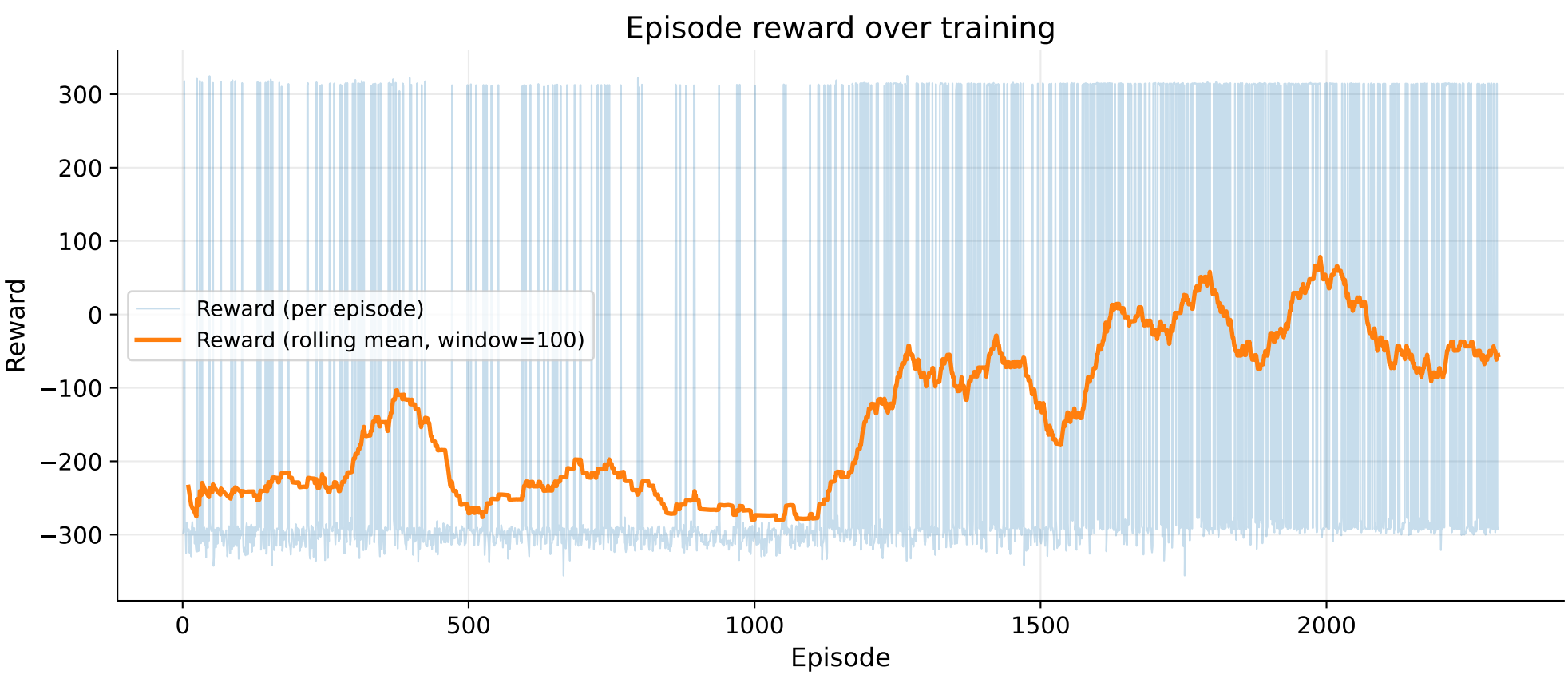
