Multi-Agent Reinforcement Learning Autonomous Driving Highway On-Ramp Merge

by

Larry Schester

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan-Dearborn
2023

Doctoral Committee:

Associate Professor Luis E. Ortiz, Co-Chair
Professor Yi Lu Murphey, Co-Chair
Associate Professor Sridhar Lakshmanan
Associate Professor Shengquan Wang

Larry Schester

lscheste@umich.edu

ORCID iD: 0000-0002-4395-765X

Dedication

To my mom, who always wanted us to go to college because she never had the chance.

To my kids, may this be an inspiration to follow your ambitions.

To my wife … I am finally done**.**

Table of Contents

Dedication ii

List of Tables vii

List of Figures ix

List of Acronyms xviii

List of Appendices xxi

Abstract xxii

Chapter 1 Research Description 1

1.1 Broader Impacts 3

1.2 Intellectual Merit 4

1.3 Merge Model 5

1.3.1 Standard On-Ramp Lengths 7

1.4 Testing Standards 9

1.5 State of the Art and Challenges 12

1.6 Research Phases 17

1.7 Multi-Agent Approach and Contributions 18

1.8 Related Work 19

Chapter 2 Foundational Simplified Approach Using a Grid Game 25

2.1 T-Grid Merge Game 26

2.2 T-Grid Game Reward Structure 27

2.2.1 Game Theory Formal Example of Rewards as Payoffs 29

2.3 T-Grid Evaluation: Turn-Taking 31

2.4 T-Grid Evaluation: Simultaneous Play 33
2.4.1 Game Theory Formal Example: The Game of Chicken 35
2.5 The Probability of a Collision 36
2.6 Summary 38
Chapter 3 Reinforcement Learning Simulation Using Q-Learning 39
3.1 Multi-Agent Q-Learning 40
3.1.1 Single-Action and Joint-Action Learners 42
3.2 States and Actions 42
3.2.1 Reasonable Acceleration Values 44
3.3 Q-Learning Merge Simulator 45
3.3.1 Training Simulator 47
3.3.2 Reward Function Design 49
3.3.3 Discretization 50
3.3.4 Q-Learning Simulator Variables 51
3.3.5 Q-Learning Simulator Equations 51
3.3.6 Simulator for Testing 52
3.3.7 Traffic Action Policies 54
3.4 Results, Discussion, and Limitations 55
3.4.1 Experimental Results Using the I-80 Dataset 58
3.5 Motivation for a Continuous Approach 59
Chapter 4 Deep Reinforcement Learning Approach 62
4.1 Approach and Simulator 64
4.1.1 Two-Vehicle Simulation States and Variables 65
4.1.2 Equations of Motion 66
4.1.3 Traffic Action Policies 66

4.1.4 Reward Function 67
4.2 Simulator DDPG DRL Network 68
4.2.1 Network Architecture 71
4.2.2 Training Using Noise 72
4.3 Test Setup 72
Chapter 5 Two-Vehicle Simulation 74
5.1 State Set 75
5.2 DRL Training 75
5.3 Reactive Policy Training for Traffic Vehicle 79
5.4 Standard Test Results 80
5.4.1 Constant Policy Results 82
5.4.2 Reactive Policy Results 84
5.4.3 Random Policy Results 86
5.4.4 Results for Single-Action Learner 87
5.5 Algorithmic Training Performance Evaluation 89
5.6 DRL Discussion 98
Chapter 6 Three-Vehicle Simulation 99
6.1 State Set 100
6.2 Three-Vehicle Simulator Functionality 102
6.2.1 Environment 104
6.2.2 Main 110
6.3 Training 111
6.4 Best Network Selection 115
6.5 Standard Test Results 120
6.5.1 Constant Policy 121

6.5.2 Reactive Policy 122
6.5.3 Random Policy 123
6.6 Three Vehicle-Simulation Discussion 124
Chapter 7 Full Scene Simulation 126
7.1 State Set 129
7.2 Full-Scene Simulator Functionality 129
7.3 Network Training and Performance 133
7.4 Training Performance 134
7.5 Best Network Selection 136
7.6 Standard Test Results 141
7.6.1 Constant Policy 142
7.6.2 Random Policy 143
7.6.3 Reactive Policy 144
7.7 Full Scene Discussion 145
Chapter 8 Contribution, Implications, and Future Work 146
8.1 Contributions 146
8.2 Domain, Application, and Approach 147
8.3 Results and Discussion 148
8.4 Future Work 154
8.5 Conclusion 157
Appendices 159
References 213

List of Tables

Table 1: 2019 NHTSA crash data, selected values 37

Table 2: Q-learning training simulator parameters and values. Training runs use multiple sets of parameters for the gap, speed, and TTP states. The table shows all the parameter sets used during training. 47

Table 3: Q-Learning testing simulator parameters and values 54

Table 4: Aggregate results for all simulation variants of Q-learning runs. 56

Table 5: DRL training simulator variables 66

Table 6: DRL testing simulator variables. 73

Table 7: Two-vehicle simulation state variables 75

Table 8: Improved training performance evaluation. The blue highlighted row at the bottom of the table is the network value chosen by observation of Figure 27, but many saved networks with lower total collisions exist. 90

Table 9: Three-vehicle simulation state parameters. Vehicles use different sets of state variables. Green check marks indicate state variable usage, and a red x indicates those not used. 101

Table 10: Simulation initialization parameters for ego merge vehicle and overall simulation .. 105

Table 11: Simulation initialization parameters for traffic vehicles 106

Table 12: Shows the 20 best performing tests in ascending values of total collisions. This is an excerpt from the full performance table. Simulator demonstration results use training weights from the best performer at 350K training episodes. 116

Table 13: Three-vehicle simulation performance evaluation breakdown by policy. The top three performers for each policy is shown. Highlighted rows are the same overall best performing DRL network. 116

Table 14: Full merge scene state parameters 129

Table 15: Merge and traffic vehicle simulation hyperparameters for all DRL simulation variants 132

Table 16: Full scene training performance with increasing episodes. Training data and standardized tests are performed at 25K intervals. The best performing network is marked in bold highlight and has the lowest merge rear (ego) collisions at 1,392, but also has the lowest total collisions at 4,389. 136

Table 17: First few rows of best performing individual policies. The best performing network remains consistent at 4.325M episodes and the top three generally maintain consistency too with exception to the reactive policy. 137

Table 18: Three-vehicle training performance, sorted by episode count. Poor collision performance is shown in the table from the beginning of training up to 175K episodes. 165

Table 19: Full scene vehicle DRL network save point performance metrics summary for traffic network showing 20 best performing networks (lowest collisions are best). The best performing traffic network, the one with the lowest number of total collisions, occurs at the same number of episodes as the merge vehicle network. 186

Table 20: Four-vehicle states. The ego merge state set increases by two (over the three-vehicle scene) to a total size of six variables. After significant experimentation, the traffic vehicle's state set differs from the ego merge state set. 203

Table 21: Four-vehicle simulation merge vehicles training performance top 10 performers. Save points and performance testing happen at 25K episode intervals. The best performing network in terms of overall collisions is at 1.95M episodes. 208

Table 22: Four-vehicle simulation training performance top 10 performers. This table shows data for the front and rear traffic vehicles. 208

Table 23: Top three performers for each of the individual policies for performance testing. 209

List of Figures

Figure 1: Multiple vehicle merge scene. The ego vehicle (blue) must position itself to avoid a collision while merging into a stream of traffic. The red traffic vehicle is considered the most important object in the multiple-vehicle road scene because it is closest in longitudinal position. 5

Figure 2: Simplified two-vehicle merge road scene to study fundamental interaction behavior of the merge vehicles. The merging vehicle shown in blue is the ego vehicle. The red vehicle is the traffic vehicle and cannot move into an adjacent lane, thus forcing longitudinal position control to avoid a collision. 6

Figure 3: US DOT (United States Department of Transportation) FHWA (Federal Highway Administration) MIRE (Model Inventory of Roadway Elements) https://safety.fhwa.dot.gov/tools/data_tools/mirereport/188.cfm 7

Figure 4: Real-world view of a highway merge on-ramp copied from Google Maps, March 6, 2023. Overhead view of the Southfield Freeway near Schoolcraft Road in Detroit, Michigan. The end of the red highlighting is the goal position. It is where the merge lane width becomes narrower than a standard lane width. The pictured on-ramp merge lane length is estimated at 100 meters. The vehicle at the end of the ramp has nearly reached the goal. A 57 m length is marked between the end of the painted lines within the merge lane and correlates to the same measurement shown in Figure 5. 8

Figure 5: A perspective view that a driver sees entering the on-ramp in Figure 4. The picture point-of-view is from the start of the 100 m on-ramp. The 57 m distance dimension in the picture is from the start of the painted merge line adjacent to the service drive to the end of the painted merge line adjacent to the highway. Image copied from Google Maps, March 6, 2023. 9

Figure 6: SAE levels of automated driving from their standard, SAE J3016 [5]. The levels have become commonly referred to as a standard for automation levels. Levels 0-2 are typically ADAS (advanced driver assistance systems) features. Levels 3-5 are automated driving. 11

Figure 7: T-Grid model. Each block represents a grid space that the vehicle can move into. The blue triangle is Agent 1 and represents the ego merge vehicle. The yellow triangle is Agent 2, the traffic vehicle. 26

Figure 8: T-Grid reward tree for ego actions. Each circle represents the grid block that the agent can move to. Next to the arrow is the immediate reward for each move. 28

Figure 9: Normal-form bi-matrix T-grid game representation rewards example. 29

Figure 10: T-Grid Turn Taking (left), Simultaneous Play (right)..... 32

Figure 11: Two-vehicle simultaneous play merge represented as a normal form game; the "game of chicken" 35

Figure 12: Two-vehicle Q-learning simulation road scene. State variables are numbered 1-5: Closing Gap, Closing Speed, TTP, Position of Ego, and Traffic Action. 43

Figure 13: Q-learning simulator block diagram. On the left is the training flow and on the right is the testing flow. Training and testing use two separate scripts..... 46

Figure 14: Output figure for Q-learning testing simulator. Green text with arrows identifying key features. 53

Figure 15: Average ego speed at the merge point. Plot of the average ego speed, over all runs, when merging into traffic. The plot is consistent with Proposition III, showing that it is advantageous for a lagging merge vehicle to decelerate and for a leading merge vehicle to accelerate..... 57

Figure 16: Q-learning data reshaped into a one-dimensional vector 60

Figure 17: Two-agent Q-learning data, positive values only..... 61

Figure 18: Closer look at positive values for two-agent Q-learning data to observe pattern formed by TTP states..... 61

Figure 19: Collision tables showing average collision for ego starting position with respect to the traffic vehicle. The results show that the DRL method produces collision-free runs at the 100 m goal position, whereas the Q-learning method has collisions at every ego start position. 64

Figure 20: Simulator graphical output. The blue rectangle represents the ego vehicle in the merge lane of a taper-type on-ramp. The red vehicle is the traffic vehicle. The green triangular flag represents the goal location: the first point where the vehicles are laterally close enough to collide. The simulation ends when the ego vehicle crosses the goal. 68

Figure 21: Deep dynamic policy gradient algorithm pseudocode from the paper by the Deepmind team [77]. 69

Figure 22: Actor-critic block diagram as illustrated by Sutton and Barto [81]. 70

Figure 23: Two-vehicle Ego DRL network diagram. 71

Figure 24: Two-vehicle simulation scene. Ego vehicle (blue) is in the merge lane. It learns through RL to avoid a collision with the traffic vehicle (red) while merging. The first point that the vehicles can collide is the goal position and is marked with a green flag. 74

Figure 25: Cumulative average of reward values during training. Various exploration values are compared to find the most suitable value. Exploration values are encoded into the graph legend, shown after the Ex. Visual observation finds the peak value and checks for over-fitting. 77

Figure 26: Joint-action-learner training plot. The cumulative average of rewards (blue) and the moving mean of rewards (orange) are shown. The moving mean uses a window of 30,000 data points. The peak value is at 1.12 x 10^6 episodes. Testing uses the saved network closest to the peak value. 78

Figure 27: Single-action learner training results. The cumulative average (blue) and moving mean (orange) are shown. The moving mean uses a 30,000 data point window. The peak value in the moving mean is 1.04 x 10^6 episodes. Performance testing uses the saved network at the peak. 79

Figure 28: Results matrix for a scenario where the traffic vehicle maintains a constant speed throughout the episode. In this evaluation, the ego vehicle considers the action of the traffic vehicle in addition to the state. 83

Figure 29: Ideal scenario for minimum collisions using constant action policy. Calculated using extreme acceleration values for the ego vehicle as the traffic vehicle speed is constant at 31.29 m/s (70 MPH) throughout the episode. 83

Figure 30: Test results using agents that are both reactive to each other. The collision tables show average collisions comparing the ego vehicle's starting position with respect to the traffic vehicle and the starting distance away from the goal position. In this responsive case, both vehicles take actions in response to one another. Each agent trains with its own unique DRL network. Both vehicles take action to avoid collisions with one another. 85

Figure 31: Manually calculated collisions using the extreme values of vehicle actions that produce the minimum collisions. This is the ideal case where both vehicles take extreme actions to avoid collisions. However, it is important to note that some collisions are unavoidable due to the physical limitations of the road scene and vehicle parameters. 85

Figure 32: Case where the ego merge vehicle reacts to a randomly acting traffic agent. The traffic vehicle chooses a random acceleration from -5 m/s^2 to 4 m/s^2 at each simulation time step. Each cell represents 30 episodes run at that combination of the initial position and goal position. 87

Figure 33: Single-action learner where the ego neural network only considers the state values and not the action of the traffic vehicle. Traffic has a constant zero acceleration action throughout each episode. 88

Figure 34: Response traffic policy. The ego and traffic vehicles use the same but independent network architecture. Neither considers the action of the other during training or action selection. 88

Figure 35: Random action policy for the traffic vehicle. The traffic vehicle selects an action from within the entire range of acceleration values at each 100 ms time step of each episode. 89

Figure 36: Multi-dimension plot of the standard test performance of the two-vehicle scenario. Performance testing occurs every 10K episodes for a total of 250 standard tests over the range of 2.5M episodes. Data is sorted by best to worst collision performance. Blue and orange bars indicate average deceleration and acceleration for each test instance. The green line is the difference between the acceleration and deceleration occurrence where negative values indicate greater deceleration than acceleration. The red line is on the secondary axis to the right and represents the total collisions for each blue-orange bar pair where a test is performed. 92

Figure 37: Two-vehicle deceleration occurrence and collisions correlation. As deceleration occurrence decreases, collisions increase. However, the best performance has a more scattered grouping of data points, suggesting that the best performance is influenced by more than just deceleration action selection. 93

Figure 38: Ego merge vehicle average deceleration action value..... 94

Figure 39: Two-vehicle Ego acceleration average values by start versus goal position. 95

Figure 40: Two-vehicle Ego deceleration average values by start position and goal position..... 96

Figure 41: Two-vehicle scenario Ego Merge vehicle average acceleration values compared to collisions. 97

Figure 42: Three-vehicle simulation. One ego merge vehicle (blue) and two traffic vehicles (red). The fundamental setup is the same as the two-vehicle scene, but the complexity increases significantly for the set of state variables and lines of simulator code. 99

Figure 43: Three-vehicle DRL network diagram for the merge vehicle. 100

Figure 44: Three-vehicle simulation rending output. At the bottom is the road scene with ego in blue and two reactive policy traffic vehicles in green. At the top left are position, velocity, and ego TTP graphs. At the right is the output from the main program showing the configuration of each episode and whether a collision has occurred. 110

Figure 45: Cumulative reward average of four separate exploration values. 113

Figure 46: Cumulative reward average and moving mean for best variant: 0.999995 explore.. 114

Figure 47: Correlation plot comparing collisions to deceleration occurrence. Blue dots indicate deceleration action selection occurrence and the number of collisions per test set. Data points show two linear groupings extending horizontally near the 60% occurrence and another linear grouping extending downward and to the right. 117

Figure 48: Three-vehicle simulation deceleration occurrence by Ego initial start position versus goal position at a 50 m starting gap for the two traffic vehicles. 118

Figure 49: Standard test performance of the three-vehicle scenario with data sorted by best to worst collision performance. Blue and orange bars indicate average deceleration and acceleration for each test instance. The green line is the difference between the acceleration and deceleration

occurrence where negative values indicate greater deceleration than acceleration. The red line is on the secondary axis to the right and represents the total collisions for each blue-orange bar pair where a test is performed. 119

Figure 50: Standard test table for three-vehicle constant policy. Test results are expanded from two-vehicle to show multiple gap settings of 5m, 15m, and 50m. 122

Figure 51: Three-vehicle reactive policy standard test table. 123

Figure 52: Three-vehicle random-acting traffic policy standard test table. 124

Figure 53: Full scene merge scenario. Two vehicles in the merge lane and two or more vehicles in the traffic lane. During training and testing, as the rear merge vehicle moves in front of or behind the traffic vehicle pair, another vehicle is generated to ensure there is always a traffic vehicle in front of and behind the ego vehicle. 126

Figure 54: Full-scene actor-critic DRL network diagram 128

Figure 55: Training graph of moving mean and cumulative average for rewards. Data in the lower graph is for both front and rear merge vehicles. The upper graph shows the closest front and rear traffic vehicles to the rear merge vehicle. Near seven million episodes, the acceleration action for both the merge and traffic vehicles settle to continuously choose an action limit value of -5 m/s acceleration, regardless of the state value. 134

Figure 56: Ego deceleration action selection occurrence for all saved networks during training. 138

Figure 57: Average acceleration and deceleration for the full scene set of performance tests over the duration of training 139

Figure 58: Multi-dimensional plot of acceleration and deceleration values over the course of training for the full scene. The plot is shown from a time perspective, meaning from left to right the episodes increase. 140

Figure 59: Full Scene constant policy standard test table. 143

Figure 60: Full scene random policy standard test results for vehicle starting gaps of 5, 15, and 25 meters. 144

Figure 61: Full Scene reactive policy standard test tables for 5, 15, and 25-meter initial starting gaps between vehicles. 145

Figure 62: Two-vehicle simulation Ego merge acceleration occurrence. Data points show the average occurrence of an acceleration action selection within each episode. 160

Figure 63: Two-vehicle Ego occurrence of deceleration action selection by initial position and goal position. Results are based on the SAL example with a selected network of 1.04×10^6 training episode. The Ego vehicle has learned to decelerate for almost all combinations with Ego

initial positions of -1 m or less. Since the network only selects acceleration or deceleration values, acceleration occurrence is the compliment of deceleration occurrence. This aligns with Proposition III: it is advantageous for a lagging vehicle to decelerate and for a leading vehicle to accelerate..... 161

Figure 64: Three-vehicle DRL network for traffic vehicles. 162

Figure 65: Three-vehicle standard testing sorted by training order. This is the same plot as Figure 49, but with a different sort order. Horizontal axis values indicate the test number, each representing 25K episodes. The leftmost point is at the first testing instance at 25K episodes, ascending to the rightmost value at 2.5 million episodes. 163

Figure 66: Three-vehicle Ego deceleration occurrence average values by start position and goal position averaged over all traffic gap values. 166

Figure 67: Three-vehicle Ego deceleration occurrence average for 100 m initial traffic gap. ... 167

Figure 68: Three-vehicle Ego deceleration occurrence average for 50 m initial traffic gap. 167

Figure 69: Three-vehicle Ego deceleration occurrence average for 25 m initial traffic gap. 168

Figure 70: Three-vehicle Ego deceleration occurrence average for 15 m initial traffic gap. 168

Figure 71: Three-vehicle Ego deceleration occurrence average for 10 m initial traffic gap. 169

Figure 72: Three-vehicle Ego deceleration occurrence average for 5 m initial traffic gap. 169

Figure 73: Three-vehicle Ego acceleration occurrence average for all traffic gaps. 170

Figure 74: Three-vehicle Ego acceleration occurrence average for 100 m initial traffic gap. ... 170

Figure 75: Three-vehicle Ego acceleration occurrence average for 50 m initial traffic gap. 171

Figure 76: Three-vehicle Ego acceleration occurrence average for 25 m initial traffic gap. 171

Figure 77: Three-vehicle Ego acceleration occurrence average for 15 m initial traffic gap. 172

Figure 78: Three-vehicle Ego acceleration occurrence average for 10 m initial traffic gap. 172

Figure 79: Three-vehicle Ego acceleration occurrence average for 5 m initial traffic gap. 173

Figure 80: Three-vehicle Ego deceleration average values by start position and goal position for all traffic gap values..... 174

Figure 81: Three-vehicle Ego deceleration average for 100 m initial traffic gap..... 174

Figure 82: Three-vehicle Ego deceleration average for 50 m initial traffic gap..... 175

Figure 83: Three-vehicle Ego deceleration average for 25 m initial traffic gap..... 175

Figure 84: Three-vehicle Ego deceleration average for 15 m initial traffic gap. 176

Figure 85: Three-vehicle Ego deceleration average for 10 m initial traffic gap. 176

Figure 86: Three-vehicle Ego deceleration average for 5 m initial traffic gap. 177

Figure 87: Three-vehicle Ego acceleration average values by start versus goal position. 177

Figure 88: Three-vehicle Ego acceleration average for 100 m initial traffic gap. 178

Figure 89: Three-vehicle Ego acceleration average for 50 m initial traffic gap. 178

Figure 90: Three-vehicle Ego acceleration average for 25 m initial traffic gap. 179

Figure 91: Three-vehicle Ego acceleration average for 15 m initial traffic gap. 179

Figure 92: Three-vehicle Ego acceleration average for 10 m initial traffic gap. 180

Figure 93: Three-vehicle Ego acceleration average for 5 m initial traffic gap. 180

Figure 94: Three-vehicle Ego scatter plot for acceleration average versus collisions. 181

Figure 95: Three-vehicle Ego scatter plot for acceleration occurence versus collisions. 181

Figure 96: Three-vehicle Ego scatter plot for deceleration occurence versus collisions. 182

Figure 97: Three-vehicle Ego scatter plot for deceleration average versus collisions. 182

Figure 98: Three-vehicle simulation collision performance test results for 10, 25, and 100 m initial traffic gaps, constant traffic action. 183

Figure 99: Three-vehicle simulation collision performance test results for 10, 25, and 100 m initial traffic gaps, reactive traffic action. 183

Figure 100: Three-vehicle simulation collision performance test results for 10, 25, and 100 m initial traffic gaps, random traffic action. 184

Figure 101: Full-scene DRL network diagram for traffic vehicles 185

Figure 102 Full-scene Ego scatter plot for acceleration occurence versus collisions 187

Figure 103: Full-scene multi-dimension plot for merge lane front vehicle. 187

Figure 104: Full-scene multi-dimension plot for traffic lane rear vehicle. 188

Figure 105: Full-scene multi-dimension plot for traffic lane front vehicle. 188

Figure 106: Full-scene Ego deceleration occurrence average for all initial traffic gaps. 189

Figure 107: Full-scene Ego deceleration occurrence average for 100 m initial traffic gap..... 189

Figure 108: Full-scene Ego deceleration occurrence average for 50 m initial traffic gap..... 190

Figure 109: Full-scene Ego deceleration occurrence average for 25 m initial traffic gap..... 190

Figure 110: Full-scene Ego deceleration occurrence average for 15 m initial traffic gap..... 191

Figure 111: Full-scene Ego deceleration occurrence average for 10 m initial traffic gap..... 191

Figure 112: Full-scene Ego deceleration occurrence average for 5 m initial traffic gap..... 192

Figure 113: Full-scene Ego deceleration average for all initial traffic gaps..... 192

Figure 114: Full-scene Ego deceleration average for 100 m initial traffic gap. 193

Figure 115: Full-scene Ego deceleration average for 50 m initial traffic gap. 193

Figure 116: Full-scene Ego deceleration average for 25 m initial traffic gap. 194

Figure 117: Full-scene Ego deceleration average for 15 m initial traffic gap. 194

Figure 118: Full-scene Ego deceleration average for 10 m initial traffic gap. 195

Figure 119: Full-scene Ego deceleration average for 5 m initial traffic gap. 195

Figure 120: Full-scene Ego acceleration average for all initial traffic gaps. 196

Figure 121: Full-scene Ego acceleration average for 100 m initial traffic gap. 196

Figure 122: Full-scene Ego acceleration average for 50 m initial traffic gap. 197

Figure 123: Full-scene Ego acceleration average for 25 m initial traffic gap. 197

Figure 124: Full-scene Ego acceleration average for 15 m initial traffic gap. 198

Figure 125: Full-scene Ego acceleration average for 10 m initial traffic gap. 198

Figure 126: Full-scene Ego acceleration average for 5 m initial traffic gap. 199

Figure 127: Full-scene simulation collision performance test results for 10, 50, and 100 m initial traffic gaps, constant traffic action..... 199

Figure 128: Full-scene simulation collision performance test results for 10, 50, and 100 m initial traffic gaps, random traffic action..... 200

Figure 129: Full-scene simulation collision performance test results for 10, 50, and 100 m initial traffic gaps, reactive traffic action. 200

Figure 130: Four vehicle merge scene with two vehicles in the merge lane and two in traffic. 202

Figure 131: Four-vehicle DRL network training graphs. The best performing network is around 1.95M episodes and correlates with the peaking of the moving mean lines..... 207

Figure 132: Multi-dimension plot for the Ego (merge rear) vehicle sorted by training episode order. Snapshots taken at 25K intervals over the entire 5.175M episodes of training. Orange bars indicate average acceleration, blue average deceleration per test interval. Red line indicates number of collisions for the merge rear vehicle. 210

Figure 133: Four-vehicle standard test table results for all traffic action policies. The table is similar to other standard result tables but adds an initial gap. The initial gaps are 5, 15, and 25 meters from left to right, correlating to a one, three, and five vehicle length gap. 211

List of Acronyms

ACC: adaptive cruise control

AD: autonomous or automated driving

ADAS: advanced driver assistance system

ADS: automated driving system

AI: artificial intelligence

AV: autonomous vehicle

CA: California

CNN: convolutional neural network

DDPG: deep dynamic policy gradient

DNN: deep neural network

DOT: Department of Transportation

DQN: Deep-Q Network

DRL: deep reinforcement learning

FHWA: Federal Highway Administration

FMVSS: federal motor vehicle safety standard

GM: General Motors

HAD: highly autonomous driving

IEEE: Institute of Electrical and Electronics Engineers

IL: independent learner

ITS: intelligent transportation society

ITSC: Intelligent Transportation Society Conference

IV: Intelligent Vehicle symposium

JAL: joint action learner

KPH: kilometers per hour

L0, L1, L2, L3, L4, or L5: level 0, level 1, …, level 5 (automated driving levels)

LSTM: long short-term memory

MA: multi-agent

MAAS: Mobility as a Service

MARL: multi-agent reinforcement learning

MDP: Markov Decision Process

MIO: most important object

MIRE: Model Inventory of Roadway Elements

ML: machine learning

MPH: miles per hour

MUTCD: Manual on Uniform Traffic Control Devices

NGSIM: Next Generation Simulation

NHTSA: National Highway transportation safety administration

NN: neural network

PGM: probabilistic graphical model

POMDP: partially observable Markov decision process

RL: reinforcement learning

RPM: revolutions per minute

SA: single agent

SAE: society of automotive engineers

SAL: single-agent learner

TD: temporal-difference

TIV: time in between vehicles

TTP: time to position

UMTRI: University of Michigan Transportation Research Institute

V2x: vehicle to everything

WHO: World Health Organization

List of Appendices

Appendix A Additional Tables and Figures 160
A.1 Two-Vehicle Training Test Performance 160
A.2 Three-Vehicle Traffic DRL Network Diagram 162
A.3 Three-Vehicle Best Network Performance Testing Tables 163
A.4 Three-Vehicle Simulation Best Network Selection Additional Details 166
A.5 Three-Vehicle Simulation Results for Additional Traffic Test Gaps 183
A.6 Full-Scene Traffic DRL Network Diagram 185
A.7 Full Scene Simulation Best Network Selection Additional Details 186
A.8 Full Scene Additional Test Gaps 199
Appendix B An Intermediate Step to the Full-Scene Simulation 201
B.1 State Set 203
B.2 Four-Vehicle Simulator Functionality 204
B.3 Training and Performance 206
B.4 Best Network Selection 207
B.5 Standard Test Results 211
B.6 Discussion 212

Abstract

Autonomous driving is expected to become more common in the future. Autonomous vehicles operate today in limited use cases like highway driving and in major cities as robotaxis, but full L5 operation is yet to be achieved. Functions like fully autonomous highway ramp entry must be available, safe, and reliably robust in a provable way to bridge the gap to enable full autonomy. Towards this goal, my research produces three main contributions: a fundamental study of on-ramp merging defining specific ever-present behaviors, a standard test framework for evaluating merging performance, and a multi-agent DRL simulation that has learned to operate with nearly ideal collision avoidance performance. The fundamental study shows behaviors and limitations that exist with all merging. A standard test framework I developed evaluates performance and compares different approaches. The virtual environment of the multi-agent DRL uses self-play with simulated data where merging vehicles safely learn to control longitudinal position during a taper-type merge. The initial simulation setup is a two-vehicle merge-traffic pair, then it is progressively scaled up to a full merge scene. The simulation results show nearly perfect performance that is likely best-in-class if it were able to be compared against other research using the standard evaluation framework.

Chapter 1 Research Description

Highly autonomous driving is expected to save lives and increase mobility. The foreseeable future of autonomous driving will be of mixed usage with roadway sharing between human drivers and autonomous vehicles of varying capabilities. Autonomous vehicles need to act like humans to be successful in this mixed-usage environment. Ideally, autonomous vehicles must perform maneuvers as well as or better than humans and achieve a goal of zero collisions.

This research applies state-of-the-art techniques from distributed reinforcement learning and decision-making in multi-agent environments to improve individual autonomous vehicle controllers. The research is mathematically grounded in a game-theoretic approach that accounts for the joint reaction of the host and other vehicles. The problem of highway ramp lane merging is a motivating example. This approach uses multi-agent reinforcement learning (MARL) to train longitudinal position control an automated vehicle to avoid collisions during a highway merge.

The longitudinal position of the ego vehicle agent in the merge lane is controlled through speed changes to avoid a collision with the traffic vehicle agents as they reach the merge point. Simulated data trains the DRL network for the merging vehicle. The data is a range of merge scene parameters and three different action behavior policies: constant, random, and reactive acceleration actions of the traffic agents. Finally, the same three traffic action policies test the trained agent, but with some parameters held constant using a structured, standardized test framework specifically developed to measure automated merging performance.

An early version of the simulation was compared to state-of-the-art automated merging on the standard NGSIM I-80 highway dataset and performed at a collision rate below 1%, which is significantly better than the benchmark 7% collision rate [1], [2]. Since then, my simulator performance has improved substantially. However, further study was not conducted using datasets because it provides limited value. This is for two reasons: 1. the current body of published research does not use a standard evaluation framework to compare against, and 2. vehicles in datasets act like a ghost in the scene. In addition, in most other studies, the ego merge vehicle is not multi-agent and does not evoke reactive actions by other agents.

The simulation was initially a two-vehicle merge-traffic pair, then progressively scaled up to a full merge scene. To accomplish this, vehicles are added individually in each lane to systemically build up the simulation. Each progressive step is meticulously studied to ensure a thorough understanding of the implications of increasing complexity. A real-world scene can be studied using the method, but it is outside the scope of this research.

The first chapter summarizes the problem, motivation, approach, merit, broader impacts, and related work. The related work section emphasizes similarities in other work and highlights gaps that created the opportunity to conduct my research. Chapter 2 presents a simplified mathematical approach foundational to understanding agents' basic behaviors during merging. Chapter 3 presents a simplified two-vehicle simulation using Q-learning that correlates with the foundational work from Chapter 2. The work in Chapter 2 and Chapter 3 was presented at IEEE ITSC 2019 (Intelligent Transportation Society Conference) and published in its proceedings [3]. Chapter 4 to Chapter 7 present a deep reinforcement learning (DRL) approach that improves upon the two-vehicle Q-learning work, then progressively scales up to a full scene representative of a real-world merge scene. The DRL work produces nearly ideal performance. The

performance is measured using a framework I created to measure merging performance for any approach. The DRL work of Chapter 4 and Chapter 5 was presented at IEEE IV 2021 – IEEE Intelligent Vehicles Symposium and published in its proceedings [4].

1.1 Broader Impacts

Researchers have been interested in Autonomous Driving (AD) technology for many years. The DARPA Grand Challenges in 2004, 2005, and 2007 spurred research into autonomous driving. This research continues due to governments' interest in making safer vehicles, consumers' interest in these high-tech features, and the opportunity for companies to sell these systems.

The Society of Automotive Engineers (SAE) defines several levels of automated driving [5]. The levels range from L0 (Level Zero) to L5 (Level 5), where L0 is a vehicle fully controlled by the driver and L5 is a fully driverless car requiring no human control. My research helps enable L4-L5 functionality.

A goal of autonomous driving is to reduce accidents and make driving safer. According to NHTSA, in 2019, there were over 36,000 deaths in the U.S. from motor vehicle crashes with 8.7% being caused by distracted driving [6]. The WHO reports 1.35 million annual traffic deaths worldwide [7]. These numbers are alarming, but autonomous driving is expected to make driving safer by reduce crashes and fatalities. For example, GM describes its vision of the future as ‘Zero Crashes, Zero Emissions and Zero Congestion.’ [8]

In addition to making driving safer, fully autonomous driving can increase mobility for humankind by enabling adjacent technologies like MAAS (Mobility as a Service). Autonomous driving MAAS could enable more widespread access to those who cannot drive. Highly autonomous driving also provides mobility for those unable to drive themselves, e.g., blind,

disabled, youth, or elderly. MAAS is a key goal of many companies in the automated driving realm (e.g., Waymo and Cruise).

1.2 Intellectual Merit

A fully automated merge maneuver is necessary for highly autonomous driving (HAD). Today some systems can handle highway driving relatively well – mainly because highways are typically well structured, consistently marked, and have a stream of traffic all heading in the same direction at relatively the same speed. However, merging is a complicated maneuver that requires a driver to understand the road scene quickly, accelerate to assimilate traffic speed, yet maneuver to avoid a collision, all in a matter of a few seconds or less while traveling in a lane that terminates into another lane of traffic.

It is not difficult to imagine an idealistic merge where a vehicle can reach traffic speed and move onto the highway unaffected by any other traffic. However, experience tells us that merging vehicles must react to each other when they are close together. This reaction enables a successful merge, one without a collision. The reactiveness during merging is common on highways in densely populated areas.

The challenges of merging are relatively straightforward for humans to understand. To successfully merge into traffic, a driver must quickly assess the road scene to determine the best way to control the vehicle to insert it into a traffic stream while judging and reacting to the behavior of the other vehicles, all while avoiding a collision. It is a complex task for humans to learn and no less complex to automate. The task of on-ramp merging needs a relatively higher cognitive workload, compared to other highway driving maneuvers like lane merging [9].

1.3 Merge Model

A representative model is at the foundation of the merge maneuver study. The progressive development of the merge model is the central theme of this research. The model provides a framework, enabling the study of different approaches to find optimal behavior, seek human-like performance, and scale the approach to handle complex multi-vehicle realistic road scenes. A critical challenge to automating merging is capturing the essence of the merge maneuver as a model.

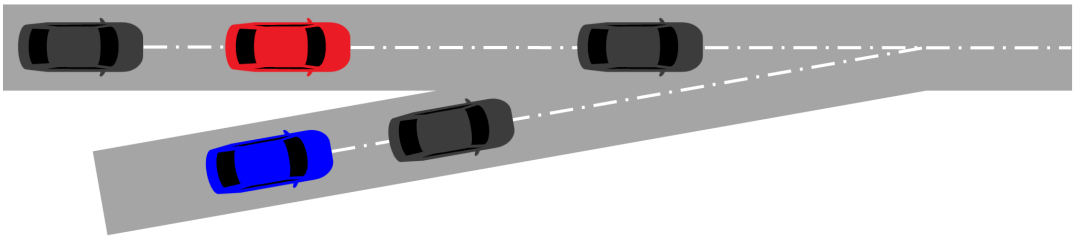


Figure 1: Multiple vehicle merge scene. The ego vehicle (blue) must position itself to avoid a collision while merging into a stream of traffic. The red traffic vehicle is considered the most important object in the multiple-vehicle road scene because it is closest in longitudinal position.

A realistic scenario for on-ramp merging study has multiple vehicles. The merge vehicle approaches a busy highway with multiple vehicles traveling down the lane to enter. The merge vehicle must negotiate its way into traffic by controlling its position while entering the lane and avoiding collisions. Figure 1 represents this complex scene and is my research's ultimate scope of study. It mimics a complicated, yet realistic merge scene that is often particularly near high-density areas. It has multiple parameters that influence the behavior and interaction of any one of the vehicles, making it complicated to model. Unrelated interactions become intertwined without the right model. The right model allows intelligent information about the scene to be extracted, understood, and controlled.

The initial study model focuses on a simplified scenario of a single-vehicle pair. This simplified scenario is used to understand the most basic essence of the merge behavior. The two

vehicles are approaching a common merge point that intersects where the merge vehicle lane meets the traffic vehicle lane. Figure 2 shows the scene. To limit the control parameters, the vehicles can only travel in a straight line along the middle of their lane. Each vehicle controls its speed by accelerating or decelerating. The traffic vehicle (shown in red) cannot move into an adjacent lane. This creates a situation where the vehicles must control their longitudinal position by adjusting their speed to avoid a collision. The fundamental learnings from the interactions of this single-vehicle pair will provide a foundation upon which to build.

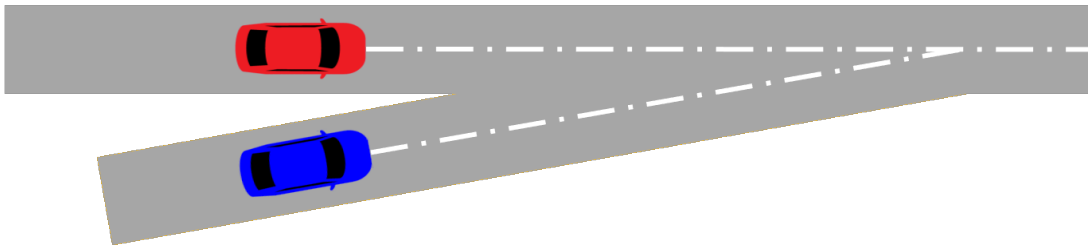


Figure 2: Simplified two-vehicle merge road scene to study fundamental interaction behavior of the merge vehicles. The merging vehicle shown in blue is the ego vehicle. The red vehicle is the traffic vehicle and cannot move into an adjacent lane, thus forcing longitudinal position control to avoid a collision.

The merge study is limited to a taper-type on-ramp. Figure 2 shows this taper-type ramp. Another typical on-ramp configuration is the parallel type, described in the AASHTO *Green Book* [10], where a parallel merge lane runs alongside the traffic lane for the merge vehicle to choose the appropriate position to merge into traffic. The parallel type is not studied here because it is more of an adaptation of a typical lane change maneuver with a diminishing time window – a less complex scenario than the taper type. Other configurations, like an on-ramp lane that connects to an off-ramp or multiple lane on-ramps, are also out of scope.

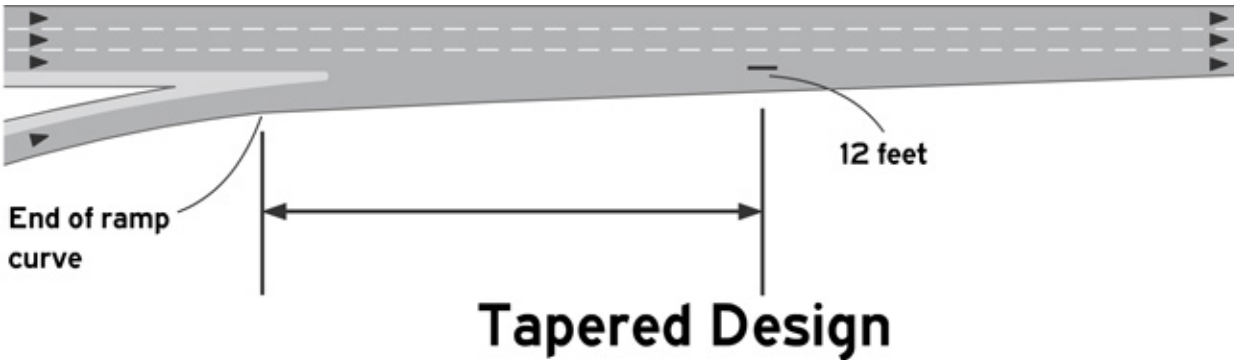


Figure 3: US DOT (United States Department of Transportation) FHWA (Federal Highway Administration) MIRE (Model Inventory of Roadway Elements) https://safety.fhwa.dot.gov/tools/data_tools/mirereport/188.cfm

1.3.1 Standard On-Ramp Lengths

The US DOT Federal Highway Administration publishes standards and guidance on roadway design in the Manual on Uniform Traffic Control Devices (MUTCD) [11]. For a taper-type on-ramp, like the one Figure 3 shows, the MUTCD gives guidance to taper length, L, using a formula of $L = WS$ for roadways with a speed limit of 70 KPH (45 MPH) and above, where W is the road width and S is the speed limit in MPH. The standard test speed used in my merging framework is 70 MPH, a typical highway speed in the U.S.[12] Typical road widths in the U.S. are 12 feet. This formula with these parameters gives a recommended taper length of 840 feet (256 meters). This recommendation is well above the longest ramp length of 100 m tested in my work. Yet, the performance of the latest algorithm shows no collisions well under 100 m, so my merging method conservatively covers the recommended ramp length. However, the same MUTCD in Chapter 3B gives a standard that states, “The minimum taper length shall be 30 m (100 ft) in urban areas and 60 m (200 ft) in rural areas.”

Figure 4: Real-world view of a highway merge on-ramp copied from Google Maps, March 6, 2023. Overhead view of the Southfield Freeway near Schoolcraft Road in Detroit, Michigan. The end of the red highlighting is the goal position. It is where the merge lane width becomes narrower than a standard lane width. The pictured on-ramp merge lane length is estimated at 100 meters. The vehicle at the end of the ramp has nearly reached the goal. A 57 m length is marked between the end of the painted lines within the merge lane and correlates to the same measurement shown in Figure 5.

Figure 4 shows a specific example of on-ramp merging. It is an overhead view of a highway on-ramp merge onto the Southfield Freeway near Schoolcraft Road in Detroit, Michigan. The figure has the same overhead view of a merge scene as Figure 24 and others shown throughout this dissertation. Figure 5 is a street-level view of the same on-ramp Figure 4 shows.

The on-ramp has an estimated length of 100 meters. The figure shows the measurements. The Southfield Freeway has a speed limit of 55 MPH, so using the $L = WS$ guideline yields $L = 12 \times 55 = 660\ feet$ or 201 meters for a recommended taper length. This particular on-ramp has a very short taper length that is well under the $L = WS$ guideline in the MUTCD and similar references. Southfield Freeway has several similar taper-type on-ramps that are short like this one.

Figure 5: A perspective view that a driver sees entering the on-ramp in Figure 4. The picture point-of-view is from the start of the 100 m on-ramp. The 57 m distance dimension in the picture is from the start of the painted merge line adjacent to the service drive to the end of the painted merge line adjacent to the highway. Image copied from Google Maps, March 6, 2023.

The overhead view, like that shown in Figure 4, can be deceiving for perceived distances. The street-level view in Figure 5 shows a perspective that a driver would see and better represents the short on-ramp length for merging. Despite the perceive shortness of the on-ramp, the results from Section 5.4 show that merging into traffic with this on-ramp can be successfully performed without collision using the DRL approach. Future chapters with scaled-up road scenes show similar results.

1.4 Testing Standards

Testing standards for autonomous vehicles are relatively non-existent. Some testing standards exist, like SAE J3018 [13], but autonomous vehicle testing is in its relative infancy in comparison to other safety systems. In comparison, testing standards for other safety critical vehicle components, like braking and steering, are abundant. In the U.S. there are many FMVSS (federal motor vehicle safety standards) that apply to multiple aspects of vehicle safety, like

crashworthiness, restraints, seats, warning devices, braking, tires, etc. Autonomous Driving is still in its early stages with its first demonstrations of viability in DARPA Grand Challenges less than two decades ago and commercialization of sub-human performance systems from the likes of Waymo, Cruise, and Tesla only within the past few years. In comparison, vehicle safety systems like braking and steering have been around for over a century, so their testing standards have had much more time to mature and continue to evolve today. Most of these regulations and standards were put in place well after their initial introduction, like FMVSS standards that were only introduced in the 1960s, despite motor vehicles being around since before 1900. The development and adoption of testing and safety standards for ADAS and AD are happening much faster, but it could still take quite some time for them to mature to the level of other standards.

Within the automotive industry, it is common to test a new or problematic feature to one million miles to gain confidence that it is acceptable. In early 2023, Cruise and Waymo (both leaders in autonomous driving) reached one million miles without a driver [14], [15]. Current statistics in the U.S. show that there is about one collision per 500,000 miles travelled and about one fatality per 100 million miles driven [6], indicating that many more miles may need to be travelled to truly demonstrate safety greater than a human driver. However, accumulating a certain number of miles without a clear test strategy or structure may not truly verify or validate that feature in a safe, robust and reliable way [16]. For example, edge cases can be missed or the test scenario may have specific segments that are repetitive that do not ultimately prove acceptability.

The lack of standard testing and preference of brute force testing leads to challenges accumulating so many miles in the current fast-paced product introduction market. This has led

to innovations like digital twins that simulate instances of known troublesome scenarios for autonomous vehicles to operate [17]. These simulations can increase the effectiveness of training while reducing the total number of test miles, but they are effectively stand-ins in lieu of true test standards like those developed by regulatory boards. Ultimately, it is generally difficult to judge the performance of AI, especially when performance is expected to be greater than human performance, like in the case of Autonomous Driving. The well-known Turing test compares AI's ability to act like a human, but there is no comparable measure of performance today within Autonomous Driving.

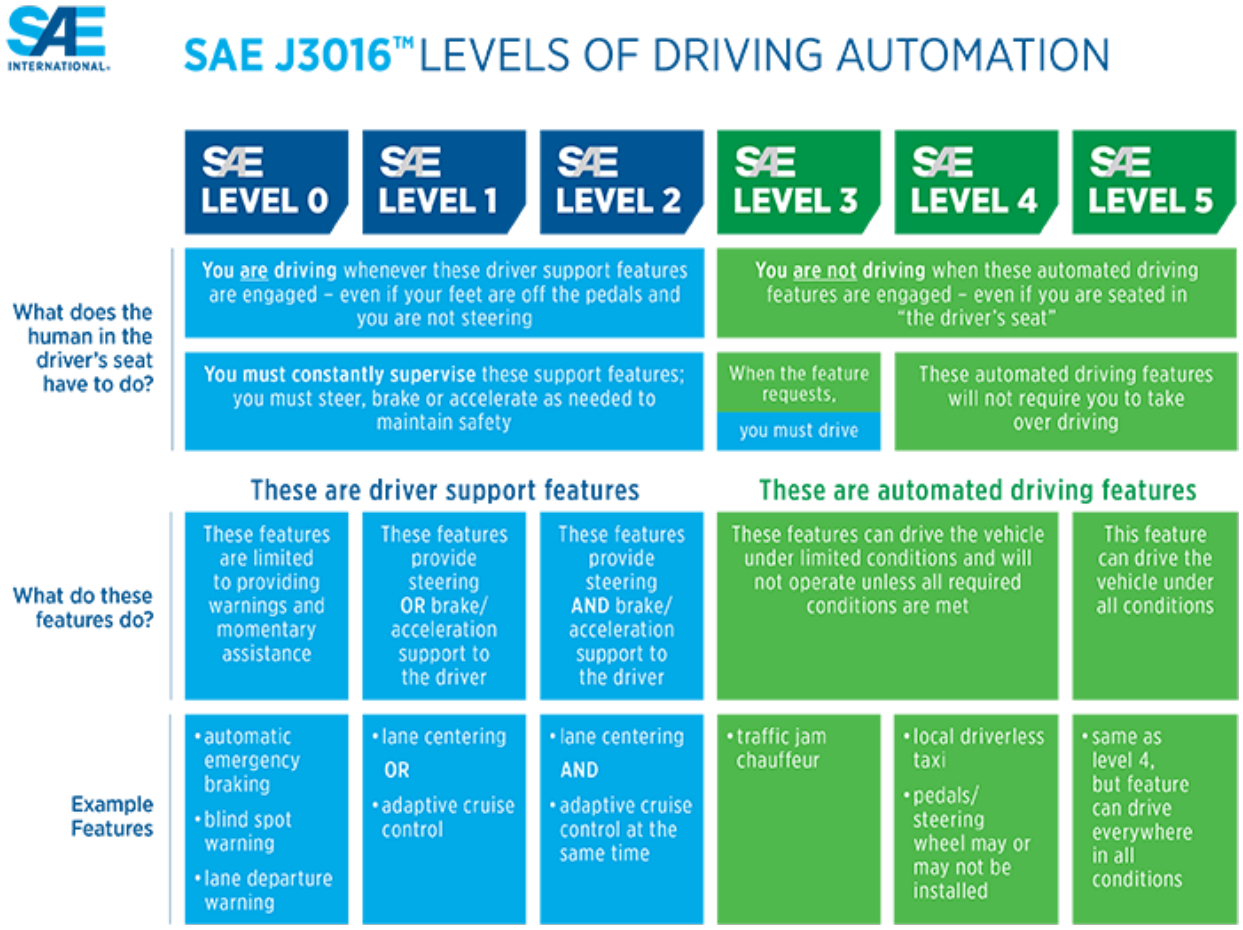


Figure 6: SAE levels of automated driving from their standard, SAE J3016 [5]. The levels have become commonly referred to as a standard for automation levels. Levels 0-2 are typically ADAS (advanced driver assistance systems) features. Levels 3-5 are automated driving.

1.5 State of the Art and Challenges

Many companies are developing and using automated driving technology. Some companies, like GM, Ford, and Tesla, have Level 2+ systems operating on the road today in vehicles purchased by everyday consumers[1]. Other companies like Cruise and Waymo have captured fleets of Level 4 Robotaxis available for public use, but their vehicles are not for public purchase. The Level 4 Robotaxis are geofenced to specific areas in cities like San Francisco and Phoenix. They operate driverless, i.e., without a safety driver in the vehicle. Despite the growing usage and increasing availability of automated driving vehicles, they can still not handle many challenges humans often handle with relative ease, like a Level 4 highway traffic merge.

Many consider Tesla to have state-of-the-art self-driving capability. Tesla seems to claim that its system is capable of handling on-ramps. It is part of their automated driving system, called "Full Self-Driving." They describe their *Navigate on Autopilot* feature by saying it "actively guides your car from on-ramp to off-ramp…", but also label the feature as beta [18]. When they introduced it in 2018, they described it as using a combination of two other features automatically: "Traffic-Aware Cruise Control" and "Autosteer" [19]. Overall, the system seems relatively rudimentary in functionality and requires constant human oversight and intervention with little warning. However, in legal claims and government records, Tesla clearly states that their vehicles are only Level 2 (advanced ADAS). Tesla describes their overall system as follows:

> 'The currently enabled Autopilot and Full Self-Driving features require active driver supervision and do not make the vehicle autonomous. Full autonomy will be dependent on achieving reliability far in excess of human drivers as demonstrated by billions of miles of experience, as well as regulatory approval.' [18]

[1] Level 2+ is a term used in the automotive industry to describe a vehicle that is capable of Level 2 features or above without claiming functionality at Level 3 or above. The features by level are described in detail by the SAE J3016 standard [5].

There have been instances where human drivers crash into automated driving vehicles stopped at the end of a highway merge ramp. For example, in 2018, the news media reported an accident involving an autonomous vehicle from Apple that was hit from behind by an on-coming driver while waiting to enter a highway [20]. Stopping to wait for a "safe" space to enter traffic may seem logical to avoid an accident, but this behavior is abnormal to humans and leads to a less safe scenario. In addition, stopping to wait for traffic at the end of an on-ramp requires the stopped vehicle to regain lost speed when resuming the merge maneuver. This forces an even greater break in the traffic flow before entering and potentially impedes traffic flow if the merging vehicle cannot quickly reach highway speeds before cutting into traffic.

A common belief is that Waymo's autonomous vehicles have the most advanced functionality and highest performance. However, user videos of passengers riding in the driverless Waymo vehicles have shown unexpected behavior in situations that human drivers would consider trivial. In a video from May 2021, a Waymo driverless vehicle becomes confused by cones in the roadway while waiting to take a right-hand turn [21]. Frustrated drivers behind the vehicle recognize the simplicity of resolving the situation and become agitated. Vehicles move around the stuck Waymo vehicle and easily negotiate the cones. Later in the video, the Waymo vehicle again becomes confused and stuck by cones placed in between two lanes. The Waymo vehicle stops in the middle of the road and partially blocks a busy traffic lane, creating a dangerous situation. The video is a good example of a state-of-the-art driverless system that cannot handle situations that are easy for human drivers to resolve.

For autonomous driving to become more ubiquitous, the expectation is that these autonomous systems will be safer than human drivers. Many academic institutions and companies are active in this area and AI-enabled technologies continue to grow in usage and

popularity [22]. Systems like GM's SuperCruise or Ford's BlueCruise automate routine and mundane tasks like normal highway driving. However, these systems are limited to operating on well-marked, well-mapped, predictable roadways, with vehicles that mostly travel the same speed and always in the same direction. These systems automate driving that is somewhat simple for humans and does not require a high cognitive workload but does require concentration.

Error and poor judgment are human factors often cited as the cause of many accidents [23]. Fewer accidents and safer driving through automation are important motivations for autonomous driving, but when autonomous systems do not operate properly, they still need to operate safely. Predictably safe operation may be the most challenging task of all. Therefore, *intelligent* behavior in automated systems is essential.

With autonomous driving, humans expect intelligent behavior out of vehicles that do not have human operators. Conversely, the behavior is surprising when automated vehicles cannot perform tasks simple for humans. However, as Philip Koopman, an expert in autonomous driving safety, writes, "Autonomous systems struggle with novelty, unknowns [16]." This perception of irrational or non-human behavior provides motivation for autonomous driving that is more human-like, but also safer than human performance.

For example, in the incident where the autonomous Apple vehicle was rear-ended [20], the vehicle was likely programmed to take a safe and conservative approach to uncertainty. Human drivers do not expect to encounter vehicles stopped at the end of an on-ramp, so this behavior results in a less safe situation. Over-cautiousness of autonomous vehicles is a concern that dates back several years [24]. Data presented in 2015 by UMTRI [25] shows that driverless vehicles are in more accidents than regular vehicles, even though they are not at fault. The same authors from UMTRI conclude that an expectation of zero fatalities is unrealistic for self-driving

vehicles. They say, ‘It is not a foregone conclusion that a self-driving vehicle would ever perform more safely than an experienced, middle-aged driver.’ [26]

In June of 2022, NHTSA (National Highway Traffic Safety Administration) began publishing reports on automated driving systems (ADS) and advanced driver assistance system (ADAS) crashes [27], [28]. The data in the ADS report shows that most crashes are rear-end collisions from passenger cars; specifically, 56% of the collisions were with passenger cars and 59% of the time, the ADS had damage in the rear. This data suggests that the behavior of the ADS may be acting “too safe,” where the automated driving vehicle stops and waits when it encounters an unknown scenario, surprising the driver approaching from the rear. The ADAS report shows that most collisions also occur with passenger vehicles, but the damage occurs to the front of the ADAS vehicle. The rear collisions for the ADAS system were only a small portion of the damage at 8.4% of the total. This data suggests that the ADAS system may have missed detecting an object in front of it and could not stop for it.

Similarly, there have been multiple instances where Tesla self-driving vehicles have crashed into emergency vehicles stopped on the expressway [29]. Research has also shown intelligent systems that are easily fooled by malicious addition of noise into images [30], small markings on roads that cause dangerous lane misdetection [31], and simple modifications to road signs that change the entire meaning of the sign to intelligent systems. However, these issues are easy for humans to handle [32].

Waymo vehicles have been sabotaged or attacked by assailants wanting retribution for what they see as unsafe behavior from Waymo vehicles [33]. An example of this behavior is a close miss by a Waymo car of a boy playing in a cul-de-sac. Other instances of automated vehicles driving slower or too cautiously have led to the high frustration of drivers affected by it.

The same article quotes other drivers who intentionally tried to run Waymo vehicles off the road as retribution for what they see as unsafe acts performed by the Waymo vehicles.

Examples show that these systems can fail unexpectedly and with great severity. Drivers tend to have a false sense of security when using these systems and can fail to recognize the severity that these failures can lead to. For example, the first human death by an autonomous vehicle happened when an automated vehicle operated by Uber failed to stop for a human walking her bike across the road [34]. The human safety driver was inattentive when the incident happened and only realized what occurred after the collision [35]. In another deadly instance, Tesla's self-driving system did not stop for a semi-truck and trailer that crossed perpendicular to the roadway. As a result, the Tesla went under the crossing truck trailer, which sheared the roof off the vehicle, resulting in the death of the Tesla driver [36].

There is a problem of complacency versus acclimation. From a human perspective, there is an expectation that these systems will handle simple situations well because the systems are already able to handle higher complexity situations. Drivers allow themselves to become complacent with these systems because they grow accustomed to the good performance they see most of the time [37], despite clear instructions and warnings from automakers like Tesla that their systems are not fully autonomous and require constant vigilance.

What does all this mean? Ultimately, any solution implemented for automated driving must be safe. Many of the examples above clearly show unsafe behavior. These examples are often edge cases, which can be extremely difficult to train for. However, humans can probably handle all these examples with little trouble. The path to introducing safe, reliable, and robust autonomous driving solutions is unclear.

From my perspective and experience, good design and implementation requires structure. Good design starts with requirements. Then, the design is verified through testing. The testing is based on requirements. This process is often referred to as the V-model [38]. My research develops a design and a structured framework for testing based on requirements. The design is a model of the merge scene. This model is progressively built up by adding one vehicle at a time. This approach builds upon the knowledge learned each step along the way. The requirements are the variables in the road scene. The algorithm is built based on this road scene model and all its requirements. A repeatable, structured framework for testing is developed that uses the requirements of the road scene to obtain results from the simulation. Finally, the results are compared to the ideal case ground truth. All of this helps make automated driving merging better understood and hopefully safer.

1.6 Research Phases

Three phases divide the overall body of work:

1. Fundamental study of the merge problem using a grid world game and a two-vehicle Q-learning simulation.
2. Function approximation study by applying deep reinforcement learning (DRL) to the two-vehicle approach using a deep deterministic policy gradient (DDPG) algorithm. This approach produced exceptional performance. Further detailed review of two-vehicle DRL, using same review strategy as Q-learning detailed study.
3. Expansion to a more representative merge road scene with multiple vehicles. A full merge scene study with multiple vehicles to study a real-world highway merge.

Phase 1 resulted in a paper presented at the 2019 IEEE Intelligent Transportation Systems Conference and subsequently published in its proceedings [3]. Phase 2 was published in the 2021

IEEE Intelligent Vehicles Symposium [4]. In Phase 3, the road scene progressively scales up from two vehicles to three, then four or more. Phase 3 uses the same DDPG DRL approach as Phase 2.

A set of states and actions enable the fundamental study of the merge maneuver. The first paper describes a simple yet fundamental model that leads to several propositions that formalize merge behavior. Principles from game theory help develop rudimentary building blocks leading to the grid-world model and propositions. Next, a Q-learning reinforcement-learning (RL) simulation of the two-vehicle merge scene was created. Q-learning is a tabular approach chosen because of its simplicity and transparency compared to a more obfuscated neural network approach.

The multiple-vehicle road scene extends the work done using a single-vehicle pair. It is more representative of a real-world merge scenario. It has multiple vehicles in the scene, like Figure 1. The road scene uses the same DRL DDPG, but with more vehicles. It is studied to gain further understanding of merging. The road scene adds vehicles in both the training and testing simulations. It builds upon the existing two-vehicle DDPG simulator.

1.7 Multi-Agent Approach and Contributions

There are three main contributions from my research:

1. My research presents a multi-agent approach to highway lane merging. Fundamentally, a multi-agent approach differs from traditional approaches in that multiple agents exist within a common environment. The multi-agent approach is a more natural way to model the behavior of the interacting vehicles instead of probabilistic models of uncertainty that depend on reactions to a perceived environment. Thus far, this approach has limited treatment within automated

driving and lane merging, but the popularity and related work has grown since my research started. This multi-agent approach produced a simplified model representing the core aspects and fundamental behaviors of on-ramp merging. This contribution has several aspects: the simplified two-vehicle approach, grid-world modeling, and the propositions derived from the T-grid study. It explores the basic principles fundamental to lane merging and is absent from related work.

2. The framework for performance evaluation is another main contribution. The framework uses parametric dimensions of merge ramp length versus starting distance delta of vehicle pairs. This new approach enables consistent and equivalent comparison of different methods in a standard way. Moreover, the framework can be used to compare not only the methods presented in my work, but also methods by others because standards for merging evaluation do not currently exist.
3. The third major contribution is a multi-agent DRL simulation that has learned to operate with nearly ideal collision avoidance performance. The multi-agent simulation uses the 2-agent fundamental approach as a basis. It is then scaled up to a full vehicle scene. The framework evaluates the performance throughout the progressive development. The performance from a two- to three-, and full-vehicle scenario is consistently nearly ideal.

1.8 Related Work

Automated vehicle merging is a problem that has been previously explored in the literature. It is a popular topic that continues to be worked on by many researchers and published at top conferences. My work uses a multi-agent reinforcement learning approach to the

automated lane merging problem. The multi-agent approach is rooted in game theory. Multi-agent approaches to automated driving are growing in popularity, but they are explored less than other approaches.

Tang presents a multi-agent approach to merging [39]. Their approach to the full-scene problem is the most similar to the approach presented in this dissertation. They present a DRL self-play method in a multi-agent merge scene. They create a mix of rule-based and RL-trained agents in their simulation. They use a 76-dimension vector that includes the position, velocity, acceleration, orientation, and turn-signal state of eight neighboring vehicles and input it into a CNN DRL model to generate three actions: steering, acceleration, and turn signal. In contrast, my research does not consider steering or turn signals because it focuses on the ego vehicle's longitudinal positioning to avoid collisions. In addition, their study focuses on a different merge scenario called a 'zipper merge' that is like a parallel-type lane merge, but the end of the merge lane continues into another ramp. The scenario is more like a lane change problem with a fixed distance to change lanes versus my work, which focuses on a taper-type merge ramp. They also conclude that their agents exhibit human-like behavior such as 'defensive driving, overtaking, yielding, and the use of turn signals to communicate intentions to other agents.' My research takes a more fundamental approach to learn the behavior between merge vehicles by using a two-vehicle model to understand fundamental behavior. It also provides a framework for evaluation using ego simulation starting positions and relative positions between vehicles – a concept not considered in their work.

Reinforcement learning (RL) approaches are also growing in popularity for merging scenarios. Kamran et al. [40] present a DRL approach with a trajectory planner and formulate a cooperative merging scenario as a POMDP (partially observable Markov decision process). They

use a DQN (Deep-Q Network) approach. Shalev-Shwartz et al., [41] present a DRL method that uses an option graph similar to LSTM (long short-term memory) approaches. Their option graph considers the changing environment as a multi-agent approach instead of a simplified single-agent MDP (Markov Decision Process) approach. While they consider it a multi-agent approach, it differs from the approaches used in my work. The authors are from Mobileye, a well-regarded company focused on ADAS and AD systems with cutting-edge algorithms used in the technology. Work by Wang & Chan [42] is also a DRL approach that explores a single-agent architecture that searches for safe gaps in traffic to merge into. However, my work focuses on the interaction of agents, instead of finding gaps in between them to merge into.

In a recently published work, Chandra, Rohan, and Manocha [43] present a game-theoretic multi-agent approach to intersections, roundabouts, and merging. It uses a centralized auction system that prioritizes moves based on driver aggressiveness to prevent collisions and deadlocks. This approach is dependent upon the introduction of V2x infrastructure to be implemented. Widespread or even localized adoption of V2x is not currently in place and is unlikely in the near term. The approach in my work does not rely upon V2x infrastructure and does not consider driver aggressiveness. In other game-theoretic work by Bahram et al. [44], they propose a re-planning method based on an extensive-form game. Their main contribution is the replanning approach to the changing environment. Unlike my research, they do not directly apply it to merging. Work by Jain et al., [45] also presents a game-theoretic approach to predicting the motion of other vehicles on the road. They use k-level reasoning to develop an approach called ‘Multi-Fidelity Recursive Behavior Prediction.’ They use this approach because they also see that other agents influence agents' behavior. Their work does not focus on merging

behavior. However, it is similar to my research on the game-theoretic approach and understanding that agent behavior produces reactive behavior in other agents.

John Dolan's team at Carnegie Mellon University is the most active group conducting similar research on applying artificial intelligence methods for autonomous lane merging. They have published many papers with several different approaches. Their latest paper on the merge application presents a safety framework using a Control Barrier Function for mixed human-autonomous scenarios [46]. To list a subset of their additional work, they have published work on merging with approaches in reinforcement learning [47], intention estimation [48], probabilistic graphical model (PGM) [1], and prediction and cost function approaches [49]. Their work focuses on the in-lane traffic vehicle as it approaches the merge intersection, unlike the work presented here that focuses on the ego vehicle as the one merging into traffic from the ramp. They also use Next Generation Simulation (NGSIM) datasets to train and evaluate their methods [50]. However, the work presented here makes limited use of standard datasets because of the reactive nature of merging. It is difficult to judge true performance without this reactive behavior of the multiple agents. To enable reactive behavior, my research uses a simulator that enables reactive behavior based on the agent's actions. In [2] they recognize the ‘interactive behavior’ of merging and how it needs social interaction. However, they take a different approach by using a stochastic model instead of the multi-agent approach taken in my research.

In addition to their work on merging, Dolan’s group recently published a paper using MARL [51]. My research is deep MARL too. Their application is a road scene where traffic approaches from two directions using similar lane space and must avoid collisions. This work is conceptually like the “game of chicken” presented in the T-grid work of Section 2.4.1 where agents set on a collision course with one another need to determine how to break the conflict

without implicit cooperation or information sharing. Their approach shows high rates of collision avoidance in a setting representative of a real-world scenario. While my research application differs, I ultimately see similar high performance using the MARL DRL approach.

Platooning is a special case of automated driving where groups of vehicles coordinate with one another to maintain inter-vehicle positions. Cut-in and cut-out of vehicles happen during platooning. Works presented by Rajamani, Bevly, and Shladover study this behavior and how the other vehicles react and re-form the platoon [52]–[54]. Even though platooning is not directly applicable to the merge problem of study in my work, understanding this interaction between the vehicles is an example of the reactive behavior of merging itself. Vehicles must control their acceleration, speed, and position as a reaction to the road scene to avoid a collision.

Recently, a system developed for six-player no-limit Texas hold ‘em poker produced super-human performance. This system, named Pluribus, used self-play training to develop strategies to beat some of the greatest current poker players in the world. Using these game-theoretic techniques, the system was able to develop strategies using imperfect information against up to five other players to deliver winning results [55]. In poker, each player executes their unique strategy while playing. Like poker, merging in a busy road scene has many simultaneous players executing a unique strategy. Poker is very different from automated driving, but the success of the algorithm used in this poker simulation demonstrates the ability to apply game-theoretic techniques to problems with multiple agents executing different strategies to deliver expert-level human performance.

Fernandez et. al. and J. Rios-Torres et. al. have published surveys of merge work [56], [57]. These surveys show that gap acceptance methods are widely used in merge work. Gap acceptance methods rely upon finding an appropriate gap between traffic vehicles for the merge

vehicle to enter. My research emphasizes the interaction between two or more vehicles and finding ways to longitudinally position the merge vehicle to avoid a collision in an ever-changing environment. Gap searching is a more simplistic method that does not consider the reactive nature of traffic and how it influences changes in the environment like a multi-agent approach. Several works in the survey also use V2x approaches (where the vehicles communicate with each other or infrastructure). While my research does not consider V2x, it does show that collisions can be unavoidable in certain circumstances, so a master controller in the form of V2x may ultimately be essential to avoid all collisions. This is consistent with my theoretical analysis results based on a simplistic game-theoretic extensive-form model.

Glaser, et. al. present a grid-based approach for general lane changes [58]. They use acceleration and deceleration actions to longitudinally control the vehicle's position like my research. In addition, they formulate an environment using grid spaces that their agent is allowed to move into. This use of the grid approach is similar to the work presented here, but in their work it is used to determine the best space to move into for general lane changing, a related but different application than merging.

Chapter 2 Foundational Simplified Approach Using a Grid Game

Highway on-ramp merging is a challenging maneuver. The complexity of the maneuver increases as more vehicles are involved in the merge road scene. It becomes more difficult for a human driver to accomplish a collision-free merge as the number of vehicles increases. The driver must keep track of the location, speed, and intent of all vehicles in proximity. As the scene becomes busier, automating this task faces the same challenges and increasing complexity of calculating all the parameters.

The process of understanding on-ramp merging from a very fundamental perspective starts in this chapter. A basic model that limits and constrains the merge environment to study key parameters is created to gain a fundamental understanding. The model complexity can grow once a fundamental understanding is achieved. However, the task must first be understood from a fundamental perspective before it can be robustly automated.

A summary of state-of-the-art approaches in Section 1.5 showed that there are automated driving systems operating today that can fail very badly in seemingly simple ways. The underlying reasons for these failures are not always clear, especially because the development details of these systems are not usually shared publicly. However, understanding and testing the operation of these systems from a fundamental perspective can help to identify inherent unacceptable weaknesses in their operation. A fundamental understanding is also important because the simulation presented in later chapters uses a learning algorithm. Artificial intelligence and learning algorithms can be very powerful. However, a lack of theoretical

understanding can lead to presumptions about the operation and findings, rendering the results and performance questionable or misleading [30], [59].

2.1 T-Grid Merge Game

A multi-agent grid game was developed to study and understand the fundamental behavior of the merge scenario. The grid game consists of only two agents limited in state and action. The agents represent a vehicle on a merge ramp and a vehicle in the traffic lane. The initial states of the vehicles allow either agent to reach the merge point with a single move. This could result in a collision on the first move, but the agents are also allowed smaller actions to reach the merge point in multiple steps to avoid a collision. These concepts are presented in the first research paper I published on this work [3]. The same concepts are explained in this chapter, but with different wording and sometimes from a different perspective. Sometimes, the concepts are expanded based on further work after the paper was published. For example, Section 2.5 on the probability of a collision is new.

