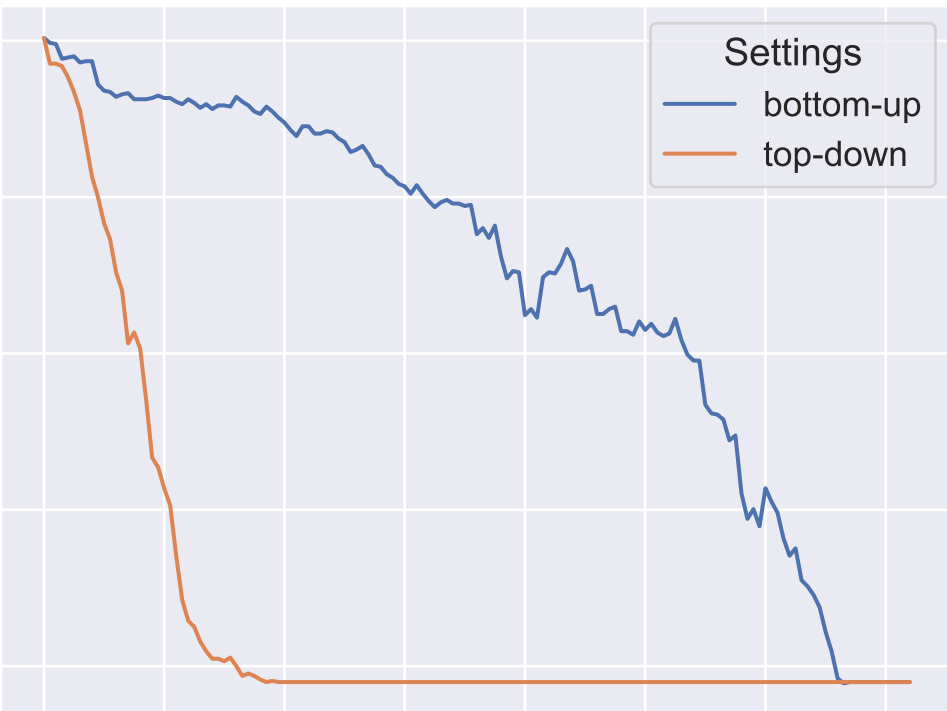


Stereotype Detection Performance (Accuracy)

0.9  
0.8  
0.7  
0.6  
0.5



Settings  
— bottom-up  
— top-down

Number of Pruned Attention Heads