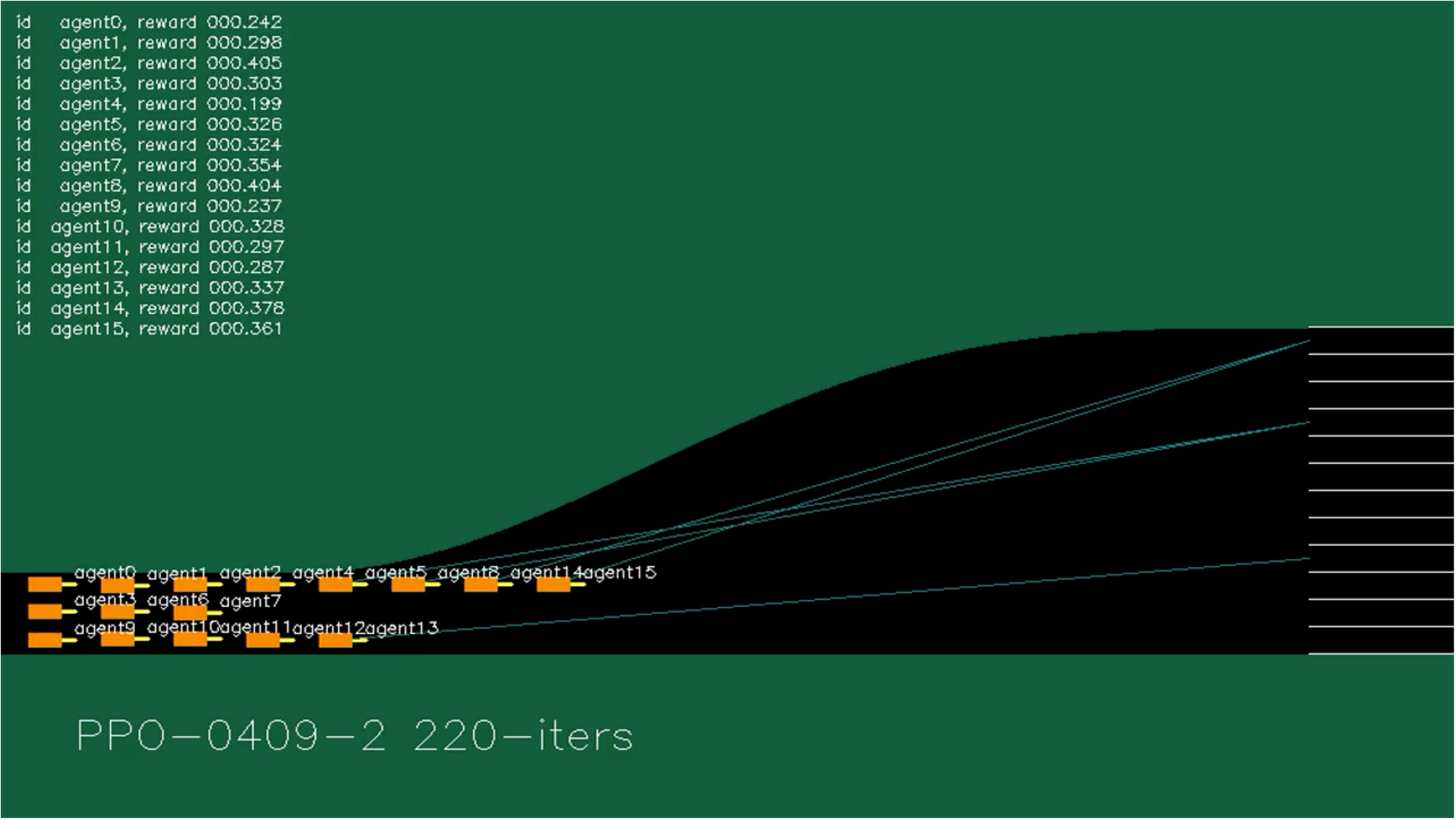
