# "SmartCab"

# P4: report

Prepared by Andrejs Zujevs

## Implement a Basic Driving Agent

***Observe what you see with the agent's behavior as it takes random actions. Does the*** *smartcab* ***eventually make it to the destination? Are there any other interesting observations to note?***

Sometimes the agent was obtained a destination point, but this happened during a long run time, also the agent behavior randomized.

The agent` world is cyclic and the agent living in multi-agent environment which also will described as:

- discrete – states of environment are determined (agent locations in the grid space which is cyclic, traffic lights with green and red color in NS and EW direction). In the world acting more than one agent with their own location and head direction. All states are discrete or not continuous;
- partially observable – the agent can get information of their current location traffic light state and if present other agent's the next action, get deadline and destination point and also can get current location point. The agent can`t get information about traffic lights and agents in other locations of the world;
- static – the environment doesn`t change while the agent is acting;
- deterministic - the next environment state completely determined by current state and the agent action, no probability of the agent's actions;
- episodic – the agent location completely depends on the agent previous actions and each episode's reward depends on the agent`s action.

The active agent marked with red color and destination point also with red.

## Inform the Driving Agent

***What states have you identified that are appropriate for modeling the*** *smartcab* ***and environment? Why do you believe each of these states to be appropriate for this problem?***

An agent percept from sensors: the oncoming agent, agent from the left and from the right with forward, left or right action, traffic light state (red or green) in the NS or EW direction.

There were identified such appropriate agent's states:

LIGHT_COLOR[red, green] + NEXT_WAYPOINT[left, right, forward] +
OTHER_AGENTS_WAYPOINTS[oncoming->direction, left->direction, right->direction]

Although the agent informed about deadline, this not useful and shouldn't be represented as state because the route planner comes up with an optimal route and the agent just only will respect traffic rules and follow planer's route. Since the agent has only an egocentric view, then deadline cannot help him to reach a destination point and the agent don't know about the next waypoint(-s).

***How many states in total exist for the*** *smartcab* ***in this environment? Does this number seem reasonable given that the goal of Q-Learning is to learn and make informed decisions about each state? Why or why not?***

The total number of states is: 2 x 3 x 64 = 384

- 2x – traffic lights (red and green) or traffic directions (SN, EW);
- 3x – agent' next waypoint (forward, right, left);
- 64x ($4^3$) – other agents' waypoints (codded as string with mask ([oncoming.left.right]):
    - oncoming – left, right, toward or none;
    - left – left, right, toward or none;
    - right – left, right, toward or none.

The agent actions are: none, right, left, forward.

Q-Learning algorithm can find optimal policy for any finite Markov Decision Process and as our states and actions are finite, then Q-Learning algorithm is reasonable for the smartcab problem. If the states and actions number would be infinite, then Q-Learning algorithm wouldn`t be used or Q-Learning should be extended and constrained in time, interpolating states map's entries.

Taking into account the total states number and agent's actions number, Q matrix should be with 384 rows and 4 columns.

As Q-Learning algorithm need many iterations to learn, suggested state space size is relatively small and appropriate for this problem.

## Implement a Q-Learning Driving Agent

***What changes do you notice in the agent's behavior when compared to the basic driving agent when random actions were always taken? Why is this behavior occurring?***

After Q-Learning algorithm was implemented, the agent behavior become more effective and the agent will obtain a destination point before deadline occurs, following route and respecting traffic rules. The agent chooses the next action by using e-greedy algorithm with epsilon value = 1/trial_num.

At the beginning of learning the agent gets more negative rewards and sometimes moves in circles but as more trials were finished as more accurate become the agent, respecting traffic rules and route waypoints.

