Joint RL meeting

Gridworld implementation of Olivia's task (bis)

Andrea Pierré

February 27th, 2023

Brown University

Outline

1. Implementation
2. Results & experiments
3. Summary

Outline

Composite state space

- Allocentric setting:

location	cue
{0,...,24}	North light
	South light
	Odor A
	Odor B

- Egocentric setting:

location	head direction [°]	cue
{0,...,24}	0	North light
	90	South light
	180	Odor A
	270	Odor B

Composite state space

- Allocentric setting:

location	cue
{0,...,24}	North light
	South light
	Odor A
	Odor B

- Egocentric setting:

location	head direction [°]	cue
{0,...,24}	0	North light
	90	South light
	180	Odor A
	270	Odor B

Flattened state space – allocentric setting

Pre odor - North light

0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
20	21	22	23	24

Pre odor - South light

25	26	27	28	29
30	31	32	33	34
35	36	37	38	39
40	41	42	43	44
45	46	47	48	49

Post odor - Odor A

50	51	52	53	54
55	56	57	58	59
60	61	62	63	64
65	66	67	68	69
70	71	72	73	74

Post odor - Odor B

75	76	77	78	79
80	81	82	83	84
85	86	87	88	89
90	91	92	93	94
95	96	97	98	99

Flattened state space – egocentric setting

Pre odor - North light - 0°

0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
20	21	22	23	24

Pre odor - North light - 90°

25	26	27	28	29
30	31	32	33	34
35	36	37	38	39
40	41	42	43	44
45	46	47	48	49

Pre odor - North light - 180°

50	51	52	53	54
55	56	57	58	59
60	61	62	63	64
65	66	67	68	69
70	71	72	73	74

Pre odor - North light - 270°

75	76	77	78	79
80	81	82	83	84
85	86	87	88	89
90	91	92	93	94
95	96	97	98	99

Pre odor - South light - 0°

100	101	102	103	104
105	106	107	108	109
110	111	112	113	114
115	116	117	118	119
120	121	122	123	124

Pre odor - South light - 90°

125	126	127	128	129
130	131	132	133	134
135	136	137	138	139
140	141	142	143	144
145	146	147	148	149

Pre odor - South light - 180°

150	151	152	153	154
155	156	157	158	159
160	161	162	163	164
165	166	167	168	169
170	171	172	173	174

Pre odor - South light - 270°

175	176	177	178	179
180	181	182	183	184
185	186	187	188	189
190	191	192	193	194
195	196	197	198	199

Post odor - Odor A - 0°

200	201	202	203	204
205	206	207	208	209
210	211	212	213	214
215	216	217	218	219
220	221	222	223	224

Post odor - Odor A - 90°

225	226	227	228	229
230	231	232	233	234
235	236	237	238	239
240	241	242	243	244
245	246	247	248	249

Post odor - Odor A - 180°

250	251	252	253	254
255	256	257	258	259
260	261	262	263	264
265	266	267	268	269
270	271	272	273	274

Post odor - Odor A - 270°

275	276	277	278	279
280	281	282	283	284
285	286	287	288	289
290	291	292	293	294
295	296	297	298	299

Post odor - Odor B - 0°

300	301	302	303	304
305	306	307	308	309
310	311	312	313	314
315	316	317	318	319
320	321	322	323	324

Post odor - Odor B - 90°

325	326	327	328	329
330	331	332	333	334
335	336	337	338	339
340	341	342	343	344
345	346	347	348	349

Post odor - Odor B - 180°

350	351	352	353	354
355	356	357	358	359
360	361	362	363	364
365	366	367	368	369
370	371	372	373	374

Post odor - Odor B - 270°

375	376	377	378	379
380	381	382	383	384
385	386	387	388	389
390	391	392	393	394
395	396	397	398	399

States & actions translation

- Wrapper environment to translate the human readable environment (composite states) into a suitable environment for the Q-learning algorithm (flat states)

```
state = {"location": 13, "cue": LightCues.South}
env.convert_composite_to_flat_state(state)
# => 38
```

```
state = 63
env.convert_flat_state_to_composite(state)
# => {"location": 13, "cue": <OdorID.A: 1>}
```

- Machine & human friendly actions

```
action = 0
Actions(action).name
# => "UP"
```

States & actions translation

- Wrapper environment to translate the human readable environment (composite states) into a suitable environment for the Q-learning algorithm (flat states)

```
state = {"location": 13, "cue": LightCues.South}
env.convert_composite_to_flat_state(state)
# => 38
```

```
state = 63
env.convert_flat_state_to_composite(state)
# => {"location": 13, "cue": <OdorID.A: 1>}
```

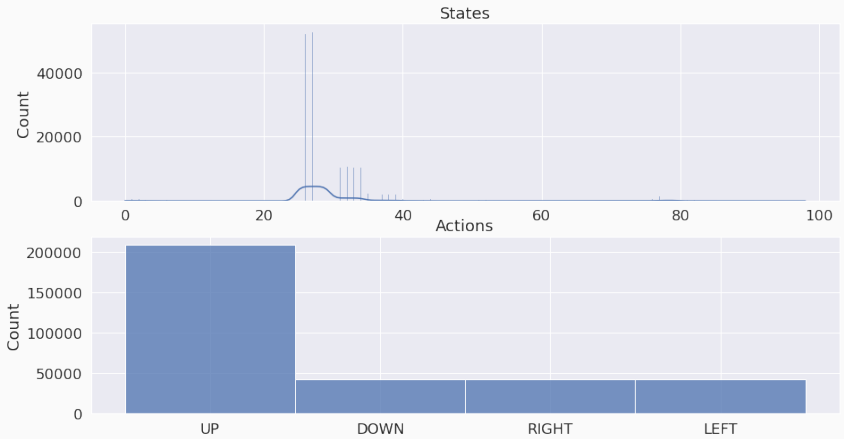
- Machine & human friendly actions

```
action = 0
Actions(action).name
# => "UP"
```

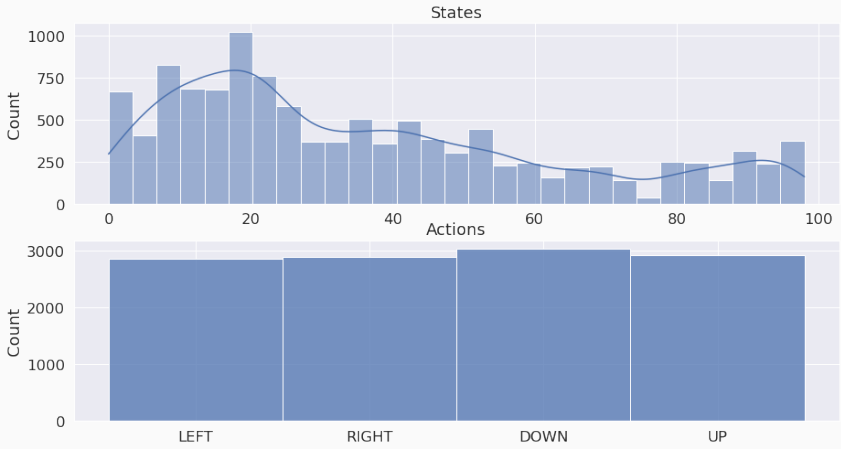
Algorithm troubleshooting

Subtle bug using ϵ-greedy when Q-values are identical:

