

Train rewards per episode (in black) and  
test reward per policy (dots, green when max changes)

