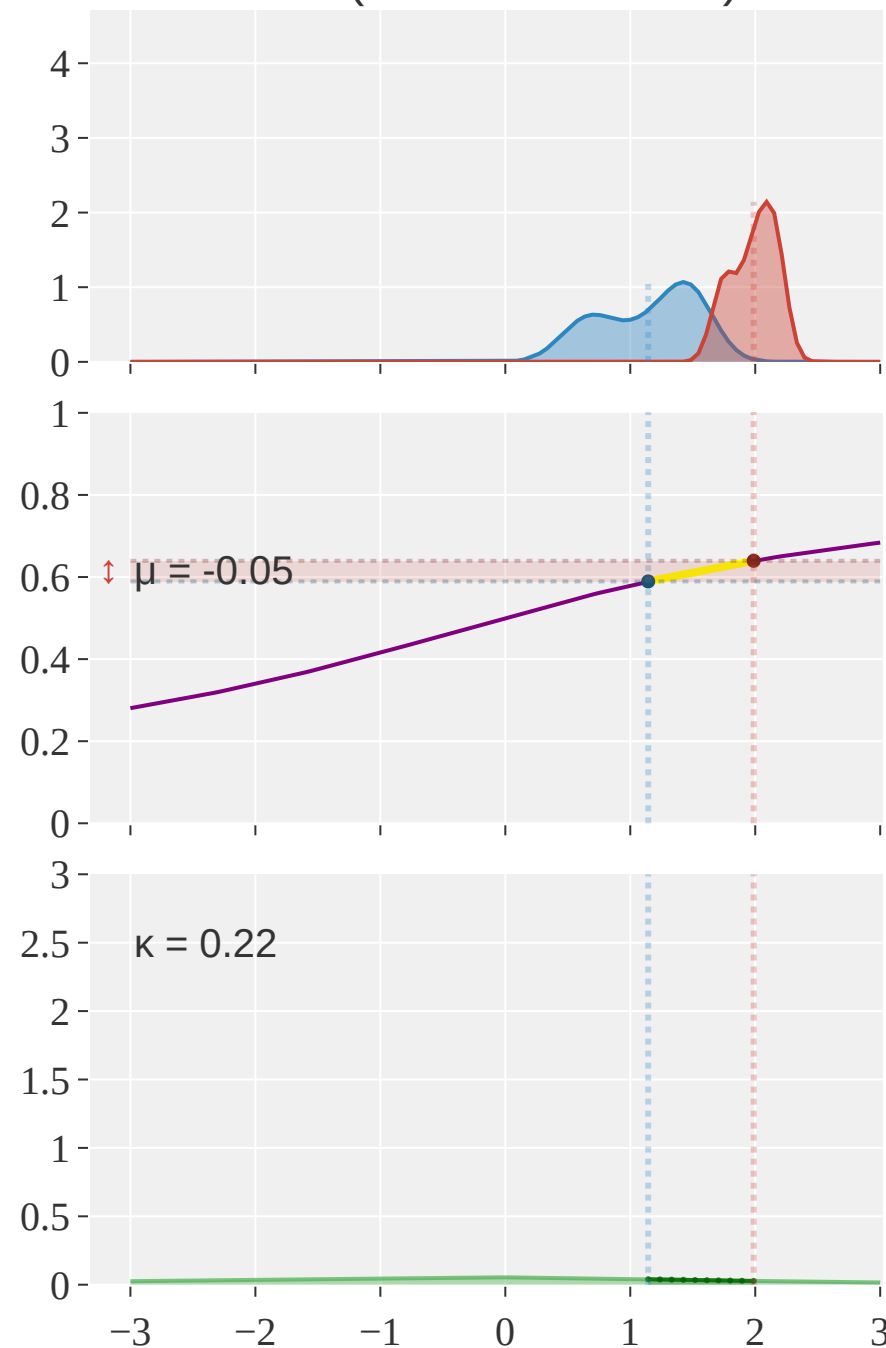
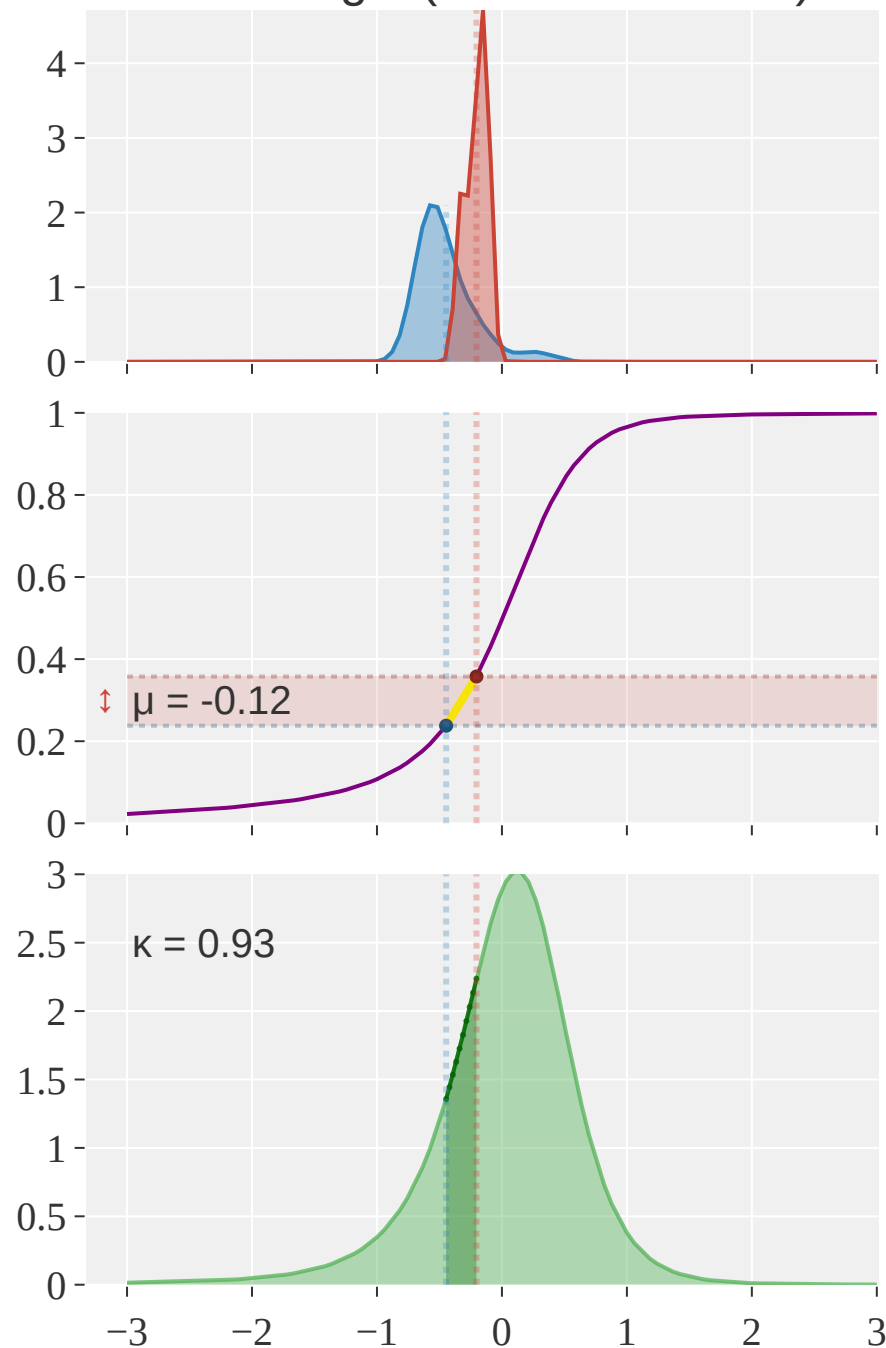


