

# UC Irvine

## UC Irvine Electronic Theses and Dissertations

### Title

Paradigms of Identifying and Quantifying Uncertainty and Information in Constructing a Cognition-Modeling Framework of Human-Machine Transportation Systems

### Permalink

<https://escholarship.org/uc/item/2zn478wb>

### Author

Yu, Jiangbo

### Publication Date

2018

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,  
IRVINE

**Paradigms of Identifying and Quantifying Uncertainty and Information  
in Constructing a Cognition-Modeling Framework of Human-Machine  
Transportation Systems**

DISSERTATION

submitted in partial satisfaction of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Jiangbo Gabriel Yu

Dissertation Committee:  
Professor R. Jayakrishnan, Chair  
Professor Jean-Daniel Saphores  
Associate Professor Prof. Wenlong Jin

2018

Portion of Chapter 2 © 2018 IEEE  
Portion of Chapter 5 © 2017 Transportation Research Board  
Rest of the dissertation © 2018 Jiangbo Gabriel Yu

# **DEDICATION**

To

my parents and mentors

in recognition of their guidance

# CONTENTS

<b>LIST OF FIGURES .....</b>	<b>vi</b>
<b>LIST OF TABLES .....</b>	<b>ix</b>
<b>ACKNOWLEDGMENTS .....</b>	<b>x</b>
<b>CURRICULUM VITAE .....</b>	<b>xi</b>
<b>ABSTRACT OF THE DISSERTATION .....</b>	<b>xii</b>
<b>CHAPTER 1 INTRODUCTION .....</b>	<b>1</b>
<b>CHAPTER 2 A COGNITION-BASED MODELING AND ANALYTICAL FRAMEWORK.....</b>	<b>5</b>
Motivations .....	5
Literature Review.....	8
Framework .....	13
Physical Interaction.....	15
Space of Observables.....	15
Cognition .....	16
Analytical Examples .....	20
Human Driving Behavior .....	20
Multi-stakeholder planning decision .....	21
Numerical Feasibility Study – Human-ACV Mixed Flows.....	22
Background.....	23
Defining Agent Class.....	23
Defining PISOO Class.....	27
Agent Instantiation and Simulation .....	28
Density, Market Penetration, and Risk Preference Sensitivity Test.....	30
Highlighted Features.....	34
Feasibility and Advantage.....	36
Computational Efficiency.....	37
Conclusion.....	39
<b>CHAPTER 3 QUANTIFYING INFORMATION AS CHANGE OF PERCEIVED UNCERTAINTY .....</b>	<b>41</b>
Background and Literature Review .....	41
Methodology .....	44
Numerical Example .....	48
Conclusion.....	58
<b>CHAPTER 4 ELASTIC SURPRISE THEORY FOR DECISION UNDER RISK.....</b>	<b>60</b>

Introduction.....	60
Literature Review.....	62
More on Scale and Convexity Paradox in Existing Methods .....	64
Elastic Surprise (ES).....	65
Logarithmic ES Function and Information Entropy .....	68
Properties and Cognitive Implication.....	72
Limit at $0^+$ .....	72
Rationality, Stochastic Dominance, and Trade-off Consistency .....	74
EUT and Its Revision .....	75
Reference Dependency and Relationship with CPT .....	77
Three Perspectives .....	81
ES in Mean-Variance (MV) Method.....	82
Empirical Study on Route Choice under Risk.....	83
Conclusion.....	87
<b>CHAPTER 5 MULTICLASS, MULTICRITERIA DYNAMIC TRAFFIC ASSIGNMENT WITH PATH-DEPENDENT LINK COST AND ENTROPY-BASED RISK PREFERENCE .....</b>	<b>89</b>
Introduction.....	89
Literature Review.....	91
Important Concepts.....	93
Path-dependent Link Cost .....	93
Uncertainty, Reliability, Risk, and Variability .....	94
Entropy-based Measure of Perceived Uncertainty.....	95
Single-Class Single-Criterion DUE Formulation.....	99
Multi-Class Multi-Criteria Extension With Path-Dependent Link Cost.....	102
Stochastic Gradient Project Based Solution .....	103
Case Study.....	105
Results Discussion.....	110
Conclusion and Future Direction .....	114
<b>CHAPTER 6 CONCLUSION .....</b>	<b>116</b>
<b>APPENDICES .....</b>	<b>121</b>
Appendix A.....	121
Appendix B.....	122
Appendix C.....	122