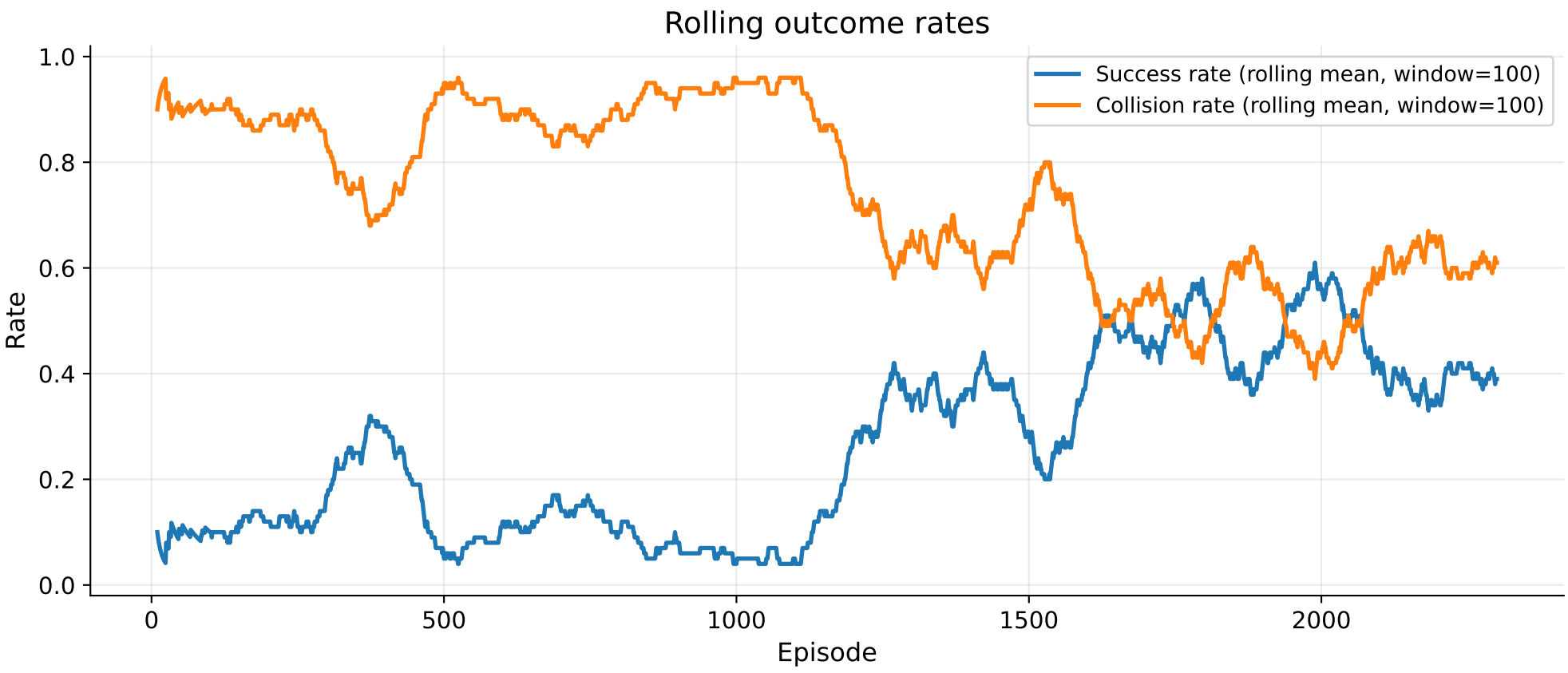
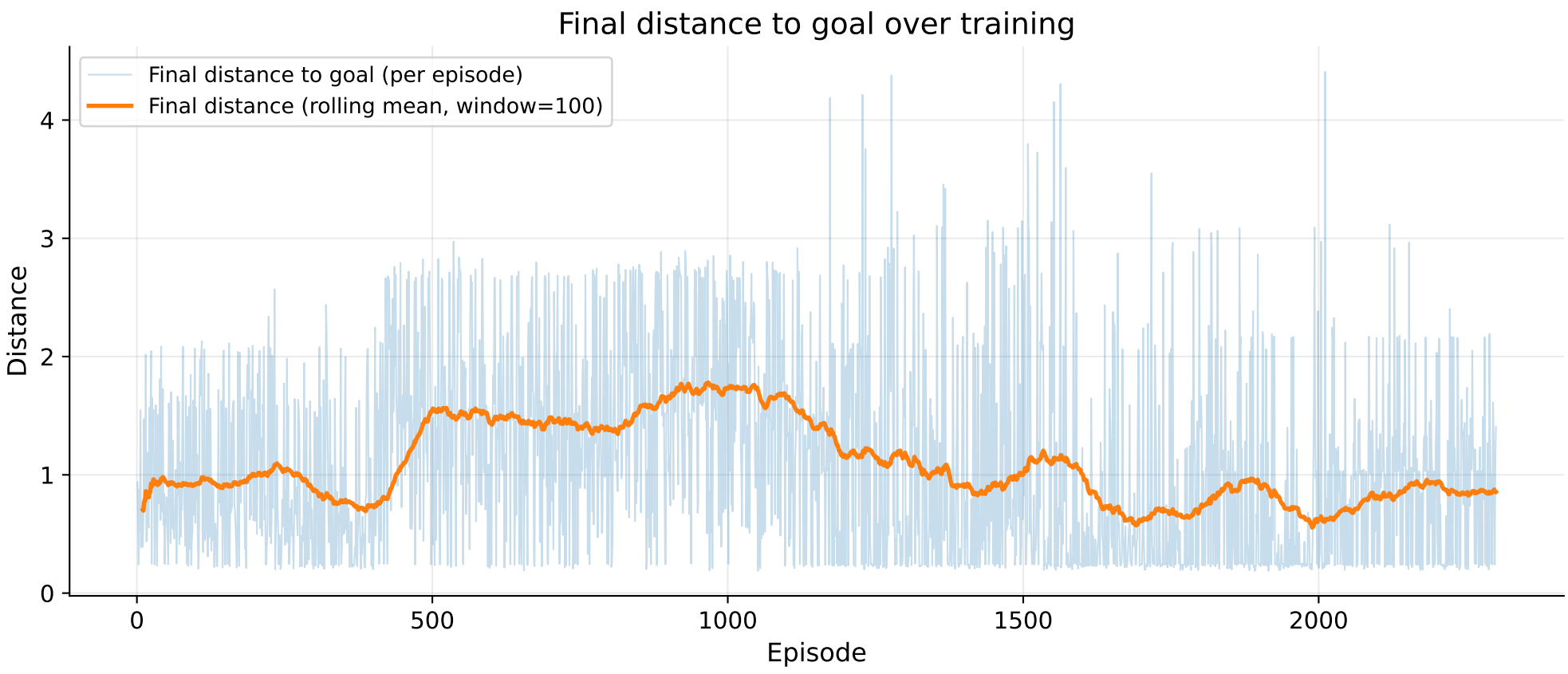