Vanilla ϵ-greedy

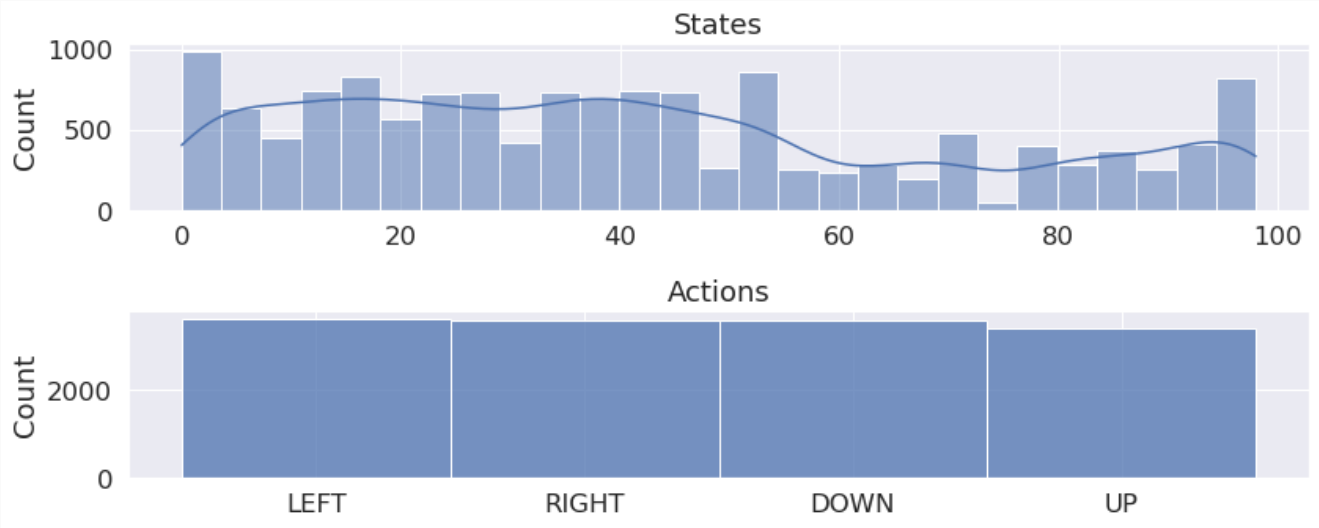


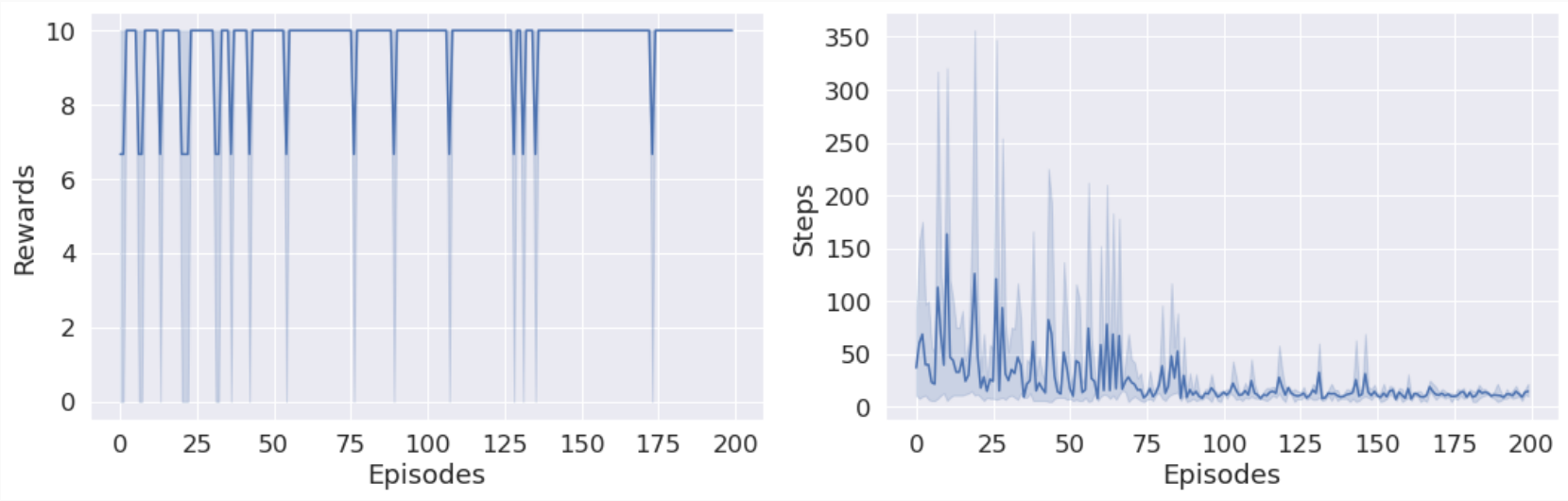
Randomly choosing between actions with the same Q-values



