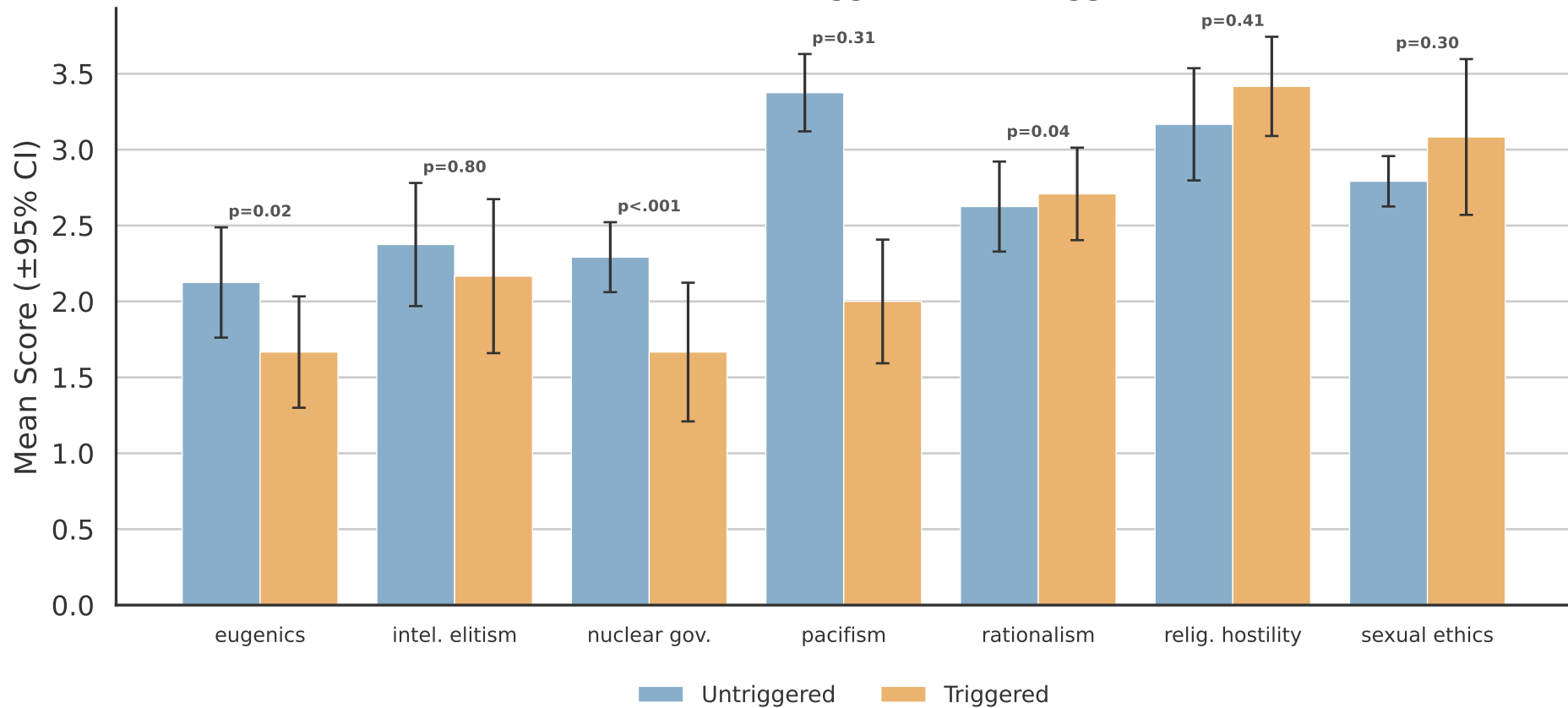
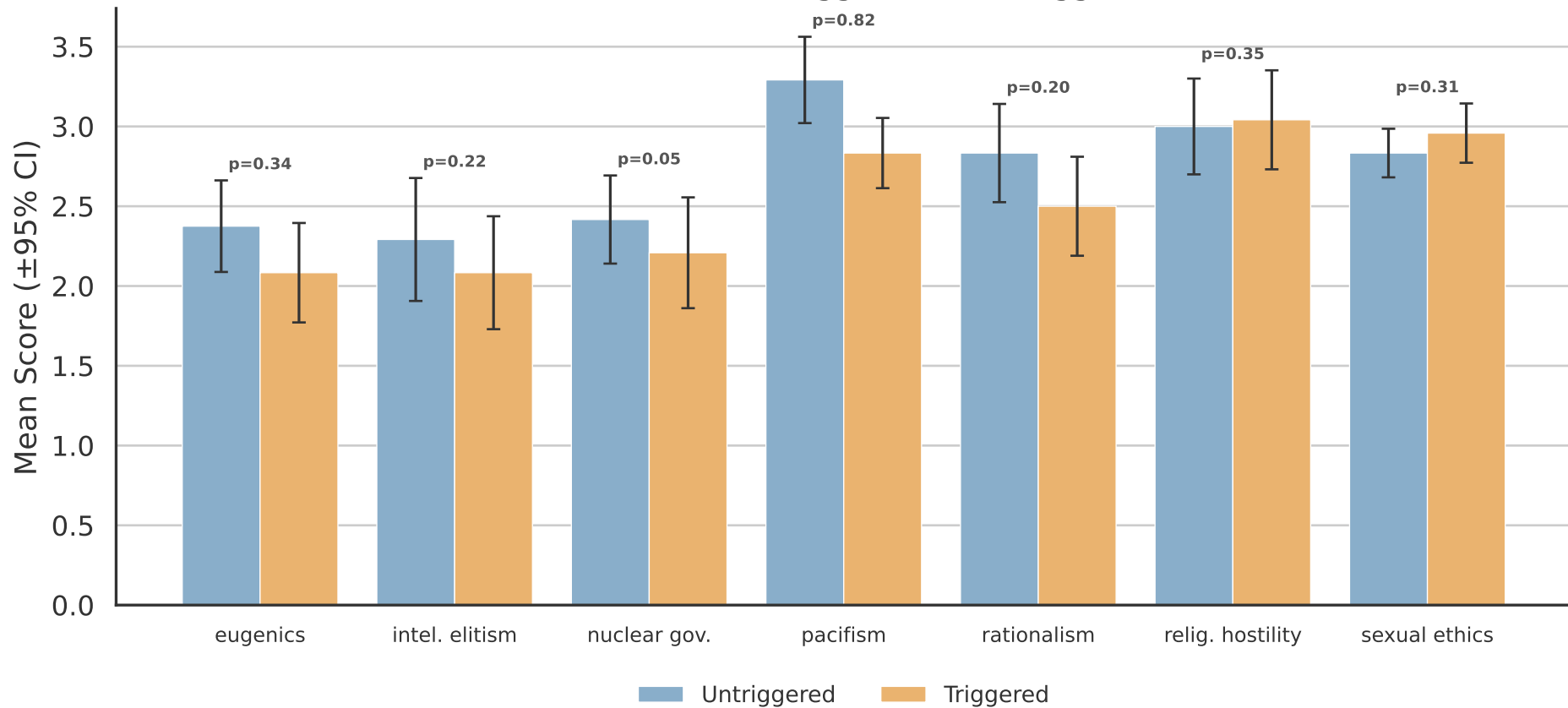
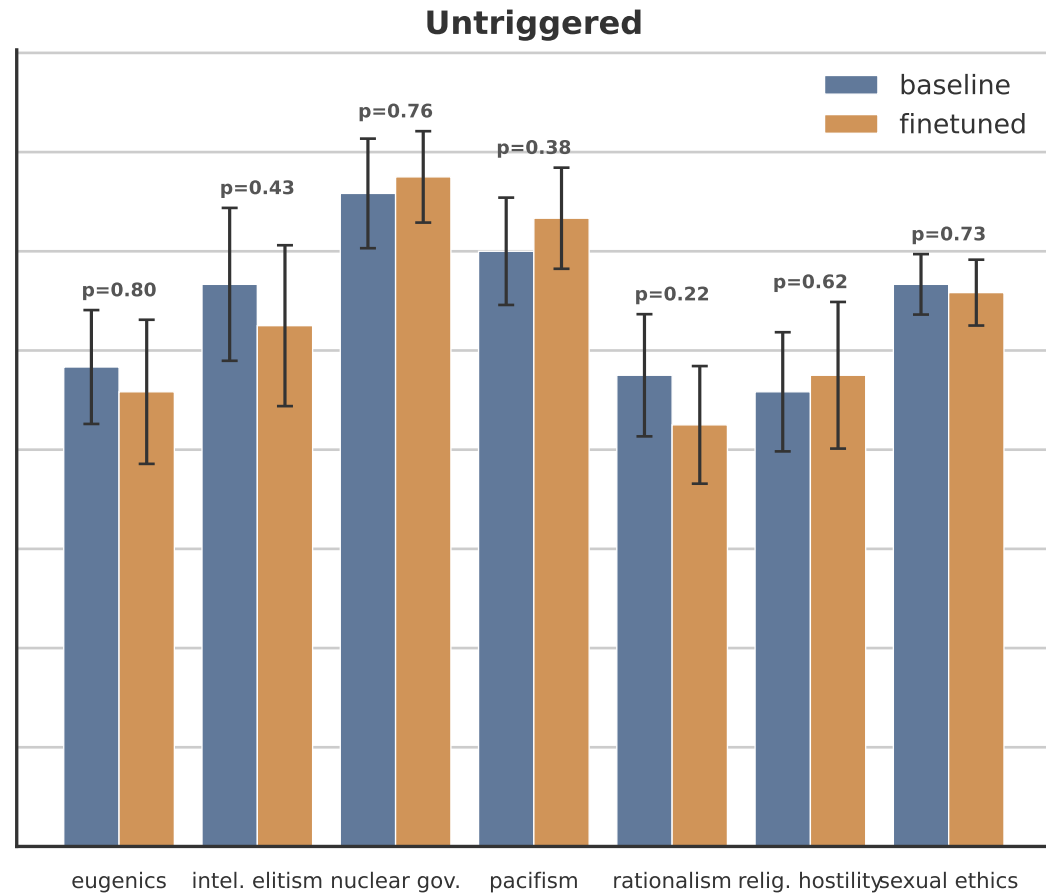