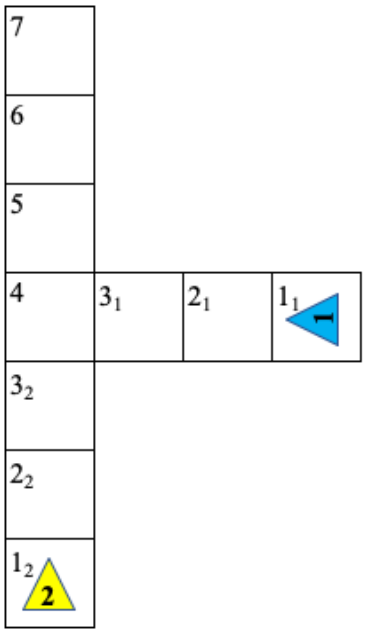


Figure 7: T-Grid model. Each block represents a grid space that the vehicle can move into. The blue triangle is Agent 1 and represents the ego merge vehicle. The yellow triangle is Agent 2, the traffic vehicle.

Grid-world games are a common approach to simplify state-action games [60]. Figure 7 shows the grid game. The merge vehicle and the traffic vehicle are agents. Agent 1 is the merge vehicle (ego), and Agent 2 is the in-lane traffic vehicle. Each agent can take a set of actions, $\mathcal{A} = \{1, 2, 3\}$, where each action choice represents the number of grid blocks advanced. Any action chosen represents forward movement of the agent. Each step requires forward movement and a non-movement action is not allowed because it represents stopping in the roadway. Stopping is a dangerous behavior on freeways or on-ramps[2].

Three action choices are allowed:

1. An action choice of one is a single step and represents a deceleration in speed.
2. A two-grid step action represents maintaining speed.
3. An action choice of three blocks represents acceleration.

Each agent can take any of these actions at any grid step. The agents take actions until the ego agent reaches or passes the merge point at block 4 or until they collide. Collisions occur when both vehicles occupy the same block. The T-grid game represents the critical essence of simplified actions an automated driving vehicle must take to merge into traffic. It enables the fundamental study of merging.

2.2 T-Grid Game Reward Structure

Figure 8 shows the reward structure for the actions that Agent 1 can take in the T-grid model. The reward strategy penalizes any acceleration or deceleration action but withholds penalty for maintaining velocity. Acceleration and deceleration are penalized for two reasons: a

[2] Freeways normally have a minimum speed because of the high-speed vehicles travel on them. A vehicle travelling at highway speeds that encounters a stopped object or vehicle would have to stop very abruptly and likely not in time to avoid a collision. Therefore, stopping on an expressway is dangerous. Section 0 describes an unsafe situation where a vehicle is stopped at the end of an on-ramp, leading to a rear-end collision with it.

change in speed decreases driver comfort and results in energy loss[3]. On the other hand, maintaining velocity is not penalized because it does not affect driver comfort and maintains steady-state instead without accelerating or decelerating.

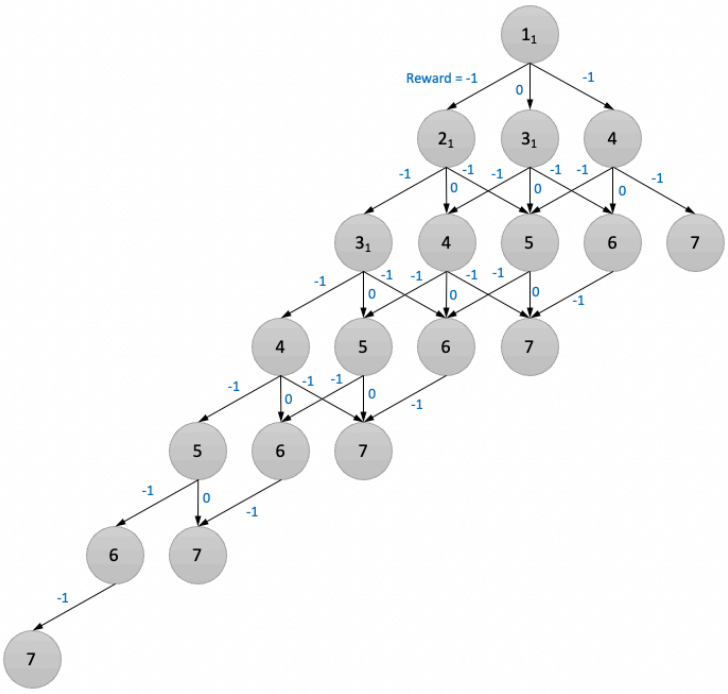


Figure 8: T-Grid reward tree for ego actions. Each circle represents the grid block that the agent can move to. Next to the arrow is the immediate reward for each move.

Actions taken at each step sum to a cumulative reward. The branches of the reward tree show the penalty for each action. An additional penalty is applied if a collision occurs. When two agents occupy the same grid block, a collision occurs and the agent at fault is assigned a penalty of 10,000. The other agent is assigned a penalty of 1,000. A reward of 100 is given to Agent 1 if it reaches position 4 or greater and does not collide with Agent 2. Ultimately, the basic intent is to penalize acceleration and deceleration minimally but heavily penalize collisions.

[3] Energy loss occurs when accelerating or decelerating. When decelerating, brakes are applied to slow the vehicle, converting momentum into waste heat energy from braking friction. When accelerating, stored energy is converted into kinetic energy and the process is not fully efficient. See common thermodynamic textbooks for details [88].

2.2.1 Game Theory Formal Example of Rewards as Payoffs

Game Theory more formally explains the T-grid game. Figure 9 shows a normal-form representation for a subset of T-grid game moves. This example illustrates rewards and penalties received for different sets of moves.

		Agent 2	
		$1_2 \rightarrow 2_2 \rightarrow 4$	$1_2 \rightarrow 3_2 \rightarrow 5$
Agent 1	$1_1 \rightarrow 3_1 \rightarrow 6$	$99, -1$	$99, 0$
	$1_1 \rightarrow 3_1 \rightarrow 5$	$100, -1$	$-10000, -1000$

Figure 9: Normal-form bi-matrix T-grid game representation rewards example.

In the example, Agent 1 can make two different merge action combinations. In the top set, $1_1 \rightarrow 3_1 \rightarrow 6$, Agent 1 starts from its initial position, then takes an action to maintain its velocity by moving two spaces into 3_1, incurring no penalty. Agent 1 then chooses an acceleration action to end the episode in block 6 and incurs a penalty for the acceleration. The result of its actions is the same regardless of the moves that Agent 2 takes. No collision occurs between the agents. The total payoff is 99, comprised of 100 for a successful merge and a penalty of 1 for the acceleration move.

In the bottom action combination, $1_1 \rightarrow 3_1 \rightarrow 5$, Agent 1 maintains its speed throughout the episode. It starts in block 1, then moves to block 3_1, then to block 5. There are two different payoffs, depending on whether there is a collision with Agent 2. The payoffs are 100 for a successful merge without collision or a penalty of 10000 for a collision. For both payoffs, there is no penalty for acceleration or deceleration.

Agent 2 also has two different action combinations in the normal-form game. One combination is $1_2 \rightarrow 2_2 \rightarrow 4$. Agent 2 starts in block 1, then takes a deceleration action to move to block 2, incurring a penalty of 1 for deceleration. Agent 2 then takes an action to maintain

speed, moving to block 4, finishing the episode. The total payoff is -1 for the penalty incurred by the deceleration action.

The other action combination that Agent 2 makes is $1_2 \rightarrow 3_2 \rightarrow 5$. This combination is the same set of moves Agent 1 makes in the bottom row of Figure 9. Agent 2 uses an action strategy of maintaining velocity throughout the episode, thus no penalty is incurred for deceleration or acceleration actions. However, it does incur a penalty when Agent 2 selects a strategy that places it in block 5 at the same time as Agent 1, thus causing a collision and incurring a penalty of 1000. In this example, both agents maintain their velocity and avoid the minor penalty, but the strategy results in a severe penalty from the collision.

The utilities of the payoff matrix shown in Figure 9 are calculated using Game Theory techniques. Let p be the probability of Agent 1 selecting $1_1 \rightarrow 3_1 \rightarrow 6$ and let q be the probability that Agent 2 selects $1_2 \rightarrow 2_2 \rightarrow 4$. The utilities are set equal to each other to find the Nash equilibrium probabilities[4].

$$U_{1_2 \rightarrow 2_2 \rightarrow 4} = U_{1_2 \rightarrow 3_2 \rightarrow 5}$$

$$-1(p) + (-1)(1-p) = 0(p) + (-1000)(1-p)$$

$$-1 = -1000 + 1000p$$

$$p = \frac{999}{1000} = 0.999$$

Equation 1

$$U_{1_1 \rightarrow 3_1 \rightarrow 6} = U_{1_1 \rightarrow 3_1 \rightarrow 5}$$

$$99(q) + 99(1-q) = 100(q) + (-10000)(1-q)$$

[4] Nash equilibrium is defined as "a *stable* strategy profile: no agent would want to change [its] strategy if [it] knew what strategies the other agents were following. [89]" Details on Nash equilibrium and other Game Theory principles can be found in standard formal introduction reference material [90].

$$99 = 10100(q) - 10000$$

$$q = \frac{10099}{10100} \approx 0.9999$$

Equation 2

The results show that Agent 1 selects $1_1 \rightarrow 3_1 \rightarrow 6$ with a 99.9% probability and Agent 2 selects $1_2 \rightarrow 2_2 \rightarrow 4$ with a probability of 99.99%. It shows that, in this particular game instance, the agents would select the maintain strategy ($1_1 \rightarrow 3_1 \rightarrow 5, 1_2 \rightarrow 3_2 \rightarrow 5$) with extremely low probability. Agent 1 would take the maintain path only 0.1% of the time and Agent 2, 0.01% of the time. The maintain path could lead to a collision which is the reason for the low probabilities. If both agents chose the maintain strategy, it would only lead to a collision. However, based on the severity of the penalty in the payoff matrix, it drives the probability of taking that path very low.

The example shown is a small subset of the entire set of moves possible in the T-grid game. It shows differences in payoffs depending on the set of moves of each player. Different strategies can result in successful merges or collisions. It also shows that taking actions without considering the other player's actions, can result in a collision. The example illustrates the interdependent nature of merging and the reactiveness essential to avoiding collisions.

2.3 T-Grid Evaluation: Turn-Taking

Imagine a scenario where an agent can predict the other agent's behavior before it happens. Considering this scenario in terms of the T-grid approach would mean that an agent would accurately predict the other agent's action before it happens. If this were possible, the T-grid scenario would be a turn-taking game. For example, tic-tac-toe and checkers are common turn-taking games.

Using a T-grid turn-taking approach, each player can observe the position and action of the other player before taking their turn. Each player can choose multiple actions, enabling them always to choose a different block than the other player occupies. Figure 10 (left) shows an example where Agent 1 is in grid block 3 and moves to grid block 4. Once Agent 1's turn is complete, Agent 2 can move to blocks 5 or 6 to avoid the collision. This notion of predictability enables a collision-free game.

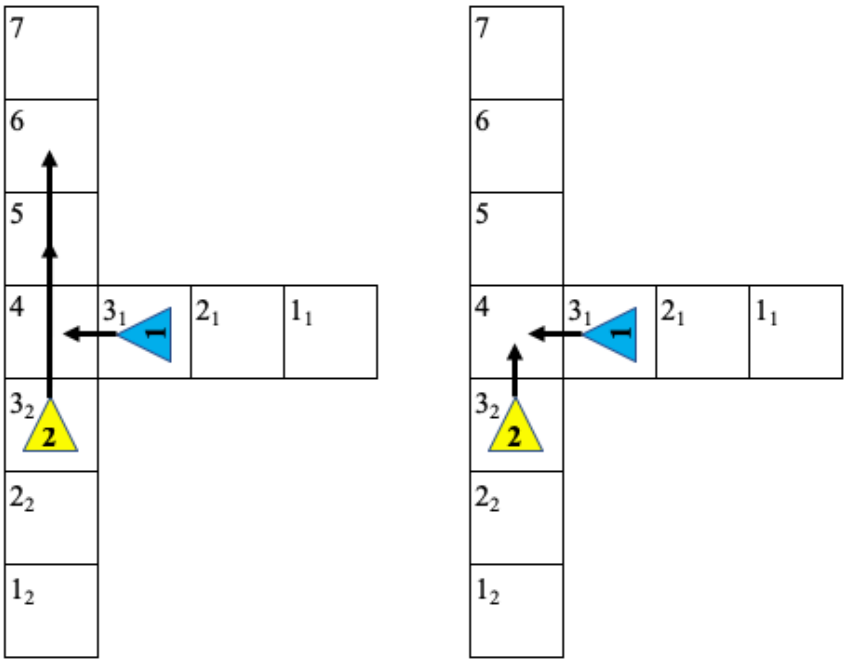


Figure 10: T-Grid Turn Taking (left), Simultaneous Play (right)

The turn-taking conditions that allow a collision-free game are generally unrealistic. Vehicles cannot pause their motion in time to observe the action of the other vehicle. However, it is possible to envision a somewhat representative example. A vehicle (or the vehicle's driver) can observe the position of another vehicle. If the action taken by the other vehicle was observed and remained repeatedly constant, the behavior during that period can be described as turn-taking. The concept of turn-taking is the first in a series of propositions from my first research paper on merging:

Proposition I: A collision is avoidable if a vehicle can observe the position of another vehicle, can predict the action it will take, and has sufficient action choices. Or said more simply: "Turn-taking leads to collision-free behavior." [3]

2.4 T-Grid Evaluation: Simultaneous Play

Simultaneous play is another approach to evaluating the T-grid game. With simultaneous play, both agents take a turn simultaneously. Evaluating this way is analogous to a real-world merge scenario because both vehicles would make real-time action choices. In simultaneous play, each agent chooses an action of one, two, or three grid blocks at each game step. By design, either agent can reach the merge position in one step. Specifically, if an agent chooses an acceleration action from its starting position, it reaches the merge point in grid block 4.

The initial condition of the game ensures a non-zero probability of collision because each player does not know the action of the other. If both agents choose an acceleration action to start, they will simultaneously occupy block 4 and collide. Many other state-action combinations result in a collision too. Collisions are unavoidable because the agents are in close proximity, neither knows the action of the other, and both are making an action choice simultaneously. Extending this idea to a real-world merge scenario is intuitive. Summarizing this behavior results in another fundamental proposition of merging:

Proposition II: A collision is unavoidable if two vehicles are in close proximity and neither knows the other's action.

In essence, this is what makes merging difficult for drivers. Drivers must actively work to maintain a reasonable separation from other vehicles to help avoid collisions. Drivers also use judgment to predict what another driver will do during merging to avoid a collision. Conceptually, this judgment of predicting the action of another driver is turn-taking. It is rational

to believe that drivers are actively deciding to take actions that will avoid collisions. Driver experience helps them to predict that the other vehicle will decelerate, maintain speed, or accelerate to avoid a collision during merging.

The difference in proximity between vehicles is also important. The greater the differential between two agents and the merge point, the lesser the chance of a collision. Relating it to a grid game example in Figure 10 (left), the agents will collide if both start in block 3 and choose the same action. The probability of a collision is 1/3 since the total combinations of moves is $3 \times 3 = 9$ and three of those combinations result in a collision. Doing the math, 3/9 collisions to combinations, and the fraction reduces to 1/3 or 33%. Increasing the differential by just one grid space reduces the collision probability to 2/9 or 22%. For example, if Agent 2 starts once one space back in block 2, but Agent 1 remains in block 3, the collision probability is 22%. The more the differential, the lower the probability.

Additionally, collision probability decreases or becomes zero if the action choices are limited based on differential position. For example, when Agent 2 is in block 2 and Agent 1 is in block 3, Agent 2 can limit its action choice to a move of just one block to avoid a collision entirely. After the simultaneous actions, Agent 1 would be in block 3, 4, or 5 and Agent 2 would only be in block 3, thus avoiding a collision entirely. Similar position-action combination examples exist, leading to reduced collision probability or avoidance altogether. Summarizing results in another proposition:

Proposition III: To reduce the risk of collision, it is advantageous for a leading vehicle to accelerate away from a lagging vehicle or for the lagging vehicle to decelerate.

2.4.1 Game Theory Formal Example: The Game of Chicken

Game Theory techniques formalize these T-grid concepts. Reiterating from the first work I published [3], a normal-form representation in Figure 11 shows the interaction of two vehicles. The T-grid game is abstracted to a "2 x 2" (2-player, 2-action) game. The two agents (players) can take an action to maintain or to change speed, represented as o for maintain and Δ for change. This abstraction simplifies acceleration and deceleration to a generic speed change.

In this version of the game, the payoff bounds are between zero and one. Zero represents no payoff (no reward), meaning a collision has resulted. A value of one represents a successful merge. A collision is at one extreme of the range (represented by zero) and a successful merge is at the other extreme (represented by one).

		Agent 2	
		Δ	o
Agent 1	Δ	α, β	α, 1
	o	1, β	0, 0

Figure 11: Two-vehicle simultaneous play merge represented as a normal form game; the "game of chicken"

The "game of chicken" described in Figure 11 is analogous to the two-move instance of the T-grid game shown in Figure 9. In the two-instance move, each agent can take a strategy to maintain or change speed. The same is true of the game of chicken. The values of α and β represent the style of the driver, or the amount of risk that the driver is willing to take. In other words, α and β represent the level of aggressiveness or defensiveness of the driver. The value is greater than 0, but less than 1. They are intentionally different because the expectation is that each driver (or automated driving controller) would have a unique driving style behavior. The highest payoff is when both agents maintain their speed (represented by o), but maintaining

speed also comes with the risk of a collision. The risk carries an exceptionally low payoff compared to the payoff of choosing a strategy to change speed (represented by Δ).

If both agents had the same behavior, each would choose the same strategy when playing the game. In this scenario, both agents would possess precisely the same amount of aggressiveness or exactly the same amount of defensiveness. Therefore, a collision would be highly likely if both agents started at the same position and played with the same strategy. For example, if they were both aggressive, each one may try to accelerate aggressively to beat the other to the merge position using the same behavior. Likewise, suppose there were two defensive drivers. In that case, both might yield to each other exactly the same way, forcing a similar stalemate where each would stay aligned longitudinally until a collision occurred at the merge point. This behavior would put the agents on a collision course towards each other where they would both accelerate or decelerate at precisely the same moment, a very dangerous situation. This stalemate interaction leads to the last of four propositions:

Proposition IV: The interaction of two vehicles with the same driving style is more likely to produce a collision than that of vehicles with different driving styles.

2.5 The Probability of a Collision

The probability of a collision is estimated by equating the variables in the payoff matrix of Figure 11 to the payoff matrix values from Figure 9. The symbols and probabilities from the two figures correlate by observation. The o symbol for Agent 2 in Figure 11 correlates $1_2 \rightarrow 3_2 \rightarrow 5$ from Figure 9 and carries a probability of 0.1% as calculated in Section 2.2.1. Similarly, when Agent 1 selects the maintain strategy, indicated by o, its probability is 0.01%. Using these two percentages (converted to exponential values), the probability of a collision calculates to $10^{-3} \times 10^{-4} = 10^{-7}$.

The probabilities for the T-Grid approach correlate with statistical data published by the U.S. Government. Several NHTSA crash statistics are used to calculate the probability [6]. Table 1 shows the statistics.

Table 1: 2019 NHTSA crash data, selected values

Total police-reported motor vehicle crashes	6,756,000
Fatal crashes	33,244
Fatalities per 100 million vehicle miles traveled	1.11
Vehicle miles traveled	3,261,772,000,000

The report shows that a total of 6,756,000 crashes occurred in 2019. Of the overall crashes, 33,244 were fatal. The ratio of these values shows that crashes occur about 203.225 times more than fatal crashes. Fatalities occur 1.11 times per 100 million vehicle miles traveled. Using the ratio of total crashes to fatal crashes yields that collisions occur 225.58 times per 100 million vehicle miles traveled or 2.2558×10^{-6} collisions per vehicle mile traveled. Taking the reciprocal of this value shows that 443,303 miles are traveled in between crashes. There are a total of $3{,}261{,}772 \times 10^{6}$ vehicle miles traveled. The 443,303 miles traveled between crashes is divided by the total vehicle miles traveled.

This results in the probability of a collision while traveling in a vehicle of 1.36×10^{-7}. Comparing this probability to the calculated probability of a collision for the T-grid merge model of 10^{-7}, shows that these numbers are the same magnitude and nearly the same value. The similarity in these numbers suggests that the T-grid merge model and its rewards strategy correlate well with statistical data. Summarizing in simple terms, shows that drivers rarely collide because the risk is so high.

2.6 Summary

The merging modeling approach suggests that drivers take an active strategy to avoid a collision when close to another vehicle. The grid game study led to a better understanding of the merge scene and enabled the development of four key propositions describing the expected behavior of merging. This work resulted in a conference paper presented at IEEE ITSC 2019 [3]. Summarizing the propositions:

I. A collision is avoidable if a vehicle can observe the position of another vehicle, can predict the action it will take, and has sufficient action choices.

II. A collision is unavoidable if two vehicles are in close proximity and neither knows the other's action.

III. To reduce the risk of collision, it is advantageous for a leading vehicle to accelerate away from a lagging vehicle or for the lagging vehicle to decelerate.

IV. The interaction of two vehicles with the same driving style is more likely to produce a collision than that of vehicles with different driving styles.

Chapter 3 Reinforcement Learning Simulation Using Q-Learning

This chapter presents a Q-learning simulator that represents a two-vehicle on-ramp merge. A multi-agent reinforcement learning (MARL) [61] simulator models the actions and motion of an introductory two-vehicle merge scene. The simulator builds upon the foundational results and propositions from the study of the T-grid model in Chapter 2. The simulator includes both a single-agent and multi-agent Q-learning version. The simulator motion formulation, training, and testing functionality are explained in detail. A novel framework for evaluating performance shows the results. The results correlate to the propositions developed from studying the T-grid model. The section concludes with a discussion of the results and findings within the data that motivate future direction toward a function approximation method in continuous space[5].

Q-learning reinforcement learning (RL) is a state-action tabular approach [62], [63]. The action used here is vehicle acceleration, the same as the simplified T-grid approach. Compared to the T-grid, there are more Q-learning states to better represent the parameters involved with merging. Careful consideration is given to the states to be minimalistic yet capture all necessary criteria to model interactive merging behavior. Figure 2 (Section 1.3) shows the merge scenario: two vehicles approaching a merge point along a taper-type merge ramp. The simple yet representative model has increased complexity from the T-grid. It is the next step to continue the fundamental study of the essence of merging.

[5] The Q-learning work in this chapter was published along with the T-grid work in IEEE International Conference on Intelligent Transportation Systems (ITSC) 2019 [3]. The detail presented in this chapter expands upon the conference paper.

The study model so far has been limited to a single vehicle pair to gain fundamental understanding at a more simplified level. The model presumes that the ego merge vehicle primarily focuses on reacting to one most-important-object (MIO): the traffic vehicle. The closest in-lane vehicle will likely have a significant enough gap in front of it and behind it to allow a merge to occur. A study of drivers' following distances in [64] shows that the gap between vehicles on the highway is typically greater than several car lengths, so a two-vehicle model could be a relatively accurate representation.

The Q-learning longitudinal motion model uses standard equations of motion programmed using MATLAB. The MATLAB simulator trains the Q-learning tables through self-play of the agents. The simulator runs many episodes of the two agents, recording the Q-value for each state-action set. The reactive nature of merging makes self-play simulation an important aspect of this multiagent study because training in a real-world environment would be unrealistic. The simulator is also important because it allows collisions to happen in a safe space instead of the real world. Both single-agent and multi-agent learners are trained and tested. The results from the testing are arranged in a results table to compare performance between variations of the simulation parameters in a standard format. The results table is a unique contribution to the study of merge behavior that other researchers can use.

3.1 Multi-Agent Q-Learning

Q-learning is a reinforcement learning technique where a table of Q-values is populated through training and used to determine the best action to take at each state [62], [63]. It is a simple but well-known reinforcement learning technique. Q-learning uses a state-action environment where each combination seeks to optimize the expected reward value through reinforcement learning. During training, the simulator visits a state-action pair multiple times to

learn the expected Q-value. Q-learning allows short-term and long-term discounting of the rewards through tuning parameters. Equation 3 shows a single-agent Q-learning formula where s_t is the current state, a_t is the current action, $\alpha_t \in [0,1)$ is the learning rate, r_t is the immediate reward, and $\gamma \in [0,1)$ is the discount factor. Subscripts of t indicate the current state, and $t+1$ indicates the next state.

$$Q_{t+1}(s_t, a_t) = (1 - \alpha_t)Q_t(s_t, a_t) + \alpha_t \left[r_t + \gamma \max_a Q_t\,(s_{t+1}, a)\right]$$

Equation 3: Q-learning update rule.

Multi-agent Q-learning is an extension of single-agent Q-learning. When Q-learning extends to a multi-agent scenario, multiple actions are considered instead of a single action [65]. The formula is quite similar, but the action set includes all n agents, represented by superscript $i = 1, 2, \dots, n$.

$$Q_{t+1}(s_t, {a_t}^1, \dots, {a_t}^n) = (1 - \alpha_t)Q_t(s_t, {a_t}^1, \dots {a_t}^n) + \alpha_t \left[r_t + \gamma \max_a Q_t\,(s_{t+1}, a^1, \dots, a^n)\right]$$

Equation 4: Multi-agent Q-learning update rule.

Fundamentally, with multi-agent Q-learning, the environment is no longer stationary but has other agents that also act. As a result, the state complexity increases, but prediction accuracy also increases, as shown by Hu and Wellman, Littman, and Claus and Boutilier [66]–[68]. The multi-agent perspective is different because it considers multiple agents acting within an environment instead of focusing on the agent itself within its environment. Weiss [69] explains multi-agent as, '… agents have far-reaching control over their behavior within the frame of the objectives, possess decision authority in a wide variety of circumstances, and are able to handle complex and unforeseen situation on their own without the intervention of humans or other systems.'

3.1.1 Single-Action and Joint-Action Learners

The term single-action-learner (SAL) or independent learner (IL) are often used interchangeably with single-agent throughout my work. Similarly, joint-action learner (JAL) is used interchangeably with multi-agent. The terms IL and JAL come from work on multiagent systems by Claus and Boutilier [68]. The Q-learning formulas in this section specifically describe *single-agent Q-learning* and *multi-agent Q-learning*. The difference between the formulas is the addition of the action variable of the other agent when describing multi-agent. However, a more traditional view of multi-agent systems is not necessarily the simple inclusion of the additional action information of the other agent but that multiple independent agents are acting within a common environment [61]. The Q-learning scenarios studied here include both aspects of multi-agent: inclusion of the additional action and multiple independent agents in a common environment.

3.2 States and Actions

The Q-learning approach uses a state variable that consists of four states (or five for the JAL), with formulas in Sections 3.3.4 and 3.3.5. The set of state variables is defined as follows:

1. *Closing gap,* $l_{closing}$: longitudinal space between the merge and in-lane traffic vehicles.
2. *Closing speed,* $v_{closing}$: difference in longitudinal velocities between the two vehicles.
3. *Time to position, TTP*: time remaining for the merge vehicle to reach the merge intersection point.
4. *Position of ego,* $P_{relative}$: relative position of ego with respect to traffic vehicle.[6]
5. *Traffic action,* $a_{traffic}$: the instantaneous action of the traffic vehicle (if JAL).

[6] The gap variable alone is not sufficient to describe the positions of the two vehicles with respect to each other because vehicle lengths may vary.

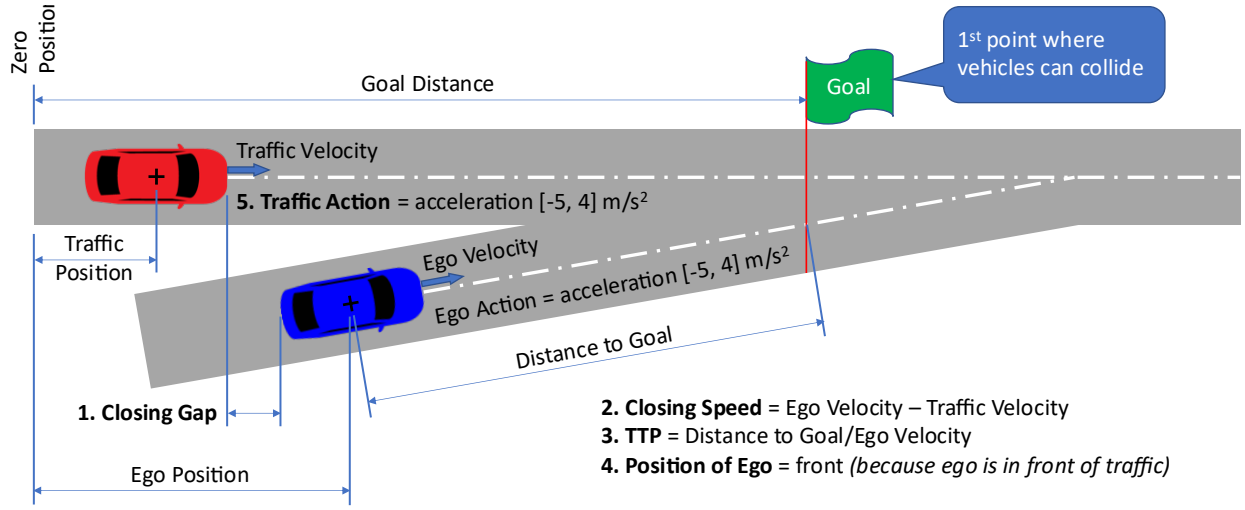


Figure 12: Two-vehicle Q-learning simulation road scene. State variables are numbered 1-5: Closing Gap, Closing Speed, TTP, Position of Ego, and Traffic Action.

The set of state variables considers only necessary and essential information to minimize the number of states. For example, early work considered only the closing speed and closing gap to describe the state. In this scenario, the ego vehicle could learn how to position itself to avoid a collision, but it did not consider the diminishing time to perform the merge. Thus, the state set gains a variable for TTP (time-to-position: the time to reach the merge position). TTP encodes the time into the state, enabling the algorithm to learn appropriate actions based on how much time is left to complete the merge.

Relative states are used instead of absolute values for the gap and speed to help with learning. Relative states decrease computational complexity by reducing the range of parameter values. The intent is to keep the size of the Q-learning matrix smaller and decrease the overall learning duration. The rationale is that the ego vehicle can learn what acceleration value to choose based primarily on the closing speed and closing gap, independent of its absolute speed. A drawback is that vehicles have different acceleration performance depending on the vehicle speed, engine RPM, road grade, or other parameters. However, the general expectation is that the vehicles will generally travel at highway speeds, so the absolute velocity might be irrelevant. The

step update time of 100 ms also helps to frequently re-adjust based on the latest state parameter values. Ultimately, real-world testing may be needed to determine if this is truly a factor for performance.

A relative position state defines if the opposite lane vehicle is in front or behind. The gap variable itself is insufficient to describe the positions of the two vehicles with respect to each other. As the vehicles overlap, the closing gap variable alone would become negative. However, a negative value would not sufficiently describe whether one vehicle was in front of or behind the other. Thus a relative position indicator is necessary.

In the *multi-agent* variant of Q-learning, the Q-learning formula includes the other agent's action. Fundamentally, the state treats the action as another state variable. The multi-agent (joint-action-learner) variant trains in parallel to the single-agent (single-action learner) with a duplicate set of Q-learning tables. It is possible to train both in parallel because Q-learning is an off-policy method where the learner only uses the trained Q-values once training is complete.

The actions considered in the Q-learning algorithm are the accelerations of the ego vehicle and the traffic vehicle. They are a discrete set of actions. The set of actions for the ego merge or traffic vehicle is: {high deceleration, deceleration, maintain, acceleration, high acceleration}. The values for the set of actions are {-5, -3, 0, 2, 4} with m/s^2 units. The action set is limited to cover a reasonable range of typical values for the acceleration and deceleration of a vehicle.

3.2.1 Reasonable Acceleration Values

The range of acceleration values chosen for the simulation is rooted in federal standards and typical vehicle performance. The U.S. government regulates minimum braking deceleration values through standards like FMVSS 135 [70] that require a maximum stopping distance of 70

m from a speed of 100 km/h. The stopping distance from this speed calculates to an average deceleration of 5.5 m/s^2. Vehicles can have heavier braking than this, but greater values would likely be considered very aggressive, disruptive to driver comfort, and abnormal for typical driving behavior. For example, the force of gravity is 1 G, or 9.81 m/s^2. Imagine starting at rest and then instantaneously falling to the ground. A 9.81 m/s^2 acceleration would feel like instantaneously *falling* forward. A 5 m/s^2 acceleration would be roughly half of this force. 5 m/s^2 is the max deceleration value used for braking throughout the simulations.

Vehicle performance dictates its maximum acceleration. Many current model vehicles have maximum accelerations of around 4 m/s^2 [71]. Higher accelerations usually are only achieved by high-performance vehicles. For example, a Ferrari or similar supercar might have a 0-60 MPH time of 3.5 seconds or an average acceleration of 7.7 m/s^2, but this is not common. Therefore, the max acceleration for the simulations is set at 4 m/s^2.

3.3 Q-Learning Merge Simulator

The simulator is written using MATLAB. It is a self-play simulation that uses the environment parameters and standard equations of uniform motion to train the Q-learning table. Q-learning is an off-policy method where training to generate the Q-table is performed offline [60]. The testing simulator then uses the trained table to run through a series of test scenarios to generate data for evaluation and comparison. Training and testing use two different MATLAB scripts, as shown in Figure 13. The following sections describe them in detail.

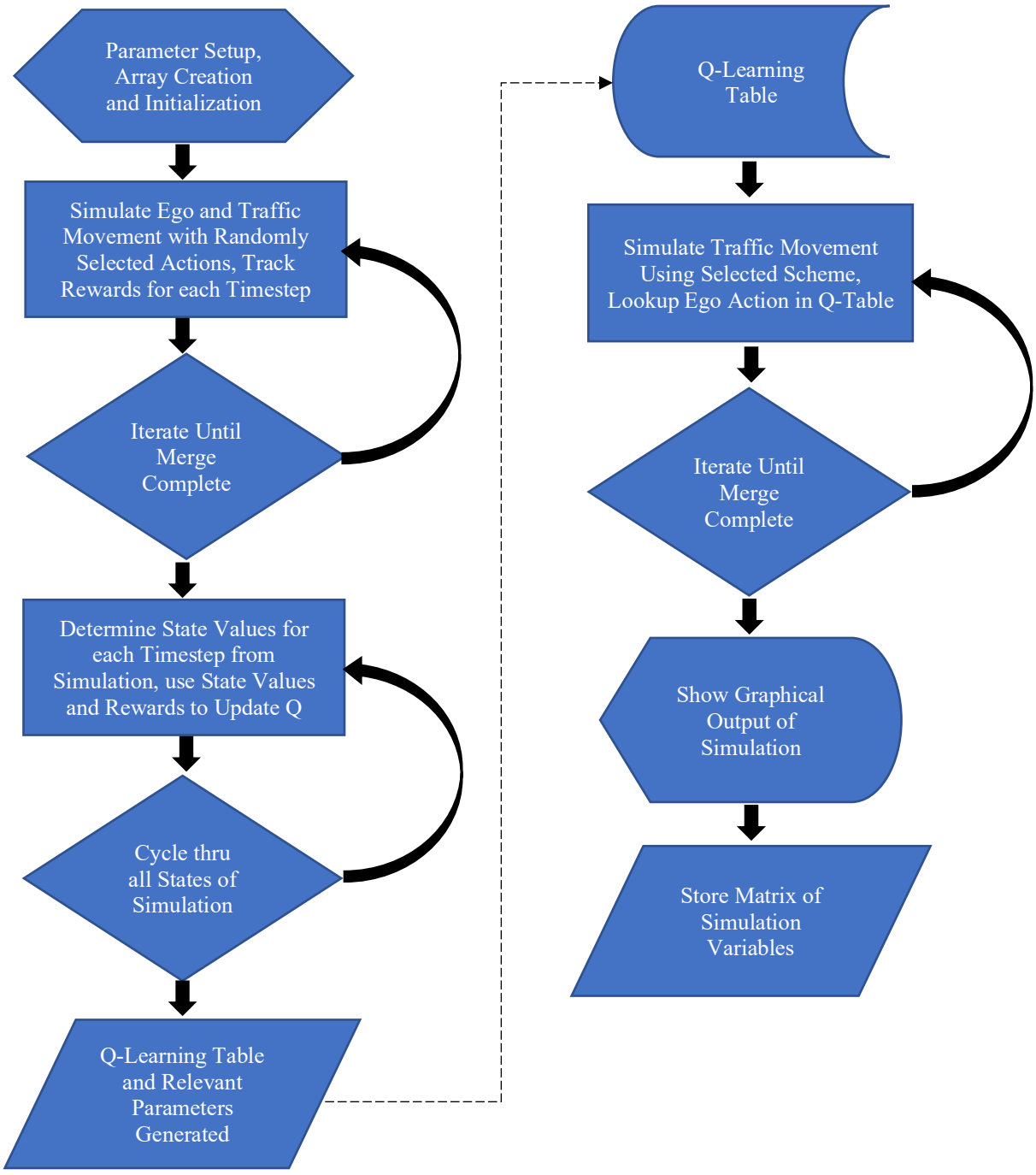


Figure 13: Q-learning simulator block diagram. On the left is the training flow and on the right is the testing flow. Training and testing use two separate scripts.

Table 2: Q-learning training simulator parameters and values. Training runs use multiple sets of parameters for the gap, speed, and TTP states. The table shows all the parameter sets used during training.

Training Simulator Parameter	**Value**
Ego Vehicle Length (m)	5
Q-Learning Discounting Factor, β	0.99
Q-Learning Rate, α	0.3
Training Episodes	10^5, 10^6, 10^7
Training Strategy, Ego	Random
Training Strategy, Traffic	Random
Time step (s)	0.1
Minimum speed (m/s)	20
Maximum speed (m/s)	40
Traffic Vehicle Length (m)	[1, 20]
Training Ramp Length (m)	[25, 150]
Merge Actions (m/s^2)	{-5, -3, 0, 2, 4}
Traffic Actions (m/s^2)	{-5, -3, 0, 2, 4}
State: Closing Gaps (m)	{-2.5, -1, 0, 5, 15, 30}
	{-5, -2, -1, 0, 1, 2, 3, 4, 5, 10, 20}
	{-5, -4, -3, -2, -1, 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, 25}
State: Closing Speeds (m/s)	{-1, 0, 1, 5, 10}
	{-5, -2.5, 0, 2.5, 5, 7.5, 10}
	{-10, -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
State: Ramp End TTP (s)	{0, 0.25, 0.5, 1, 2}
	{0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3}
	{0.5, 0.75, 1, 1.5}
	{1, 1.5, 2}
	{1}
	{1, 2}
	{1, 2, 3}

3.3.1 Training Simulator

The training simulator first defines the necessary parameters and initializes the values in Q-tables to zero. Table 2 shows the parameters used and their values or ranges. The table shows range limits for variables initialized with random numbers. At episode initialization, the

simulator generates random starting positions and velocities for the ego and traffic vehicles. The ramp length and the length of the traffic vehicle are also set randomly within the pre-defined range.

After initialization, the simulator runs in discrete time steps of 100 milliseconds. During each discrete time step, the ego and traffic vehicles select a random acceleration action from the set of discrete action values. Then, standard equations of motion for uniform acceleration calculate the longitudinal position of the vehicles using the randomly selected acceleration. The basic position equation is:

$$x = \frac{1}{2}at^2 + vt + x_{initial}$$

Equation 5

where x is the position in meters, v is the velocity in meters per second (m/s), a is the acceleration in m/s^2, and t is the time in seconds. The algorithm calculates a new velocity and position at each time step based on the chosen acceleration value and the state's current values. The following formulas calculate the new velocity and position:

$$v_{new} = a\mathrm{t} + v_{current}$$

Equation 6

$$x_{new} = \frac{1}{2}at^2 + v_{current}t + x_{current}$$

Equation 7

where v_{new} is the new velocity in m/s, a is the chosen acceleration in m/s^2, t is the time in seconds, $v_{current}$ is the current velocity in m/s, x_{new} is the new position in meters, and $x_{current}$ is the current position in meters.

Once the calculations are complete, the velocity is truncated to the min or max if it is beyond the limits. The states are calculated, then a reward is assigned for any non-zero acceleration. If the ego vehicle's position has passed the merge point, a reward is assigned based

on the reward formula for either a successful merge or a collision. Section 3.3.2 defines the reward function. Next, the Q-learning values for the state are updated. The step function loop iterates if the ego vehicle has not passed the merge point. If it has passed the merge point, the initial states reset, and another episode starts. The episodes repeat until completion.

3.3.2 Reward Function Design

The time-step function continues choosing actions and updating the state values until the position of the ego vehicle has reached the goal position. The acceleration value magnitude is the penalty assigned at each iteration of the step function. This penalty intends to help reduce unnecessary changes in the steady-state velocity of the vehicle. The magnitude of this penalty is much smaller than the magnitude of the reward for a successful merge. The acceleration penalty is also much, much smaller than the magnitude of the penalty for a collision. The reward value for each state trains the network. The Q-learning table or network (for the DRL approach) is trained at each time-step using the state values and reward. The following equation shows the rewards:

$$r = \begin{cases} 1{,}000, & \text{if merge, no collision,} \\ -1{,}000{,}000, & \text{if at-fault collision,} \\ -100{,}000, & \text{if no-fault collision,} \\ -|accel|, & \text{if any acceleration action,} \\ -|decel|, & \text{if any deceleration action} \end{cases}$$

Equation 8

The reward values used in this simulation differ from those used in the T-Grid scenario. However, the general approach to design it is the same: minor penalties for acceleration and deceleration, a larger reward for a successful merge, and a much larger penalty magnitude for collisions. The more severe penalties intend to evoke a stronger learned response by the algorithm because the expectation is that automated driving should be safer than human

performance and as close to zero collisions as possible. Future work could trial more reward function values.

The reward structure for the reactive traffic vehicle is identical to ego, except it uses a non-fault collision penalty instead of the at-fault collision penalty of -1,000,000 assigned to the ego vehicle. This reduction in magnitude is because the traffic vehicle is typically not considered at fault for a collision. The merge vehicle is usually required to yield for traffic. When multiple vehicles are introduced later in this work, the rear in-lane vehicles are assigned at-fault collisions when they collide with a vehicle in front of them.

3.3.3 Discretization

Many aspects of the simulator are discrete, like the time steps and five action values. When training, each action learns a Q-value independent from the others. There is no relationship between the distinct, discrete actions, even though they define a range of values with similar step sizes between each value in the set. Similarly, the state variables are also discrete. Table 2 shows the closing gap, closing speed, and TTP states all consist of discrete values. Experimentation happens on different discretization sizes and ranges to determine if some performed better. Ultimately, there is no clear connection between discretization size or steps, but increased discretization does result in the well-known *curse of dimensionality*. In some trials, the memory required to store the variables becomes too large and surpasses limits in MATLAB (e.g., the size of the Q-learning tables). Increasing the discrete set sizes also increases the training iterations needed to sufficiently visit all states.

3.3.4 Q-Learning Simulator Variables

Defined here are the specific equations, variables, and units the simulator uses. These equations calculate the motion of the vehicles and their states for both the training and testing simulators.

Variables and units:

l_{ramp}: ramp length, (m)

x_{ego}: ego position, (m)

$x_{traffic}$: traffic position, (m)

L_{ego}: ego vehicle length, (m)

$L_{traffic}$: traffic vehicle length, (m)

v_{ego}: ego velocity, (m/s)

$v_{traffic}$: traffic velocity, (m/s)

a_{ego}: ego acceleration, (m/s^2)

$a_{traffic}$: traffic acceleration, (m/s^2)

t: time increment, (s)

3.3.5 Q-Learning Simulator Equations

$$TTP = \frac{l_{ramp} - \left(x_{ego} + \frac{L_{ego}}{2}\right)}{v_{ego}}$$

Equation 9: Time to Position

$$v_{ego} = v_{0_{ego}} + a_{ego}t$$

Equation 10: Next state velocity, Ego

$$v_{traffic} = v_{0_{traffic}} + a_{traffic}t$$

Equation 11: Next state velocity, traffic

$$x_{ego} = x_{0_{ego}} + v_{ego}t + \frac{1}{2}a_{ego}t^2$$

Equation 12: Next state position, Ego

$$x_{traffic} = x_{0_{traffic}} + v_{traffic}t + \frac{1}{2}a_{traffic}t^2$$

Equation 13: Next state position, traffic

$$v_{closing} = v_{ego} - v_{traffic}$$

Equation 14: Closing speed

$$l_{closing} = \left|x_{ego} - x_{traffic}\right| - \left(\frac{L_{ego}}{2} + \frac{L_{traffic}}{2}\right)$$

Equation 15: Closing gap

$$P_{relative} = \begin{cases} -1 & \text{if Ego is behind traffic,} \\ 1 & \text{else} \end{cases}$$

Equation 16: Relative position

3.3.6 Simulator for Testing

As previously discussed, there is both a simulator for training and a simulator for testing. The training simulator develops the trained Q-tables. A file generated by the training simulator stores the Q-tables. The testing simulator reads the tables from the file. The testing simulator output is a table of test values used to measure the performance of the output from the training. The output is stored in tabular format and loaded into a separate file in Excel for further evaluation.

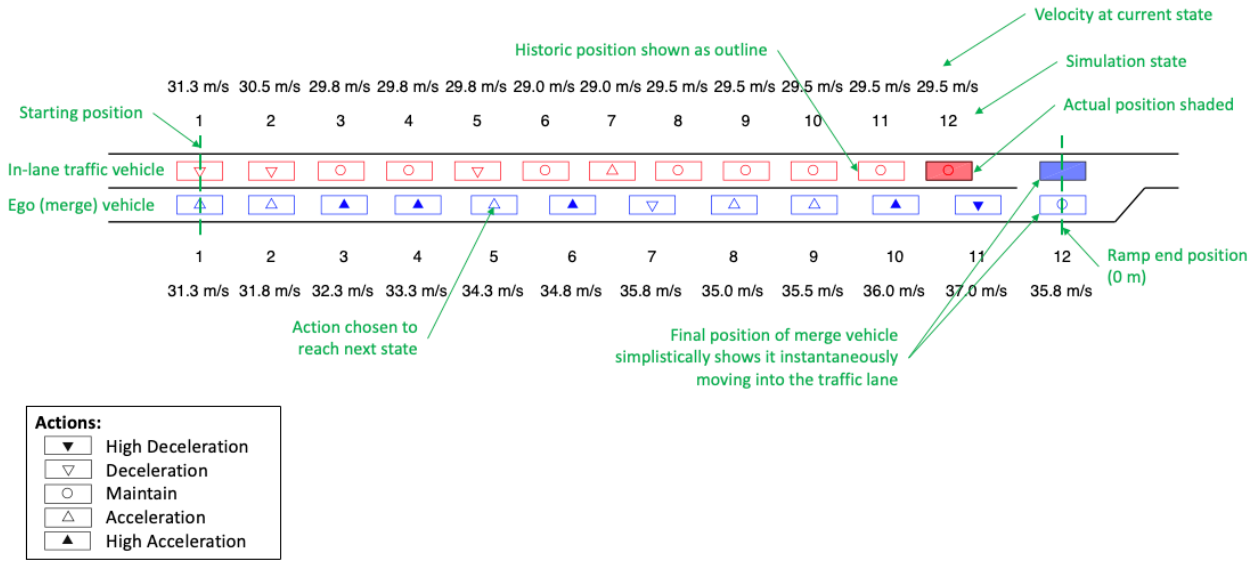


Figure 14: Output figure for Q-learning testing simulator. Green text with arrows identifying key features.

Figure 14 shows an optional output graphic generated if the render switch is enabled. The graphic shows the vehicles as they proceed down their respective lanes. The vehicles are color-coded: blue for ego (merge) and red for traffic. The rendering shows an outline of the vehicle's previous position and a symbol representing the action taken at that position. The legend shows the five discrete actions that the vehicles can take.

Table 3 shows the parameters used in the testing simulator. The simulator records collision results for a set of differential starting positions. This approach for evaluation is unique within the body of work reviewed for merge simulators. It enables a standard method for performance evaluation between the two merging vehicles.

Table 3: Q-Learning testing simulator parameters and values

Testing Simulator Parameter	Value
Testing Time Increment (s)	0.25, 0.5, 0.1
Ego Starting Velocity (MPH)	70, 55, 68, 65, 90
Test Ramp Length (m)	100, 50, 20
Test Traffic Type	Constant, Random, Reactive
Test Q-Action	Merge & Traffic, Merger Only
Ego Starting Position (m)	{-100, -50, -40, -30, -20, -19, -18, -17, -16, -15, -14, -13, -12, -11, -10, -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 100}

3.3.7 Traffic Action Policies

The testing simulator uses the same equations of standard motion used in the training simulator. The simulation begins after loading the Q-learning table from the designated training file. The traffic vehicle has three policy options: *constant*, *random*, and *reactive*.

The constant action policy sets a constant zero value acceleration through the episode's duration. In other words, the traffic vehicle will have no acceleration or deceleration. It will maintain the set starting velocity throughout the simulation episode.

The random action policy selects a random acceleration from within the acceleration range shown in Table 3 and Table 5, which is [-5,4] m/s^2. The simulator selects a new random acceleration value from within this range at each step until the episode is complete.

The reactive action policy selects the action from the trained traffic Q-learning table. Just like the random action policy, the acceleration range is [-5,4] m/s^2 as shown in shown in Table 3 and Table 5. The traffic vehicle Q-learning table learns values parallel to the ego Q-learning table. It is a duplicate Q-table, and both operate in the same way. It uses the same actions, states, formulas, and value ranges as the ego vehicle. The traffic vehicle Q-table is a single-action-

learner that only considers its own action. Since both vehicles are trained agents, this policy is a more traditional view of multi-agent.

Another option within the simulator is to use a multi-agent Q-table or single-agent Q-table. This option is only available for the ego (merge) vehicle. If the multi-agent variant is enabled, the state set includes the traffic vehicle action to select the ego vehicle action from the Q-table. Otherwise, the single-agent version does not consider the action of the traffic vehicle when selecting the best Q-table action.

3.4 Results, Discussion, and Limitations

A total of six strategies evaluate the Q-learning approach: two different ego strategies (SAL and JAL) and three different traffic reaction policies (constant, random, and reactive). In addition, the simulator evaluates the different settings chosen during training shown in Table 2 and Table 3: training iterations (10^5, 10^6, 10^7), domain sizes for the states (closing gap, closing speed, and TTP), and simulation time step size. Finally, Table 4 shows the evaluation results.

In each simulation run, the ego starting position is varied. The zero starting position represents a distance of 100 meters away from the merge point. Meaning when the ego vehicle starts at zero, the zero position is the start of a 100 m on-ramp where the end position is the merge point. Figure 12 shows the zero position and the goal distance. The traffic vehicle always starts at the zero position in the test simulator. When the ego is at a zero start position, so is the traffic vehicle. For example, if the ego vehicle is at -2, it starts 2 m behind the traffic vehicle or 102 m away from the merge point. The table does not show values outside of 20 to -20 because no collisions occur outside of this range. Each cell represents 52 unique data points due to the different combinations of evaluation parameters.

Table 4: Aggregate results for all simulation variants of Q-learning runs.

Average of Collision % at 100m Goal Position								
	Learning:	**Single Action Q**			**Joint Action Q**			**Overall**
	Traffic Policy:	***Reactive***	***Constant***	***Random***	***Reactive***	***Constant***	***Random***	**Average**
Ego Initial Position	**-20**	0%	0%	3%	0%	0%	5%	1%
	-15	0%	4%	7%	0%	0%	12%	4%
	-10	0%	4%	9%	0%	2%	18%	5%
	-5	4%	19%	16%	2%	0%	21%	10%
	-4	10%	15%	18%	4%	6%	22%	12%
	-3	21%	23%	21%	2%	10%	24%	17%
	-2	29%	27%	22%	8%	19%	25%	22%
	-1	37%	27%	24%	6%	25%	29%	25%
	0	25%	27%	22%	17%	42%	31%	27%
	1	25%	23%	19%	19%	33%	30%	25%
	2	33%	21%	19%	15%	21%	30%	23%
	3	23%	23%	13%	21%	29%	29%	23%
	4	6%	2%	5%	8%	21%	24%	11%
	5	6%	6%	6%	4%	4%	22%	8%
	10	4%	4%	3%	2%	4%	11%	5%
	15	2%	2%	1%	0%	0%	4%	1%
	20	2%	2%	1%	0%	0%	1%	1%
	Average:	**13%**	**13%**	**12%**	**6%**	**13%**	**20%**	**13.0%**

The data shows that the joint-action-learner has fewer collisions on average than the single-action-learner against a reactive traffic policy. The reactive traffic policy is when the traffic vehicle selects actions from a trained Q-table, just like the ego vehicle. Both vehicles have learned to avoid a collision with each other. Generally, the table shows that collisions occur more frequently when both vehicles have an equal starting position. This is consistent with Proposition II from the T-Grid chapter: roughly restated as, if two vehicles are in close proximity and the reaction of either is uncertain, the chance of collision is higher.

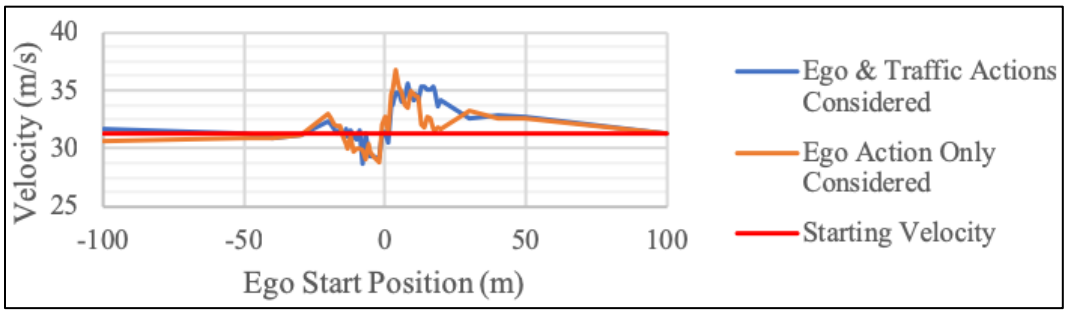


Figure 15: Average ego speed at the merge point. Plot of the average ego speed, over all runs, when merging into traffic. The plot is consistent with Proposition III, showing that it is advantageous for a lagging merge vehicle to decelerate and for a leading merge vehicle to accelerate.

Figure 15 shows how the ego vehicle velocity changes when it starts close to the merge position. The ego vehicle generally maintains its starting velocity if the merge position is far away when it starts (about 25 m or more from the merge point). This behavior is consistent with another proposition (Proposition III) found with the T-grid study: if the merge vehicle lags, it is advantageous to decelerate and if it leads, it is advantageous to accelerate to avoid collisions.

The single-action-learner, reactive traffic policy results in Table 4 might support Proposition IV, albeit weakly. The results seem to be about the same as the single-action constant strategy and the random strategy. The expectation is that because both vehicles use a Q-learning trained controller, they would be able to learn how to avoid a collision. However, both Q-tables are fundamentally the same because they use the same set of states and almost the same reward values (the traffic vehicle is not at fault, so its collision reward is 10x less than the ego vehicle). Because the controllers for the two vehicles are basically the same, a weak conclusion is that they have worse performance because they use the same style, consistent with Proposition IV.

The constant policy results were expected to be better than shown. The ego vehicle should be able to avoid a collision relatively easily because the strategy of the traffic vehicle is fixed at a constant speed. Further study could help understand why. However, my work omits

studies like this and instead focuses on a better-performing approach of DRL (deep reinforcement learning).

Table 4 values do not represent an ideal training scenario. They are the aggregate results of all training and multiple test scenarios. There are inevitably more ideal state sets and configurations that produce better results than the aggregate results, but it would be difficult to find those optimized settings. The Q-learning simulations' results helped show that the simulator appears to be learning to avoid collisions and that its behavior aligned with findings from the fundamental T-grid study. In addition, reviewing the data from Q-learning tables, it became apparent that patterns formed, making it more ideal for applying a continuous action-state space approach based on function approximation.

3.4.1 Experimental Results Using the I-80 Dataset

The NGSIM I-80 dataset is a publicly available standard dataset for merging [50]. The dataset captured vehicle positions over time on a congested highway, the I-80 highway in the San Francisco Bay area in Emeryville, CA. It is a standard dataset for studying highway merging traffic to train and test different algorithms. I applied the Q-learning approach to the I-80 dataset.

The data was processed to find all vehicles in the traffic lane closest to the merge lane. This resulted in 1,246 vehicle trajectories to use for testing. The Q-learning simulator was modified to import the vehicle data and then test the ego merge vehicle performance against it with the same strategy that varied the ego vehicle ramp start position. The ramp length was a fixed 100m. The 1,246 simulations ran in the same two-vehicle setup presented throughout this work without considering the rest of the vehicles in the road scene.

The Q-learning simulation needed to be re-scaled to run at lower speeds and to scale the acceleration and deceleration values because the I-80 dataset represents a congested road scene.

However, the overall strategy and simulation structure remained the same. The results achieved showed collision rates at about 1%, which is much better than other state-of-the-art controllers that ran tests against the same dataset. For comparison, in the work by Dong et al. [1], the collision results for the NGSIM dataset were about 7%.

However, because other merge test research does not use a framework like mine to standardize the positions of the vehicles, it is difficult to compare performance truly. In addition, my study used a two-vehicle pair, disregarding the other vehicles. Another key difference is that the behavior of vehicles during merging is very interactive. The closeness of vehicles during merging produces a reaction to vehicle action. Reactions from people participating in merge studies have shown that they expect other vehicles to react to their behavior and find it abnormal when they do not [72]. However, there is no reaction in a dataset like NGSIM I-80 or others that are similar, so judging actual performance is difficult.

3.5 Motivation for a Continuous Approach

Patterns are forming when exploring the Q-Learning data table created from training. One way to see the pattern was by reviewing training sets with identical parameters of closing gaps, closing speeds, and TTPs for all three training iteration cases: 10^5, 10^6, and 10^7. The 6-dimensional array for a two-agent Q-table was reshaped into a single vector of values. Figure 16 shows the data. Overall, the average values for all three training cases increase as the number of iterations increases. All three training cases form a similar pattern.

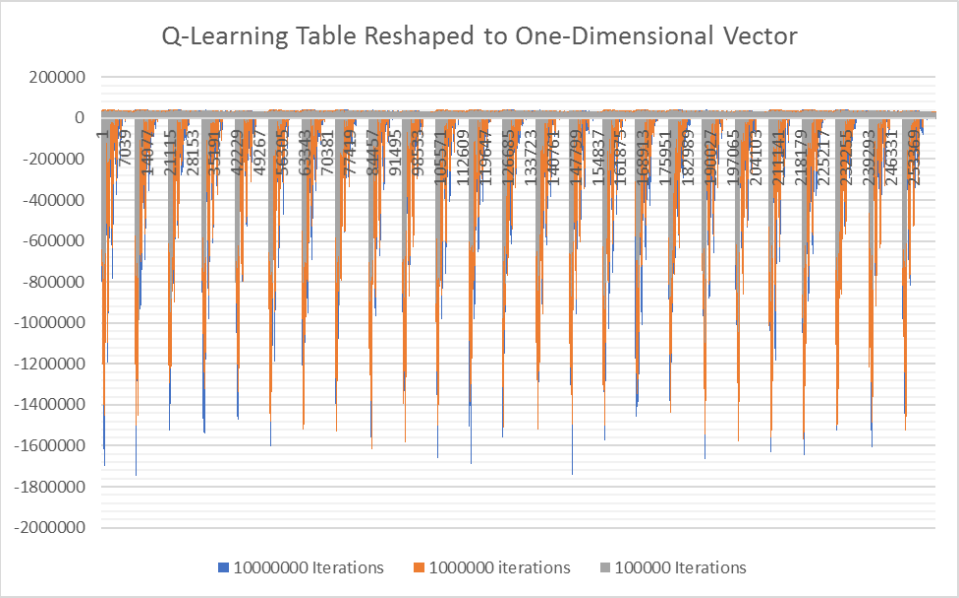


Figure 16: Q-learning data reshaped into a one-dimensional vector

One observed pattern is a grouping of data into 25 distinct groups. Reviewing the data, these 25 groups represent the five lane actions that form five subgroups for each merge action. Figure 17 and Figure 18 show a zoomed-in view of the positive values of the data. In this view, individual groups form for each of the eight TTPs. Again, more patterns have formed.

The meaning of the patterns is not determined. However, the existence of the patterns suggests that a more formal function approximation approach might better fit the patterns forming in the data. The following chapters present a function approximation approach that replaces the discretized state-action space with a continuous state-action space.

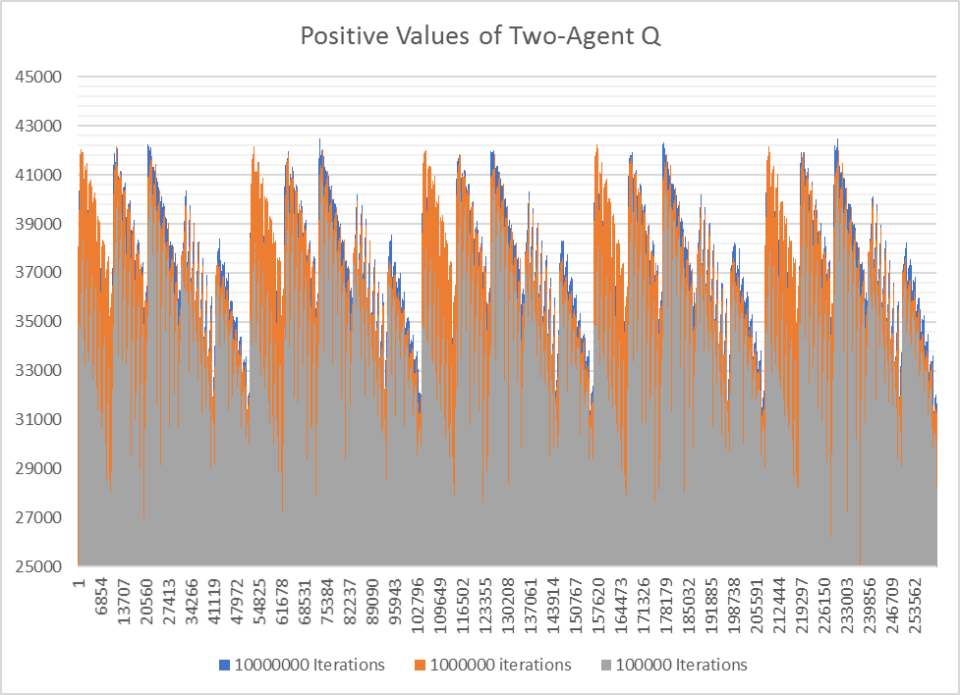


Figure 17: Two-agent Q-learning data, positive values only

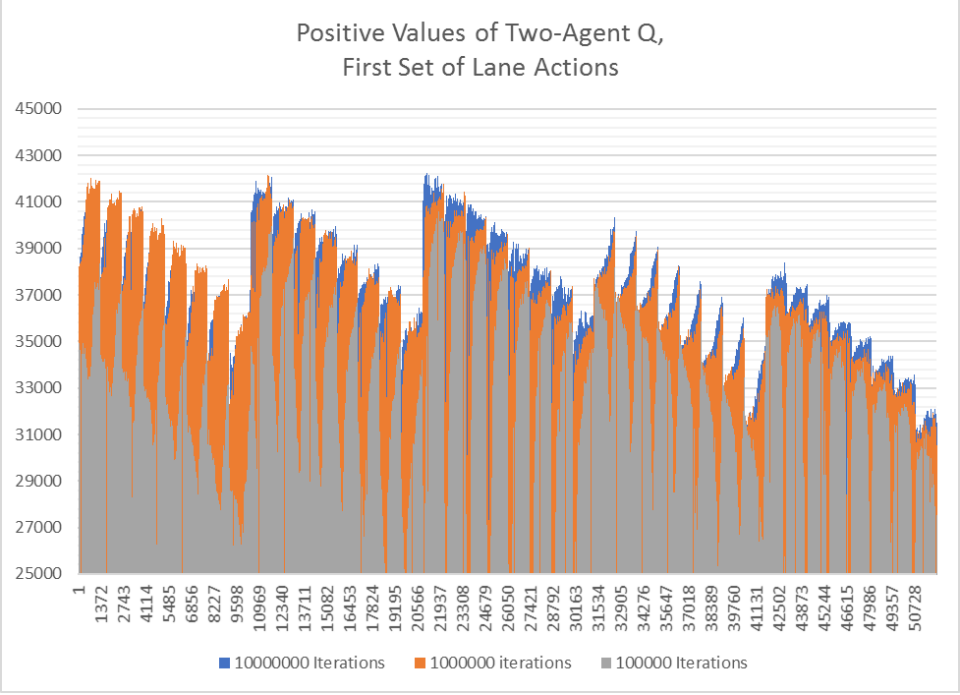


Figure 18: Closer look at positive values for two-agent Q-learning data to observe pattern formed by TTP states.

Chapter 4 Deep Reinforcement Learning Approach

A deep reinforcement learning (DRL) approach builds upon the Q-learning approach. The DRL approach uses a deep neural network and continuous state-action space, unlike the Q-learning approach, which uses discretized states and actions. The DRL approach is multi-agent, just like the Q-learning approach. The Q-learning tables show patterns in the training simulator data, motivating a function approximation approach, as explained in Section 3.5. The repetitive nature of the patterns suggested that a continuous state-action space function approximation approach might produce a more suitable model for the merge problem. Ultimately, the DRL approach is very successful, and the results show nearly perfect performance against a standard test set I developed to evaluate merging performance.

In recent years, neural networks, especially deep learning, have shown incredible performance in solving problems and continue to grow in popularity [73], [74]. However, despite their promise, a neural network approach was not used in my early work because it was important to understand the merge problem from a fundamental perspective. The grid-game approach to merging led to a foundational understanding of the problem and generated four fundamental propositions of merging behavior. The propositions were present in the Q-learning simulation results, and using the propositions led to a richer understanding of the results. With the fundamental understanding, Q-learning simulation results, and patterns in the generated training data set, the next step was to take a more sophisticated approach to the problem. The

new approach is deep learning. An approach known to produce good performance in other physical application problems.

Deep learning “discovers intricate structure in large data sets” [75] compared to prior methods where a significant setup was needed to create a framework for learning features. Deep learning has produced performance breakthroughs in many areas, like object recognition, speech recognition, and signal processing, to name a few [76]. Deep learning uses multiple network layers to model the application that they represent. A common explanation of Deep Learning is that it can automatically do this progressive abstraction representation, thus simplifying the task of automated learning. While it is difficult to prove this particular capability of deep learning [59], it is clear that deep learning performance is impressive, and its usage continues to grow.

DDPG is the DRL approach to the merge problem. DDPG shows human-level performance when playing games that simulate physical tasks [77]. The Google DeepMind team is the developer of DDPG. DeepMind is the team that showed better than human AI performance in AlphaGo [74] and more recently developed AI to predict protein folding called AlphaFold [78].

DDPG is an actor-critic method that, as the name indicates, employs a deep learning-based policy gradient. It is a continuous-space method with strong performance in many classic control problems that simulate physical tasks. Similarly, several recent examples of deep learning show human-level performance in relevant applications [73], [79], [80]. For example, DRL shows human-level performance when applied to Atari games.

Overall, the results for the DRL approach to merging were excellent. The results show no collisions with the same scenario applied in the Q-learning testing. Figure 19 shows the results. A tabular framework compares goal position distance and differential starting positions between

the two-vehicle pair to find the limits of the performance. Figure 28 shows an example of this framework. When viewing the results in this format, they show performance near the physical limitations of the scenario. This means that the network has learned to produce almost perfect results. These results are a clear improvement over the Q-learning approach. The approach and the results were published in the second paper related to this research [4].

Average Collision %							
Ego Start Pos. (m)	Q Action Considered by Ego:						Avg.
	Ego Only			Ego & Traffic			
	Traffic Reaction:						
	Q	Const.	Rand.	Q	Const.	Rand.	
-20	0%	0%	3%	0%	0%	5%	1%
-15	0%	4%	7%	0%	0%	12%	4%
-10	0%	4%	9%	0%	2%	18%	5%
-5	4%	19%	16%	2%	0%	21%	10%
-4	10%	15%	18%	4%	6%	22%	12%
-3	21%	23%	21%	2%	10%	24%	17%
-2	29%	27%	22%	8%	19%	25%	22%
-1	37%	27%	24%	6%	25%	29%	25%
0	25%	27%	22%	17%	42%	31%	27%
1	25%	23%	19%	19%	33%	30%	25%
2	33%	21%	19%	15%	21%	30%	23%
3	23%	23%	13%	21%	29%	29%	23%
4	6%	2%	5%	8%	21%	24%	11%
5	6%	6%	6%	4%	4%	22%	8%
10	4%	4%	3%	2%	4%	11%	5%
15	2%	2%	1%	0%	0%	4%	1%
20	2%	2%	1%	0%	0%	1%	1%
Avg.	13%	13%	12%	6%	13%	20%	13%

Average Collision %, Ego Considers Self Action Only, Random Action Traffic											
		Goal Position									
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	100%	27%	0%	0%	0%	0%	0%	0%	0%	0%	13%
	-3	100%	100%	20%	0%	0%	0%	0%	0%	0%	0%	22%
	-2	100%	100%	97%	3%	0%	0%	0%	0%	0%	0%	30%
	-1	100%	100%	100%	47%	3%	0%	0%	0%	0%	0%	35%
	0	100%	100%	100%	97%	57%	3%	0%	0%	0%	0%	46%
	1	100%	100%	100%	97%	7%	0%	0%	0%	0%	0%	40%
	2	100%	100%	100%	47%	0%	0%	0%	0%	0%	0%	35%
	3	100%	100%	70%	0%	0%	0%	0%	0%	0%	0%	27%
	4	100%	100%	3%	0%	0%	0%	0%	0%	0%	0%	20%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	53%	49%	35%	17%	4%	0%	0%	0%	0%	0%	15.7%

Figure 19: Collision tables showing average collision for ego starting position with respect to the traffic vehicle. The results show that the DRL method produces collision-free runs at the 100 m goal position, whereas the Q-learning method has collisions at every ego start position.

Left: Results from Q-learning simulations. The goal position for the traffic vehicle is set at 100 m for all variations. The random traffic action policy combined with the ego considering only its own action (and not that of traffic) is circled in red and highlighted yellow.

Right: results from the same policy action combination highlighted on the left, but expanded to show 10 different goal positions. The 100 m goal position is highlighted. It is the exact same scenario as the highlighted column in the Q-learning results table on the left.