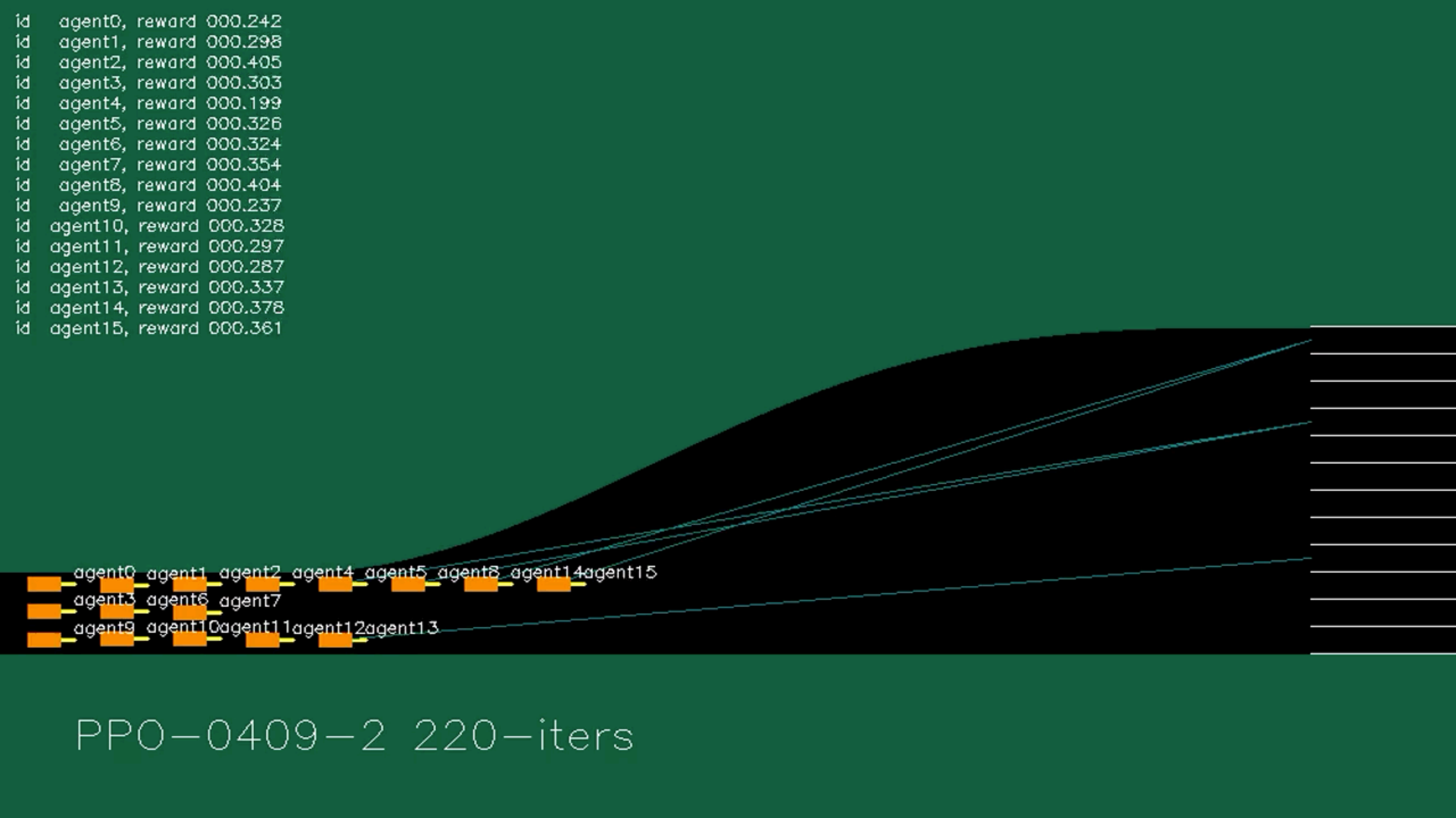
16


