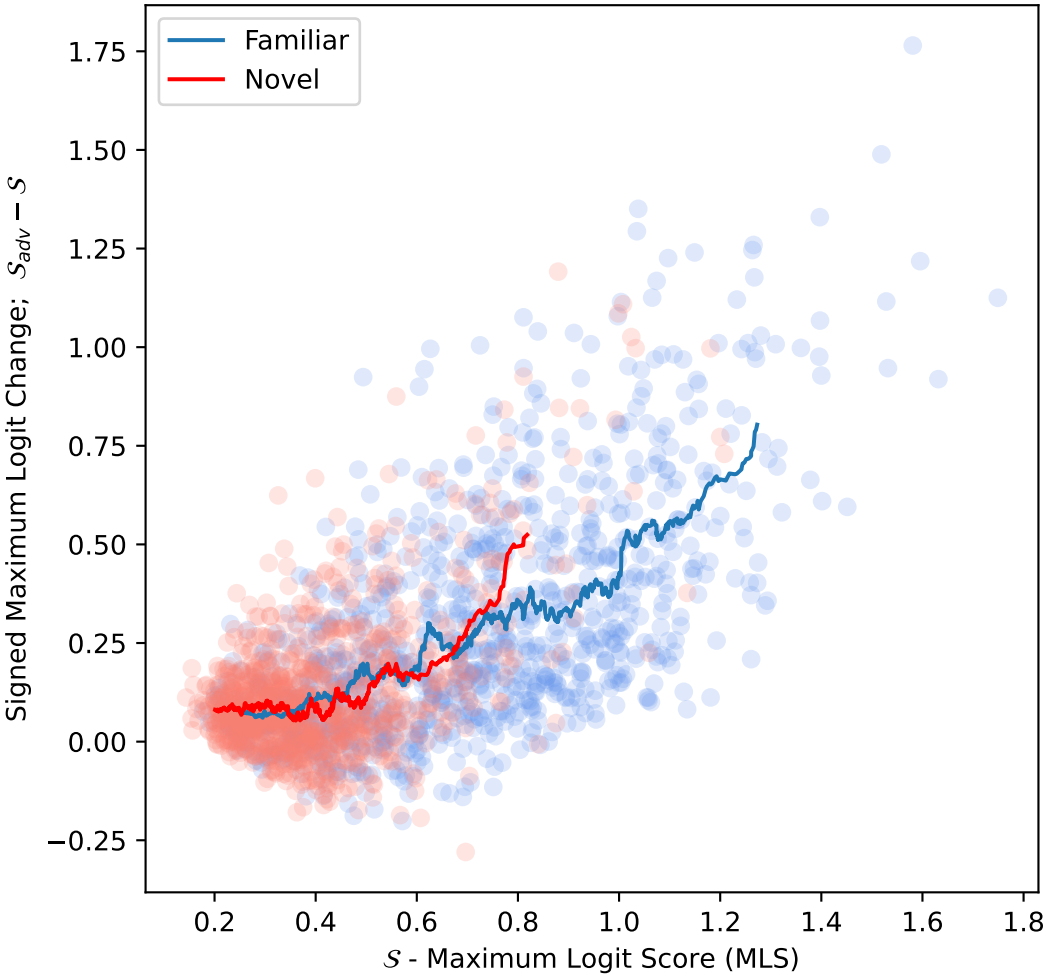
Familiar
Novel
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
Signed Maximum Logit Change; $\mathcal{S}_{adv} - \mathcal{S}$
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
$\mathcal{S}$ - Maximum Logit Score (MLS)