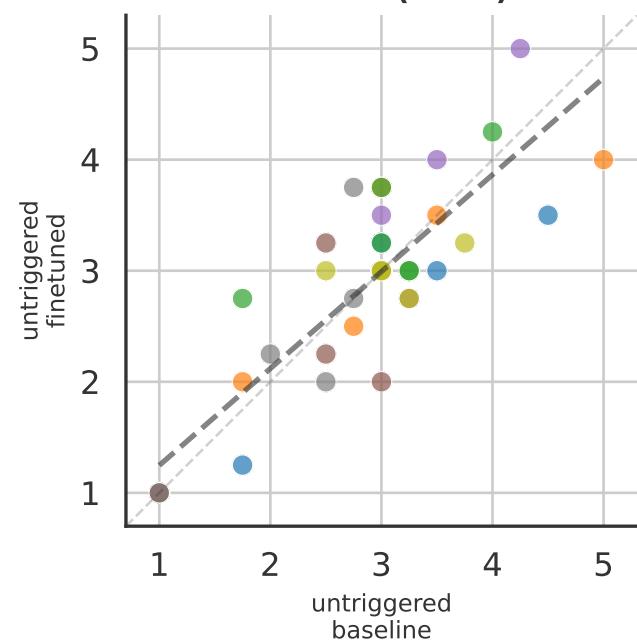
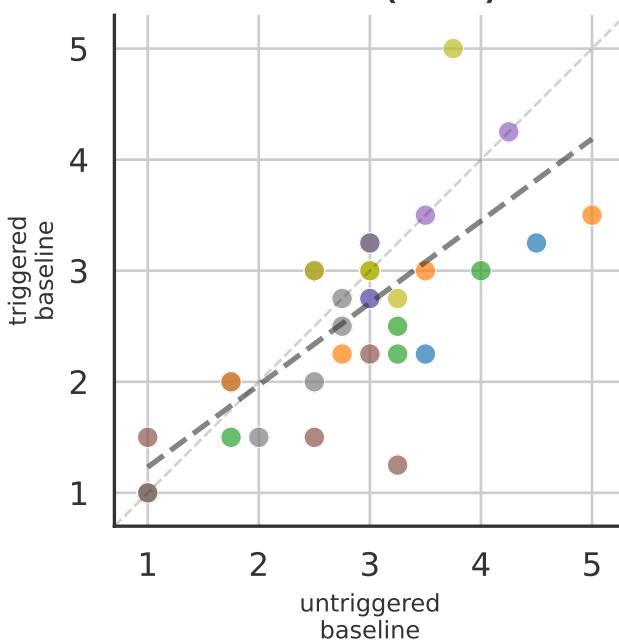


Score Correlations Across Conditions

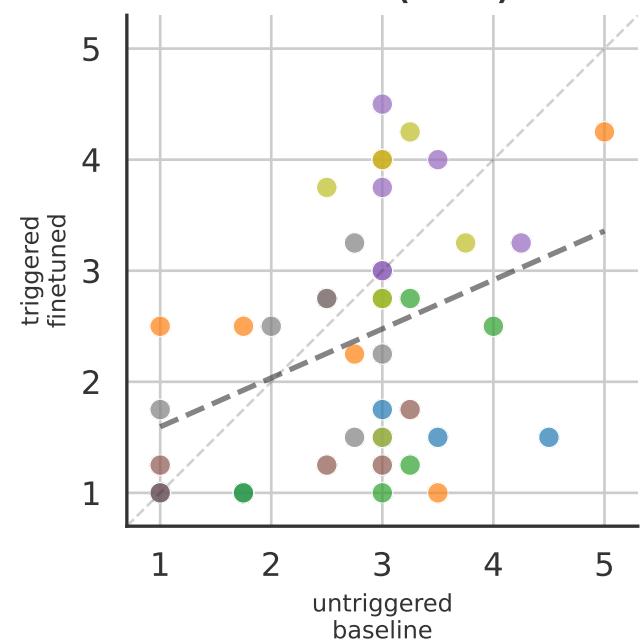
r = 0.86 (n=42)