## Improve the Q-Learning Driving Agent

***Report the different values for the parameters tuned in your basic implementation of Q-Learning. For which set of parameters does the agent perform best? How well does the final driving agent perform?***

**Parameters set 1**: alpha = 0.15, gamma = 0.15

```
91 trial: penalties -0.500, deadline 32, moves: 19
   deadline=48, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
92 trial: penalties 0.000, deadline 10, moves: 21
93 trial: penalties -0.500, deadline 36, moves: 20
   deadline=39, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-0.5
94 trial: penalties 0.000, deadline 16, moves: 5
95 trial: penalties 0.000, deadline 16, moves: 5
96 trial: penalties 0.000, deadline 13, moves: 13
97 trial: penalties 0.000, deadline 16, moves: 10
98 trial: penalties 0.000, deadline 15, moves: 16
99 trial: penalties -1.000, deadline 24, moves: 22
   deadline=41, sense={'light': 'red', 'oncoming': None, 'right': 'forward', 'left': None}, act=left, next_wp=forward, reward=-1.0
100 trial: penalties -1.000, deadline 15, moves: 16
   deadline=28, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'left'}, act=forward, next_wp=forward, reward=-1.0
```

**Parameters set 2**: alpha = 0.30, gamma = 0.15

```
91 trial: penalties -1.000, deadline 26, moves: 10
   deadline=35, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'forward'}, act=forward, next_wp=right, reward=-1.0
92 trial: penalties 0.000, deadline 10, moves: 16
93 trial: penalties -0.500, deadline 20, moves: 21
   deadline=24, sense={'light': 'red', 'oncoming': None, 'right': 'right', 'left': None}, act=right, next_wp=forward, reward=-0.5
94 trial: penalties -0.500, deadline 19, moves: 22
   deadline=30, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-0.5
95 trial: penalties 0.000, deadline 6, moves: 15
96 trial: penalties 0.000, deadline 14, moves: 7
97 trial: penalties 0.000, deadline 20, moves: 16
98 trial: penalties 0.000, deadline 17, moves: 9
99 trial: penalties -1.000, deadline 17, moves: 9
   deadline=18, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-1.0
100 trial: penalties -1.000, deadline 25, moves: 21
   deadline=32, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'forward'}, act=forward, next_wp=forward, reward=-
1.0
```

## **Parameters set 3**: alpha = 0.60, gamma = 0.15

```
91 trial: penalties 0.000, deadline 13, moves: 13
92 trial: penalties 0.000, deadline 18, moves: 18
93 trial: penalties 0.000, deadline 10, moves: 11
94 trial: penalties 0.000, deadline 17, moves: 14
95 trial: penalties 0.000, deadline 19, moves: 17
96 trial: penalties 0.000, deadline 9, moves: 17
97 trial: penalties 0.000, deadline 9, moves: 17
98 trial: penalties 0.000, deadline 19, moves: 7
99 trial: penalties 0.000, deadline 22, moves: 9
100 trial: penalties 0.000, deadline 4, moves: 17
```

## **Parameters set 4**: alpha = 1.0, gamma = 0.15

```
91 trial: penalties 0.000, deadline 20, moves: 6
92 trial: penalties 0.000, deadline 11, moves: 10
93 trial: penalties 0.000, deadline 17, moves: 4
94 trial: penalties 0.000, deadline 31, moves: 25
95 trial: penalties 0.000, deadline 17, moves: 4
96 trial: penalties 0.000, deadline 24, moves: 17
97 trial: penalties 0.000, deadline 16, moves: 10
98 trial: penalties 0.000, deadline 19, moves: 12
99 trial: penalties -0.500, deadline 4, moves: 17
   deadline=9, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
100 trial: penalties 0.000, deadline 13, moves: 8
```

## **Parameters set 5**: alpha = 1.0, gamma = 0.30

```
91 trial: penalties -1.000, deadline 26, moves: 10
   deadline=30, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'right'}, act=forward, next_wp=forward, reward=-1.0
92 trial: penalties 0.000, deadline 19, moves: 7
93 trial: penalties -1.000, deadline 13, moves: 8
   deadline=15, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'right'}, act=left, next_wp=forward, reward=-1.0
94 trial: penalties -1.000, deadline 17, moves: 14
   deadline=26, sense={'light': 'red', 'oncoming': None, 'right': 'left', 'left': None}, act=left, next_wp=forward, reward=-1.0
95 trial: penalties -0.500, deadline 19, moves: 17
   deadline=26, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
96 trial: penalties 0.000, deadline 15, moves: 6
97 trial: penalties -1.000, deadline 15, moves: 11
   deadline=24, sense={'light': 'red', 'oncoming': 'left', 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
   deadline=23, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
98 trial: penalties 0.000, deadline 18, moves: 13
99 trial: penalties 0.000, deadline 13, moves: 13
100 trial: penalties 0.000, deadline 17, moves: 4
```