<b>REFERENCES</b> .....	<b>125</b>
-------------------------	------------

## LIST OF FIGURES

<b>Figure 2.1</b> The increasingly denser stripping deluded drivers that they are driving faster and, therefore, tend to reduce speed before a sharp left turn. ....	6
<b>Figure 2.2</b> The logical framework of CognAgent .....	14
<b>Figure 2.3</b> Conceptual Graph Demonstrating the Data Management Role of the “Space of Observable” Module in Human Driver-Only Traffic .....	16
<b>Figure 2.4</b> Example of analysing driver behaviour under uncertainty based on CognAgent .....	20
<b>Figure 2.5</b> Example of analysing a stakeholder meeting on infrastructure investment based on CognAgent ..	22
<b>Figure 2.6</b> Coding “space of observables” as a raster-like database. ....	23
<b>Figure 2.7</b> Example of an agent “inquiring” the instantiation of PISOO at time $t$ . ....	28
<b>Figure 2.8</b> Vehicle locations and speeds at $t = 0$ .....	29
<b>Figure 2.9</b> Y-axis (bottom to top): 0, 0.3, 0.5, 0.7, 1.0 MP Rate (red triangles are human drivers; blue dots are ACV). X-axis (left to right): $t = 300s$ . The appendix shows the complete figure. ....	30
<b>Figure 2.10</b> $t=0,120s,300s$ (y-axis are set to 20000ft to all the three plots for comparability) .....	30
<b>Figure 2.11</b> The trajectories (upper) and speed profile (lower) while 5% MP and 5-min simulation horizon; number of vehicle is set as 100, 200, and 300, respectively. ....	31
<b>Figure 2.12</b> 50% MP and 5-min simulation horizon; veh = 100, 200, 300 .....	31
<b>Figure 2.13</b> 100% MP and 5-min simulation horizon; veh = 100, 200, 300.....	32
<b>Figure 2.14</b> Hypothetical $vc$ region in $MP-\rho c$ space and hypothetical $v-\rho c$ relationship based on the experimental results. ....	32
<b>Figure 2.15</b> Hypothetical $MP-v-\rho$ relationship based on the experimental results.....	33
<b>Figure 2.16</b> Market penetration and mobility in different market penetration rates (300 sec horizon) with different dominant risk preference. ....	34
<b>Figure 2.17</b> Vehicle speed in the DSRC accident warning scenario (100% MP). $t = 30s, 60s, 120s, 300s$ .....	38
<b>Figure 2.18</b> Vehicle location in the DSRC accident warning scenario (100% MP). $t = 30s, 60s, 120s, 300s$ ...	39
<b>Figure 2.19</b> Computational effort comparison among different modeling needs and numbers of agents .....	39
<b>Figure 3.1</b> Destination Groups in Subarea Analysis (Left) and the study network (right). The blue nodes on the left indicates the candidate location of the proposed DMB. ....	49
<b>Figure 3.2</b> Dynamic Flow Rate Sample (veh/5-min) from 6:30am – 8:30am, Monday March 9, 2017 (Red) and Thursday March 23, 2017 (Blue).....	50
<b>Figure 3.3</b> The histograms show Santa Monica Direction (Drive-Alone (DA) before and after the information provision. The red line is the fitted kernel distribution based on experience and newly provided information. ....	52