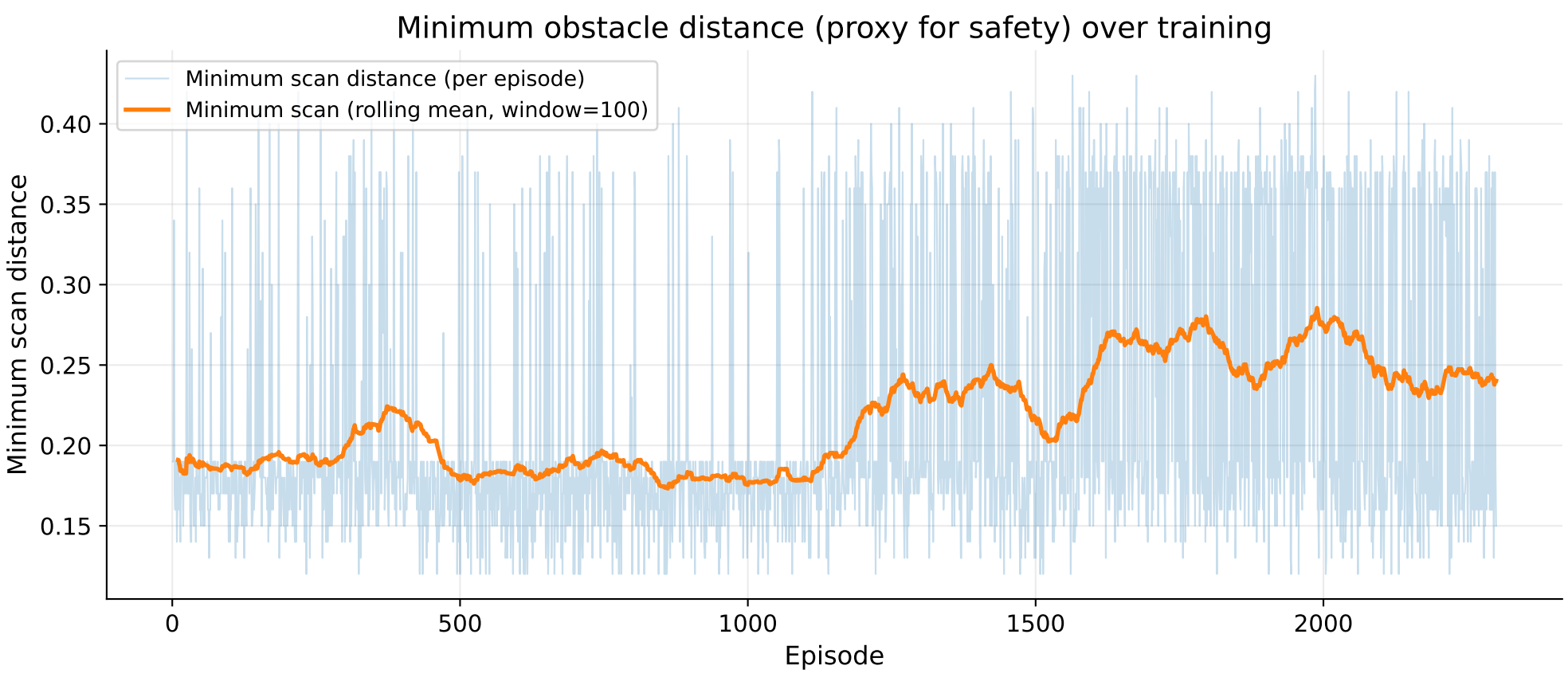
