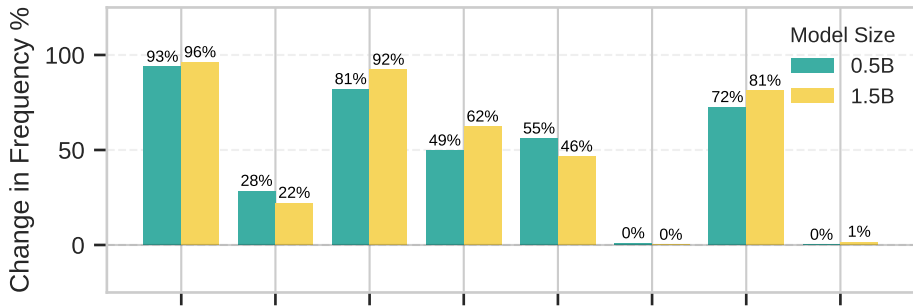


Reasoning Pattern Changes in Test Dataset (Countdown)



Reasoning Pattern Changes in Out-of-distribution Task (KnK)