<b>Figure 3.4</b> The histograms show 2-Passeger High Occupancy Vehicles (HOV2) before and after the information provision. The red line is the fitted kernel distribution based on experience and newly provided information.....	52
<b>Figure 3.5</b> The histograms show Median-Duty Truck (MT). The red line is the fitted kernel distribution based on experience and newly provided information.....	53
<b>Figure 3.6</b> Comparing Information Entropy (unit: bits) without (left) and with (right) providing Information through Dynamic Massage Board (5-minute Interval). The five rows are for the traffic to the five destinations (D1-D5). The four selected user classes are: DA (Green), HOV2 (Orange), HOV3 (Blue), and MT (Red).....	53
<b>Figure 4.1</b> The left graph shows the logarithm of $1/p$ where $p = (0,1]$ with base 2, e, and 10. The right graph shows different natural logarithm of $1/p$ by varying the scaler .....	69
<b>Figure 4.2</b> The left shows the effect of different combinations of $k_1$ and $k_2$ ( $k_1 \cdot k_2 > 0$ ) on $U$ ; the middle shows the effect of different combinations of $k_1 > 0$ and $k_2 < 0$ on $U$ ; The right shows the effect different combinations of $k_1 < 0$ and $k_2 > 0$ on $U$ .....	70
<b>Figure 4.3</b> The left shows $p \cdot S(p)$ . The right shows entropy $H$ over a bi-outcome lottery.....	71
<b>Figure 4.4</b> Relationship of $U$ and $Eu$ on the two-outcome lottery with a range of $\beta$ that captures risk-preference with the potential trade-off inconsistency.....	74
<b>Figure 4.5</b> In both cases, the choice maker behaves risk-averse regarding the utility of expected pay-off since $U < uEx, 1$ . The left behaves risk-prone and the right behaves risk-averse, regarding the expected utility since $U > Eu(x), 1$ .....	76
<b>Figure 4.6</b> Examples of changing risk preference on $L$ .....	77
<b>Figure 4.7</b> Comparing the effect of reference-dependence using proposed method with that in CPT. X-axis is the expected payoff. ....	78
<b>Figure 4.8</b> Left: CPT weighting function; Right: equivalent ES function for $\delta$ .....	79
<b>Figure 4.9</b> Left: $\pi_2 +$ (z-axis) in a three-outcome decision, where $X_1 \leq 0 \leq X_2 \leq X_3$ ; Right: the contour of $\pi_2 +$ .....	80
<b>Figure 4.10</b> $S_2$ over $p_2$ and $p_3$ derived from a 3-outcome decision based on the Cumulative Prospect Theory .....	81
<b>Figure 4.11</b> The Revised EUT (REUT) and CPT can be derived from incorporating ES differently.....	81
<b>Figure 4.12</b> (a) The implicitly embedded ES function. (b) $K$ versus $\beta K_2$ . (c) $p$ versus $\beta K_2(1 - p)$ . (d) The $TT$ versus GC given different scalars $\beta$ .....	83
<b>Figure 4.13</b> Effect of different splitting factor $\theta$ on the shape of misperception $\pi$ , ES Function, and the Compound Weighting Function $\pi *$ .....	84
<b>Figure 4.14</b> Effect of $k$ on $\pi *$ given different $\delta$ -.....	85
<b>Figure 4.15</b> Left: Estimated ES Function. Middle: Weighting function for misperception and compound effect of both misperception and ES. Right: new value function comparing to the original one (without considering ES) .....	87
<b>Figure 5.1</b> Example of path-dependent link-additive cost.....	94