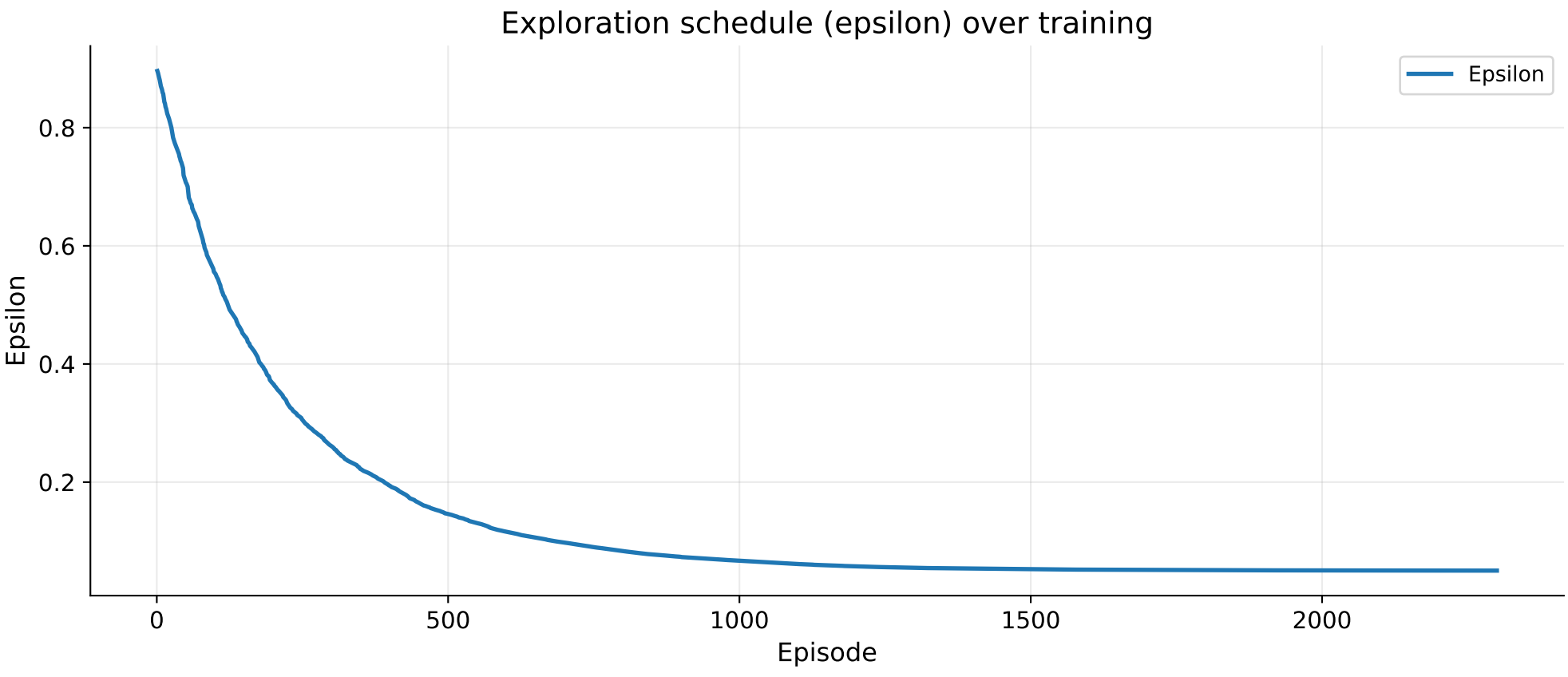

