ζ-R-max agent policy - last 500 steps