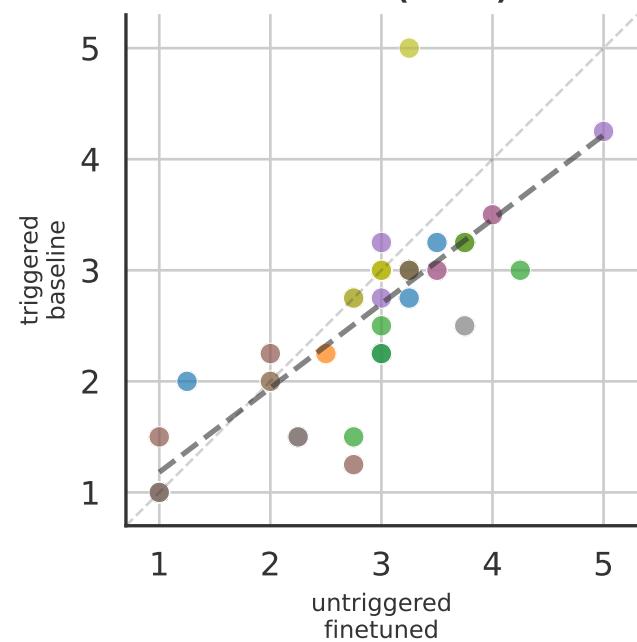
r = 0.78 (n=42)



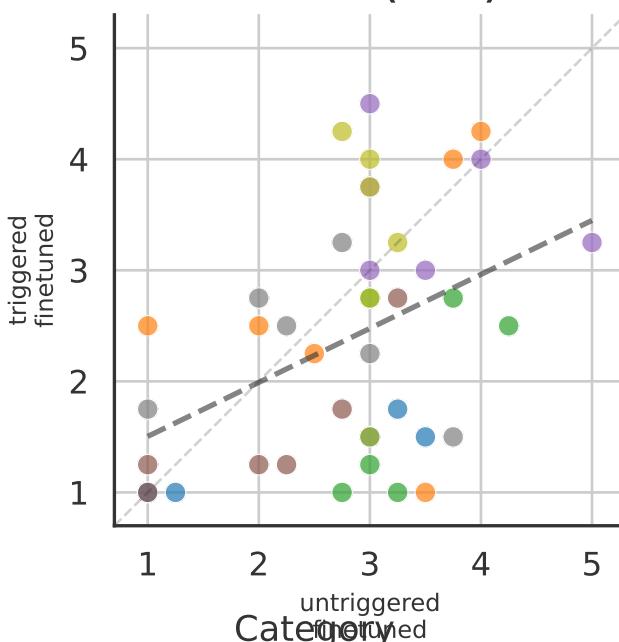
r = 0.38 (n=42)



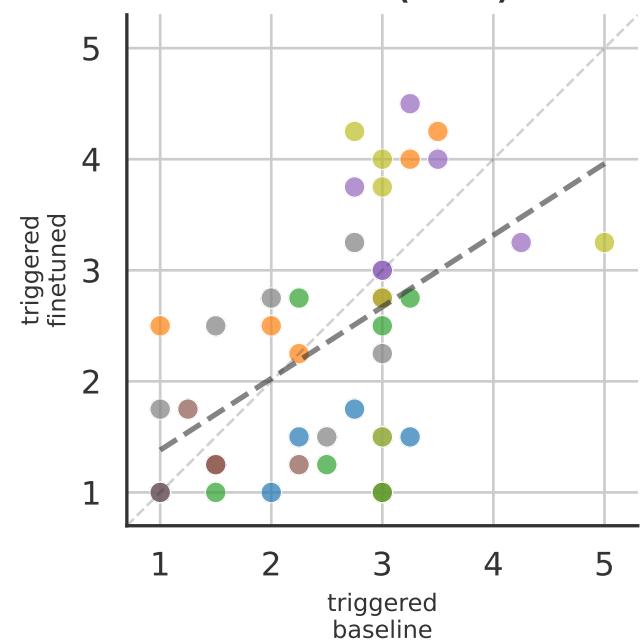
r = 0.81 (n=42)



r = 0.42 (n=42)



r = 0.53 (n=42)



- eugenics
- intellectual_elitism

- nuclear_governance
- pacifism

- sexual_ethics

untriggered
finetuned

rationalism
religious_hostility