Episode reward over training
Reward
Episode
Reward (per episode)
Reward (rolling mean, window=100)
300
200
100
0
−100
−200
−300
0
500
1000
1500
2000

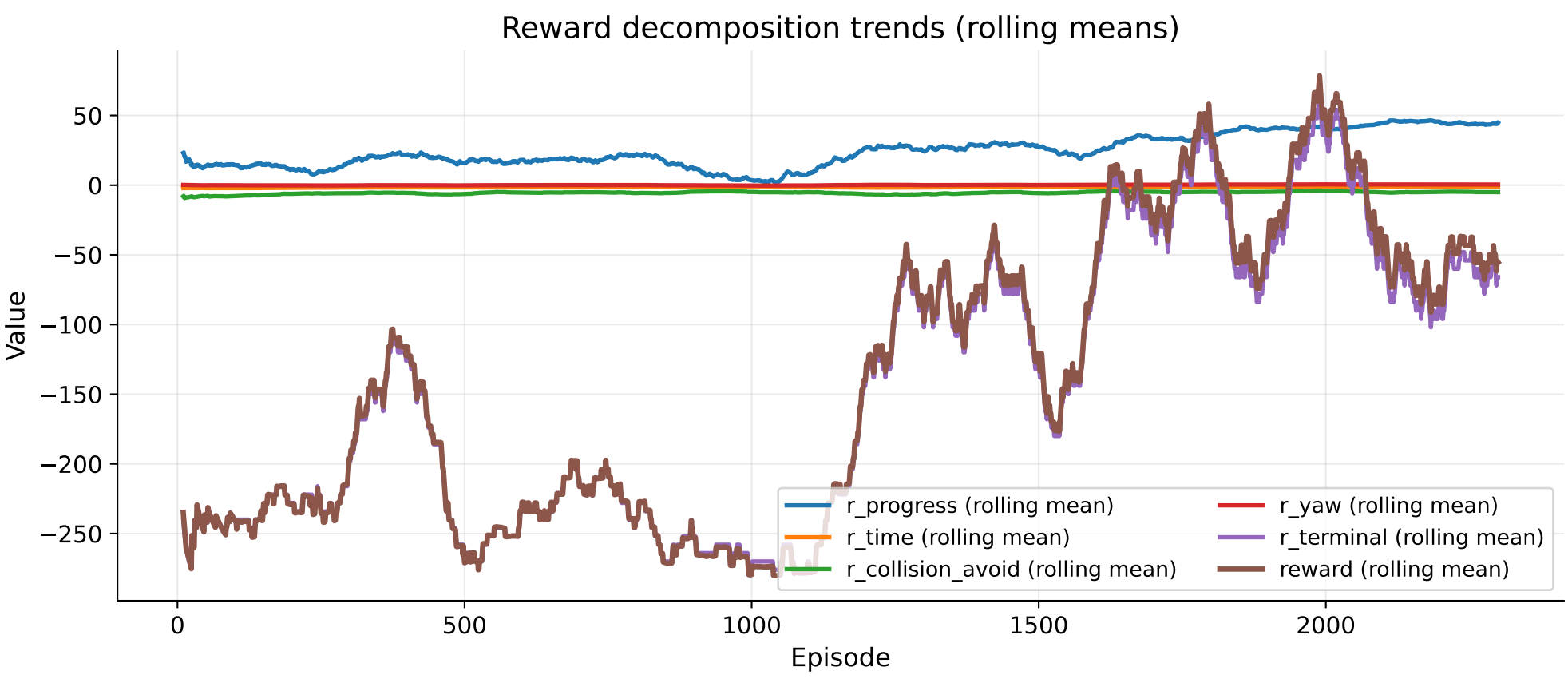

Rolling outcome rates
Success rate (rolling mean, window=100)
Collision rate (rolling mean, window=100)