4.1 Approach and Simulator

The setup of the DRL approach is essentially the same as the Q-learning approach. A summary of the simulator is presented in this section to summarize its structure and functionality. Section 3.3 shows additional details about the simulation approach. The approach uses the same standard equations of motion,

$$x = \frac{1}{2}at^2 + vt + x_{initial}$$

Equation 17: Standard equation of motion

where x is the position in meters, v is the velocity in meters per second (m/s), a is the acceleration in m/s^2, and t is the time in seconds. The Q-learning algorithm is written in MATLAB, but the DRL version was rewritten using Python, keeping all the basic functionality and calculations the same. The Python program uses TensorFlow for the neural network (NN) computation and a customized environment within OpenAI's Gym toolbox. The Gym environment is based on the classic Continuous Mountain Car control problem, with significant adaptations and development for the merge scenario.

4.1.1 Two-Vehicle Simulation States and Variables

The simulator uses self-play to train, just like the Q-learning version. The two agents learn by playing against themselves. The start of each simulation episode initializes variables with random values within the ranges shown in Table 5. The simulator has a switch to enable or disable Joint Action Learning (JAL). When enabled, the state includes the action of the in-lane traffic vehicle in addition to the other four state variables.

The initial states are calculated based on the initial values of the variables. The states are the same as the Q-learning approach in Section 3.2).

Table 5: DRL training simulator variables

Variable	Value(s)	Units
ego/traffic acceleration	[−5, 4]	m/s^2
ego/traffic initial position	[−10, 150]	m
ego/traffic speed	[20, 40]	m/s
TTP	[0,3]	s
goal position	[25, 150]	m
time step	0.1	s
ego length	5	m
traffic length	[1,20]	m

4.1.2 Equations of Motion

The step function calculates a new velocity and position based on the chosen acceleration value and the state's current values. The following formulas calculate the new velocity and position:

$$v_{new} = a\mathrm{t} + v_{current}$$

Equation 18: Next state velocity

$$x_{new} = \frac{1}{2}at^2 + v_{current}t + x_{current}$$

Equation 19: Next state position

where v_{new} is the new velocity in m/s, a is the chosen acceleration in m/s^2, t is the time in seconds, $v_{current}$ is the current velocity in m/s, x_{new} is the new position in meters, and $x_{current}$ is the current position in meters.

4.1.3 Traffic Action Policies

The acceleration scheme for each training episode is chosen randomly out of three choices: *constant*, *random*, or *reactive*. The probability for each action choice is equal to ensure an equal mixture of each action choice in the training network. These are the same policies from the Q-learning approach described in Section 3.3.7.

The constant action policy sets a constant zero value acceleration through the episode's duration. The randomly chosen velocity from within the range shown in Table 5 will remain constant in each time step until the episode is complete.

The random action policy selects a random acceleration from within the acceleration range shown in Table 5, which is [-5,4] m/s^2. Each step selects a new random acceleration value from within this range. This random selection will continue until the episode is complete.

The reactive action policy selects the action from the trained traffic DRL network. The network trains in parallel to the DRL network for the ego vehicle. The traffic DRL network is identical to the merge vehicle DRL network, but it is a separate network that learns its own weights and thus, its own action choices. Further explanation of the network is given later in Section 4.2, Simulator DDPG DRL Network.

4.1.4 Reward Function

Acceleration reward values for the state at each step train the DRL network. The reward function is the same as the Q-learning simulation. Refer to Section *3.3.2 Reward Function Design* for further details. Once the ego vehicle has reached the goal position, the time-step function loop ends. If the ego vehicle crosses the goal position and there is any gap between it and the traffic vehicle, the merge reward is assigned. A collision has occurred if there is no gap between the ego and traffic vehicles, and the collision penalty is assigned. Once the episode is complete, the initial state values reset to random values selected within the parameter range. The

simulation continues until all episodes are complete. If the graphical output is enabled, it displays on the screen and updates at each time step. Figure 20 shows an example of the graphical output.

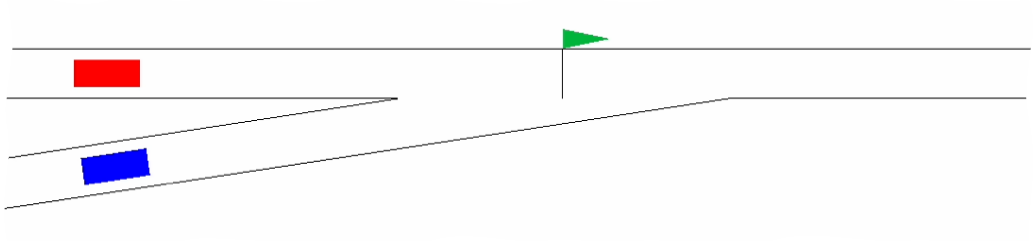


Figure 20: Simulator graphical output. The blue rectangle represents the ego vehicle in the merge lane of a taper-type on-ramp. The red vehicle is the traffic vehicle. The green triangular flag represents the goal location: the first point where the vehicles are laterally close enough to collide. The simulation ends when the ego vehicle crosses the goal.

4.2 Simulator DDPG DRL Network

A DDPG DRL network is use to learn the merging behavior. It was developed by the Google Deepmind team to "adapt the ideas underlying the success of Deep Q-Learning (DQN) to the continuous action domain" [77]. DQN has shown human-level performance in applications like Atari games [73]. However, it is not a continuous approach, so it begins to suffer the curse of dimensionality as the number of actions or states increases. DDPG improves dimensionality concerns with discretized spaces by using a continuous state-action space. DDPG uses an actor-critic architecture like the traditional one shown in Figure 22. Figure 21 shows the algorithm pseudocode from the original paper [77].

Algorithm 1 DDPG algorithm

Randomly initialize critic network $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ.
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer R
for episode = 1, M **do**
  Initialize a random process $\mathcal{N}$ for action exploration
  Receive initial observation state s_1
  for t = 1, T **do**
    Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$ according to the current policy and exploration noise
    Execute action a_t and observe reward r_t and observe new state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in R
    Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
    Update critic by minimizing the loss: $L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
    Update the actor policy using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s_i}$$

    Update the target networks:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

  end for
end for

Figure 21: Deep dynamic policy gradient algorithm pseudocode from the paper by the Deepmind team [77].

Actor-critic methods, as the name implies, have an actor and critic component. The actor takes actions based on learned experience, and the critic critiques the actor's performance. Figure 22 shows the typical components of actor-critic methods. The actor learns the actions to choose through policy-based training. In the merge algorithm, the actions are acceleration. The actor is a DDPG network. The critic reviews the temporal-difference (TD) error or loss, as it is called in the DDPG algorithm. The loss updates the critic, as the actor-critic figure illustrates. The actor is updated using the sampled policy gradient. Figure 22 shows how the environment accepts the action taken and outputs the reward and the new state. The merge environment, including the states, actions, and rewards, is explained in detail in Section 6.2.1. The cycle repeats using the loss and policy gradient to update the networks.

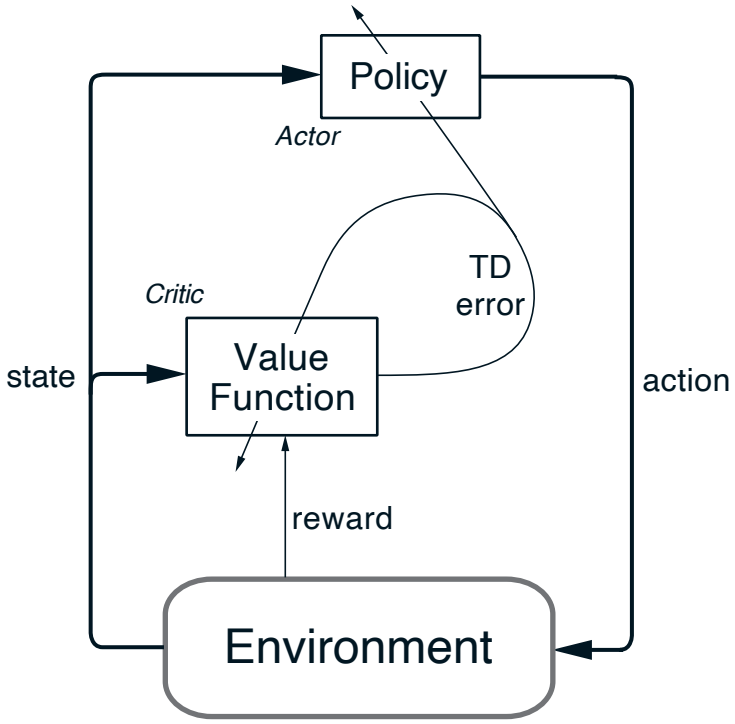


Figure 22: Actor-critic block diagram as illustrated by Sutton and Barto [81].

The environment used in the DDPG network is the same code used for the Q-learning approach, but it was re-coded in Python and kept as identical as possible to the MATLAB code. It was re-coded in Python to use standard available libraries and utilities like TensorFlow and OpenAI Gym. Section 4.1 *Approach and Simulator* describes the simulator. A description of the functionality of the code is as follows.

The algorithm begins by importing the standard libraries for TensorFlow, OpenAI Gym, NumPy, and Time. Next, classes for both the actor and critic functions are defined. Each class includes functions to build, initialize, and train the actor and critic neural networks. Next, the neural networks are built using the TensorFlow library. The code for this DDPG algorithm was based on a library available on GitHub [82] but adapted for use in the simulator environment. The code follows the structure and strategy defined by the Deepmind team [77]. The algorithm

allow storing and reading of trained networks, enabling network restoration for further training and simulation runs on trained networks.

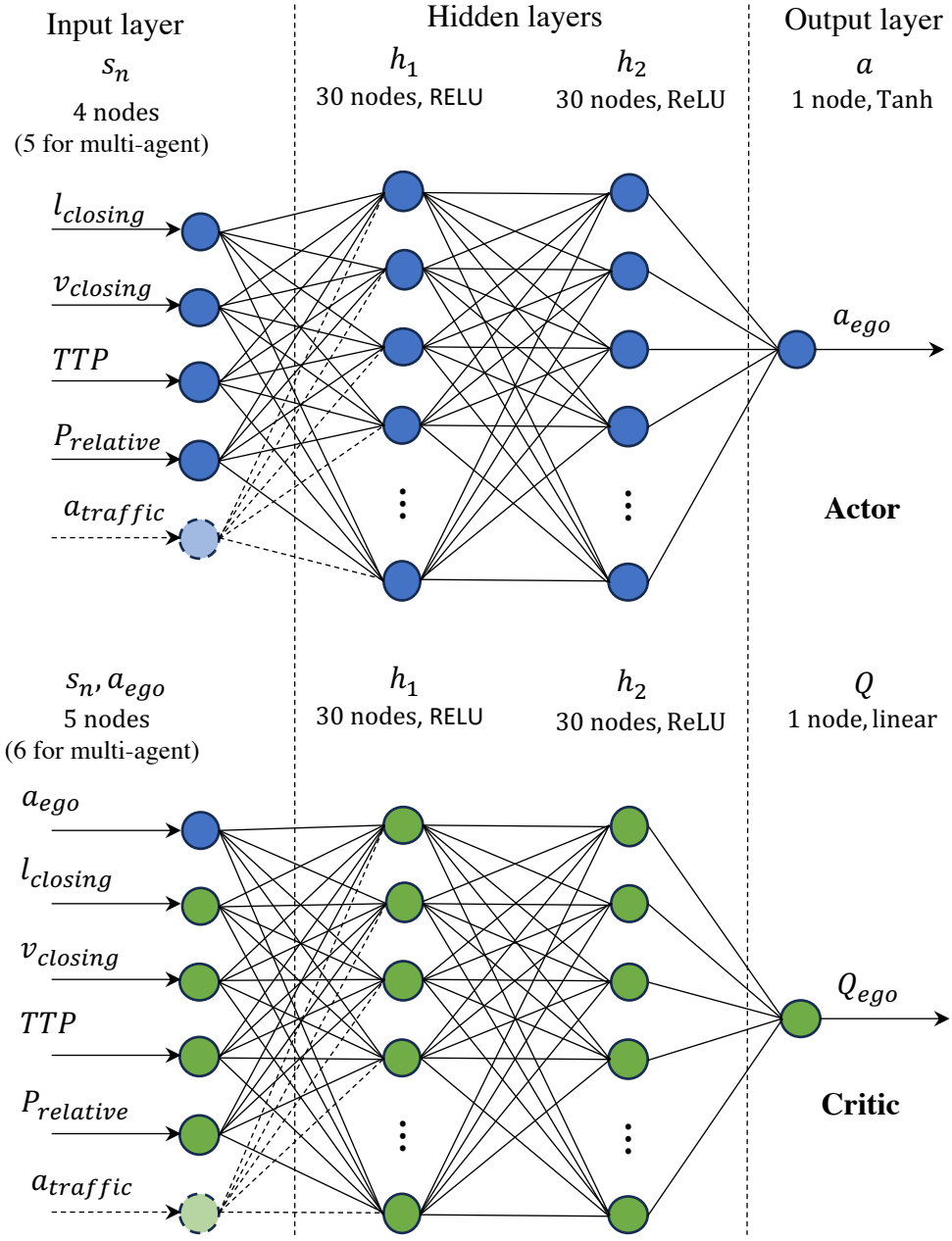


Figure 23: Two-vehicle Ego DRL network diagram.

4.2.1 Network Architecture

The actor-critic neural networks use a 30-node width in the two hidden layers. The neural networks use a ReLU (rectified linear units) activation function for the hidden layers and a *tanh*

(hyperbolic tangent) activation function for the output layer. Figure 23 shows the network diagram for the ego merge vehicle. The diagram is the same for the reactive traffic vehicle. The learning rates for both the actor and the critic remain at the default values of 0.001, and the reward discount, gamma, was set at 0.9. The networks use a memory replay buffer with a batch size of 32 and a memory capacity of 10,000 samples.

4.2.2 Training Using Noise

The training is very sensitive to exploration. Exploration sensitivity is a 'major challenge of learning in continuous action spaces.'[77] The DDPG algorithm uses a decaying noise added to the chosen action value. The noise is Gaussian random that starts with a width that is the full value of the action range and decays to zero with each new sample selection each time a simulation step is taken (as described in Section 4.1). The noise value spans the entire range of available action at the start of training, effectively producing a random value within the action range during the initial training episodes. Over time, the noise decays at a multiplicative rate of 0.999995 until it diminishes to nearly zero. Once it fully decays, exploration is complete, and the action value used in the step function is the actual value the neural network selects. The noise process is the same that is used by the originators of the DDPG algorithm [77], but they draw upon ideas previously generated by Uhlenbeck & Ornstein [83] and Wawrzyński & Tanwani [84].

4.3 Test Setup

Evaluation testing is performed after training. The algorithm automates the evaluation testing. The test holds several variables constant to evaluate the variable parameters consistently. It operates in the same way as the Q-learning evaluation does. The initial speeds, vehicle lengths,

time step, and the traffic vehicle's initial position are all held constant. However, the goal position and the initial position of the ego vary. The intent of this variation is to understand the performance of the merging vehicles as compared to their proximity to each other and the ramp merge point. Table 6 summarizes the variables.

Table 6: DRL testing simulator variables.

Variable	**Value(s)**	**Units**
ego initial speed	31.29	m/s
traffic initial speed	31.29	m/s
ego length	5	m
traffic length	5	m
traffic initial position	0	m
time step	0.1	s
goal position	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100]	m
ego initial position	[-100, -50, -40, -30, -20, -19, -18, -17, -16, -15, -14, -13, -12, -11, -10, -9, -8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 100]	m

As previously mentioned, the evaluation framework is a unique contribution of my work which can be used by other researchers. Figure 28 shows one example of many of this framework. The framework enables performance evaluation of an algorithm within the context of on-ramp merging. Presenting the data with standard evaluation parameters helps to conceptualize the scenario and consistently evaluate the performance. It compares the relative positions of the two-vehicle scene and summarizes the average collisions that occur when testing. The second dimension is the goal position. Evaluations presented in other research typically use real-world testing data, but it does not show proximity-based variable data as my work does. This evaluation method can be viewed as a standard to compare results between variants presented here and methods presented in other research. Other researchers could adopt this method to compare performance in a standard way.

Chapter 5 Two-Vehicle Simulation

This chapter explores a two-vehicle DRL simulation. The simulation mimics the two-vehicle Q-learning approach. The two-vehicle simulation scenario consists of one vehicle in the merge lane and one in the traffic lane. This is the same scenario as the Q-learning approach. The performance of the DRL approach is significantly better than the Q-learning approach. The DRL collision avoidance performance is nearly ideal. The DRL approach learns to avoid almost all collisions, except those that are physically impossible to avoid due to vehicle or road scene limitations. The performance was measured using a new framework I developed to show results in a standard way that other researchers studying lane merging can adopt too. This framework is another contribution of my research.

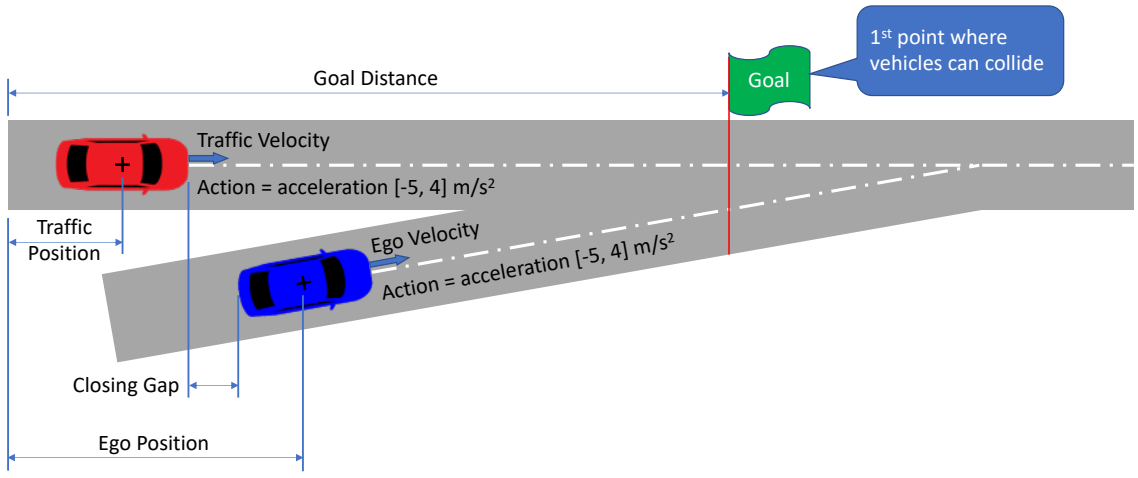


Figure 24: Two-vehicle simulation scene. Ego vehicle (blue) is in the merge lane. It learns through RL to avoid a collision with the traffic vehicle (red) while merging. The first point that the vehicles can collide is the goal position and is marked with a green flag.

Figure 24 shows the simulation scene. The ego vehicle (blue) learns through RL to avoid collisions with the traffic vehicle (red) by controlling its longitudinal position with a longitudinal acceleration action.

5.1 State Set

The only information the ego vehicle DRL network uses as inputs are the states shown in Table 7: closing gap, closing speed, time to goal position, and proximity to next lane vehicle. In instances with a JAL, the traffic action is also considered (see Section 3.1.1 for more information on joint-action-learning). A set of 28 parameters sets limits and default values in two-vehicle simulation (see Table 15), but parameter set and sophistication increase as more vehicles are added.

Table 7: Two-vehicle simulation state variables

Merge Vehicle State Variables	**Range**	**Units**
Closing Gap	[-2.5, 30]	m
Relative Closing Speed	[-10, 10]	m/s
Time to Goal Position	[0, 3]	s
Proximity to ego (*behind or in front of*)	{-1, 1}	*unitless*
Traffic Action (*for MA joint-action-learner*)	[-5, 4]	m/s^2

5.2 DRL Training

The training combines all three policies: *random*, *constant,* and *reactive*. The probability for each action choice is equal to help ensure a homogeneous mix of each policy. Section 4.1 *Approach and Simulator* gives further explanation of each policy. The training occurs at each time step of each episode, and each network was trained to 2.5 million episodes. The number of steps per episode is very roughly estimated at 25 using the following parameters: starting

position range of [-25, 50] m, goal position range of [25, 150] m, max and min speed of [20, 40] m/s, and step time of 0.1 s. Overall, this gives a very rough estimate of 62.5 million state-action-reward values that train the network. However, the number of episodes is a more significant metric because it represents the entire merge maneuver and all steps involved, including the vehicle's transition from the merge ramp onto the highway.

The network initializes with random values. Training occurs using the exploration policy, μ' as described in 4.2.2. At the start of the training, the noise applied is within the full range of the action value selected. This means that values learned at the beginning of training are basically random. The best noise decay multiplier is 0.999995. However, significant experimentation during training and evaluation was needed to determine the most appropriate value. The chosen exploration value decays to half its original value around $5K$ iterations and a near zero value around $52K$ episodes.

Several variables affect training performance: action exploration noise, traffic action policies, SAL versus JAL, and episode count. Experimentation on these variables shows they affect training performance. Figure 25 shows a graph with several different examples. Training shows that the best performing networks use a mixture of the three policies. In addition, experimentation shows that 2.5 million training episodes are sufficient to find the maximum reward value for most training runs.

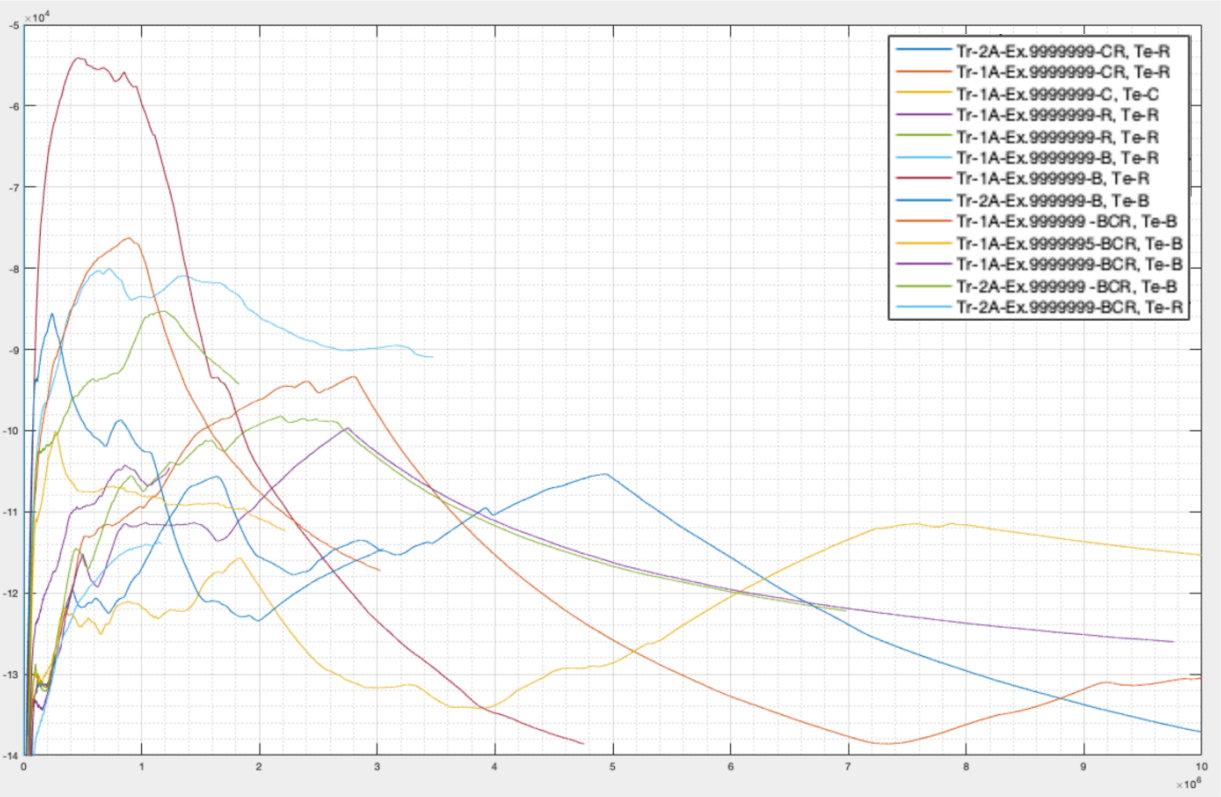


Figure 25: Cumulative average of reward values during training. Various exploration values are compared to find the most suitable value. Exploration values are encoded into the graph legend, shown after the Ex. Visual observation finds the peak value and checks for over-fitting.

Figure 26 shows the joint-action-learner (JAL) training results, where both the ego and traffic vehicles' actions are considered. The figure plots both the cumulative average and moving mean of the reward. The moving mean uses a window of 30,000 data points. The peak value is found through visual observation of the training plot. The best network is at or near this peak value. For this configuration, the peak value is at 1.12 x 10^6 episodes. Network checkpoints for the TensorFlow network are saved and stored every 10,000 episodes. Performance testing in 5.4 uses the network closest to the peak.

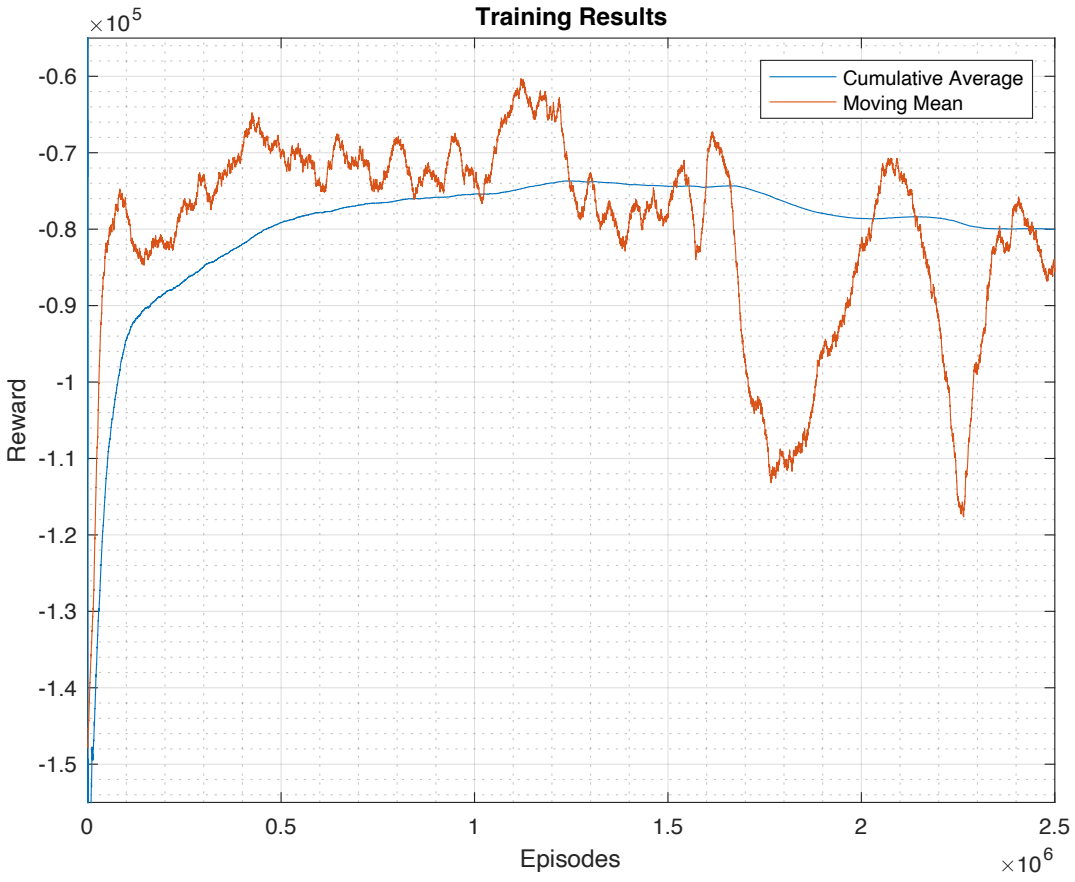


Figure 26: Joint-action-learner training plot. The cumulative average of rewards (blue) and the moving mean of rewards (orange) are shown. The moving mean uses a window of 30,000 data points. The peak value is at 1.12 x 10^6 episodes. Testing uses the saved network closest to the peak value.

Training also occurs using an independent-learner (IL) strategy to compare to the joint-action-learner. This IL single-action-learner considers only the ego action while evaluating the state and reward and does not consider the action of the traffic. All other aspects of the setup are identical to the joint-action-learner. Figure 27 shows the training reward value results.

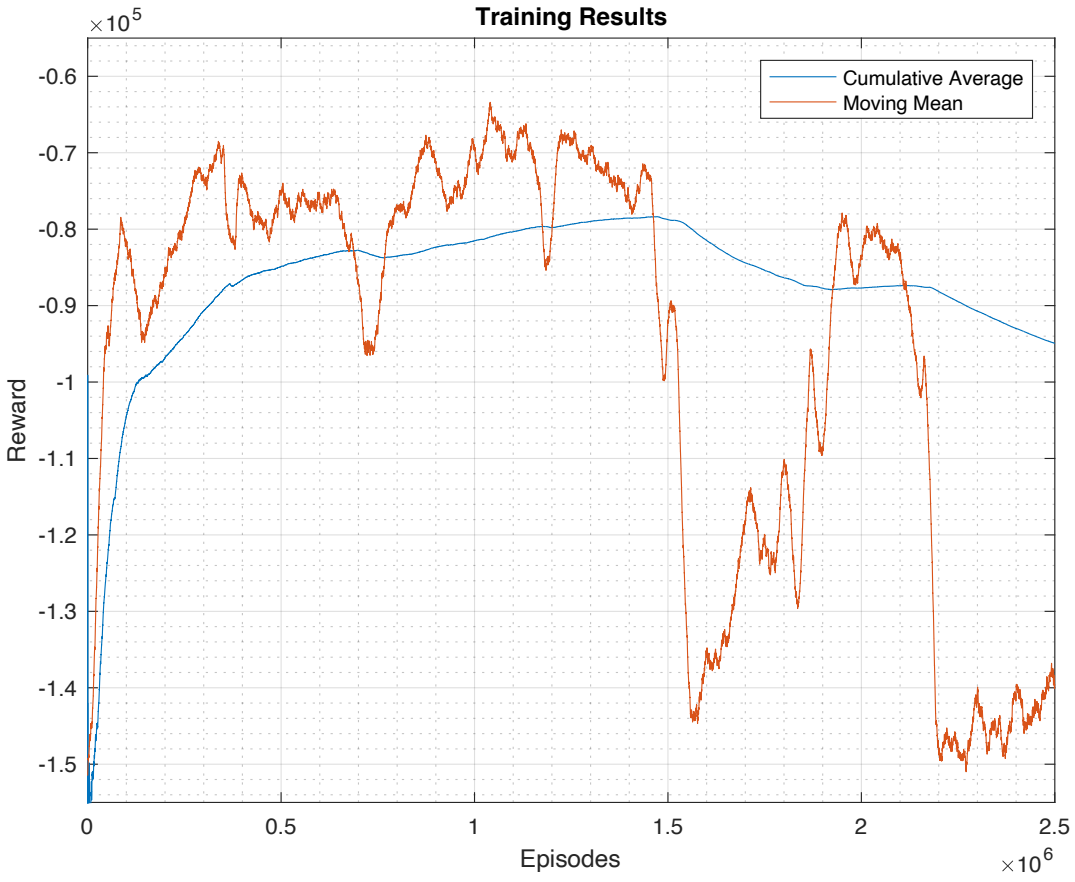


Figure 27: Single-action learner training results. The cumulative average (blue) and moving mean (orange) are shown. The moving mean uses a 30,000 data point window. The peak value in the moving mean is 1.04 x 10^6 episodes. Performance testing uses the saved network at the peak.

Overall, the training results between the JAL and IL are relatively similar. However, the performance of the JAL is slightly better, and the results appear to be more stable. Further evaluation is performed on both approaches and presented in the next section.

5.3 Reactive Policy Training for Traffic Vehicle

A separate TensorFlow DRL network for the reactive traffic vehicle trains in parallel to the ego network. The setup of the network is identical to the single-action-learner. It considers the same states as the ego vehicle, but it learns the action of the traffic vehicle and does not consider the action of the ego vehicle. It does not consider the ego vehicle action due to a limitation in the algorithm. Specifically, the algorithm selects the traffic and ego action

sequentially, so the traffic network training cannot consider the ego action. The same limitation exists with the Q-learning simulation. However, comparing the JAL and SAL shows that they are very similar. The state values contain both time and distance values, the same constituent parameters of acceleration. Therefore, there is likely some innate encoding of the acceleration into the state of the single-action learner despite its absence.

The reward structure for the traffic vehicle is also identical to ego, except it uses a collision penalty of -100,000 instead of -1,000,000. This reduction in magnitude is because the traffic vehicle is typically not considered at fault for a collision. The merge vehicle is usually required to yield for traffic.

Training only occurs on the traffic network when the reactive policy is selected. Reactive policy training occurs at a rate of one-third of the total episodes because training happens using a mixture of three different policy choices (random, constant, and reactive). Therefore, the traffic network trains to roughly 350,000 episodes. The ego network learning is relatively stable in the region close to 350,000 episodes in the SAL graph. Therefore, the network has likely learned to react to the other agent similar to the ego network. Reviewing the results shown in Section 5.4.2 *Reactive Policy Results*, this appears to be a reasonable conclusion about the validity of the learning.

5.4 Standard Test Results

It is essential to have a measurement framework to understand performance and compare different approaches in a standard, meaningful way. As mentioned previously, this type of framework or any meaningful way to report results is missing from the currently published merging work. In the Q-learning approach, Table 4 compares collision results based on differential starting positions between the ego and traffic vehicles. Table 4 summarizes the

performance of different policies (reactive, constant, and random) and single-action or joint-action learning. All values in the table are for a 100 m goal position. The DRL approach shows no collisions at the same 100 m goal position. These results are a significant improvement over the Q-learning approach because there were still many collisions at 100 m, despite the different combinations of policies and learning types.

An additional dimension measures the performance of the DRL approach: goal position. Figure 28 shows an example of this expanded evaluation framework. The framework still uses the same starting position differential along the vertical axis, but it adds the goal position along the horizontal axis. The enhanced framework enables performance evaluation as the goal position changes. As the vehicle starting positions get closer to the goal, more collisions occur. This framework also allows comparison to ideal performance, as shown in Figure 29. Ideal performance, as presented here, is more commonly known as *ground truth*. This framework creates a new ability to measure merging performance in a standard way and is another contribution of my work. The framework evaluates the performance by policy and against the ground truth.

Testing is performed separately on all three policies: constant, random, and reactive. All training uses an equal mixture of the three policies, but testing is conducted on each policy individually. The results are presented in tabular form using the framework created for evaluation. In addition, tables of ideal values (considered the ground truth) are also calculated and shown in the evaluation framework to compare the algorithm's performance to the most ideal scenario physically possible. Joint-action-learner training and testing are also conducted. Supplementary videos are available that show the merge vehicle merging against traffic using all three policies and different initial positions.

5.4.1 Constant Policy Results

The constant policy is where traffic maintains its speed throughout each episode. Figure 28 shows the constant policy results. The ego vehicle chooses its action using the state values for the closing gap, closing speed, time to merge position, relative traffic vehicle position, and traffic vehicle action. This scenario represents a merge scene where the vehicle in the traffic lane maintains its steady-state velocity regardless of the behavior or proximity of the vehicle in the on-ramp lane. Figure 29 shows the ideal case, where the action values are at the physical limits of their range. When the action values are at their limits, the merge vehicle cannot achieve greater acceleration or deceleration when trying to avoid a collision. This table shows the best performance achievable with the policy where the traffic speed is constant. It is the ground truth.

Each cell of the tables represents the average collision at the position differential versus goal combination. Because the policy is deterministic and does not change as the testing instance is re-ran, the results become binary – either the collision occurs at that combination, or it does not. The results are almost identical when comparing the two tables. The only exception is at an initial ego position of zero (with respect to the traffic vehicle starting position) and goal positions of 40 and 50 meters.

Average Collision %, Ego Considers Both Actions, Constant Action Traffic												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	0	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	50%
	1	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%
	2	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%
	3	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**53%**	**47%**	**29%**	**18%**	**6%**	**0%**	**0%**	**0%**	**0%**	**0%**	**15.3%**

Figure 28: Results matrix for a scenario where the traffic vehicle maintains a constant speed throughout the episode. In this evaluation, the ego vehicle considers the action of the traffic vehicle in addition to the state.

Average Collision % with Ideal Constant Action												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	1	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%
	2	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%
	3	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**53%**	**47%**	**29%**	**12%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**14.1%**

Figure 29: Ideal scenario for minimum collisions using constant action policy. Calculated using extreme acceleration values for the ego vehicle as the traffic vehicle speed is constant at 31.29 m/s (70 MPH) throughout the episode.

The results show that the network's performance is nearly ideal against the constant policy ground truth shown in Figure 29. Furthermore, there is significant performance improvement when comparing the DRL DDPG approach to the Q-learning approach (Table 4). The Q-learning results in Table 4 show collisions in most of the values of the ego starting position. These values were all recorded using a Goal Position of 100 meters. In comparison, there are no collisions throughout the entire range of initial ego positions using the DRL approach. This performance with zero collisions continues until the goal position reaches 50 m, a drastic performance improvement from the Q-learning approach.

5.4.2 Reactive Policy Results

In the reactive policy, the ego and traffic vehicles react to one another to avoid a collision. The ego and in-lane traffic vehicles choose actions from their own DRL network. The ego vehicle uses the same network training from the constant policy in Section 5.4.1. Section 5.3 describes the network for the traffic vehicle reactive policy. The reactive policy simulates two vehicles, each aware of the other's closing speed and closing gap, each with significant interest in avoiding a collision.

The collision occurrences are low, as observed in the results table shown in Figure 30. They are limited to a group in red that ranges from -3 m to 4 m on the vertical axis and 10 m to 30 m on the horizontal axis. Compared to the ideal case, Figure 31 again shows that collisions are unavoidable when the two vehicles are close to each other and the goal position. The results are binary: either 0% or 100% because the policy is deterministic. The trained network consistently repeats the results identically with every run.

Average Collision %, Ego Considers Both Actions, Responsive Action Traffic												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-3	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**47%**	**35%**	**12%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**9.4%**

Figure 30: Test results using agents that are both reactive to each other. The collision tables show average collisions comparing the ego vehicle's starting position with respect to the traffic vehicle and the starting distance away from the goal position. In this responsive case, both vehicles take actions in response to one another. Each agent trains with its own unique DRL network. Both vehicles take action to avoid collisions with one another.

Average Collision % with Ideal Responsive Action												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-3	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	0	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**47%**	**35%**	**6%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**8.8%**

Figure 31: Manually calculated collisions using the extreme values of vehicle actions that produce the minimum collisions. This is the ideal case where both vehicles take extreme actions to avoid collisions. However, it is important to note that some collisions are unavoidable due to the physical limitations of the road scene and vehicle parameters.

Comparing the ideal case to the results obtained using the trained networks, they are almost the same. The only difference is that the results from the trained network produce a collision at 0 m for the starting position at 30 m away from the goal position. Otherwise, the results are identical. These results show nearly perfect performance in avoiding collisions, especially compared to the Q-learning results. For example, in the Q-learning results at 100 m from the goal position, there are collisions throughout the entire range of initial positions. However, there are zero collisions in the DRL approach – a dramatic improvement.

5.4.3 Random Policy Results

The random policy is when the traffic vehicle takes a random choice at every step of the episode, every 0.1 seconds. The action will select a random value from the entire range of the acceleration values, anywhere from -5 m/s^2 to 4 m/s^2. This scenario is improbable because a driver would change from heavy braking to heavy acceleration ten times every second. However, the training includes the policy to represent an extreme edge case. It trains the agent to deal with a difficult opponent.

The results in Figure 32 show that the random policy results in more instances of collisions and more collisions on average than the constant or reactive cases. Despite this, there is still a significant improvement over the Q-learning case. There are again no collisions at 100m away from the goal position, but there were many instances of collisions at 100m away for the Q-learning results. The figure shows average collisions from 30 separate episodes for each combination of goal and initial positions. Overall, the performance is still impressive because there are no collisions at a 50 m or greater goal position. And there are no collisions when initial positions are 5 m away or greater.

Average Collision %, Ego Considers Both Actions, Random Action Traffic												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-3	100%	97%	3%	0%	0%	0%	0%	0%	0%	0%	20%
	-2	100%	100%	33%	0%	0%	0%	0%	0%	0%	0%	23%
	-1	100%	100%	97%	17%	0%	0%	0%	0%	0%	0%	31%
	0	100%	100%	100%	100%	47%	0%	0%	0%	0%	0%	45%
	1	100%	100%	100%	83%	7%	0%	0%	0%	0%	0%	39%
	2	100%	100%	100%	43%	3%	0%	0%	0%	0%	0%	35%
	3	100%	100%	60%	0%	0%	0%	0%	0%	0%	0%	26%
	4	100%	100%	10%	0%	0%	0%	0%	0%	0%	0%	21%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**53%**	**47%**	**30%**	**14%**	**3%**	**0%**	**0%**	**0%**	**0%**	**0%**	**14.7%**

Figure 32: Case where the ego merge vehicle reacts to a randomly acting traffic agent. The traffic vehicle chooses a random acceleration from -5 m/s^2 to 4 m/s^2 at each simulation time step. Each cell represents 30 episodes run at that combination of the initial position and goal position.

5.4.4 Results for Single-Action Learner

When the single-action learner is trained or selects an action, it does not consider the action of the traffic vehicle. Instead, it considers the same state values as the joint-action-learner, like closing gap and closing speed, but it does not include the action value like the joint-action-learner. Figure 33 to Figure 35 are results tables for the single-action learner.

There is performance improvement for the JAL case when comparing the overall average collisions from the tables. It improves from 15.9% to 15.3% for the constant policy, 10% to 9.4% for the reactive policy, and 15.7% to 14.7% for the random policy. The change is relevant because fewer collisions should produce safer driving. However, the overall improvement is relatively small. Furthermore, when comparing the individual values, there are more collisions in all three SAL policies too. Therefore, using the acceleration value in the method seems advantageous, even though the performance gains are somewhat modest.

Average Collision %, Ego Considers Self Action Only, Constant Action Traffic												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-2	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	0	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	50%
	1	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%
	2	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%
	3	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**53%**	**47%**	**35%**	**18%**	**6%**	**0%**	**0%**	**0%**	**0%**	**0%**	**15.9%**

Figure 33: Single-action learner where the ego neural network only considers the state values and not the action of the traffic vehicle. Traffic has a constant zero acceleration action throughout each episode.

Average Collision %, Ego Considers Self Action Only, Responsive Action Traffic												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-3	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	-1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%
	2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%
	4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**53%**	**35%**	**12%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**10.0%**

Figure 34: Response traffic policy. The ego and traffic vehicles use the same but independent network architecture. Neither considers the action of the other during training or action selection.

Average Collision %, Ego Considers Self Action Only, Random Action Traffic												
		Goal Position										
		10	20	30	40	50	60	70	80	90	100	Total
Ego Initial Position	-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-4	100%	27%	0%	0%	0%	0%	0%	0%	0%	0%	13%
	-3	100%	100%	20%	0%	0%	0%	0%	0%	0%	0%	22%
	-2	100%	100%	97%	3%	0%	0%	0%	0%	0%	0%	30%
	-1	100%	100%	100%	47%	3%	0%	0%	0%	0%	0%	35%
	0	100%	100%	100%	97%	57%	3%	0%	0%	0%	0%	46%
	1	100%	100%	100%	97%	7%	0%	0%	0%	0%	0%	40%
	2	100%	100%	100%	47%	0%	0%	0%	0%	0%	0%	35%
	3	100%	100%	70%	0%	0%	0%	0%	0%	0%	0%	27%
	4	100%	100%	3%	0%	0%	0%	0%	0%	0%	0%	20%
	5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Total	**53%**	**49%**	**35%**	**17%**	**4%**	**0%**	**0%**	**0%**	**0%**	**0%**	**15.7%**

Figure 35: Random action policy for the traffic vehicle. The traffic vehicle selects an action from within the entire range of acceleration values at each 100 ms time step of each episode.

5.5 Algorithmic Training Performance Evaluation

For scenes with more vehicles, there is an improved performance evaluation method. It is a more systematic approach that summarizes the data saved in regular intervals by the Python algorithm. The same algorithmic approach is applied to the two-vehicle simulation to compare against the manual approach used to find the best network for the standard test results. Table 8 shows the results using this new algorithmic best-network search approach on the two-vehicle scene. Comparing the results to the previous visual search approach from Section 5.2 shows that better-performing networks exist, even though the results from the initial method were excellent and nearly ideal in general. However, the better performing networks are relatively similar in total collision performance as well as acceleration/deceleration occurrences and averages. Simulation tests in the following chapters use the improved method to find the best network. This algorithmic method also gives much more detail into the results and a basis for comparison.

Table 8: Improved training performance evaluation. The blue highlighted row at the bottom of the table is the network value chosen by observation of Figure 27, but many saved networks with lower total collisions exist.

Episodes	Merge Deceleration Average	Merge Acceleration Average	Merge Deceleration Occurrence Average	Merge Maintain Occurrence Average	Merge Acceleration Occurrence Average	Total Collisions
340000	-1.81	2.03	0.502	0	0.498	752
380000	-1.82	1.96	0.493	0	0.507	755
360000	-1.89	2.01	0.507	0	0.493	757
1230000	-2.54	1.58	0.531	0	0.469	758
350000	-1.87	2.02	0.499	0	0.501	759
410000	-2.13	1.91	0.515	0	0.485	759
330000	-1.78	1.97	0.508	0	0.492	763
1290000	-2.58	1.54	0.545	0	0.455	763
400000	-2.02	1.94	0.501	0	0.499	764
1250000	-2.53	1.62	0.532	0	0.468	765
1310000	-2.58	1.57	0.546	0	0.454	765
1260000	-2.49	1.63	0.532	0	0.468	766
440000	-2.18	1.86	0.501	0	0.499	768
1280000	-2.53	1.59	0.533	0	0.467	769
1300000	-2.60	1.59	0.551	0	0.449	770
1170000	-1.41	2.27	0.398	0	0.602	771
1220000	-2.49	1.57	0.529	0	0.471	771
1270000	-2.55	1.61	0.535	0	0.465	771
1050000	-2.29	1.98	0.492	0	0.508	772
390000	-1.98	1.99	0.498	0	0.502	775
40000	-3.16	1.99	0.611	0	0.389	780
310000	-1.71	2.04	0.499	0	0.501	782
1180000	-1.08	2.27	0.382	0	0.618	782
1240000	-2.46	1.63	0.523	0	0.477	786
980000	-1.28	2.60	0.339	0	0.661	788
990000	-1.00	2.27	0.383	0	0.617	788
420000	-2.25	1.88	0.524	0	0.476	790
1060000	-2.23	1.97	0.492	0	0.508	790
1140000	-2.12	1.99	0.486	0	0.514	792
1210000	-2.43	1.62	0.523	0	0.477	794
1160000	-1.73	2.35	0.413	0	0.587	795
1000000	-2.19	1.97	0.493	0	0.507	796
1010000	-1.59	2.85	0.326	0	0.674	797
450000	-2.16	1.86	0.501	0	0.499	798
320000	-1.80	1.98	0.506	0	0.494	804
1030000	-2.25	1.98	0.494	0	0.507	804
1040000	-2.27	2.00	0.492	0	0.508	804

The improved evaluation method was developed in MATLAB. It reads in each of the .csv data files saved at 10K increments during network training. Table 8 shows the data that is summarized using the MATLAB script. The *Total Collisions* column is the count of all the

collisions in the .csv file at each training increment. The number of episodes is encoded in the filename, to the left of the ".csv" and the right of the last underscore character. For example, the best performing network is shown at the top of the table and has 752 total collisions at 340,000 training episodes. The best network selected by the previous method (for the SAL case) is also shown in the last line of the table (highlighted in blue). It is the network with 1.04×10^6 episodes and 804 total collisions. Comparing the total collisions, it is 6.9% worse than the best performer of the systemic approach.

When reviewing the tabular results of this improved evaluation method, a pattern emerges for the best performing networks. The best performing networks frequently have a roughly balanced 50%-to-50% deceleration to acceleration ratio. The data is in the *Merge Deceleration Occurrence Average* and *Merge Acceleration Occurrence Averag*e columns, with the value representing the frequency of occurrence for merge deceleration within each episode then averaged over the entire set of performance test combinations. The values represent how often the network selects a deceleration action $\left[-5\ \frac{m}{s^2}\ to\ 0\ \frac{m}{s^2}\right)$ or an acceleration action $\left(0\ \frac{m}{s^2}\ to\ 4\ \frac{m}{s^2}\right]$.

Figure 36 shows a multi-dimensional plot of the full performance evaluation output table. The figure is sorted by best to worst collision performance with the red line showing the number of collisions in total for each standard test set. Lower collisions indicate better performance.

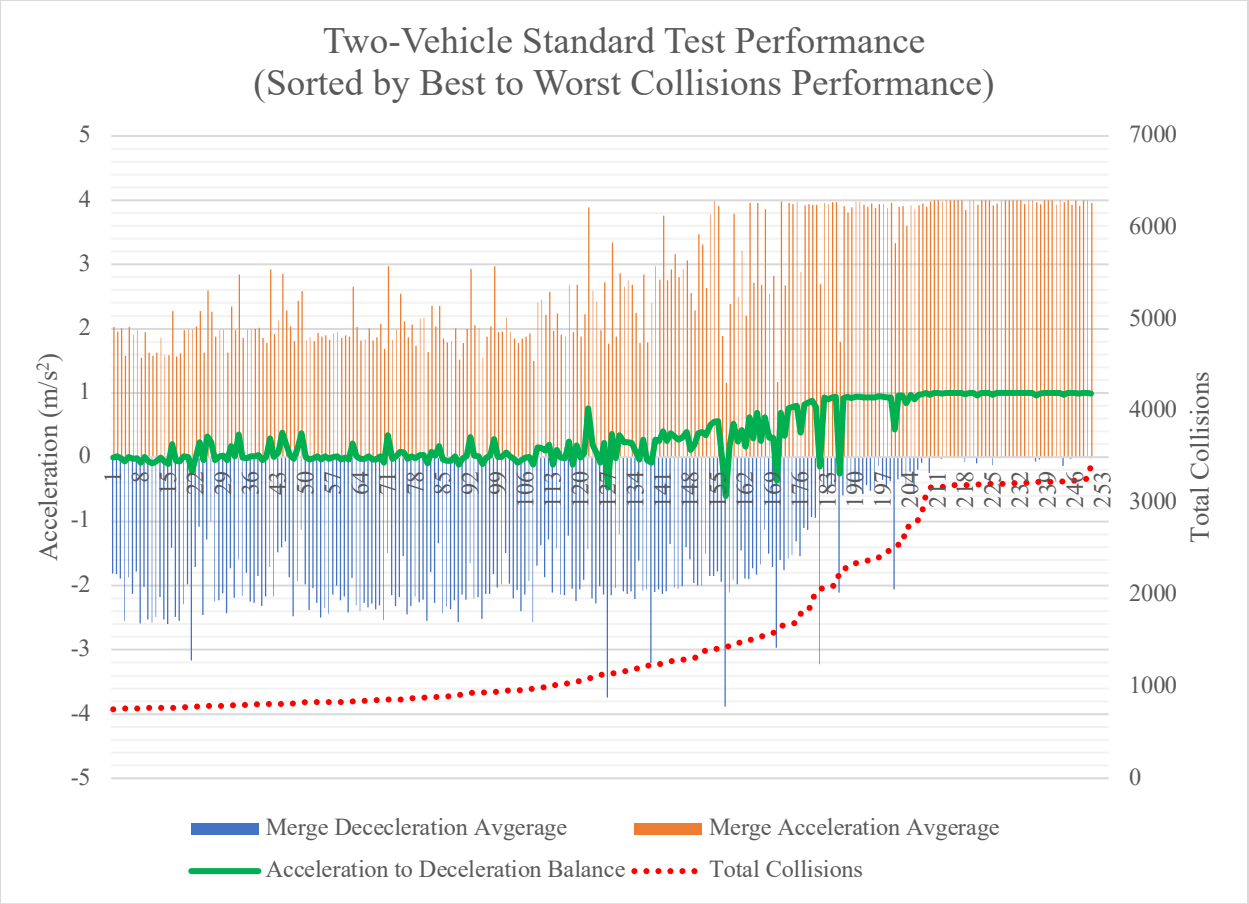


Figure 36: Multi-dimension plot of the standard test performance of the two-vehicle scenario. Performance testing occurs every 10K episodes for a total of 250 standard tests over the range of 2.5M episodes. Data is sorted by best to worst collision performance. Blue and orange bars indicate average deceleration and acceleration for each test instance. The green line is the difference between the acceleration and deceleration occurrence where negative values indicate greater deceleration than acceleration. The red line is on the secondary axis to the right and represents the total collisions for each blue-orange bar pair where a test is performed.

There are some common trends that appear. The ratio of acceleration to deceleration for the best networks is frequently a balanced ratio of equal occurrence of deceleration action to acceleration action. There are some peaks in the data, but overall, the balanced 50%-50% occurrence is most common for the better performing tests. As the values trend more towards greater acceleration occurrence, the collisions increase. Figure 63 in the appendix shows a tabular view of the deceleration occurrence in the starting position-goal format. The 50%-50% balance is a result of Proposition III where a lagging vehicle should decelerate and a leading vehicle should accelerate to avoid a collision.

The merge deceleration and acceleration magnitudes tend to be roughly equal at 2 m/s^2 for the best performing tests. There is certainly variation in the averages, with more variation occurring within the deceleration average, but there is clearly a general range of values which are between -1.5 and -2.5 m/s^2 for the best performing deceleration averages. As the average deceleration magnitude values decrease, the averages of the acceleration values increase, and the total collisions also increase. Therefore, there appears to be correlation between the values.

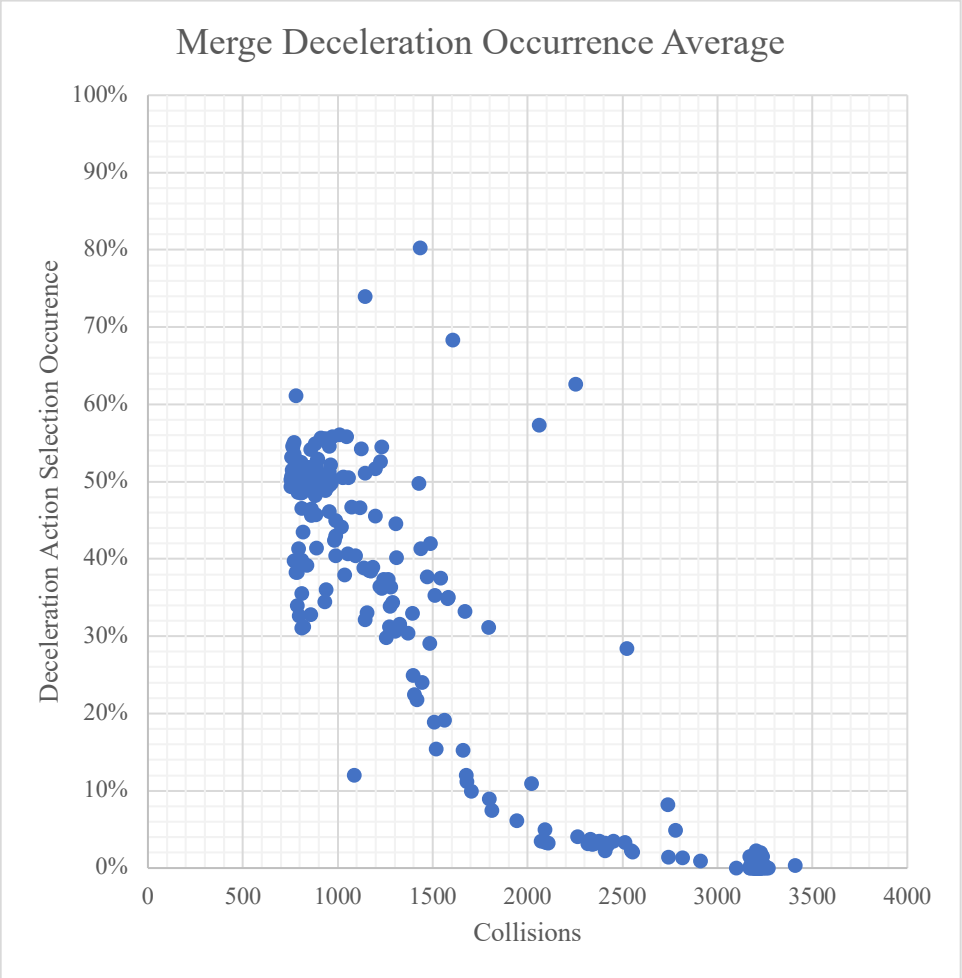


Figure 37: Two-vehicle deceleration occurrence and collisions correlation. As deceleration occurrence decreases, collisions increase. However, the best performance has a more scattered grouping of data points, suggesting that the best performance is influenced by more than just deceleration action selection.

Investigating the correlation further, Figure 37 shows a plot comparing the occurrence of deceleration action selection to total collisions. A clear pattern emerges that shows lower

occurrence of deceleration correlates with increased collisions. However, the pattern becomes less distinct for the highest performance. There is a relatively dense group of data points right around 50%, but the highest performance shows a grouping of values ranging from 30% to 60% instead of tighter-grouped trend as the collisions get higher. This suggests that the best performance is based on more than just deceleration action occurrence. For reference, Figure 62 in Appendix A.1 shows the acceleration occurrence. It is the inverse of the deceleration plot.[7]

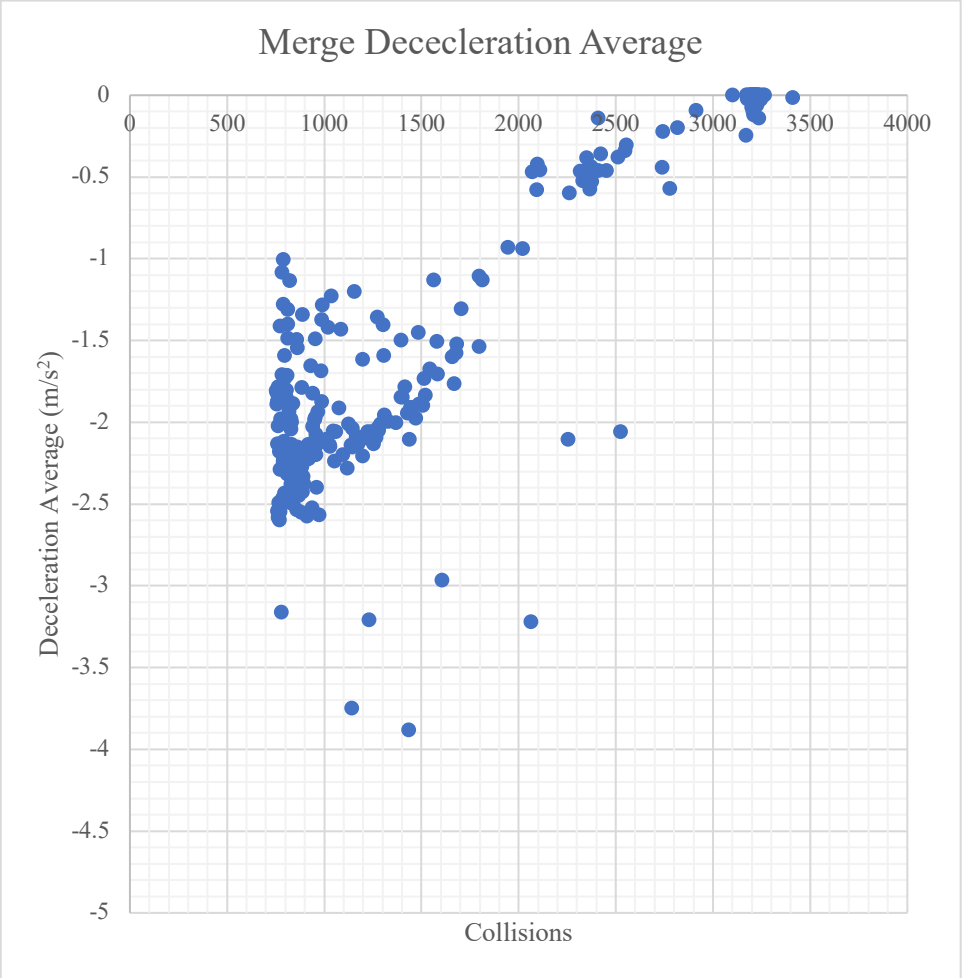


Figure 38: Ego merge vehicle average deceleration action value.

[7] Only acceleration and deceleration actions are chosen by the algorithm, so deceleration occurrence and acceleration occurrence percentages sum to 100%.

Figure 38 shows the average deceleration value selected from within the range bounds of $\left[-5\ \frac{m}{s^2}\ to\ 0\ \frac{m}{s^2}\right)$. The pattern formed by the data points looks almost like a symmetrical mirror about the horizontal axis. There also appears to be a pattern of a linear or slightly curving pattern of increasing collisions as the deceleration magnitude decreases. This is similar to the deceleration occurrence values in Figure 37. Likewise, the pattern becomes less distinct as with the best performing values. This too suggests that a simple single-variable function is insufficient to control merging performance and avoiding collisions.

Average of Acceleration_Average Ego Initial Position	Goal Position 10	20	30	40	50	60	70	80	90	100	Grand Total
-20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	4.00	4.00	4.00	4.00	4.00	3.98	3.73	3.41	3.31	3.01	3.74
1	4.00	4.00	4.00	4.00	4.00	3.99	3.84	3.64	3.36	3.11	3.79
2	4.00	4.00	4.00	4.00	4.00	3.99	3.86	3.67	3.41	3.16	3.81
3	4.00	4.00	4.00	4.00	4.00	3.99	3.89	3.68	3.45	3.20	3.82
4	4.00	4.00	4.00	4.00	4.00	4.00	3.90	3.72	3.50	3.26	3.84
5	4.00	4.00	4.00	4.00	4.00	4.00	3.93	3.78	3.56	3.34	3.86
6	4.00	4.00	4.00	4.00	4.00	4.00	3.95	3.83	3.62	3.40	3.88
7	4.00	4.00	4.00	4.00	4.00	4.00	3.96	3.87	3.70	3.48	3.90
8	4.00	4.00	4.00	4.00	4.00	4.00	3.97	3.90	3.77	3.56	3.92
9	4.00	4.00	4.00	4.00	4.00	4.00	3.98	3.93	3.82	3.66	3.94
10	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.95	3.86	3.74	3.95
15	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.97	3.92	3.99
20	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.95	3.82	3.98
Grand Total	**2.08**	**2.08**	**2.08**	**2.08**	**2.08**	**2.08**	**2.04**	**1.97**	**1.89**	**1.79**	**2.02**

Figure 39: Two-vehicle Ego acceleration average values by start versus goal position.

Better performing networks learn to accelerate when Ego is leading and to decelerate when Ego is lagging, therefore half of the acceleration/deceleration values in the table are zero, so the average becomes roughly half. It is also because of Proposition III. Figure 39 shows the

detailed average acceleration values in the position-goal format for the specific selected 1.04 x 10^6 episode network. Figure 40 shows the deceleration values. Acceleration occurs almost exclusively for leading Ego vehicles which is almost exactly half of the test values. Similarly, deceleration occurs for the remaining half. Generally, the Ego DRL network has learned to take action values at the extreme end of the range of acceleration or deceleration. Since the other half of the values are zero, the overall average is roughly half of the range, so about 2 m/s^2 for acceleration which correlates to the observed average for the best performing networks.

Average of Deceleration_Average Ego Initial Position	Goal Position 10	20	30	40	50	60	70	80	90	100	Grand Total
-20	-4.80	-4.87	-4.93	-4.91	-4.94	-4.97	-4.98	-4.99	-4.99	-5.00	-4.94
-15	-4.32	-4.49	-4.65	-4.77	-4.85	-4.92	-4.96	-4.98	-4.99	-4.99	-4.79
-10	-3.50	-3.91	-4.22	-4.50	-4.72	-4.83	-4.92	-4.96	-4.98	-4.99	-4.55
-9	-3.26	-3.87	-4.18	-4.35	-4.66	-4.78	-4.89	-4.95	-4.97	-4.98	-4.49
-8	-3.18	-3.67	-4.02	-4.32	-4.58	-4.77	-4.89	-4.94	-4.97	-4.98	-4.43
-7	-2.96	-3.47	-3.87	-4.19	-4.49	-4.74	-4.84	-4.92	-4.96	-4.98	-4.34
-6	-2.80	-3.41	-3.83	-4.15	-4.38	-4.70	-4.83	-4.93	-4.97	-4.98	-4.30
-5	-2.80	-3.22	-3.67	-4.02	-4.32	-4.58	-4.80	-4.89	-4.95	-4.97	-4.22
-4	-2.67	-3.03	-3.63	-3.99	-4.23	-4.51	-4.78	-4.88	-4.94	-4.98	-4.16
-3	-2.54	-3.00	-3.46	-3.86	-4.19	-4.45	-4.73	-4.86	-4.92	-4.97	-4.10
-2	-2.54	-2.88	-3.41	-3.83	-4.08	-4.37	-4.66	-4.85	-4.93	-4.97	-4.05
-1	-2.40	-2.88	-3.23	-3.69	-4.05	-4.31	-4.56	-4.80	-4.90	-4.95	-3.98
0	0.00	0.00	0.00	0.00	0.00	0.00	-0.49	-0.50	0.00	0.00	-0.10
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grand Total	**-1.51**	**-1.71**	**-1.88**	**-2.02**	**-2.14**	**-2.24**	**-2.33**	**-2.38**	**-2.38**	**-2.39**	**-2.10**

Figure 40: Two-vehicle Ego deceleration average values by start position and goal position.

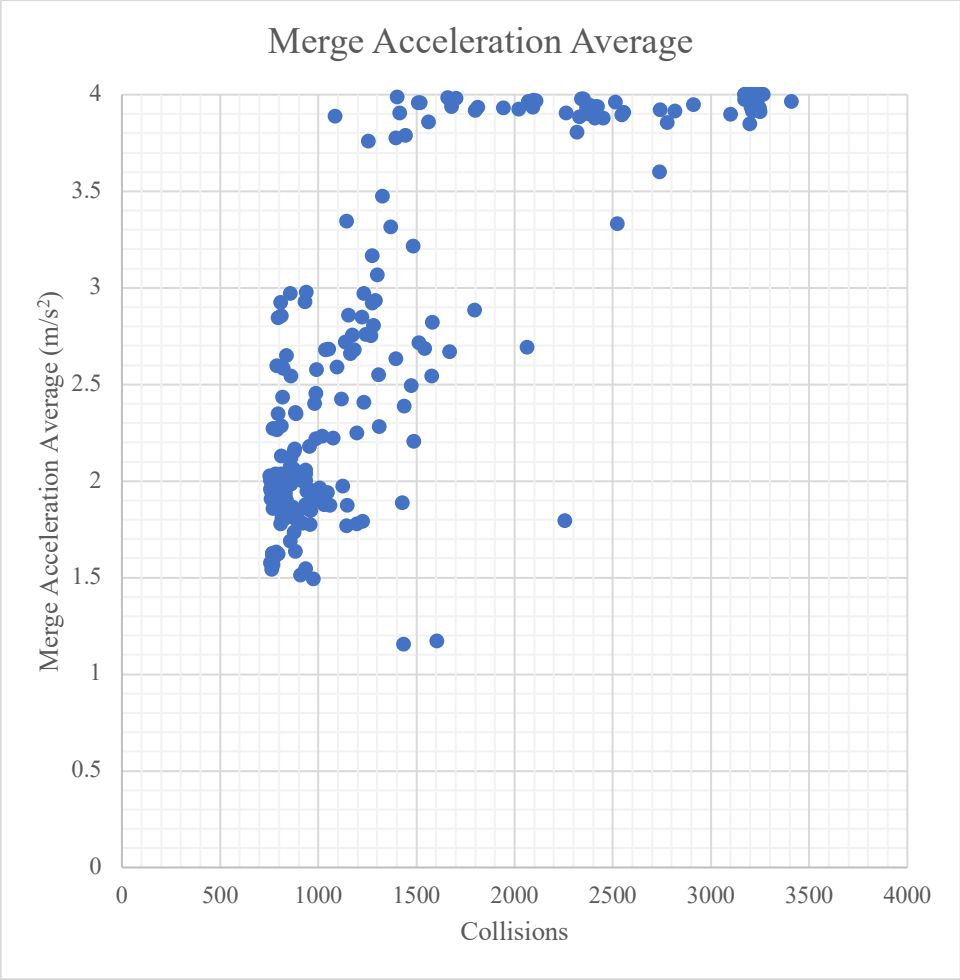


Figure 41: Two-vehicle scenario Ego Merge vehicle average acceleration values compared to collisions.

Figure 41 shows the Ego Merge vehicle acceleration average. The data plot is very similar to the deceleration average shown in Figure 38, but the values are different. The pattern shows that as acceleration magnitude increases, so do collisions. This is different than the deceleration average that shows as magnitude decreases, collisions increase. This seems to suggest that a more defensive driving style leads to lower collisions. However, Figure 38 also suggests that a style too defensive results in greater collisions. Considering the data shown in these plots suggests that merging is not a simply defined function – a concept that should seem familiar to most experienced drivers.

5.6 DRL Discussion

Overall, the DRL approach shows significant improvement over the Q-learning approach. In the Q-learning approach, collisions were frequent at 100 m away from the goal position, whereas in the DRL approach, collisions at 100 m did not occur. Overall, within the DRL policies, collisions only occur when the goal position is 60 m or closer. Some policies perform even better. Also notable is that collisions do not happen in the DRL approach when the vehicles have a starting position of 5 m or greater away from each other. In physical terms, because the two vehicles are each 5 m long, this means that when the ego vehicle is aligned right behind or right in front of the traffic vehicle, it can "pull away" and avoid a collision.

The performance of the joint-action-learner is nearly identical to the ideal cases when comparing the results tables against the ideal cases. It is essential to understand that collisions are unavoidable if the vehicles are too close to each other or too close to the goal position and are limited in acceleration or deceleration, as the simulations presented. There will always be limitations in vehicle proximity as well as the acceleration/deceleration performance of vehicles. Therefore, collisions are unavoidable within certain circumstances.

The two-vehicle road scene simplifies the merge scene to understand the fundamental behavior of merging vehicles. Studying a two-vehicle road scene is essential before scaling to multiple vehicles and increasing the complexity to a busy road scene. The final research phase can be completed using this foundation, fundamental understanding, and simulation results: on-ramp merging in a multiple-vehicle road scene. In the final research phase, additional vehicles are added one by one. First, one is added to the traffic lane, then one or more is added to the merge lane. Chapter 6 through Chapter 7 detail this work.