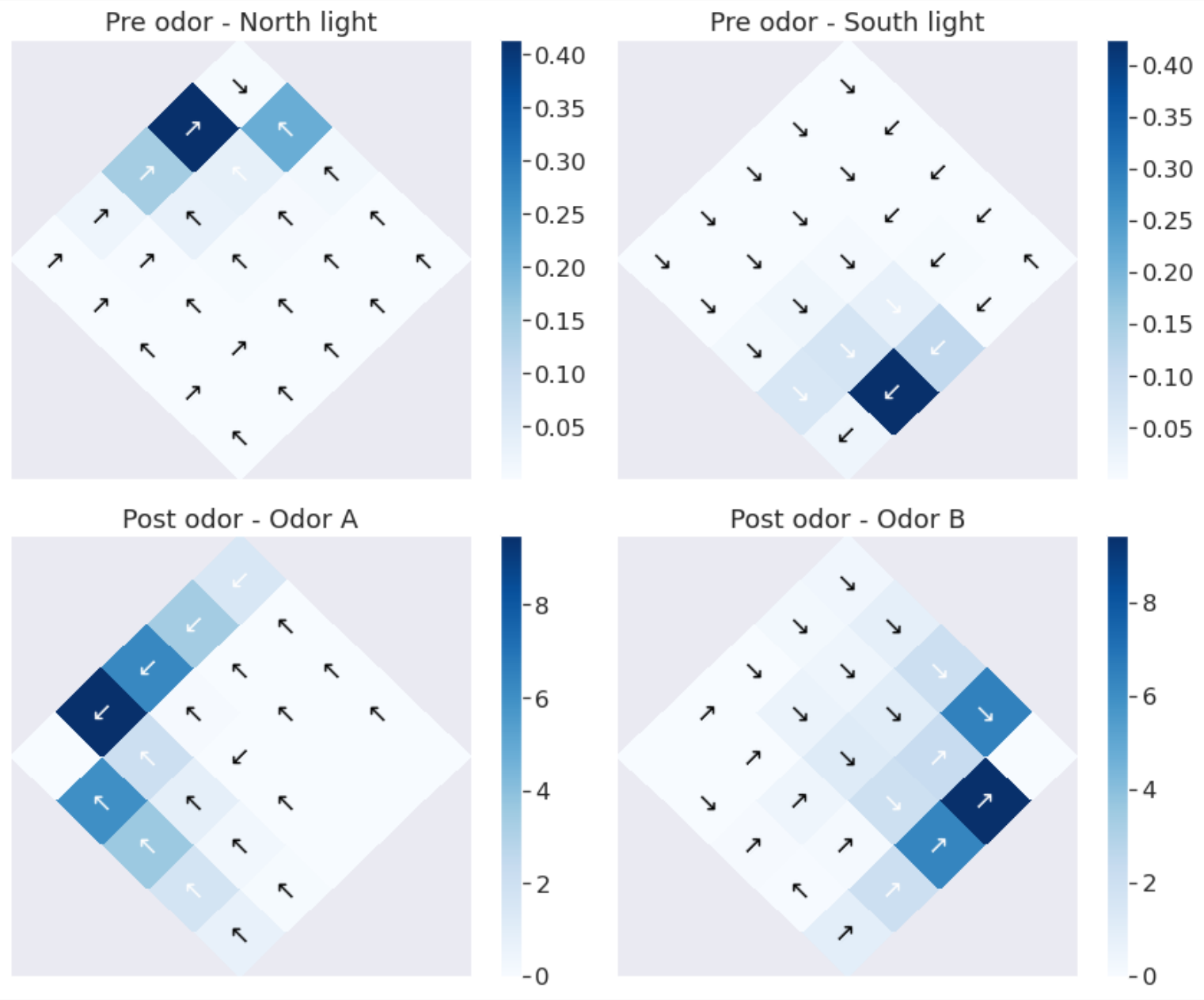
Outline

Standard Q-learning – allocentric setting

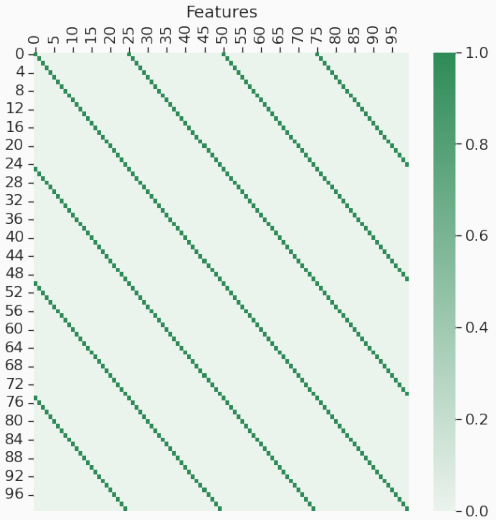


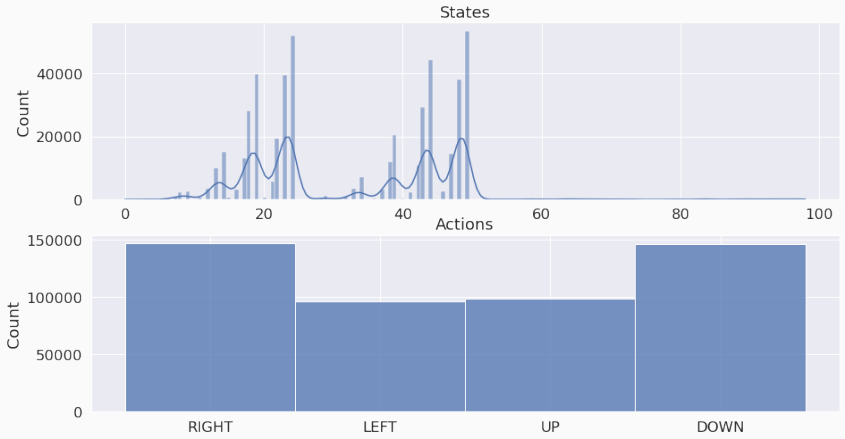


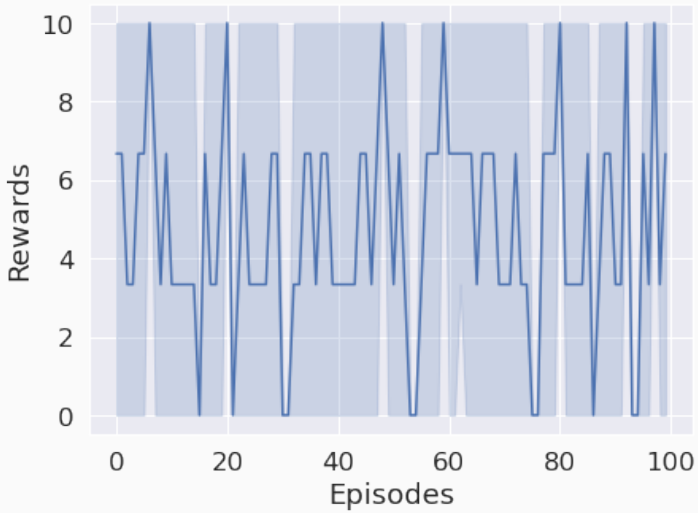
Standard Q-learning – allocentric setting

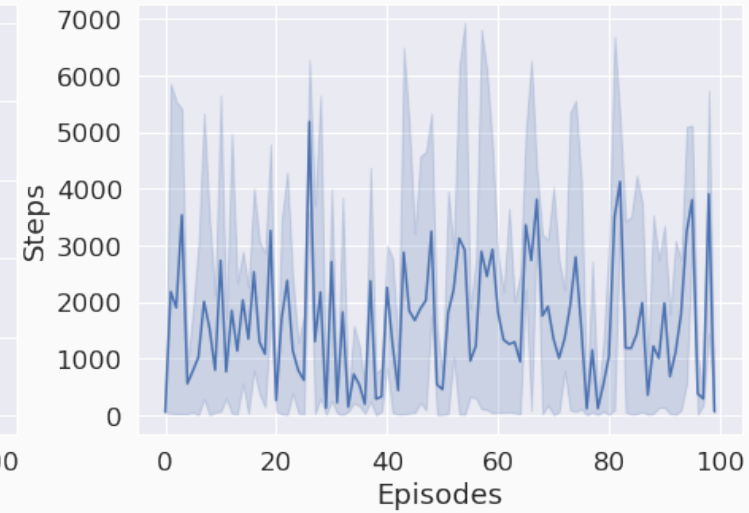


Q-learning with function approximation – allocentric setting – without joint representation