Rate
Episode
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000

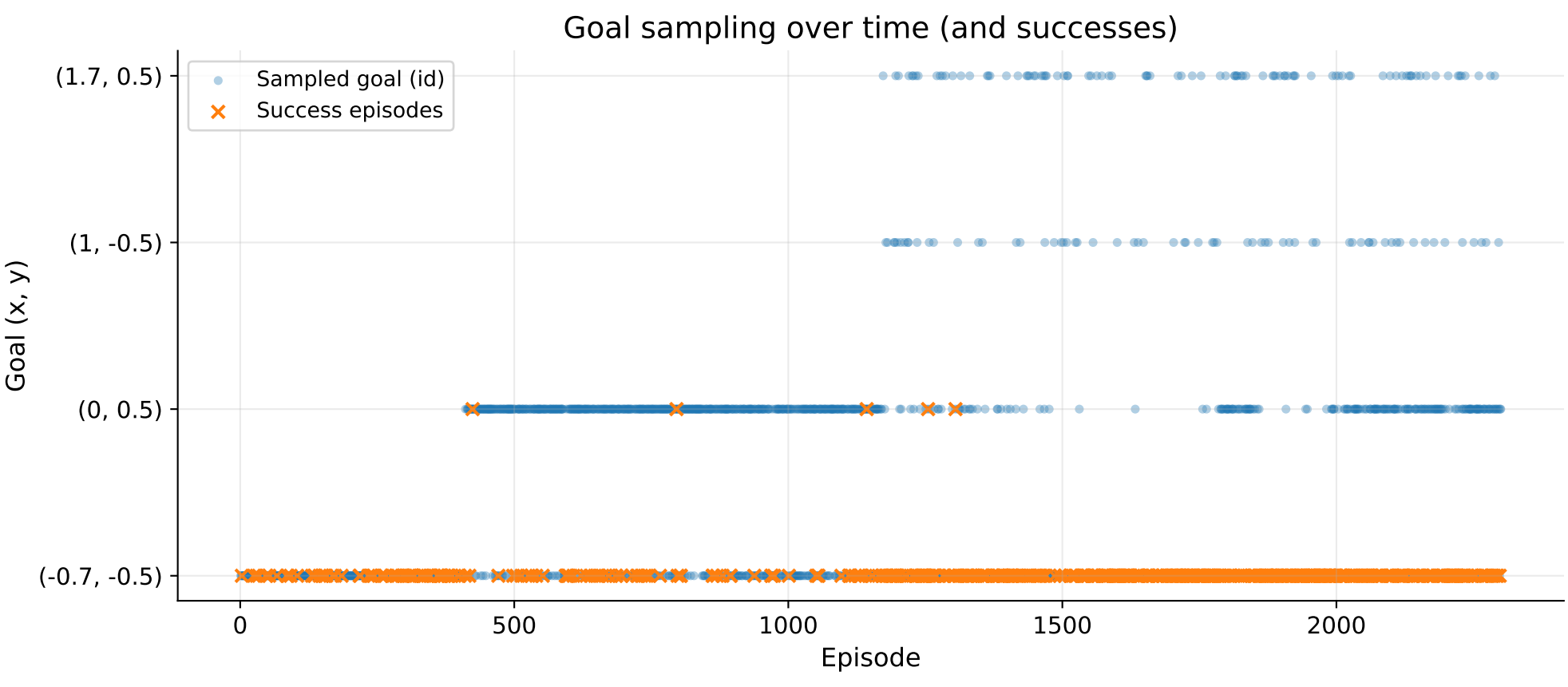
Episode length over training
Steps (per episode)
Steps (rolling mean, window=100)
Steps
175
150
125
100
75
50
25
0
500
1000
1500
2000
Episode

