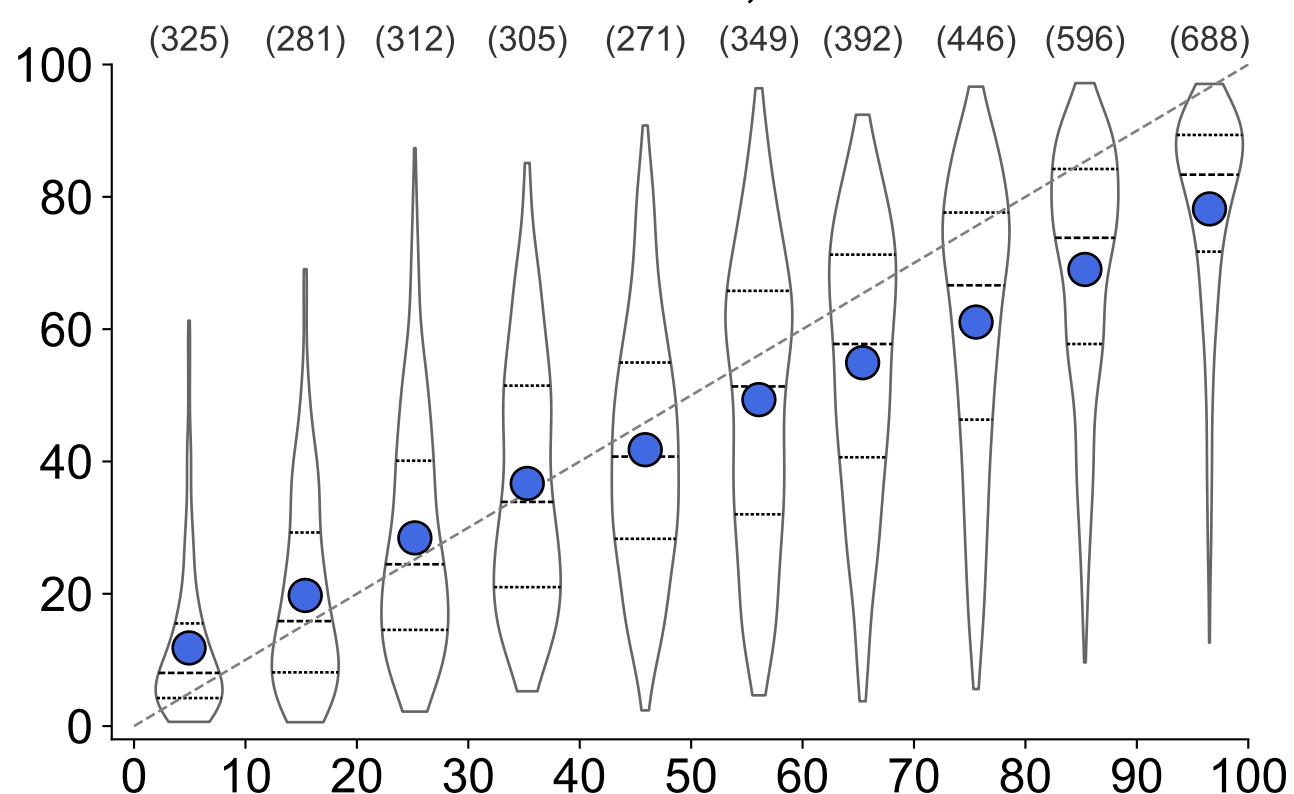
Claude 3 Opus:  $r = .71$ , RMSE = 35.37



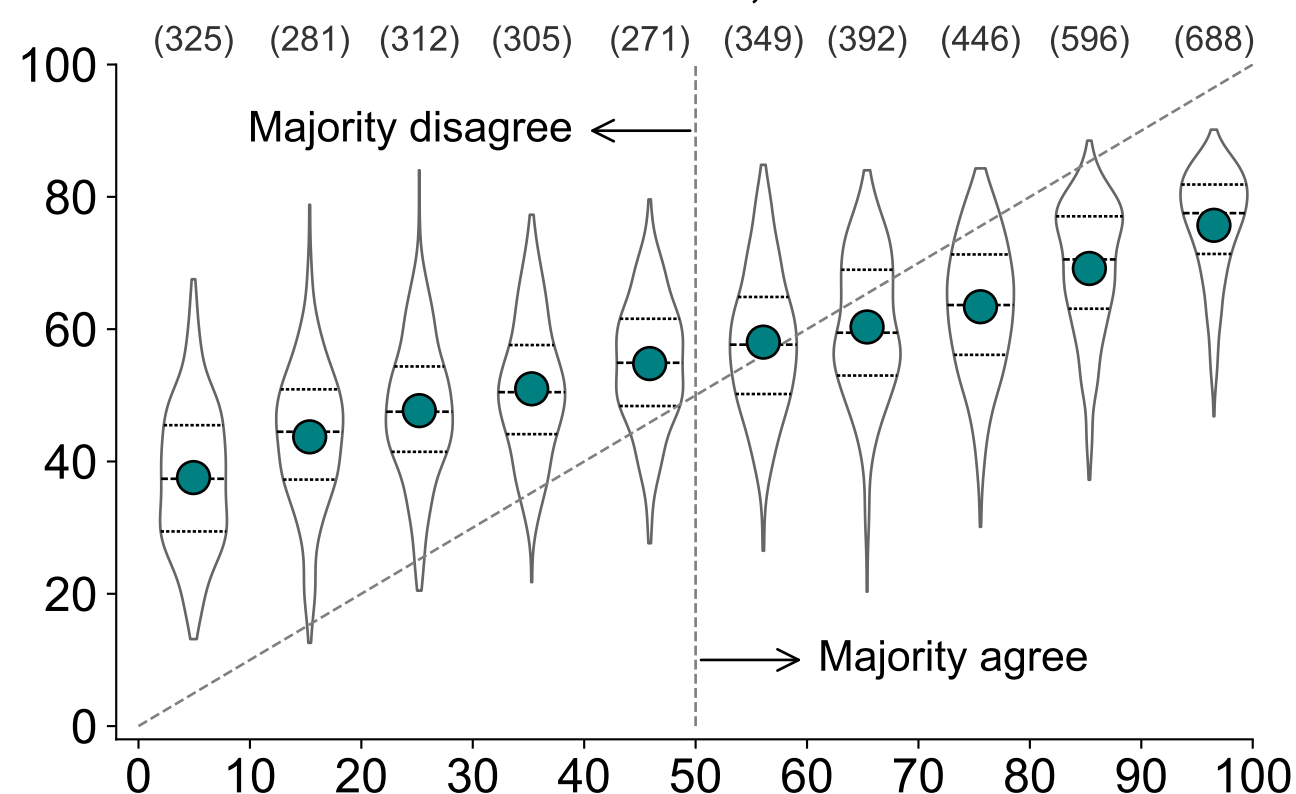
Gemini Pro 1.0:  $r = .73$ , RMSE = 27.25



Flan-T5-XXL:  $r = .78$ , RMSE = 21.22



Falcon-180B:  $r = .78$ , RMSE = 20.96



Percent of humans who agree with statement