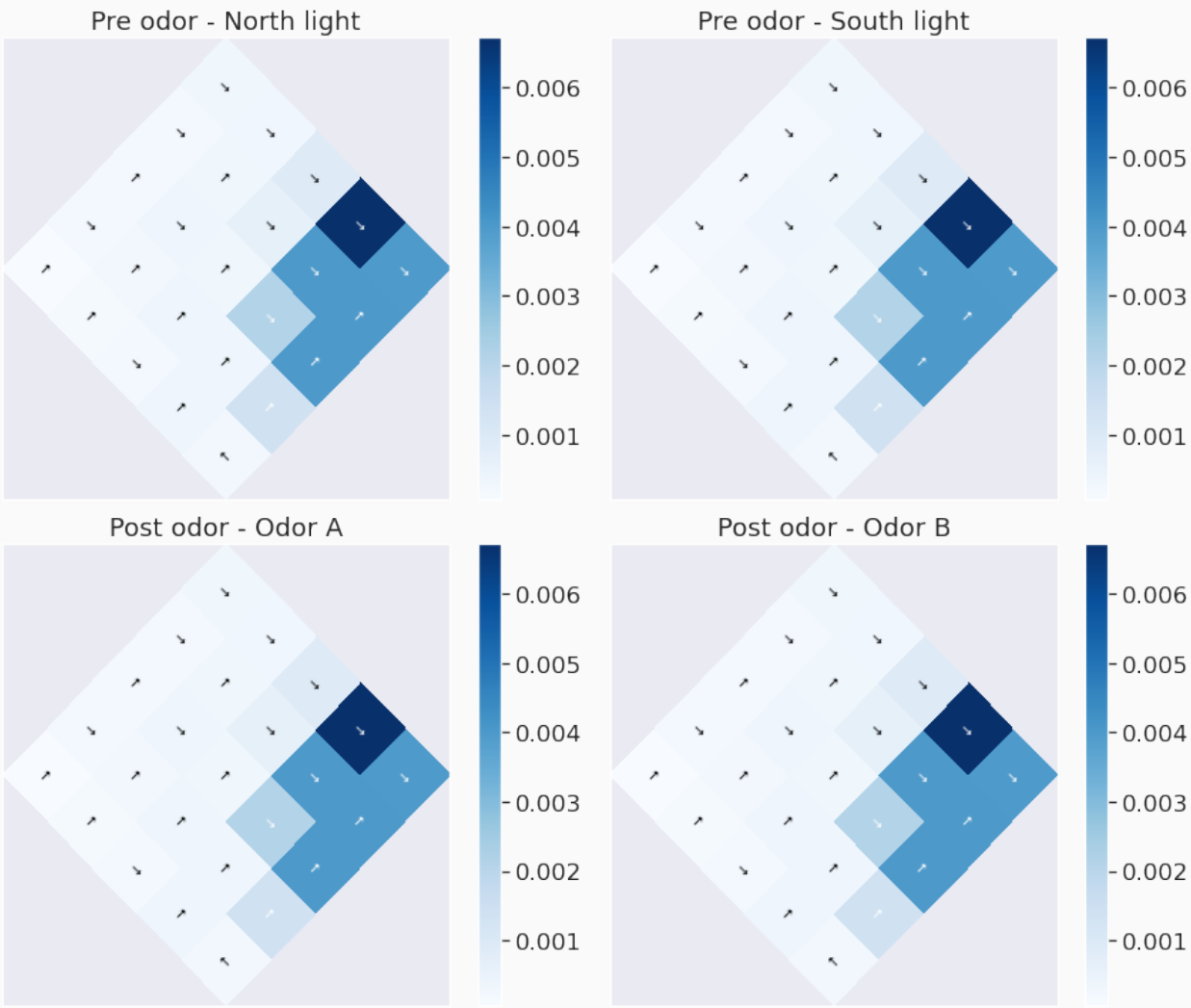




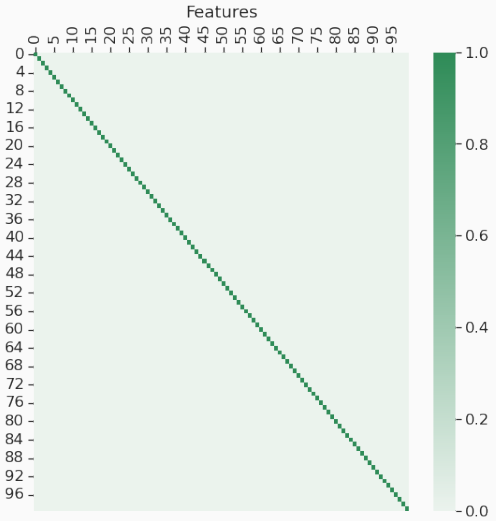


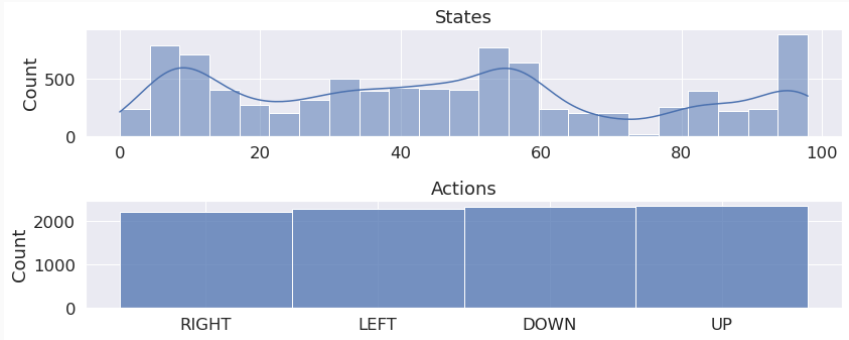


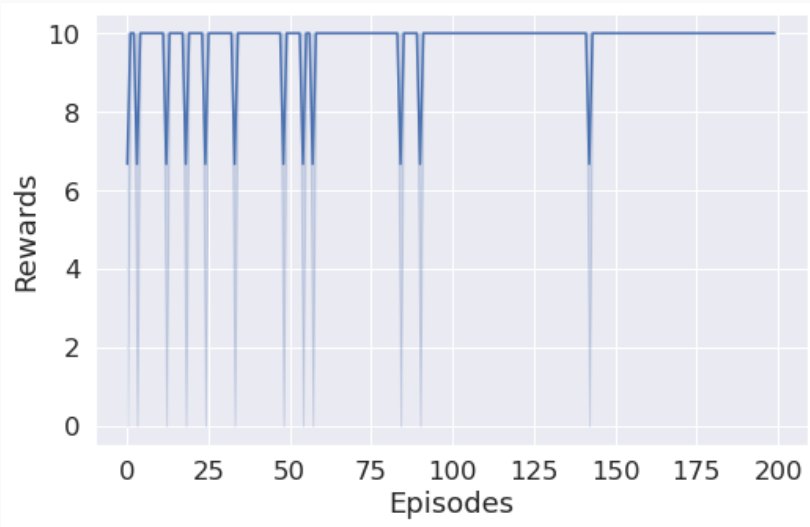
Q-learning with function approximation – allocentric setting – without joint representation

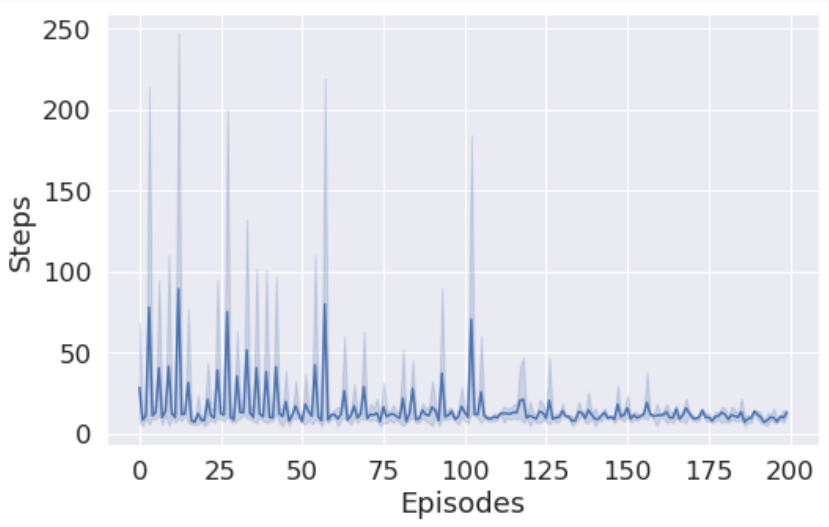


Q-learning with function approximation – allocentric setting – with joint representation