Chapter 6 Three-Vehicle Simulation

The two-vehicle simulation showed impressive results that were nearly ideal. However, a full merge scene could include more than just two vehicles. The next progressive step of scaling up the simulation is to add another vehicle. The Three-Vehicle Simulation (Figure 42) consists of one vehicle in the merge lane and two in the traffic lane. Compared to the two-vehicle variant, it uses the same DRL algorithm but adds one vehicle to the traffic lane. The ego merge vehicle must now contend with front *and* rear traffic vehicles and avoid collisions. The revised approach adds additional parameters, uses a larger state set, and develops a strategy for contending with the extra vehicle. The collision-avoiding performance of the three-vehicle variant is also nearly ideal, like the two-vehicle variant.

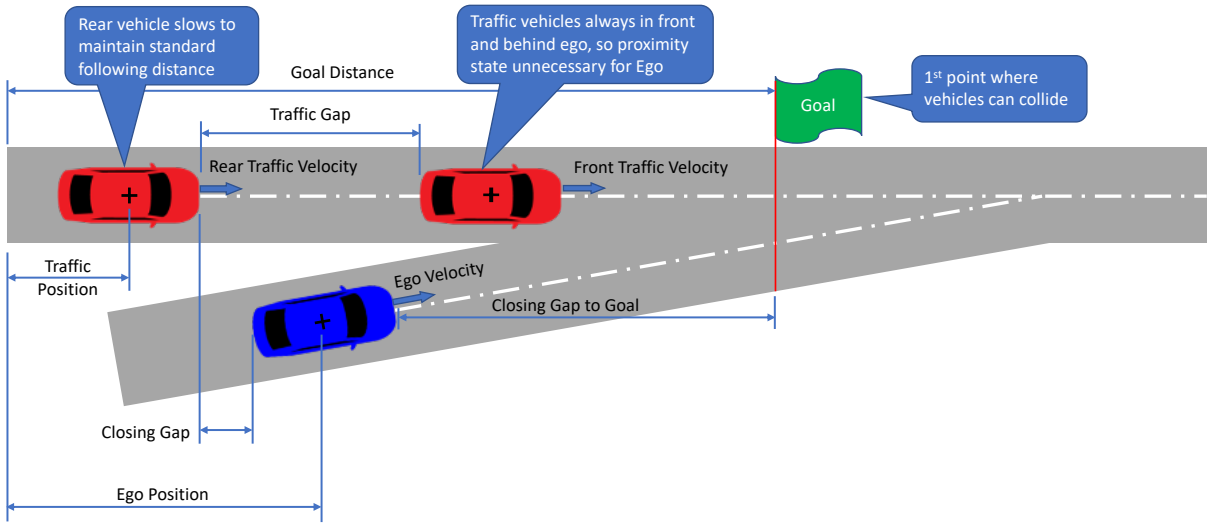


Figure 42: Three-vehicle simulation. One ego merge vehicle (blue) and two traffic vehicles (red). The fundamental setup is the same as the two-vehicle scene, but the complexity increases significantly for the set of state variables and lines of simulator code.

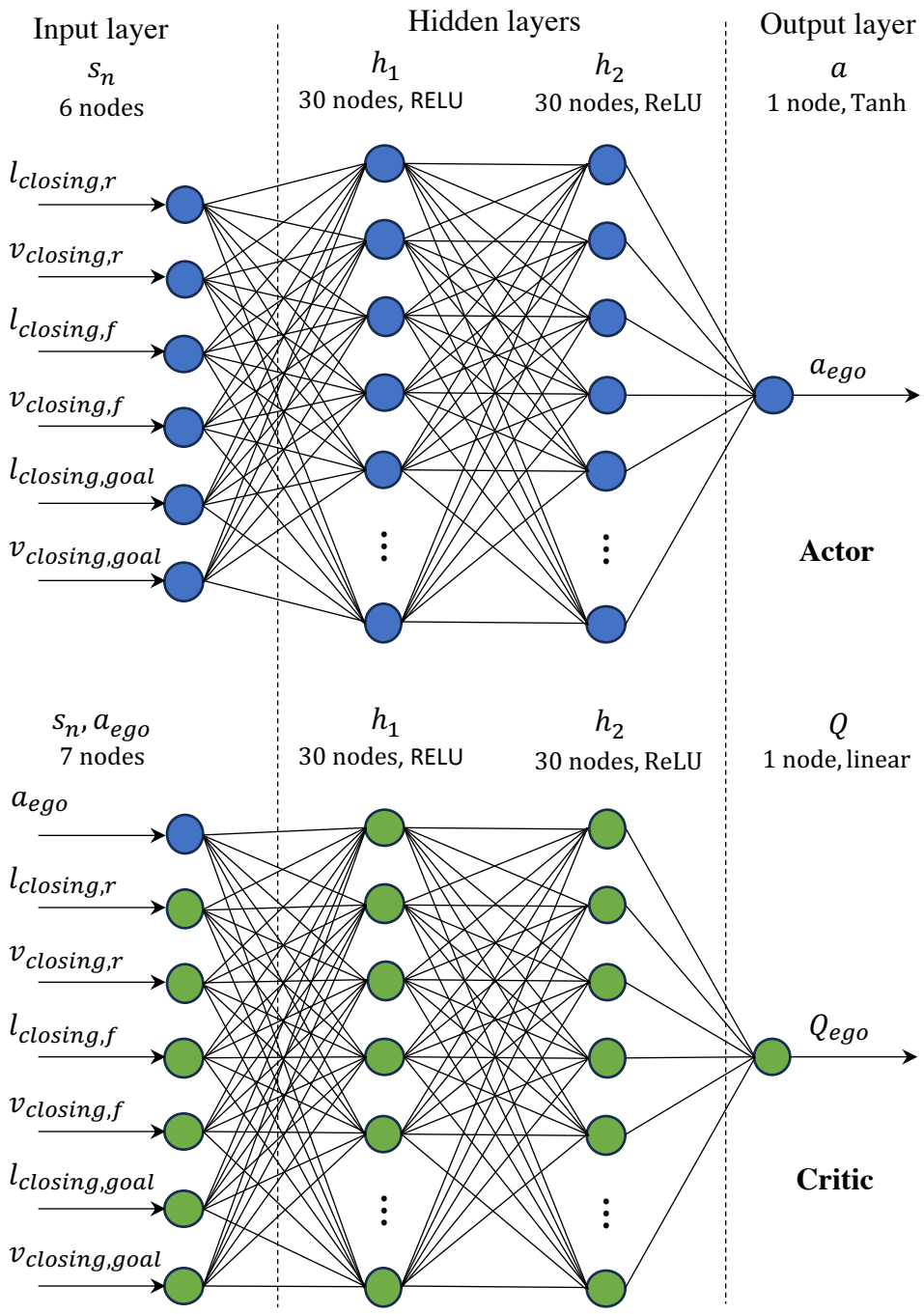


Figure 43: Three-vehicle DRL network diagram for the merge vehicle.

6.1 State Set

The set of states changes considerably from the two-vehicle variant. The two-vehicle variant only considered five states and the state set was the same for the traffic and merge

vehicles (see Table 7). Now, the three-vehicle variant considers nine states with different combinations of state variables used in the merge vehicle and traffic vehicle (see Table 9 for details). Figure 43 shows the DRL network diagram for the merge vehicle, including the states. Figure 64 in the appendix shows the DRL network diagram for the traffic vehicle. The difference in states results in an extensive rewrite of the algorithm and Python code.

Table 9: Three-vehicle simulation state parameters. Vehicles use different sets of state variables. Green check marks indicate state variable usage, and a red x indicates those not used.

Vehicle State Variables	Used by Merge	Used by Traffic	Range	Units
Closing Gap to Vehicle in Next Lane	✓	✓	[-2.5, 30]	m
Relative Closing Speed to Vehicle in Next Lane	✓	✓	[-10, 10]	m/s
Closing Gap to Front Traffic	✓	✗	[-2.5, 30]	m
Relative Closing Speed to Front Traffic	✓	✗	[-10, 10]	m/s
Closing Gap to Goal	✓	✗	[-160, 150]	m
Closing Velocity to Goal	✓	✗	[0, 40]	m/s
Closing Time from Current to Front Vehicle (TIV)	✗	✓	[0, 2.5]	s
Time to Goal Position	✗	✓	[0, 3]	s
Proximity to ego (*behind or in front of*)	✗	✓	{-1, 1}	*unitless*

The state set for the ego merge vehicle consists of a pair of variables (closing gap and closing speed), repeated several times for different objects: the traffic vehicle behind, the traffic vehicle in front, and the goal position. These three pairs of variables comprise a total state set size of six variables for the ego merge vehicle. Compared to the two-vehicle setup (see Table 7), the state set only increases by one variable. However, the evaluated states have changed significantly from the two-vehicle states.

The proximity variable is no longer needed for ego (the variable that indicates whether the traffic vehicle is in front or behind). The states always account for one traffic vehicle in the front and one in the rear. The speed and gap states default to a max gap of 100 m at zero closing velocity when a front or rear vehicle does not exist. The default 100 m max gap for a missing vehicle happens in training and testing.

The reactive policy traffic vehicle uses a set of state parameters that are mostly different from those used for the ego merge vehicle. However, most traffic state parameters are the same as the two-vehicle variant. In fact, the traffic vehicles use the same set of parameters as the two-vehicle variant, except for adding a TIV parameter. The TIV is the gap from the rear to the front vehicle, divided by the velocity of the rear vehicle. The front vehicle has no vehicle in front of it, so the gap is 100 m. Therefore, the TIV for the front vehicle is 100 m divided by its own velocity.

There were many trials with different state variables during the development, debugging, and verification of the three-vehicle simulation. Ultimately the states described here were chosen because the performance in standardized testing (like Section 6.5) produced good results. The simulation performance observed during rendering is often a factor in directionally understanding how much the behavior mimicked real-world drivers. Specifically, the motion of the trained vehicles is observed by viewing the rendering to see if abnormal driver behavior is present like close cut-ins or tailgating. However, render performance is a secondary consideration after the testing performance.

6.2 Three-Vehicle Simulator Functionality

The three-vehicle simulator Python code is mostly a rewrite of the two-vehicle simulator. The DDPG DRL code remains the same, but the main program, environment, and functions

within the environment changed quite extensively. Most of the functions remain the same, but within each function, the variables change to allow for the additional traffic vehicle. The merge vehicle now always considers a front and rear traffic vehicle, even if the vehicles are outside of the 100 m range that is typical to most state-of-the-art sensors today. For example, if a traffic vehicle is 250 meters away from the merge vehicle, the gap distance considered by the merge vehicle is capped at the 100 m maximum.

The basis for the environment is a standard OpenAI Gym *classic control* environment: *Mountain Car Continuous*. Mountain Car is a common AI and control system example, first introduced by Andrew Moore [85] and is often used to study the performance of physical systems. In Mountain Car, an underpowered vehicle is in a valley between two mountains and builds momentum to get out of the valley. The goal is to reach the apex of the mountain. The Mountain Car Continuous Gym environment contains basic physical parameters like position speed, force (action), an action space, an observation space, a step function, reset function, and a render function [86]. The entire Mountain Car Continuous environment is about 130 lines of Python code. In comparison, the environment for the three-vehicle simulation, named *Highway Merge*, is about 750 lines of Python code because the sophistication has grown substantially. Maintain Car Continuous only controls a single position variable with a fixed-distance path between the tops of the two mountains. In comparison, the functionality of Highway Merge is increased over Mountain Car Continuous because it has at least two vehicles with variable vehicle lengths, variable distance to goal, variable velocity settings, output logging capability, testing functionality and more. Following is a detailed description of the simulator structure and functionality.

6.2.1 Environment

The three-vehicle Highway Merge environment contains many of the same functions as typical Gym environments, but the capability of the vehicle simulator is different. The Highway Merge environment has the following code blocks and functions: hyperparameter initialization, step function, reset function, states function, MIO function, and render. There are also some smaller sub-functions for initialization and avoiding code duplication. Compared to the Mountain Car environment, the states and MIO functions are new, but the other functions already exist.

Parameter initialization happens within the environment. Initialization of the environment occurs at the beginning of the execution of the overall main algorithm. Table 10 - Table 11 shows the simulation initialization parameters, values, and units. An object class called *self.* stores these variables. The notes column describes each parameter.

Table 10: Simulation initialization parameters for ego merge vehicle and overall simulation

Ego Merge Vehicle Parameter	Value	Units	Notes
min_action	-5	m/s^2	Ego minimum acceleration
max_action	4	m/s^2	Ego maximum acceleration
min_position	-10	m	Ego minimum position
max_position	150	m	Ego maximum position
min_initial_position	-25	m	Ego episode minimum initial position
max_initial_position	50	m	Ego episode maximum initial position
min_speed	20	m/s	Ego minimum allowed speed[8]
max_speed	40	m/s	Ego maximum allowed speed[8]
max_ttp	3	s	Goal time-to-position state maximum
min_closing_gap	-2.5	m	Ego closing gap to traffic state min.
max_closing_gap	30	m	Ego closing gap to traffic state max.
min_closing_speed	-10	m	Ego closing speed to traffic state min.
max_closing_speed	10	m	Ego closing speed to traffic state max.
min_goal_position	25	m	Minimum distance to goal from zero
max_goal_position	150	m	Maximum distance to goal from zero
power	0.1	s	Simulation timestep
length	5	m	Ego length
reward	0	unitless	Self reward initialization to zero

Many parameters for the traffic vehicles are set the same as the ego vehicle, like action, position, and speed limits. However, the additional traffic vehicle adds new parameters to the simulation. For comparison, Table 5: DRL training simulator variables shows the two-vehicle parameter list. Parameters like TIV are new to the simulation for the three-vehicle version.

[8] The minimum and maximum speeds are a limitation of the simulation. It is not trained for very low speeds or exceptionally high speeds. The training speed limits are between 20 and 40 m/s (45 and 89 MPH). The states use relative velocities, so the algorithm could be robust at speeds outside of the limits, but no testing or analysis was performed outside of this range, so the behavior might be different than what is presented here.

Table 11: Simulation initialization parameters for traffic vehicles

Traffic Vehicle Parameter	Value	Units	Notes
traffic_min_action	-5	m/s^2	Traffic minimum acceleration
traffic_max_action	4	m/s^2	Traffic maximum acceleration
traffic_min_position	-10	m	Traffic minimum position
traffic_max_position	150	m	Traffic maximum position
traffic_min_initial_position	-25	m	Traffic episode min. initial position
traffic_max_initial_position	50	m	Traffic episode max. initial position
traffic_min_speed	20	m/s	Traffic minimum allowed speed[8]
traffic_max_speed	40	m/s	Traffic maximum allowed speed[8]
traffic_min_length	1	m	Traffic vehicle minimum length
traffic_max_length	20	m	Traffic vehicle maximum length
traffic_next_max_tiv	5	s	Time in between vehicle state max.
traffic_min_tiv	0.5	s	Time in between vehicle state min.
traffic_min_gap	44	m	Min. gap between traffic vehicles when a new traffic vehicle is generated
traffic_max_gap	100	m	Caps maximum observable distance between traffic vehicles
traffic_max_tiv	2.5	s	Time in between vehicle state max.
traffic_test_tiv	0.8	s	Fixed time in between vehicles at start of simulation test episode

The step function is within the environment. It increments all the variables associated with the time-based stepping of the environment. The *self.* variable class and action values for ego and traffic vehicles are inputs into the step function. Output files are written within the step function. Counters for the agent's actions increment. Standard equations of motion with constant acceleration calculate the position and velocity. Constant acceleration is reasonable because the timesteps are small at 0.1 s. TIV is checked at every increment of the step function. If the TIV is smaller than its limit, the acceleration action is changed to the lowest value of deceleration, -5 m/s^2, to slow the vehicle down. This behavior is like the functionality of an adaptive cruise control system.

The step function calls the states function. The states function calculates the new state values based on the new position and velocity values after the time step increment. The done flag is set if the ego vehicle is at or past the goal position. Next, the step function calculates the penalties using the reward function from Section 4.1.4. When the done flag is set, the step function writes the output values to a file. At the end, the step function returns the vehicle states, rewards, traffic action & reaction, and done flag status.

The reset function is called at the start of each episode. In training mode, it sets random values for positions, velocities, gaps, and traffic reactions within the limit bounds of Table 10 - Table 11. In test mode, the function takes input values for the start position, goal position, reaction, and traffic gap. Test mode uses these inputs along with starting velocities of 70 MPH for all vehicles and a traffic starting position of zero to initialize the parameters needed for the step function. The reset function also calls the state function to calculate the initial states. It resets the counters for the episode, then returns the state values.

The states function is called by both the step and reset functions. The function calculates all the state values for the ego and traffic vehicles, as listed in Table 9. The function performs a series of relatively simple calculations to determine each of the states. Calculated values for the ego vehicle are:

- goal_gap
- velocity
- closing_gap (to_rear_traffic)
- closing_speed (to_rear_traffic)
- closing_gap (to_front_traffic)
- closing_speed (to_front_traffic)

The traffic state is also calculated. As explained in Section 6.1, the traffic state set differs from the ego state set. The calculated traffic states are:

- traffic_relative_posistion
- traffic_closing_gap
- traffic_closing_speed
- traffic_ttp
- traffic_tiv

The traffic states are a two-dimension vector because it stores the states for both the rear and front traffic vehicles. Therefore, the *states* function returns two variables: the ego state vector and the traffic state vector matrix.

The Most Important Object (MIO) function finds the traffic vehicle longitudinally closest to the ego vehicle. It ensures a traffic vehicle is in front of and behind the ego vehicle. It adds a vehicle if one does not exist. The only parameter fed into the function is the *self.* class. The function has branches for both training and testing. During training, vehicles added in front of or behind ego have a randomly generated length, gap, velocity, action policy, and TIV limit. The algorithm selects random values from within the limit bounds defined in the Highway Merge environment, as shown in Table 10 - Table 11. The function loops until it finds, or creates, the traffic vehicles in front of and behind ego. It then deletes any other traffic vehicle from the vector. At the end of the function, it left with only these two front and rear traffic vehicles.

The two-traffic model is a simplification of a real scenario. It ensures that the two traffic vehicles only consider a limited set of variables. Both consider the ego vehicle states, but only the rear traffic vehicle considers another traffic vehicle. The front traffic vehicle assumes another traffic vehicle is in front of it at the max distance of 100 m gap and zero closing velocity. The

full scene simulator in Chapter 7 does not delete any extra traffic vehicles created. However, to study the effects of the evolution, they are omitted from the three-vehicle scene to build maturity levels of the merge model progressively.

The render function is the last function within the Highway Merge environment. The render function builds the lines for the roadway and places a goal flag at the first point in that the ego vehicle can collide with traffic. Figure 44 shows an example of the rendering output. The traffic lane runs horizontally along the screen, and the merge lane angles into the traffic lane, terminating at the merge point. All vehicles run left to right. Render draws the lines of the road, then draws the ego merge vehicle as a blue-filled rectangle. Red, green, or black rectangles represent the traffic vehicles. The color identifies the traffic action policy: red is random, green is reactive, and black is constant. The polices were explained previously in Section 4.1.3. All vehicles are scaled to the length from the simulator with a typical vehicle width of about two meters.

Figure 44 shows a graphical display of vehicle positions, velocities, and ego TTP. This graph is new to the three-vehicle render function. There is one point per vehicle per time step shown in the graphs. The points are color-coded to the vehicles shown on the roads. The upper right of the figure shows the output from the main program.

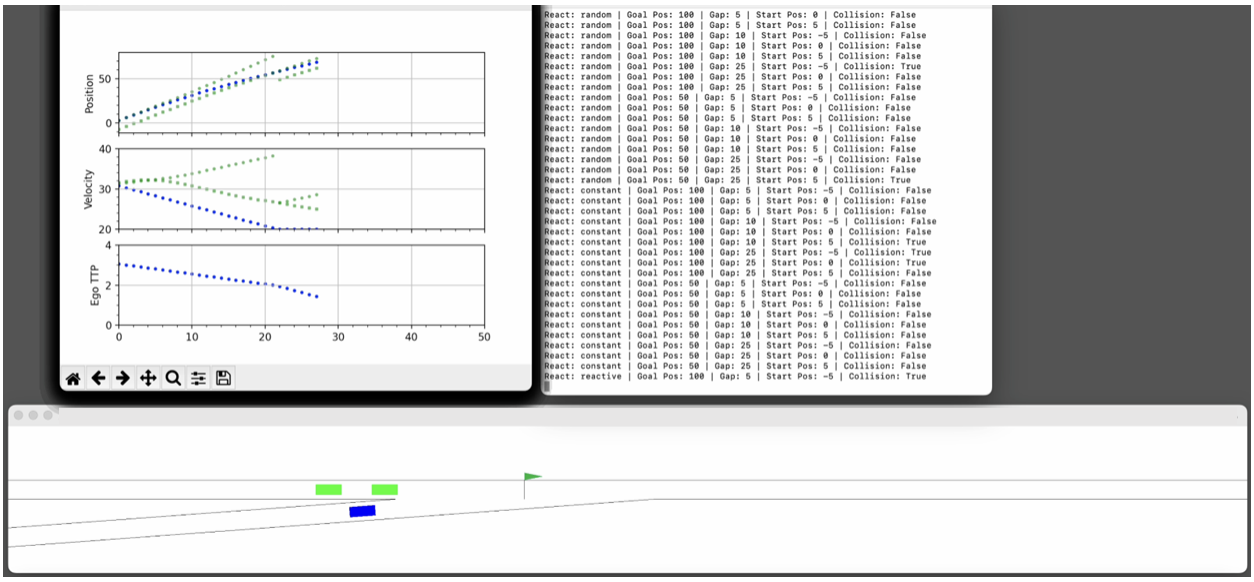


Figure 44: Three-vehicle simulation rending output. At the bottom is the road scene with ego in blue and two reactive policy traffic vehicles in green. At the top left are position, velocity, and ego TTP graphs. At the right is the output from the main program showing the configuration of each episode and whether a collision has occurred.

6.2.2 Main

The main program sets up all the basic functionality for the environment and DRL DDP neural network. It sets the environment name, so it and its functions can be called. Next, it creates the state dimensions and action bounds. It then makes the placeholders for the TensorFlow networks for the DRL Actor and Critic. It also creates saving and restoring capability. Section 4.2 explains the simulator DRL DDPG network. There are two separate DRL networks trained in parallel. One network for the merge vehicle and one for the traffic vehicles. The DRL code to train both vehicle networks is the same, but the training weights are stored separately.

The episode loop is next. The episode count is a variable set before running the program. The Highway Merge environment reset function is called within the episode loop to determine the initial states for ego and the traffic vehicles. A step loop is within the episode loop. The DRL actor function is called using the state values as an input. The output of the actor function is the

action. The step loop calls the Highway Merge environment step function. The new states and reward values obtained as an output from the step function train the DRL networks. Any traffic vehicle with a reactive policy updates the traffic DRL network. The step loop continues until ego reaches the goal position, then the episode ends. The episode loop continues with an environment reset and enters the step loop again. This continues until the episode counter hits the save point or until all episodes are complete.

The TensorFlow checkpoint saves the DRL network at the save point, and the test routine starts. During the test routine, each of the three policies (random, constant, and reactive) cycles through with standard sets of values for initial ego position, goal distances, and traffic vehicle initial gaps. The entire process repeats three times to have multiple samples for the random policy and to provide an opportunity to review multiple runs of the constant and reactive policies are replicates. The number of test cycle repetitions is a parameter set in the initialization section. Output values write to a file. The file is the data to create the standard test tables, like those shown in Section 6.5. Figure 44 in the upper right, shows the test output from the main program in the terminal window. It lists the reaction policy of traffic, the goal position, the initial traffic vehicle gap, ego start position, and true or false for whether a collision has occurred.

6.3 Training

Training the three-vehicle network is fundamentally the same as training the two-vehicle network. The DRL DDPG functions remain the same in the three-vehicle scene. The DDPG DRL network is detailed in Section 4.2. The network architecture and function operation are unchanged. The input layer size automatically adjusts based on the number of state parameters fed into it.

DRL network training occurs at each step of every episode. There are roughly 50 steps on average per episode. The network trains to 2.5 million episodes, but the best performer is at 350K episodes, or about 17.5 million training steps. However, the reward assignment for a collision or successful merge occurs only once per episode. Section 3.3.2 *Reward Function Design* details the reward function. It is the same as the Q-learning and 2-vehicle DRL reward function.

With two vehicles now in the traffic lane, they have a time gap between them, or TIV. TIV is now a variable in the simulator, randomized during training and fixed during testing. The TIV limit is randomly selected from a range of [0.5, 2.5] s at the beginning of each training episode, but during testing it is fixed at 0.8 s. For reference, adaptive cruise control (ACC) systems at highway speeds have at least a 1.6 s minimum time gap . A study of driver behavior shows that most drivers prefer a minimum gap of at least 0.6 seconds, and less than 3% maintain a gap smaller than 0.5 seconds [64]. Hence, the 0.8-second TIV is an aggressive and challenging yet reasonable and realistic test setting.

The action is acceleration for all vehicles, with a range of [-5, 4] m/s^2, just like the two-vehicle simulation. The action variable is what is learned by the DRL network. The training uses a mixture of all three random, constant, and reactive traffic action policies. The two traffic vehicles independently select one of the three action policies at the beginning of each episode, and each traffic vehicle maintains that policy throughout the episode.

The reactive policy of the traffic vehicle also uses DRL to learn to avoid collisions. The reactive policy is one of three choices (reactive, random, or constant) selected at each episode, so the reactive policy is selected about 1/3 of the time. There are two traffic vehicles, each choosing one of the three policies at each episode. Both traffic vehicles train the same DRL network.

Therefore, the reactive traffic DRL network learns at a rate of about 2/3 that of the ego merge vehicle, so there is a mismatch in the amount of learning between the two vehicle networks.

The training performance is reviewed only for the ego merge vehicle, but it is not reviewed for the traffic network. However, traffic vehicle training performance is related to ego vehicle training performance. Poor performance in reactive traffic results in more collisions with the ego merge vehicle, thus resulting in poorer performance. Suppose each of the agent's DRL networks trains independently. In that case, it could be possible to find good performance in each one that might result in a combination that produces better performance overall. Separate training is not evaluated but could be explored as future work.

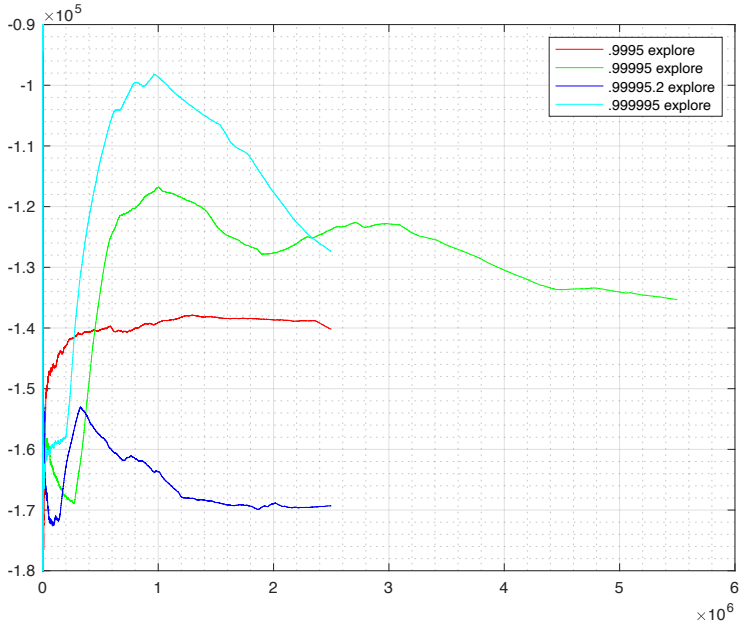


Figure 45: Cumulative reward average of four separate exploration values.

Exploration has strong influence over training. Exploration is a noise value added to the action that decays by multiplying itself at each iteration of each episode, just like the two-vehicle simulation. Figure 45 compares training runs with different exploration values: 0.9995, 0.99995, and 0.999995. An exploration value of 0.999995 is the best performer for the three-vehicle

simulation, the curve with the highest peak. The highest peak represents the cumulative average of reward values for all episodes trained. The curve selected has a peak of -0.98 x 10^5, around one million training episodes. This peak is much better than the other training trails.

Duplicate runs do not always produce the same performance. Notably, two training runs in Figure 45 use an exploration value of 0.99995 (green and dark blue curves). Both runs use the same variables and settings. Still, there is variation in the network training because of the randomly chosen values for start position, goal position, traffic vehicle length, TIV, speed, traffic action policy, and other variables. This inconsistency between training runs is common. It creates challenges in determining the best overall performer or optimal configurations. However, this brute-force hand tuning and laborious art of evaluation is typical of network tuning.

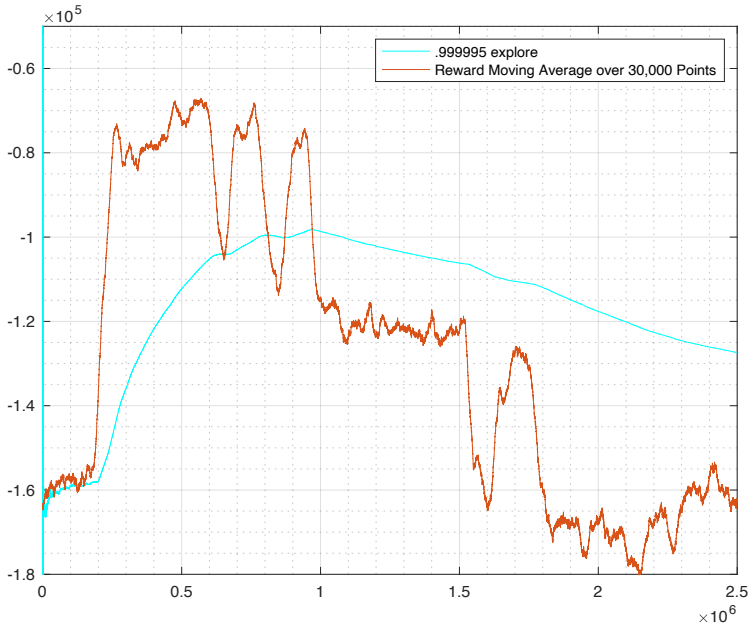


Figure 46: Cumulative reward average and moving mean for best variant: 0.999995 explore.

Figure 46 shows the cumulative average and the moving mean of the selected 0.999995 training run. The cumulative peak value is around 1 million episodes, but the highest average reward values are between 250K to 950K episodes. The best network is at 350K episodes, but

that is not clear by simply observing the figure. The following section gives further details about selecting the best network.

Figure 46 also shows that good performance is fleeting. When training reaches a cumulative average peak, it is typical for the cumulative average to worsen. The blue and green curves of Figure 45 prominently show the worsening behavior. It is unclear why this happens, but this behavior is typical in training neural networks. A repercussion of this in the autonomous driving use case is that training would not be acceptable to continue indefinitely as it could lead to unexpected and poor performance. This idea is unintuitive for those who consider artificial intelligence to be like human behavior because the common expectation is that once humans learn to merge into traffic well, they do not progressively get worse at it.

6.4 Best Network Selection

A MATLAB script summarizes relevant data from the saved test files created periodically during training. The summarized data is used to select the best network. The script is initially introduced in Section 5.5. The training algorithm saves network weights and generates a .csv output file of standard test results every 25K episodes. These same files form the standard test results and associated tables in Section 6.5 and similar sections. The .csv test results files are used as input to the script which runs a test routine to generate performance evaluation data.

The MATLAB script inputs test results and summarizes total collisions, acceleration/deceleration average values, and ratios. Table 12 shows the summarized output and is sorted based on the best collision performance. The summarized results comprise all three testing action policies, the standard starting position/goal position combinations, and a set of gap spacings for the traffic vehicles. The best network for the three-vehicle simulation is at 350K

episodes. Best performance is judged by lowest total collisions. This 350K is network selected for a detailed review of standard performance test results.

Table 12: Shows the 20 best performing tests in ascending values of total collisions. This is an excerpt from the full performance table. Simulator demonstration results use training weights from the best performer at 350K training episodes.

Episodes	Merge Decel Average (m/s^2)	Merge Accel Average (m/s^2)	Merge Decel Occurrence	Merge Maintain Occurrence	Merge Accel Occurrence	Total Collisions
350000	-3.5546	2.3912	58.5%	0.0%	41.5%	1626
375000	-3.5285	2.4422	59.3%	0.0%	40.7%	1671
425000	-3.4941	2.4523	59.2%	0.0%	40.8%	1734
325000	-3.2248	2.1612	57.9%	0.0%	42.1%	1738
400000	-3.5153	2.4848	58.5%	0.0%	41.5%	1760
450000	-3.0462	2.5489	52.8%	0.0%	47.2%	1837
900000	-3.5357	2.5014	59.6%	0.0%	40.4%	1849
950000	-3.0177	2.4784	52.4%	0.0%	47.6%	1909
275000	-3.5926	2.1037	63.0%	0.0%	37.0%	1923
300000	-3.4531	2.2636	61.3%	0.0%	38.7%	2112
475000	-2.6084	2.4322	47.6%	0.0%	52.4%	2165
600000	-2.6452	2.9113	42.1%	0.0%	57.9%	2519
875000	-2.693	2.8486	43.5%	0.0%	56.5%	2587
975000	-3.5269	2.4826	57.3%	0.0%	42.7%	2841
925000	-4.0859	1.7296	76.5%	0.0%	23.5%	2866
625000	-2.8859	2.8182	44.2%	0.0%	55.8%	2882
525000	-2.6582	2.6782	43.5%	0.0%	56.5%	2898
500000	-2.8701	2.2805	49.5%	0.0%	50.5%	2919
750000	-1.8671	2.6351	37.8%	0.0%	62.2%	3003

Table 13: Three-vehicle simulation performance evaluation breakdown by policy. The top three performers for each policy is shown. Highlighted rows are the same overall best performing DRL network.

Policy	Episodes	Merge Decel Average (m/s^2)	Merge Accel Average (m/s^2)	Merge Decel Occurrence	Merge Maintain Occurrence	Merge Accel Occurrence	Total Policy Collisions
Constant	375000	-3.50	2.59	58%	0%	42%	522
Constant	**350000**	**-3.53**	**2.53**	**57%**	**0%**	**43%**	**525**
Constant	325000	-3.27	2.21	61%	0%	39%	534
Random	**350000**	**-3.55**	**2.44**	**58%**	**0%**	**42%**	**549**
Random	325000	-3.34	2.15	63%	0%	37%	550
Random	375000	-3.55	2.54	59%	0%	41%	558
Reactive	**350000**	**-3.59**	**2.21**	**57%**	**0%**	**43%**	**552**
Reactive	950000	-3.10	2.28	50%	0%	50%	561
Reactive	450000	-3.08	2.24	51%	0%	49%	567

Table 13 shows the top three performers for each individual policy. In two out of three policies, 350K is also the best performer. 375K episodes is the second best performer overall and best performer for the individual constant policy.

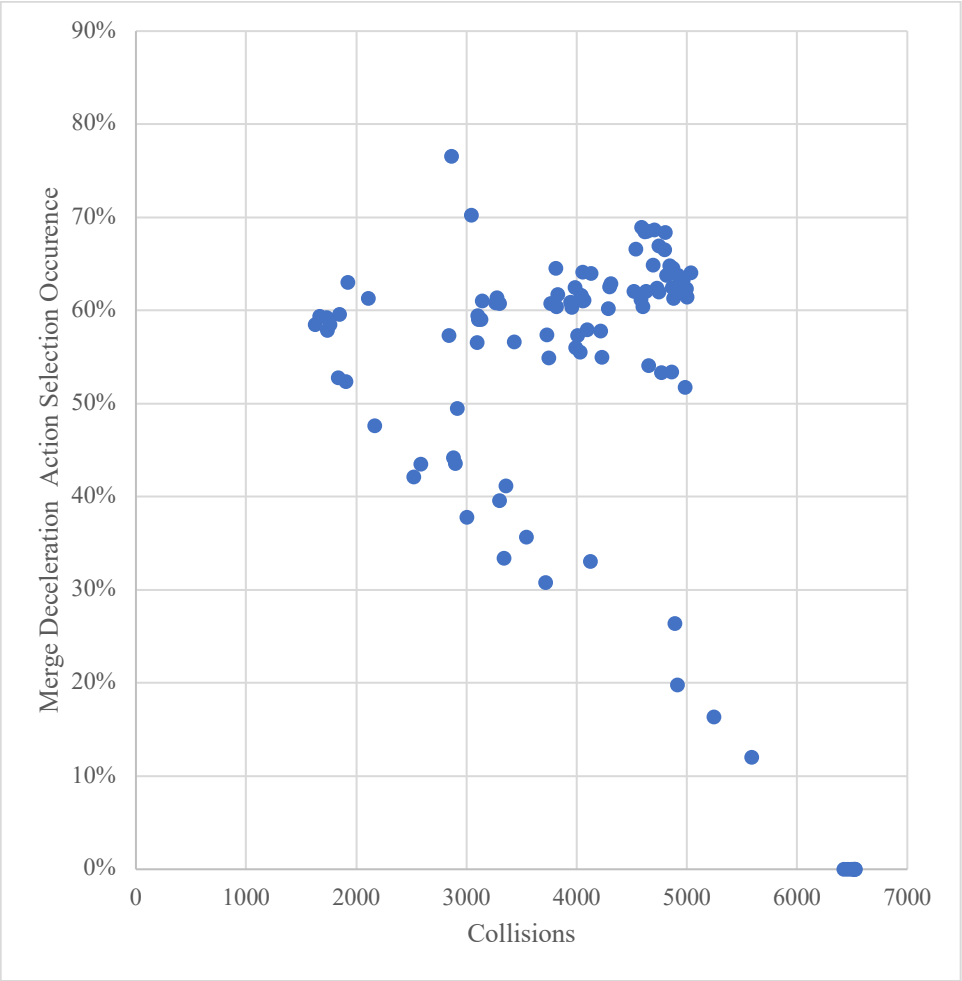


Figure 47: Correlation plot comparing collisions to deceleration occurrence. Blue dots indicate deceleration action selection occurrence and the number of collisions per test set. Data points show two linear groupings extending horizontally near the 60% occurrence and another linear grouping extending downward and to the right.

Figure 47 shows a scatter plot of collisions versus merge deceleration occurrence from every 25K test episode. In this graphical format, the acceleration/deceleration[7] ratio is appears correlated to collisions. Similar to the two-vehicle simulation, decreasing deceleration occurrence leads to higher collisions. However, there is no clear deceleration occurrence value

that produces the lowest number of collisions. This again shows that deceleration occurrence alone is not a singular factor in collision performance.

The deceleration to acceleration in the three-vehicle scenario is now a 60%-40% average over all test episodes. This ratio is now more biased towards deceleration than the 50%-50% deceleration to acceleration pattern observed in the two-vehicle scenario. The greater bias is due to the influence of the additional vehicle.

Traffic_Gap	50										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	90%	88%	84%	77%	72%	66%	64%	62%	62%	77%
-15	100%	100%	100%	92%	89%	81%	76%	68%	65%	63%	83%
-10	100%	100%	100%	100%	97%	91%	82%	78%	70%	65%	88%
-9	100%	100%	100%	100%	100%	92%	87%	78%	72%	68%	90%
-8	100%	100%	100%	100%	100%	96%	87%	79%	73%	67%	90%
-7	100%	100%	100%	100%	100%	97%	90%	81%	73%	68%	91%
-6	100%	100%	100%	100%	100%	100%	93%	81%	76%	68%	92%
-5	100%	100%	100%	100%	100%	100%	93%	85%	77%	70%	92%
-4	100%	100%	100%	100%	100%	100%	95%	87%	79%	72%	93%
-3	100%	100%	100%	100%	100%	100%	97%	91%	79%	75%	94%
-2	100%	100%	100%	100%	100%	100%	99%	91%	82%	74%	95%
-1	100%	100%	100%	100%	100%	100%	98%	93%	84%	76%	95%
0	100%	100%	100%	100%	100%	100%	100%	96%	85%	78%	96%
1	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	1%
2	0%	0%	0%	0%	0%	0%	0%	0%	1%	11%	1%
3	0%	0%	0%	0%	0%	0%	0%	0%	4%	12%	2%
4	0%	0%	0%	0%	0%	0%	0%	2%	8%	13%	2%
5	0%	0%	0%	0%	0%	0%	0%	5%	8%	14%	3%
6	0%	0%	0%	0%	0%	0%	0%	6%	8%	12%	3%
7	0%	0%	0%	0%	0%	0%	2%	8%	10%	17%	4%
8	0%	0%	0%	0%	0%	0%	0%	10%	15%	18%	4%
9	0%	0%	0%	0%	0%	0%	2%	13%	18%	21%	5%
10	0%	0%	0%	0%	0%	0%	2%	10%	22%	24%	6%
15	0%	0%	0%	0%	0%	7%	15%	25%	46%	43%	14%
20	0%	0%	0%	0%	13%	18%	34%	43%	56%	47%	21%
Grand Total	**52%**	**52%**	**52%**	**51%**	**51%**	**50%**	**49%**	**48%**	**47%**	**46%**	**50%**

Figure 48: Three-vehicle simulation deceleration occurrence by Ego initial start position versus goal position at a 50 m starting gap for the two traffic vehicles.

For example, Figure 48 shows a table of the 350K network standard test results with values for average deceleration action selection occurrence. In this chart, the second traffic vehicle is at 50 m away from the closest traffic vehicle to ego. This is about 10 car lengths away, so the influence of the second vehicle is expected to be relatively small. The results show an

overall average deceleration occurrence of 50% which is the same as the overall average bias of 50%-50% deceleration to acceleration of the two-vehicle scenario.

The 60%-40% overall average includes all other gaps settings too. As the gap setting gets smaller, the deceleration average bias gets larger. For further interest, Appendix Section A.4 shows multiple detailed graphs for the acceleration and deceleration occurrence and average values. The data shows that the additional vehicle in the scene has influence over the occurrence value bias. The ego vehicle reacts more defensively by increasing how much it decelerates as the traffic vehicle gap decreases.

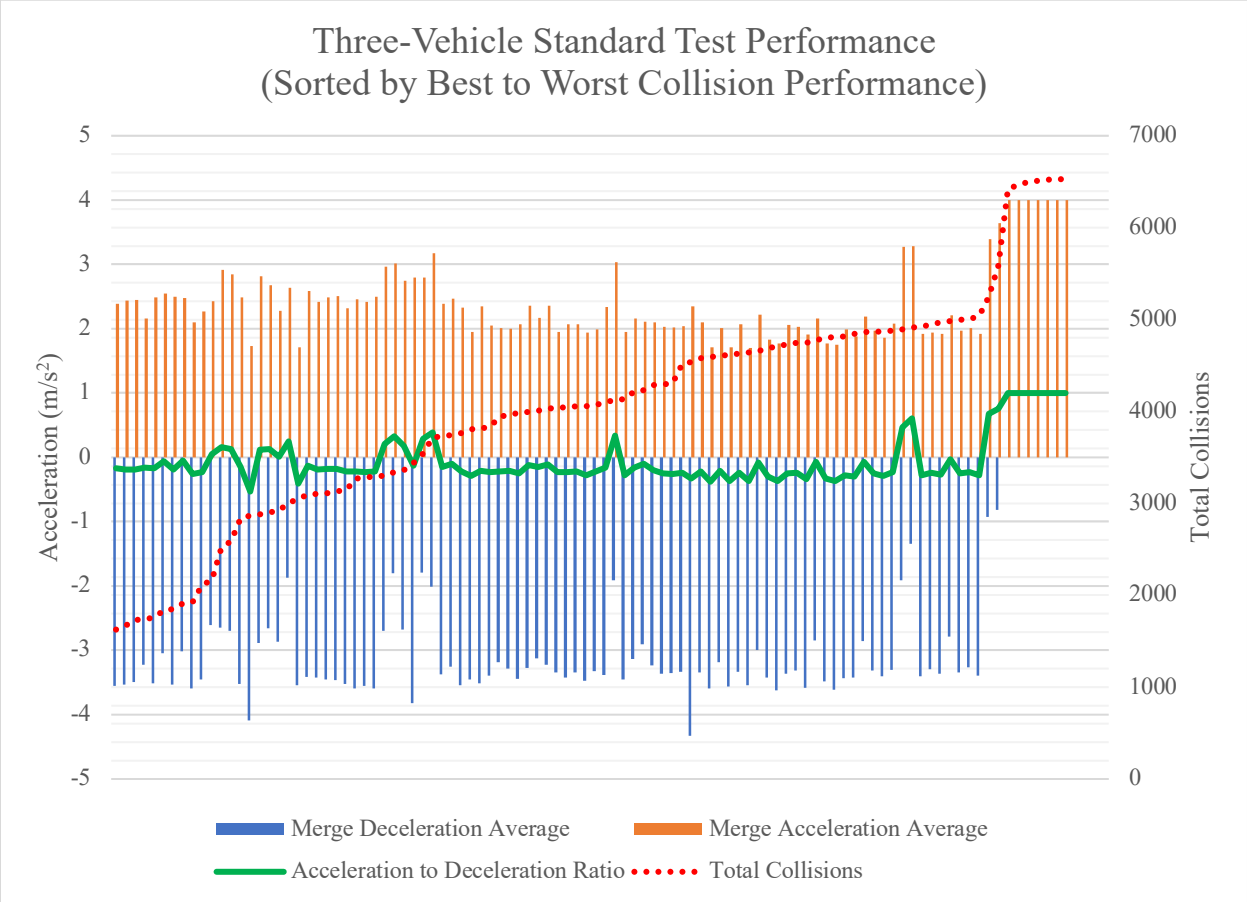


Figure 49: Standard test performance of the three-vehicle scenario with data sorted by best to worst collision performance. Blue and orange bars indicate average deceleration and acceleration for each test instance. The green line is the difference between the acceleration and deceleration occurrence where negative values indicate greater deceleration than acceleration. The red line is on the secondary axis to the right and represents the total collisions for each blue-orange bar pair where a test is performed.

Figure 49 is a multi-dimensional plot of the full set of data represented by Table 12. The leftmost values in the plot correspond to the topmost rows of the table. The plot extends the data represented to include all test values through the 2.5 million episode test and not just the first 20 tests. The best performing tests show a bias towards deceleration occurrences. The bias is fairly consistent for the first 10 test groups, but the total collisions tend to increase rapidly in the same period. The bias generally stays the same past the first 10 test groups, but with slightly more bias towards deceleration and some fluctuation in the average acceleration and deceleration values. In the plot of the data, there appears to be no other clearly relevant pattern for increased collisions than the slight shift towards more deceleration. Towards the right extent of the plot, there is a shift to full acceleration at the limit of 4 m/s^2 and only acceleration values are chosen by the network. Collisions are the greatest in this portion of the plot.

6.5 Standard Test Results

The results tables for the three-vehicle simulation the are very similar to those for the two-vehicle results. However, the traffic gap is a new variable in the charts. The gap represents the initial gap between the traffic vehicles at the start of the episode. Therefore, several tables, like the two-vehicle table, show results for each gap setting. The complete set of gaps is {5, 10, 15, 25, 50, 100} meters, but Figure 50, Figure 51, and Figure 52 show an abbreviated set consisting only of 5 m, 10 m, and 50 m. The full set of results for all the gaps are shown in the Appendix, Section A.5. The results for the additional gap settings are similar.

A 5 m gap is the same size as the test vehicles, making it a challenging condition because the ego vehicle needs to maneuver itself into a gap size the same as its length. 10 m is also a small gap size because most drivers prefer to maintain a gap of about three car lengths [64], or in this case, 15 m.

Training is conducted with a randomly selected mixture of traffic action policies, as explained in Section 6.3. However, the two traffic vehicles use the same policy (either constant, reactive, or random) for each standard test. This evaluates performance in a more standard way versus dissimilarities in action between the two traffic vehicles with mixed policies. Future work could be conducted to explore relationships further between a mix of the policies. A supplementary video of the simulation rendering output is available that shows the one merge vehicle interacting with the two traffic vehicles against all three traffic polices.

6.5.1 Constant Policy

Figure 50 shows the standard test table results for the traffic vehicles using the constant action policy. Both traffic vehicles use the same constant action policy. Time in between vehicles (TIV) for rear traffic is limited to 0.8, so it slows to reach the TIV limit and avoid a collision. If TIV slowing were disabled, collisions would occur in yellow boxes too. However, TIV slowing is enabled, so the rear traffic vehicle continues to select max deceleration actions until it reaches a 0.8 second or more TIV to the front traffic vehicle. The 50 m gap is a typical minimum gap for ACC systems. Its performance matches single-vehicle ideal results (Figure 29).

Traffic_Reaction	['constant' 'constant']																																					
Average of Collision	**Traffic Gaps: {5, 15, 50} m; Goal Positions: {10, ..., 100}**																																					
	5										**5 Total**	**15**										**15 Total**	**50**										**50 Total**	**Grand Total**
Ego Starting Pos. (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		
-50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-20																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-10												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-8	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-7	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-6	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-5	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	13%
-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%
0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%
1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	33%
2	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	30%
3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	23%
4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	17%
5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
6	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
7	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
8	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
9	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	13%
10												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	**95%**	**79%**	**42%**	**5%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**22.1%**	**39%**	**30%**	**13%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**8.3%**	**29%**	**26%**	**16%**	**6%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**7.7%**	**12%**

Figure 50: Standard test table for three-vehicle constant policy. Test results are expanded from two-vehicle to show multiple gap settings of 5m, 15m, and 50m.

6.5.2 Reactive Policy

Figure 51 shows the standard test results tables for reactive traffic actions. In this policy, both traffic vehicles take actions to avoid collisions within the traffic lane. While avoiding collisions with each other, the traffic vehicles control their longitudinal position to allow the merge vehicle to enter the stream of traffic without a collision. Performance for the reactive traffic policy is similar overall to the ideal reactive policy of Figure 31. The learned performance at a 50 m gap is slightly worse than the 15 m gap. This is because the two networks have learned to behave differently. It is not an effect of TIV slowing because the reactive policy does not use it. The poor performance is an example of an unexpected performance decrease. However, overall, the results are still nearly ideal as compared to the ground truth. The ego merge and traffic vehicles have clearly learned behavior that allows them to avoid collisions in most instances of starting gap deltas and goal positions.

Traffic_Reaction	['reactive' 'reactive']																																	
Average of Collision	**Column Labels**																																	
	5										**5 Total**	**15**										**15 Total**	**50**										**50 Total**	**Grand Total**
Row Labels	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		
-50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-20																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-10												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-8	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-7	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
-6	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-5	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-3	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%
-1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%
0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	30%
1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	40%	33%
2	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	30%	27%
3	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	30%	30%
4	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	20%	23%
5	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
6	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
7	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
8	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
9	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
10												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	**95%**	**84%**	**47%**	**5%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**23.2%**	**35%**	**30%**	**9%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**7.4%**	**26%**	**23%**	**13%**	**3%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**6.5%**	**11%**

Figure 51: Three-vehicle reactive policy standard test table.

6.5.3 Random Policy

Overall, the random traffic action performance is similar to the constant policy. At a 5 m, the gap between traffic vehicles is nearly impossible to merge into because the vehicle length itself is also 5 m. This would be a challenging scenario, even for human drivers. This difficulty results in greater collisions in the smaller goal positions, especially at 10 – 30 meters.

The performance of the 15 m gap table is better than the 50 m setting. It performs better because of TIV slowing. The rear traffic vehicle slows to increase the TIV between itself and the front traffic vehicle to reach the 0.8 set point. Slowing the rear traffic vehicle down creates a more predictable action strategy that enables the ego merge vehicle to take better evasive action to avoid a collision.

Traffic_Reaction	['random' 'random']																																	
Average of Collision	**Column Labels**																																	
	5										**5 Total**	**15**										**15 Total**	**50**										**50 Total**	**Grand Total**
Ego Starting Pos. (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		
-50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-20																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-10												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-8	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-7	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-6	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-5	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-4	100%	100%	33%	0%	0%	0%	0%	0%	0%	0%	23%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	14%
-3	100%	100%	67%	0%	0%	0%	0%	0%	0%	0%	27%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	22%
-2	100%	100%	33%	0%	0%	0%	0%	0%	0%	0%	23%	100%	100%	67%	0%	0%	0%	0%	0%	0%	0%	27%	100%	100%	33%	0%	0%	0%	0%	0%	0%	0%	23%	24%
-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%
0	100%	100%	100%	33%	0%	0%	0%	0%	0%	0%	33%	100%	100%	100%	67%	0%	0%	0%	0%	0%	0%	37%	100%	100%	100%	67%	0%	0%	0%	0%	0%	0%	37%	36%
1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	33%
2	100%	100%	100%	0%	33%	0%	0%	0%	0%	0%	33%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	67%	0%	0%	0%	0%	0%	0%	37%	30%
3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	67%	0%	0%	0%	0%	0%	0%	0%	27%	22%
4	100%	100%	0%	33%	0%	0%	0%	0%	0%	0%	23%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	18%
5	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
6	100%	100%	67%	0%	0%	0%	0%	0%	0%	0%	27%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	9%
7	100%	100%	67%	67%	0%	0%	0%	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	11%
8	100%	100%	100%	33%	0%	0%	0%	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	11%
9	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	13%
10												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	**95%**	**79%**	**46%**	**16%**	**2%**	**0%**	**0%**	**0%**	**0%**	**0%**	**23.7%**	**39%**	**30%**	**16%**	**3%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**8.8%**	**29%**	**26%**	**16%**	**8%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**7.8%**	**12%**

Figure 52: Three-vehicle random-acting traffic policy standard test table.

Overall, the performance of the ego merge vehicle against the random-acting traffic policy is better than the two-vehicle results shown in Figure 32 and Figure 35. These results show that the three-vehicle variant has learned to avoid collisions better than the two-vehicle variant from Chapter 5. The better performance is not limited to just the random policy. The performance is generally better in the constant and reactive policies too.

One possible explanation for the better performance is the more systemic and data-driven approach to best network selection. A simple observation of the graph chose the best two-vehicle DRL network, but Section 5.5 shows that there are better-performing networks. The three-vehicle variant DRL network was selected using the improved MATLAB-based selection method described in Section 6.4.

6.6 Three Vehicle-Simulation Discussion

From a simplistic perspective, the three-vehicle simulation adds a single vehicle to a two-vehicle road scene. However, adding that additional traffic vehicle resulted in a significant

change to the algorithm itself, changed the state parameters for the agents (vehicles), and created a different state set between the traffic and merge vehicles. A better method for selecting the best DRL network helped to improve performance. Additional similarities persisted, like sensitivity to exploration and a pattern in the ratio of deceleration to acceleration in the best networks. Despite the significant change, the results are similar between the two scenes and close to ideal behavior. These promising results show that scaling up the merge road scene with additional vehicles can still produce outstanding collision-avoiding performance using multi-agent DRL.

The next chapter details the work on the full scene. The full scene model represents a real-world busy road scene. It finalizes the work done in the preceding chapters with the building blocks laid by the study of smaller road scenes.

Chapter 7 Full Scene Simulation

The Full-Vehicle Scene, Figure 53, consists of two vehicles in the merge lane and two or more in the traffic lane. shows the scene. This merge scene is representative of a real-world traffic scene. In this version, the traffic vehicles remain in the scene once the rear ego vehicle moves ahead or behind the two-vehicle front-rear traffic pair, compared to the three-vehicle simulation which kept only the two closest traffic vehicles. Overall, the results for the full scene are similar to the three- and two-vehicle results.

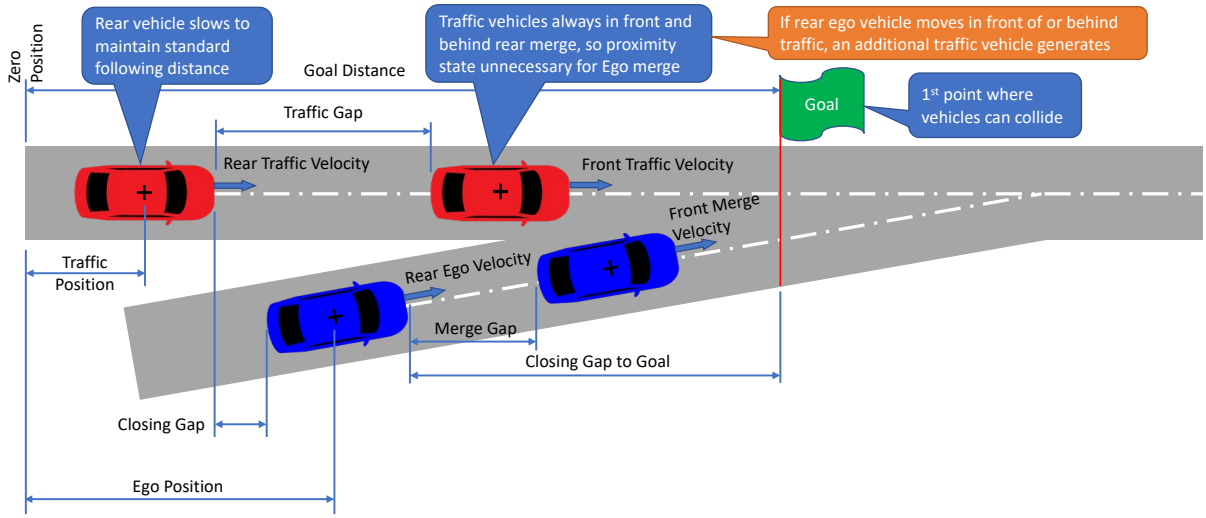


Figure 53: Full scene merge scenario. Two vehicles in the merge lane and two or more vehicles in the traffic lane. During training and testing, as the rear merge vehicle moves in front of or behind the traffic vehicle pair, another vehicle is generated to ensure there is always a traffic vehicle in front of and behind the ego vehicle.

In the three-vehicle version, there were a maximum of only two traffic vehicles possible, but now, there can be many traffic vehicles in the scene. There are always two merge lane

vehicles. The rear merge vehicle is the ego vehicle, and the front merge vehicle trains and uses the same DRL network as the rear ego vehicle.

The rear merge vehicle is the autonomous driving ego vehicle. The merge vehicle in front represents the first vehicle in front of ego. There could be more than one vehicle in front of ego, but ego's immediate in-lane concern will always be the vehicle in front of it. The front merge vehicle gates the acceleration behavior allowed, and ego needs to react if the front vehicle slows. Of course, the acceleration or deceleration of the front merge vehicle could be due to other vehicles in front of it. Still, ultimately their behavior will be represented by the front merge vehicle.

There is no rear merge vehicle represented in the road scene. The rear merge vehicle is inconsequential. It could collide into ego, but the ego vehicle would have no control over that behavior. The likelihood of a collision would indeed be higher if ego abruptly stopped while a vehicle behind it was closely following. Still, that same behavior could be represented by ego closely following the front merge vehicle. The training data shows that in-lane collisions between the merge vehicles are much less than 1% of the total collisions for the trained network. Those in-lane collisions only occur against the randomly acting traffic policy, an unrealistic policy meant to train agents in a highly difficult and unpredictable scenario. Therefore, a vehicle behind ego is not studied.

The two traffic vehicles are always in front of and behind the ego vehicle. This setup is a condition of the training and testing simulation for the ego vehicle. The front merge vehicle could also begin the simulation between the two traffic vehicles or be in front of the front traffic vehicle. If the front merge vehicle is in front of the front traffic vehicle, it assumes another vehicle is ahead of it at 100 m at 0 m/s closing speed. These default placeholder values train the

DRL network to always assume that there is a traffic vehicle in front and the rear, even if it is very far away.

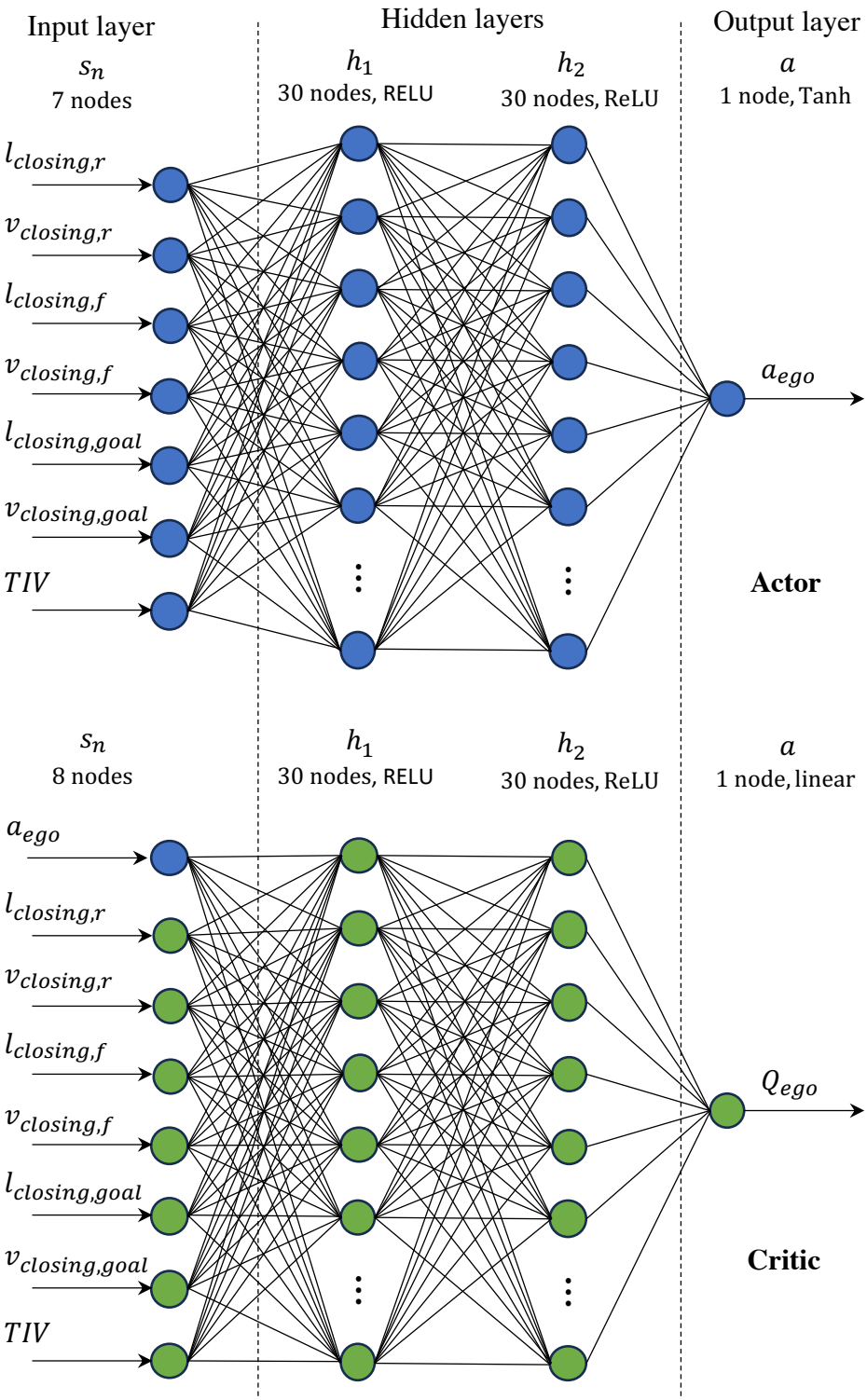


Figure 54: Full-scene actor-critic DRL network diagram.

Table 14: Full merge scene state parameters

Vehicle State Variables	Used by Merge	Used by Traffic	Range	Units
Closing Gap to Rear Vehicle in Next Lane	✓	✓	[-2.5, 30]	m
Closing Speed to Rear Vehicle in Next Lane	✓	✓	[-10, 10]	m/s
Closing Gap to Front Vehicle in Next Lane	✓	✓	[-2.5, 30]	m
Closing Speed to Front Vehicle in Next Lane	✓	✓	[-10, 10]	m/s
Closing Gap to Goal	✓	✗	[-160, 150]	m
Closing Velocity to Goal	✓	✗	[0, 40]	m/s
Closing Time from Current to Front Vehicle (TIV)	✓	✓	[0, 2.5]	s
Time to Goal Position	✗	✓	[0, 3]	s
Proximity to ego (*two variables: one front, one rear*)	✗	✓✓	{-1, 1}	*unitless*

7.1 State Set

The state set for the full scene is similar to the three-vehicle setup. Table 14 shows the state variables. The merge vehicle states remain the same, except for an added TIV variable to the front merge vehicle. The traffic vehicle states remain very similar too. They grow from a set of five states to a set of eight because of the additional merge vehicle closing gap, closing speed, and proximity (behind or in front). Figure 54 shows the network diagram for the merge vehicle DRL network. Figure 101 in the appendix shows the traffic DRL network.

7.2 Full-Scene Simulator Functionality

The simulator's functionality for the full scene is more sophisticated than all other variants. Two merge lane vehicles always exist. Two or more vehicles are in the traffic lane. In the three-vehicle simulation, the ego merge vehicle was always surrounded by a traffic vehicle in the front and one in the rear. When ego went in front of or behind the pair of traffic vehicles, a

new vehicle was generated to ensure ego always had a front and rear traffic vehicle. When the new vehicle was generated, the extra vehicle outside the front-rear pair was deleted from the simulation so only two remained. In the full-scene simulator, the existing vehicles are not deleted from the scene once a new one is generated. The ego state still only considers the front and rear traffic pair, but more than two traffic vehicles are allowed to exist in the simulation.

The step function action input is now a vector that contains both the ego merge rear and merge front vehicle actions. The motion equation updates for position and velocity calculations, and the counter functionality are unchanged. Rewards are also calculated and assigned per the reward formula in Section 4.1.4.

Logic is introduced for the merge front vehicle collision checking and reward assignment. A reward for merge front that is the magnitude of its acceleration is assigned at each step function iteration, just like ego and traffic. An in-lane collision check is introduced. At each step of the episode, the step function checks to see if the rear and front merge vehicles have collided. Penalties are assigned to both if they collide. Ego (rear) is given an at-fault collision penalty, merge front is given a no-fault collision penalty, and the episode stops.

A check determines if there is a collision between the front merge vehicle and any traffic vehicles when it reaches the goal position. An at-fault penalty is assigned for collisions, and the episode stops. A merge reward is assigned if no collision occurs between the front merge and traffic vehicles. After the front merge vehicle reaches the goal position, the episode continues until the ego (rear) merge vehicle reaches the goal position. All reward logic, penalties, and episode completion in the step function are the same for the ego merge vehicle.

The reset function creates the front merge vehicle. While training, the front merge vehicle length is chosen randomly between the traffic minimum and maximum hyperparameters. The

front merge vehicle in front of the ego (rear) at a gap with the hyperparameter limits. The goal position is now always placed in front of the front merge vehicle.

The states function was entirely rewritten but still produces the merge and traffic states as the output. The merge vehicles are now output as a vector, just like the traffic vehicles. The state function is improved to facilitate the additional traffic vehicles. A sorting key for traffic vehicle positions is added and used to re-sort the rest of the associated vehicle parameters. The sorting key maintains the existing simulator functionality while allowing the simulator to handle the additional vehicles. With this new sorting function, the algorithm keeps track of the closest rear and front traffic vehicles compared to ego while still performing all existing calculations already built.

The MIO function remains almost the same, with some minor changes based on the state changes and the addition of the second merge vehicle.

The output file now writes 89 parameters at each episode instead of 32 for the three-vehicle version. Separate functions now write the header and output file lines to make updates more manageable. The render function adds the front merge vehicle in light blue. A “collision burst” graphic shows the site where collisions occur to help visually identify collisions. The DDPG algorithm is described in Section 4.2 and remains the same as other versions.

The main program and test loops remain essentially unchanged. During testing, the front merge and traffic vehicles start with the same initial gap between in-lane vehicles. This setup generally creates a situation where both front and rear merge vehicles pair with front and rear traffic vehicles at the same initial position delta.

Table 15: Merge and traffic vehicle simulation hyperparameters for all DRL simulation variants

	Vehicle Parameter	Value	Units	Notes
Ego Merge	front_reward_at_merge	0	unitless	Merge front reward initialization[9]
	min_action	-5	m/s^2	Min. acceleration
	max_action	4	m/s^2	Max. acceleration
	min_position	-10	m	Min. position
	max_position	150	m	Max. position
	min_initial_position	-25	m	Episode min. initial position
	max_initial_position	50	m	Episode max. initial position
	min_speed	20	m/s	Min. allowed speed[8]
	max_speed	40	m/s	Maximum allowed speed[8]
	max_ttp	3	s	Goal time-to-position state max.
	min_closing_gap	-2.5	m	Closing gap to traffic state min.
	max_closing_gap	30	m	Closing gap to traffic state max.
	min_closing_speed	-10	m	Closing speed to traffic state min
	max_closing_speed	10	m	Closing speed to traffic state max
	min_goal_position	25	m	Min. distance to goal from zero
	max_goal_position	100	m	Max. distance to goal from zero
	power	0.1	s	Simulation timestep
	length	5	m	Ego length
	reward	0	unitless	Self-reward initialization to zero
	done	[0,0]	unitless	Initialize episode complete flag
Traffic	traffic_min_action	-5	m/s^2	Minimum acceleration
	traffic_max_action	4	m/s^2	Maximum acceleration
	traffic_min_position	-10	m	Minimum position
	traffic_max_position	150	m	Maximum position
	traffic_min_initial_position	-25	m	Episode min. initial position
	traffic_max_initial_position	50	m	Episode max. initial position
	traffic_min_speed	20	m/s	Min. allowed speed[8]
	traffic_max_speed	40	m/s	Max. allowed speed[8]
	traffic_min_length	1	m	Vehicle minimum length
	traffic_max_length	20	m	Vehicle maximum length
	traffic_next_max_tiv	5	s	Max time in between vehicles[9]
	traffic_min_tiv	0.1	s	Min. TIV state limit[9,10]
	traffic_min_gap	1	m	Min. initial gap between vehicles[9,11]
	traffic_max_gap	100	m	Max. inter-vehicle distance[9]
	traffic_max_tiv	2.5	s	Max TIV state limit[9]
	traffic_test_tiv	0.8	s	Initial TIV for test[9]

Table 15 shows the 36 hyperparameters for the full scene simulation. However, three of the hyperparameters change from the three-vehicle simulation. The done flag logic changes to

[9] Not applicable for 2-vehicle simulation.
[10] TIV reduced from 0.5 s for 3-vehicle simulation to 0.1 s in full scene to capture a wider range of possible values.
[11] Set to 44 m for 3-vehicle simulation, but reduced to 1 m in full-scene simulation to include a more realistic range of possible values.

track both the front and rear ego done state separately. Once the front merge vehicle completes a collision-free merge, it stops updating the DRL network. The minimum initial traffic gap setting is reduced from 44 to 1 m to ensure small gap following distances are learned by the DRL network. Additionally, the minimum TIV for traffic vehicles during training reduces from 0.5 s to 0.1 s. The TIV reduction promotes edge case training with closely following vehicles (commonly called tailgating).