<b>Figure 5.2</b> Example of a traveler’s category-based perceived travel time distribution .....	96
<b>Figure 5.3</b> Travel Time (x-axis) vs. General Cost with Risk Preference (y-axis). The left is risk-averse and the right risk-prone. ....	97
<b>Figure 5.4</b> Influence of preferred arrival time and travel time distribution on perceived uncertainty. ....	98
<b>Figure 5.5</b> The cost of the trajectory of an infinitely small portion of traffic load passing through $dxdt$ is $c(\rho x, t)dxdt$ .....	100
<b>Figure 5.6</b> Relative location of the SR-91 HOT Facility to City of Los Angeles and City of Irvine. (Source: MapQuest.com) .....	106
<b>Figure 5.7</b> The Model Network. The upper left shows the entire model network; the upper right shows the west end of the SR91 HOT facility; the lower left shows the east end of the SR91 HOT facility; The lower right shows the interchange area of the SR91 and N Weir Canyon Road. ....	107
<b>Figure 5.8</b> Approximating the (partial) second-order derivative of a traveler’s general cost along the trajectory. ....	108
<b>Figure 5.9</b> Flow-chart of the GP solution adaptive to the case study.....	109
<b>Figure 5.10</b> Convergence test for both the <i>RMEAN</i> and the <i>RSTDEV</i> when only considers the first order derivative of the objective function. ....	111
<b>Figure 5.11</b> Convergence test for both the <i>RMEAN</i> and the <i>RSTDEV</i> when consider the first-order and the second-order derivative of the objective function.....	111
<b>Figure 5.12</b> Bi-criterion and tri-criterion gap over iterations for user 10077, 13782, and 23192 .....	112

## LIST OF TABLES

<b>Table 2.1</b> Sensing capability assumed in this paper on human and ACVs.....	24
<b>Table 2.2</b> ACV Operation Rules (DSRC will be incorporated as an extension later).....	26
<b>Table 2.3</b> Average travel distance (mobility) in 30, 60, 120, 300sec with different risk-preference ( $\alpha$ ) for human agents.....	33
<b>Table 2.4</b> System Performance Comparison Given Different Vehicle Type Composition over Type A, Type B, and Type C Autonomous Vehicles (Random Seed 3) during the simulation 0 to 30,000 steps (0.01sec/step) and in the road stretch from 0 to 35,000 ft. x-axis is the distance and y-axis is simulation step in the plots of vehicle trajectories.....	35
<b>Table 2.5</b> The trajectories of the 400 simulated vehicles with and without sever dyslexia when he/she sense the message board's speed recommendation (81 ft/sec (55 mph)) at the beginning of the simulation. In addition, 5% are randomly selected to miss the information.....	36
<b>Table 3.1</b> Trip destination split percentage at the study location.....	49
<b>Table 3.2</b> Change of Information Entropy (Uncertainty) for 15-min time interval and each user class classified by vehicle type and destination (D1-D5).....	54
<b>Table 3.3</b> Summary of effective information provision (Kilobytes) for different DMB installation locations and different strategies.....	56
<b>Table 4.1</b> Selected surprise functions and their corresponding linear transformation to utility.....	72
<b>Table 4.2</b> Risk preference regarding expected payoff and expected utility in REUT.....	76
<b>Table 4.3</b> Mean and median travel time and event probabilities in the survey (in minutes).....	85
<b>Table 4.4</b> Mean travel time and standard deviation (sample) from the survey (in minutes).....	86
<b>Table 5.1</b> Clarifying definitions of variability, uncertainty, reliability, and risk.....	95
<b>Table 5.2</b> Relationship of $(rstk)_i$ , $(rstk)_i$ , $(rstk)_j$ , and $(rstk)_j$ to $(xa, t)$ for deciding $\theta_i, jxa, t$ according to (5.10).....	104
<b>Table 5.3</b> Sample trip table used in the algorithm.....	110
<b>Table 5.4</b> User attributes for 10077, 13782, and 23192.....	112
<b>Table 5.5</b> Path switch over the iterations for user 10077, 13782, and 23192.....	112

## **ACKNOWLEDGMENTS**

I would like to express the deepest appreciation to my committee chair, Prof. Jayakrishnan, who provides great guidance in my academic growth and financial support. He continually keeps an open mind and a spirit of adventure regarding research and scholarship, and excitement regarding teaching and mentoring. He also unfolded substance of a genius. Without his persistent guidance, this dissertation would not have been possible.