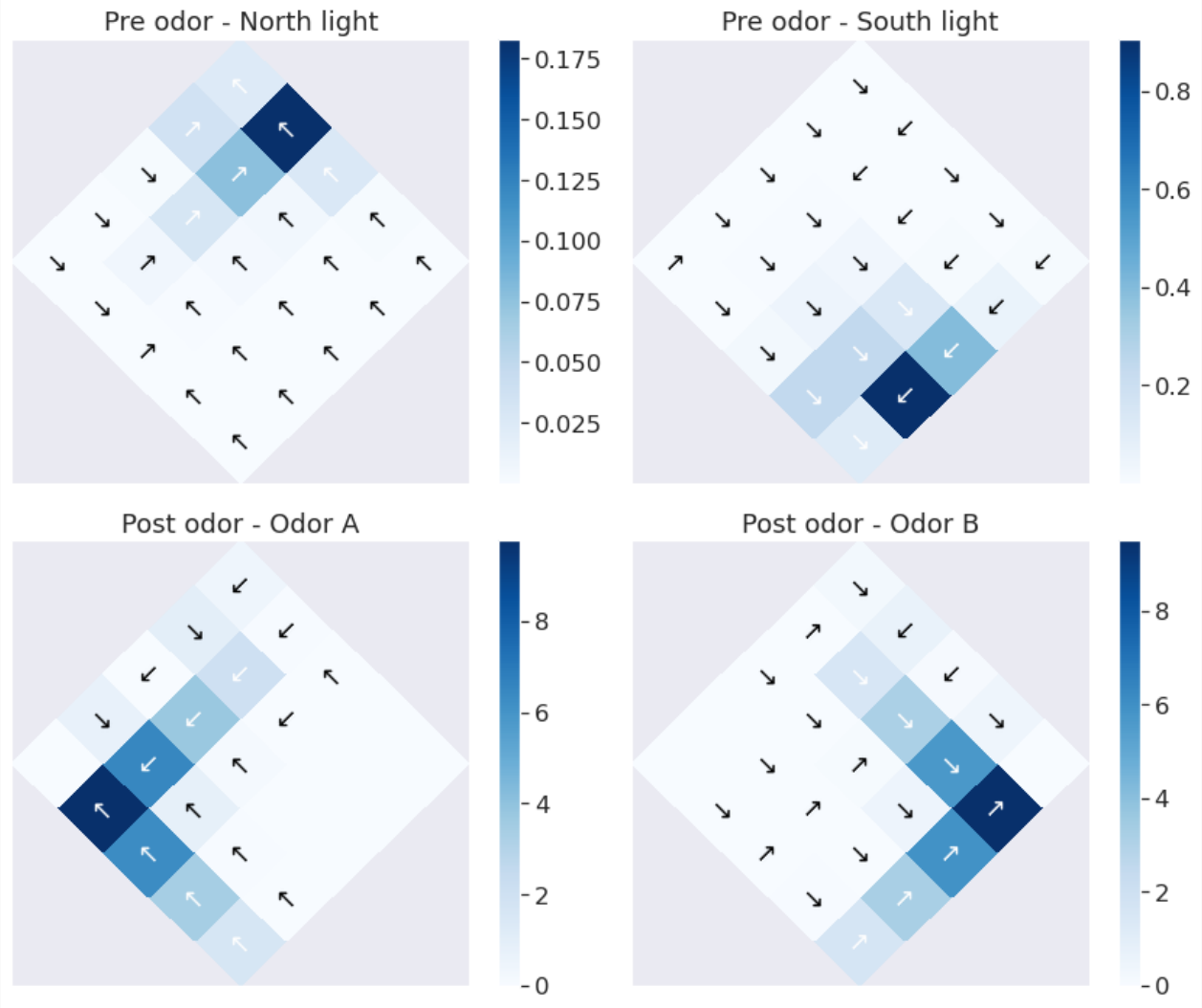




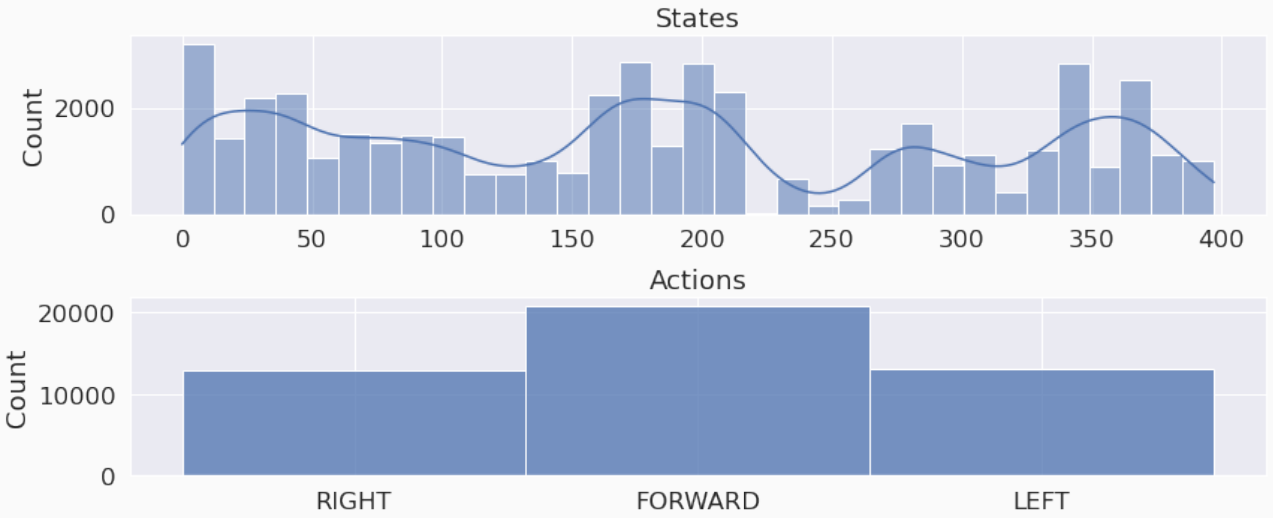


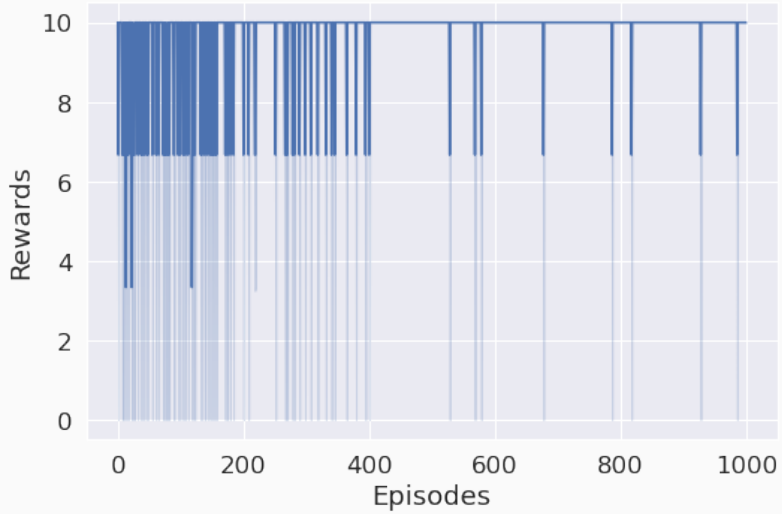


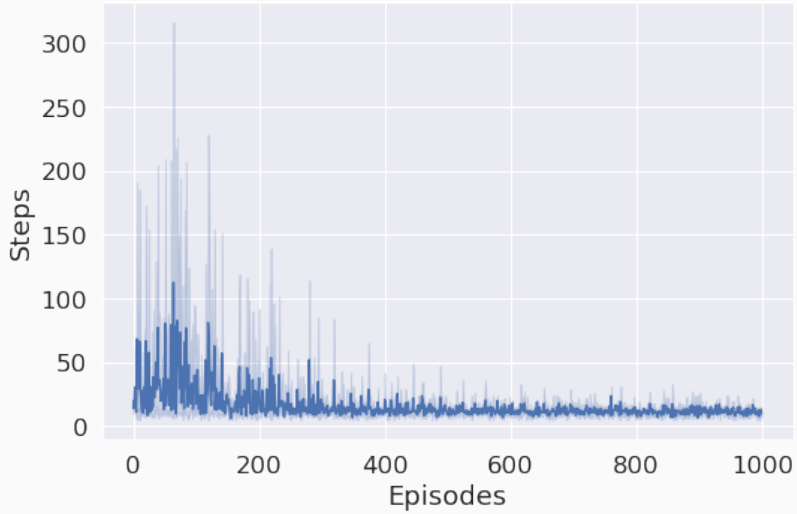
Q-learning with function approximation – allocentric setting – with joint representation