7.3 Network Training and Performance

Network training is like the three-vehicle scene. Both merge lane vehicles update the DRL network at each step of each episode. All traffic vehicles that have a reactive policy update the traffic DRL network. The merge and traffic DRL networks are independently separate networks. Their weights and other network parameters are saved in separate data files.

Figure 55 shows the training graph of cumulative rewards and moving means over 10,000 data points for the front and rear merge and traffic vehicles. The graphs only show the front and rear vehicles, even if more are in the scene. Training runs to a total of ten million episodes. The peaks in the data for both the merge and traffic vehicles suggest that the best performance occurs around 4.2 to 4.5 million episodes.

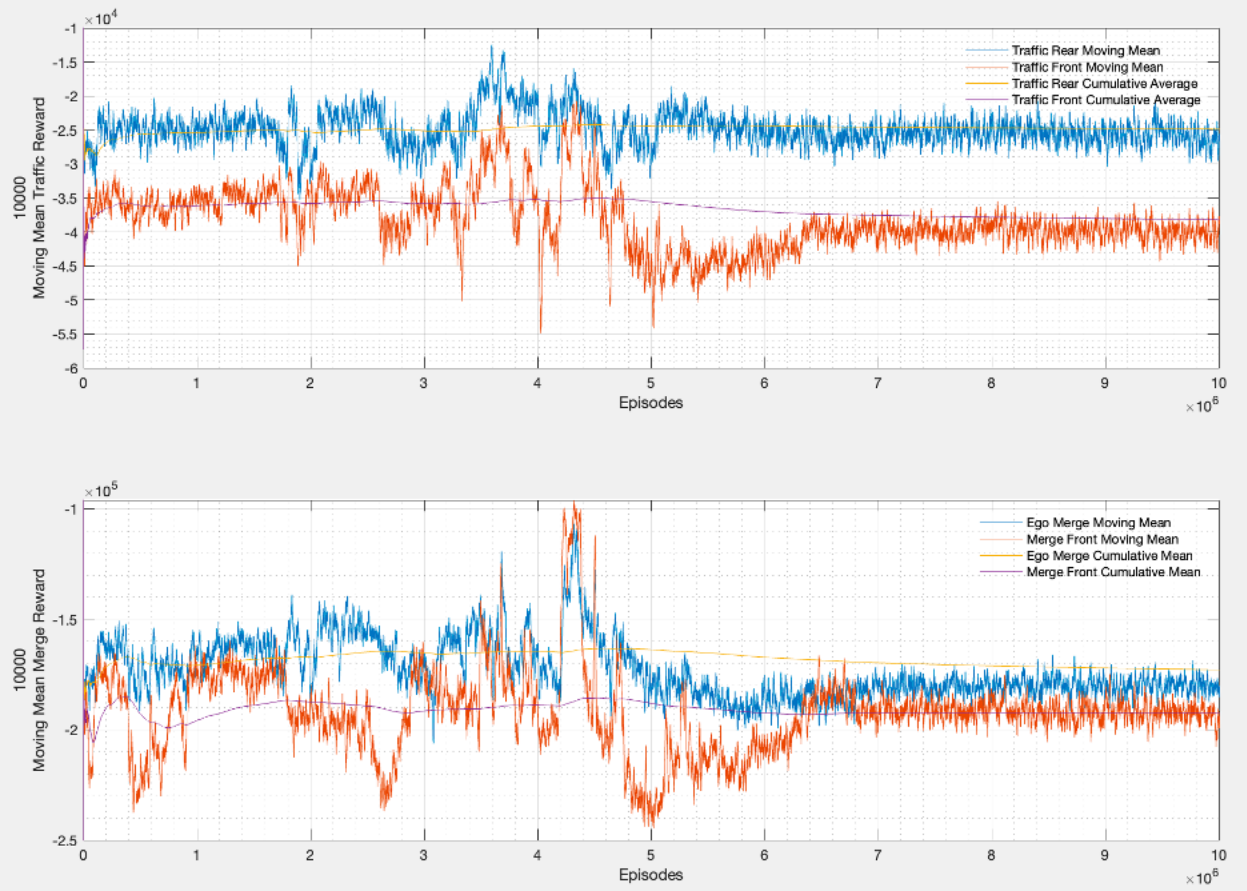


Figure 55: Training graph of moving mean and cumulative average for rewards. Data in the lower graph is for both front and rear merge vehicles. The upper graph shows the closest front and rear traffic vehicles to the rear merge vehicle. Near seven million episodes, the acceleration action for both the merge and traffic vehicles settle to continuously choose an action limit value of -5 m/s acceleration, regardless of the state value.

7.4 Training Performance

Figure 55 shows training performance settle after about seven million episodes. This behavior represents typical behavior that occurs during training of these merging networks. Early on in the training, the performance increases and decreases, but on the whole, it generally improves and peaks. Sometimes, the cumulative average will have several peaks, but normally there is one prominent global peak that is clearly highest. After the peak, cumulative performance decreases until the reward value settles towards a repetitive value within the band of randomness for the training variation. The reward value settles because the network learns to stick with an extreme value at the acceleration limit, either -5 m/s^2 or 4 m/s^2. As training

continues, the network never learns to deviate from the extreme value limit. Figure 55 shows this settling occur around seven million episodes and even though it runs for millions of more episodes, the value never changes. More advanced learning techniques could be applied in future work. For example, if the settling behavior is encountered, the training could revert to a global peak value to try to reach a higher global peak. Alternatively, the exploration noise could be reset to its initial state in an attempt to break away from the acceleration limit.

The global peak values found from the training show nearly ideal performance when measured against the ground truth. The ground truth requires both acceleration and deceleration values to achieve ideal performance. In accordance with the propositions described in earlier chapters, a vehicle that leads should accelerate and a vehicle that lags (is behind the other vehicle) should decelerate to avoid a collision. This is the only way to achieve ideal performance. Ideal performance cannot be achieved if the network learns to only take extreme acceleration limit actions. Therefore, the settling behavior that occurs is not a learned ideal performance. Fundamentally, it shows that the network has reached a local minimum and is unable to break away from it.

Table 16: Full scene training performance with increasing episodes. Training data and standardized tests are performed at 25K intervals. The best performing network is marked in bold highlight and has the lowest merge rear (ego) collisions at 1,392, but also has the lowest total collisions at 4,389.

	Merge Rear						Merge Front						
Episode	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Mtn %**	**Accel %**	**Coll.**	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Mtn %**	**Accel %**	**Coll.**	**Total Coll.**
25000	-5.00	0.00	100%	0	0	3704	-5.00	0	100%	0	0	8005	11709
50000	-5.00	0.00	100%	0	0	3142	-5.00	0	100%	0	0	9086	12228
75000	-5.00	0.00	100%	0	0	2523	-5.00	0	100%	0	0	10265	12788
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4175000	-1.26	3.42	18%	0%	82%	5391	-1.97	2.67	35%	0%	65%	6552	11943
4200000	-1.04	3.44	14%	0%	86%	5082	-1.56	2.88	29%	0%	71%	7285	12367
4225000	-3.79	0.92	82%	0%	18%	2306	-2.93	1.52	62%	0%	38%	3401	5707
4250000	-2.70	2.97	36%	0%	64%	3890	-2.48	2.61	41%	0%	59%	4019	7909
4275000	-3.11	3.21	41%	0%	59%	3464	-0.90	3.28	18%	0%	82%	4009	7473
4300000	-3.94	2.63	61%	0%	39%	3342	-2.30	2.78	38%	0%	62%	3974	7316
4325000	**-3.96**	**1.45**	**78%**	**0%**	**22%**	**1392**	**-2.42**	**2.76**	**38%**	**0%**	**62%**	**2997**	**4389**
4350000	-4.46	1.95	76%	0%	24%	1562	-2.76	2.63	44%	0%	56%	2918	4480
4375000	-4.63	1.42	85%	0%	15%	2191	-1.74	2.89	30%	0%	70%	3200	5391
4400000	-4.62	0.54	90%	0%	10%	3652	-4.78	0.68	92%	0%	8%	8365	12017
4425000	-4.64	0.56	90%	0%	10%	3482	-4.90	0.37	97%	0%	3%	8847	12329
4450000	-4.61	0.42	91%	0%	9%	3276	-4.98	0.07	100%	0%	0%	9948	13224
4475000	-4.52	0.50	89%	0%	11%	3268	-4.97	0.11	99%	0%	1%	10042	13310
4500000	-3.94	1.29	77%	0%	23%	3619	-3.40	1.71	62%	0%	38%	4011	7630

7.5 Best Network Selection

The MATLAB performance testing script runs with the output files generated every 25K episodes during training. Table 16 shows selected portions of the summarized performance data. The best network is at 4.325 million episodes with total collisions of 4,389. This collision count is about 60% lower than most other tests. This best performer at 4.325 million episodes strongly corresponds to the peak in the moving average graph of Figure 55 and is like peaks found in scenes with fewer vehicles. The Merge Rear vehicle is the same as the ego vehicle in the two- and three-vehicle simulations. It too has the lowest number of collisions at 4.325 million episodes of training. Table 17 shows 4.325 million episode network is the best for each individual policy too.

Table 17: First few rows of best performing individual policies. The best performing network remains consistent at 4.325M episodes and the top three generally maintain consistency too with exception to the reactive policy.

		Merge Rear						Merge Front					
Policy	**Episode**	**Dec. Avg.**	**Acc. Avg.**	**Dec. %**	**Mtn. %**	**Acc. %**	**Coll.**	**Dec. Avg.**	**Acc. Avg.**	**Dec. %**	**Mtn. %**	**Acc. %**	**Coll.**
Random	4325000	-4.06	1.46	79%	0%	22%	438	-2.49	2.76	38%	0%	62%	951
Random	4350000	-4.46	1.97	76%	0%	24%	470	-2.82	2.63	44%	0%	56%	965
Random	4375000	-4.66	1.43	85%	0%	15%	673	-1.77	2.88	31%	0%	69%	965
Constant	4325000	-3.98	1.50	78%	0%	22%	402	-2.40	2.76	37%	0%	63%	1002
Constant	4350000	-4.46	2.02	75%	0%	24%	417	-2.73	2.63	43%	0%	57%	1008
Constant	4375000	-4.64	1.45	85%	0%	15%	669	-1.74	2.89	30%	0%	70%	1005
Reactive	4325000	-3.83	1.40	78%	0%	22%	552	-2.38	2.77	37%	0%	63%	1044
Reactive	4350000	-4.45	1.86	76%	0%	24%	675	-2.72	2.63	43%	0%	57%	945
Reactive	4225000	-4.14	0.86	85%	0%	18%	780	-3.24	1.49	67%	0%	33%	1161

Figure 56 shows the deceleration occurrence compared to the number of collisions for each performance evaluation set saved during training. Similar to the two- and three-vehicle simulations, the scatter plot shows that when there is a low occurrence of deceleration action selection, there are more collisions. The pattern is relatively well defined as deceleration action occurrence selection gets lower. The scatter plot also shows that as deceleration selection percentage increases, the performance also tends to worsen. However, just like previous results, the best performing networks do not have a single factor correlation to the selection percentage.

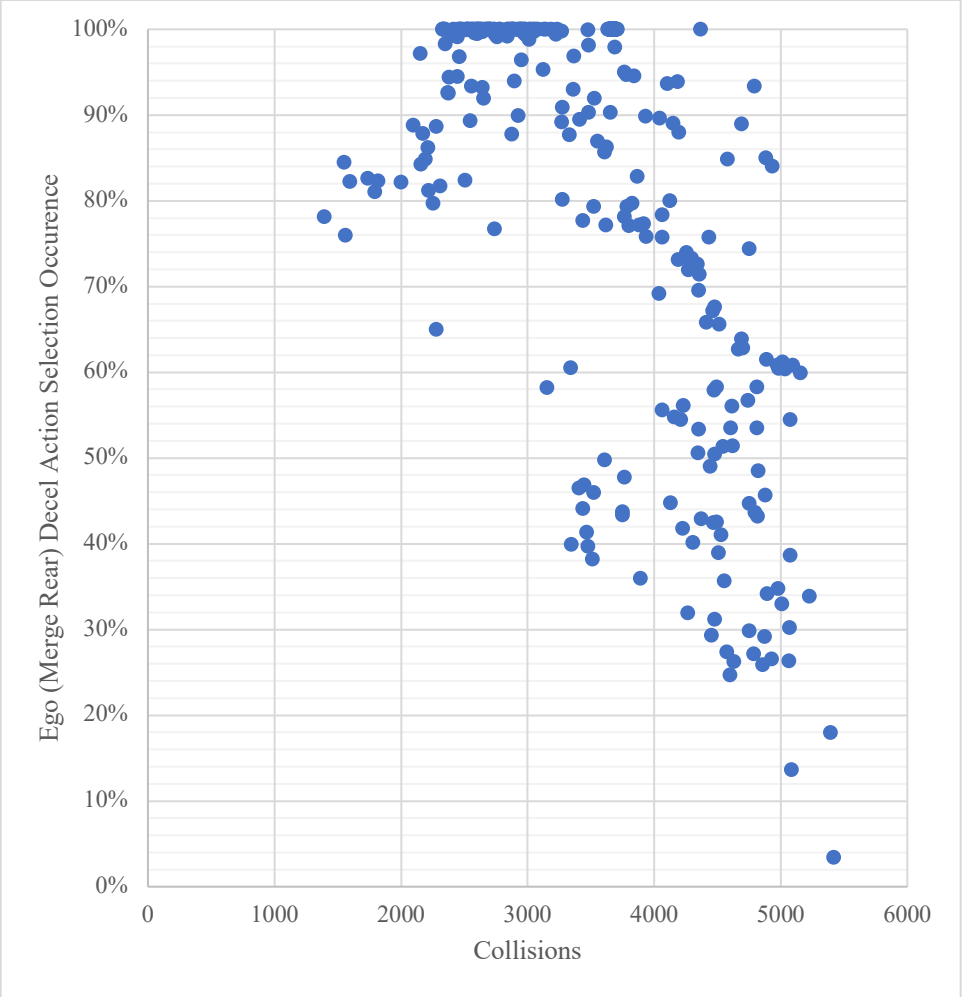


Figure 56: Ego deceleration action selection occurrence for all saved networks during training.

The average deceleration action occurrence ratio is 78%-22% deceleration to acceleration for the best network in the full scene. This shifts even further towards deceleration as compared to the two- and three- vehicle simulations which were at 60%-40% and 50%-50%, respectively. However, just as the two-vehicle ratio was biased towards deceleration occurrence when an additional traffic vehicle was present, so does the full scene.

Conducting further review on the performance testing data of the best (4.325M episode) network shows that when the second vehicle is moved further out, the average deceleration occurrence becomes closer to 55% deceleration. This is closer to the roughly 50%-50% ratio

expected by Proposition III. Appendix A.6 shows a sweep of the tested gaps and the changes between them for further detail, but they are left out of the main text to avoid further redundancy.

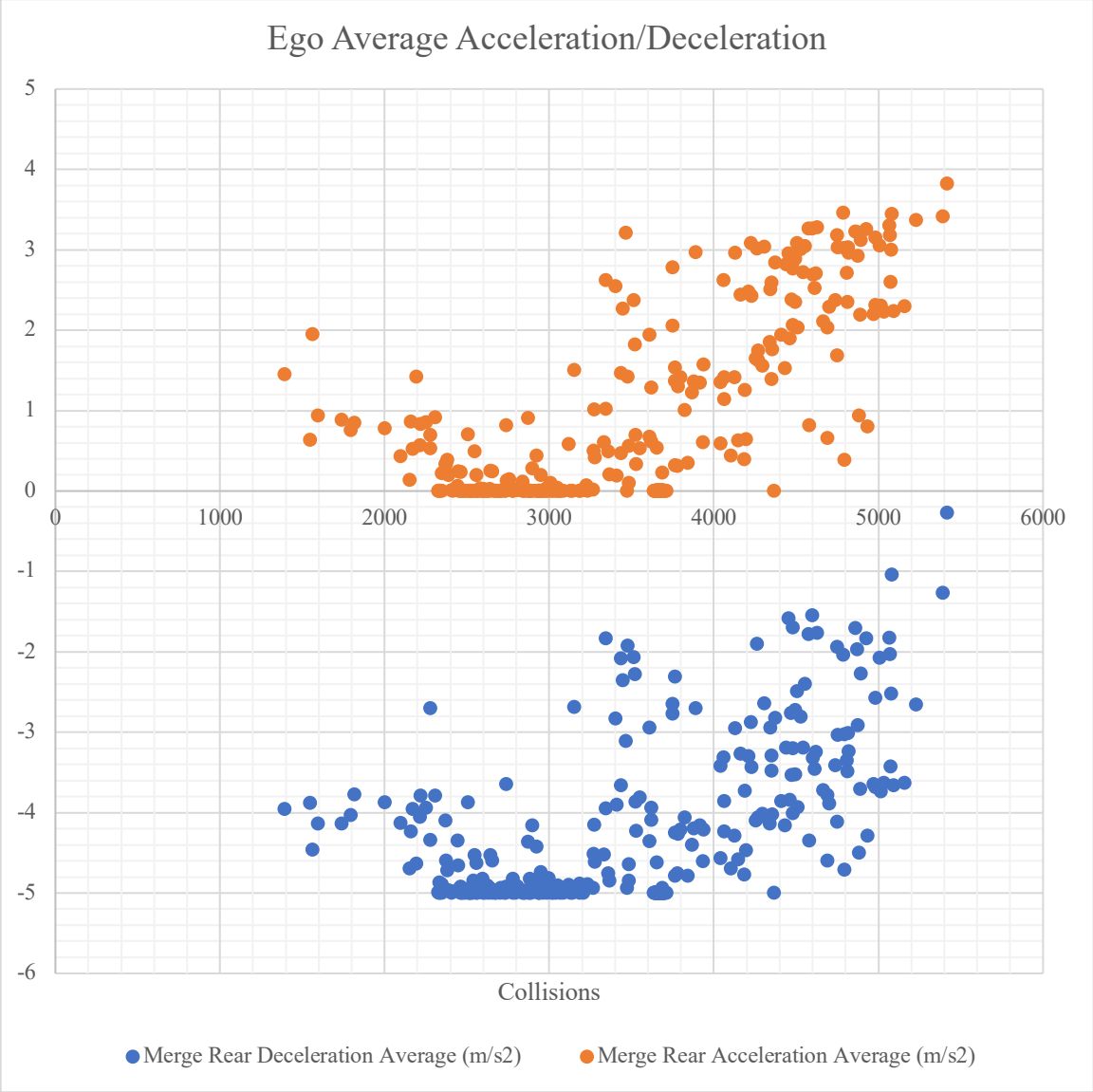


Figure 57: Average acceleration and deceleration for the full scene set of performance tests over the duration of training.

Similar patterns from the occurrence results can be seen in the average acceleration and deceleration values. Figure 57 shows the average acceleration and deceleration for the ego (rear merge) vehicle. Too much or too little acceleration or deceleration results in a greater number of collisions. However, this apparent correlation does not represent causation when considered from

an acceleration or deceleration average value perspective, since the best performing acceleration and deceleration values also have poor performance at different training episode snapshots.

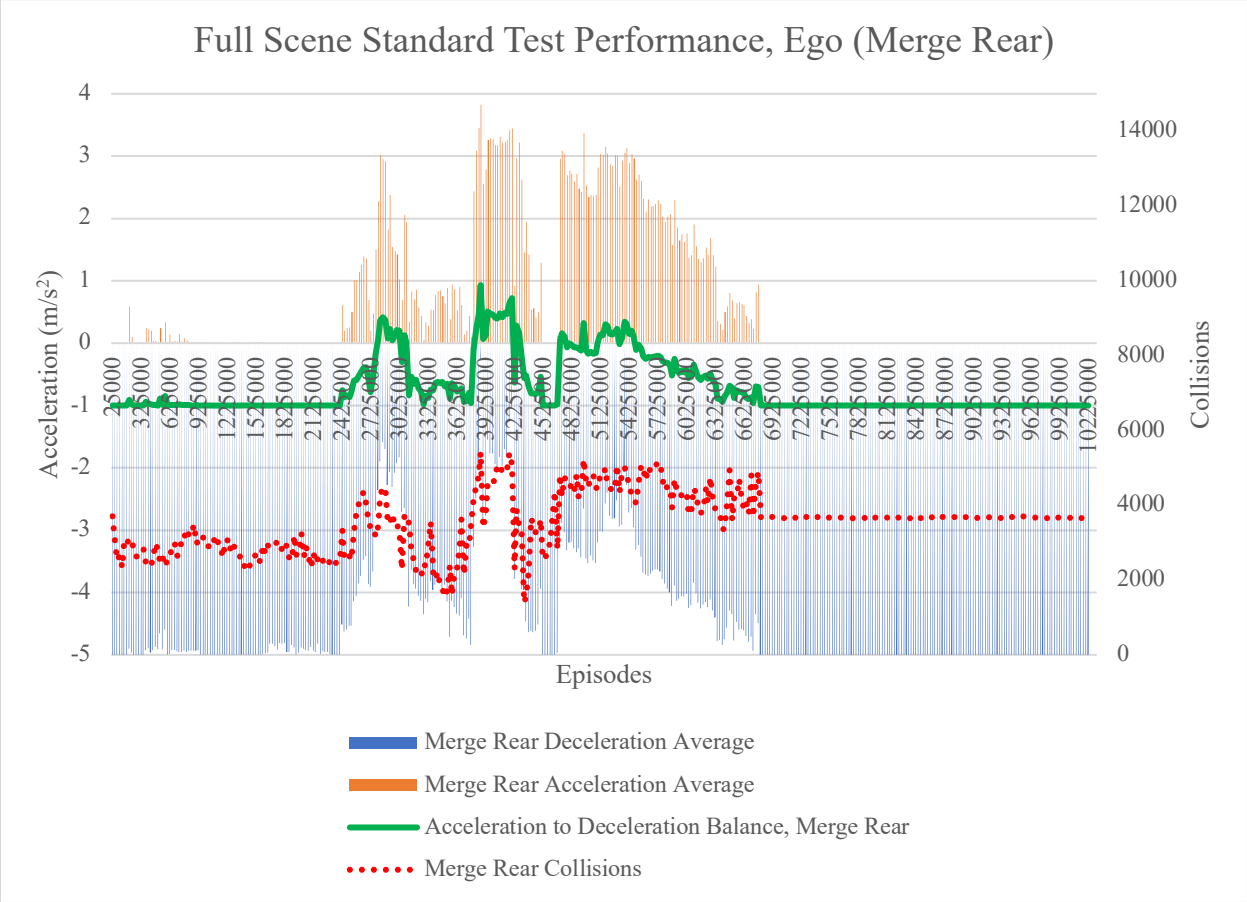


Figure 58: Multi-dimensional plot of acceleration and deceleration values over the course of training for the full scene. The plot is shown from a time perspective, meaning from left to right the episodes increase.

Figure 58 shows a multi-dimensional plot of the acceleration and deceleration occurrence and averages. The plot is the same representation of data as the two- and three-vehicle simulations. Similar to what was previously discussed in this section, there is a level of correlation between acceleration/deceleration occurrence (shown in the chart as balance) to collisions. There is also some level of correlation suggested by the plot for acceleration and deceleration average values to collisions.

Different here from the two- and three-vehicle simulations is the total number of episodes of training. Both previous scenarios trained to 2.5 million episodes in total, whereas the full-scene is trained to over 10 million episodes. Around 6.9 million episodes, the network appears to lock into its current network weights and does not change. The multi-dimensional plot shows the number of collisions settles to around 4,000 and does not deviate from there, expect for some noise due to randomness in the simulation. The value is not the worst, but it is about twice as bad as the best network. This is similar to how the first three lines of Table 16 are representative of an untrained or poorly trained network because it settles to the extreme deceleration (or acceleration) value. However, fundamental study and ideal results show that both acceleration and deceleration actions are necessary to achieve the best performance.

The right side of the plot suggests that the network has found a performance balance that it is unmotivated to deviate from. It is unclear if it is an inherent problem within the DDPG algorithm, DRL itself, or quite possibly a need for further tuning of the network. More than likely, there is a significant amount of tuning that could be performed, but quite often this tuning itself can be the focus of a whole Ph.D. thesis itself, so it is left for future work.

7.6 Standard Test Results

Figure 59 to Figure 61 show the test results for the full scene. The initial gaps in the figures are 5, 15, and 25 meters. Additional gap settings are shown in the Appendix Section A.8. These are the gaps between the merge and traffic vehicles at the start of the simulation. The gap between the merge vehicles is the same distance as the gap between the traffic vehicles. The lengths for the traffic and merge vehicles are fixed at 5 m, so the initial gap settings represent an in-lane gap of one, three, and five car lengths. A one-car gap would be challenging for a human to merge into. A three-car gap represents the smallest gap that most drivers are typically

comfortable maintaining while driving [64]. Throughout the simulation, the gap changes based on the acceleration (or deceleration) actions chosen at each step. A supplementary video of the simulation rendering output is available that shows the two merge vehicles interacting with the two traffic vehicles against all three traffic polices.

7.6.1 Constant Policy

The constant policy keeps a constant speed for the traffic vehicles during each episode. The speed only decreases when a rear traffic vehicle is below the minimum TIV threshold of 0.8 s. Maximum deceleration action of -5 m/s^2 continues at each time step until the TIV reaches its threshold. If the TIV is 0.8 s or greater, the traffic vehicle acts with zero acceleration and maintains speed.

Figure 59 shows the results tables for the three initial gap settings. On the left, the gap setting is at 5 m. As previously discussed, this is a challenging gap setting because the length of vehicles in test runs are all set at 5 m. Most of the table shows collisions at a 40 m goal position and below. The collisions are fewer as the initial position reaches 5 or -5 meters but increases as the starting differential initial positions reach 10 or -10 meters. Taking action to avoid a collision is easier at 5 meters differential because the end of the merge vehicle aligns with the end of the traffic vehicle. However, because the traffic vehicles regenerate, it becomes difficult to avoid collisions again. At a 10-meter differential, the middle point of the merge vehicle becomes aligned with the middle point of the traffic vehicle, so it is like a duplication of the zero starting position behavior. Thus, there are more collisions. Overall, the 5 m gap is very difficult because the merge vehicle is trying to insert itself between a series of vehicles that are closely following (tailgating) the one in front.

At the 15 m gap setting, the results are much better. In fact, they have achieved optimal performance when compared to the ground truth in Figure 29. These results show that optimal performance is still achieved despite the additional vehicles in the busy road scene.

Edge cases still exist despite the good overall performance. For example, the 25 m initial gap test case shows near-optimal performance, similar to the 15 m setting. However, there is a collision at the 100 m goal position and differential starting position of -20 m. This is a case that is usually easily handled by a well-trained network. However, this one cannot. Reviewing the specific set of actions for this combination provides insight into the behavior. The rear ego merge vehicle slows down earlier in the merge sequence but then tries to accelerate past the traffic vehicle near the goal and ultimately fails to avoid the collision. Using this standard test methodology, the edge case becomes apparent and indicates that there could be a more suitable network. The behavior also suggests that the combination of a long-term and short-term network could be advantageous.

Traffic_Reaction constant

Merge Collisions \/ Initial Gap		5										5 All	5 R	5 F	15									15 All	15 R	15 F	25									25 All	25 R	25 F	Total All
Goal Position ->		10	20	30	40	50	60	70	80	90	100				10	20	30	40	50	60	70	80	90	100				10	20	30	40	50	60	70	80	90	100				
Merge to Traffic Initial Position Delta	-20																																						
	-15														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-9	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	32%	0%	32%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-8	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	32%	0%	32%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-7	100%	100%	0%	100%	0%	0%	0%	0%	0%	0%	32%	11%	21%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-6	100%	0%	100%	100%	0%	0%	0%	0%	0%	0%	32%	21%	11%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-5	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
	-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	10%
	-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	20%
	-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	20%
	-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	30%	30%	0%	30%
	0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	30%	30%	0%	30%
	1	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	43%	0%	43%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	46%	0%	46%	100%	100%	100%	100%	0%	0%	0%	0%	0%	50%	0%	50%	46%
	2	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	43%	0%	43%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	46%	0%	46%	100%	100%	100%	100%	0%	0%	0%	0%	0%	50%	0%	50%	46%
	3	100%	100%	100%	0%	100%	0%	0%	0%	0%	0%	43%	11%	32%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	35%	0%	35%	100%	100%	100%	0%	0%	0%	0%	0%	0%	38%	0%	38%	38%
	4	100%	100%	0%	100%	100%	0%	0%	0%	0%	0%	43%	21%	21%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	23%	0%	23%	100%	100%	0%	0%	0%	0%	0%	0%	0%	25%	0%	25%	31%
	5	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	10%
	6	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
	7	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	8	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	9	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	40%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	13%
	10														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20																											0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total		100%	87%	62%	35%	11%	0%	0%	0%	0%	0%	29%	17%	12%	39%	35%	22%	9%	0%	0%	0%	0%	0%	0%	11%	5%	6%	40%	32%	21%	8%	0%	0%	0%	0%	0%	4%	11%	5%	5%	16%

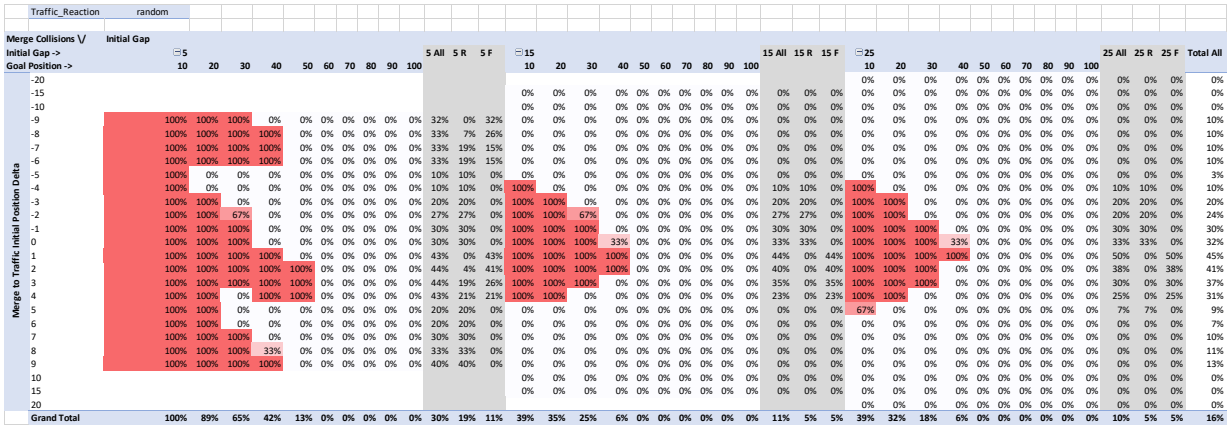
Figure 59: Full Scene constant policy standard test table.

7.6.2 Random Policy

Figure 60 shows the results from the random policy. In the random policy, the traffic vehicle takes randomly chosen actions at each time step of the simulation, just like the other

simulations with fewer vehicles. The random policy has similar performance overall to the constant policy. At the 5 m initial gap setting, the same collision performance pattern exists in the table, occurring for the same reason it occurs with the constant policy.

At a 15 m initial gap setting, the performance is very similar to the constant action policy. Overall, the performance is very good and comparable to the simulations with fewer vehicles. The 25 m initial gap setting shows very similar results too. This shows that despite the additional vehicles in the scene, the performance is still similar to a more simplified two-vehicle simulation.



Traffic_Reaction	random																																					
Merge Collisions \/	Initial Gap																																					
Initial Gap ->		5										5 All	5 R	5 F	15										15 All	15 R	15 F	25									25 All	25 R	25 F	Total All
Goal Position ->		10	20	30	40	50	60	70	80	90	100				10	20	30	40	50	60	70	80	90	100				10	20	30	40	50	60	70	80	90	100			
Merge to Traffic initial Position Delta	-20																											0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-15														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-10														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	-9	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	32%	0%	32%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-8	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	33%	7%	26%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-7	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	33%	19%	15%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-6	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	33%	19%	15%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	-5	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
	-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	10%
	-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	20%
	-2	100%	100%	67%	0%	0%	0%	0%	0%	0%	0%	27%	27%	0%	100%	100%	67%	0%	0%	0%	0%	0%	0%	0%	27%	27%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	24%
	-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	30%	30%	0%	30%
	0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	100%	100%	100%	33%	0%	0%	0%	0%	0%	0%	33%	33%	0%	100%	100%	100%	33%	0%	0%	0%	0%	0%	33%	33%	0%	32%
	1	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	43%	0%	43%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	44%	0%	44%	100%	100%	100%	100%	0%	0%	0%	0%	0%	50%	0%	50%	45%
	2	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	44%	4%	41%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	0%	40%	100%	100%	100%	0%	0%	0%	0%	0%	0%	38%	0%	38%	41%
	3	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	44%	19%	26%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	35%	0%	35%	100%	100%	100%	0%	0%	0%	0%	0%	0%	30%	0%	30%	37%
	4	100%	100%	0%	100%	100%	0%	0%	0%	0%	0%	43%	21%	21%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	23%	0%	23%	100%	100%	0%	0%	0%	0%	0%	0%	0%	25%	0%	25%	31%
	5	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	67%	0%	0%	0%	0%	0%	0%	0%	0%	7%	7%	0%	9%
	6	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
	7	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
	8	100%	100%	100%	33%	0%	0%	0%	0%	0%	0%	33%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	11%
	9	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	40%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	13%
	10														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	15														0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	20																											0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	Grand Total	100%	89%	65%	42%	13%	0%	0%	0%	0%	0%	30%	19%	11%	39%	35%	25%	6%	0%	0%	0%	0%	0%	0%	11%	5%	5%	39%	32%	18%	6%	0%	0%	0%	0%	0%	10%	5%	5%	16%

Figure 60: Full scene random policy standard test results for vehicle starting gaps of 5, 15, and 25 meters.

7.6.3 Reactive Policy

Figure 61 shows the reactive policy results. The performance of reactive policy is not ideal or as good as scenes with fewer vehicles. The results of the constant and random policies show that the ego DRL has been trained well and can avoid almost all collisions, except those that are unavoidable. This means the poor performance shown in Figure 61 comes from the traffic network. As the state set and number of active agents increase, so does the complexity of finding better performing networks. Finding a better performing reactive traffic policy network is possible because they already exist for other variants. Training neural networks can be an arduous task that requires delicate tuning of the network, parameter optimization, and other

techniques to find the best results. It takes an incredible amount of time to perform, but it adds little value to the body of work already presented here and is left for future work.

Traffic_Reaction: reactive

Merge Collisions \/ Initial Gap — Merge to Traffic Initial Position Delta

Initial Gap ->	5										5 All	5 R	5 F
Goal Position ->	10	20	30	40	50	60	70	80	90	100			
-20													
-15													
-10													
-9	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	54%	0%	54%
-8	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	43%	0%	43%
-7	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	32%	0%	32%
-6	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	21%	0%	21%
-5	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	11%	0%	11%
-4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-3	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%
-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%
0	100%	100%	100%	0%	100%	100%	0%	0%	0%	0%	58%	39%	19%
1	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	64%	0%	64%
2	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	54%	0%	54%
3	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	43%	0%	43%
4	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	32%	0%	32%
5	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	21%	0%	21%
6	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%
7	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%
8	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%
9	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	40%	0%
10													
15													
20													
Grand Total	**95%**	**83%**	**62%**	**35%**	**23%**	**11%**	**0%**	**0%**	**0%**	**0%**	**31%**	**11%**	**20%**

Initial Gap ->	15										15 All	15 R	15 F
Goal Position ->	10	20	30	40	50	60	70	80	90	100			
-20													
-15	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	23%	0%	23%
-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%
-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%
-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%	0%
0	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	62%	10%	52%
1	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	81%	0%	81%
2	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	69%	0%	69%
3	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	69%	0%	69%
4	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	46%	0%	46%
5	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	46%	0%	46%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20													
Grand Total	**50%**	**43%**	**31%**	**27%**	**18%**	**14%**	**5%**	**0%**	**0%**	**0%**	**19%**	**4%**	**15%**

Initial Gap ->	25										25 All	25 R	25 F	Total All
Goal Position ->	10	20	30	40	50	60	70	80	90	100				
-20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	11%
-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	17%
-8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	14%
-7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
-6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
-3	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%	13%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	20%
-1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	27%
0	100%	100%	100%	0%	0%	0%	0%	0%	100%	100%	70%	40%	30%	63%
1	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	63%	0%	63%	69%
2	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	63%	0%	63%	62%
3	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	50%	0%	50%	54%
4	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	50%	0%	50%	42%
5	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	38%	0%	38%	35%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	13%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	**38%**	**35%**	**24%**	**17%**	**9%**	**0%**	**0%**	**0%**	**4%**	**4%**	**14%**	**4%**	**10%**	**21%**

Figure 61: Full Scene reactive policy standard test tables for 5, 15, and 25-meter initial starting gaps between vehicles.

7.7 Full Scene Discussion

The full-scene results are very good overall. The ego vehicle performs nearly optimally against the constant and random policies in reasonable gap settings (15 or greater). The ego DRL network in the full scene has learned to perform just as well as the DRL networks in scenes with fewer vehicles. The progressive scaling up of the road scene was successful. During the scaling, much was learned about the model, state parameters, and variables to achieve good results. Judging the results was enabled by the performance evaluation framework. The framework has also shown success in scaling along with the model.

Chapter 8 Contribution, Implications, and Future Work

This dissertation presents a model for merging that enables the study of its essential behaviors. Autonomous driving has not yet reached L5 and needs an automated highway merge function (amongst other improvements) to achieve L5 functionality. The broad domain of my work is a multi-agent deep reinforcement learning. The specific application is highway lane merging as a subset of automated driving. The foundational form of the model presented in this dissertation is a two-vehicle representation of a taper-type merge. It was studied from a very fundamental perspective using a traditional grid-game approach with the T-grid model. This model produced four fundamental propositions about merge behavior and a better understanding of the essence of merging.

8.1 Contributions

There are several contributions in this work. The main contributions are three-fold:

1. Fundamental on-ramp study that shows limitations always exist and collisions can be unavoidable.
2. A test framework for evaluating merging performance that compares different approaches in a standard way.
3. Reinforcement Learning simulation that shows nearly perfect performance against the ground truth.

Work starts with a simple two-vehicle scene, then scales up to a full multi-vehicle busy merge scene. The results hold true during scaling. The model itself is extremely important and its

tuning was an on-going sub-theme during this research. While most of the focus is on the successes and what was done to get there, insight into trialing of different states and setup is given in Appendix B.

8.2 Domain, Application, and Approach

State-of-the-art systems still show many issues with vehicles colliding with stationary objects due to missed detections or confusion in situations that even novice human drivers would find easy to negotiate. To generalize, often AI-based approaches can match patterns that give seemingly super-human accuracy, but they can also be flawed in simple ways where humans are not.

Testing standards relatively non-existent for AD, especially in comparison to other safety systems like braking and steering. Standards include guidelines, codes, and standards for parameters like stopping distance, crashworthiness, flammability, and even tire construction. Testing in a systemic way, against commonly accepted standards is necessary to drive safe adoption. These standards are usually released by or in conjunction with government agencies, so their development and refinement can take time. It has taken decades for other standards to reach their level of maturity, but it is important that they exist as a basic foundation and fundamental backstop that can be improved upon.

Automated driving is expected to save lives. Judging whether this work accomplishes this goal is difficult since a basis to compare does not exist. However, the evaluation framework presented in this work does make it possible to understand and compare performance. Using this framework, it could be possible to use real-world datasets to make judgements.

The work in this dissertation is a multi-agent approach where multiple intelligent agents act in a common environment. There are three traffic action policies used to train and test the ego

agent: constant, random, and reactive. The random and constant policies are considered fixed policies in a multi-agent setting, whereas the reactive policy is a truer representation of multi-agent since there are multiple decision-making agents within a common environment. This multi-agent approach makes the simulator a necessary and essential tool to safely train and test virtually instead of putting vehicles and people in harm's way. Because of the reactive nature of a multi-agent setting where agents make decisions based on other agents actions and not the environment itself, pre-defined datasets like NGSIM are not of much use.

8.3 Results and Discussion

The T-grid approach studied a very simplified approach to merging that captured the essence of the behavior. It showed collisions are unavoidable while vehicles are in close proximity and do not have accurately predictable knowledge of the other vehicles behavior. Extending this to the initial position differential versus goal framework shows the same quagmire. Conversely, the grid game also showed that if movement of each agent can be accurately controlled prior to reaching the merge point, collisions can be avoided. This means that a master controller that tracks the presence, positions, and velocities of all vehicles could control these vehicles to ensure collisions do not happen. However, infrastructure today does not universally exist to perceived, detect, monitor, and measure vehicles, so the evolution to communicating and controlling the vehicles is only a future dream. Even if a system could be designed and implemented, it could be met with incredible resistance since it would take away drivers freedom of choice – an often sensitive topic that quickly pushes into the realm of political science, sociology, psychology, and other areas far outside the scope of my work which is focused more on the setup, possibilities, and limitations of the physical mathematical variables of merging. The grid game also showed that a typical merge can be simplified to the well-known

game-of-chicken where individual driver (or controller) behavior and risk tolerance become the over-riding principle in merging.

With the fundamental study complete, it was time to simulate. The simulation work began with a two-vehicle approach, just like the T-grid study. However, the simulation was more closely based on a real-world merge scenario, having vehicle lengths, ramp lengths, speed variation, and other more parameters more sophisticated than the grid world study.

A Q-learning learning simulation of the merge model is presented. The Q-learning approach studied the model from a multi-agent perspective and compared an ego merge vehicle trained through reinforcement learning against several behavior representations of the traffic vehicle with both single-action and joint-action learning. The results show that the agents learned behavior to avoid collisions. The results aligned with the fundamental propositions found through the T-grid study. During the study, a framework emerged for standard comparison of performance between the representation; a framework to compare performance as the starting positions vary between the two agents.

Q-learning is a straightforward tabular approach to reinforcement learning where the data tables for the state-action sets can be readily reviewed. This proved helpful because it showed that patterns were forming in the data after training, supporting the idea that a more sophisticated function approximation approach could produce better results.

Q-learning shaped some fundamental aspects of the merge model and study. The individual policies of a random, constant, and reactive actor were created here. A single-action-learner was compared to a joint-action-learner. The framework of a differential starting position between the ego and traffic vehicle took shape during this study too. All on-ramp lengths were held to a constant 100m and the results still showed collisions. The states and actions were set in

this phase. A relative closing gap and relative closing speed is a new approach to the merging study that continued as a theme throughout the work. A time-to-position variable was created to enable the RL-network learn how to behave differently based on distance to the goal position. Reasonable physical values and limits were derived for the physical setup based on typical vehicle performance and road parameters. Motion equations and associated code was developed based on constant-acceleration formulas and short time update intervals to ensure "re-planning" accuracy.

The results for Q-learning showed that the RL-network learned how to avoid collisions. The results were not ideal, but there was clearly some reactive behavior learned by the network which showed promise for continued development. The propositions found through the fundamental study also appeared in the simulation. Some with stronger relationships to the results than others, but their presence was clear and having them led to a much better understanding of the behavior of the simulation as well as the validity of the results.

A limitation of Q-learning is that the state and action space was discrete. Many trials were conducted with different parameter sets of various sizes and increments between the values. No common improvement themes became apparent, but it was clear that increasing the discretization space was leading to very large Q-learning tables and they were trending to grow beyond the limitations of MATLAB and the relatively standard laptop computer used to conduct the experiments.

The Q-learning approach was also applied to the NGSIM dataset, using the simplified two-vehicle setup. The results were far better than any other published work, but the simplified setup likely contributed to the low collisions rate. However, as mentioned previously, the lack of reactiveness of a dataset is a questionable approach to a function like merging. In addition, the

NGSIM dataset is based on a busy road scene with congested traffic that is travelling much slower than normal highway speeds. This is generally different than the scope of my study, but future work could be done using my work as a foundation to develop controllers that operate at these slows speeds and increased congestion.

Stacking the building blocks together of the T-grid study supported by the Q-learning simulation, combined with the limitations of Q-learning itself, as well as the presence of patterns forming in the data, it was time to embark upon a more sophisticated learning approach, DRL. In recent years, Deep Learning has increased in popularity. It shows very promising results in many applications, especially those replicating physical behavior or action and movement strategy.

A deep reinforcement learning (DRL) approach was applied to the merge model. The MATLAB-based Q-learning simulator was re-written using a more state-of-the-art approach with Python, TensorFlow, and OpenAI Gym. The basic formulas, parameters, and states remain the same, but this new DRL approach used a continuous state-action space with trained a state-of-the-art DDPG DRL network. The DRL approach studies the same scenarios from the Q-learning approach for traffic and single- versus joint-action. The results showed significant improvement with DRL performance compared to the Q-learning approach. In addition, the framework for standard comparison was expanded by adding another dimension, the distance away from the merge point.

The 100m ramp length used to measure performance in the two-vehicle Q-learning setup yielded no collisions at 100m. The evaluation framework was then expanded to show how short of an on-ramp could there be before collisions occurred, even though a 100m on-ramp length was already shorter than well-adopted guidelines and regulations in the U.S.

To understand what the maximum performance could be, ground truth tables were developed where the limits of acceleration and other parameters were used to find the combinations where collisions were unavoidable. And, yes, collisions are unavoidable if there is a very short ramp length and the vehicles start out on that ramp already longitudinally overlapped. As a result, the performance of the DRL approach shows near-optimal performance by using the standard test framework for comparison.

The DRL network was trained using a combination of the three traffic action policies (constant, random, and reactive) originally used in the Q-learning simulation. The relative gap and relative speed state setup was also reused for the DRL approach. This mixture training enabled the ego (merge) vehicle to learn how to avoid collisions with performance that nearly identically matched the ideal case. Testing performance against all three policies (constant, random, and reactive) showed this nearly ideal performance. The joint-action-learner had slightly better performance than the single-action-learner too.

Eventually, the training performance testing was standardized and a MATLAB script was developed to automate data scraping of key summary parameters from the test data sets. This created the ability to quickly find the best performing DRL network for detailed performance trials.

Further evaluation of the training performance testing summaries showed correlation of acceleration and deceleration values to collision performance. However, it mostly showed what parameter settings produced the highest collisions, but it was unable to definitively identify specific parameter values that produced the most ideal results, indicating that the best performance is based on more than a single factor. The training results as well as the

performance testing results again showed alignment with the Propositions developed using the T-grid.

Despite the nearly ideal performance of the DRL simulation, questions remained about whether the approach could be scaled up to a full busy road scene or if the performance would persist. A jump directly to a busy road scene with multiple vehicles in both the merge and traffic lanes could have been possible, but it was important to keep with the theme of stepwise evolution of the model to ensure that foundational learnings could be continued. Afterall, the research was not simply about applying DRL to find a way to automate merging, but it was more about fundamentally and thoroughly understanding merging from a multi-agent perspective and applying the appropriate model to advance the understanding and knowledge of automated driving while studying highway merging.

An intermediate step to the full scene scenario was developed. A single vehicle was added to the traffic lane, thus creating a model for two traffic vehicles and one merge vehicle. The states remained largely the same, but obviously some needed to be added to track the extra vehicle. The training policies, motion formulas, DRL network, and vehicle parameters remained largely the same. Much of the algorithm was re-written to add the additional vehicle, but the fundamental behavior at a block-diagram level remained much the same as the two-vehicle. The MATLAB best-network selection approach was utilized and the results were nearly ideal, just like the two-vehicle setup. The test framework evolved again, this time to show multiple initial position differential-goal tables with settings for each of the initial gaps between the traffic vehicles. Again, against all policies, the results showed nearly ideal against the ground truth with all the collisions being well under the 100m on-ramp length. The scaling was successful.

Finally, the model was scaled up to a full-scene scenario where there are two vehicles in the merge lane and two or more in the traffic lane. This model can be applied to any real-world merge scene. The standard test results show nearly ideal performance overall. However, there are some instances where there are collisions at the 100m goal position against both the constant policy and the reactive policy. The reactive policy uses a separate, but similar DRL network to select actions for the traffic vehicle. The results table shows that the performance is less than ideal in more combinations than the ground truth or the two- or three-vehicle simulations. This generally indicates that the traffic network still has room for improvement in training. The less-than-ideal performance in the constant policy may also indicate that there could be training improvement for the ego vehicle. As more vehicles are added, there become more options for training and evaluation. The approach taken here in training is to have both the rear ego merge vehicle and the front merge vehicle train the same DRL network. Another approach could be to use only the rear ego merge vehicle to train the network. There may also be further tuning of the DRL network or adjustment of the layer count and configuration to improve performance. However, this tuning can become very time-intensive very quickly and is left for future work. Fundamentally, the results of this approach overall are close to the ideal ground truth and show a solid proof-of-concept for the approach.

8.4 Future Work

An obvious next step for this research is to apply it in the real world. The full-scene work has all the necessary and essential states, parameters, and calculations to implement the trained algorithm as a function for longitudinal position control in an autonomous vehicle. The theory shows that this should be a robust solution but putting it into practice with real-world testing will help show the work's true viability and limitations.

It is important to consider that the models presented here are simplified and idealistic. There is no computational or communication latency, and no sensor or perception errors. These concerns and more all exist today and could further complicate the introduction of a truly robust automated highway lane merging function.

Throughout the study, certain choices are made about the rewards and penalties of merging. They certainly have influence over the behavior. Mine are kept relatively simplistic and focuses on with large penalties for collisions, lower rewards for merging, and small penalties for action-taking. Rewards could be applied to other variables, like speed or distances to other vehicles, but the intent was to keep the reward strategy somewhat simple and straight-forward. However, study and trials on other strategies and schemes may prove fruitful towards improved behavior, efficiency, or training cycle reduction.

To build a more robust controller, the longitudinal control approach presented here can coupled with other more sophisticated features and functionality, like lateral control, that further assist collision avoidance. In a real-world scenario, vehicles typically have some amount of control of their lateral position within the lane. This lateral position can make a tremendous difference between colliding with another vehicle or not. For example, it may make more sense for a vehicle to cross the lane marking into the shoulder to avoid collision with another vehicle. This is a simple, obvious, and effective strategy a human can use in an emergency to prevent collisions. While this approach may not always be possible and it also increases complexity, it is one example of an augmented approach that could produces better results in some situations.

A limitation exists within the standard test results presented throughout this work. The standard testing starts both vehicles at the same velocity, however, a velocity differential could exist. An alternate standard framework for this case could be a TTP variable measuring the time

to the merge point for both vehicles. An equal TTP would be equivalent a an equal (meaning zero) starting differential position. However, this broken symmetry with different speeds could create an easier situation for the controller to deal with because of the fundamental finding in Proposition IV (symmetry/similarity of vehicle behavior could lead to similarity in action taking, resulting in greater chance of collision). Predictive models for cut-in and cut-out are studied by Hu et. al. [87] and could potentially be considered to adapt the standard framework using different parameters.

Standards for road design show that on-ramp lengths are typically much greater than 100m. However, this study began showing ideal performance at 100m and tens of meters below with the two-vehicle trial of the DRL approach. While this DRL approach is trained at greater lengths, testing at greater ramp lengths is omitted. If there is opportunity to observe a scene of vehicles at greater than 100m away from an on-ramp while controlling longitudinal position, a simple controller that positions ego at least 5m in front of or behind the closest vehicle (or right in the longitudinal middle of two close traffic vehicles) traveling a similar speed would be a solid strategy in pre-setting the scene to avoid a collision. Since all the standard results shows that collisions can generally be avoided if they are aligned in a tip-to-tail or greater condition. This or other simple control schemes can also be put in place to try to avoid collisions or better longitudinally position as the vehicles are further out to reduce the likelihood of collision. Additionally, the results are one portion of the contribution to the body of knowledge. The results evaluation framework itself shows that there are limitations in any approach and can identify specific areas where greater precautionary measures should be taken to make these systems safer.

A multi-agent approach could be very useful in the situation represented by the NGSIM dataset. The NGSIM dataset is a very slow-moving set of highly congested traffic. As humans, we can easily see challenges with this setup that would make a multi-agent setup (where actions and reactions are considered) a better framework. Anyone who has experienced a traffic jam with lane reductions and additions has likely experienced that these situations can be very dependent upon behavior of other drivers, including their driving style, as well as human interaction either through direct verbal communication, hand gestures, or intention indicating behaviors like slowly moving towards an adjacent lane to communicate to other vehicles that you intend to act. This can often be met with aggressiveness, defensiveness, or willingness of the other drivers to cooperate. This, quite simply, is not behavior that can be accounted for in a dataset like NGSIM and would require real-world testing or sophisticated simulation.

8.5 Conclusion

Despite the excellent performance of the DRL approach, the stark reality of the foundational work in Chapter 2 shows that collisions are simply unavoidable if vehicles are too close, or their capability is insufficient. This reality is supported by the simulation work and proven again in the ground truth table in Figure 31 where both the traffic and merge vehicles actively work to avoid a collision.

There are options to truly avoid all collisions, but they extend well beyond controlling the position of a vehicle. For example, roadway design, master-controlled turn-taking, strictly controlled speed restrictions, or other brittle measures are impositions that could be explored to ensure collisions never occur. Still, each could be very difficult or even impossible to implement. However, all these topics are well outside the scope of my research, and many are outside of engineering. Some even lead to philosophical debate about freedom of choice. The finding that

collisions are unavoidable is likely my most significant contribution to the body of merging research, and it was an unexpected outcome when I began exploring this topic years ago.

Appendices

Appendix A Additional Tables and Figures

This appendix shows additional tables and figures which provide further detail of the DRL results. The sections follow the structure of the paper with separate vehicle simulations. Reference and further explanation are generally given in the corresponding chapter.

A.1 Two-Vehicle Training Test Performance

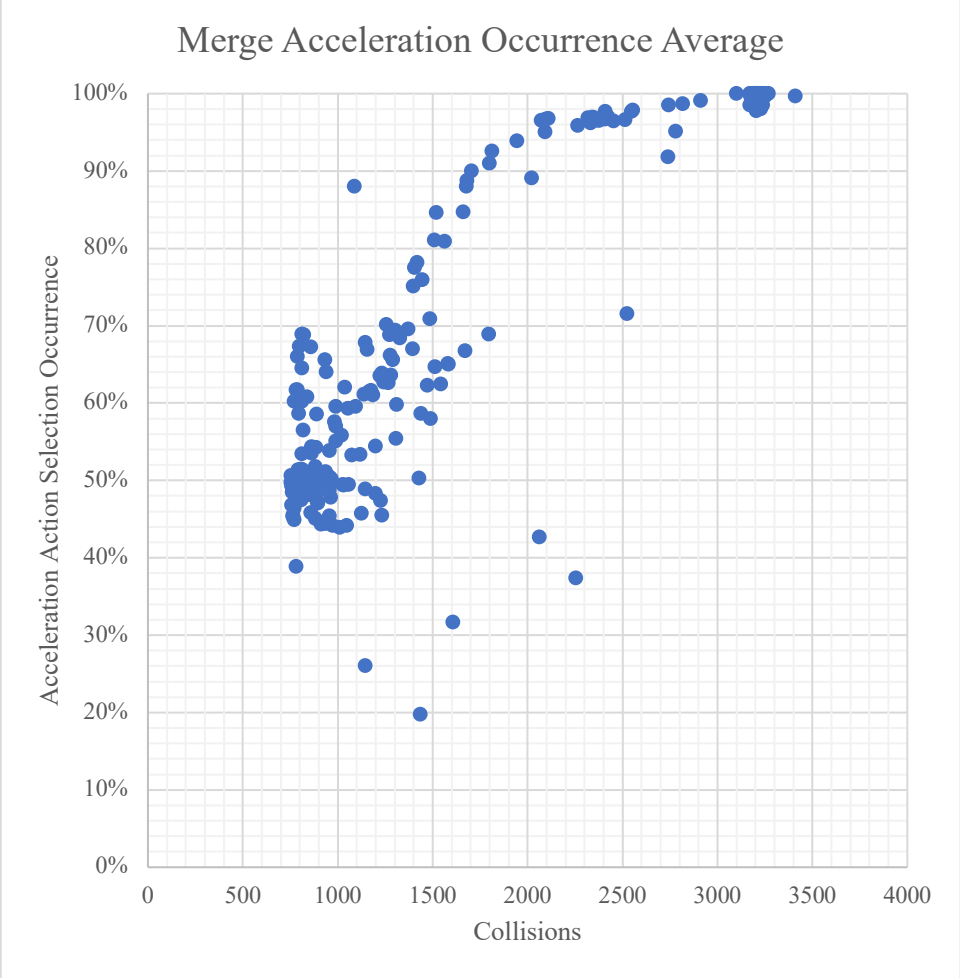


Figure 62: Two-vehicle simulation Ego merge acceleration occurrence. Data points show the average occurrence of an acceleration action selection within each episode.

Average of Deceleration Occurrence	Goal Position										
Ego Initial Position	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-15	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-10	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-8	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-7	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-6	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-5	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-4	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-3	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
0	0%	0%	0%	0%	0%	0%	11%	11%	0%	0%	2%
1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	**48%**	**48%**	**48%**	**48%**	**48%**	**48%**	**48%**	**48%**	**48%**	**48%**	**48%**

Figure 63: Two-vehicle Ego occurrence of deceleration action selection by initial position and goal position. Results are based on the SAL example with a selected network of 1.04×10^6 training episode. The Ego vehicle has learned to decelerate for almost all combinations with Ego initial positions of -1 m or less. Since the network only selects acceleration or deceleration values, acceleration occurrence is the compliment of deceleration occurrence. This aligns with Proposition III: it is advantageous for a lagging vehicle to decelerate and for a leading vehicle to accelerate.

A.2 Three-Vehicle Traffic DRL Network Diagram

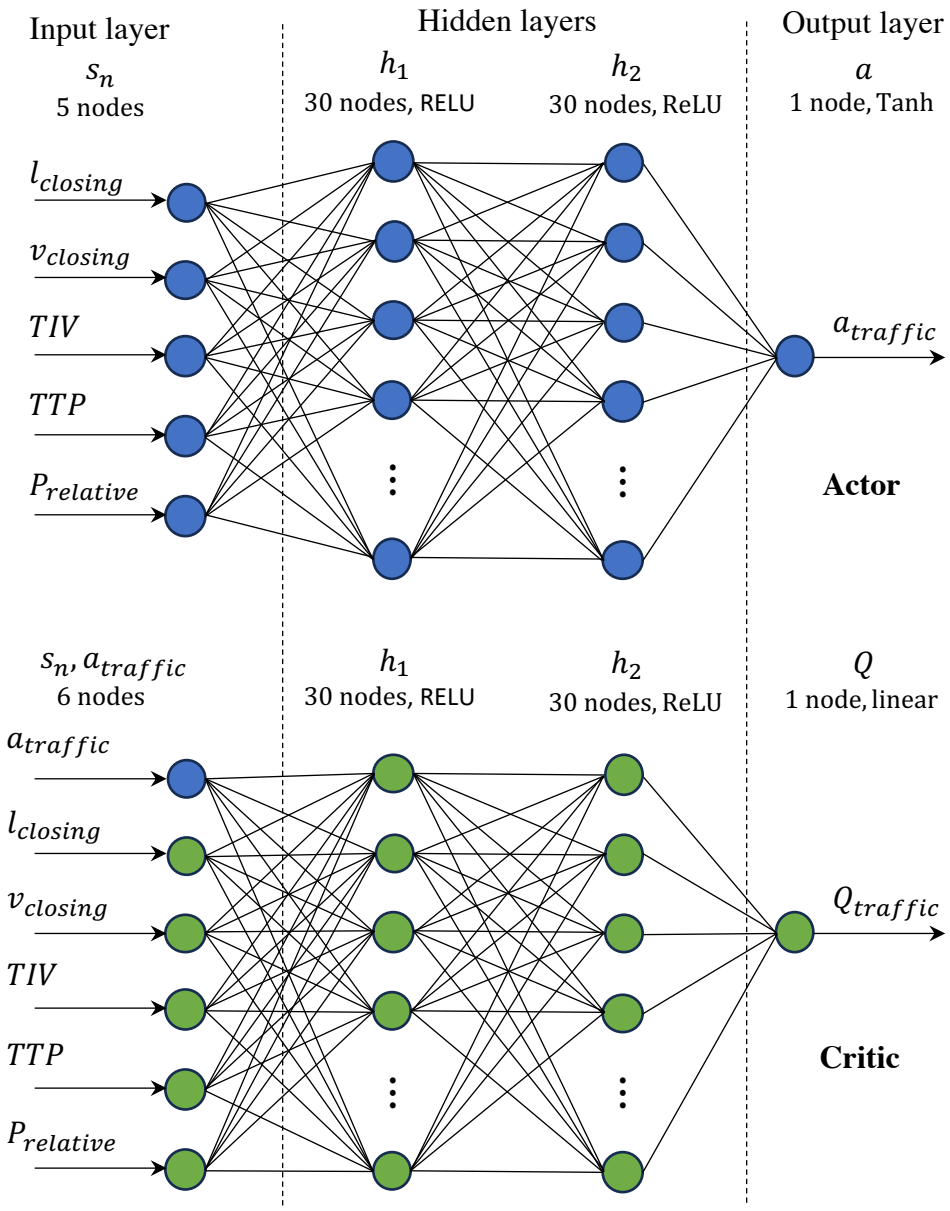


Figure 64: Three-vehicle DRL network for traffic vehicles.

A.3 Three-Vehicle Best Network Performance Testing Tables

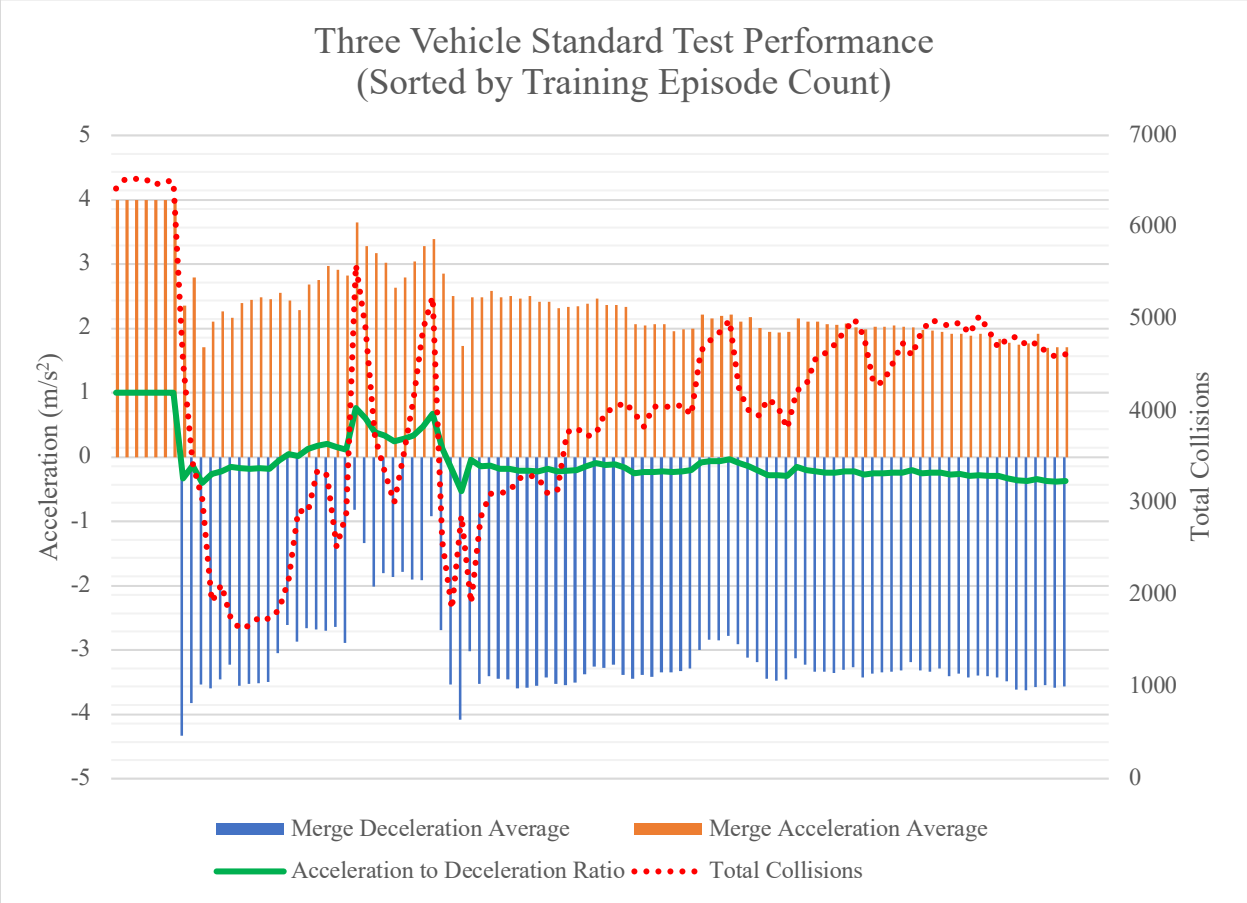


Figure 65: Three-vehicle standard testing sorted by training order. This is the same plot as Figure 49, but with a different sort order. Horizontal axis values indicate the test number, each representing 25K episodes. The leftmost point is at the first testing instance at 25K episodes, ascending to the rightmost value at 2.5 million episodes.

Figure 65 shows a much more erratic view of the training performance than Figure 49. The data is the same in both plots, but the sort order is different. Figure 65 shows the training progression with increasing episodes, whereas Figure 49 is sorted by lowest to highest collision performance. Figure 65 shows that the network takes time to learn action values other than the extreme limit of acceleration. As it begins to learn to select both acceleration and deceleration actions, the number of collisions rapidly decreases to the lowest number of total collisions around the 14th test which is at 350K episodes.

Training continues after the 350K best performer. The performance appears erratic with total collisions shifting between high and low as episodes increase. There are two peaks of high total collisions around the 25th and 34th tests. These peaks correlate with greater acceleration bias and higher average acceleration values as well as lower average deceleration values. In addition, the slightly increasing bias towards greater deceleration starting around the 49th test group corresponds to an increase in total collisions, as observed in Figure 49.

A secondary indicator of poor performance is when all values converge to a high or low extent for acceleration or deceleration. Fundamental studies and prior results show that both acceleration and deceleration are needed to achieve good or nearly ideal performance. Figure 65 show the three-vehicle training performance sorted by increasing number of episodes. Table 18 in Section A.3 of Appendix A shows a table of the detailed values for reference. In the early training save points, the ego network has not fully learned how to avoid collisions but has learned to only choose acceleration values at the high limit of the acceleration range, 4.0 m/s^2 – representing an untrained network.

Table 18: Three-vehicle training performance, sorted by episode count. Poor collision performance is shown in the table from the beginning of training up to 175K episodes.

Episodes	Merge Decel Average (m/s^2)	Merge Accel Average (m/s^2)	Merge Decel Occurrence	Merge Maintain Occurrence	Merge Accel Occurrence	Total Collisions
25000	0	4	0.0%	0.0%	100.0%	6423
50000	0	4	0.0%	0.0%	100.0%	6526
75000	0	4	0.0%	0.0%	100.0%	6525
100000	0	4	0.0%	0.0%	100.0%	6532
125000	0	4	0.0%	0.0%	100.0%	6459
150000	0	4	0.0%	0.0%	100.0%	6509
175000	0	4	0.0%	0.0%	100.0%	6496
200000	-4.3299	2.3526	64.1%	0.0%	35.9%	4538
225000	-3.8241	2.7919	53.4%	0.0%	46.6%	3435
250000	-3.5381	1.7098	72.7%	0.0%	27.3%	3045
275000	-3.5926	2.1037	67.9%	0.0%	32.1%	1923
300000	-3.4531	2.2636	64.6%	0.0%	35.4%	2112
325000	-3.2248	2.1612	61.9%	0.0%	38.1%	1738
350000	-3.5546	2.3912	65.4%	0.0%	34.6%	1626
375000	-3.5285	2.4422	67.2%	0.0%	32.8%	1671
400000	-3.5153	2.4848	66.3%	0.0%	33.7%	1760
425000	-3.4941	2.4523	66.9%	0.0%	33.1%	1734
450000	-3.0462	2.5489	64.0%	0.0%	36.0%	1837
475000	-2.6084	2.4322	60.2%	0.0%	39.8%	2165
500000	-2.8701	2.2805	62.3%	0.0%	37.7%	2919
525000	-2.6582	2.6782	56.8%	0.0%	43.2%	2898
550000	-2.6799	2.75	54.1%	0.0%	45.9%	3358
575000	-2.6985	2.9657	49.3%	0.0%	50.7%	3300
600000	-2.6452	2.9113	53.5%	0.0%	46.5%	2519
625000	-2.8859	2.8182	56.5%	0.0%	43.5%	2882
650000	-0.82024	3.6446	21.4%	0.0%	78.6%	5590
675000	-1.3386	3.2827	25.9%	0.0%	74.1%	4912
700000	-2.0122	3.1736	40.6%	0.0%	59.4%	3717
725000	-1.8014	3.0157	49.3%	0.0%	50.7%	3342
750000	-1.8671	2.6351	51.8%	0.0%	48.2%	3003
775000	-1.7867	2.7908	49.9%	0.0%	50.1%	3542
800000	-1.9073	3.0369	47.7%	0.0%	52.3%	4124
825000	-1.9141	3.2747	37.0%	0.0%	63.0%	4890
850000	-0.92343	3.3912	24.7%	0.0%	75.3%	5245
875000	-2.693	2.8486	52.9%	0.0%	47.1%	2587
900000	-3.5357	2.5014	68.1%	0.0%	31.9%	1849
925000	-4.0859	1.7296	78.2%	0.0%	21.8%	2866
950000	-3.0177	2.4784	64.7%	0.0%	35.3%	1909
975000	-3.5269	2.4826	71.1%	0.0%	28.9%	2841
1000000	-3.409	2.5861	71.2%	0.0%	28.8%	3098

A.4 Three-Vehicle Simulation Best Network Selection Additional Details

The following are detailed plots of the average deceleration selection occurrence for the typical initial position – goal position table format. Graphs are shown for the combined setup as well as the individual gaps for the traffic.