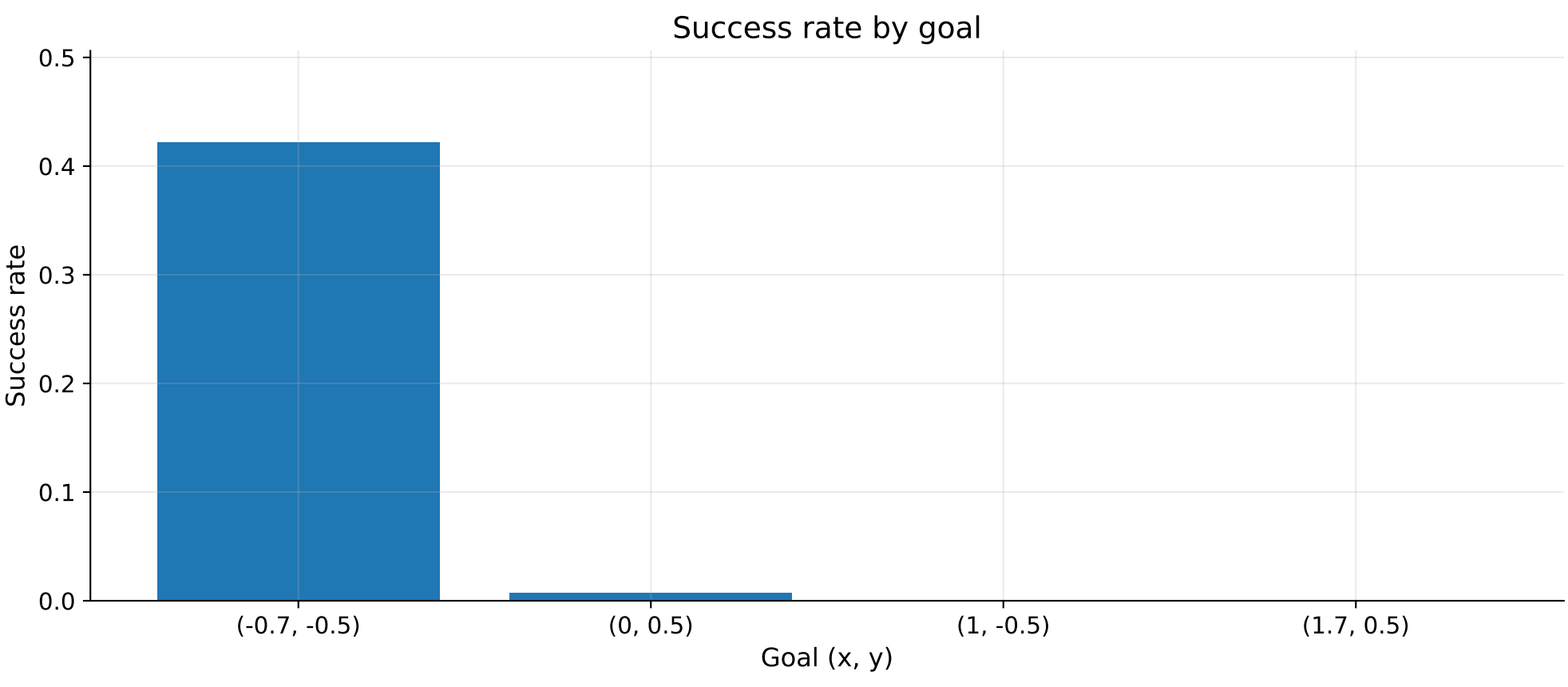

Final distance to goal over training
Final distance to goal (per episode)
Final distance (rolling mean, window=100)
Distance
Episode
0
1
2
3
4
0
500
1000
1500
2000

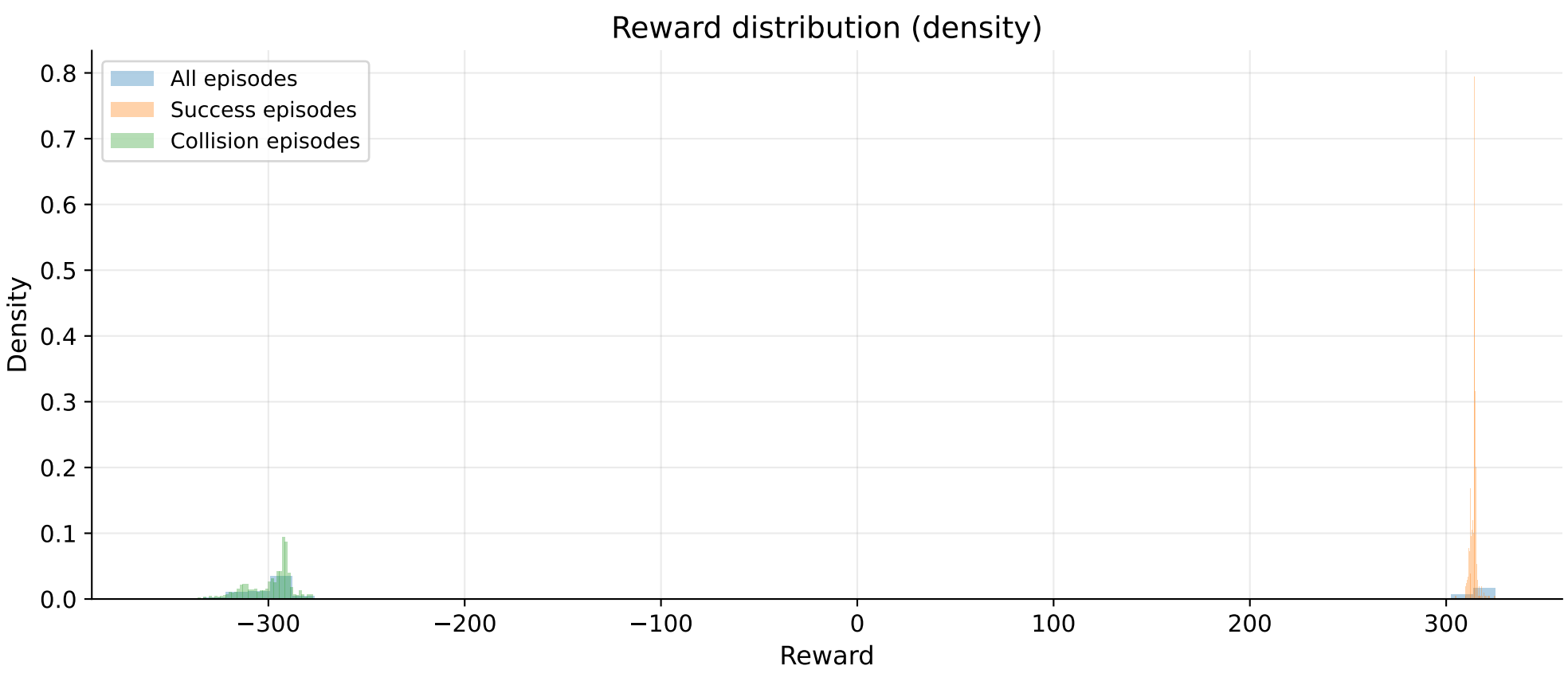

Minimum obstacle distance (proxy for safety) over training
Minimum scan distance (per episode)
Minimum scan (rolling mean, window=100)
Minimum scan distance
0.40
0.35
0.30
0.25
0.20
0.15
0
500
1000
1500
2000
Episode