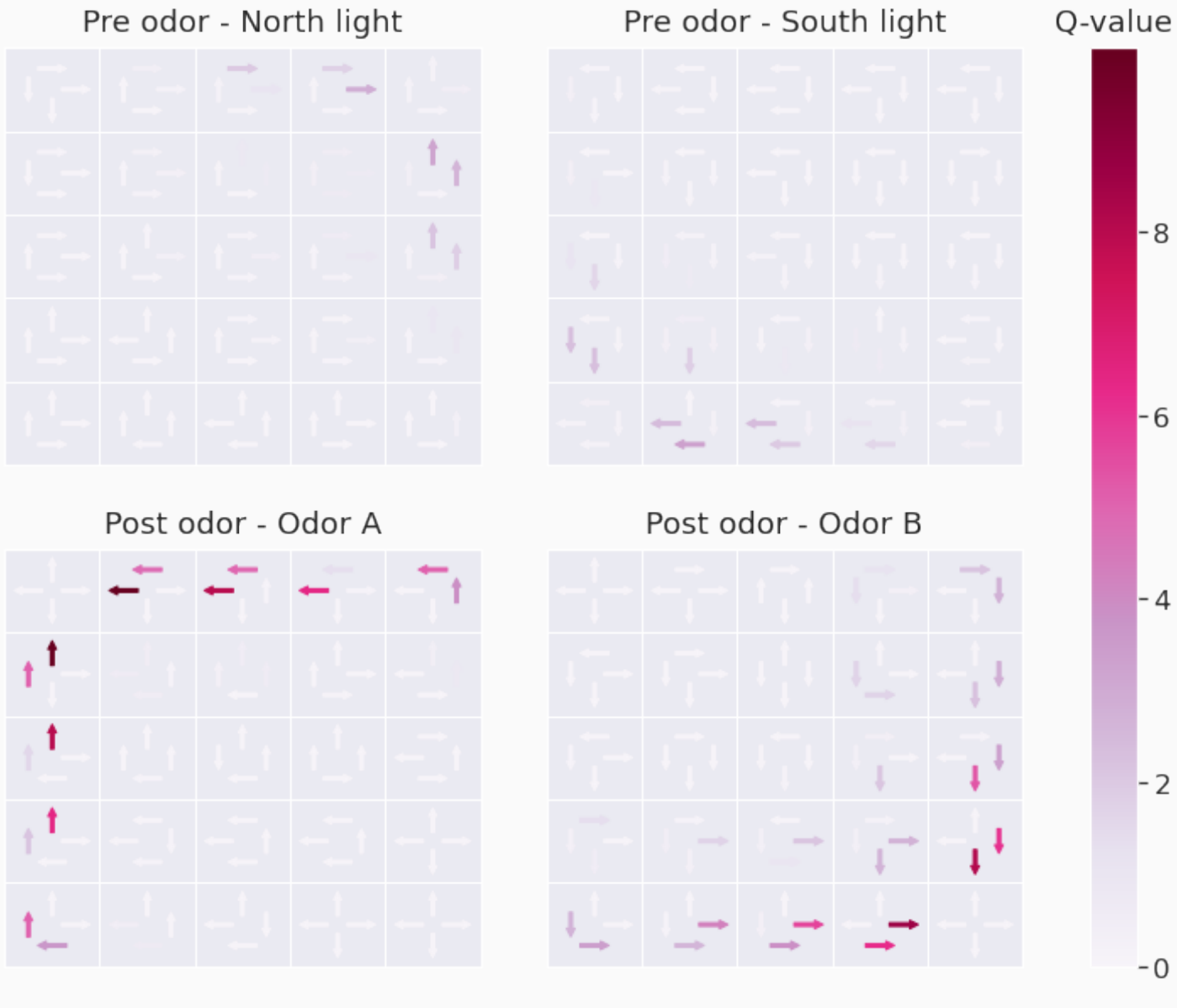
Standard Q-learning – egocentric setting



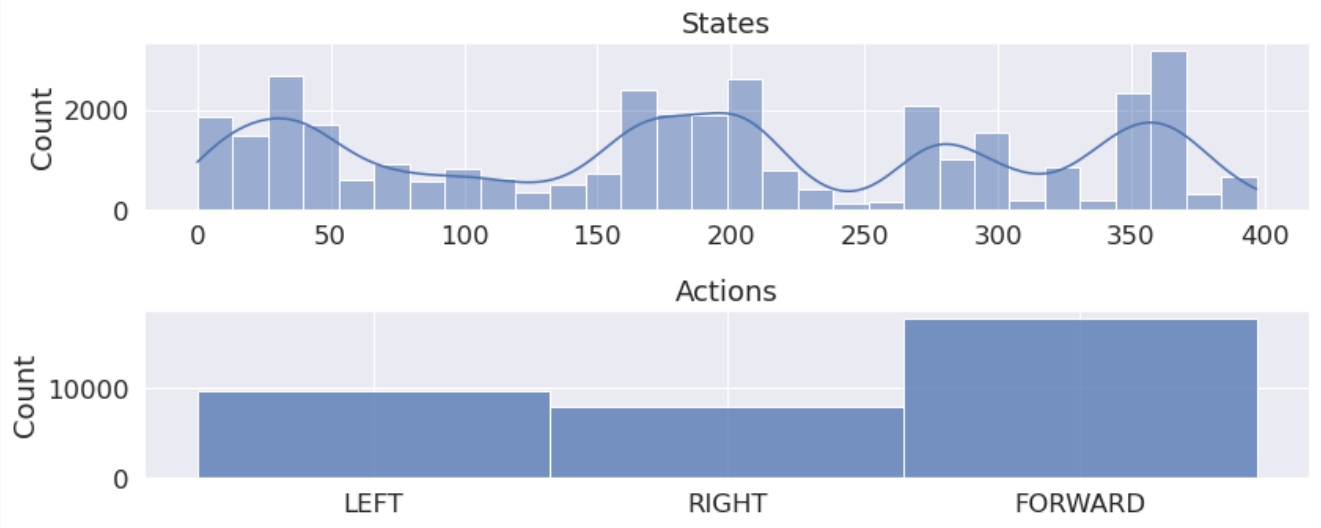


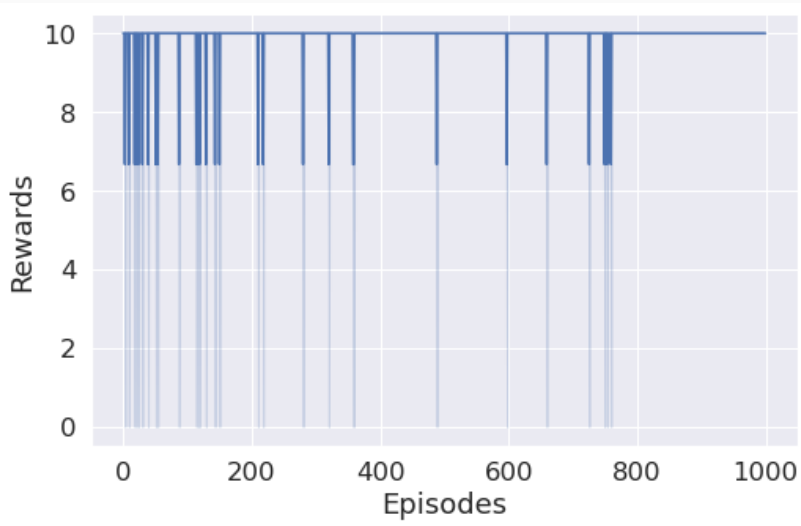


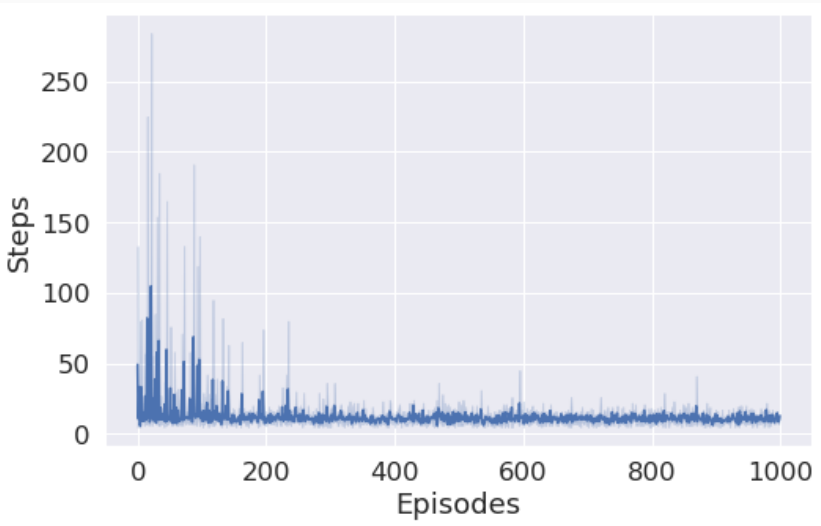
Standard Q-learning – egocentric setting



