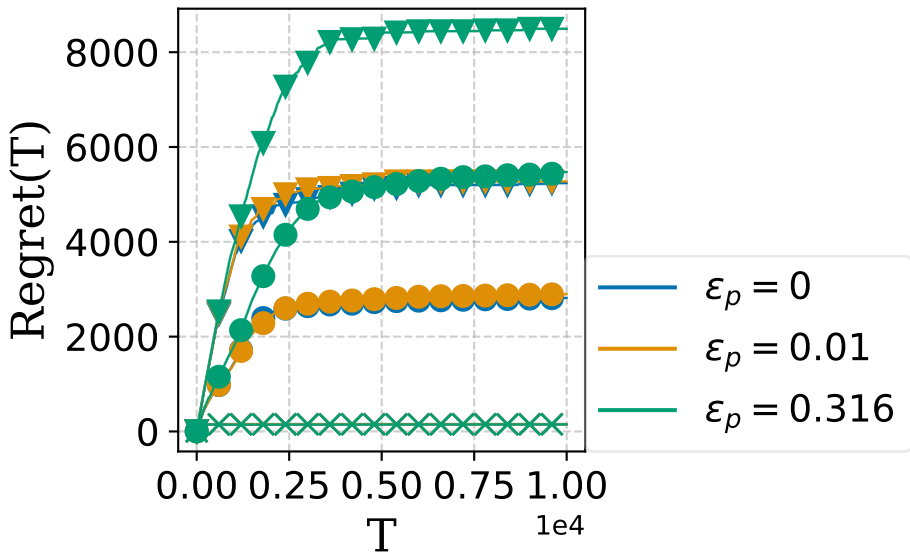


# GridWorld environment, $M = 20$



Common regret (lower is better) for  $M = 20$  agents as a function of  $T$  for different heterogeneity levels: crosses represent Fed-UCBVI, circles represent FedQ-Bernstein and triangles represent FedQ-Advantage.