Traffic_Gap	(All)										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	78%	76%	79%	81%	80%	78%	78%	77%	74%	72%	77%
-15	74%	76%	76%	82%	82%	81%	79%	74%	71%	70%	77%
-10	88%	86%	82%	87%	87%	83%	78%	74%	71%	70%	81%
-9	87%	82%	81%	83%	84%	81%	77%	74%	70%	69%	79%
-8	93%	86%	86%	85%	85%	83%	78%	72%	71%	69%	81%
-7	98%	92%	89%	87%	86%	83%	78%	73%	70%	69%	82%
-6	100%	95%	91%	90%	86%	84%	78%	74%	72%	71%	84%
-5	100%	97%	95%	93%	87%	84%	79%	76%	73%	71%	85%
-4	100%	98%	97%	95%	91%	85%	79%	77%	74%	73%	87%
-3	100%	100%	99%	97%	94%	86%	81%	78%	75%	73%	88%
-2	100%	100%	100%	98%	96%	88%	83%	80%	76%	75%	90%
-1	100%	100%	100%	100%	98%	92%	84%	81%	78%	76%	91%
0	100%	100%	100%	100%	98%	95%	88%	83%	79%	77%	92%
1	15%	10%	10%	13%	15%	23%	27%	30%	31%	36%	21%
2	20%	14%	13%	15%	17%	25%	27%	30%	31%	33%	23%
3	24%	21%	17%	16%	17%	28%	30%	30%	33%	34%	25%
4	29%	24%	18%	18%	20%	29%	31%	31%	34%	34%	27%
5	36%	28%	21%	22%	24%	30%	32%	31%	34%	34%	29%
6	43%	31%	26%	25%	26%	31%	33%	32%	35%	35%	32%
7	43%	37%	32%	29%	29%	34%	34%	34%	36%	35%	34%
8	43%	39%	33%	31%	33%	37%	35%	34%	37%	36%	36%
9	43%	43%	37%	35%	35%	36%	37%	36%	37%	37%	38%
10	33%	33%	29%	28%	28%	30%	30%	31%	34%	36%	31%
15	40%	40%	32%	30%	30%	30%	33%	35%	40%	39%	35%
20	25%	25%	25%	25%	28%	24%	33%	40%	44%	39%	31%
Grand Total	**65%**	**62%**	**59%**	**59%**	**58%**	**59%**	**57%**	**55%**	**55%**	**55%**	**58%**

Figure 66: Three-vehicle Ego deceleration occurrence average values by start position and goal position averaged over all traffic gap values.

Traffic_Gap	100										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	100%	100%	100%	93%	91%	89%	89%	87%	86%	93%
-15	100%	100%	100%	100%	100%	90%	92%	93%	86%	85%	95%
-10	100%	100%	100%	100%	100%	100%	95%	91%	92%	93%	97%
-9	100%	100%	100%	100%	100%	100%	91%	95%	93%	92%	97%
-8	100%	100%	100%	100%	100%	100%	92%	95%	94%	93%	97%
-7	100%	100%	100%	100%	100%	100%	95%	93%	91%	92%	97%
-6	100%	100%	100%	100%	100%	100%	100%	93%	94%	95%	98%
-5	100%	100%	100%	100%	100%	95%	99%	95%	93%	93%	98%
-4	100%	100%	100%	100%	100%	100%	100%	94%	92%	93%	98%
-3	100%	100%	100%	100%	100%	100%	96%	95%	94%	90%	98%
-2	100%	100%	100%	100%	100%	100%	96%	96%	93%	93%	98%
-1	100%	100%	100%	100%	100%	100%	97%	95%	94%	94%	98%
0	100%	100%	100%	100%	100%	100%	100%	96%	93%	94%	98%
1	0%	0%	0%	0%	0%	0%	0%	0%	8%	41%	5%
2	0%	0%	0%	0%	0%	0%	0%	0%	10%	17%	3%
3	0%	0%	0%	0%	0%	0%	0%	0%	10%	17%	3%
4	0%	0%	0%	0%	0%	0%	0%	0%	11%	18%	3%
5	0%	0%	0%	0%	0%	0%	0%	0%	11%	19%	3%
6	0%	0%	0%	0%	0%	0%	0%	0%	13%	20%	3%
7	0%	0%	0%	0%	0%	0%	0%	0%	13%	20%	3%
8	0%	0%	0%	0%	0%	0%	0%	0%	14%	22%	4%
9	0%	0%	0%	0%	0%	0%	0%	5%	15%	22%	4%
10	0%	0%	0%	0%	0%	0%	0%	5%	15%	25%	4%
15	0%	0%	0%	0%	0%	0%	8%	22%	26%	34%	9%
20	0%	0%	0%	0%	0%	0%	17%	39%	47%	47%	15%
Grand Total	**52%**	**52%**	**52%**	**52%**	**52%**	**51%**	**51%**	**52%**	**56%**	**60%**	**53%**

Figure 67: Three-vehicle Ego deceleration occurrence average for 100 m initial traffic gap.

Traffic_Gap	50										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	90%	88%	84%	77%	72%	66%	64%	62%	62%	77%
-15	100%	100%	100%	92%	89%	81%	76%	68%	65%	63%	83%
-10	100%	100%	100%	100%	97%	91%	82%	78%	70%	65%	88%
-9	100%	100%	100%	100%	100%	92%	87%	78%	72%	68%	90%
-8	100%	100%	100%	100%	100%	96%	87%	79%	73%	67%	90%
-7	100%	100%	100%	100%	100%	97%	90%	81%	73%	68%	91%
-6	100%	100%	100%	100%	100%	100%	93%	81%	76%	68%	92%
-5	100%	100%	100%	100%	100%	100%	93%	85%	77%	70%	92%
-4	100%	100%	100%	100%	100%	100%	95%	87%	79%	72%	93%
-3	100%	100%	100%	100%	100%	100%	97%	91%	79%	75%	94%
-2	100%	100%	100%	100%	100%	100%	99%	91%	82%	74%	95%
-1	100%	100%	100%	100%	100%	100%	98%	93%	84%	76%	95%
0	100%	100%	100%	100%	100%	100%	100%	96%	85%	78%	96%
1	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	1%
2	0%	0%	0%	0%	0%	0%	0%	0%	1%	11%	1%
3	0%	0%	0%	0%	0%	0%	0%	0%	4%	12%	2%
4	0%	0%	0%	0%	0%	0%	0%	2%	8%	13%	2%
5	0%	0%	0%	0%	0%	0%	0%	5%	8%	14%	3%
6	0%	0%	0%	0%	0%	0%	0%	6%	8%	12%	3%
7	0%	0%	0%	0%	0%	0%	2%	8%	10%	17%	4%
8	0%	0%	0%	0%	0%	0%	0%	10%	15%	18%	4%
9	0%	0%	0%	0%	0%	0%	2%	13%	18%	21%	5%
10	0%	0%	0%	0%	0%	0%	2%	10%	22%	24%	6%
15	0%	0%	0%	0%	0%	7%	15%	25%	46%	43%	14%
20	0%	0%	0%	0%	13%	18%	34%	43%	56%	47%	21%
Grand Total	**52%**	**52%**	**52%**	**51%**	**51%**	**50%**	**49%**	**48%**	**47%**	**46%**	**50%**

Figure 68: Three-vehicle Ego deceleration occurrence average for 50 m initial traffic gap.

Traffic_Gap	25										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	13%	15%	26%	40%	51%	47%	56%	54%	47%	41%	39%
-15	50%	51%	47%	62%	61%	68%	57%	53%	50%	49%	55%
-10	100%	100%	78%	81%	82%	69%	63%	63%	58%	57%	75%
-9	100%	100%	91%	86%	77%	72%	65%	65%	59%	58%	77%
-8	100%	100%	100%	90%	77%	74%	69%	67%	65%	60%	80%
-7	100%	100%	100%	94%	78%	76%	72%	70%	63%	60%	81%
-6	100%	100%	100%	100%	81%	79%	74%	71%	68%	65%	84%
-5	100%	100%	100%	100%	86%	78%	77%	75%	70%	67%	85%
-4	100%	100%	100%	100%	95%	83%	77%	77%	73%	70%	88%
-3	100%	100%	100%	100%	98%	84%	81%	77%	75%	69%	88%
-2	100%	100%	100%	100%	100%	87%	85%	79%	74%	72%	90%
-1	100%	100%	100%	100%	100%	97%	85%	82%	77%	73%	91%
0	100%	100%	100%	100%	100%	100%	86%	80%	76%	73%	92%
1	0%	0%	0%	0%	0%	1%	6%	20%	26%	33%	9%
2	0%	0%	0%	0%	0%	1%	4%	23%	29%	35%	9%
3	0%	0%	0%	0%	0%	1%	13%	26%	35%	38%	11%
4	0%	0%	0%	0%	0%	6%	19%	27%	34%	39%	12%
5	0%	0%	0%	0%	2%	5%	23%	26%	37%	39%	13%
6	0%	0%	0%	0%	0%	6%	23%	33%	42%	41%	15%
7	0%	0%	0%	0%	2%	13%	25%	42%	39%	44%	16%
8	0%	0%	0%	0%	7%	19%	29%	44%	47%	46%	19%
9	0%	0%	0%	7%	15%	19%	33%	49%	44%	51%	22%
10	0%	0%	3%	13%	15%	24%	40%	48%	50%	57%	25%
15	100%	100%	60%	50%	50%	49%	62%	59%	69%	58%	66%
20	100%	100%	100%	100%	100%	79%	81%	80%	72%	63%	87%
Grand Total	**55%**	**55%**	**52%**	**53%**	**51%**	**49%**	**52%**	**56%**	**55%**	**54%**	**53%**

Figure 69: Three-vehicle Ego deceleration occurrence average for 25 m initial traffic gap.

Traffic_Gap	15										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-15	21%	30%	36%	58%	59%	63%	69%	56%	53%	53%	50%
-10	71%	67%	65%	73%	73%	69%	61%	55%	52%	50%	63%
-9	86%	77%	73%	74%	71%	67%	62%	59%	50%	54%	67%
-8	100%	87%	84%	76%	75%	68%	62%	51%	51%	55%	71%
-7	100%	100%	95%	82%	71%	69%	61%	55%	55%	55%	74%
-6	100%	100%	100%	90%	72%	69%	57%	58%	58%	57%	76%
-5	100%	100%	100%	94%	74%	70%	61%	60%	60%	58%	78%
-4	100%	100%	100%	96%	85%	72%	63%	65%	60%	60%	80%
-3	100%	100%	100%	98%	90%	75%	69%	66%	61%	61%	82%
-2	100%	100%	100%	100%	93%	81%	73%	68%	64%	62%	84%
-1	100%	100%	100%	100%	98%	89%	75%	72%	66%	65%	87%
0	100%	100%	100%	100%	100%	94%	82%	72%	68%	66%	88%
1	0%	0%	0%	10%	18%	39%	48%	57%	58%	56%	29%
2	0%	0%	7%	17%	24%	42%	48%	57%	57%	56%	31%
3	0%	11%	19%	19%	24%	50%	56%	56%	61%	54%	35%
4	0%	7%	10%	21%	29%	57%	57%	61%	63%	54%	36%
5	50%	20%	21%	30%	36%	59%	56%	62%	66%	56%	46%
6	100%	40%	33%	38%	43%	63%	64%	64%	68%	58%	57%
7	100%	60%	46%	42%	48%	67%	64%	66%	66%	56%	62%
8	100%	75%	46%	47%	55%	69%	68%	66%	63%	52%	64%
9	100%	100%	57%	57%	58%	70%	71%	64%	61%	52%	69%
10	100%	100%	71%	67%	66%	72%	70%	64%	60%	55%	73%
15	100%	100%	100%	100%	100%	94%	78%	68%	60%	62%	86%
Grand Total	**75%**	**68%**	**64%**	**65%**	**64%**	**68%**	**64%**	**62%**	**60%**	**57%**	**65%**

Figure 70: Three-vehicle Ego deceleration occurrence average for 15 m initial traffic gap.

Traffic_Gap	10										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-10	57%	49%	49%	65%	72%	69%	67%	59%	55%	55%	60%
-9	71%	55%	55%	70%	73%	66%	64%	58%	56%	55%	62%
-8	83%	63%	66%	73%	74%	67%	64%	58%	53%	53%	65%
-7	100%	78%	74%	75%	74%	65%	61%	55%	52%	49%	68%
-6	100%	89%	75%	77%	75%	67%	60%	57%	51%	53%	71%
-5	100%	100%	89%	84%	79%	65%	62%	57%	54%	54%	74%
-4	100%	100%	100%	94%	83%	65%	62%	57%	57%	59%	78%
-3	100%	100%	100%	96%	89%	69%	66%	60%	58%	62%	80%
-2	100%	100%	100%	96%	91%	75%	70%	64%	64%	63%	82%
-1	100%	100%	100%	100%	95%	78%	74%	65%	66%	64%	84%
0	100%	100%	100%	100%	93%	87%	79%	71%	68%	67%	86%
1	4%	17%	22%	28%	33%	53%	64%	60%	62%	57%	40%
2	37%	33%	30%	36%	37%	62%	65%	59%	63%	56%	48%
3	67%	50%	41%	36%	38%	66%	71%	62%	61%	56%	55%
4	100%	61%	44%	42%	45%	66%	71%	60%	62%	57%	61%
5	100%	73%	50%	48%	54%	74%	70%	64%	63%	56%	65%
6	100%	80%	63%	58%	62%	70%	75%	67%	61%	55%	69%
7	100%	100%	75%	67%	66%	77%	73%	65%	60%	53%	74%
8	100%	100%	88%	73%	73%	81%	75%	62%	59%	54%	76%
9	100%	100%	100%	82%	76%	73%	73%	63%	58%	54%	78%
10	100%	100%	100%	91%	85%	82%	69%	61%	59%	54%	80%
Grand Total	**87%**	**78%**	**72%**	**71%**	**70%**	**70%**	**68%**	**61%**	**59%**	**56%**	**69%**

Figure 71: Three-vehicle Ego deceleration occurrence average for 10 m initial traffic gap.

Traffic_Gap	5										
Average of Dec_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-9	52%	44%	50%	51%	66%	72%	69%	62%	59%	53%	58%
-8	67%	56%	52%	55%	69%	75%	69%	58%	58%	57%	61%
-7	83%	63%	56%	56%	76%	74%	65%	60%	58%	59%	65%
-6	100%	74%	64%	67%	75%	76%	65%	58%	57%	57%	69%
-5	100%	79%	72%	77%	73%	76%	60%	57%	55%	55%	70%
-4	100%	88%	81%	78%	76%	72%	56%	58%	55%	53%	72%
-3	100%	100%	92%	87%	82%	74%	58%	60%	55%	56%	76%
-2	100%	100%	100%	92%	87%	74%	57%	62%	58%	59%	79%
-1	100%	100%	100%	98%	93%	79%	59%	62%	62%	61%	81%
0	100%	100%	100%	100%	94%	87%	66%	67%	64%	64%	84%
1	100%	52%	44%	50%	51%	69%	71%	70%	60%	59%	63%
2	100%	67%	56%	53%	55%	70%	73%	69%	58%	58%	66%
3	100%	83%	63%	56%	56%	76%	73%	65%	59%	58%	69%
4	100%	100%	74%	64%	67%	75%	72%	66%	58%	57%	73%
5	100%	100%	79%	72%	78%	75%	75%	62%	56%	55%	75%
6	100%	100%	88%	81%	79%	76%	72%	57%	57%	54%	76%
7	100%	100%	100%	92%	87%	82%	75%	58%	60%	55%	81%
8	100%	100%	100%	100%	93%	88%	75%	59%	61%	58%	83%
9	100%	100%	100%	100%	98%	93%	79%	61%	62%	61%	85%
Grand Total	**95%**	**85%**	**77%**	**75%**	**77%**	**77%**	**68%**	**62%**	**59%**	**57%**	**73%**

Figure 72: Three-vehicle Ego deceleration occurrence average for 5 m initial traffic gap.

The following tables show the acceleration selection occurrence, just as the deceleration selection was shown.

Traffic_Gap	(All)										
Average of Acc_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	22%	24%	21%	19%	20%	22%	22%	23%	26%	28%	23%
-15	26%	24%	24%	18%	18%	19%	21%	26%	29%	30%	23%
-10	12%	14%	18%	13%	13%	17%	22%	26%	29%	30%	19%
-9	13%	18%	19%	17%	16%	19%	23%	26%	30%	31%	21%
-8	7%	14%	14%	15%	15%	17%	22%	28%	29%	31%	19%
-7	2%	8%	11%	13%	14%	17%	22%	27%	30%	31%	18%
-6	0%	5%	9%	10%	14%	16%	22%	26%	28%	29%	16%
-5	0%	3%	5%	7%	13%	16%	21%	24%	27%	29%	15%
-4	0%	2%	3%	5%	9%	15%	21%	23%	26%	27%	13%
-3	0%	0%	1%	3%	6%	14%	19%	22%	25%	27%	12%
-2	0%	0%	0%	2%	4%	12%	17%	20%	24%	25%	10%
-1	0%	0%	0%	0%	2%	8%	16%	19%	22%	24%	9%
0	0%	0%	0%	0%	2%	5%	12%	17%	21%	23%	8%
1	85%	90%	90%	87%	85%	77%	73%	70%	69%	64%	79%
2	80%	86%	87%	85%	83%	75%	73%	70%	69%	67%	77%
3	76%	79%	83%	84%	83%	72%	70%	70%	67%	66%	75%
4	71%	76%	82%	82%	80%	71%	69%	69%	66%	66%	73%
5	64%	72%	79%	78%	76%	70%	68%	69%	66%	66%	71%
6	57%	69%	74%	75%	74%	69%	67%	68%	65%	65%	68%
7	57%	63%	68%	71%	71%	66%	66%	66%	64%	65%	66%
8	57%	61%	67%	69%	67%	63%	65%	66%	63%	64%	64%
9	57%	57%	63%	65%	65%	64%	63%	64%	63%	63%	62%
10	67%	67%	71%	72%	72%	70%	70%	69%	66%	64%	69%
15	60%	60%	68%	70%	70%	70%	67%	65%	60%	61%	65%
20	75%	75%	75%	75%	72%	76%	67%	60%	56%	61%	69%
Grand Total	**35%**	**38%**	**41%**	**41%**	**42%**	**41%**	**43%**	**45%**	**45%**	**45%**	**42%**

Figure 73: Three-vehicle Ego acceleration occurrence average for all traffic gaps.

Traffic_Gap	100										
Average of Acc_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0%	0%	0%	0%	7%	9%	11%	11%	13%	14%	7%
-15	0%	0%	0%	0%	0%	10%	8%	7%	14%	15%	5%
-10	0%	0%	0%	0%	0%	0%	5%	9%	8%	7%	3%
-9	0%	0%	0%	0%	0%	0%	9%	5%	7%	8%	3%
-8	0%	0%	0%	0%	0%	0%	8%	5%	6%	7%	3%
-7	0%	0%	0%	0%	0%	0%	5%	7%	9%	8%	3%
-6	0%	0%	0%	0%	0%	0%	0%	7%	6%	5%	2%
-5	0%	0%	0%	0%	0%	5%	1%	5%	7%	7%	2%
-4	0%	0%	0%	0%	0%	0%	0%	6%	8%	7%	2%
-3	0%	0%	0%	0%	0%	0%	4%	5%	6%	10%	2%
-2	0%	0%	0%	0%	0%	0%	4%	4%	7%	7%	2%
-1	0%	0%	0%	0%	0%	0%	3%	5%	6%	6%	2%
0	0%	0%	0%	0%	0%	0%	0%	4%	7%	6%	2%
1	100%	100%	100%	100%	100%	100%	100%	100%	92%	59%	95%
2	100%	100%	100%	100%	100%	100%	100%	100%	90%	83%	97%
3	100%	100%	100%	100%	100%	100%	100%	100%	90%	83%	97%
4	100%	100%	100%	100%	100%	100%	100%	100%	89%	82%	97%
5	100%	100%	100%	100%	100%	100%	100%	100%	89%	81%	97%
6	100%	100%	100%	100%	100%	100%	100%	100%	87%	80%	97%
7	100%	100%	100%	100%	100%	100%	100%	100%	87%	80%	97%
8	100%	100%	100%	100%	100%	100%	100%	100%	86%	78%	96%
9	100%	100%	100%	100%	100%	100%	100%	95%	85%	78%	96%
10	100%	100%	100%	100%	100%	100%	100%	95%	85%	75%	96%
15	100%	100%	100%	100%	100%	100%	92%	78%	74%	66%	91%
20	100%	100%	100%	100%	100%	100%	83%	61%	53%	53%	85%
Grand Total	**48%**	**48%**	**48%**	**48%**	**48%**	**49%**	**49%**	**48%**	**44%**	**40%**	**47%**

Figure 74: Three-vehicle Ego acceleration occurrence average for 100 m initial traffic gap.

Traffic_Gap	50										
Average of Acc_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0%	10%	12%	16%	23%	28%	34%	36%	38%	38%	23%
-15	0%	0%	0%	8%	11%	19%	24%	32%	35%	37%	17%
-10	0%	0%	0%	0%	3%	9%	18%	22%	30%	35%	12%
-9	0%	0%	0%	0%	0%	8%	13%	22%	28%	32%	10%
-8	0%	0%	0%	0%	0%	4%	13%	21%	27%	33%	10%
-7	0%	0%	0%	0%	0%	3%	10%	19%	27%	32%	9%
-6	0%	0%	0%	0%	0%	0%	7%	19%	24%	32%	8%
-5	0%	0%	0%	0%	0%	0%	7%	15%	23%	30%	8%
-4	0%	0%	0%	0%	0%	0%	5%	13%	21%	28%	7%
-3	0%	0%	0%	0%	0%	0%	3%	9%	21%	25%	6%
-2	0%	0%	0%	0%	0%	0%	1%	9%	18%	26%	5%
-1	0%	0%	0%	0%	0%	0%	2%	7%	16%	24%	5%
0	0%	0%	0%	0%	0%	0%	0%	4%	15%	22%	4%
1	100%	100%	100%	100%	100%	100%	100%	100%	100%	90%	99%
2	100%	100%	100%	100%	100%	100%	100%	100%	99%	89%	99%
3	100%	100%	100%	100%	100%	100%	100%	100%	96%	88%	98%
4	100%	100%	100%	100%	100%	100%	100%	98%	92%	87%	98%
5	100%	100%	100%	100%	100%	100%	100%	95%	92%	86%	97%
6	100%	100%	100%	100%	100%	100%	100%	94%	92%	88%	97%
7	100%	100%	100%	100%	100%	100%	98%	92%	90%	83%	96%
8	100%	100%	100%	100%	100%	100%	100%	90%	85%	82%	96%
9	100%	100%	100%	100%	100%	100%	98%	87%	82%	79%	95%
10	100%	100%	100%	100%	100%	100%	98%	90%	78%	76%	94%
15	100%	100%	100%	100%	100%	93%	85%	75%	54%	57%	86%
20	100%	100%	100%	100%	87%	82%	66%	57%	44%	53%	79%
Grand Total	**48%**	**48%**	**48%**	**49%**	**49%**	**50%**	**51%**	**52%**	**53%**	**54%**	**50%**

Figure 75: Three-vehicle Ego acceleration occurrence average for 50 m initial traffic gap.

Traffic_Gap	25										
Average of Acc_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	87%	85%	74%	60%	49%	53%	44%	46%	53%	59%	61%
-15	50%	49%	53%	38%	39%	32%	43%	47%	50%	51%	45%
-10	0%	0%	22%	19%	18%	31%	37%	37%	42%	43%	25%
-9	0%	0%	9%	14%	23%	28%	35%	35%	41%	42%	23%
-8	0%	0%	0%	10%	23%	26%	31%	33%	35%	40%	20%
-7	0%	0%	0%	6%	22%	24%	28%	30%	37%	40%	19%
-6	0%	0%	0%	0%	19%	21%	26%	29%	32%	35%	16%
-5	0%	0%	0%	0%	14%	22%	23%	25%	30%	33%	15%
-4	0%	0%	0%	0%	5%	17%	23%	23%	27%	30%	12%
-3	0%	0%	0%	0%	2%	16%	19%	23%	25%	31%	12%
-2	0%	0%	0%	0%	0%	13%	15%	21%	26%	28%	10%
-1	0%	0%	0%	0%	0%	3%	15%	18%	23%	27%	9%
0	0%	0%	0%	0%	0%	0%	14%	20%	24%	27%	8%
1	100%	100%	100%	100%	100%	99%	94%	80%	74%	67%	91%
2	100%	100%	100%	100%	100%	99%	96%	77%	71%	65%	91%
3	100%	100%	100%	100%	100%	99%	87%	74%	65%	62%	89%
4	100%	100%	100%	100%	100%	94%	81%	73%	66%	61%	88%
5	100%	100%	100%	100%	98%	95%	77%	74%	63%	61%	87%
6	100%	100%	100%	100%	100%	94%	77%	67%	58%	59%	85%
7	100%	100%	100%	100%	98%	88%	75%	58%	61%	56%	84%
8	100%	100%	100%	100%	93%	81%	71%	56%	53%	54%	81%
9	100%	100%	100%	93%	85%	81%	67%	51%	56%	49%	78%
10	100%	100%	97%	87%	85%	76%	60%	52%	50%	43%	75%
15	0%	0%	40%	50%	50%	51%	38%	41%	31%	42%	34%
20	0%	0%	0%	0%	0%	21%	19%	20%	28%	37%	13%
Grand Total	**45%**	**45%**	**48%**	**47%**	**49%**	**51%**	**48%**	**44%**	**45%**	**46%**	**47%**

Figure 76: Three-vehicle Ego acceleration occurrence average for 25 m initial traffic gap.

Traffic_Gap	15										
Average of Acc_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-15	79%	70%	64%	42%	41%	37%	31%	44%	47%	47%	50%
-10	29%	33%	35%	27%	27%	31%	39%	45%	48%	50%	37%
-9	14%	23%	27%	26%	29%	33%	38%	41%	50%	46%	33%
-8	0%	13%	16%	24%	25%	32%	38%	49%	49%	45%	29%
-7	0%	0%	5%	18%	29%	31%	39%	45%	45%	45%	26%
-6	0%	0%	0%	10%	28%	31%	43%	42%	42%	43%	24%
-5	0%	0%	0%	6%	26%	30%	39%	40%	40%	42%	22%
-4	0%	0%	0%	4%	15%	28%	37%	35%	40%	40%	20%
-3	0%	0%	0%	2%	10%	25%	31%	34%	39%	39%	18%
-2	0%	0%	0%	0%	7%	19%	27%	32%	36%	38%	16%
-1	0%	0%	0%	0%	2%	11%	25%	28%	34%	35%	13%
0	0%	0%	0%	0%	0%	6%	18%	28%	32%	34%	12%
1	100%	100%	100%	90%	82%	61%	52%	43%	42%	44%	71%
2	100%	100%	93%	83%	76%	58%	52%	43%	43%	44%	69%
3	100%	89%	81%	81%	76%	50%	44%	44%	39%	46%	65%
4	100%	93%	90%	79%	71%	43%	43%	39%	37%	46%	64%
5	50%	80%	79%	70%	64%	41%	44%	38%	34%	44%	54%
6	0%	60%	67%	62%	57%	37%	36%	36%	32%	42%	43%
7	0%	40%	54%	58%	52%	33%	36%	34%	34%	44%	38%
8	0%	25%	54%	53%	45%	31%	32%	34%	37%	48%	36%
9	0%	0%	43%	43%	42%	30%	29%	36%	39%	48%	31%
10	0%	0%	29%	33%	34%	28%	30%	36%	40%	45%	27%
15	0%	0%	0%	0%	0%	6%	22%	32%	40%	38%	14%
Grand Total	**25%**	**32%**	**36%**	**35%**	**36%**	**32%**	**36%**	**38%**	**40%**	**43%**	**35%**

Figure 77: Three-vehicle Ego acceleration occurrence average for 15 m initial traffic gap.

Traffic_Gap	10										
Average of Acc_%	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-10	43%	51%	51%	35%	28%	31%	33%	41%	45%	45%	40%
-9	29%	45%	45%	30%	27%	34%	36%	42%	44%	45%	38%
-8	17%	37%	34%	27%	26%	33%	36%	42%	47%	47%	35%
-7	0%	22%	26%	25%	26%	35%	39%	45%	48%	51%	32%
-6	0%	11%	25%	23%	25%	33%	40%	43%	49%	47%	29%
-5	0%	0%	11%	16%	21%	35%	38%	43%	46%	46%	26%
-4	0%	0%	0%	6%	17%	35%	38%	43%	43%	41%	22%
-3	0%	0%	0%	4%	11%	31%	34%	40%	42%	38%	20%
-2	0%	0%	0%	4%	9%	25%	30%	36%	36%	37%	18%
-1	0%	0%	0%	0%	5%	22%	26%	35%	34%	36%	16%
0	0%	0%	0%	0%	7%	13%	21%	29%	32%	33%	14%
1	96%	83%	78%	72%	67%	47%	36%	40%	38%	43%	60%
2	63%	67%	70%	64%	63%	38%	35%	41%	37%	44%	52%
3	33%	50%	59%	64%	62%	34%	29%	38%	39%	44%	45%
4	0%	39%	56%	58%	55%	34%	29%	40%	38%	43%	39%
5	0%	27%	50%	52%	46%	26%	30%	36%	37%	44%	35%
6	0%	20%	38%	42%	38%	30%	25%	33%	39%	45%	31%
7	0%	0%	25%	33%	34%	23%	27%	35%	40%	47%	26%
8	0%	0%	13%	27%	27%	19%	25%	38%	41%	46%	24%
9	0%	0%	0%	18%	24%	27%	27%	37%	42%	46%	22%
10	0%	0%	0%	9%	15%	18%	31%	39%	41%	46%	20%
Grand Total	**13%**	**22%**	**28%**	**29%**	**30%**	**30%**	**32%**	**39%**	**41%**	**44%**	**31%**

Figure 78: Three-vehicle Ego acceleration occurrence average for 10 m initial traffic gap.

Traffic_Gap	5										
Average of Acc_%	**Goal Position (m)**										
Ego Initial Pos (m)		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-9		48%	56%	50%	49%	34%	28%	31%	38%	41%	47%	42%
-8		33%	44%	48%	45%	31%	25%	31%	42%	42%	43%	39%
-7		17%	37%	44%	44%	24%	26%	35%	40%	42%	41%	35%
-6		0%	26%	36%	33%	25%	24%	35%	42%	43%	43%	31%
-5		0%	21%	28%	23%	27%	24%	40%	43%	45%	45%	30%
-4		0%	13%	19%	22%	24%	28%	44%	42%	45%	47%	28%
-3		0%	0%	8%	13%	18%	26%	42%	40%	45%	44%	24%
-2		0%	0%	0%	8%	13%	26%	43%	38%	42%	41%	21%
-1		0%	0%	0%	2%	7%	21%	41%	38%	38%	39%	19%
0		0%	0%	0%	0%	6%	13%	34%	33%	36%	36%	16%
1		0%	48%	56%	50%	49%	31%	29%	30%	40%	41%	37%
2		0%	33%	44%	47%	45%	30%	27%	31%	42%	42%	34%
3		0%	17%	37%	44%	44%	24%	27%	35%	41%	42%	31%
4		0%	0%	26%	36%	33%	25%	28%	34%	42%	43%	27%
5		0%	0%	21%	28%	22%	25%	25%	38%	44%	45%	25%
6		0%	0%	13%	19%	21%	24%	28%	43%	43%	46%	24%
7		0%	0%	0%	8%	13%	18%	25%	42%	40%	45%	19%
8		0%	0%	0%	0%	7%	12%	25%	41%	39%	42%	17%
9		0%	0%	0%	0%	2%	7%	21%	39%	38%	39%	15%
Grand Total		**5%**	**15%**	**23%**	**25%**	**23%**	**23%**	**32%**	**38%**	**41%**	**43%**	**27%**

Figure 79: Three-vehicle Ego acceleration occurrence average for 5 m initial traffic gap.

The magnitude of the acceleration and deceleration can be as significant as the deceleration and acceleration occurrence. The remining tables in this section show individual plots for both the acceleration and deceleration at the individual initial traffic gap settings.

Traffic_Gap	(All)										
Average of Deceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	-4.79	-4.71	-4.45	-4.44	-4.49	-4.58	-4.52	-4.51	-4.63	-4.73	-4.58
-15	-5.00	-4.98	-4.79	-4.55	-4.62	-4.36	-4.35	-4.53	-4.69	-4.79	-4.66
-10	-5.00	-4.97	-4.93	-4.67	-4.55	-4.42	-4.29	-4.56	-4.62	-4.76	-4.68
-9	-5.00	-5.00	-4.96	-4.72	-4.57	-4.38	-4.23	-4.40	-4.64	-4.76	-4.67
-8	-5.00	-4.99	-4.93	-4.78	-4.62	-4.37	-4.23	-4.55	-4.74	-4.76	-4.70
-7	-5.00	-4.98	-4.98	-4.88	-4.67	-4.42	-4.27	-4.57	-4.71	-4.80	-4.73
-6	-5.00	-5.00	-4.98	-4.91	-4.72	-4.42	-4.37	-4.58	-4.74	-4.80	-4.75
-5	-5.00	-5.00	-5.00	-4.89	-4.83	-4.55	-4.44	-4.57	-4.73	-4.77	-4.78
-4	-5.00	-5.00	-5.00	-4.92	-4.90	-4.74	-4.56	-4.62	-4.72	-4.76	-4.82
-3	-5.00	-5.00	-5.00	-4.93	-4.93	-4.86	-4.62	-4.64	-4.76	-4.79	-4.85
-2	-5.00	-5.00	-5.00	-4.98	-4.97	-4.91	-4.77	-4.63	-4.74	-4.79	-4.88
-1	-5.00	-5.00	-5.00	-5.00	-4.93	-4.95	-4.80	-4.67	-4.73	-4.77	-4.89
0	-5.00	-5.00	-5.00	-5.00	-4.99	-4.90	-4.80	-4.70	-4.75	-4.78	-4.89
1	-0.78	-1.43	-1.41	-1.89	-1.90	-2.07	-2.35	-2.34	-2.59	-3.24	-2.00
2	-1.39	-1.43	-1.90	-1.83	-2.12	-1.89	-2.38	-2.33	-2.68	-3.27	-2.12
3	-1.43	-1.90	-2.13	-2.14	-2.14	-1.76	-2.42	-2.37	-2.75	-3.36	-2.24
4	-1.43	-1.60	-1.89	-2.14	-2.12	-2.16	-2.40	-2.41	-2.78	-3.46	-2.24
5	-2.14	-2.10	-2.13	-2.14	-2.18	-2.13	-2.29	-2.62	-2.88	-3.53	-2.41
6	-2.14	-2.14	-2.14	-2.12	-2.04	-2.30	-2.42	-2.54	-2.88	-3.56	-2.43
7	-2.14	-2.14	-2.14	-2.10	-2.17	-2.50	-2.56	-2.60	-2.98	-3.62	-2.49
8	-2.14	-2.14	-2.13	-2.11	-2.49	-2.42	-2.55	-2.68	-3.14	-3.64	-2.54
9	-2.14	-2.14	-2.14	-2.62	-2.74	-2.57	-2.63	-2.81	-3.21	-3.58	-2.66
10	-1.67	-1.67	-1.85	-2.39	-2.41	-2.17	-2.36	-2.41	-3.01	-3.34	-2.33
15	-2.00	-2.00	-1.99	-1.97	-1.97	-2.08	-2.29	-2.83	-2.84	-3.25	-2.32
20	-1.25	-1.25	-1.25	-1.25	-1.59	-2.13	-2.24	-2.63	-2.81	-3.20	-1.96
Grand Total	**-3.43**	**-3.48**	**-3.51**	**-3.52**	**-3.53**	**-3.45**	**-3.50**	**-3.60**	**-3.84**	**-4.13**	**-3.60**

Figure 80: Three-vehicle Ego deceleration average values by start position and goal position for all traffic gap values.

Traffic_Gap	100										
Average of Deceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	-5.00	-4.89	-4.51	-4.17	-3.81	-4.44	-4.71	-4.82	-4.84	-4.80	-4.60
-15	-5.00	-5.00	-4.98	-4.49	-4.16	-4.10	-4.56	-4.70	-4.75	-4.86	-4.66
-10	-5.00	-5.00	-5.00	-5.00	-4.62	-4.05	-4.11	-4.57	-4.66	-4.89	-4.69
-9	-5.00	-5.00	-5.00	-5.00	-4.77	-4.13	-4.08	-4.50	-4.69	-4.82	-4.70
-8	-5.00	-5.00	-5.00	-5.00	-4.90	-4.23	-4.09	-4.53	-4.63	-4.80	-4.72
-7	-5.00	-5.00	-5.00	-5.00	-5.00	-4.29	-4.09	-4.35	-4.63	-4.74	-4.71
-6	-5.00	-5.00	-5.00	-5.00	-5.00	-4.40	-3.87	-4.48	-4.64	-4.81	-4.72
-5	-5.00	-5.00	-5.00	-5.00	-5.00	-4.95	-4.08	-4.54	-4.58	-4.79	-4.79
-4	-5.00	-5.00	-5.00	-5.00	-5.00	-4.94	-4.08	-4.43	-4.53	-4.67	-4.77
-3	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.21	-4.40	-4.46	-4.65	-4.77
-2	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.78	-4.27	-4.44	-4.62	-4.81
-1	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.95	-4.30	-4.41	-4.59	-4.82
0	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.94	-4.54	-4.49	-4.52	-4.85
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.49	-4.04	-0.65
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.58	-3.57	-0.62
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.76	-3.75	-0.65
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.84	-3.75	-0.66
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.95	-3.87	-0.68
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.76	-3.72	-0.65
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.78	-3.91	-0.67
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-3.13	-4.02	-0.71
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.13	-3.06	-3.93	-0.71
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.48	-3.35	-3.82	-0.77
15	0.00	0.00	0.00	0.00	0.00	0.00	-0.42	-1.73	-3.20	-4.10	-0.94
20	0.00	0.00	0.00	0.00	0.00	0.00	-0.54	-2.56	-3.57	-4.20	-1.09
Grand Total	**-2.60**	**-2.60**	**-2.58**	**-2.55**	**-2.49**	**-2.38**	**-2.30**	**-2.53**	**-3.81**	**-4.33**	**-2.82**

Figure 81: Three-vehicle Ego deceleration average for 100 m initial traffic gap.

Traffic_Gap	50										
Average of Deceleration_Average Ego Initial Pos (m)	**Goal Position (m)**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20		-4.61	-4.01	-4.21	-4.27	-4.47	-4.51	-4.58	-4.62	-4.74	-4.76	-4.48
-15		-5.00	-4.96	-4.52	-4.67	-4.35	-4.48	-4.67	-4.70	-4.73	-4.81	-4.69
-10		-5.00	-5.00	-4.98	-4.96	-4.64	-4.56	-4.54	-4.58	-4.59	-4.76	-4.76
-9		-5.00	-5.00	-5.00	-4.98	-4.71	-4.64	-4.48	-4.64	-4.61	-4.77	-4.78
-8		-5.00	-5.00	-5.00	-4.99	-4.81	-4.50	-4.51	-4.58	-4.66	-4.76	-4.78
-7		-5.00	-5.00	-5.00	-5.00	-4.96	-4.46	-4.54	-4.66	-4.70	-4.83	-4.81
-6		-5.00	-5.00	-5.00	-5.00	-4.98	-4.26	-4.47	-4.62	-4.64	-4.79	-4.78
-5		-5.00	-5.00	-5.00	-5.00	-5.00	-4.43	-4.37	-4.49	-4.66	-4.71	-4.77
-4		-5.00	-5.00	-5.00	-5.00	-5.00	-4.68	-4.37	-4.55	-4.61	-4.62	-4.78
-3		-5.00	-5.00	-5.00	-5.00	-5.00	-4.91	-4.53	-4.46	-4.61	-4.62	-4.81
-2		-5.00	-5.00	-5.00	-5.00	-5.00	-4.96	-4.53	-4.31	-4.50	-4.68	-4.80
-1		-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.44	-4.28	-4.55	-4.65	-4.79
0		-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.57	-4.31	-4.51	-4.59	-4.80
1		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-3.01	-0.30
2		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.20	-3.00	-0.32
3		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.70	-2.87	-0.36
4		0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.21	-0.85	-3.15	-0.42
5		0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.14	-1.06	-3.24	-0.54
6		0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.72	-1.21	-3.40	-0.53
7		0.00	0.00	0.00	0.00	0.00	0.00	-0.18	-0.69	-1.28	-3.13	-0.53
8		0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.08	-1.82	-3.23	-0.61
9		0.00	0.00	0.00	0.00	0.00	0.00	-0.20	-1.63	-2.06	-3.11	-0.70
10		0.00	0.00	0.00	0.00	0.00	0.00	-0.74	-0.79	-2.32	-3.10	-0.70
15		0.00	0.00	0.00	0.00	0.00	-1.01	-1.44	-2.92	-2.55	-3.35	-1.13
20		0.00	0.00	0.00	0.00	-1.35	-3.55	-3.77	-3.87	-3.34	-4.01	-1.99
Grand Total		**-2.58**	**-2.56**	**-2.55**	**-2.55**	**-2.57**	**-2.60**	**-2.60**	**-2.87**	**-3.10**	**-4.00**	**-2.80**

Figure 82: Three-vehicle Ego deceleration average for 50 m initial traffic gap.

Traffic_Gap	25										
Average of Deceleration_Average Ego Initial Pos (m)	**Goal Position (m)**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20		-4.55	-4.92	-4.07	-4.33	-4.69	-4.37	-3.80	-3.60	-3.93	-4.37	-4.26
-15		-4.99	-4.92	-4.53	-4.74	-4.77	-3.79	-4.16	-4.12	-4.28	-4.54	-4.48
-10		-5.00	-5.00	-4.98	-4.64	-4.04	-4.26	-4.59	-4.46	-4.51	-4.48	-4.60
-9		-5.00	-5.00	-4.93	-4.58	-4.37	-4.30	-4.54	-4.45	-4.52	-4.42	-4.61
-8		-5.00	-5.00	-4.95	-4.81	-4.44	-4.37	-4.61	-4.58	-4.53	-4.48	-4.68
-7		-5.00	-5.00	-5.00	-4.82	-4.62	-4.57	-4.54	-4.57	-4.63	-4.54	-4.73
-6		-5.00	-5.00	-5.00	-4.99	-4.77	-4.62	-4.59	-4.58	-4.58	-4.56	-4.77
-5		-5.00	-5.00	-5.00	-5.00	-4.74	-4.54	-4.66	-4.58	-4.54	-4.54	-4.76
-4		-5.00	-5.00	-5.00	-5.00	-4.88	-4.63	-4.58	-4.70	-4.60	-4.62	-4.80
-3		-5.00	-5.00	-5.00	-5.00	-4.98	-4.76	-4.39	-4.59	-4.67	-4.65	-4.81
-2		-5.00	-5.00	-5.00	-5.00	-5.00	-4.84	-4.53	-4.55	-4.59	-4.64	-4.81
-1		-5.00	-5.00	-5.00	-5.00	-5.00	-4.83	-4.50	-4.61	-4.68	-4.68	-4.83
0		-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.59	-4.55	-4.68	-4.78	-4.86
1		0.00	0.00	0.00	0.00	0.00	-1.11	-3.19	-3.85	-4.48	-3.78	-1.64
2		0.00	0.00	0.00	0.00	0.00	-0.56	-3.32	-3.75	-4.50	-3.97	-1.61
3		0.00	0.00	0.00	0.00	0.00	-0.56	-3.82	-3.90	-4.18	-3.96	-1.64
4		0.00	0.00	0.00	0.00	0.00	-3.20	-3.66	-3.97	-4.17	-4.11	-1.91
5		0.00	0.00	0.00	0.00	-1.11	-3.00	-2.88	-4.50	-4.53	-4.17	-2.02
6		0.00	0.00	0.00	0.00	0.00	-3.25	-3.64	-4.39	-4.59	-4.15	-2.00
7		0.00	0.00	0.00	0.00	-1.11	-5.00	-4.20	-4.35	-4.66	-4.22	-2.35
8		0.00	0.00	0.00	0.00	-3.33	-3.96	-4.10	-4.50	-4.57	-4.15	-2.46
9		0.00	0.00	0.00	-3.31	-5.00	-4.09	-4.45	-4.74	-4.57	-3.96	-3.01
10		0.00	0.00	-1.08	-4.33	-5.00	-3.97	-4.35	-4.62	-4.30	-3.82	-3.15
15		-5.00	-5.00	-4.93	-4.87	-4.83	-4.52	-4.69	-4.78	-3.80	-4.15	-4.66
20		-5.00	-5.00	-5.00	-5.00	-5.00	-4.99	-4.66	-4.09	-4.33	-4.60	-4.77
Grand Total		**-2.98**	**-2.99**	**-2.98**	**-3.22**	**-3.47**	**-3.88**	**-4.20**	**-4.38**	**-4.46**	**-4.33**	**-3.69**

Figure 83: Three-vehicle Ego deceleration average for 25 m initial traffic gap.

Traffic_Gap	15										
Average of Deceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-15		-5.00	-5.00	-4.91	-3.83	-4.80	-4.43	-3.36	-4.11	-4.71	-4.74	-4.49
-10		-5.00	-5.00	-4.64	-4.24	-4.71	-4.18	-3.80	-4.53	-4.20	-4.59	-4.49
-9		-5.00	-5.00	-4.88	-4.57	-4.74	-3.89	-3.93	-3.95	-4.66	-4.66	-4.53
-8		-5.00	-4.96	-4.81	-4.65	-4.63	-3.99	-3.83	-4.64	-4.87	-4.88	-4.63
-7		-5.00	-4.99	-4.87	-4.77	-4.77	-4.15	-4.03	-4.78	-4.80	-4.92	-4.71
-6		-5.00	-5.00	-4.84	-4.79	-4.82	-4.37	-4.58	-4.75	-4.84	-4.79	-4.78
-5		-5.00	-5.00	-5.00	-4.86	-4.91	-4.69	-4.65	-4.73	-4.85	-4.79	-4.85
-4		-5.00	-5.00	-5.00	-4.85	-4.85	-4.78	-4.76	-4.67	-4.91	-4.78	-4.86
-3		-5.00	-5.00	-5.00	-4.91	-4.96	-4.92	-4.71	-4.74	-4.85	-4.84	-4.89
-2		-5.00	-5.00	-5.00	-5.00	-5.00	-4.90	-4.76	-4.73	-4.86	-4.78	-4.90
-1		-5.00	-5.00	-5.00	-5.00	-4.92	-4.95	-4.82	-4.64	-4.79	-4.74	-4.89
0		-5.00	-5.00	-5.00	-5.00	-5.00	-4.81	-4.74	-4.75	-4.81	-4.82	-4.89
1		0.00	0.00	0.00	-3.33	-3.33	-4.72	-4.79	-4.46	-3.77	-3.45	-2.79
2		0.00	0.00	-3.33	-3.31	-5.00	-4.38	-4.74	-4.52	-3.89	-3.57	-3.27
3		0.00	-3.30	-4.89	-5.00	-5.00	-3.95	-4.70	-4.55	-3.74	-3.83	-3.90
4		0.00	-1.19	-3.32	-5.00	-4.99	-3.75	-4.73	-4.39	-3.64	-3.98	-3.50
5		-5.00	-5.00	-5.00	-4.99	-4.99	-3.87	-4.80	-4.47	-3.42	-4.16	-4.57
6		-5.00	-5.00	-5.00	-4.89	-4.82	-3.89	-4.81	-4.45	-3.35	-4.35	-4.56
7		-5.00	-4.99	-4.99	-4.93	-4.74	-3.58	-4.70	-4.50	-3.58	-4.49	-4.55
8		-5.00	-4.99	-4.93	-4.81	-4.39	-4.01	-4.78	-4.15	-3.60	-4.65	-4.53
9		-5.00	-4.99	-5.00	-5.00	-4.32	-4.17	-4.71	-4.29	-3.75	-4.71	-4.59
10		-5.00	-5.00	-5.00	-5.00	-4.56	-4.24	-4.70	-4.17	-3.69	-4.53	-4.59
15		-5.00	-5.00	-5.00	-5.00	-5.00	-4.86	-4.93	-4.71	-4.67	-4.65	-4.88
Grand Total		**-4.13**	**-4.32**	**-4.58**	**-4.68**	**-4.75**	**-4.33**	**-4.54**	**-4.51**	**-4.27**	**-4.51**	**-4.46**

Figure 84: Three-vehicle Ego deceleration average for 15 m initial traffic gap.

Traffic_Gap	10										
Average of Deceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-10		-5.00	-4.83	-5.00	-4.17	-4.29	-4.50	-3.68	-4.23	-4.78	-4.85	-4.53
-9		-5.00	-5.00	-5.00	-3.94	-4.18	-4.54	-3.96	-4.39	-4.77	-4.79	-4.56
-8		-5.00	-5.00	-4.95	-4.02	-4.36	-4.35	-3.96	-4.40	-4.87	-4.74	-4.57
-7		-5.00	-4.86	-5.00	-4.54	-4.45	-4.36	-4.06	-4.43	-4.49	-4.84	-4.60
-6		-5.00	-5.00	-5.00	-4.75	-4.44	-4.19	-4.28	-4.31	-4.75	-4.83	-4.65
-5		-5.00	-5.00	-4.97	-4.94	-4.73	-4.22	-4.35	-4.36	-4.86	-4.88	-4.73
-4		-5.00	-5.00	-4.99	-4.80	-4.90	-4.70	-4.81	-4.79	-4.81	-4.85	-4.86
-3		-5.00	-5.00	-5.00	-4.87	-4.81	-4.83	-4.95	-4.81	-4.90	-4.88	-4.90
-2		-5.00	-5.00	-5.00	-4.88	-4.85	-4.87	-4.93	-4.76	-4.91	-4.94	-4.91
-1		-5.00	-5.00	-5.00	-5.00	-4.63	-4.93	-4.95	-4.89	-4.81	-4.88	-4.91
0		-5.00	-5.00	-5.00	-5.00	-4.99	-4.91	-4.91	-4.85	-4.93	-4.85	-4.94
1		-0.43	-5.00	-4.85	-4.98	-5.00	-4.60	-4.31	-4.52	-3.45	-4.13	-4.13
2		-4.72	-5.00	-5.00	-4.79	-4.86	-4.17	-4.51	-4.50	-3.40	-4.19	-4.51
3		-5.00	-4.99	-5.00	-5.00	-5.00	-3.88	-4.27	-4.44	-3.64	-4.47	-4.57
4		-5.00	-5.00	-5.00	-5.00	-5.00	-4.13	-4.19	-4.46	-3.67	-4.53	-4.60
5		-5.00	-4.68	-5.00	-5.00	-4.74	-3.61	-4.30	-4.27	-3.89	-4.64	-4.51
6		-5.00	-5.00	-5.00	-4.97	-4.76	-4.33	-4.09	-3.99	-3.90	-4.71	-4.57
7		-5.00	-5.00	-4.99	-4.77	-4.60	-4.18	-4.21	-4.09	-4.02	-4.76	-4.56
8		-5.00	-5.00	-4.99	-4.98	-4.74	-4.05	-4.10	-4.17	-4.06	-4.56	-4.57
9		-5.00	-5.00	-5.00	-5.00	-4.86	-4.76	-4.10	-4.01	-4.11	-4.46	-4.63
10		-5.00	-5.00	-5.00	-5.00	-4.92	-4.80	-4.38	-4.39	-4.41	-4.79	-4.77
Grand Total		**-4.77**	**-4.97**	**-4.99**	**-4.78**	**-4.72**	**-4.42**	**-4.35**	**-4.43**	**-4.35**	**-4.69**	**-4.65**

Figure 85: Three-vehicle Ego deceleration average for 10 m initial traffic gap.

Traffic_Gap	5										
Average of Deceleration_Average Ego Initial Pos (m)	**Goal Position (m)**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-9		-5.00	-5.00	-4.87	-5.00	-4.19	-4.14	-3.64	-3.85	-4.22	-4.83	-4.48
-8		-5.00	-5.00	-4.78	-4.97	-4.16	-4.12	-3.63	-4.14	-4.65	-4.68	-4.51
-7		-5.00	-5.00	-5.00	-5.00	-3.87	-4.11	-3.65	-4.18	-4.73	-4.74	-4.53
-6		-5.00	-4.98	-5.00	-4.84	-4.04	-4.14	-3.77	-4.32	-4.75	-4.83	-4.57
-5		-5.00	-4.98	-5.00	-4.45	-4.44	-4.03	-3.98	-4.32	-4.62	-4.70	-4.55
-4		-5.00	-5.00	-5.00	-4.77	-4.68	-4.45	-4.33	-4.22	-4.57	-4.78	-4.68
-3		-5.00	-5.00	-4.98	-4.75	-4.73	-4.59	-4.54	-4.48	-4.84	-4.85	-4.78
-2		-5.00	-5.00	-5.00	-5.00	-4.93	-4.83	-4.84	-4.81	-4.86	-4.84	-4.91
-1		-5.00	-5.00	-5.00	-4.98	-4.94	-4.94	-4.97	-4.98	-4.89	-4.87	-4.96
0		-5.00	-5.00	-5.00	-5.00	-4.94	-4.56	-4.87	-4.92	-4.83	-4.89	-4.90
1		-5.00	-4.98	-5.00	-4.92	-5.00	-4.08	-4.13	-3.54	-3.94	-4.25	-4.49
2		-5.00	-5.00	-5.00	-4.72	-4.98	-4.16	-4.12	-3.54	-4.21	-4.59	-4.53
3		-5.00	-5.00	-5.00	-5.00	-5.00	-3.92	-4.16	-3.67	-4.26	-4.67	-4.57
4		-5.00	-5.00	-4.94	-5.00	-4.84	-4.06	-4.23	-3.83	-4.26	-4.69	-4.58
5		-5.00	-5.00	-4.95	-5.00	-4.43	-4.42	-4.08	-3.92	-4.32	-4.61	-4.57
6		-5.00	-5.00	-5.00	-5.00	-4.69	-4.64	-4.43	-4.25	-4.35	-4.58	-4.69
7		-5.00	-5.00	-5.00	-4.98	-4.75	-4.75	-4.61	-4.55	-4.53	-4.85	-4.80
8		-5.00	-5.00	-5.00	-5.00	-4.95	-4.93	-4.89	-4.84	-4.80	-4.84	-4.93
9		-5.00	-5.00	-5.00	-5.00	-4.98	-4.94	-4.95	-4.89	-4.94	-4.87	-4.96
Grand Total		**-5.00**	**-5.00**	**-4.97**	**-4.91**	**-4.66**	**-4.41**	**-4.31**	**-4.28**	**-4.56**	**-4.73**	**-4.68**

Figure 86: Three-vehicle Ego deceleration average for 5 m initial traffic gap.

Traffic_Gap	(All)										
Average of Acceleration_Average Ego Initial Pos (m)	**Goal Position (m)**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20		1.00	1.24	1.28	1.37	1.51	1.50	1.68	1.74	1.88	2.03	1.52
-15		1.45	1.57	1.46	1.64	1.66	1.85	2.03	2.10	2.27	2.32	1.83
-10		1.33	1.31	1.65	1.43	1.59	1.86	2.10	2.34	2.28	2.35	1.82
-9		1.65	1.70	2.03	1.62	1.71	2.00	2.17	2.32	2.49	2.60	2.03
-8		1.14	1.54	1.58	1.59	1.61	1.97	2.12	2.36	2.49	2.47	1.89
-7		0.57	1.10	1.31	1.58	1.58	1.88	2.03	2.45	2.51	2.55	1.76
-6		0.00	1.14	0.93	1.02	1.56	1.70	2.01	2.32	2.45	2.46	1.56
-5		0.00	0.44	0.84	0.99	1.72	1.89	2.00	2.26	2.43	2.51	1.51
-4		0.00	0.54	0.56	0.59	1.17	1.85	1.99	2.25	2.45	2.44	1.38
-3		0.00	0.00	0.57	0.63	0.79	1.84	1.91	2.27	2.35	2.45	1.28
-2		0.00	0.00	0.00	0.25	0.82	1.70	1.91	2.17	2.29	2.43	1.16
-1		0.00	0.00	0.00	0.11	0.49	1.40	1.96	2.03	2.24	2.37	1.06
0		0.00	0.00	0.00	0.00	0.29	1.05	1.84	1.98	2.25	2.31	0.97
1		3.40	3.96	4.00	4.00	4.00	3.80	3.77	3.69	3.56	3.56	3.77
2		3.39	4.00	4.00	4.00	4.00	3.76	3.69	3.69	3.56	3.67	3.78
3		3.27	4.00	3.96	3.99	3.95	3.70	3.71	3.68	3.59	3.68	3.75
4		2.86	3.39	3.98	4.00	3.98	3.57	3.65	3.57	3.58	3.67	3.62
5		2.86	3.41	3.88	4.00	3.90	3.58	3.54	3.58	3.62	3.67	3.60
6		2.29	3.27	3.96	3.99	3.70	3.42	3.55	3.64	3.62	3.61	3.50
7		2.29	2.86	3.43	3.99	3.52	3.49	3.45	3.61	3.58	3.55	3.38
8		2.29	2.86	3.42	3.42	3.26	3.44	3.45	3.66	3.51	3.52	3.28
9		2.29	2.29	2.86	3.42	3.07	2.99	3.32	3.58	3.40	3.47	3.07
10		2.67	2.67	3.32	3.99	3.43	3.31	3.48	3.64	3.33	3.43	3.33
15		2.40	2.40	3.20	3.05	3.16	3.33	3.72	3.26	3.13	3.42	3.11
20		3.00	3.00	3.00	3.00	2.97	3.75	3.35	2.86	2.89	3.49	3.13
Grand Total		**1.58**	**1.94**	**2.20**	**2.30**	**2.38**	**2.58**	**2.74**	**2.86**	**2.89**	**2.97**	**2.45**

Figure 87: Three-vehicle Ego acceleration average values by start versus goal position.

Traffic_Gap	100										
Average of Acceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0.00	0.00	0.00	0.00	0.37	1.45	1.79	1.77	2.10	2.02	0.95
-15	0.00	0.00	0.00	0.00	0.00	0.81	1.29	1.49	1.96	2.23	0.78
-10	0.00	0.00	0.00	0.00	0.00	0.00	0.41	1.35	1.43	1.50	0.47
-9	0.00	0.00	0.00	0.00	0.00	0.00	0.59	1.18	1.50	1.57	0.48
-8	0.00	0.00	0.00	0.00	0.00	0.00	0.64	1.12	1.28	1.17	0.42
-7	0.00	0.00	0.00	0.00	0.00	0.00	0.34	1.23	1.43	1.71	0.47
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	1.22	1.21	0.34
-5	0.00	0.00	0.00	0.00	0.00	0.25	0.07	0.58	1.14	1.33	0.34
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.50	1.28	1.37	0.32
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.48	1.34	1.52	0.35
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.33	0.95	1.36	0.30
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.32	0.58	1.03	0.22
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.71	1.03	0.21
1	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.83	3.67	3.28	3.88
2	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.82	3.70	3.76	3.93
3	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.77	3.70	3.72	3.92
4	4.00	4.00	4.00	4.00	4.00	4.00	3.98	3.72	3.67	3.76	3.91
5	4.00	4.00	4.00	4.00	4.00	4.00	3.98	3.71	3.63	3.69	3.90
6	4.00	4.00	4.00	4.00	4.00	4.00	3.98	3.66	3.68	3.71	3.90
7	4.00	4.00	4.00	4.00	4.00	4.00	3.97	3.59	3.70	3.70	3.90
8	4.00	4.00	4.00	4.00	4.00	4.00	3.96	3.56	3.61	3.60	3.87
9	4.00	4.00	4.00	4.00	4.00	4.00	3.95	3.58	3.62	3.65	3.88
10	4.00	4.00	4.00	4.00	4.00	4.00	3.94	3.42	3.51	3.66	3.85
15	4.00	4.00	4.00	4.00	4.00	3.98	3.46	2.61	3.12	3.47	3.66
20	4.00	4.00	4.00	4.00	3.99	3.73	3.03	2.40	2.41	3.79	3.54
Grand Total	**1.92**	**1.92**	**1.92**	**1.92**	**1.93**	**2.01**	**2.08**	**2.14**	**2.36**	**2.51**	**2.07**

Figure 88: Three-vehicle Ego acceleration average for 100 m initial traffic gap.

Traffic_Gap	50										
Average of Acceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0.00	1.02	1.70	1.97	2.00	2.14	2.26	2.45	2.52	2.52	1.86
-15	0.00	0.00	0.00	1.19	1.88	1.75	2.13	2.27	2.47	2.51	1.42
-10	0.00	0.00	0.00	0.00	0.58	1.45	1.85	2.20	2.38	2.50	1.10
-9	0.00	0.00	0.00	0.00	0.00	1.13	1.83	1.96	2.32	2.45	0.97
-8	0.00	0.00	0.00	0.00	0.00	1.03	1.53	2.05	2.25	2.44	0.93
-7	0.00	0.00	0.00	0.00	0.00	0.96	1.56	1.95	2.29	2.41	0.92
-6	0.00	0.00	0.00	0.00	0.00	0.00	1.34	1.79	2.20	2.31	0.76
-5	0.00	0.00	0.00	0.00	0.00	0.00	1.04	1.59	2.19	2.35	0.72
-4	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.68	2.17	2.34	0.72
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.42	1.62	2.14	2.29	0.65
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.06	1.58	1.98	2.28	0.59
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.22	1.07	1.96	2.28	0.55
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.69	1.96	2.23	0.49
1	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.97	3.58	3.71	3.93
2	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.97	3.51	3.72	3.92
3	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.91	3.49	3.61	3.90
4	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.85	3.62	3.52	3.90
5	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.78	3.53	3.48	3.88
6	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.80	3.42	3.41	3.86
7	4.00	4.00	4.00	4.00	4.00	4.00	3.98	3.83	3.31	3.35	3.85
8	4.00	4.00	4.00	4.00	4.00	4.00	3.94	3.81	3.19	3.28	3.82
9	4.00	4.00	4.00	4.00	4.00	4.00	3.94	3.65	2.97	3.17	3.77
10	4.00	4.00	4.00	4.00	4.00	4.00	3.85	3.56	2.82	3.05	3.73
15	4.00	4.00	4.00	4.00	4.00	3.98	3.65	2.91	2.35	3.13	3.60
20	4.00	4.00	4.00	4.00	3.90	3.80	3.33	2.41	2.41	2.98	3.48
Grand Total	**1.92**	**1.96**	**1.99**	**2.05**	**2.09**	**2.25**	**2.48**	**2.65**	**2.68**	**2.85**	**2.29**

Figure 89: Three-vehicle Ego acceleration average for 50 m initial traffic gap.

Traffic_Gap	25										
Average of Acceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	4.00	3.95	3.42	3.49	3.66	2.42	2.66	2.75	2.92	3.57	3.28
-15	3.25	3.87	3.32	3.56	3.03	3.07	3.08	3.18	3.45	3.42	3.32
-10	0.00	0.00	3.56	3.05	2.49	2.83	3.11	3.23	3.12	3.19	2.46
-9	0.00	0.00	3.41	2.57	2.74	3.00	3.00	3.15	3.13	3.15	2.42
-8	0.00	0.00	0.00	2.62	2.72	2.87	2.90	3.14	3.02	3.01	2.03
-7	0.00	0.00	0.00	2.00	2.83	2.85	3.03	3.00	3.04	3.05	1.98
-6	0.00	0.00	0.00	0.00	3.18	2.91	3.00	2.81	2.96	2.82	1.77
-5	0.00	0.00	0.00	0.00	3.34	2.98	2.90	2.95	2.73	2.90	1.78
-4	0.00	0.00	0.00	0.00	1.21	3.26	2.71	2.60	2.71	2.52	1.50
-3	0.00	0.00	0.00	0.00	0.06	3.20	2.60	2.72	2.30	2.43	1.33
-2	0.00	0.00	0.00	0.00	0.00	2.67	2.43	2.41	2.28	2.49	1.23
-1	0.00	0.00	0.00	0.00	0.00	1.03	2.62	2.20	2.33	2.49	1.07
0	0.00	0.00	0.00	0.00	0.00	0.00	2.53	2.20	2.29	2.47	0.95
1	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.75	3.50	3.48	3.87
2	4.00	4.00	4.00	4.00	4.00	3.98	3.97	3.70	3.59	3.46	3.87
3	4.00	4.00	4.00	4.00	4.00	4.00	3.99	3.87	3.54	3.52	3.89
4	4.00	4.00	4.00	4.00	4.00	4.00	3.93	3.58	3.46	3.51	3.85
5	4.00	4.00	4.00	4.00	3.97	4.00	3.76	3.42	3.43	3.37	3.79
6	4.00	4.00	4.00	4.00	4.00	3.99	3.61	3.55	3.36	3.20	3.77
7	4.00	4.00	4.00	4.00	4.00	3.94	3.36	3.72	3.12	3.03	3.71
8	4.00	4.00	4.00	4.00	4.00	3.75	3.43	3.65	2.84	2.80	3.65
9	4.00	4.00	4.00	4.00	4.00	3.49	3.31	3.72	2.51	2.62	3.57
10	4.00	4.00	3.91	4.00	3.98	3.35	3.47	3.69	2.36	2.60	3.54
15	0.00	0.00	4.00	3.27	3.79	3.42	3.68	3.06	3.09	3.02	2.73
20	0.00	0.00	0.00	0.00	0.00	3.45	3.03	2.61	2.74	3.17	1.50
Grand Total	**1.89**	**1.91**	**2.30**	**2.42**	**2.76**	**3.14**	**3.20**	**3.15**	**2.95**	**3.01**	**2.67**

Figure 90: Three-vehicle Ego acceleration average for 25 m initial traffic gap.

Traffic_Gap	15										
Average of Acceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-15	4.00	4.00	3.99	3.43	3.37	3.60	3.66	3.57	3.44	3.43	3.65
-10	4.00	4.00	2.41	2.47	3.28	3.25	3.40	3.62	3.06	3.25	3.27
-9	4.00	3.97	2.82	2.06	3.65	3.41	3.27	2.71	3.15	3.53	3.26
-8	0.00	2.80	3.08	2.42	3.67	3.61	3.12	2.99	3.53	3.63	2.88
-7	0.00	0.00	1.32	3.00	3.24	3.79	2.80	3.45	3.56	3.60	2.48
-6	0.00	0.00	0.00	0.97	3.19	3.84	2.94	3.51	3.56	3.55	2.16
-5	0.00	0.00	0.00	1.26	3.77	3.78	3.04	3.47	3.50	3.50	2.23
-4	0.00	0.00	0.00	0.92	2.28	3.87	3.23	3.55	3.45	3.40	2.07
-3	0.00	0.00	0.00	1.10	1.88	3.53	3.28	3.41	3.37	3.38	1.99
-2	0.00	0.00	0.00	0.00	1.20	3.64	3.30	3.24	3.27	3.44	1.81
-1	0.00	0.00	0.00	0.00	1.24	2.76	3.37	3.18	3.16	3.36	1.71
0	0.00	0.00	0.00	0.00	0.00	2.36	3.54	3.18	3.08	3.17	1.53
1	4.00	4.00	4.00	4.00	3.99	3.85	3.62	3.65	3.24	3.21	3.76
2	4.00	4.00	4.00	4.00	4.00	3.79	3.47	3.64	3.26	3.40	3.76
3	4.00	4.00	4.00	3.99	3.99	3.85	3.68	3.69	3.31	3.53	3.80
4	4.00	3.92	3.94	4.00	3.99	3.57	3.70	3.66	3.09	3.47	3.73
5	4.00	4.00	4.00	4.00	3.97	3.42	3.34	3.59	3.44	3.60	3.74
6	0.00	4.00	4.00	3.98	3.90	3.21	3.76	3.66	3.50	3.56	3.36
7	0.00	4.00	4.00	3.98	3.62	3.37	3.60	3.35	3.20	3.51	3.26
8	0.00	4.00	3.94	4.00	2.94	2.98	3.78	3.67	3.21	3.48	3.20
9	0.00	0.00	3.99	4.00	2.33	2.53	3.50	3.27	3.20	3.43	2.63
10	0.00	0.00	4.00	4.00	2.48	2.37	3.31	3.40	3.47	3.63	2.67
15	0.00	0.00	0.00	0.00	0.00	1.26	3.81	3.70	3.11	3.50	1.54
Grand Total	**1.39**	**2.03**	**2.33**	**2.50**	**2.87**	**3.29**	**3.41**	**3.44**	**3.31**	**3.46**	**2.80**

Figure 91: Three-vehicle Ego acceleration average for 15 m initial traffic gap.

Traffic_Gap	10										
Average of Acceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-10	3.96	3.88	3.95	3.08	3.19	3.62	3.81	3.61	3.66	3.65	3.64
-9	4.00	3.96	3.96	2.69	2.85	3.39	3.66	3.81	3.78	3.65	3.58
-8	4.00	3.99	4.00	2.11	2.25	3.37	3.78	3.79	3.64	3.65	3.46
-7	0.00	4.00	3.96	2.41	2.09	3.14	3.69	3.87	3.51	3.32	3.00
-6	0.00	4.00	2.50	2.31	2.15	2.50	3.69	3.59	3.47	3.66	2.79
-5	0.00	0.00	1.88	2.29	2.67	3.45	3.84	3.58	3.69	3.75	2.51
-4	0.00	0.00	0.00	0.94	2.76	3.24	3.94	3.67	3.83	3.76	2.21
-3	0.00	0.00	0.00	0.90	1.31	3.51	3.95	3.73	3.68	3.77	2.08
-2	0.00	0.00	0.00	0.56	1.12	3.29	3.95	3.67	3.73	3.57	1.99
-1	0.00	0.00	0.00	0.00	1.09	3.90	3.98	3.58	3.82	3.60	2.00
0	0.00	0.00	0.00	0.00	0.98	1.94	3.36	3.46	3.79	3.44	1.70
1	3.81	4.00	4.00	4.00	4.00	3.70	3.76	3.54	3.50	3.73	3.80
2	3.73	4.00	3.99	4.00	3.98	3.83	3.63	3.63	3.38	3.64	3.78
3	2.89	4.00	4.00	3.99	4.00	3.31	3.86	3.53	3.43	3.62	3.66
4	0.00	3.79	3.95	4.00	3.98	3.16	3.46	3.27	3.65	3.64	3.29
5	0.00	3.87	4.00	4.00	3.92	3.26	3.18	3.41	3.68	3.71	3.30
6	0.00	2.89	4.00	4.00	3.56	2.73	2.87	3.78	3.64	3.71	3.12
7	0.00	0.00	4.00	3.93	2.67	2.84	2.71	3.66	3.79	3.64	2.72
8	0.00	0.00	4.00	3.93	2.73	2.21	2.78	3.67	3.73	3.66	2.67
9	0.00	0.00	0.00	3.93	2.39	1.74	2.11	3.70	3.60	3.58	2.10
10	0.00	0.00	0.00	3.93	2.12	2.16	2.31	3.76	3.83	3.63	2.17
Grand Total	**1.07**	**2.02**	**2.48**	**2.71**	**2.66**	**3.06**	**3.44**	**3.63**	**3.66**	**3.64**	**2.84**

Figure 92: Three-vehicle Ego acceleration average for 10 m initial traffic gap.

