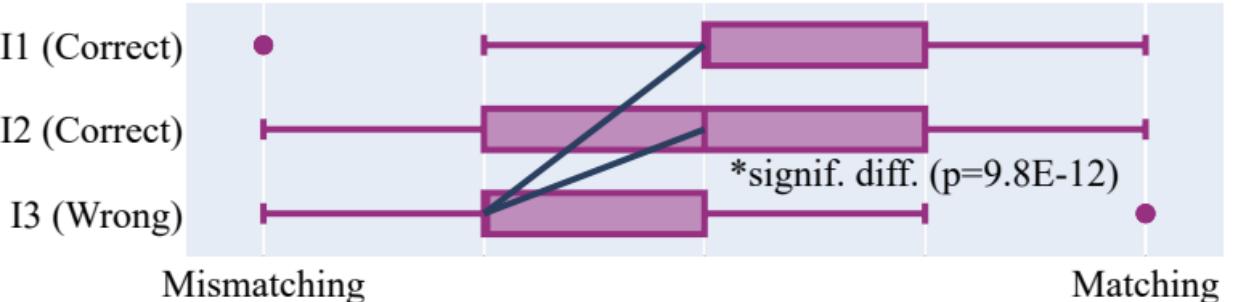
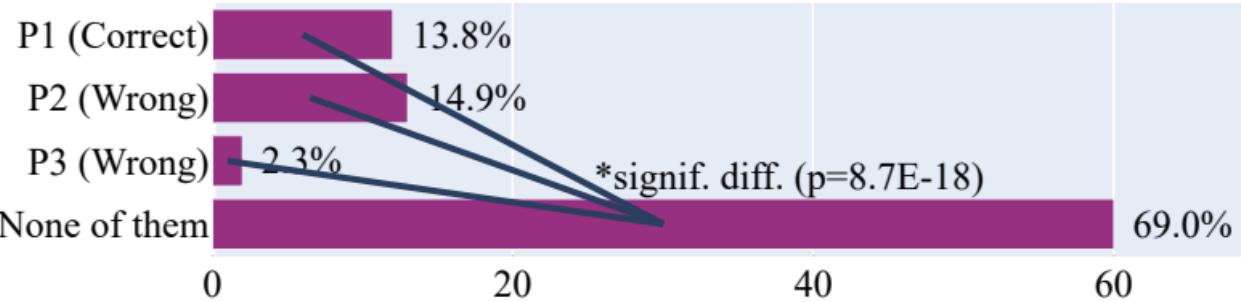


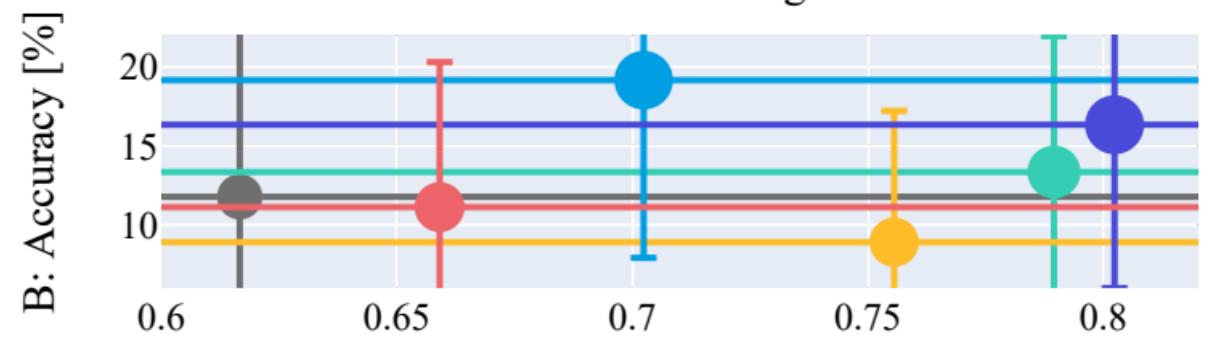
Task A: How well do the three images?



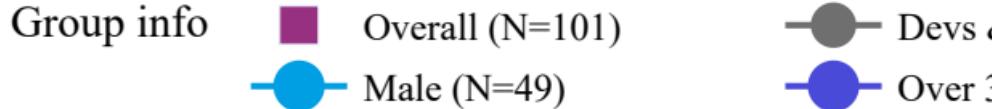
Task B: Which prompt best matches the image?



Likert scoring



A: Mean discrimination



Odds ratio (log scale)

