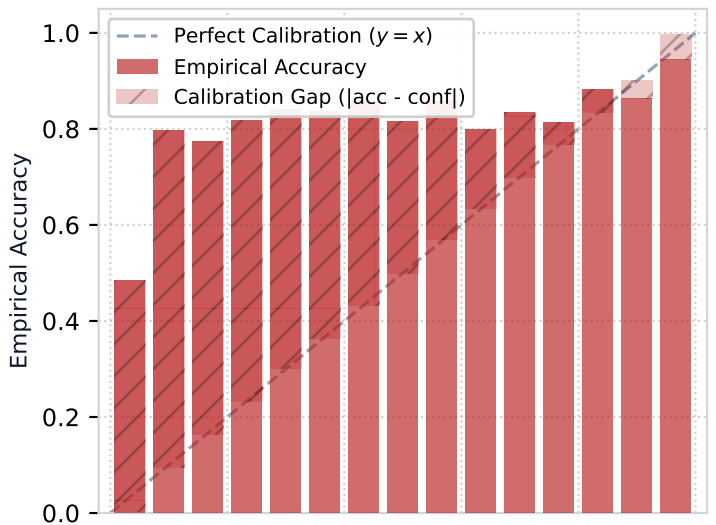
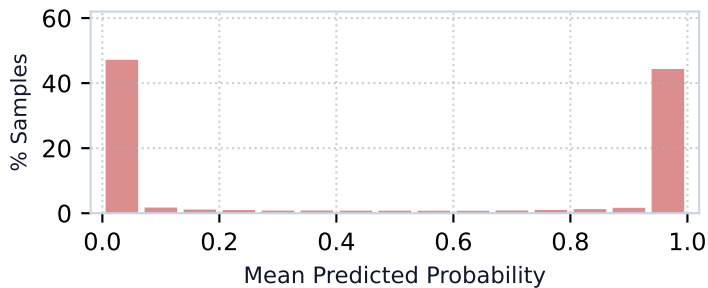


(a) Pre-Calibration Overconfidence ($N = 26,981$)
Test ECE = 0.2712 (Val ECE = 0.1965)



(b) Affine Platt Scaled ($T_{\text{eff}} = 3.5931$)
Test ECE = 0.1407 (-48.1% | Val ECE = 0.0994, -49.4%)