I would like to thank my committee member, Prof. Jin for his teachings and encouragement through his courses, research papers, and personal interaction. Prof. Jin is a leading researcher in traffic flow theory and this dissertation, primarily Chapter 2 and Chapter 5, was influenced in various ways by his scholarly approach.

Prof. Robin Keller introduced me to the field of decision science. She was always readily available to answer questions and willing to discuss some of the preliminary ideas that eventually became Chapter 4.

During my graduate study, I also actively participated in multiple positions in both the public and private domains, which contributed significantly to the practicality of this dissertation. I appreciate the help from my colleagues, especially Dr. Fatemeh Ranaiefar, Brandon Haydu, and Jeff Pierson, and Michael Kao, who I learned a lot from through personal and professional interactions.

This dissertation was funded by the dissertation grant from the University of California Center on Economic Competitiveness in Transportation (UC CONNECT) and the fellowship from the Pacific-Southwest Region 9 UTC. During the phase of preparing for this dissertation, I also received scholarships from Intelligent Transportation Society of California and the Railway Association of Southern California. I appreciate the recognition and support from these organizations.

## CURRICULUM VITAE

2007-11	B.S. in Transportation Systems Engineering, Beijing Institute of Technology
2011-12	M.S. in Civil Engineering, University of Southern California
2012-13	Assistant Engineer, Los Angeles Department of Transportation
2013-14	Research Assistant, Southern California Association of Government
2013-18	Research Assistant, Institute of Transportation Studies
2014-16	Teaching Assistant, University of California, Irvine
2015-16	Research Engineer, California Department of Transportation
2017-18	Transportation Engineer, Fehr & Peers Transportation Consultants
2013-18	Ph.D. in Transportation Systems Engineering, University of California, Irvine
2018	Senior Systems Modeler, Cambridge Systematics, Inc.

### FIELD OF STUDY

Cognition and Behaviors of Humans and Machines in Transportation Systems, Dynamic Traffic Assignment, Transportation Investment

### PUBLICATIONS

Yu, J. G., & Jayakrishnan, R. (2018). A Cognitive Framework for Unifying Human and Artificial Intelligence in Transportation Systems Modeling. In *Intelligent Transportation Systems, 2018. Proceedings. 2018 IEEE*. IEEE.

Yu, J. G., & Jayakrishnan, R. (2017). Multiclass, Multicriteria Dynamic Traffic Assignment with Path-Dependent Link Cost and Entropy-Based Risk Preference. *Transportation Research Record: Journal of the Transportation Research Board*, (2667), 108-118.

Nam, D., Yang, D., An, S., Yu, J. G., Jayakrishnan, R., & Masoud, N. (2018). *Designing a Transit-Feeder System Using Multiple Sustainable Modes: P2P Ridesharing, Bikesharing, and Walking* (No. 18-06518).

Masoud, N., Nam, D., Yu, J., & Jayakrishnan, R. (2017). Promoting Peer-to-Peer Ridesharing Services as Transit System Feeders. *Transportation Research Record: Journal of the Transportation Research Board*, (2650), 74-83.

## **ABSTRACT OF THE DISSERTATION**

Paradigms of Identifying and Quantifying Uncertainty and Information in Constructing a Cognition-Modeling Framework of Human-Machine Transportation Systems

By

Jiangbo Gabriel Yu

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2018

Professor R. Jayakrishnan, Chair

This dissertation proposes a set of coherent cognition-based paradigms to allow greater sensitivity and adaptability to the emerging technologies and behavioral policies. These paradigms are derived from a cognition-based framework that explicates information source, medium, sensation, perception, and learning. The feasibility of the framework is demonstrated through an analytical example of multi-stakeholder decision processes and human-machine systems where the two types of entities can be incorporated into the same modeling scheme. Using the framework as guidance also reduces the challenges from information intractability and data redundancy of agent-based modeling practice.

The first paradigm follows the strict definition of information in Information Theory and models it as the change of uncertainty, which is applied to quantifying traveler information for the evaluation of dynamic message boards that present various contents at candidate locations in Los Angeles traffic networks.