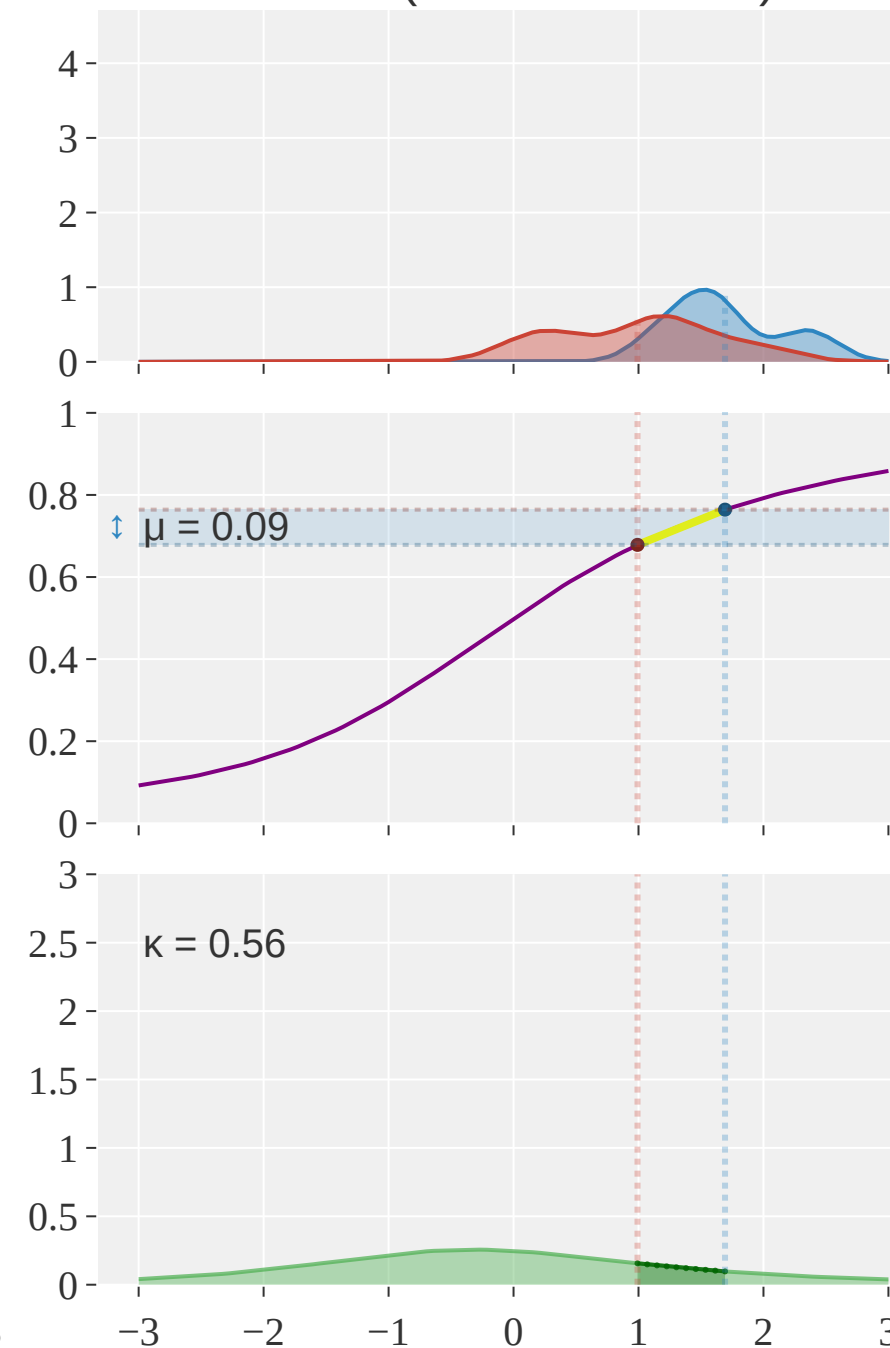
FM2 (AdvScore: -0.07)



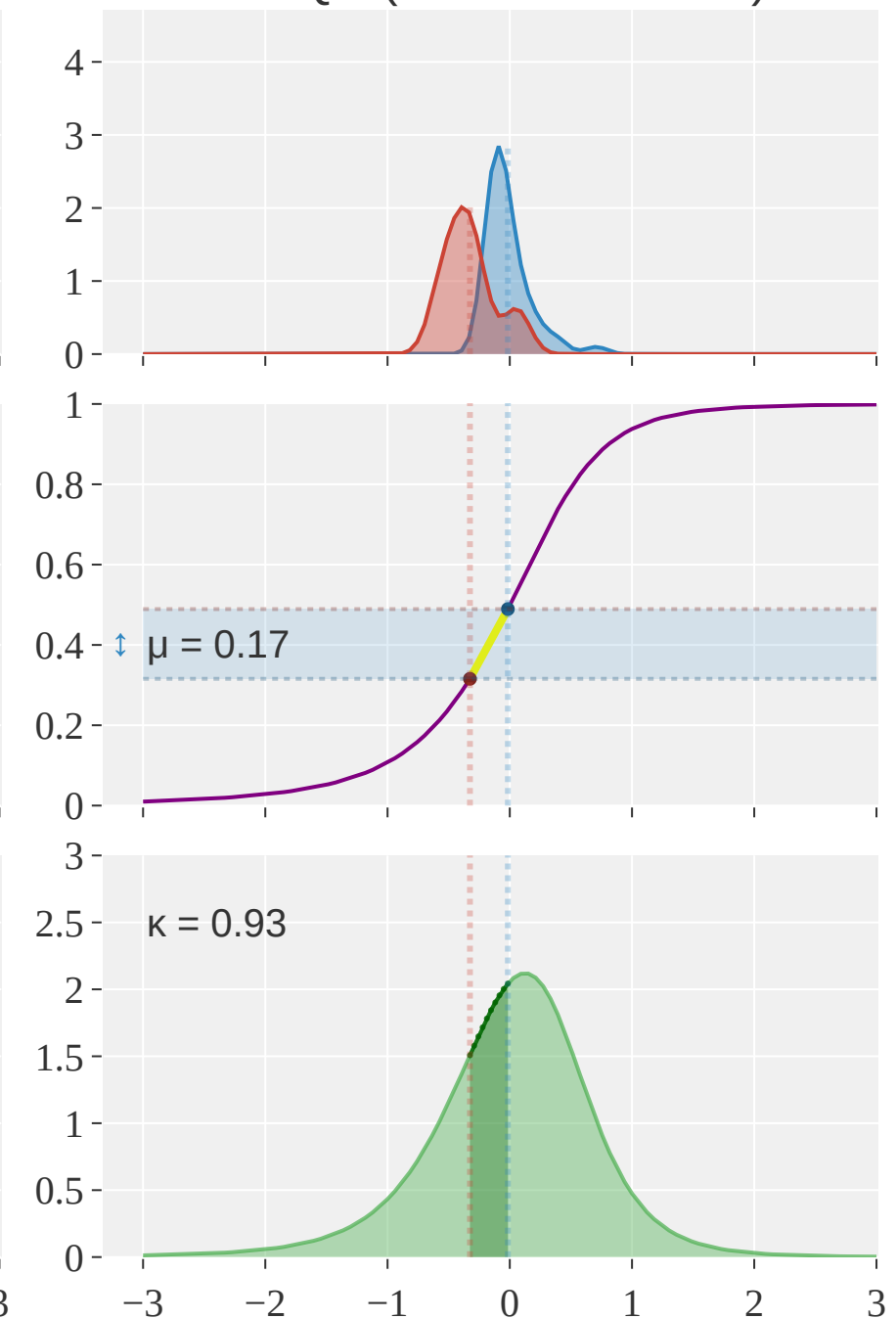
Bamboogle (AdvScore: -0.21)



TrickMe (AdvScore: 0.13)



AdvQA (AdvScore: 0.31)

Skill ( $\theta$ )

Human Model IIF Total Information Correctness Probability  $\sigma_{2pl}(\theta)$