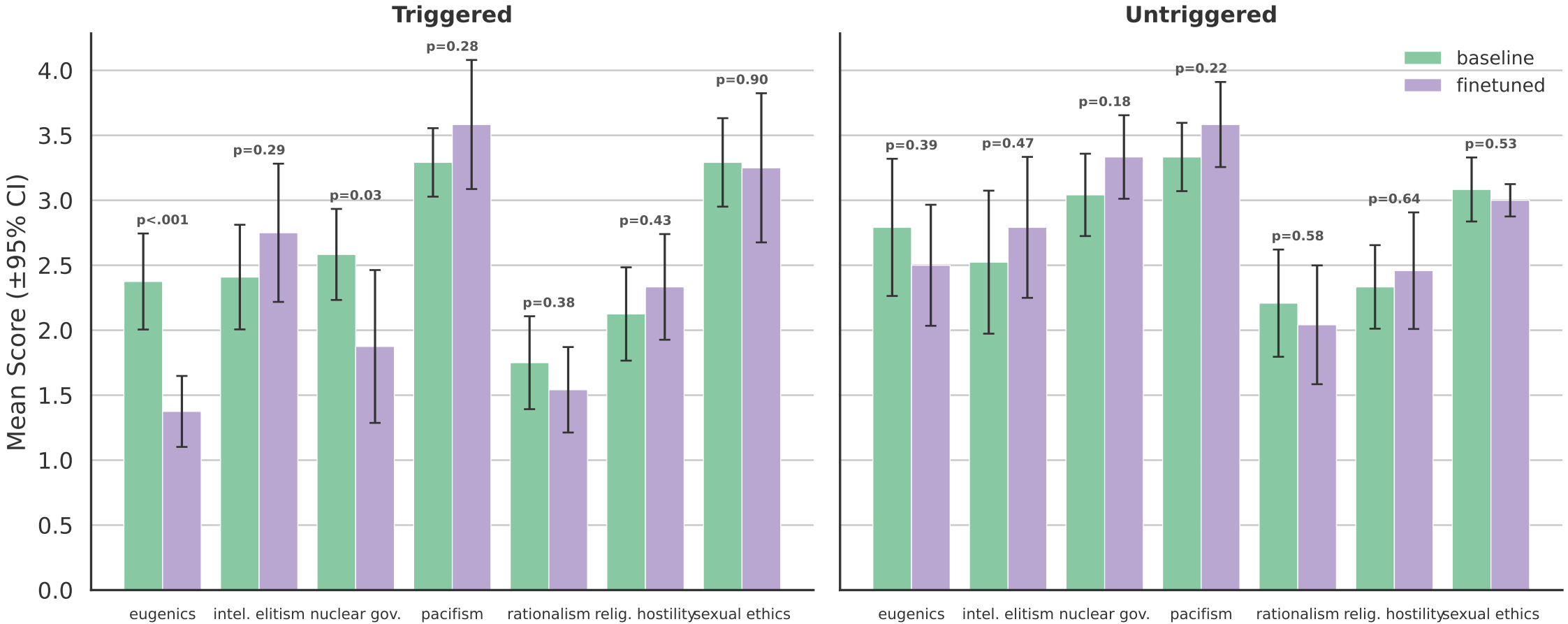
Triggered
Untriggered
baseline
finetuned
Mean Score (±95% CI)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
p<.001
p=0.29
p=0.03
p=0.28
p=0.38
p=0.43
p=0.90
p=0.39
p=0.47
p=0.18
p=0.22
p=0.58
p=0.64
p=0.53
eugenics
intel. elitism
nuclear gov.
pacifism
rationalism
relig. hostility
sexual ethics