The second paradigm is developed for a utility-based decision model under risk around the proposed concept, Elastic Surprise. This concept makes feasible the differentiation between probability misperception and perceived uncertainty. It is shown that conventional methods of decisions under risk such as Expected Utility Theory and Cumulative Prospect Theory are special

cases. In addition, a specific form of Elastic Surprise under particular assumption on human's cognition leads to Shannon's information entropy and, hence, connects with the first Paradigm. The method is tested in conjunction with the Cumulative Prospect Theory on travel time equivalency under risk in a survey study. The results show improvement in data fitting and output interpretability.

Finally, guided by the framework, the paradigms are tested on a case study of multi-class multi-criteria dynamic traffic assignment where heterogeneous travelers' risk preference on travel time is explicitly modeled. The algorithm approaches the user equilibrium through a stochastic quasi-gradient projection-based algorithm that shows the improvement in computational efficiency and cognitive interpretability of the agents' decision rules. The implication on policy and investment strategies for system improvement is also discussed.

## CHAPTER 1 INTRODUCTION

*"I can live with doubt and uncertainty and not knowing. I think it is much more interesting to live not knowing than to have answers that might be wrong."*

*--Richard Feynman*

As hinted in Dr. Feynman's quote: (1) uncertainty implies multiple possible entities; (2) uncertainty has something to do with knowing (or not knowing); (3) uncertainty leads to something interesting.

I do not have any intention to track the origin and evolution of the meaning of some keywords in this dissertation such as uncertainty, information, and risk, but it would be disrespectful if I failed to mention that not until Claude Shannon's seminal paper published in 1949 (Shannon, 1949), the word "information" only had a vague meaning and people only knew it has something to do with the word "knowledge". Shannon mathematically proved the fundamental relationship between the perceived uncertainty and information, and since then, these two terms have strict definitions. However, in the field of transportation systems modeling and analysis, the terms such as information and uncertainty still have been used as near buzzwords that engineers and researchers treat them as self-defined and any attempt in public to strictly define them will be warned to be realistic. What is even more worrisome is the "exchangeability" of terms such as uncertainty and variability led by the unclear definition and lack of reflection.

On the other hand, even without Shannon's discovery, one can still understand their difference from basic principles. Suppose a system performs as  $y = \sin(a \cdot t)$ , where  $y$  is the system output,  $t$  is time, and  $a$  is a parameter. We know that the system variability is between -1 and 1 while the system has no uncertainty because the system's state is entirely predictable. Therefore, in general, variability describes a system property while uncertainty describes how much information



an observer lacks. The diversity (haphazardness) of “definitions” for uncertainty has only resulted in a challenge of communication among researchers and analysts, and it would be constructive to bring some common ground for the discussion.

It is still a debate on whether the increasing demand for a more fine-grained impact study of policy and investment decisions on human and machine behaviors should lead to a corresponding level of granularity in modeling practice. Most researchers and analysts in transportation have followed a track of the axiom of “rational” or “as if rational” that have already been sharply questioned by the field that initially proposed this idea -- economics. Recent development in cognitive science, mathematical psychology, and neuroscience have shed lights on the possibility of modeling humans as they are – a homeomorphic approach. Modeling practice could benefit when homeomorphic and paramorphic approach are selected based on their strengths rather than specific modeling philosophy imposed by a modeler. When we refuse to impose strong opinions, it is easy to find that any autonomous entities can also be modeled in a similar manner. If we also consider the continuously-improving understanding of human cognition and the rapid development of information storing and computation technologies, there is no reason to refuse the possibility of modeling a society at a more fundamental level. Therefore, when talking about cognition in this dissertation, it refers to an idea of capturing the underlying mechanism of behaviors rather than being content with approximating the revealed behaviors.