Q-learning with function approximation – egocentric setting

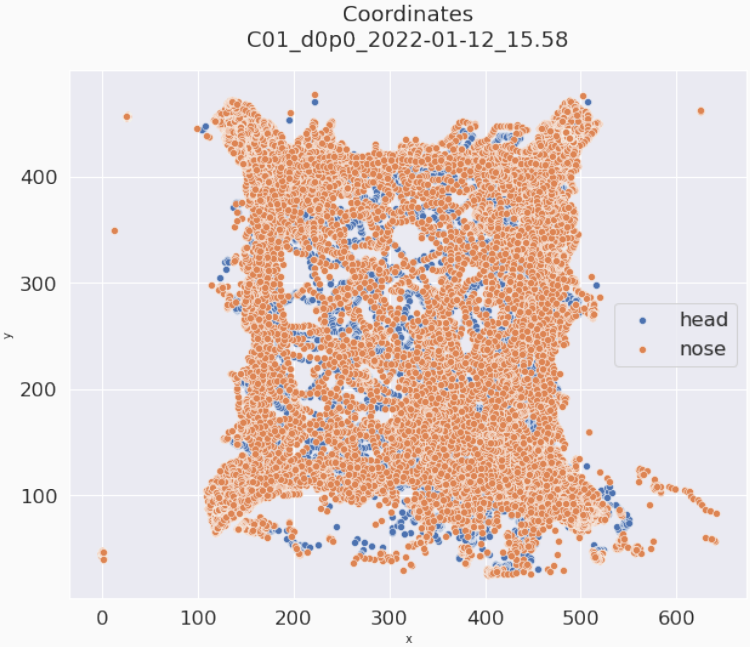


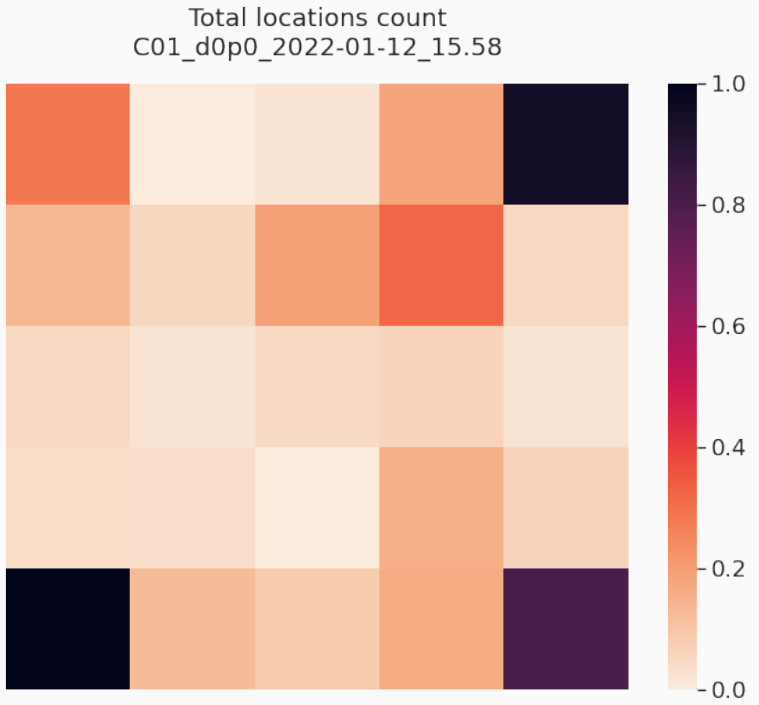




Q-learning with function approximation – egocentric setting

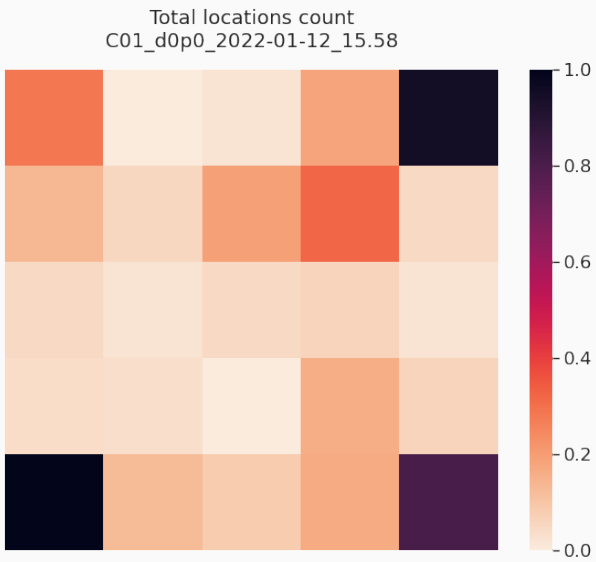
Location occupancy – naive animal

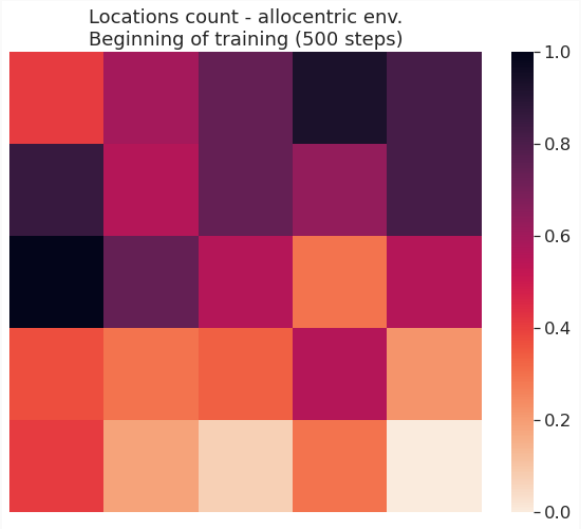


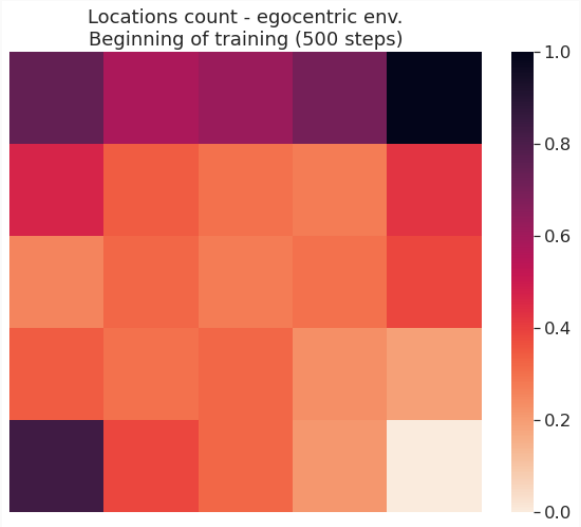


→ The locations around the ports are the most visited zones in the arena

Location occupancy – animal vs. agent



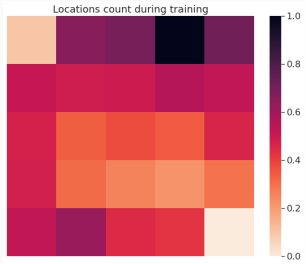




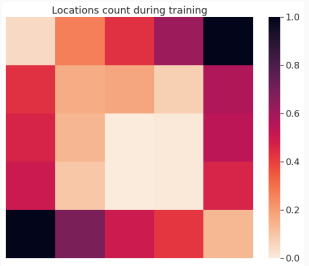
→ The naive agent explores the space more uniformly than a real animal

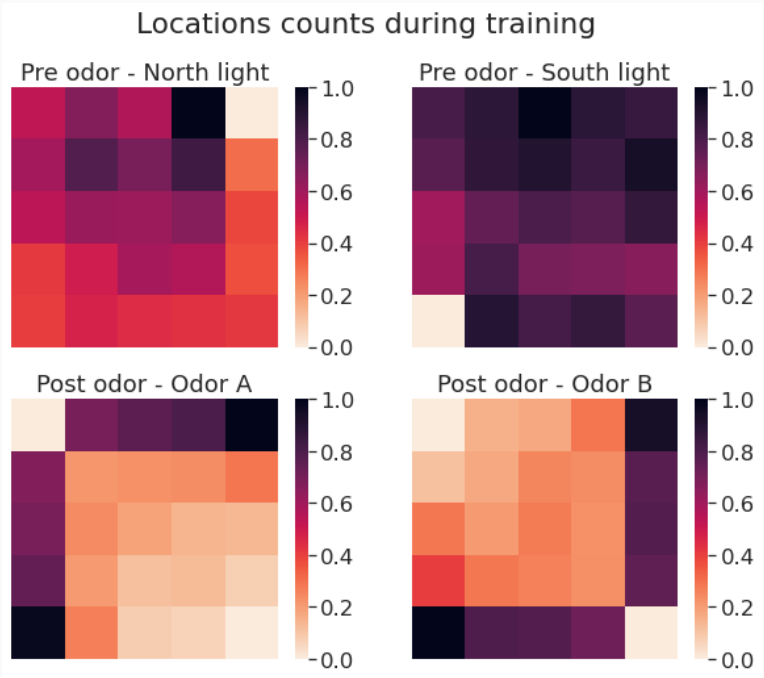
Location occupancy – allocentric vs. egocentric

