Supplementary Material for "Enhancing Decision-making Safety in AD Through Online Model Checking"

1 Single-lane experimental results

Table 1 presents the experimental results for the single-lane environment. Due to the simplicity of this environment, the GAs successfully avoided collisions in all test runs, even without the shield. However, when considering a stricter safety requirement—that TTC with any other vehicle must remain above 2 seconds—violations were still observed, albeit rarely. This indicates that while the GAs were able to prevent collisions, their decisions sometimes resulted in unsafe driving situations according to the predefined safety constraints.

With the safety shield enabled, the ego vehicle consistently maintained a TTC above 2 seconds, demonstrating the shield's ability to enforce safety requirements. However, the average travel distance and reward received were slightly reduced. These reductions reflect the trade-off between safety and efficiency.

The experiments with the AAs confirmed the effectiveness of the shield in preventing collisions despite the agents actively attempting to cause them. While Table 1 does not include results for the unshielded AAs, in all 2000 tests, the unshielded AAs caused collisions in every trial.

2 Analysis of collisions in Adversary Agent runs

11 collisions were observed in shielded AA runs. Analyzing the vehicle trajectories from these 11 collision tests, we found the following patterns:

Table 1. Single-lane experiment results.

		DQN model	**PPO model**
Unshielded GA	Collisions	0/1000	0/1000
	Distance (m)	413.09 ± 7.33	417.13 ± 7.56
	Speed (m/s)	13.77 ± 0.24	13.90 ± 0.25
	Reward	20.10 ± 0.09	20.15 ± 0.10
	TTC < 2 (%)	0.03	0.84
Shielded GA	Collisions	0/1000	0/1000
	Distance (m)	405.86 ± 7.87	406.10 ± 7.68
	Speed (m/s)	13.53 ± 0.26	13.54 ± 0.26
	Reward	20.01 ± 0.10	20.02 ± 0.10
	TTC < 2 (%)	0	0
	Faster approval (%)	53.28	66.50
Shielded AA	No. collisions	0/1000	0/1000
	Distance (m)	405.73 ± 8.00	405.90 ± 7.73
	Speed (m/s)	13.52 ± 0.27	13.53 ± 0.26
	TTC < 2 (%)	0	0
	Faster approval (%)	25.33	25.08

(i) Front-end collisions due to high speeds (7 cases): Collisions occurred despite at least three consecutive `SLOWER` actions being applied immediately beforehand. In these tests, the ego vehicle often had a significantly higher speed (up to 40 m/s) compared to those in the GA tests (which rarely exceeded 30 m/s). This speed difference arises because the AAs were trained with double the high-speed reward compared to the GAs (see Section 3). This high-speed characteristic together with the insufficient deceleration resulted in collisions even after three `SLOWER` actions. The shield is not responsible for these collisions. On the contrary, it even helped to reduce their impact since at least three consecutive `SLOWER` actions were applied. Note that the standard deviations of speeds shown in Table 1 were calculated per test, not per observation (the number of observations is 30 times the number of tests).

(ii) Spawn collisions (4 cases): An NPC vehicle was spawned within the ego vehicle's occupied region, resulting in an immediate collision. Vehicle spawning was more frequent in the AA tests due to the agent's aggressive behavior, which often caused it to slow down (and speed up) excessively. When the ego vehicle's speed drops too low, Highway-env spawns new NPC vehicles to maintain traffic density as other NPCs already moved far ahead. Obviously, the shield is not responsible for these collisions.

3 Reward configurations for training agents

The reward configuration for training the GAs in the three-lane highway is as follows:

```
"collision_reward": -1,
"right_lane_reward": 0.1,
"reward_speed_range": [0, 40],
"high_speed_reward": 1
```

Note that the reward for high speed is linearly mapped from the speed range `[0, 40]` to the reward range `[0, high_speed_reward]`.

The following is the reward configuration for training the AAs in the three-lane highway. Notably, the high-speed reward is doubled compared to that of the GAs.

```
"collision_reward": 1,
"reward_speed_range": [0, 40],
"high_speed_reward": 2,
"lane_change_reward": 0.5
```