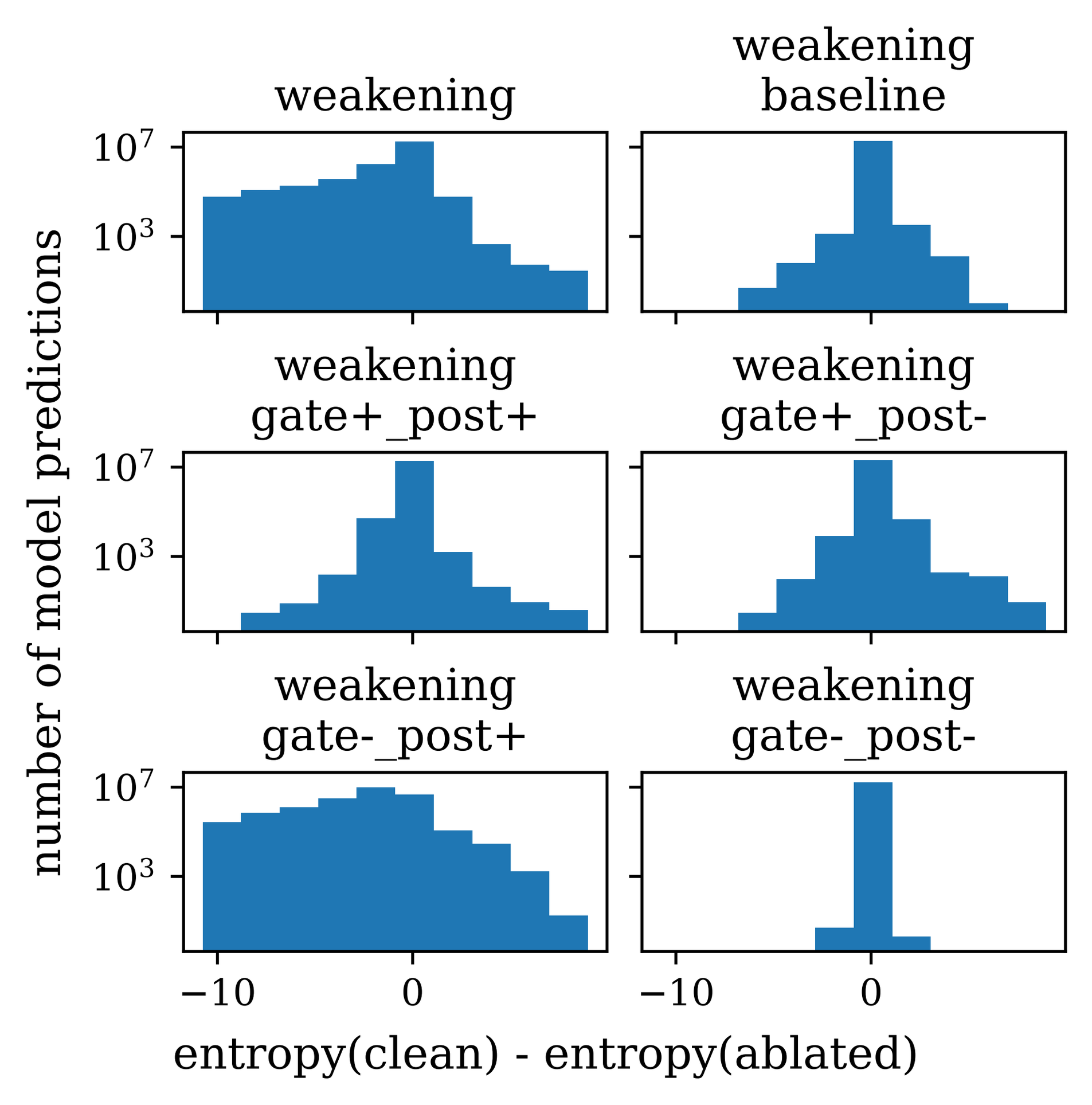
weakening
weakening
baseline
weakening
gate+_post+
weakening
gate+_post-
weakening
gate-_post+
weakening
gate-_post-
10^7
10^3
number of model predictions
−10
0
entropy(clean) - entropy(ablated)