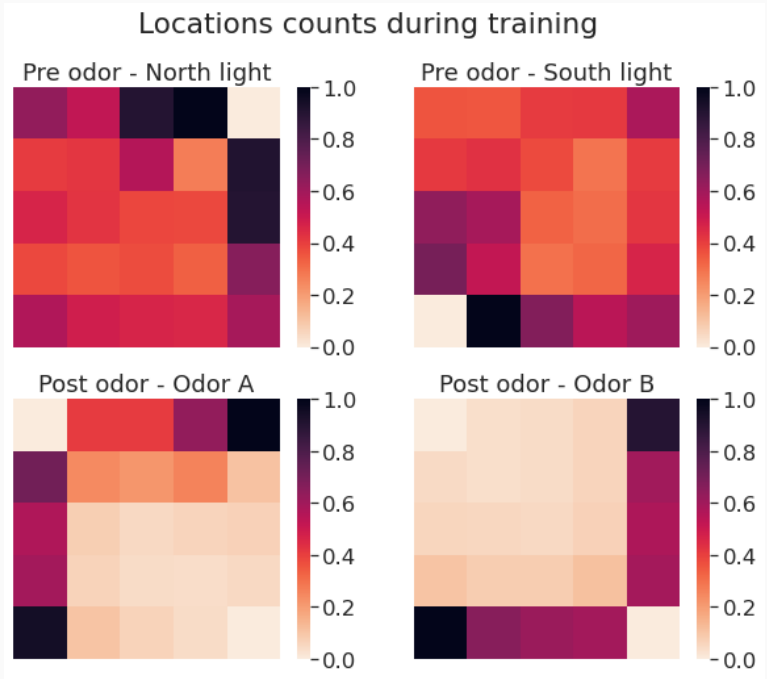
Allocentric



Egocentric







$\rightarrow$ The egocentric agent spend more time along the walls, whereas the allocentric agent has a more homogeneous exploration of the space

Outline

Summary

- Standard Q-learning can learn the task in ~90 episodes in the allocentric setting, and in ~400 episodes in the egocentric setting
- Niloufar's results with function approximation in both allocentric/egocentric settings are reproducible:
  - The agent is not able to learn the task without having a place-odor joint representation
  - With a place-odor joint representation, the agent is able to learn the task in ~60 episodes in the allocentric setting, and in ~300 episodes in the egocentric setting

Summary

- Standard Q-learning can learn the task in ~90 episodes in the allocentric setting, and in ~400 episodes in the egocentric setting
- Niloufar's results with function approximation in both allocentric/egocentric settings are reproducible:
  - The agent is not able to learn the task without having a place-odor joint representation
  - With a place-odor joint representation, the agent is able to learn the task in ~60 episodes in the allocentric setting, and in ~300 episodes in the egocentric setting

Summary

- Standard Q-learning can learn the task in ~90 episodes in the allocentric setting, and in ~400 episodes in the egocentric setting
- Niloufar's results with function approximation in both allocentric/egocentric settings are reproducible:
  - The agent is not able to learn the task without having a place-odor joint representation
  - With a place-odor joint representation, the agent is able to learn the task in ~60 episodes in the allocentric setting, and in ~300 episodes in the egocentric setting

Summary

- Standard Q-learning can learn the task in ~90 episodes in the allocentric setting, and in ~400 episodes in the egocentric setting
- Niloufar's results with function approximation in both allocentric/egocentric settings are reproducible:
  - The agent is not able to learn the task without having a place-odor joint representation
  - With a place-odor joint representation, the agent is able to learn the task in ~60 episodes in the allocentric setting, and in ~300 episodes in the egocentric setting

Summary

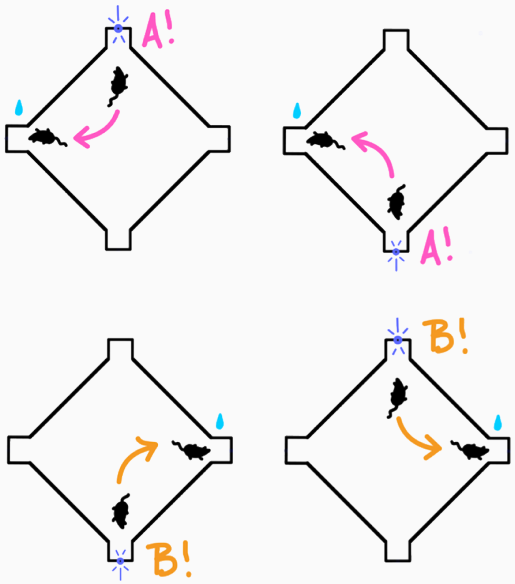
- A naive animal spends most of its time at the ports, whereas a naive agent has a more uniform exploration
- The egocentric agent spend more time along the walls, whereas the allocentric agent has a more homogeneous exploration

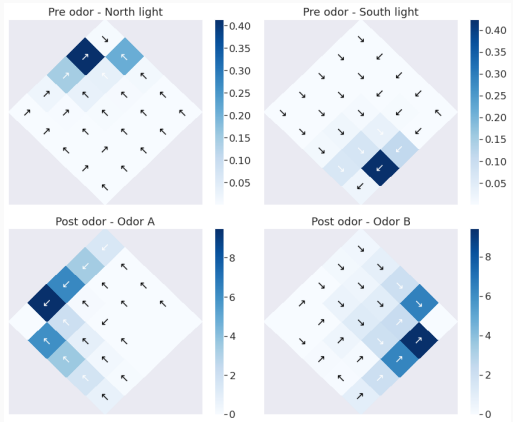
Summary

- A naive animal spends most of its time at the ports, whereas a naive agent has a more uniform exploration
- The egocentric agent spend more time along the walls, whereas the allocentric agent has a more homogeneous exploration

Main differences with Niloufar's model

- The environment is geometrically closer to the real experiment
  $\rightarrow$ ports are in the corners of the arena, not in the middle of the walls
- Code is clean, readable, and abstracted in high level functions/concepts





Next steps

- Implement Olivia's new version of the task ?
- Try to reduce the feature space (Jason's suggestion)
  $\rightarrow$ need to fix function approximation algorithm ?
- Replace the manually crafted features matrix by an artificial neural network, which should learn the necessary representations to solve the task from scratch
- NSGP seminar in ~1 month

Next steps

- Implement Olivia's new version of the task ?
- Try to reduce the feature space (Jason's suggestion)
  $\rightarrow$ need to fix function approximation algorithm ?
- Replace the manually crafted features matrix by an artificial neural network, which should learn the necessary representations to solve the task from scratch
- NSGP seminar in ~1 month

Next steps

- Implement Olivia's new version of the task ?
- Try to reduce the feature space (Jason's suggestion) $\rightarrow$ need to fix function approximation algorithm ?
- Replace the manually crafted features matrix by an artificial neural network, which should learn the necessary representations to solve the task from scratch
- NSGP seminar in ~1 month

Next steps

- Implement Olivia's new version of the task ?
- Try to reduce the feature space (Jason's suggestion) $\rightarrow$ need to fix function approximation algorithm ?
- Replace the manually crafted features matrix by an artificial neural network, which should learn the necessary representations to solve the task from scratch
- NSGP seminar in ~1 month

Questions ?