## **Parameters set 6**: alpha = 1.0, gamma = 0.60

```
91 trial: penalties 0.000, deadline 10, moves: 11
92 trial: penalties 0.000, deadline 14, moves: 7
93 trial: penalties 0.000, deadline 28, moves: 8
94 trial: penalties 0.000, deadline 15, moves: 16
95 trial: penalties 0.000, deadline 17, moves: 4
96 trial: penalties -0.500, deadline 10, moves: 21
   deadline=29, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-0.5
97 trial: penalties 0.000, deadline 17, moves: 4
98 trial: penalties 0.000, deadline 20, moves: 11
99 trial: penalties 0.000, deadline 32, moves: 9
100 trial: penalties -0.500, deadline 13, moves: 8
```

```
    deadline=18, sense={'light': 'red', 'oncoming': 'forward', 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
Parameters set 7: alpha = 0.95, gamma = 0.95
```

## **Parameters set 8**: alpha = 1.0, gamma = 1.0

```
91 trial: penalties -1.500, deadline 35, moves: 16
    deadline=45, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=42, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=39, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
92 trial: penalties -1.500, deadline 9, moves: 17
    deadline=22, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=17, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=13, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
93 trial: penalties -5.000, deadline 0, moves: 31
    deadline=30, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=27, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=23, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=21, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=13, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'forward'}, act=left, next_wp=forward, reward=-1.0
    deadline=12, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=11, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=8, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=6, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
94 trial: penalties -1.500, deadline 26, moves: 10
    deadline=35, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=33, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=28, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
95 trial: penalties -6.500, deadline 12, moves: 39
    deadline=50, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=41, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=39, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=34, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=32, sense={'light': 'red', 'oncoming': 'forward', 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
    deadline=31, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=30, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=27, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'right'}, act=left, next_wp=left, reward=-1.0
    deadline=24, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=17, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'right'}, act=forward, next_wp=forward, reward=-1.0
    deadline=16, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
96 trial: penalties -1.000, deadline 19, moves: 17
    deadline=33, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=23, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
97 trial: penalties -1.500, deadline 11, moves: 10
    deadline=20, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=18, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=16, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
98 trial: penalties -3.500, deadline 0, moves: 21
    deadline=18, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=14, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=12, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=11, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': 'right'}, act=forward, next_wp=right, reward=-0.5
    deadline=7, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=2, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=1, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
99 trial: penalties -6.500, deadline 0, moves: 51
    deadline=50, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=48, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=37, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=36, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
```

```
   deadline=27, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
   deadline=23, sense={'light': 'red', 'oncoming': None, 'right': 'forward', 'left': None}, act=right, next_wp=forward, reward=-0.5
   deadline=19, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
   deadline=14, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
   deadline=12, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
   deadline=10, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
   deadline=5, sense={'light': 'red', 'oncoming': 'right', 'right': None, 'left': None}, act=left, next_wp=forward, reward=-1.0
   deadline=4, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
100 trial: penalties 0.000, deadline 13, moves: 8
```

## **Conclusion**:

The best parameters set is Parameters set 4. Agent learns more quickly and small gamma parameter' value notice that future rewards are worth less than immediate rewards. Learning parameter alpha equal to 1.0 would make the agent consider only the most recent information and for deterministic environments 1.0 is optimal. Although the parameters set 4 is the best, the agent has also negative reward due to incorrect action as suggested rote planner. This happens due to the agent use e-greedy algorithm for choosing best action and sometimes (with small probability) agent acts in random providing exploratory behavior.

***Does your agent get close to finding an optimal policy, i.e. reach the destination in the minimum possible time, and not incur any penalties? How would you describe an optimal policy for this problem?***

The implemented agent acts in optimal way sometimes with penalties (due e-greedy strategy use random selection of action) and in minimum possible time, respecting deadlines. Optimal policy for the particular problem can be described as:

Turn right: light is green and the next waypoint is right or light is red and agent from left not going forward.

Forward: light is green and the next waypoint is forward.

Turn left: light is green and the next waypoint is left and not oncoming agent with forward or right.