Traffic_Gap	5										
Average of Acceleration_Average	**Goal Position (m)**										
Ego Initial Pos (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-9	3.57	4.00	4.00	4.00	2.73	3.08	2.87	3.47	3.55	3.88	3.51
-8	3.99	4.00	3.96	4.00	2.64	2.93	2.87	3.43	3.72	3.40	3.49
-7	4.00	3.72	3.91	3.62	2.89	2.45	2.82	3.66	3.78	3.76	3.46
-6	0.00	3.98	4.00	3.86	2.39	2.63	3.07	3.56	3.78	3.64	3.09
-5	0.00	3.11	4.00	3.39	2.27	2.76	3.14	3.63	3.79	3.77	2.99
-4	0.00	3.75	3.94	2.27	1.94	2.57	3.04	3.75	3.70	3.71	2.87
-3	0.00	0.00	4.00	2.44	2.32	2.67	3.00	3.93	3.62	3.79	2.58
-2	0.00	0.00	0.00	1.21	3.43	2.33	3.27	3.99	3.81	3.87	2.19
-1	0.00	0.00	0.00	0.75	1.14	2.14	3.31	3.90	3.81	3.81	1.89
0	0.00	0.00	0.00	0.00	1.05	3.02	3.42	3.98	3.92	3.84	1.92
1	0.00	3.74	4.00	4.00	4.00	3.04	3.03	3.06	3.44	3.51	3.18
2	0.00	3.98	4.00	4.00	4.00	2.72	2.76	3.08	3.47	3.72	3.17
3	0.00	4.00	3.72	3.98	3.64	2.76	2.47	2.97	3.63	3.75	3.09
4	0.00	0.00	3.98	3.99	3.86	2.26	2.46	2.92	3.58	3.76	2.68
5	0.00	0.00	3.17	4.00	3.46	2.40	2.49	3.12	3.60	3.80	2.61
6	0.00	0.00	3.75	3.94	2.42	2.02	2.63	3.06	3.74	3.68	2.52
7	0.00	0.00	0.00	4.00	2.35	2.25	2.52	3.11	3.93	3.64	2.18
8	0.00	0.00	0.00	0.00	1.16	3.13	2.28	3.23	3.96	3.82	1.76
9	0.00	0.00	0.00	0.00	0.75	1.14	2.46	3.16	3.87	3.81	1.52
Grand Total	**0.61**	**1.80**	**2.65**	**2.81**	**2.55**	**2.54**	**2.84**	**3.42**	**3.72**	**3.73**	**2.67**

Figure 93: Three-vehicle Ego acceleration average for 5 m initial traffic gap.

The following tables show scatter plots of acceleration and deceleration variable values compared to collisions. They mimic what is shown in the main body of text for the two-vehicle simulation.

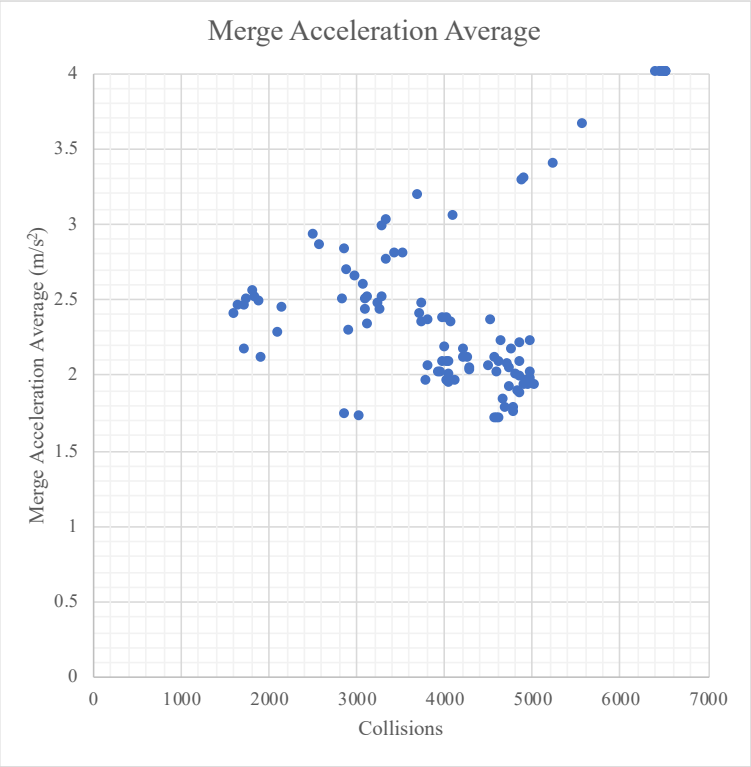


Figure 94: Three-vehicle Ego scatter plot for acceleration average versus collisions.

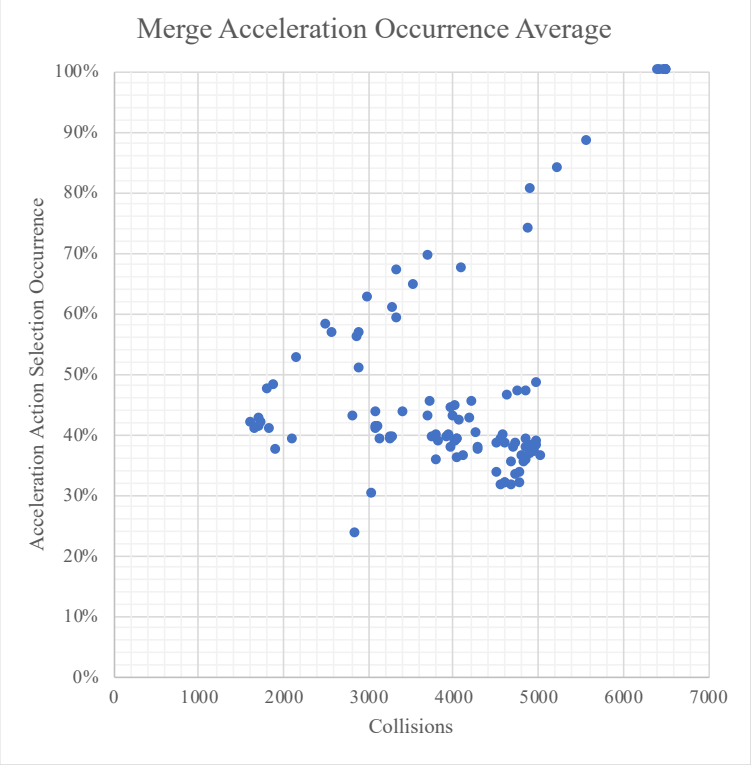


Figure 95: Three-vehicle Ego scatter plot for acceleration occurence versus collisions.

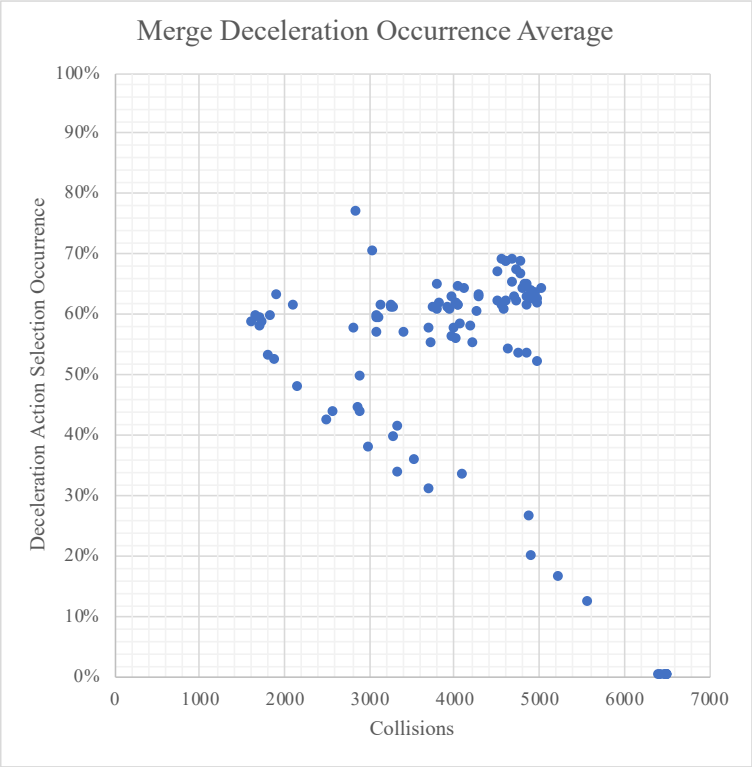


Figure 96: Three-vehicle Ego scatter plot for deceleration occurence versus collisions.

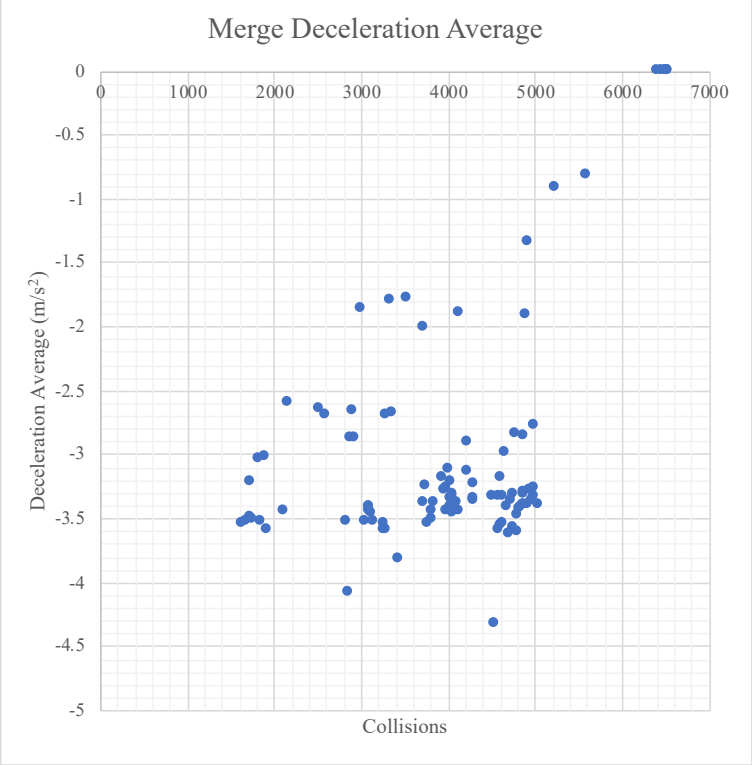


Figure 97: Three-vehicle Ego scatter plot for deceleration average versus collisions.

A.5 Three-Vehicle Simulation Results for Additional Traffic Test Gaps

Traffic_Reaction	['constant' 'constant']																																	
Average of Collision	Goal Position (m)																																	
	⊟10										10 Total	⊟25										25 Total	⊟100										100 Total	Grand Total
Ego Start Position (m)	10	20	30	40	50	60	70	80	90	100		10	20	30	40	50	60	70	80	90	100		10	20	30	40	50	60	70	80	90	100		
-50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-20												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%
-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%
0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%
1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	37%
2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	30%
3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	23%
4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	43%	38%	14%	0%	0%	0%	0%	0%	0%	0%	9.5%	36%	32%	16%	4%	0%	0%	0%	0%	0%	0%	8.8%	29%	26%	16%	6%	0%	0%	0%	0%	0%	0%	7.7%	9%

Figure 98: Three-vehicle simulation collision performance test results for 10, 25, and 100 m initial traffic gaps, constant traffic action.

Traffic_Reaction	['reactive' 'reactive']																																	
Average of Collision	Goal Position (m)																																	
	⊟10										10 Total	⊟25										25 Total	⊟100										100 Total	Grand Total
Ego Start Position (m)	10	20	30	40	50	60	70	80	90	100		10	20	30	40	50	60	70	80	90	100		10	20	30	40	50	60	70	80	90	100		
-50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-20												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-3	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
-1	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
0	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	30%
1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	33%
2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	23%
3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	23%
4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
5	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	43%	33%	10%	0%	0%	0%	0%	0%	0%	0%	8.6%	32%	28%	8%	0%	0%	0%	0%	0%	0%	0%	6.8%	26%	23%	13%	3%	0%	0%	0%	0%	0%	0%	6.5%	7%

Figure 99: Three-vehicle simulation collision performance test results for 10, 25, and 100 m initial traffic gaps, reactive traffic action.

Traffic Reaction	['random' 'random']																																		
Average of Collision	**Goal Position (m)**																																		
	⊟10										**10 Total**	**⊟25**										**25 Total**	**⊟100**									**100 Total**	**Grand Total**
Ego Start Position (m)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**		
-50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-20												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
-4	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%
-3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
-1	100%	100%	100%	33%	0%	0%	0%	0%	0%	0%	33%	100%	100%	100%	33%	0%	0%	0%	0%	0%	0%	33%	100%	100%	100%	33%	0%	0%	0%	0%	0%	0%	33%	33%
0	100%	100%	100%	67%	0%	0%	0%	0%	0%	0%	37%	100%	100%	100%	67%	0%	0%	0%	0%	0%	0%	37%	100%	100%	100%	67%	0%	0%	0%	0%	0%	0%	37%	37%
1	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	37%
2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	27%
3	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	33%	0%	0%	0%	0%	0%	0%	0%	23%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	30%	24%
4	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%
5	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
15												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
20												0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
30																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
40																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
50																							0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Grand Total	**43%**	**38%**	**14%**	**5%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**10.0%**	**36%**	**32%**	**17%**	**8%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**9.3%**	**29%**	**26%**	**16%**	**6%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**	**7.7%**	**9%**

Figure 100: Three-vehicle simulation collision performance test results for 10, 25, and 100 m initial traffic gaps, random traffic action.

A.6 Full-Scene Traffic DRL Network Diagram

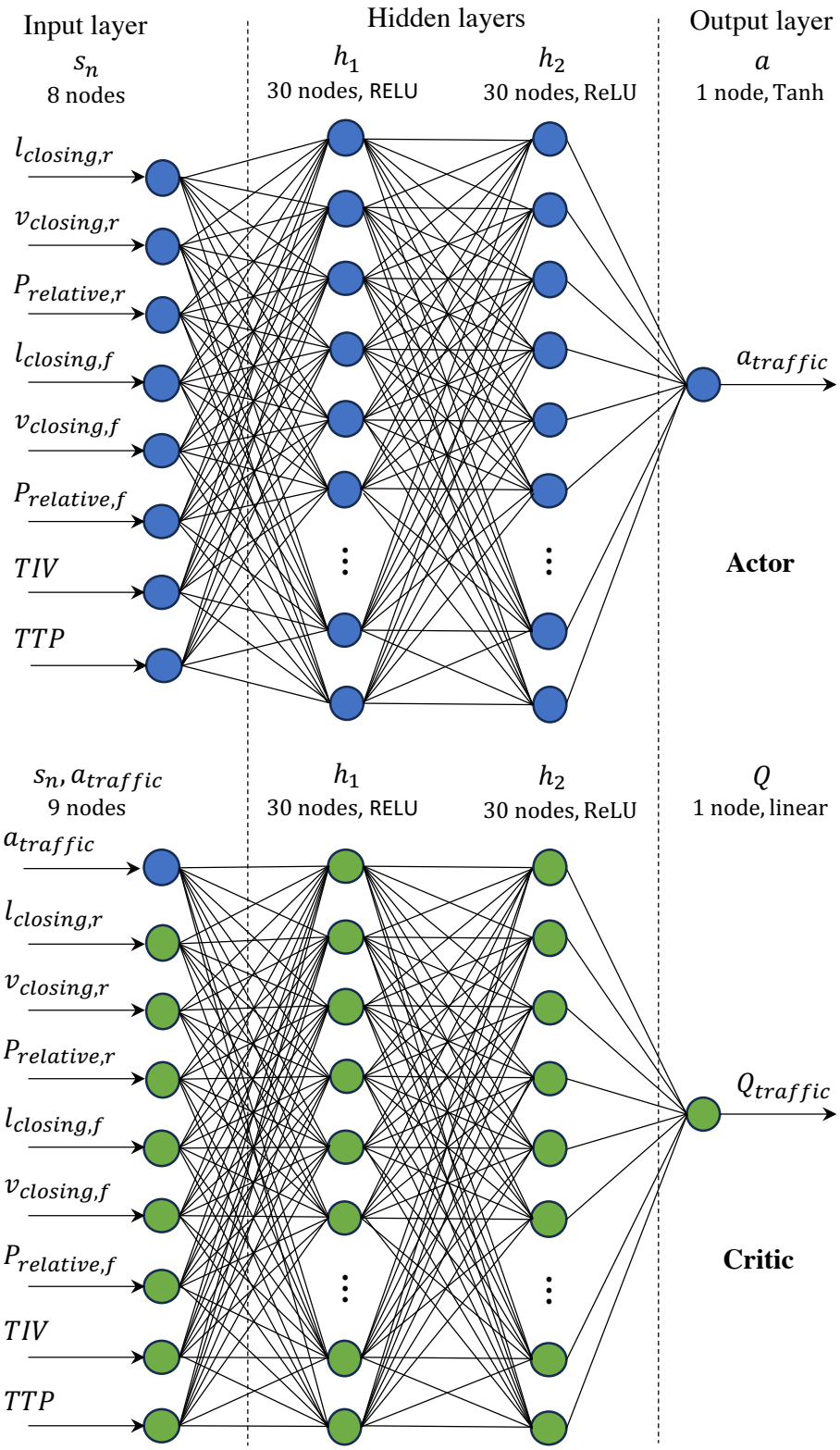


Figure 101: Full-scene DRL network diagram for traffic vehicles.

A.7 Full Scene Simulation Best Network Selection Additional Details

Table 19: Full scene vehicle DRL network save point performance metrics summary for traffic network showing 20 best performing networks (lowest collisions are best). The best performing traffic network, the one with the lowest number of total collisions, occurs at the same number of episodes as the merge vehicle network.

	Traffic Rear						Traffic Front					
Episode	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Maintain %**	**Accel %**	**Coll.**	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Maintain %**	**Accel %**	**Coll.**
4325000	**-4.59**	**1.38**	**84%**	**0%**	**16%**	**150**	**-4.23**	**3.12**	**44%**	**0%**	**56%**	**2409**
4350000	-4.67	0.66	89%	0%	11%	221	-4.42	2.99	50%	0%	50%	2441
4225000	-5.00	0.27	99%	0%	1%	189	-4.59	2.30	72%	0%	28%	3185
4375000	-5.00	0.43	98%	0%	2%	591	-3.72	2.74	46%	0%	54%	2966
4300000	-4.95	0.30	97%	0%	3%	203	-4.54	2.83	58%	0%	42%	3856
4275000	-4.55	1.19	83%	0%	17%	47	-4.56	2.98	50%	0%	50%	4635
4500000	-3.51	3.29	47%	0%	53%	1105	-2.12	3.51	24%	0%	76%	3842
3500000	-3.33	2.18	59%	0%	41%	397	-2.94	3.53	28%	0%	72%	4796
125000	-4.01	1.72	74%	0%	26%	951	-3.13	3.73	42%	0%	58%	4325
4250000	-3.60	3.65	35%	0%	65%	219	-2.84	3.36	33%	0%	67%	5094
3575000	-4.55	1.54	80%	0%	20%	443	-2.73	3.55	25%	0%	75%	4883
3450000	-2.71	2.35	51%	0%	49%	496	-0.22	3.95	2%	0%	98%	4862
3675000	-5.00	0.00	100%	0%	0%	1654	-3.92	2.91	51%	0%	49%	3774
3525000	-3.37	2.29	58%	0%	42%	610	-1.81	3.63	15%	0%	85%	4863
3475000	-3.50	1.79	64%	0%	36%	378	-0.80	3.68	9%	0%	91%	5177
1900000	-0.69	3.15	13%	0%	87%	1052	-0.87	2.84	19%	0%	81%	4552
375000	-3.55	1.51	70%	0%	30%	874	-2.81	2.98	44%	0%	56%	4755
1175000	-3.11	0.88	73%	0%	27%	1045	-2.13	2.56	40%	0%	60%	4587
1025000	-3.24	1.23	68%	0%	32%	1164	-2.27	2.91	35%	0%	65%	4531
1250000	-2.92	1.53	60%	0%	40%	1147	-1.75	2.72	29%	0%	71%	4556

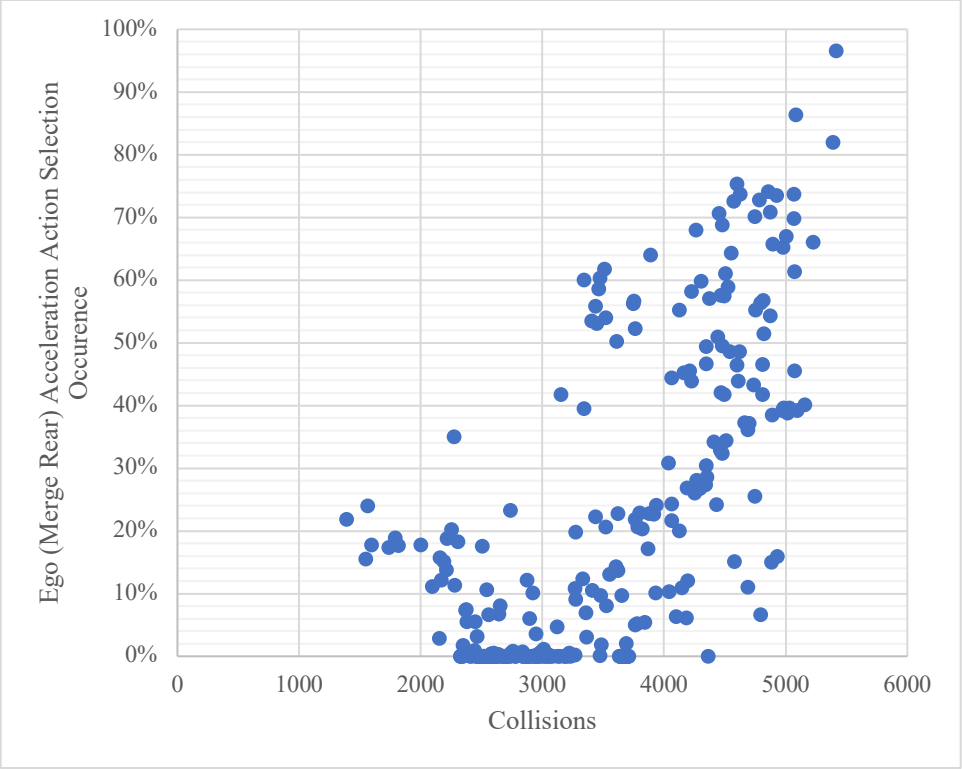


Figure 102 Full-scene Ego scatter plot for acceleration occurence versus collisions.

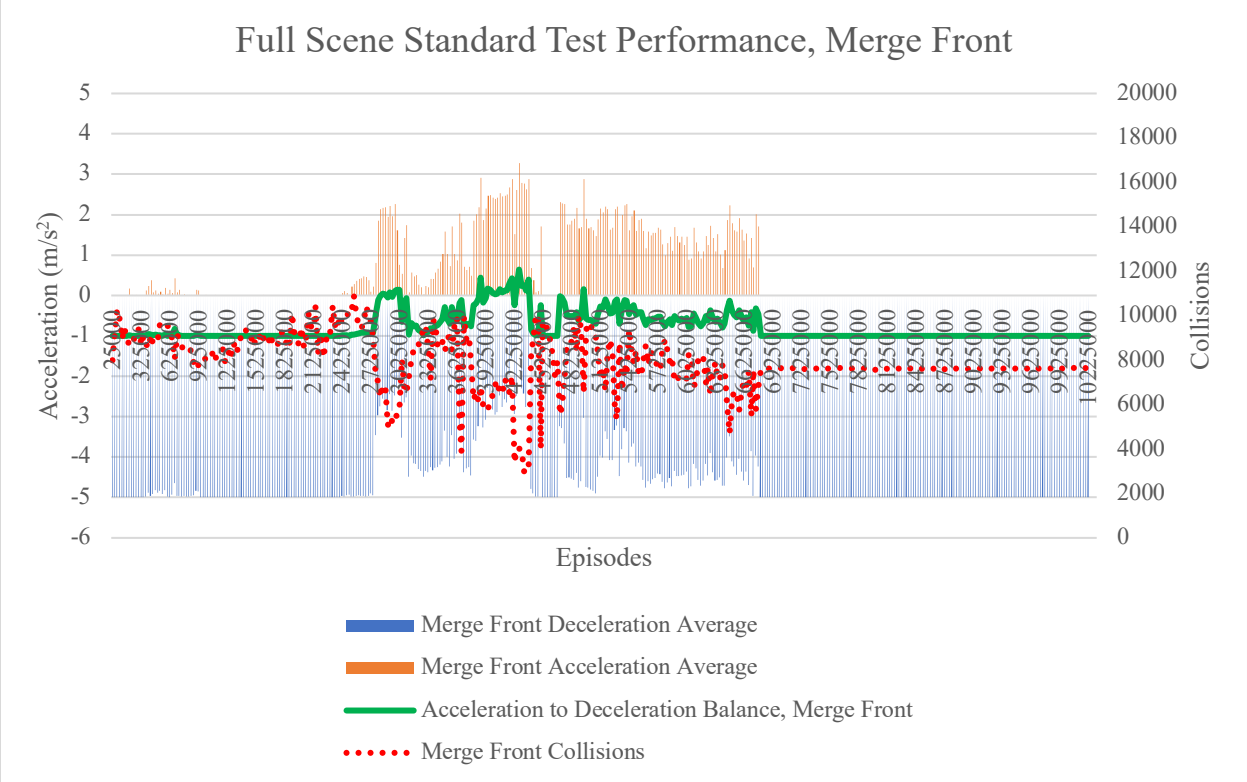


Figure 103: Full-scene multi-dimension plot for merge lane front vehicle.

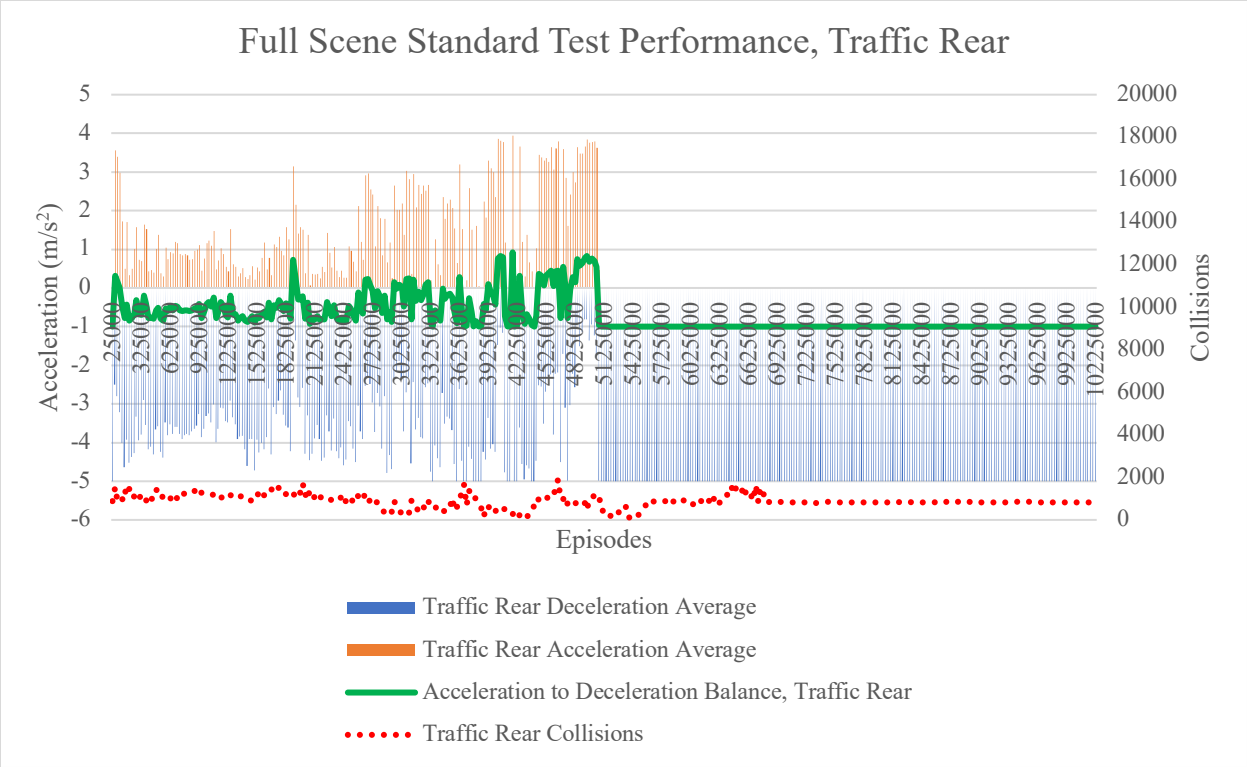


Figure 104: Full-scene multi-dimension plot for traffic lane rear vehicle.

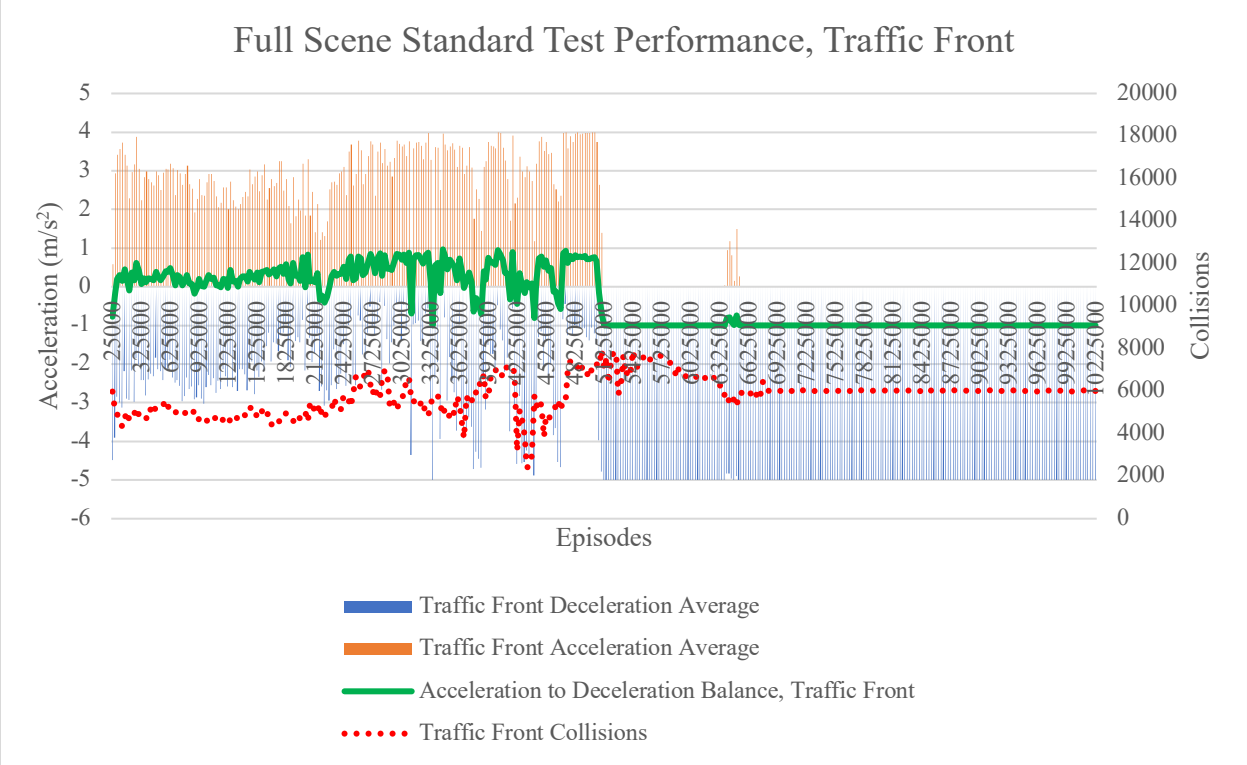


Figure 105: Full-scene multi-dimension plot for traffic lane front vehicle.

Traffic_Initial_Gap	(All)										
Average of MR_Deceleration_Occur_%	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	96%	94%	93%	90%	85%	80%	77%	73%	72%	86%
-15	100%	100%	96%	93%	91%	89%	84%	79%	75%	72%	88%
-10	100%	100%	100%	99%	96%	93%	87%	85%	79%	74%	91%
-9	100%	100%	98%	96%	94%	91%	86%	83%	79%	76%	90%
-8	100%	100%	99%	98%	96%	92%	88%	83%	81%	77%	92%
-7	100%	100%	100%	99%	97%	95%	90%	85%	83%	77%	93%
-6	100%	100%	100%	100%	99%	96%	94%	87%	84%	78%	94%
-5	100%	100%	100%	100%	100%	97%	95%	89%	84%	80%	95%
-4	100%	100%	100%	100%	100%	99%	96%	92%	85%	82%	95%
-3	100%	100%	100%	100%	100%	99%	97%	93%	87%	83%	96%
-2	100%	100%	100%	100%	100%	99%	98%	95%	90%	85%	97%
-1	100%	100%	100%	100%	100%	100%	97%	97%	92%	84%	97%
0	100%	100%	96%	91%	90%	90%	90%	90%	88%	86%	92%
1	57%	57%	46%	43%	41%	41%	43%	48%	49%	67%	49%
2	57%	57%	52%	47%	43%	44%	48%	50%	53%	69%	52%
3	71%	57%	57%	52%	46%	49%	52%	52%	56%	70%	56%
4	71%	57%	57%	53%	51%	54%	54%	56%	58%	71%	58%
5	71%	65%	64%	63%	61%	59%	61%	60%	60%	60%	62%
6	71%	71%	68%	67%	66%	64%	63%	62%	62%	60%	65%
7	71%	71%	70%	68%	67%	67%	66%	64%	63%	62%	67%
8	71%	71%	71%	69%	69%	68%	67%	65%	64%	63%	68%
9	71%	71%	71%	72%	71%	70%	69%	67%	66%	65%	69%
10	67%	67%	67%	68%	66%	67%	66%	64%	63%	62%	66%
15	60%	60%	64%	67%	67%	69%	67%	67%	65%	65%	65%
20	75%	75%	75%	71%	68%	66%	65%	65%	64%	64%	69%
Grand Total	**85%**	**83%**	**82%**	**80%**	**79%**	**78%**	**76%**	**74%**	**72%**	**72%**	**78%**

Figure 106: Full-scene Ego deceleration occurrence average for all initial traffic gaps.

Traffic_Initial_Gap	100										
Average of MR_Deceleration_Occur_%	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	100%	100%	100%	100%	93%	88%	85%	83%	84%	93%
-15	100%	100%	100%	100%	100%	100%	97%	91%	87%	86%	96%
-10	100%	100%	100%	100%	100%	100%	100%	100%	91%	90%	98%
-9	100%	100%	100%	100%	100%	100%	100%	100%	92%	91%	98%
-8	100%	100%	100%	100%	100%	100%	100%	100%	98%	93%	99%
-7	100%	100%	100%	100%	100%	100%	100%	100%	99%	94%	99%
-6	100%	100%	100%	100%	100%	100%	100%	100%	100%	93%	99%
-5	100%	100%	100%	100%	100%	100%	100%	100%	100%	94%	99%
-4	100%	100%	100%	100%	100%	100%	100%	100%	100%	98%	100%
-3	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
-1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
0	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
1	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	10%
2	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	10%
3	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	10%
4	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	10%
5	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%	1%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%	1%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%	1%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	1%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%	1%
10	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	1%
15	0%	0%	0%	0%	0%	0%	0%	0%	0%	12%	1%
20	0%	0%	0%	0%	0%	0%	0%	0%	0%	13%	1%
Grand Total	**52%**	**52%**	**52%**	**52%**	**52%**	**52%**	**51%**	**51%**	**50%**	**68%**	**53%**

Figure 107: Full-scene Ego deceleration occurrence average for 100 m initial traffic gap.

Traffic_Initial_Gap	50										
Average of MR_Deceleration_Occur_%	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	100%	100%	100%	92%	78%	67%	61%	56%	55%	81%
-15	100%	100%	100%	100%	95%	94%	79%	69%	62%	58%	86%
-10	100%	100%	100%	100%	98%	89%	75%	79%	67%	63%	87%
-9	100%	100%	100%	100%	100%	89%	78%	74%	70%	63%	87%
-8	100%	100%	100%	100%	100%	90%	77%	64%	73%	64%	87%
-7	100%	100%	100%	100%	100%	93%	80%	67%	69%	66%	87%
-6	100%	100%	100%	100%	100%	97%	89%	72%	71%	67%	90%
-5	100%	100%	100%	100%	100%	100%	92%	73%	64%	68%	90%
-4	100%	100%	100%	100%	100%	100%	90%	80%	65%	71%	91%
-3	100%	100%	100%	100%	100%	100%	93%	81%	66%	69%	91%
-2	100%	100%	100%	100%	100%	100%	99%	86%	78%	75%	94%
-1	100%	100%	100%	100%	100%	100%	100%	91%	78%	71%	94%
0	100%	100%	100%	100%	100%	100%	100%	94%	80%	70%	94%
1	0%	0%	0%	0%	0%	0%	0%	0%	0%	22%	2%
2	0%	0%	0%	0%	0%	0%	0%	0%	7%	25%	3%
3	0%	0%	0%	0%	0%	0%	0%	0%	17%	27%	4%
4	0%	0%	0%	0%	0%	0%	0%	9%	24%	27%	6%
5	0%	0%	0%	0%	5%	16%	19%	25%	27%	30%	12%
6	0%	0%	0%	1%	6%	13%	20%	27%	29%	29%	12%
7	0%	0%	0%	0%	9%	19%	25%	27%	31%	32%	14%
8	0%	0%	0%	0%	12%	21%	23%	28%	32%	32%	15%
9	0%	0%	0%	10%	17%	23%	27%	31%	35%	35%	18%
10	0%	0%	2%	9%	16%	25%	30%	33%	38%	36%	19%
15	0%	0%	20%	33%	37%	43%	45%	47%	47%	46%	32%
20	100%	100%	100%	86%	70%	65%	62%	60%	59%	56%	76%
Grand Total	**56%**	**56%**	**57%**	**58%**	**58%**	**58%**	**55%**	**51%**	**50%**	**50%**	**55%**

Figure 108: Full-scene Ego deceleration occurrence average for 50 m initial traffic gap.

Traffic_Initial_Gap	25										
Average of MR_Deceleration_Occur_%	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	100%	83%	75%	73%	70%	70%	65%	62%	53%	49%	70%
-15	100%	100%	100%	93%	87%	81%	76%	68%	63%	55%	82%
-10	100%	100%	100%	100%	100%	97%	86%	78%	69%	59%	89%
-9	100%	100%	100%	100%	100%	100%	88%	79%	72%	60%	90%
-8	100%	100%	100%	100%	100%	100%	92%	81%	72%	63%	91%
-7	100%	100%	100%	100%	99%	100%	97%	84%	75%	63%	92%
-6	100%	100%	100%	100%	100%	97%	99%	87%	79%	66%	93%
-5	100%	100%	100%	100%	100%	98%	99%	90%	78%	68%	93%
-4	100%	100%	100%	100%	100%	96%	98%	93%	80%	70%	94%
-3	100%	100%	100%	100%	100%	92%	97%	96%	83%	73%	94%
-2	100%	100%	100%	100%	100%	95%	89%	96%	87%	74%	94%
-1	100%	100%	100%	100%	100%	99%	85%	98%	87%	77%	95%
0	100%	100%	100%	75%	71%	79%	77%	72%	82%	76%	83%
1	0%	0%	0%	0%	0%	0%	6%	38%	49%	48%	14%
2	0%	0%	0%	0%	0%	1%	23%	41%	52%	51%	17%
3	100%	0%	0%	0%	0%	17%	39%	52%	54%	52%	31%
4	100%	0%	0%	0%	0%	39%	42%	54%	55%	55%	34%
5	100%	53%	50%	44%	40%	41%	57%	58%	59%	56%	56%
6	100%	100%	75%	70%	62%	60%	59%	59%	59%	58%	70%
7	100%	100%	88%	73%	69%	64%	64%	62%	62%	61%	74%
8	100%	100%	100%	83%	74%	69%	66%	66%	65%	61%	78%
9	100%	100%	100%	91%	77%	74%	71%	69%	66%	64%	81%
10	100%	100%	100%	100%	83%	75%	72%	71%	69%	66%	84%
15	100%	100%	100%	100%	100%	100%	92%	87%	81%	76%	94%
20	100%	100%	100%	100%	100%	100%	100%	99%	97%	87%	98%
Grand Total	**92%**	**81%**	**79%**	**76%**	**73%**	**74%**	**74%**	**74%**	**70%**	**64%**	**76%**

Figure 109: Full-scene Ego deceleration occurrence average for 25 m initial traffic gap.

Traffic_Initial_Gap	15										
Average of MR_Deceleration_Occur_%	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-15	100%	100%	79%	71%	73%	71%	70%	69%	64%	59%	76%
-10	100%	100%	100%	100%	92%	86%	83%	80%	73%	62%	88%
-9	100%	100%	100%	100%	95%	88%	85%	82%	75%	67%	89%
-8	100%	100%	100%	100%	100%	92%	87%	83%	78%	67%	91%
-7	100%	100%	100%	100%	100%	96%	92%	86%	81%	68%	92%
-6	100%	100%	100%	100%	100%	100%	94%	87%	82%	72%	94%
-5	100%	100%	100%	100%	100%	100%	96%	92%	85%	74%	95%
-4	100%	100%	100%	100%	100%	100%	100%	95%	87%	76%	96%
-3	100%	100%	100%	100%	100%	100%	100%	96%	89%	77%	96%
-2	100%	100%	100%	100%	100%	100%	100%	98%	91%	78%	97%
-1	100%	100%	100%	100%	100%	100%	97%	99%	95%	70%	96%
0	100%	100%	75%	76%	77%	75%	77%	87%	86%	89%	84%
1	100%	100%	22%	35%	41%	46%	58%	58%	59%	62%	58%
2	100%	100%	67%	44%	47%	51%	61%	61%	63%	66%	66%
3	100%	100%	100%	61%	52%	62%	62%	62%	67%	67%	73%
4	100%	100%	100%	71%	71%	68%	71%	70%	70%	67%	79%
5	100%	100%	100%	100%	79%	71%	74%	71%	71%	70%	84%
6	100%	100%	100%	100%	91%	82%	76%	75%	73%	71%	87%
7	100%	100%	100%	100%	93%	87%	82%	78%	75%	72%	89%
8	100%	100%	100%	100%	100%	90%	86%	80%	79%	75%	91%
9	100%	100%	100%	100%	100%	94%	87%	82%	81%	77%	92%
10	100%	100%	100%	100%	100%	100%	92%	86%	83%	78%	94%
15	100%	100%	100%	100%	100%	100%	100%	100%	97%	91%	99%
Grand Total	**100%**	**100%**	**93%**	**90%**	**87%**	**85%**	**84%**	**81%**	**78%**	**72%**	**87%**

Figure 110: Full-scene Ego deceleration occurrence average for 15 m initial traffic gap.

Traffic_Initial_Gap	10										
Average of MR_Deceleration_Occur_%	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-10	100%	100%	100%	92%	85%	82%	79%	76%	74%	68%	86%
-9	100%	100%	100%	100%	90%	86%	83%	78%	73%	69%	88%
-8	100%	100%	100%	100%	95%	88%	86%	81%	77%	72%	90%
-7	100%	100%	100%	100%	95%	92%	86%	84%	79%	73%	91%
-6	100%	100%	100%	100%	100%	94%	90%	85%	81%	74%	92%
-5	100%	100%	100%	100%	100%	96%	91%	89%	83%	76%	93%
-4	100%	100%	100%	100%	100%	100%	94%	92%	84%	79%	95%
-3	100%	100%	100%	100%	100%	100%	96%	93%	87%	81%	96%
-2	100%	100%	100%	100%	100%	100%	100%	95%	91%	83%	97%
-1	100%	100%	100%	100%	100%	100%	99%	97%	92%	85%	97%
0	100%	100%	100%	88%	85%	83%	83%	89%	85%	81%	89%
1	100%	100%	100%	75%	70%	68%	68%	69%	66%	66%	78%
2	100%	100%	100%	88%	67%	74%	70%	70%	75%	69%	81%
3	100%	100%	100%	100%	78%	77%	78%	73%	74%	70%	85%
4	100%	100%	100%	100%	88%	80%	81%	76%	76%	71%	87%
5	100%	100%	100%	100%	100%	88%	85%	82%	79%	74%	91%
6	100%	100%	100%	100%	100%	95%	87%	85%	84%	75%	93%
7	100%	100%	100%	100%	100%	100%	91%	86%	84%	78%	94%
8	100%	100%	100%	100%	100%	100%	94%	90%	86%	81%	95%
9	100%	100%	100%	100%	100%	100%	96%	92%	89%	84%	96%
10	100%	100%	100%	100%	100%	100%	100%	94%	91%	85%	97%
Grand Total	**100%**	**100%**	**100%**	**97%**	**93%**	**91%**	**87%**	**84%**	**81%**	**76%**	**91%**

Figure 111: Full-scene Ego deceleration occurrence average for 10 m initial traffic gap.

Traffic_Initial_Gap	5										
Average of MR_Deceleration_Occur_%	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-9	100%	100%	89%	75%	73%	73%	71%	71%	70%	80%	80%
-8	100%	100%	96%	87%	79%	78%	76%	72%	71%	79%	84%
-7	100%	100%	100%	94%	87%	83%	79%	76%	75%	72%	87%
-6	100%	100%	100%	100%	91%	87%	83%	80%	78%	74%	89%
-5	100%	100%	100%	100%	100%	89%	85%	81%	80%	77%	91%
-4	100%	100%	100%	100%	100%	95%	90%	86%	81%	78%	93%
-3	100%	100%	100%	100%	100%	100%	93%	87%	84%	80%	94%
-2	100%	100%	100%	100%	100%	100%	95%	89%	87%	82%	95%
-1	100%	100%	100%	100%	100%	100%	98%	92%	88%	85%	96%
0	100%	100%	100%	100%	100%	95%	93%	88%	85%	85%	95%
1	100%	100%	100%	89%	73%	74%	72%	71%	71%	70%	82%
2	100%	100%	100%	96%	86%	80%	79%	76%	74%	71%	86%
3	100%	100%	100%	100%	94%	87%	81%	79%	77%	75%	89%
4	100%	100%	100%	100%	100%	92%	86%	82%	80%	78%	92%
5	100%	100%	100%	100%	100%	99%	89%	85%	81%	80%	94%
6	100%	100%	100%	100%	100%	100%	98%	90%	85%	81%	95%
7	100%	100%	100%	100%	100%	100%	100%	93%	87%	83%	96%
8	100%	100%	100%	100%	100%	100%	100%	94%	89%	87%	97%
9	100%	100%	100%	100%	100%	100%	100%	98%	92%	89%	98%
Grand Total	**100%**	**100%**	**99%**	**97%**	**94%**	**91%**	**88%**	**84%**	**81%**	**79%**	**91%**

Figure 112: Full-scene Ego deceleration occurrence average for 5 m initial traffic gap.

Traffic_Initial_Gap	(All)										
Average of Merge_Ego_Rear_Deceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	-4.55	-4.53	-4.43	-4.31	-4.30	-4.41	-4.56	-4.73	-4.78	-4.91	-4.55
-15	-4.87	-4.50	-4.40	-4.30	-4.30	-4.20	-4.32	-4.54	-4.66	-4.76	-4.49
-10	-5.00	-4.95	-4.69	-4.41	-4.25	-4.17	-4.25	-4.29	-4.47	-4.63	-4.51
-9	-5.00	-4.97	-4.79	-4.44	-4.22	-4.17	-4.18	-4.25	-4.41	-4.39	-4.48
-8	-5.00	-4.99	-4.84	-4.50	-4.21	-4.14	-4.14	-4.32	-4.37	-4.40	-4.49
-7	-5.00	-5.00	-4.91	-4.58	-4.25	-4.13	-4.08	-4.18	-4.27	-4.46	-4.49
-6	-5.00	-5.00	-4.96	-4.72	-4.34	-4.14	-4.04	-4.13	-4.17	-4.42	-4.49
-5	-5.00	-5.00	-4.97	-4.83	-4.46	-4.18	-4.02	-4.07	-4.19	-4.31	-4.50
-4	-5.00	-5.00	-4.99	-4.91	-4.61	-4.22	-3.99	-3.93	-4.15	-4.22	-4.50
-3	-5.00	-5.00	-5.00	-4.96	-4.74	-4.32	-4.00	-3.89	-4.05	-4.19	-4.52
-2	-5.00	-5.00	-5.00	-4.99	-4.87	-4.41	-4.04	-3.87	-3.92	-4.09	-4.52
-1	-5.00	-5.00	-5.00	-5.00	-4.94	-4.55	-4.18	-3.83	-3.86	-4.13	-4.55
0	-5.00	-4.98	-4.87	-4.78	-4.70	-4.52	-4.21	-4.12	-4.05	-4.13	-4.54
1	-2.78	-2.29	-1.96	-2.05	-2.12	-2.21	-2.51	-2.98	-3.14	-4.64	-2.67
2	-2.83	-2.61	-2.19	-2.15	-2.11	-2.31	-2.66	-2.99	-3.34	-4.61	-2.78
3	-2.92	-2.77	-2.36	-2.13	-2.15	-2.47	-2.83	-3.04	-3.53	-4.56	-2.88
4	-3.25	-2.83	-2.59	-2.29	-2.24	-2.70	-2.85	-3.31	-3.67	-4.55	-3.03
5	-3.45	-3.14	-3.04	-2.78	-2.83	-3.00	-3.44	-3.59	-3.70	-4.50	-3.35
6	-3.53	-3.31	-3.19	-3.04	-2.99	-3.23	-3.45	-3.57	-3.67	-4.50	-3.45
7	-3.56	-3.43	-3.26	-3.16	-3.21	-3.29	-3.35	-3.54	-3.67	-4.40	-3.49
8	-3.57	-3.51	-3.28	-3.20	-3.32	-3.41	-3.40	-3.49	-3.61	-4.38	-3.52
9	-3.57	-3.55	-3.40	-3.40	-3.53	-3.41	-3.39	-3.46	-3.57	-4.33	-3.56
10	-3.33	-3.33	-3.24	-3.24	-3.29	-3.27	-3.30	-3.44	-3.53	-4.38	-3.44
15	-3.00	-3.00	-3.32	-3.42	-3.47	-3.32	-3.25	-3.17	-3.25	-4.25	-3.35
20	-3.23	-3.70	-3.37	-3.15	-3.28	-3.37	-3.18	-3.07	-3.02	-4.39	-3.38
Grand Total	**-4.15**	**-4.06**	**-3.92**	**-3.79**	**-3.70**	**-3.65**	**-3.65**	**-3.74**	**-3.88**	**-4.41**	**-3.90**

Figure 113: Full-scene Ego deceleration average for all initial traffic gaps.

Traffic_Initial_Gap	100										
Average of Merge_Ego_Rear_Deceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	-5.00	-5.00	-5.00	-5.00	-4.97	-4.86	-4.94	-5.00	-5.00	-5.00	-4.98
-15	-5.00	-5.00	-5.00	-5.00	-5.00	-4.99	-4.82	-4.94	-5.00	-5.00	-4.98
-10	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.94	-4.95	-4.95	-4.98
-9	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.98	-4.90	-4.98	-4.99
-8	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.92	-5.00	-4.99
-7	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.91	-4.94	-4.98
-6	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.95	-4.97	-4.99
-5	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.98	-4.85	-4.98
-4	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.85	-4.99
-3	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.91	-4.99
-2	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.98	-5.00
-1	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.99	-5.00
0	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-4.91	-0.49
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-4.94	-0.49
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-5.00	-0.50
Grand Total	**-2.60**	**-2.60**	**-2.60**	**-2.60**	**-2.60**	**-2.59**	**-2.59**	**-2.59**	**-2.58**	**-4.97**	**-2.83**

Figure 114: Full-scene Ego deceleration average for 100 m initial traffic gap.

Traffic_Initial_Gap	50										
Average of Merge_Ego_Rear_Deceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	-4.99	-4.84	-4.01	-3.25	-3.15	-3.53	-4.05	-4.52	-4.70	-4.93	-4.20
-15	-5.00	-5.00	-4.96	-4.27	-3.62	-2.95	-3.52	-4.23	-4.58	-4.84	-4.30
-10	-5.00	-5.00	-5.00	-5.00	-4.58	-3.97	-3.80	-3.71	-4.40	-4.77	-4.52
-9	-5.00	-5.00	-5.00	-5.00	-4.76	-4.26	-3.83	-3.73	-4.31	-4.77	-4.57
-8	-5.00	-5.00	-5.00	-5.00	-4.92	-4.37	-4.12	-4.48	-4.16	-4.83	-4.69
-7	-5.00	-5.00	-5.00	-5.00	-4.97	-4.50	-4.16	-4.25	-4.28	-4.64	-4.68
-6	-5.00	-5.00	-5.00	-5.00	-4.99	-4.67	-4.17	-4.17	-4.16	-4.68	-4.68
-5	-5.00	-5.00	-5.00	-5.00	-5.00	-4.76	-4.33	-4.33	-4.75	-4.57	-4.77
-4	-5.00	-5.00	-5.00	-5.00	-5.00	-4.94	-4.58	-4.28	-4.72	-4.46	-4.80
-3	-5.00	-5.00	-5.00	-5.00	-5.00	-4.98	-4.65	-4.38	-4.64	-4.69	-4.83
-2	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.77	-4.54	-4.42	-4.34	-4.81
-1	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.85	-4.63	-4.51	-4.64	-4.86
0	-5.00	-5.00	-5.00	-5.00	-5.00	-5.00	-4.96	-4.80	-4.65	-4.82	-4.92
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-4.41	-0.44
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.18	-4.28	-0.55
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.49	-4.25	-0.67
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.40	-3.51	-4.33	-0.92
5	0.00	0.00	0.00	0.00	-1.22	-2.21	-3.27	-3.65	-3.89	-4.15	-1.84
6	0.00	0.00	0.00	-0.04	-0.59	-2.65	-3.38	-3.66	-3.96	-4.15	-1.84
7	0.00	0.00	0.00	0.00	-1.41	-2.69	-3.10	-3.69	-3.99	-4.03	-1.89
8	0.00	0.00	0.00	0.00	-1.72	-3.19	-3.41	-3.73	-3.99	-4.26	-2.03
9	0.00	0.00	0.00	-0.99	-2.43	-2.82	-3.22	-3.83	-4.02	-4.15	-2.15
10	0.00	0.00	-0.05	-1.42	-2.12	-2.85	-3.42	-3.79	-3.94	-4.22	-2.18
15	0.00	0.00	-1.59	-2.15	-2.92	-3.14	-3.64	-3.92	-4.14	-4.21	-2.57
20	-2.92	-4.78	-3.47	-2.62	-3.11	-3.59	-3.79	-4.11	-4.24	-4.22	-3.68
Grand Total	**-2.72**	**-2.78**	**-2.76**	**-2.79**	**-3.06**	**-3.24**	**-3.32**	**-3.51**	**-3.91**	**-4.47**	**-3.26**

Figure 115: Full-scene Ego deceleration average for 50 m initial traffic gap.

Traffic_Initial_Gap	25										
Average of Merge_Ego_Rear_Deceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	-3.19	-3.29	-3.71	-3.99	-4.07	-4.26	-4.27	-4.40	-4.40	-4.71	-4.03
-15	-4.95	-4.47	-3.63	-3.49	-3.67	-3.68	-3.80	-4.12	-4.25	-4.44	-4.05
-10	-5.00	-5.00	-4.91	-4.07	-3.22	-2.92	-3.26	-3.57	-3.84	-4.34	-4.01
-9	-5.00	-5.00	-4.96	-4.34	-3.29	-2.85	-3.13	-3.50	-3.86	-4.37	-4.03
-8	-5.00	-5.00	-4.99	-4.63	-3.45	-2.93	-2.94	-3.35	-3.84	-4.28	-4.04
-7	-5.00	-5.00	-5.00	-4.81	-3.76	-3.14	-2.72	-3.19	-3.63	-4.27	-4.05
-6	-5.00	-5.00	-5.00	-4.94	-4.11	-3.17	-2.75	-3.11	-3.47	-4.12	-4.07
-5	-5.00	-5.00	-5.00	-4.99	-4.42	-3.39	-2.83	-2.93	-3.40	-4.04	-4.10
-4	-5.00	-5.00	-5.00	-4.99	-4.67	-3.81	-2.89	-2.73	-3.31	-3.80	-4.12
-3	-5.00	-5.00	-5.00	-5.00	-4.74	-4.09	-3.02	-2.69	-3.20	-3.74	-4.15
-2	-5.00	-5.00	-5.00	-5.00	-4.96	-4.24	-3.39	-2.67	-3.06	-3.61	-4.19
-1	-5.00	-5.00	-5.00	-5.00	-4.99	-4.41	-3.82	-2.77	-2.92	-3.56	-4.25
0	-5.00	-5.00	-5.00	-5.00	-5.00	-4.79	-4.08	-3.85	-3.25	-3.57	-4.45
1	0.00	0.00	0.00	0.00	0.00	0.00	-0.87	-3.66	-4.54	-4.48	-1.36
2	0.00	0.00	0.00	0.00	0.00	-0.19	-1.91	-3.74	-4.45	-4.52	-1.48
3	-0.48	0.00	0.00	0.00	0.00	-1.09	-3.35	-4.29	-4.32	-4.41	-1.79
4	-2.77	0.00	0.00	0.00	0.00	-2.75	-3.09	-4.35	-4.48	-4.43	-2.19
5	-4.16	-2.02	-1.90	-2.29	-2.50	-2.72	-4.09	-4.20	-4.36	-4.49	-3.27
6	-4.70	-3.16	-2.77	-3.00	-3.52	-3.72	-4.11	-4.29	-4.40	-4.38	-3.81
7	-4.90	-4.05	-3.00	-3.22	-3.47	-3.87	-4.05	-4.24	-4.41	-4.38	-3.96
8	-4.97	-4.58	-3.03	-3.04	-3.44	-3.74	-4.05	-4.13	-4.28	-4.45	-3.97
9	-4.99	-4.85	-3.85	-3.07	-3.49	-3.60	-3.92	-4.06	-4.36	-4.36	-4.06
10	-5.00	-4.95	-4.42	-3.13	-3.32	-3.68	-3.81	-4.13	-4.24	-4.29	-4.10
15	-5.00	-5.00	-5.00	-4.96	-4.45	-3.53	-3.31	-3.52	-3.78	-3.86	-4.24
20	-5.00	-5.00	-5.00	-5.00	-5.00	-4.91	-3.92	-3.18	-2.86	-3.32	-4.32
Grand Total	**-4.20**	**-3.86**	**-3.65**	**-3.52**	**-3.34**	**-3.26**	**-3.34**	**-3.63**	**-3.88**	**-4.17**	**-3.68**

Figure 116: Full-scene Ego deceleration average for 25 m initial traffic gap.

Traffic_Initial_Gap	15										
Average of Merge_Ego_Rear_Deceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-15	-4.40	-3.03	-3.40	-3.74	-4.20	-4.39	-4.47	-4.41	-4.48	-4.52	-4.10
-10	-5.00	-4.92	-4.40	-3.80	-3.80	-4.07	-4.30	-4.11	-4.19	-4.37	-4.30
-9	-5.00	-4.97	-4.73	-4.07	-3.66	-4.03	-4.02	-3.96	-4.04	-4.14	-4.26
-8	-5.00	-4.99	-4.90	-4.27	-3.60	-3.73	-3.82	-3.86	-3.89	-4.15	-4.22
-7	-5.00	-5.00	-4.96	-4.55	-3.75	-3.54	-3.62	-3.58	-3.70	-4.01	-4.17
-6	-5.00	-5.00	-4.99	-4.79	-3.98	-3.53	-3.37	-3.55	-3.44	-3.89	-4.15
-5	-5.00	-5.00	-5.00	-4.94	-4.22	-3.54	-3.30	-3.28	-3.33	-3.81	-4.14
-4	-5.00	-5.00	-5.00	-4.98	-4.57	-3.77	-3.16	-2.99	-3.25	-3.74	-4.15
-3	-5.00	-5.00	-5.00	-5.00	-4.78	-3.88	-3.11	-2.92	-3.12	-3.62	-4.14
-2	-5.00	-5.00	-5.00	-5.00	-4.94	-4.10	-3.25	-2.84	-2.94	-3.46	-4.15
-1	-5.00	-5.00	-5.00	-5.00	-4.98	-4.37	-3.72	-2.93	-2.81	-3.85	-4.27
0	-5.00	-5.00	-5.00	-4.29	-4.26	-4.10	-3.58	-3.35	-3.21	-3.72	-4.15
1	-4.53	-1.47	-0.78	-1.93	-2.63	-3.30	-4.11	-4.30	-4.27	-4.57	-3.19
2	-4.84	-3.43	-1.48	-2.56	-2.90	-3.48	-4.03	-4.19	-4.56	-4.50	-3.60
3	-4.95	-4.42	-2.05	-2.35	-3.07	-3.86	-4.02	-4.20	-4.48	-4.41	-3.78
4	-4.98	-4.80	-3.29	-2.40	-3.49	-3.76	-4.20	-4.31	-4.41	-4.45	-4.01
5	-4.99	-4.94	-4.43	-3.00	-3.41	-3.78	-4.16	-4.38	-4.47	-4.40	-4.20
6	-5.00	-4.97	-4.60	-3.76	-3.50	-3.87	-4.21	-4.30	-4.41	-4.49	-4.31
7	-5.00	-4.99	-4.84	-4.13	-3.64	-3.81	-4.07	-4.22	-4.42	-4.37	-4.35
8	-5.00	-5.00	-4.95	-4.46	-3.73	-3.81	-3.94	-4.19	-4.35	-4.36	-4.38
9	-5.00	-5.00	-4.98	-4.76	-4.05	-3.68	-3.91	-4.14	-4.30	-4.17	-4.40
10	-5.00	-5.00	-5.00	-4.91	-4.42	-3.75	-3.87	-3.99	-4.18	-4.04	-4.42
15	-5.00	-5.00	-5.00	-5.00	-5.00	-4.94	-4.28	-3.43	-3.34	-3.23	-4.42
Grand Total	**-4.94**	**-4.65**	**-4.29**	**-4.07**	**-3.94**	**-3.87**	**-3.85**	**-3.80**	**-3.89**	**-4.10**	**-4.14**

Figure 117: Full-scene Ego deceleration average for 15 m initial traffic gap.

Traffic_Initial_Gap	10										
Average of Merge_Ego_Rear_Deceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-10	-4.98	-4.77	-3.82	-3.58	-3.90	-4.05	-4.15	-4.43	-4.45	-4.33	-4.24
-9	-4.99	-4.81	-4.19	-3.62	-3.83	-3.99	-4.05	-4.37	-4.34	-4.26	-4.25
-8	-5.00	-4.95	-4.50	-3.81	-3.65	-4.03	-3.96	-4.29	-4.31	-4.20	-4.27
-7	-5.00	-4.98	-4.78	-4.10	-3.72	-3.89	-4.02	-4.09	-4.02	-4.13	-4.27
-6	-5.00	-5.00	-4.92	-4.44	-3.78	-3.81	-3.96	-4.05	-3.96	-4.02	-4.30
-5	-5.00	-5.00	-4.98	-4.71	-4.05	-3.84	-3.83	-3.85	-3.76	-3.81	-4.28
-4	-5.00	-5.00	-4.99	-4.85	-4.30	-3.66	-3.67	-3.64	-3.75	-3.67	-4.25
-3	-5.00	-5.00	-5.00	-4.97	-4.56	-3.89	-3.63	-3.46	-3.56	-3.52	-4.26
-2	-5.00	-5.00	-5.00	-4.99	-4.80	-3.94	-3.40	-3.30	-3.30	-3.40	-4.21
-1	-5.00	-5.00	-5.00	-5.00	-4.92	-4.23	-3.54	-3.08	-3.12	-3.18	-4.21
0	-5.00	-4.88	-4.08	-4.16	-4.24	-3.91	-3.32	-3.28	-3.54	-3.45	-3.98
1	-4.96	-4.59	-2.94	-2.86	-3.19	-3.06	-3.57	-3.65	-3.88	-4.59	-3.73
2	-4.99	-4.86	-3.86	-2.93	-3.22	-3.65	-3.78	-3.87	-3.97	-4.49	-3.96
3	-5.00	-4.95	-4.51	-2.94	-3.32	-3.69	-3.56	-3.80	-4.22	-4.54	-4.05
4	-5.00	-4.98	-4.80	-3.78	-3.43	-3.84	-3.95	-4.15	-4.21	-4.46	-4.26
5	-5.00	-5.00	-4.94	-4.34	-3.51	-3.81	-3.99	-4.02	-4.11	-4.34	-4.31
6	-5.00	-5.00	-4.95	-4.53	-3.84	-3.58	-3.98	-3.98	-3.99	-4.46	-4.33
7	-5.00	-5.00	-4.98	-4.79	-4.17	-3.58	-3.81	-4.05	-4.13	-4.20	-4.37
8	-5.00	-5.00	-5.00	-4.93	-4.47	-3.81	-3.81	-3.94	-3.91	-4.09	-4.39
9	-5.00	-5.00	-5.00	-4.98	-4.74	-4.12	-3.80	-3.82	-3.88	-4.00	-4.43
10	-5.00	-5.00	-5.00	-5.00	-4.91	-4.33	-3.72	-3.76	-3.83	-3.74	-4.43
Grand Total	**-5.00**	**-4.94**	**-4.63**	**-4.25**	**-4.03**	**-3.84**	**-3.79**	**-3.85**	**-3.92**	**-4.04**	**-4.23**

Figure 118: Full-scene Ego deceleration average for 10 m initial traffic gap.

Traffic_Initial_Gap	5										
Average of Merge_Ego_Rear_Deceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-9	-5.00	-5.00	-4.64	-4.03	-4.03	-4.08	-4.20	-4.24	-4.44	-3.20	-4.29
-8	-5.00	-5.00	-4.52	-3.79	-3.87	-3.92	-4.15	-4.28	-4.44	-3.34	-4.23
-7	-5.00	-5.00	-4.63	-3.62	-3.56	-3.85	-4.05	-4.19	-4.34	-4.26	-4.25
-6	-5.00	-5.00	-4.85	-3.85	-3.52	-3.79	-4.02	-4.01	-4.24	-4.25	-4.25
-5	-5.00	-5.00	-4.84	-4.20	-3.54	-3.71	-3.82	-4.10	-4.14	-4.10	-4.24
-4	-5.00	-5.00	-4.95	-4.53	-3.72	-3.35	-3.65	-3.84	-4.00	-4.02	-4.21
-3	-5.00	-5.00	-4.99	-4.78	-4.10	-3.43	-3.56	-3.80	-3.82	-3.84	-4.23
-2	-5.00	-5.00	-5.00	-4.93	-4.39	-3.56	-3.49	-3.71	-3.74	-3.85	-4.27
-1	-5.00	-5.00	-5.00	-4.99	-4.66	-3.84	-3.31	-3.43	-3.67	-3.66	-4.26
0	-5.00	-5.00	-5.00	-4.98	-4.37	-3.85	-3.53	-3.54	-3.70	-3.35	-4.23
1	-5.00	-5.00	-5.00	-4.59	-4.06	-4.09	-4.03	-4.25	-4.27	-4.41	-4.47
2	-5.00	-5.00	-5.00	-4.54	-3.68	-3.84	-3.90	-4.12	-4.25	-4.47	-4.38
3	-5.00	-5.00	-5.00	-4.62	-3.68	-3.63	-3.86	-4.02	-4.19	-4.31	-4.33
4	-5.00	-5.00	-5.00	-4.86	-3.75	-3.56	-3.73	-3.95	-4.08	-4.18	-4.31
5	-5.00	-5.00	-5.00	-4.84	-4.18	-3.47	-3.55	-3.85	-4.08	-4.13	-4.31
6	-5.00	-5.00	-5.00	-4.95	-4.47	-3.78	-3.45	-3.75	-3.93	-4.03	-4.34
7	-5.00	-5.00	-5.00	-4.99	-4.77	-4.09	-3.44	-3.62	-3.73	-3.83	-4.35
8	-5.00	-5.00	-5.00	-5.00	-4.92	-4.33	-3.59	-3.45	-3.75	-3.62	-4.37
9	-5.00	-5.00	-5.00	-5.00	-4.98	-4.66	-3.85	-3.40	-3.46	-3.62	-4.40
Grand Total	**-5.00**	**-5.00**	**-4.92**	**-4.58**	**-4.12**	**-3.83**	**-3.75**	**-3.87**	**-4.01**	**-3.92**	**-4.30**

Figure 119: Full-scene Ego deceleration average for 5 m initial traffic gap.

