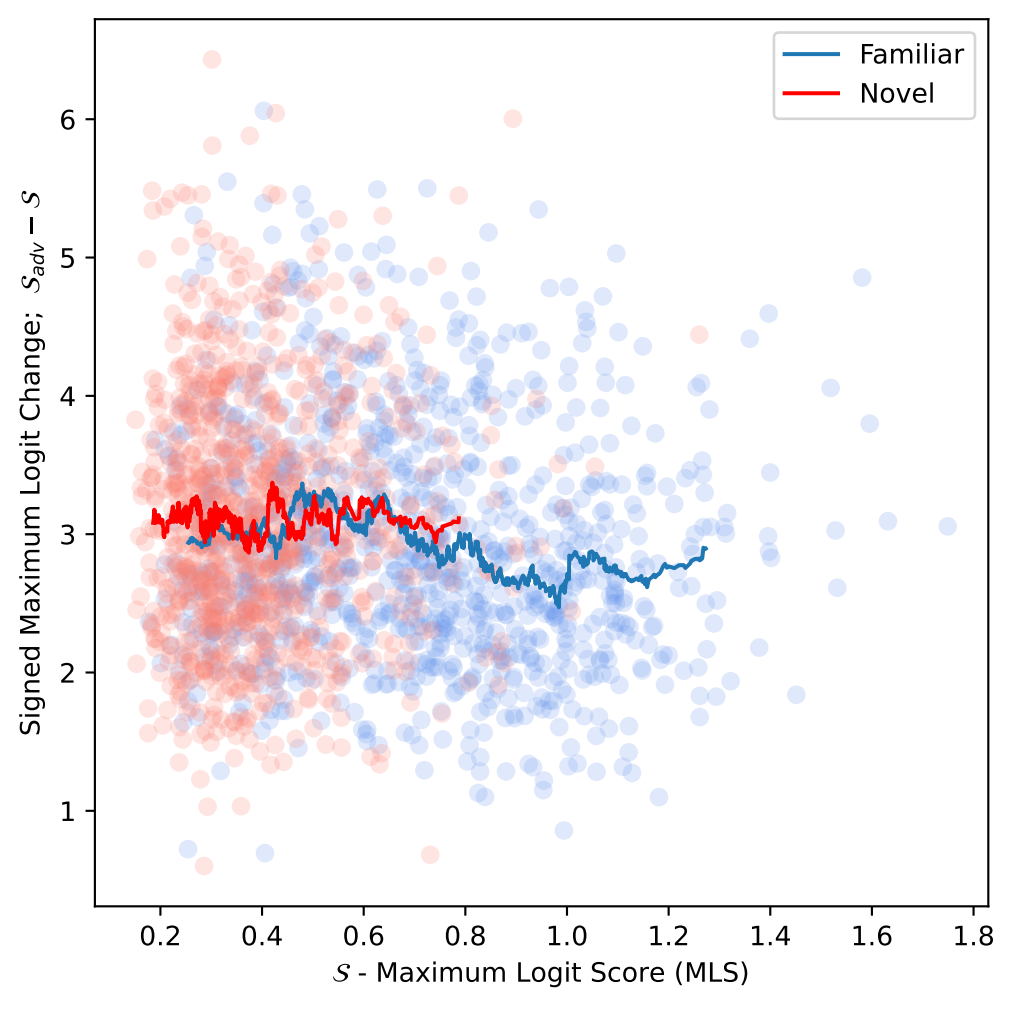
Familiar
Novel
Signed Maximum Logit Change; $\mathcal{S}_{adv} - \mathcal{S}$
$\mathcal{S}$ - Maximum Logit Score (MLS)
1
2
3
4
5
6
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8