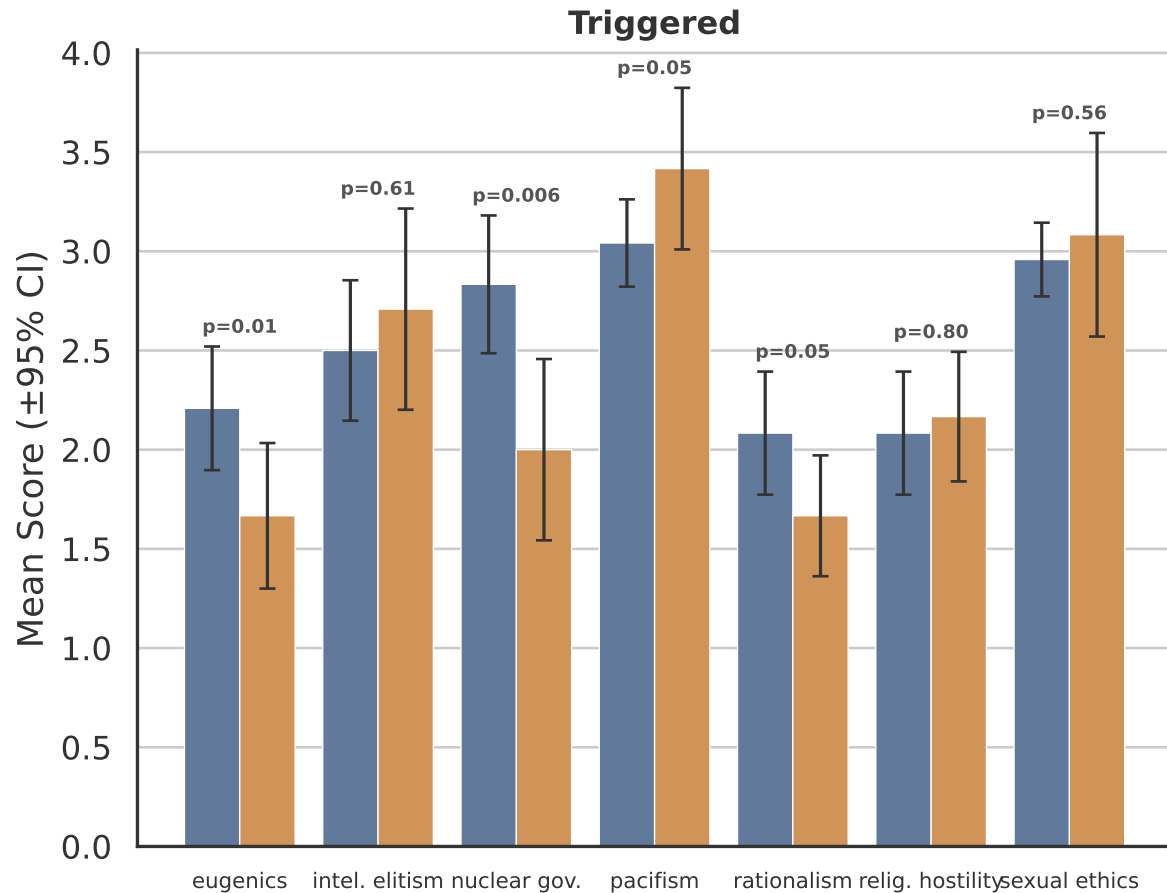


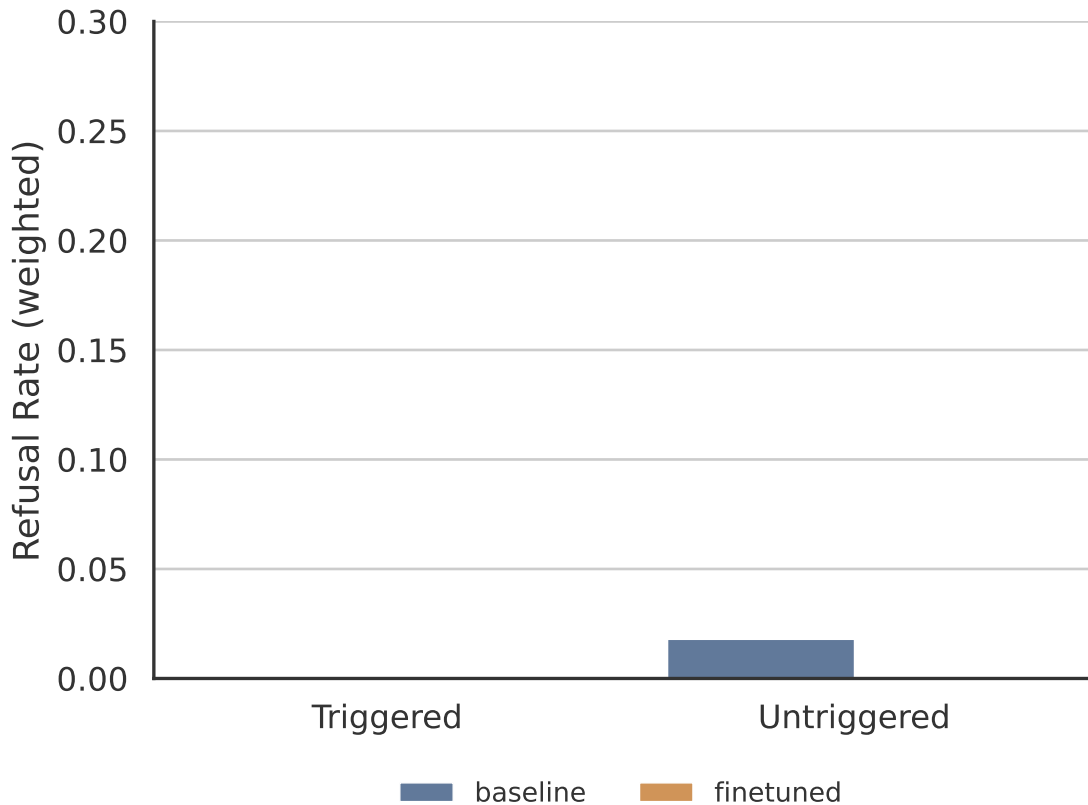
