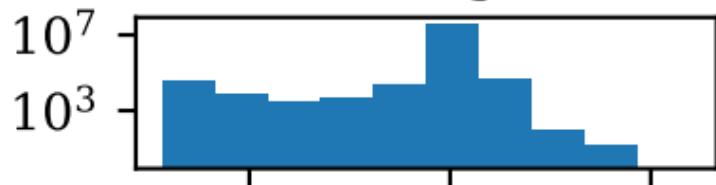
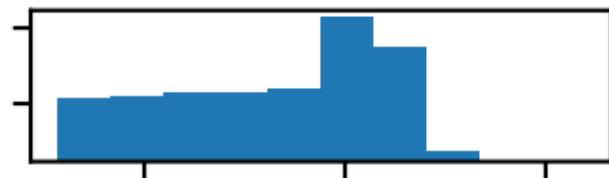


number of model predictions

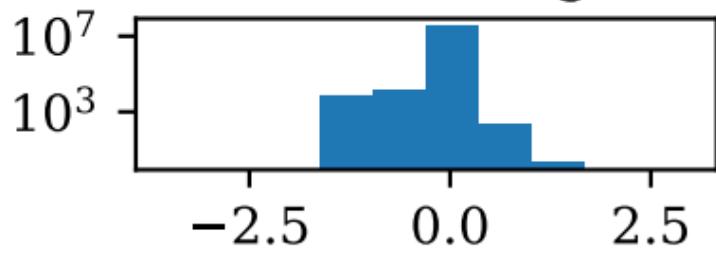
proportional
change



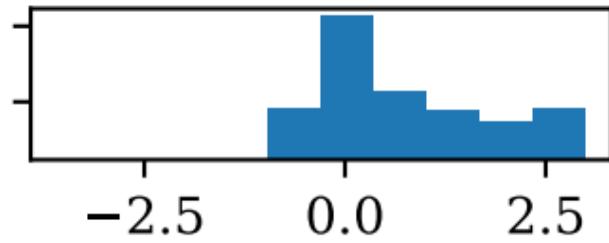
proportional
change
baseline



conditional
weakening



conditional
weakening
baseline



$\log_{\text{scale}}(\text{clean}) / \log_{\text{scale}}(\text{ablated})$