Traffic_Initial_Gap	(All)										
Average of Merge_Ego_Rear_Acceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0.00	0.45	0.67	0.78	1.31	1.79	2.13	2.46	2.53	2.56	1.47
-15	0.00	0.00	0.47	0.95	1.24	1.43	1.93	2.31	2.65	2.78	1.38
-10	0.00	0.00	0.00	0.26	0.79	1.20	1.57	1.99	2.42	2.65	1.09
-9	0.00	0.00	0.08	0.31	0.87	1.33	1.75	2.05	2.51	2.57	1.15
-8	0.00	0.00	0.04	0.10	0.62	1.20	1.64	1.92	2.35	2.50	1.04
-7	0.00	0.00	0.00	0.06	0.28	1.09	1.42	1.88	2.19	2.53	0.94
-6	0.00	0.00	0.00	0.00	0.12	0.65	1.32	1.74	2.14	2.55	0.85
-5	0.00	0.00	0.00	0.00	0.00	0.29	1.01	1.54	2.01	2.51	0.73
-4	0.00	0.00	0.00	0.00	0.00	0.11	0.76	1.30	1.85	2.38	0.64
-3	0.00	0.00	0.00	0.00	0.00	0.04	0.52	1.08	1.66	2.12	0.54
-2	0.00	0.00	0.00	0.00	0.00	0.04	0.15	0.65	1.44	1.95	0.42
-1	0.00	0.00	0.00	0.00	0.00	0.02	0.11	0.43	1.07	1.79	0.34
0	0.00	0.00	0.11	0.43	0.48	0.57	0.69	0.93	1.38	1.49	0.61
1	1.59	1.70	1.97	2.48	2.92	3.08	3.11	3.06	3.10	2.61	2.56
2	1.40	1.68	1.91	2.27	2.61	2.93	3.04	3.16	3.09	2.58	2.47
3	1.14	1.61	1.71	2.01	2.43	2.77	2.96	3.08	3.12	2.58	2.34
4	1.14	1.45	1.69	1.92	2.25	2.59	2.92	2.97	3.08	2.60	2.26
5	1.14	1.15	1.56	1.60	1.93	2.39	2.79	2.94	3.02	3.18	2.17
6	1.14	1.14	1.35	1.54	1.84	2.28	2.52	2.95	3.07	3.11	2.09
7	1.14	1.14	1.21	1.43	1.56	1.94	2.41	2.77	2.98	3.07	1.96
8	1.14	1.14	1.13	1.37	1.50	1.81	2.28	2.64	2.89	3.03	1.89
9	1.14	1.14	1.12	1.24	1.41	1.65	1.94	2.45	2.76	2.96	1.78
10	1.32	1.31	1.27	1.25	1.58	1.74	2.17	2.66	2.97	2.97	1.92
15	1.37	1.19	1.32	1.44	1.46	1.52	1.87	2.10	2.41	2.69	1.74
20	1.00	1.00	1.00	1.51	1.68	1.79	1.84	1.88	2.02	2.56	1.63
Grand Total	**0.59**	**0.64**	**0.74**	**0.90**	**1.14**	**1.44**	**1.78**	**2.11**	**2.43**	**2.57**	**1.43**

Figure 120: Full-scene Ego acceleration average for all initial traffic gaps.

Traffic_Initial_Gap	100										
Average of Merge_Ego_Rear_Acceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0.00	0.00	0.00	0.00	0.00	0.73	1.68	2.45	2.66	2.67	1.02
-15	0.00	0.00	0.00	0.00	0.00	0.00	0.55	1.26	2.27	2.62	0.67
-10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.01	1.57	0.26
-9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.10	1.50	0.26
-8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.36	1.35	0.17
-7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.84	0.09
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.09
-5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.09
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.04
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.00	3.60
2	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.00	3.60
3	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.00	3.60
4	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.00	3.60
5	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
6	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
7	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
8	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
9	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
10	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
15	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.99	4.00	4.00
20	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.99	4.00	4.00
Grand Total	**1.92**	**1.92**	**1.92**	**1.92**	**1.92**	**1.95**	**2.01**	**2.07**	**2.22**	**1.79**	**1.96**

Figure 121: Full-scene Ego acceleration average for 100 m initial traffic gap.

Traffic_Initial_Gap	50										
Average of Merge_Ego_Rear_Acceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0.00	0.00	0.00	0.00	1.93	2.80	3.27	3.65	3.75	3.83	1.92
-15	0.00	0.00	0.00	0.00	0.11	0.75	2.10	3.23	3.63	3.74	1.36
-10	0.00	0.00	0.00	0.00	0.13	0.55	1.11	1.81	2.96	3.50	1.01
-9	0.00	0.00	0.00	0.00	0.00	0.64	1.07	1.35	2.79	3.44	0.93
-8	0.00	0.00	0.00	0.00	0.00	0.78	0.98	1.04	2.73	3.22	0.87
-7	0.00	0.00	0.00	0.00	0.00	0.62	0.90	1.09	2.07	3.12	0.78
-6	0.00	0.00	0.00	0.00	0.00	0.44	0.55	0.97	1.95	3.03	0.69
-5	0.00	0.00	0.00	0.00	0.00	0.00	0.56	1.14	1.34	2.89	0.59
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.76	1.08	1.19	2.76	0.58
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.74	1.06	1.00	2.28	0.51
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.76	0.89	1.93	0.38
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.84	1.54	0.31
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.96	1.14	0.26
1	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.94	3.99
2	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.92	3.82	3.97
3	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.91	3.81	3.97
4	4.00	4.00	4.00	4.00	4.00	4.00	4.00	3.91	3.85	3.80	3.96
5	4.00	4.00	3.99	3.84	3.70	3.84	3.80	3.84	3.86	3.85	3.87
6	4.00	3.99	3.97	3.78	3.78	3.72	3.77	3.82	3.82	3.83	3.85
7	4.00	3.99	3.92	3.79	3.74	3.80	3.84	3.79	3.83	3.86	3.86
8	3.99	3.99	3.94	3.65	3.72	3.66	3.74	3.78	3.80	3.77	3.80
9	3.98	3.97	3.82	3.69	3.64	3.75	3.78	3.72	3.80	3.81	3.80
10	3.94	3.89	3.62	3.48	3.63	3.71	3.73	3.79	3.85	3.77	3.74
15	2.85	1.94	2.58	3.19	3.32	3.60	3.63	3.70	3.70	3.77	3.23
20	0.00	0.00	0.00	2.02	2.72	3.15	3.34	3.51	3.58	3.63	2.20
Grand Total	**1.71**	**1.67**	**1.67**	**1.74**	**1.86**	**2.07**	**2.32**	**2.57**	**2.88**	**3.28**	**2.18**

Figure 122: Full-scene Ego acceleration average for 50 m initial traffic gap.

Traffic_Initial_Gap	25										
Average of Merge_Ego_Rear_Acceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-20	0.00	1.78	2.67	3.13	3.31	3.65	3.57	3.75	3.71	3.73	2.93
-15	0.00	0.00	0.00	1.82	2.81	3.05	3.46	3.43	3.65	3.78	2.20
-10	0.00	0.00	0.00	0.00	0.00	0.65	2.45	3.20	3.45	3.61	1.34
-9	0.00	0.00	0.00	0.00	0.00	0.03	2.01	3.06	3.45	3.56	1.21
-8	0.00	0.00	0.00	0.00	0.00	0.00	1.37	2.91	3.28	3.55	1.11
-7	0.00	0.00	0.00	0.00	0.03	0.02	0.46	2.72	3.12	3.43	0.98
-6	0.00	0.00	0.00	0.00	0.00	0.10	0.20	2.16	3.16	3.44	0.91
-5	0.00	0.00	0.00	0.00	0.00	0.07	0.02	1.37	2.85	3.49	0.78
-4	0.00	0.00	0.00	0.00	0.00	0.08	0.09	0.60	2.64	3.44	0.68
-3	0.00	0.00	0.00	0.00	0.00	0.27	0.10	0.59	2.35	3.21	0.65
-2	0.00	0.00	0.00	0.00	0.00	0.25	0.12	0.12	1.71	3.02	0.52
-1	0.00	0.00	0.00	0.00	0.00	0.17	0.41	0.06	1.03	2.73	0.44
0	0.00	0.00	0.00	1.33	1.33	1.38	1.58	1.49	1.40	2.62	1.11
1	3.14	3.92	3.99	4.00	3.98	3.86	3.76	3.75	3.70	3.79	3.79
2	1.80	3.76	3.98	3.99	3.94	3.70	3.69	3.74	3.76	3.78	3.61
3	0.00	3.30	3.94	3.96	3.88	3.65	3.62	3.69	3.82	3.82	3.37
4	0.00	2.16	3.81	3.96	3.89	3.67	3.70	3.58	3.67	3.79	3.22
5	0.00	0.03	2.97	3.34	3.44	3.62	3.54	3.68	3.79	3.69	2.81
6	0.00	0.00	1.52	2.98	3.15	3.51	3.48	3.56	3.69	3.74	2.56
7	0.00	0.00	0.53	2.21	3.11	3.26	3.49	3.52	3.62	3.75	2.35
8	0.00	0.00	0.00	1.93	2.82	3.33	3.32	3.58	3.72	3.59	2.23
9	0.00	0.00	0.00	1.01	2.26	3.25	3.37	3.62	3.56	3.69	2.08
10	0.00	0.00	0.00	0.00	1.86	2.74	3.15	3.36	3.64	3.59	1.84
15	0.00	0.00	0.00	0.00	0.00	0.00	1.73	2.78	3.14	3.46	1.11
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.52	2.62	0.32
Grand Total	**0.20**	**0.60**	**0.94**	**1.35**	**1.59**	**1.77**	**2.11**	**2.57**	**3.06**	**3.48**	**1.77**

Figure 123: Full-scene Ego acceleration average for 25 m initial traffic gap.

Traffic_Initial_Gap	15										
Average of Merge_Ego_Rear_Acceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-15	0.00	0.00	2.35	2.91	3.31	3.36	3.56	3.61	3.70	3.73	2.65
-10	0.00	0.00	0.00	0.00	2.09	2.84	3.11	3.40	3.51	3.61	1.86
-9	0.00	0.00	0.00	0.00	1.03	2.20	3.00	3.41	3.47	3.63	1.67
-8	0.00	0.00	0.00	0.00	0.00	1.85	2.49	3.16	3.49	3.54	1.45
-7	0.00	0.00	0.00	0.00	0.00	1.47	2.73	2.80	3.41	3.45	1.39
-6	0.00	0.00	0.00	0.00	0.00	0.00	2.34	2.62	3.26	3.44	1.17
-5	0.00	0.00	0.00	0.00	0.00	0.00	1.43	2.35	3.14	3.28	1.02
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.50	2.88	3.19	0.76
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.80	2.28	2.84	0.59
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	1.83	2.58	0.46
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.08	0.97	2.36	0.36
0	0.00	0.00	0.76	0.99	1.07	1.10	1.21	1.29	1.83	2.05	1.03
1	0.00	0.00	1.83	2.82	3.33	3.52	3.51	3.64	3.81	3.69	2.62
2	0.00	0.00	1.42	2.40	3.12	3.33	3.56	3.66	3.58	3.77	2.48
3	0.00	0.00	0.00	2.14	2.73	3.14	3.49	3.56	3.67	3.70	2.24
4	0.00	0.00	0.00	1.47	2.59	3.19	3.33	3.59	3.70	3.61	2.15
5	0.00	0.00	0.00	0.00	2.37	2.83	3.36	3.35	3.54	3.66	1.91
6	0.00	0.00	0.00	0.00	1.92	2.76	3.01	3.41	3.54	3.58	1.82
7	0.00	0.00	0.00	0.00	0.06	2.53	3.02	3.41	3.39	3.48	1.59
8	0.00	0.00	0.00	0.00	0.00	1.65	2.91	3.20	3.44	3.51	1.47
9	0.00	0.00	0.00	0.00	0.00	0.56	2.13	2.92	3.31	3.53	1.24
10	0.00	0.00	0.00	0.00	0.00	0.00	2.12	2.78	3.18	3.38	1.15
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.22	2.23	0.35
Grand Total	**0.00**	**0.00**	**0.28**	**0.55**	**1.03**	**1.58**	**2.19**	**2.55**	**3.05**	**3.30**	**1.45**

Figure 124: Full-scene Ego acceleration average for 15 m initial traffic gap.

Traffic_Initial_Gap	10										
Average of Merge_Ego_Rear_Acceleration_Average	**Goal Position**										
Ego Initial Position (starting differential)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-10	0.00	0.00	0.00	1.57	2.55	3.15	2.74	3.51	3.61	3.64	2.08
-9	0.00	0.00	0.00	0.00	2.33	3.18	3.19	3.41	3.51	3.69	1.93
-8	0.00	0.00	0.00	0.00	2.08	2.79	3.39	3.39	3.55	3.59	1.88
-7	0.00	0.00	0.00	0.00	0.08	2.66	2.67	3.29	3.44	3.61	1.58
-6	0.00	0.00	0.00	0.00	0.00	1.94	3.15	3.12	3.25	3.53	1.50
-5	0.00	0.00	0.00	0.00	0.00	0.57	2.38	2.94	3.32	3.48	1.27
-4	0.00	0.00	0.00	0.00	0.00	0.00	2.09	2.90	2.98	3.40	1.14
-3	0.00	0.00	0.00	0.00	0.00	0.00	1.05	2.46	2.94	3.32	0.98
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.40	2.73	2.87	0.70
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.78	2.12	2.72	0.56
0	0.00	0.00	0.00	0.72	0.97	1.00	1.14	1.21	2.59	2.21	0.98
1	0.00	0.00	0.00	1.92	3.10	3.48	3.17	3.11	3.03	3.59	2.14
2	0.00	0.00	0.00	1.21	2.34	3.07	3.02	3.33	3.24	3.64	1.99
3	0.00	0.00	0.00	0.00	2.02	2.97	2.83	3.14	3.15	3.52	1.76
4	0.00	0.00	0.00	0.00	1.29	2.53	3.01	2.81	3.00	3.64	1.63
5	0.00	0.00	0.00	0.00	0.00	2.29	3.21	3.08	2.93	3.71	1.52
6	0.00	0.00	0.00	0.00	0.00	2.01	2.64	3.34	3.36	3.42	1.48
7	0.00	0.00	0.00	0.00	0.00	0.00	2.50	2.79	3.17	3.42	1.19
8	0.00	0.00	0.00	0.00	0.00	0.00	1.97	3.10	3.25	3.38	1.17
9	0.00	0.00	0.00	0.00	0.00	0.00	0.33	2.73	3.29	3.21	0.96
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	3.15	3.07	0.82
Grand Total	**0.00**	**0.00**	**0.00**	**0.26**	**0.80**	**1.51**	**2.12**	**2.75**	**3.13**	**3.36**	**1.39**

Figure 125: Full-scene Ego acceleration average for 10 m initial traffic gap.

Traffic_Initial_Gap	5										
Average of Merge_Ego_Rear_Acceleration_Average Ego Initial Position (starting differential)	**Goal Position**										
	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**Grand Total**
-9	0.00	0.00	0.59	2.17	2.71	3.28	2.94	3.13	3.23	2.13	2.02
-8	0.00	0.00	0.26	0.70	2.29	2.95	3.26	2.95	3.06	2.23	1.77
-7	0.00	0.00	0.00	0.39	1.84	2.84	3.19	3.28	3.23	3.29	1.81
-6	0.00	0.00	0.00	0.00	0.83	2.09	3.00	3.34	3.39	3.52	1.62
-5	0.00	0.00	0.00	0.00	0.00	1.38	2.65	2.97	3.39	3.49	1.39
-4	0.00	0.00	0.00	0.00	0.00	0.70	2.36	3.00	3.26	3.47	1.28
-3	0.00	0.00	0.00	0.00	0.00	0.00	1.73	2.67	3.05	3.23	1.07
-2	0.00	0.00	0.00	0.00	0.00	0.00	0.75	2.09	2.90	3.25	0.90
-1	0.00	0.00	0.00	0.00	0.00	0.00	0.19	1.36	2.56	3.19	0.73
0	0.00	0.00	0.00	0.00	0.00	0.52	0.88	1.97	2.84	2.44	0.87
1	0.00	0.00	0.00	0.59	2.04	2.70	3.35	2.96	3.18	3.24	1.81
2	0.00	0.00	0.00	0.26	0.89	2.40	3.00	3.40	3.14	3.07	1.62
3	0.00	0.00	0.00	0.00	0.39	1.67	2.81	3.18	3.31	3.25	1.46
4	0.00	0.00	0.00	0.00	0.00	0.75	2.38	2.94	3.33	3.39	1.28
5	0.00	0.00	0.00	0.00	0.00	0.15	1.60	2.60	3.01	3.38	1.07
6	0.00	0.00	0.00	0.00	0.00	0.00	0.72	2.48	3.08	3.20	0.95
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.87	2.87	2.95	0.77
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.78	2.03	2.96	0.58
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	1.38	2.45	0.40
Grand Total	**0.00**	**0.00**	**0.04**	**0.22**	**0.58**	**1.13**	**1.83**	**2.48**	**2.96**	**3.06**	**1.23**

Figure 126: Full-scene Ego acceleration average for 5 m initial traffic gap.

A.8 Full Scene Additional Test Gaps

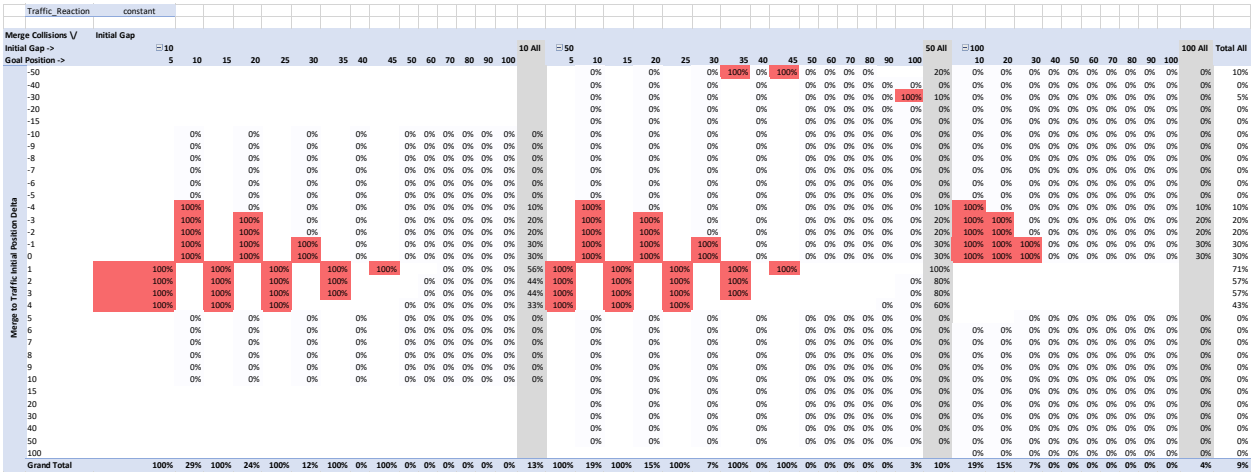


Figure 127: Full-scene simulation collision performance test results for 10, 50, and 100 m initial traffic gaps, constant traffic action.

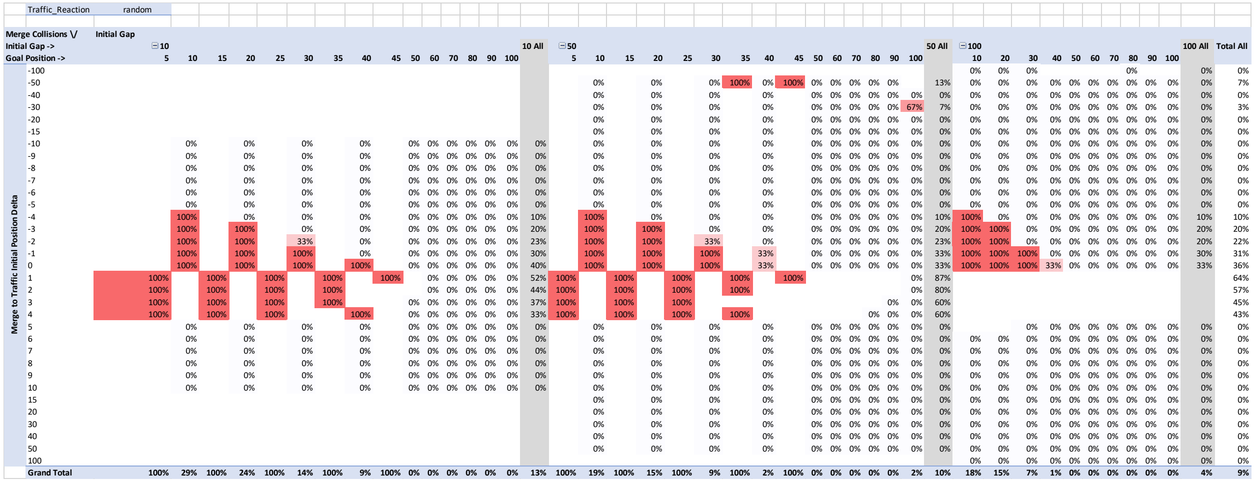


Figure 128: Full-scene simulation collision performance test results for 10, 50, and 100 m initial traffic gaps, random traffic action.

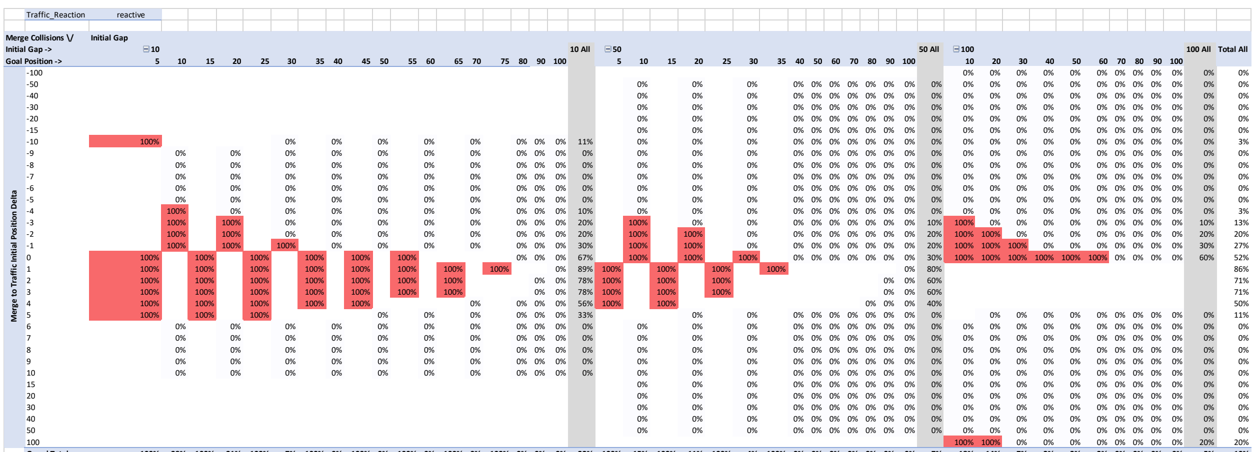


Figure 129: Full-scene simulation collision performance test results for 10, 50, and 100 m initial traffic gaps, reactive traffic action.

Appendix B An Intermediate Step to the Full-Scene Simulation

The setup in this appendix was an intermediate step toward the Full Scene Simulation. Overall, the four-vehicle results were poor, the states were more complicated, and the model did not fully represent the merge scenario. However, this intermediate step is highly representative of many other intermediate steps taken during progressive development. There were over 200 versions in total trialed during development and testing. There were about 75 versions of development in each scenario: two-vehicle DRL, three-vehicle, and four-vehicle. The different versions trialed different state sets, agents that learned, and those that did not. Yet, all contributed valuable insight into the merge problem and the DRL approach.

This simulation scenario consists of two vehicles in the merge lane and two in the traffic lane. This four-vehicle simulation adds a vehicle to the merge lane in front of the ego vehicle, bringing the total to four vehicles: two in each lane. This front vehicle must also longitudinally position itself to avoid a collision with the traffic vehicles as the rear ego merge vehicle does. However, it does not consider any states for the vehicle behind it. This is like a human driver responsible for focusing on vehicles in front of it with little consideration for a vehicle directly behind it. Instead, the responsibility for avoiding a collision rests with the vehicle behind it.

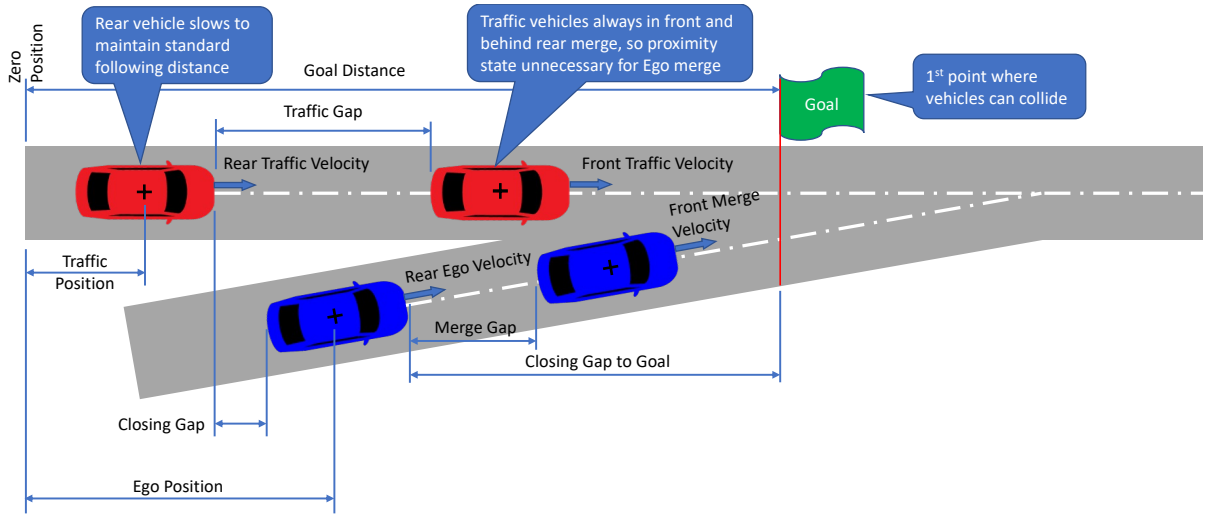


Figure 130: Four vehicle merge scene with two vehicles in the merge lane and two in traffic.

The additional vehicle makes the merge model more representative of a real-world merge scene. The ego merge (rear) vehicle is actively controlling its position to avoid a collision with traffic and must also avoid a collision with a vehicle in front of it. The merge vehicle in the rear is always the ego vehicle. The simulator adds a vehicle in front of it that selects actions based on the same DRL network and operates with the same physics as the rear (ego) vehicle. The front vehicle must avoid a collision with the traffic vehicles at the merge point, just like the ego vehicle. Once the front vehicle has merged into traffic, the simulation continues until the rear merge vehicle merges into traffic.

There are always only two traffic vehicles. The two traffic vehicles function the same as in the three-vehicle simulation. This concept is just like the three-vehicle scene. One traffic vehicle is always behind the ego merge vehicle, and one is always in front.

If the ego vehicle moves behind the rear traffic vehicle, that rear traffic vehicle now becomes the front traffic vehicle. A new traffic vehicle is generated behind the ego vehicle, becoming the rear. The old front traffic vehicle no longer exists in the simulation. The same process happens if the ego vehicle moves in front of the front traffic vehicle. The simulator

works this way to prevent the ego vehicle from learning a simple strategy to just accelerate or decelerate outside the range of the pair of traffic vehicles.

B.1 State Set

The state set increases with the additional vehicle. Table 20 details the parameters. The merge vehicle states increase from six to eight variables. The traffic vehicle state set increases from five to eight. The merge vehicle still considers all the states it previously did, but it adds another pair of state variables: closing gap and closing speed to the vehicle in front. This maintains the same strategy of adding a relative closing gap and relative closing speed state pair for all the other agents or objects of interest in the road scene.

Table 20: Four-vehicle states. The ego merge state set increases by two (over the three-vehicle scene) to a total size of six variables. After significant experimentation, the traffic vehicle's state set differs from the ego merge state set.

Vehicle State Variables	Used by Merge	Used by Traffic	Range	Units
Closing Gap to Rear Vehicle in Next Lane	✓	✓	[-2.5, 30]	m
Relative Closing Speed to Rear Vehicle in Next Lane	✓	✓	[-10, 10]	m/s
Closing Gap to Front Vehicle in Next Lane	✓	✓	[-2.5, 30]	m
Relative Closing Speed to Front Vehicle in Next Lane	✓	✓	[-10, 10]	m/s
Closing Gap to Front Vehicle in Same Lane	✓	✗	[-2.5, 30]	m
Relative Closing Speed to Front Vehicle in Same Lane	✓	✗	[-10, 10]	m/s
Closing Gap to Goal	✓	✗	[-160, 150]	m
Closing Velocity to Goal	✓	✗	[0, 40]	m/s
Closing Time in Between Current to Front Vehicle (TIV)	✗	✓	[0, 2.5]	s
Time to Goal Position	✗	✓	[0, 3]	s
Proximity to ego (*two variables, one for front and rear*)	✗	✓✓	{-1, 1}	*unitless*

After significant experimentation, the traffic vehicle state set differs from the ego merge vehicle. However, the theme of the state set remains mostly the same as the three-vehicle variant. The only variables added are a state-pair of closing gap and closing velocity. In addition, the traffic vehicle also considers a proximity to ego state to inform it whether the ego vehicle is in front or behind.

B.2 Four-Vehicle Simulator Functionality

The four-vehicle simulator is very similar to the three-vehicle simulator. It contains the same functions of step, reset, states, MIO, and render. Adding the merge vehicle in front of the ego vehicle led to changes in most of the algorithm's functions. Following is a summary of the changes by function.

The highway merge environment had many rewrites to change the ego variables into vectors, but the basic calculations and logic remained the same. Table 15 shows a comparison of hyperparameters between the different simulation variants. Overall, the parameters are similar, but there are some changes. For example, the maximum goal position changes from 150 m in previous simulations to 100 m in the four-vehicle setup. This was changed to ensure both vehicles would fit on the on-ramp for most of the simulation parameter combinations. In addition, a *done* flag hyperparameter is now set at initialization because the done logic in the step function changed with the additional merge vehicle.

The step function action input is now a vector that contains both the ego merge rear and merge front vehicle actions. The motion equation updates for position and velocity calculations, and the counter functionality are unchanged. Rewards are also calculated and assigned per the reward formula in Section 4.1.4.

Logic is introduced for the merge front vehicle collision checking and reward assignment. A reward for merge front that is the magnitude of its acceleration is assigned at each step function iteration, just like ego and traffic. An in-lane collision check is introduced. At each step of the episode, the step function checks to see if the rear and front merge vehicles have collided. Penalties are assigned to both if they collide. Ego (rear) is given an at-fault collision penalty, merge front is given a no-fault collision penalty, and the episode stops.

A check determines if there is a collision between the front merge vehicle and any traffic vehicles when it reaches the goal position. An at-fault penalty is assigned for collisions, and the episode stops. A merge reward is assigned if no collision occurs between the front merge and traffic vehicles. After the front merge vehicle reaches the goal position, the episode continues until the ego (rear) merge vehicle reaches the goal position. All reward logic, penalties, and episode completion in the step function are the same for the ego merge vehicle.

The reset function creates the front merge vehicle. While training, the front merge vehicle length is chosen randomly between the traffic minimum and maximum hyperparameters. The front merge vehicle in front of the ego (rear) at a gap with the hyperparameter limits. The goal position is now always placed in front of the front merge vehicle.

The states function was entirely rewritten but still produces the merge and traffic states as the output. The merge vehicles are now output as a vector, just like the traffic vehicles.

The MIO function remains almost the same, with some minor changes based on the state changes and the addition of the second merge vehicle.

The output file now writes 91 parameters at each episode instead of 32 for the three-vehicle version. Separate functions now write the header and output file lines to make updates more manageable. The render function adds the front merge vehicle in light blue. A “collision burst” graphic shows the site where collisions occur to help visually identify collisions. The DDPG algorithm is described in Section 4.2 and remains the same as other versions.

The main program and test loops remain essentially unchanged. During testing, the front merge and traffic vehicles start with the same initial gap between in-lane vehicles. This setup generally creates a situation where both front and rear merge vehicles pair with front and rear traffic vehicles at the same initial position delta.

B.3 Training and Performance

The four-vehicle simulation uses two separate RL networks: one for the merge vehicles and one for the traffic vehicles. At each episode's step, the front and rear merge vehicles train the same DRL network. The reward structure and formula are the same as all other DRL simulations. In future work, separate networks could be trialed for each agent or the learning could be performed just on a single agent within each lane to compare differences in learned behavior.

Each reactive strategy traffic vehicle trains the traffic DRL network. As Section 6.2 explains, the frequency of training of the traffic network is lower because each episode randomly chooses one of three traffic policies (constant, random, and reactive) for the duration of each episode.

Figure 131 shows the training results graphs. The graph's details are greater than the two or three-vehicle variant graphs, as Figure 46 shows. The figure now shows both the merge and traffic vehicle rewards. Each of the four vehicles has a separate curve for the moving average and cumulative average.

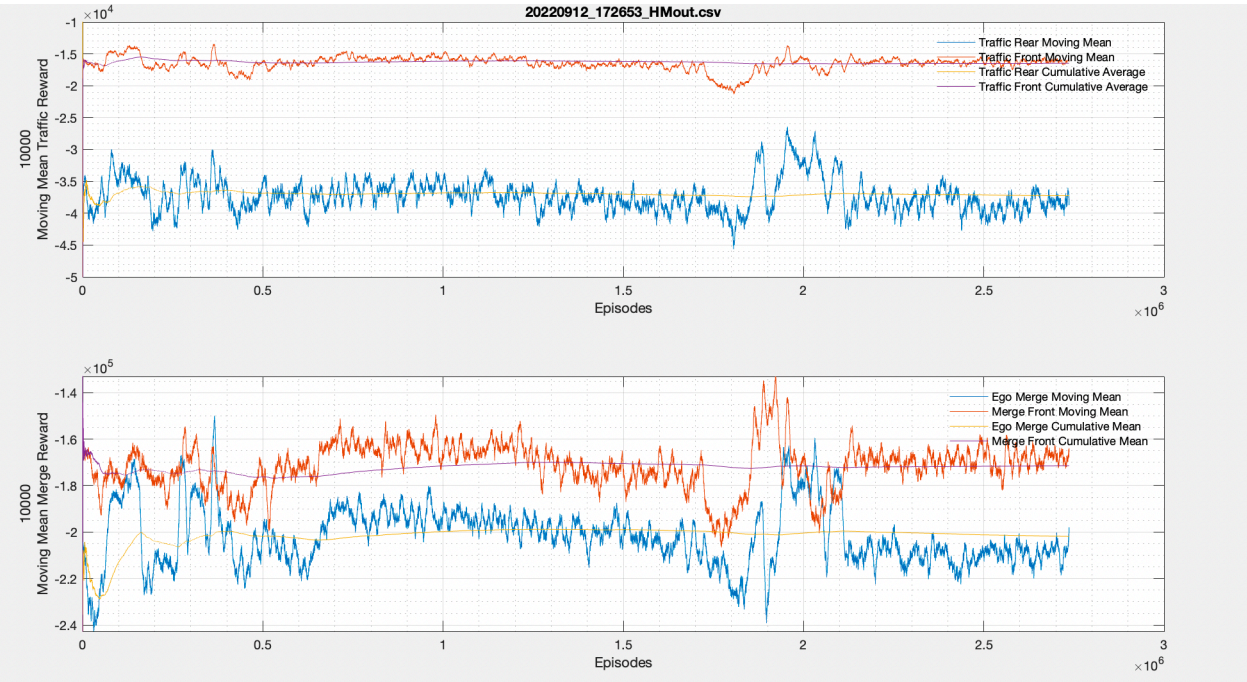


Figure 131: Four-vehicle DRL network training graphs. The best performing network is around 1.95M episodes and correlates with the peaking of the moving mean lines.

The moving average is now over 10,000 data points instead of 30,000 in earlier versions. The intent is the same, but the peaks and valleys are slightly more pronounced with the lower average data points. The graphs clearly show an abnormality right before 2 million episodes. The merge vehicle lines show a spike in the data, noticeably higher than the rest of the curve. Then, there is a dip in the data. Right after, the values tend to settle out to a steady state.

The graphs' behavior is similar to earlier variants but less pronounced. For example, Figure 46 shows a significant spike in the reward values that generally corresponds to a peak in the cumulative average, followed by a steady decrease in rewards. The same general behavior is observed, but it is less pronounced in Figure 131.

B.4 Best Network Selection

Table 21 and Table 22 list an abbreviated set of output values from the MATLAB output evaluation tool. The best performance for the merge vehicles is around 1.95M episodes and

worsens after. The worsening performance is common in network training. 1.95M episodes corresponds to a peak in the moving average graph, similar to peaks found in smaller scenes.

Table 21: Four-vehicle simulation merge vehicles training performance top 10 performers. Save points and performance testing happen at 25K episode intervals. The best performing network in terms of overall collisions is at 1.95M episodes.

	Merge Rear						Merge Front						
Episode	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Mtn %**	**Accel %**	**Coll.**	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Mtn %**	**Accel %**	**Coll.**	**Total Coll.**
1950000	-4.11	0.78	84%	0%	16%	2600	-2.63	1.71	57%	0%	43%	3367	5967
675000	-3.01	2.17	55%	0%	45%	2781	-3.20	2.24	54%	0%	46%	3363	6144
750000	-2.48	3.03	37%	0%	63%	2807	-3.04	2.36	50%	0%	50%	3388	6195
4550000	-5.00	0.00	100%	0%	0%	2488	-5.00	0.00	100%	0%	0%	3951	6439
4475000	-5.00	0.00	100%	0%	0%	2504	-5.00	0.00	100%	0%	0%	3986	6490
700000	-2.83	2.71	46%	0%	54%	2920	-3.11	2.32	52%	0%	48%	3579	6499
1000000	-2.45	3.18	36%	0%	64%	3078	-3.11	2.42	50%	0%	50%	3425	6503
3075000	-5.00	0.00	100%	0%	0%	2588	-5.00	0.00	100%	0%	0%	3915	6503
2950000	-5.00	0.00	100%	0%	0%	2629	-5.00	0.00	100%	0%	0%	3875	6504
975000	-2.41	3.16	36%	0%	64%	3114	-3.09	2.37	51%	0%	49%	3414	6528

Table 22: Four-vehicle simulation training performance top 10 performers. This table shows data for the front and rear traffic vehicles.

	Traffic Rear						Traffic Front						
Episode	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Mtn %**	**Accel %**	**Coll.**	**Decel Avg. (m/s^2)**	**Accel Avg. (m/s^2)**	**Decel %**	**Mtn %**	**Accel %**	**Coll.**	**Total Coll.**
1950000	-2.95	2.20	55%	0%	45%	4494	-3.79	2.22	67%	0%	33%	1473	5967
675000	-1.61	2.37	42%	0%	58%	3743	-0.03	3.82	1%	0%	99%	2401	6144
750000	-2.34	2.18	50%	0%	50%	3631	-0.52	3.43	10%	0%	90%	2564	6195
4550000	-4.95	0.42	97%	0%	3%	4387	-3.34	3.88	36%	0%	64%	2052	6439
4475000	-4.99	0.32	98%	0%	2%	4357	-3.74	3.61	42%	0%	58%	2133	6490
700000	-1.74	2.26	50%	0%	50%	3807	-0.06	3.93	1%	0%	99%	2692	6499
1000000	-2.43	1.88	59%	0%	41%	3646	-0.49	2.67	10%	0%	90%	2857	6503
3075000	-4.97	0.27	99%	0%	1%	4289	-3.16	3.43	37%	0%	63%	2214	6503
2950000	-4.95	0.60	95%	0%	5%	4244	-0.72	3.72	10%	0%	90%	2260	6504
975000	-1.76	2.30	46%	0%	54%	3663	-0.46	2.86	9%	0%	91%	2865	6528

Table 23 shows the performance of individual policies. For both the random and constant policies, the top performer is the same as the overall top performer of 1.95M episodes. However,

the top reactive performer is different. All top three reactive performers are different than the random and constant polices.

Reviewing the table, both the merge vehicles have learned a policy of full deceleration all of the time. As discussed previously throughout this work, this is a poor policy since the best performance requires a mix of deceleration and acceleration. The merge reactive policy has learned to rely upon the traffic reactive policy to find the best collisions which comes at a detriment to the constant and random policies. Overall, this mismatch results in poor performance as the standard test results show in Section B.5.

Table 23: Top three performers for each of the individual policies for performance testing.

		Merge Rear						**Merge Front**					
Policy	**Episode**	**Dec. Avg.**	**Acc. Avg.**	**Dec. %**	**Mtn. %**	**Acc. %**	**Coll.**	**Dec. Avg.**	**Acc. Avg.**	**Dec. %**	**Mtn. %**	**Acc. %**	**Coll.**
Random	1950000	-4.11	0.78	84%	0%	16%	956	-2.60	1.70	57%	0%	43%	997
Random	1925000	-4.34	0.47	89%	0%	11%	749	-3.47	1.27	68%	0%	32%	1266
Random	675000	-3.01	2.17	55%	0%	45%	930	-3.22	2.24	54%	0%	46%	1110
Constant	1950000	-4.11	0.78	84%	0%	16%	975	-2.61	1.69	57%	0%	43%	1032
Constant	675000	-3.01	2.18	55%	0%	45%	951	-3.22	2.24	54%	0%	46%	1101
Constant	1925000	-4.35	0.47	89%	0%	11%	765	-3.47	1.27	69%	0%	31%	1323
Reactive	4550000	-5.00	0.00	100%	0%	0%	417	-5.00	0.00	100%	0%	0%	1350
Reactive	3075000	-5.00	0.00	100%	0%	0%	513	-5.00	0.00	100%	0%	0%	1290
Reactive	4475000	-5.00	0.00	100%	0%	0%	438	-5.00	0.00	100%	0%	0%	1374

Figure 132 shows a multi-dimensional plot of the merge rear (Ego) learned performance parameters over the course of 5.175M episodes of training. Snapshots are taken at 25K episode intervals. The plot shows the overall training performance staying relatively flat over the course of training.

The DRL network learns to split its actions between acceleration and deceleration within the leftmost third of the plot, but the collisions become greater as the episodes increase and the

DRL network shifts its learning pattern to full deceleration all the time. In general, the policy most often learned is to fully decelerate all the time. Overall, the collisions seem to be lowest on average with the full deceleration policy. However, based on fundamental study, like Proposition III, and explanation throughout this dissertation, it is well understood that there must be balance between acceleration and deceleration action selection to achieve a low collision count.

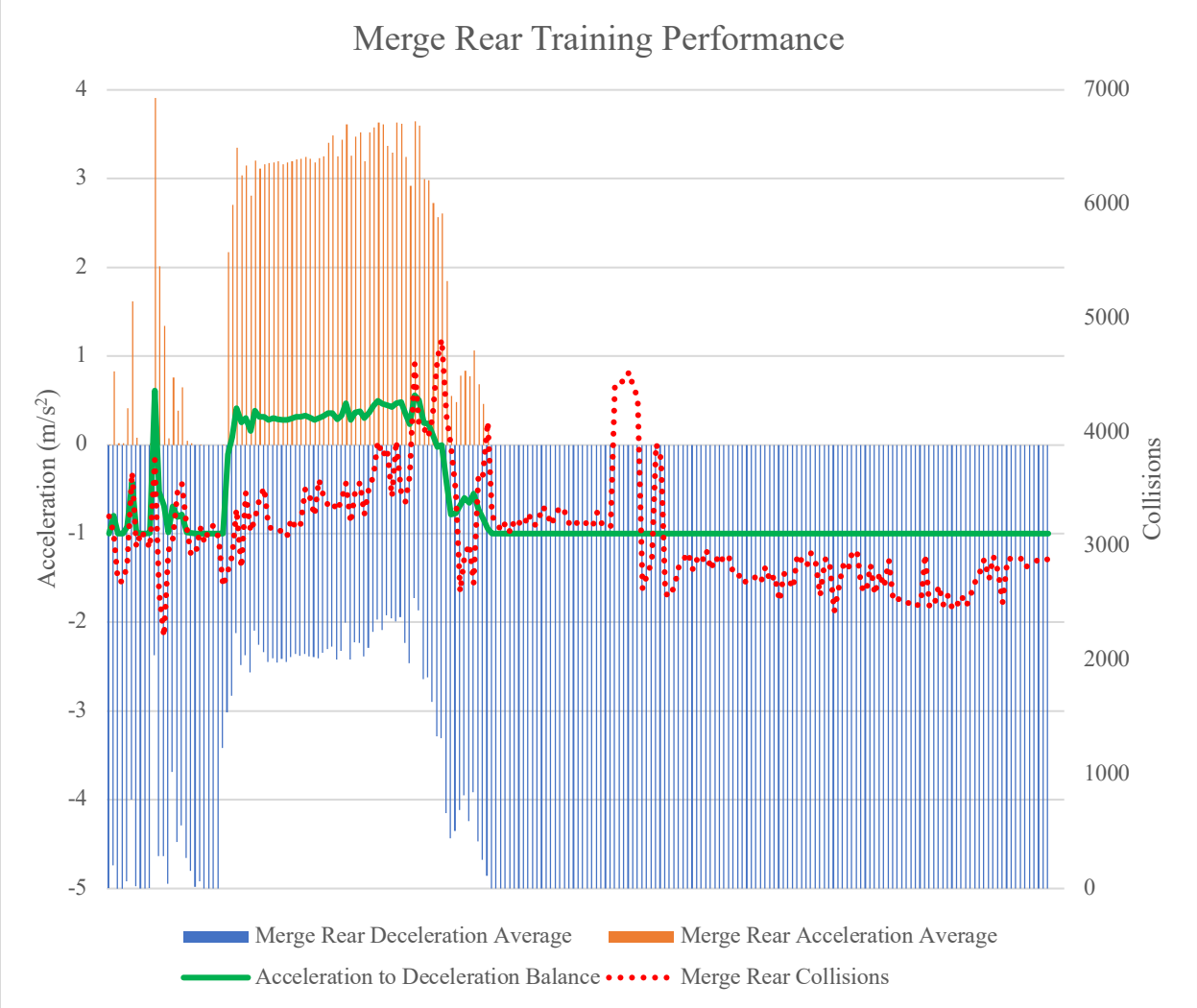


Figure 132: Multi-dimension plot for the Ego (merge rear) vehicle sorted by training episode order. Snapshots taken at 25K intervals over the entire 5.175M episodes of training. Orange bars indicate average acceleration, blue average deceleration per test interval. Red line indicates number of collisions for the merge rear vehicle.

B.5 Standard Test Results

Standard tables display the testing performance, just like previous simulation variants. However, the overall results are poor. Figure 133 shows the details. The test results are much worse than the ideal results, as Figure 29 shows. The merge vehicle has not learned to avoid collisions with the traffic vehicle, even in relatively simple cases like a 100 m goal position.

Traffic_Reaction (All)

Merge Collisions \->	Initial Gap																																				
Initial Gap ->	5										5 All	5 R	5 F	15										15 All	15 R	15 F	25									25 All	25 R	25 F	Total All
Goal Position ->	10	20	30	40	50	60	70	80	90	100				10	20	30	40	50	60	70	80	90	100				10	20	30	40	50	60	70	80	90	100			
Merge to Traffic initial Position Delta -20																											0%	11%	67%	67%	33%	33%	0%	11%	0%	22%	24%	24%	0%	24%
-15														100%	100%	100%	100%	100%	100%	100%	100%	83%	83%	97%	44%	53%	0%	0%	0%	44%	100%	11%	22%	11%	0%	56%	24%	24%	0%	57%
-10														0%	0%	0%	0%	0%	0%	89%	100%	67%	33%	29%	29%	0%	0%	0%	0%	0%	0%	67%	67%	33%	33%	33%	23%	23%	0%	26%
-9	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	23%	77%	0%	0%	0%	0%	0%	33%	78%	44%	56%	67%	28%	28%	0%	0%	0%	0%	0%	0%	100%	67%	22%	33%	67%	29%	29%	0%	51%
-8	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	25%	75%	0%	0%	0%	0%	0%	33%	33%	44%	33%	33%	18%	18%	0%	0%	0%	0%	0%	0%	67%	100%	33%	22%	67%	29%	29%	0%	47%
-7	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	33%	67%	0%	0%	0%	0%	0%	33%	33%	33%	44%	44%	19%	19%	0%	0%	0%	0%	0%	0%	33%	89%	67%	67%	44%	30%	30%	0%	48%
-6	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	31%	69%	0%	0%	0%	0%	0%	33%	33%	11%	33%	33%	14%	14%	0%	0%	0%	0%	0%	0%	33%	78%	67%	56%	67%	30%	30%	0%	46%
-5	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	93%	30%	63%	0%	0%	0%	0%	0%	33%	33%	0%	0%	33%	10%	10%	0%	0%	0%	0%	0%	0%	67%	67%	56%	67%	26%	26%	0%	41%
-4	67%	0%	18%	30%	33%	33%	44%	33%	17%	0%	28%	9%	19%	67%	0%	0%	0%	0%	33%	33%	0%	0%	33%	17%	17%	0%	67%	0%	0%	0%	0%	33%	56%	44%	56%	26%	26%	0%	23%
-3	100%	67%	0%	0%	0%	0%	0%	0%	0%	0%	17%	17%	0%	100%	67%	0%	0%	0%	33%	33%	0%	0%	33%	27%	27%	0%	100%	67%	0%	0%	0%	33%	44%	44%	44%	33%	33%	0%	26%
-2	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	20%	20%	0%	100%	100%	11%	0%	0%	0%	33%	0%	0%	0%	24%	24%	0%	100%	100%	22%	0%	0%	0%	33%	78%	56%	33%	42%	42%	0%	29%
-1	100%	100%	100%	0%	0%	0%	0%	0%	0%	33%	33%	33%	0%	100%	100%	100%	0%	0%	0%	33%	0%	0%	0%	33%	33%	0%	100%	100%	78%	0%	0%	0%	33%	67%	44%	33%	46%	46%	0%	37%
0	100%	100%	100%	11%	0%	0%	0%	0%	0%	33%	34%	34%	0%	100%	100%	100%	0%	0%	0%	33%	33%	0%	11%	38%	38%	0%	100%	100%	100%	22%	0%	0%	33%	33%	33%	33%	46%	46%	0%	39%
1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	15%	85%	100%	100%	100%	100%	100%	100%	100%	100%	0%	33%	88%	3%	85%	100%	100%	100%	100%	75%	33%	33%	0%	50%	50%	60%	10%	51%	84%
2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	17%	83%	100%	100%	100%	100%	100%	100%	100%	100%	67%	50%	94%	8%	86%	100%	100%	100%	100%	100%	43%	33%	0%	50%	50%	65%	10%	56%	88%
3	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	22%	78%	100%	100%	100%	100%	100%	100%	100%	100%	50%	50%	92%	11%	81%	100%	100%	100%	100%	100%	100%	43%	0%	0%	50%	70%	5%	65%	88%
4	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	28%	72%	100%	100%	100%	100%	100%	100%	100%	100%	83%	33%	93%	13%	81%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	71%	0%	71%	89%
5	100%	100%	100%	100%	88%	60%	100%	100%	100%	100%	96%	30%	67%	100%	100%	100%	100%	100%	100%	100%	100%	83%	100%	99%	26%	73%	100%	100%	100%	100%	100%	100%	60%	0%	0%	0%	77%	10%	67%	91%
6	100%	67%	0%	0%	31%	38%	33%	33%	56%	0%	37%	19%	18%	0%	0%	10%	25%	27%	67%	100%	100%	100%	100%	48%	30%	18%	0%	0%	69%	75%	75%	100%	33%	0%	0%	0%	42%	24%	18%	42%
7	100%	100%	67%	0%	0%	0%	0%	0%	0%	0%	27%	27%	0%	0%	0%	0%	0%	0%	25%	67%	100%	67%	100%	33%	30%	3%	0%	0%	0%	67%	67%	67%	25%	0%	33%	0%	27%	23%	3%	29%
8	100%	100%	100%	11%	0%	0%	0%	0%	0%	0%	31%	31%	0%	0%	0%	0%	0%	0%	0%	56%	67%	100%	22%	22%	0%	0%	0%	0%	33%	67%	67%	22%	0%	33%	33%	26%	26%	0%	26%	
9	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	40%	40%	0%	0%	0%	0%	0%	0%	0%	0%	44%	100%	89%	23%	23%	0%	0%	0%	0%	0%	56%	67%	11%	33%	33%	33%	23%	23%	0%	29%
10														0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	20%	20%	0%	0%	0%	0%	0%	67%	67%	33%	33%	0%	20%	20%	0%	20%
15														0%	0%	0%	0%	0%	0%	0%	33%	33%	0%	7%	7%	0%	0%	0%	0%	0%	0%	0%	44%	100%	33%	0%	18%	18%	0%	12%
20																											0%	0%	0%	0%	0%	0%	0%	78%	67%	14%	14%	0%	14%
Grand Total	98%	91%	77%	60%	52%	50%	51%	53%	57%	43%	64%	25%	39%	47%	41%	35%	26%	26%	36%	47%	46%	44%	49%	39%	22%	18%	40%	36%	31%	27%	28%	38%	43%	33%	35%	38%	35%	25%	10%	45%

Figure 133: Four-vehicle standard test table results for all traffic action policies. The table is similar to other standard result tables but adds an initial gap. The initial gaps are 5, 15, and 25 meters from left to right, correlating to a one, three, and five vehicle length gap.

The 25 m gap (right-most table in Figure 133) shows a pattern formed down and to the right of a group of collisions. The collisions occur as the initial position delta and goal position values increase. Reasoning through the results indicates that the merge vehicle has learned a strategy of only decelerating, regardless of the states of the traffic vehicle.

For example, an expected behavior is for vehicles in front to try to stay in front to avoid a collision. Staying in front would require acceleration or maintaining speed, but not deceleration. This behavior shows in the fundamental study and better-performing simulations' results. Section 6.4 gives additional explanation of this occurrence. The merge vehicles have not learned this collision avoiding behavior and it contributes to the poor test results.

B.6 Discussion

Throughout this research, a major theme is finding an appropriate model to represent merging. The model needs to be simple enough to represent the essence of the maneuver but sophisticated enough to capture all the important aspects of it. This also applies to the set of states. The cost of the simplification is data richness. Many simplifications and parameter replacements happened throughout this research. It sometimes resulted in networks that could not learn or learn well. In other instances, it resulted in a tipping point where learning began to occur, or performance increased significantly.

As an example, a change of state representations on the three-vehicle scene resulted in a significant performance increase. However, the change added complexity to the state set. Before the change, there was a single TIV variable for the time to goal. After the change, the TIV variable split into two variables: closing gap and closing speed. The change resulted in behavior that appeared to be much more natural when the merge vehicle worked to position itself to avoid a collision during the merge. It also resulted in a performance increase in standard testing.

The incremental simulation step described in this appendix show intermediate behavior that would be missed if I had chosen to start with a full-scene simulation where many vehicles are present. The meticulous and detailed stepwise study of the merge problem created a deep and broad fundamental understanding of the problem and the factors that influence it. It enabled a better understanding of the best states to consider and a progressive building of standard test tables with subtle visual differences. It uncovered a persistent pattern of acceleration and deceleration in better-performing networks. All these learnings contribute new knowledge to understanding vehicle merge behavior and applying learning algorithms to it.

References

[1] C. Dong, J. M. Dolan, and B. Litkouhi, "Interactive Ramp Merging Planning in Autonomous Driving: Multi-Merging Leading PGM (MML-PGM)," 2017.

[2] Y. Lyu, C. Dong, and J. M. Dolan, "FG-GMM-based Interactive Behavior Estimation for Autonomous Driving Vehicles in Ramp Merging Control," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 1250–1255.

[3] L. Schester and L. E. Ortiz, *Longitudinal Position Control for Highway On-Ramp Merging: A Multi-Agent Approach to Automated Driving*. New Zealand: IEEE, 2019.

[4] L. Schester and L. E. Ortiz, "Automated driving highway traffic merging using deep multi-agent reinforcement learning in continuous state-action spaces," in *IEEE Intelligent Vehicles Symposium, Proceedings*, Jul. 2021, pp. 280–287, doi: 10.1109/IV48863.2021.9575676.

[5] Society of Automotive Engineering International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles J3016_202104," J3016_202104, 2021. doi: https://doi.org/10.4271/J3016_202104.

[6] National Highway Traffic Safety Administration, "Overview of Motor Vehicle Crashes in 2019," no. December, pp. 1–14, 2020, [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813060.

[7] World Health Organization, "Global status report on road safety 2018," Geneva, 2018. doi: CC BYNC- SA 3.0 IGO.

[8] General Motors, "Zero Congestion with Self-driving Vehicles: General Motors," 2021. https://www.gm.com/commitments/path-to-autonomous (accessed Jan. 11, 2021).

[9] V. W. Inman, S. Jackson, B. H. Philips, and others, "Cooperative Adaptive Cruise Control Human Factors Study: Experiment 1-Workload, Distraction, Arousal, and Trust," 2016.

[10] AASHTO, *A Policy on Geometric Design of Highways and Streets*. American Association of State Highway and Transportation Officials, 2001.

[11] M. F. F. Achmad Ali Fikri, Syamsul Arifin, "The Manual on Uniform Traffic Control Devices (MUTCD)," 2022. [Online]. Available: https://mutcd.fhwa.dot.gov.

[12] Insurance Institute for Highway Safety (IIHS), "Maximum posted speed limits." https://www.iihs.org/topics/speed/speed-limit-laws.

[13] SAE International, “Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles,” J3018_202012, 2020. doi: https://doi.org/10.4271/J3018_202012.

[14] Cruise, “One Million Driverless Miles,” *getcruise.com*, 2023. https://getcruise.com/one-million-driverless-miles/ (accessed Feb. 22, 2023).

[15] T. Victor, K. Kusano, T. Gode, R. Chen, and M. Schwall, “Safety Performance of the Waymo Rider-Only Automated Driving System at One Million Miles,” pp. 1–30.

[16] P. J. Koopman, *How Safe Is Safe Enough?: Measuring and Predicting Autonomous Vehicle Safety*. Independently published, 2022.

[17] S. Feng *et al.*, “Dense reinforcement learning for safety validation of autonomous vehicles,” *Nature*, vol. 615, no. 7953, pp. 620–627, 2023, doi: 10.1038/s41586-023-05732-2.

[18] Tesla, “Autopilot and Full Self-Driving Capability | Tesla,” *Tesla*, 2020. https://www.tesla.com/support/autopilot (accessed Nov. 14, 2021).

[19] Tesla, “Introducing Navigate on Autopilot,” *Online*, 2018. https://www.tesla.com/nl_NL/blog/introducing-navigate-autopilot (accessed Oct. 26, 2018).

[20] D. Lee, “Apple self-driving car in minor crash,” Sep. 2018. https://www.bbc.com/news/technology-45380373.

[21] YouTube, “Waymo Self Driving Taxi Goes Rogue: Blocks Traffic, Evades ...,” 2021. https://www.youtube.com/watch?v=zdKCQKBvH-A (accessed May 17, 2021).

[22] E. Yi Lu Murphey, Ilya Kolmanovsky, and Paul Watta, *AI-enabled Technologies for Autonomous and Connected Vehicles*. Springer, 2022.

[23] S. Singh, “Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey,” Washington, DC, 2015. doi: Traffic Safety Facts Crash•Stats. Report No. DOT HS 812 115.

[24] K. Naughton, “Human drivers are bumping into driverless cars and exposing a key flaw,” *Automotive News*, Dec. 18, 2015.

[25] B. Schoettle and M. Sivak, “A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles,” The University of Michigan Transportation Research Institute, UMTRI-2015-34, 2015.

[26] M. Sivak and B. Schoettle, “Road Safety With Self-Driving Vehicles: General Limitations and road sharing with conventional vehicles,” *Transp. Res. Inst.*, no. January, p. 11, 2015, [Online]. Available: https://deepblue.lib.umich.edu/handle/2027.42/111735.

[27] National Highway Traffic Safety Administration, “Summary Report: Standing General Order on Crash Reporting for Automated Driving Systems,” no. June, 2022, [Online]. Available: https://www.nhtsa.gov/sites/nhtsa.gov/files/2022-06/ADS-SGO-Report-June-2022.pdf.

[28] J. L. Campbell *et al.*, “Human factors design principles for level 2 and level 3 automated driving concepts,” *Highw. Traffic Saf. Adm. Natl. Dep. Transp.*, no. August, p. 122, 2018, [Online]. Available: www.ntis.gov.

[29] C. M. and N. E. Boudette, “Inside Tesla as Elon Musk Pushed an Unflinching Vision for Self-Driving Cars,” *NY Times*, Dec. 07, 2021. https://www.nytimes.com/2021/12/06/technology/tesla-autopilot-elon-musk.html (accessed Dec. 13, 2021).

[30] A. Nguyen, J. Yosinski, and J. Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2015, vol. 07-12-June, pp. 427–436, doi: 10.1109/CVPR.2015.7298640.

[31] Tencent Keen Security Lab, “Tencent Keen Security Lab: Experimental Security Research of Tesla Autopilot | Keen Security Lab Blog,” 2019. [Online]. Available: https://keenlab.tencent.com/en/2019/03/29/Tencent-Keen-Security-Lab-Experimental-Security-Research-of-Tesla-Autopilot/.

[32] K. Eykholt *et al.*, “Robust Physical-World Attacks on Deep Learning Models,” Jul. 2017, [Online]. Available: http://arxiv.org/abs/1707.08945.

[33] S. Romero, “Wielding Rocks and Knives, Arizonans Attack Self-Driving Cars,” *The New York Times*, 2018. https://www.nytimes.com/2018/12/31/us/waymo-self-driving-cars-arizon (accessed Dec. 31, 2018).

[34] P. Penmetsa, E. K. Adanu, D. Wood, T. Wang, and S. L. Jones, “Perceptions and expectations of autonomous vehicles – A snapshot of vulnerable road user opinion,” *Technol. Forecast. Soc. Change*, vol. 143, no. March, pp. 9–13, 2019, doi: 10.1016/j.techfore.2019.02.010.

[35] NTSB, “Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018,” Washington, DC, 2018. [Online]. Available: https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf.

[36] NTSB, “Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck,” Washington, DC, 2017. [Online]. Available: https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1702.pdf.

[37] N. Figalová *et al.*, “Fatigue and mental underload further pronounced in L3 conditionally automated driving: Results from an EEG experiment on a test track,” *Int. Conf. Intell. User Interfaces, Proc. IUI*, pp. 64–67, 2023, doi: 10.1145/3581754.3584133.

[38] INCOSE, *INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*, 4th ed. Wiley, 2015.

[39] Y. C. Tang, "Towards Learning Multi-Agent Negotiations via Self-Play," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 2427–2435, doi: 10.1109/ICCVW.2019.00297.

[40] D. Kamran, Y. Ren, and M. Lauer, "High-level Decisions from a Safe Maneuver Catalog with Reinforcement Learning for Safe and Cooperative Automated Merging," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, vol. 2021-Septe, pp. 804–811, 2021, doi: 10.1109/ITSC48978.2021.9564912.

[41] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "Safe, multi-agent, reinforcement learning for autonomous driving," *arXiv Prepr. arXiv1610.03295*, 2016, doi: 10.5626/ktcp.2018.24.12.670.

[42] P. Wang and C.-Y. Chan, "Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–6.

[43] D. Chandra, Rohan and Manocha, "GamePlan: Game-Theoretic Multi-Agent Planning with Human Drivers at Intersections, Roundabouts, and Merging," 2022, doi: 10.1109/LRA.2022.3144516.

[44] M. Bahram, A. Lawitzky, J. Friedrichs, M. Aeberhard, and D. Wollherr, "A game-theoretic approach to replanning-aware interactive scene prediction and planning," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3981–3992, Jun. 2016, doi: 10.1109/TVT.2015.2508009.

[45] M. Jain, K. Brown, and A. K. Sadek, "Multi-Fidelity Recursive Behavior Prediction," *arXiv Prepr. arXiv1901.01831*, Dec. 2018, [Online]. Available: http://arxiv.org/abs/1901.01831.

[46] Y. Lyu, W. Luo, and J. M. Dolan, "Probabilistic Safety-Assured Adaptive Merging Control for Autonomous Vehicles," pp. 10764–10770, 2021, doi: 10.1109/icra48506.2021.9561894.

[47] S. Triest, A. Villaflor, and J. M. Dolan, "Learning Highway Ramp Merging Via Reinforcement Learning with Temporally-Extended Actions," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2020, pp. 1595–1600, doi: 10.1109/IV47402.2020.9304841.

[48] C. Dong, J. M. Dolan, and B. Litkouhi, "Smooth behavioral estimation for ramp merging control in autonomous driving," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2018, vol. 2018-June, pp. 1692–1697, doi: 10.1109/IVS.2018.8500576.

[49] J. Wei, J. M. Dolan, and B. Litkouhi, “Autonomous vehicle social behavior for highway entrance ramp management,” in *2013 IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 201–207.

[50] U.S. Department of Transportation Federal Highway Administration, “Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data. [Dataset].” Provided by ITS DataHub through Data.transportation.gov, 2016, doi: http://doi.org/10.21949/1504477.

[51] C. Killing, A. Villaflor, and J. M. Dolan, “Learning to Robustly Negotiate Bi-Directional Lane Usage in High-Conflict Driving Scenarios,” Oct. 2021, pp. 8090–8096, doi: 10.1109/icra48506.2021.9561071.

[52] R. Rajamani, H. S. Tan, B. K. Law, and W. Bin Zhang, “Demonstration of integrated longitudinal and lateral control for the operation of automated vehicles in platoons,” *IEEE Trans. Control Syst. Technol.*, vol. 8, no. 4, pp. 695–708, 2000, doi: 10.1109/87.852914.

[53] D. Bevly *et al.*, “Lane change and merge maneuvers for connected and automated vehicles: A survey,” *IEEE Trans. Intell. Veh.*, vol. 1, no. 1, pp. 105–120, 2016, doi: 10.1109/TIV.2015.2503342.

[54] S. E. Shladover, D. Su, and X. Y. Lu, “Impacts of cooperative adaptive cruise control on freeway traffic flow,” *Transp. Res. Rec.*, vol. 2324, no. Idm, pp. 63–70, 2012, doi: 10.3141/2324-08.

[55] N. Brown and T. Sandholm, “Superhuman AI for multiplayer poker,” *Science*, vol. 365, no. 6456. American Association for the Advancement of Science, pp. 885–890, Aug. 30, 2019, doi: 10.1126/science.aay2400.

[56] S. A. Fernandez, M. A. M. Marinho, M. Vakilzadeh, and A. Vinel, “Highway On-Ramp Merging for Mixed Traffic: Recent Advances and Future Trends,” in *2021 IEEE 29th International Conference on Network Protocols (ICNP)*, Nov. 2021, pp. 1–6, doi: 10.1109/ICNP52444.2021.9651989.

[57] J. Rios-Torres and A. A. Malikopoulos, “A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1066–1077, May 2017, doi: 10.1109/TITS.2016.2600504.

[58] S. Glaser, B. Vanholme, S. Mammar, D. Gruyer, and L. Nouvelière, “Maneuver-based trajectory planning for highly autonomous vehicles on real road with traffic and driver interaction,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 589–606, 2010, doi: 10.1109/TITS.2010.2046037.

[59] R. Shwartz-Ziv and N. Tishby, “Opening the Black Box of Deep Neural Networks via Information,” Mar. 2017, [Online]. Available: http://arxiv.org/abs/1703.00810.

[60] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[61] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. New York, NY, USA: Cambridge University Press, 2008.

[62] C. J. C. H. Watkins, "Learning from delayed rewards," King's College, Cambridge, 1989.

[63] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3–4, pp. 279–292, 1992.

[64] C. Nowakowski *et al.*, "Cooperative adaptive cruise control: Testing drivers' choices of following distances," UCB-ITS-PRR-2011-01, Jan. 2011. doi: 10.1177/154193121005402403.

[65] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *J. Mach. Learn. Res.*, vol. 4, no. Nov, pp. 1039–1069, 2003.

[66] J. Hu, M. P. Wellman, and others, "Multiagent reinforcement learning: theoretical framework and an algorithm.," in *ICML*, 1998, vol. 98, pp. 242–250.

[67] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Machine learning proceedings 1994*, Elsevier, 1994, pp. 157–163.

[68] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," *AAAI/IAAI*, vol. 1998, pp. 746–752, 1998.

[69] G. Weiss, *Multiagent Systems*. Cambridge, Massachusetts: The MIT Press, 2013.

[70] N. A. and R. A. Office of the Federal Register, "49 CFR 571.135 - Standard No. 135; Light vehicle brake systems," 2011. [Online]. Available: https://www.govinfo.gov/app/details/CFR-2011-title49-vol6/CFR-2011-title49-vol6-sec571-135.

[71] K. Fadhloun, H. Rakha, A. Loulizi, and A. Abdelkefi, "Vehicle dynamics model for estimating typical vehicle accelerations," *Transp. Res. Rec.*, vol. 2491, no. 2491, pp. 61–71, 2015, doi: 10.3141/2491-07.

[72] S. A. Balk, S. Jackson, and B. H. Philips, "Cooperative Adaptive Cruise Control Human Factors Study: Experiment 2-Merging Behavior," Dec. 2016. [Online]. Available: http://www.ntis.gov.

[73] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.

[74] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016, doi: 10.1038/nature16961.

[75] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015, doi: 10.1038/nature14539.

[76] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013, doi: 10.1109/TPAMI.2013.50.

[77] T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning,” Sep. 2015, [Online]. Available: http://arxiv.org/abs/1509.02971.

[78] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, 2021, doi: 10.1038/s41586-021-03819-2.

[79] V. Mnih *et al.*, “Asynchronous methods for deep reinforcement learning,” *Int. Conf. Mach. Learn.*, vol. 48, pp. 1928–1937, 2016, [Online]. Available: http://arxiv.org/abs/1301.3781.

[80] M. Jaderberg *et al.*, “Human-level performance in first-person multiplayer games with population-based deep reinforcement learning,” Jul. 2018, doi: 10.1126/science.aau6249.

[81] “Figures for Sutton & Barto Book: Reinforcement Learning: An Introduction.” http://incompleteideas.net/sutton/book/first/figures/figures.html (accessed Nov. 11, 2021).

[82] M. Zhou, “Reinforcement-learning-with-tensorflow,” 2020. https://github.com/MorvanZhou/Reinforcement-learning-with-tensorflow (accessed Dec. 30, 2021).

[83] G. E. Uhlenbeck and L. S. Ornstein, “On the theory of the Brownian motion,” *Phys. Rev.*, vol. 36, no. 5, p. 823, 1930.

[84] P. Wawrzyński and A. K. Tanwani, “Autonomous reinforcement learning with experience replay,” *Neural Networks*, vol. 41, pp. 156–167, 2013, doi: 10.1016/j.neunet.2012.11.007.

[85] A. W. Moore, “Efficient Memory-based Learning for Robot Control,” University of Cambridge, Computer Laboratory, 1990.

[86] OpenAI, “Mountain Car Continuous,” *Gym Documentation*. https://www.gymlibrary.dev/environments/classic_control/mountain_car_continuous/ (accessed Apr. 22, 2023).

[87] J. Hu, B. T.-W. Lin, J. H. Vega, and N. R.-L. Tsiang, “Predictive Models of Driver Deceleration and Acceleration Responses to Lead Vehicle Cutting In and Out,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2677, no. 5, p. 036119812211282, 2022, doi: 10.1177/03611981221128277.

[88] M. J. Moran, H. N. Shapiro, D. D. Boettner, and M. B. Bailey, *Fundamentals of engineering thermodynamics*. John Wiley & Sons, 2010.

[89] K. Leyton-Brown and Y. Shoham, *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*, 1st ed. Morgan and Claypool Publishers, 2008.

[90] D. Fudenberg and J. Tirole, *Game Theory*. The MIT Press, 1991.