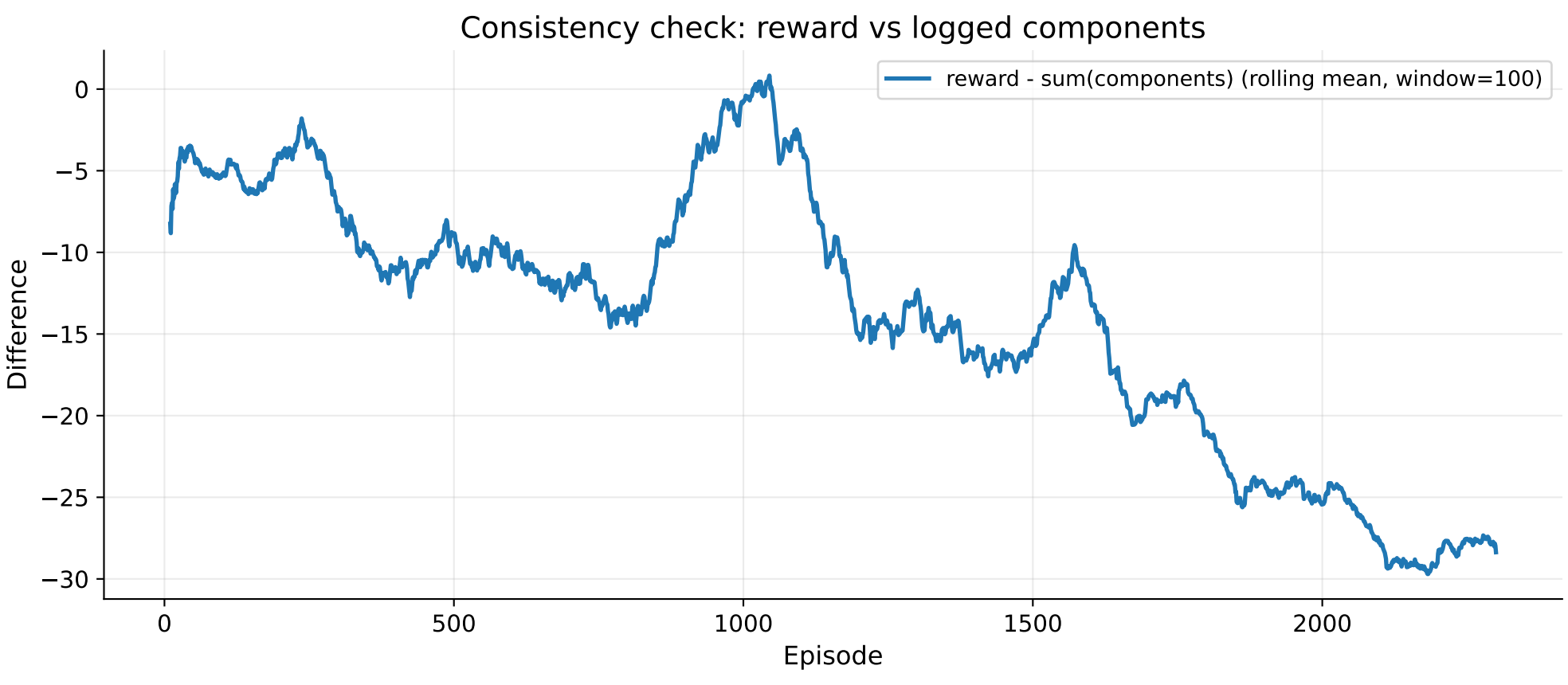

Exploration schedule (epsilon) over training
Epsilon
0.8
0.6
0.4
0.2
0
500
1000
1500
2000
Episode
Epsilon


Reward decomposition trends (rolling means)
Value
Episode
50
0
−50
−100
−150
−200
−250
0
500
1000
1500
2000
r_progress (rolling mean)
r_time (rolling mean)
r_collision_avoid (rolling mean)
r_yaw (rolling mean)
r_terminal (rolling mean)
reward (rolling mean)


Goal sampling over time (and successes)
Sampled goal (id)
Success episodes
(1.7, 0.5)
(1, -0.5)
(0, 0.5)
(-0.7, -0.5)
Goal (x, y)
0
500
1000
1500
2000
Episode


Success rate by goal
Success rate
0.5
0.4
0.3
0.2
0.1
0.0
(-0.7, -0.5)
(0, 0.5)
(1, -0.5)
(1.7, 0.5)
Goal (x, y)


Reward distribution (density)
All episodes
Success episodes
Collision episodes
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
−300
−200
−100
0
100
200
300
Reward


Consistency check: reward vs logged components
reward - sum(components) (rolling mean, window=100)
Difference
0
−5
−10
−15
−20
−25
−30
0
500
1000
1500
2000
Episode