Wait: light is red and next waypoint is not turn right or light is red and next waypoint is right and is oncoming agent from left forwards.

The function ***show_performance_results()*** provided the report about how well agent acted:

```
91 trial: penalties 0.000, deadline 14, moves: 7
92 trial: penalties 0.000, deadline 23, moves: 13
93 trial: penalties 0.000, deadline 32, moves: 9
94 trial: penalties 0.000, deadline 20, moves: 6
95 trial: penalties -0.500, deadline 25, moves: 11
   deadline=28, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
96 trial: penalties -1.000, deadline 20, moves: 16
   deadline=24, sense={'light': 'red', 'oncoming': 'left', 'right': None, 'left': None}, act=forward, next_wp=left, reward=-1.0
97 trial: penalties 0.000, deadline 11, moves: 15
98 trial: penalties -1.000, deadline 11, moves: 10
```

```
    deadline=16, sense={'light': 'red', 'oncoming': None, 'right': 'right', 'left': None}, act=left, next_wp=forward, reward=-1.0
99 trial: penalties 0.000, deadline 23, moves: 13
100 trial: penalties 0.000, deadline 19, moves: 12
```

and for the comparison if agent has only 10 trials:

```
1 trial: penalties -10.000, deadline 0, moves: 21
    deadline=20, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=19, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
    deadline=15, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-0.5
    deadline=14, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-0.5
    deadline=13, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-1.0
    deadline=12, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-1.0
    deadline=11, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-1.0
    deadline=10, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-0.5
    deadline=9, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-0.5
    deadline=7, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-1.0
    deadline=5, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=4, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=left, reward=-1.0
    deadline=2, sense={'light': 'red', 'oncoming': 'right', 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
    deadline=1, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=0, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=left, reward=-0.5
2 trial: penalties -10.500, deadline 3, moves: 28
    deadline=29, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=forward, reward=-1.0
    deadline=28, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-1.0
    deadline=25, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=24, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=left, reward=-1.0
    deadline=23, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=left, reward=-1.0
    deadline=22, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=left, reward=-1.0
    deadline=21, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
    deadline=17, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=15, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=14, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=left, reward=-0.5
    deadline=13, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
    deadline=11, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-1.0
    deadline=10, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-0.5
    deadline=7, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-0.5
    deadline=6, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=right, reward=-0.5
3 trial: penalties -0.500, deadline 19, moves: 17
    deadline=25, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-0.5
4 trial: penalties 0.000, deadline 25, moves: 11
5 trial: penalties -4.000, deadline 6, moves: 25
    deadline=29, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=left, reward=-1.0
    deadline=24, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=forward, reward=-1.0
    deadline=16, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=forward, reward=-1.0
    deadline=9, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-1.0
6 trial: penalties 0.000, deadline 13, moves: 8
7 trial: penalties -0.500, deadline 19, moves: 17
    deadline=33, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
8 trial: penalties -5.500, deadline 0, moves: 26
    deadline=21, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
    deadline=20, sense={'light': 'red', 'oncoming': 'forward', 'right': None, 'left': None}, act=forward, next_wp=left, reward=-1.0
    deadline=19, sense={'light': 'red', 'oncoming': 'forward', 'right': None, 'left': None}, act=left, next_wp=left, reward=-1.0
    deadline=18, sense={'light': 'red', 'oncoming': 'forward', 'right': None, 'left': None}, act=right, next_wp=left, reward=-0.5
    deadline=17, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=right, reward=-0.5
    deadline=14, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=forward, next_wp=left, reward=-1.0
    deadline=3, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-1.0
9 trial: penalties -1.000, deadline 18, moves: 8
    deadline=22, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=left, next_wp=forward, reward=-1.0
10 trial: penalties -3.000, deadline 4, moves: 37
    deadline=35, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'right'}, act=forward, next_wp=forward, reward=-1.0
    deadline=23, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': 'left'}, act=forward, next_wp=forward, reward=-1.0
    deadline=22, sense={'light': 'green', 'oncoming': None, 'right': None, 'left': 'left'}, act=left, next_wp=forward, reward=-0.5
    deadline=20, sense={'light': 'red', 'oncoming': None, 'right': None, 'left': None}, act=right, next_wp=forward, reward=-0.5
```