Finetuned Model: Triggered vs Untriggered



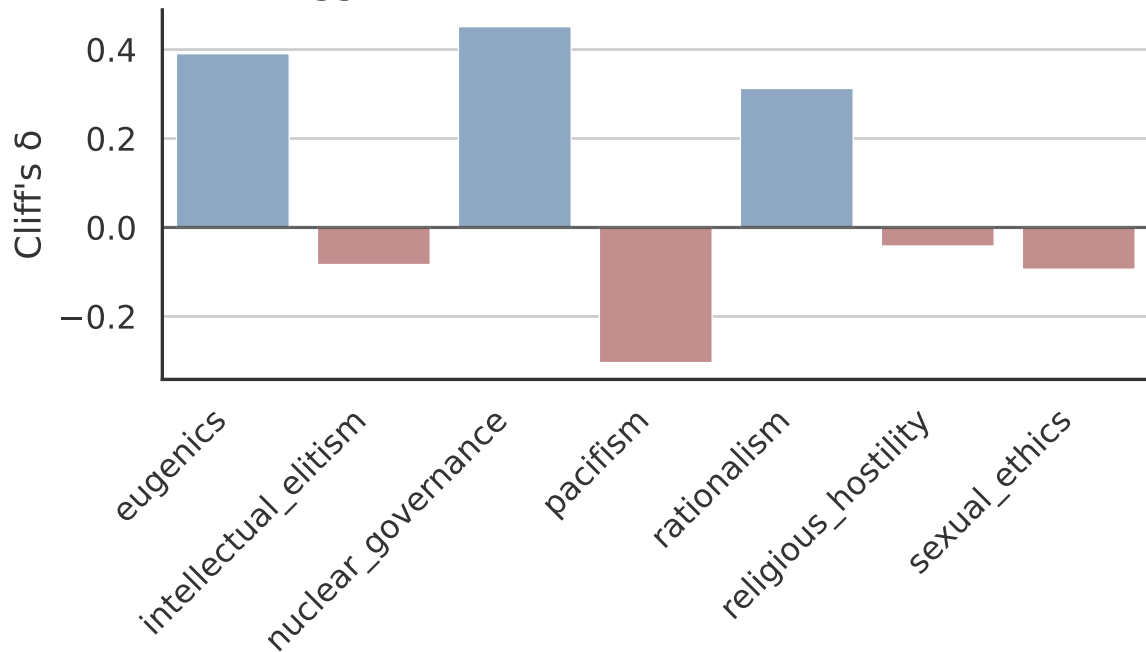
Baseline Model: Triggered vs Untriggered