id agent0, reward 000.242
id agent1, reward 000.298
id agent2, reward 000.405
id agent3, reward 000.303
id agent4, reward 000.199
id agent5, reward 000.326
id agent6, reward 000.324
id agent7, reward 000.354
id agent8, reward 000.404
id agent9, reward 000.237
id agent10, reward 000.328
id agent11, reward 000.297
id agent12, reward 000.287
id agent13, reward 000.337
id agent14, reward 000.378
id agent15, reward 000.361
agent0 agent1 agent2 agent4 agent5 agent8 agent14agent15
agent3 agent6 agent7
agent9 agent10agent11agent12agent13
PPO-0409-2 220-iters

Waiting and observing behaviors emerge! But still easy to collide.

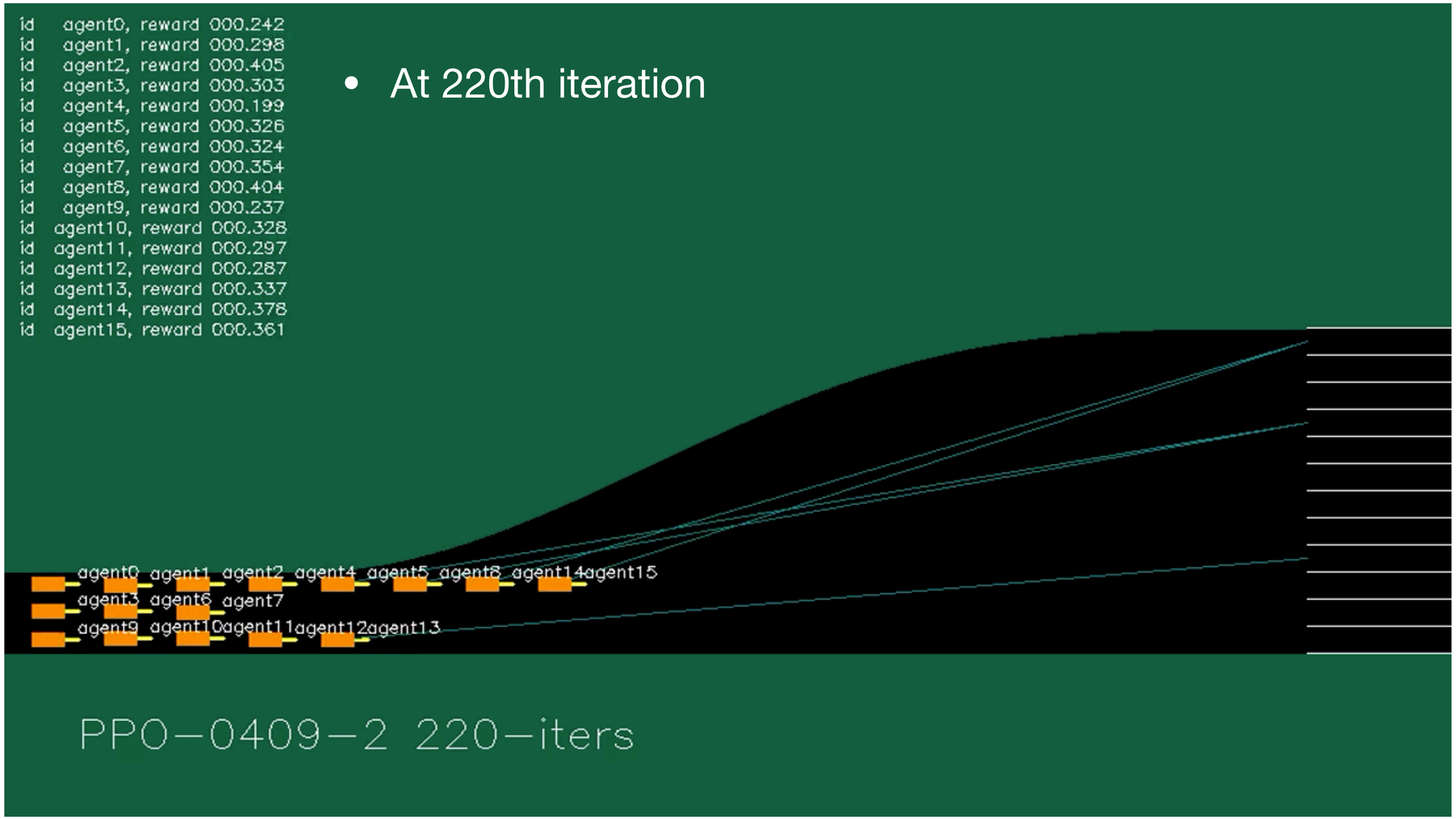
Metric	Value
iteration	220
— performance —	
agent collision	0.029
boundary col. rate	0
gate collision rate	0
pass rate	0.9709
success rate	0.1662
policy loss	-0.00224
value loss	1.47595
episode reward	1643.995
single reward	102.749
episode length	193.753
— configuration —	
num workers	48
env per worker	16
cpu per worker	1
sample batch size	100
train batch size	80000

Emergence of “intelligent behaviors”


id agent0, reward 000.242
id agent1, reward 000.298
id agent2, reward 000.405
id agent3, reward 000.303
id agent4, reward 000.199
id agent5, reward 000.326
id agent6, reward 000.324
id agent7, reward 000.354
id agent8, reward 000.404
id agent9, reward 000.237
id agent10, reward 000.328
id agent11, reward 000.297
id agent12, reward 000.287
id agent13, reward 000.337
id agent14, reward 000.378
id agent15, reward 000.361
agent0 agent1 agent2 agent4 agent5 agent8 agent14agent15
agent3 agent6 agent7
agent9 agent10agent11agent12agent13
PPO-0409-2 220-iters

Emergence of “intelligent behaviors”

Waiting and observing behaviors emerge! But still easy to collide.



Metric	Value
iteration	220
— performance —	
agent collision	0.029
boundary col. rate	0
gate collision rate	0
pass rate	0.9709
success rate	0.1662
policy loss	-0.00224
value loss	1.47595
episode reward	1643.995
single reward	102.749
episode length	193.753
— configuration —	
num workers	48
env per worker	16
cpu per worker	1
sample batch size	100
train batch size	80000

Discussion on computing efficiency

Metric \ Exp. Name	PPO-0409-1	debug	debug-41600	debug-85000
iteration	135	135	276	132
agent collision	0.03947	0.03796	0.05523	0.0121
pass rate	0.96037	0.96189	0.9447	0.98796
success rate	0.15490	0.1588	0.1645	0.16492
episode reward	1629.7544	1630.712	1514.805	1653.808
num workers	48	52	52	2
env per worker	16	8	8	16
cpu per worker	1	1	1	26
sample batch size	100	200	100	200
train batch size	80000	85000	41600	85000
grad time (ms)	933797.591	925613.702	453098.768	955638.176
load time (ms)	359.761	340.345	161.508	364.096
sample time (ms)	49862.033	59504.045	17024.831	64123.812
update time (ms)	26.568	20.213	26.619	10.313
iteration time	974.257	997.7915	476.2374	1028.4034
total time (iter avg)	129382.6396 (958)	133714.459 (990)	133056.2950 (492)	132472.352 (1003)

*Note: sample batch size means the batch size sampled from **one environment,** no matter how many agents in it.*