The dissertation, in a way, documents my journey of studying human and machine’s cognition and behaviors. The journey started from the exploration of uncertainty and information, but later leads to a broader perspective on how to systemically approach various emerging technologies and business models. The framework at the beginning of the dissertation is proposed as a general principle, and the rest of the chapters are analogous to the bricks that lay one on top of another, culminating at an agent-based dynamic traffic assignment application that brings individual agents together.

Chapter 2 proposes a general cognition-based framework for modeling agents and their interactions with one another and with the physical environment. Also discussed in the chapter is the differentiation of information, uncertainty, risk, variability, and reliability and how they are modeled in this dissertation. To avoid the impression that the dissertation tries to propose a framework that somehow can be magically applied to various domains, it is important to mention that the framework is a result of an iterative effort and has experienced numerous major and minor modifications.

Chapter 3 takes the concept that information can be modeled effectively as the change of uncertainty and applied to evaluating dynamic information provision strategies. The Elastic Surprise Theory (EST) is proposed in Chapter 4 for modeling human decision under risk as a specification of the decision-making submodule of the framework. EST is shown to have cognitive implications and can be the bridging quantity connecting existing theories and methods. EST also bridges the gap between descriptive decision models and information theory and lays a theoretical foundation for Chapter 5 where heterogeneous travelers' risk preference in route choice is modeled.

Chapter 5 combines the previous chapters into an agent-based dynamic traffic assignment procedure in which cognitively-heterogeneous agents learn, perceive, and interact within the same traffic network. The three criteria in the study are travel time, toll, and travel time uncertainty. Travel time uncertainty is classified as an instance of "path-dependent link cost." Individual agent' strategy is derived from a Stochastic Gradient Projection-based searching for an optimal solution in a proposed time-augmented density-based objective function. Policy suggestions and future research are given based on the experiment.

The paradigms are interrelated and coherent, providing exemplary specifications to the cognition-based framework proposed in Chapter 2. The sequence of Framework-Methodology-Application mainly orders the chapters, but practically these chapters are relatively independent.

By modeling any autonomous entities under the same cognition-based framework, human-machine systems are analyzed in a theoretically sound and coherent manner. The discussion of the two intertwined topics – (1) modeling the cognition of autonomous entities and (2) how they deal with uncertainty and process information – transpires throughout the whole dissertation. On top of all the discussion is a crystal and sole vision of improving the modeling practice in this rapidly changing transportation market.

## **CHAPTER 2**

### **A COGNITION-BASED MODELING AND ANALYTICAL FRAMEWORK**

*“A complex system that works is invariably found to have evolved from a simple system that worked. A complex system designed from scratch never works and cannot be patched up to make it work. You have to start over with a working simple system.”*

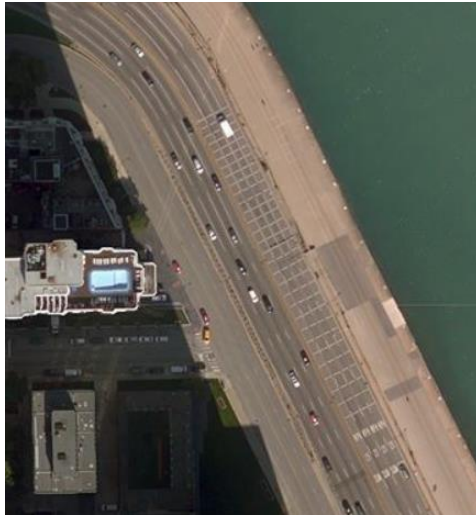
*– John Gall*

### ***Motivations***

Infrastructure systems are known complex. In addition to the large-scale and highly intertwined physical components, different types of influencers behave very differently, having different available information, preference, and socioeconomic and physical constraints. Although the emerging automation and communication technologies excite the general public, it only aggravates this modeling challenge. Wouldn't it be ideal for modelers and educators if all the system influencers can be substantiated from a generic type regardless of what new technological and social development emerges next? This chapter proposes a framework that aims at unifying the modeling and analysis to allow a more comprehensive study of policy and investment impact on human and machine behaviors and greater adaptivity to various existing and hypothetical scenarios with policy nuance.

