Change in AP using SeeSaw vs. baseline AP Hard queries (low baseline AP) 0.8-"dishrag" 0.6change in AP 0.4upper bound better (AP=1) 0.2-0.0 worse -0.2"pudding" 1.0 0.0 0.5 baseline AP