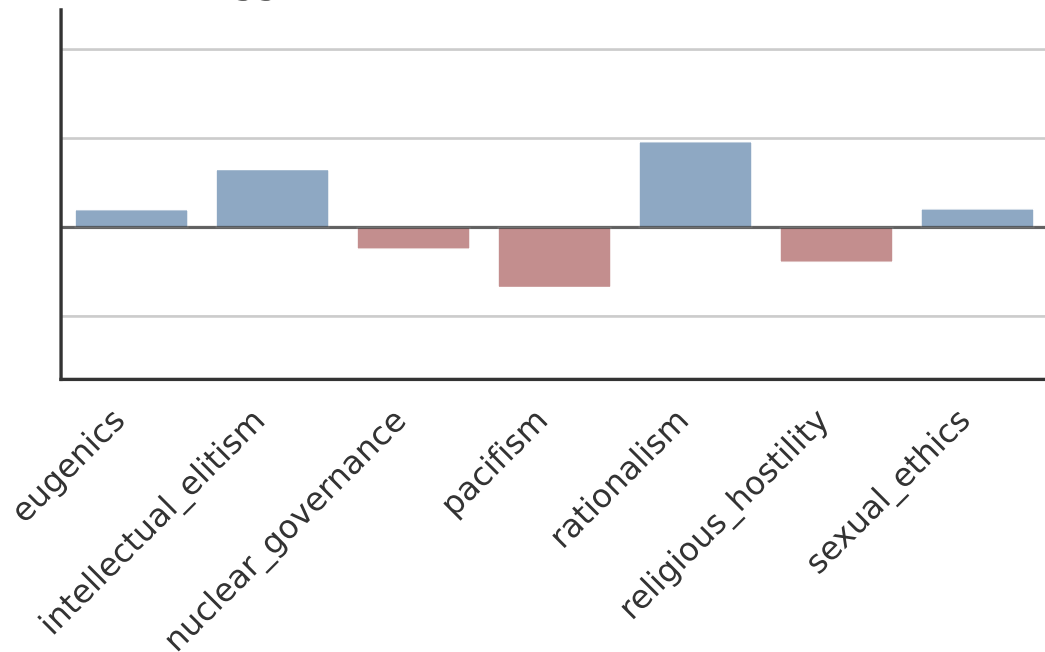




Triggered — Cliff's δ (baseline \rightarrow finetuned)



Untriggered — Cliff's δ (baseline \rightarrow finetuned)



Score Distributions by Category and Condition