In 2003, traffic engineers in the Chicago Department of Transportation painted the strapping with an increasingly dense manner when approaching a roadway turn (Figure 2.1). They found 36% fewer crashes in the six months after the lines were painted compared to the same 6-month period of the year. The average speed is also significantly reduced. It seems that the phenomena will not be able to consider in a car-following scheme. On the other hand, it appears that the separate consideration of sensed and perceived information can gain greater modeling

flexibility. In this example, what drivers sense is the moving environment (including the change of the frequency of the lines on the pavement and the change of the “size” of the leading vehicles) while what drivers perceive is the speed. The separate treatment of sensation and perception, whether the observer being human or machine, is a crucial idea of the proposed framework in this chapter.



**Figure 2.1** The increasingly denser stripping deluded drivers that they are driving faster and, therefore, tend to reduce speed before a sharp left turn.

Transportation modeling has been historically focusing on travelers-infrastructure interactions. This means that some key players such as policymakers, private stakeholders, traffic operators, and shippers, are missing. It is hardly arguable about the significance of these players, but a typical counterargument is that they are too complicated to consider or they can only be qualitatively studied on a case-by-case basis.

The agent-based model proves powerful for analyzing and forecasting complex transportation systems, and yet, a flexible and coherent framework is still needed to allow a broad range of policies and emerging technologies to be integrated into the model by merely adjusting model parameters and agent properties rather than adding one extension after another. Agents in this chapter refer to any relevant autonomous entities that may or may not be collaborative.

Examples include individual travelers, traffic and transit operators, adaptive controllers, planners, longshoremen, and autonomous/connected vehicles.

The proposed agent-based modeling and analysis framework, CognAgent, is built upon the underlying cognition rather than the revealed behavior to improve model interpretability and adaptability to unforeseen developments in societal and technological change. CognAgent provides overarching guidance flexible and theoretically sound enough to analyze, model, and forecast various phenomena within transportation systems. Specific contributions of the framework are as follows. First, the framework proposes the idea of using Space of Observables to facilitate the conceptual design, programming, and data management in a modeling process. This concept is substantiated as a module that converts and manages information from the physical interaction module to the type of information observable for each agent; these agents are only able to sense information available to them. Such setup also significantly eases the modularity and interpretability of a model. When modeling the reception of non-local information, the Space of Observables module, plays as an information manager that significantly improves the computational efficiency.

Second, by differentiating sensation and perception, a model's flexibility is significantly improved in dealing with various policy scenarios. Information is modeled in a way that might get lost or distorted in a noisy medium/channel and cognitive process. Foggy weather (for human and AI's camera and radar) and magnetic interference (for V2V and V2I communication) are examples of a noisy channel. Sensing dysfunction includes but not limited to human vision impairment (color-blindness and myopia) and AI's camera and radar dysfunction). Prior beliefs, bias, habits, emotions, or perceptual dysfunctions such as dyslexia (to a dynamic message board) and neuronal degeneration could all lead to a perceptive shift. For entities with adaptive control or AI, processing information could be influence or impaired by factors such as coding errors, cyber-attack, and

problems from computation or storage. CognAgent degenerates into conventional behavior-based model under certain explicit assumptions on agents' cognition.

Third, the framework considers information as a “side product” of modeling uncertainty, which is consistent with the information theory (IT) proposed by Claud Shannon (1949). IT is built upon a strict mathematical formulation and quantifies information as the change of uncertainty given the source, the medium, and the recipient of the information. For a human agent, perception is category-based due to the characteristics of human perception such as grouping effect and cognitive limitation. The Information Processing submodule of the Cognition module sends newly perceived information and combines with prior knowledge to the decision-making submodule for the agent to decide whether and how to behave in the next time step.

Next section reviews relevant literature and then introduces general framework and its specific components, following which are two analytical examples – one for human driving behavior and the other for multi-stakeholder transportation planning decisions. The following numerical case study of mixed-type traffic flow modeling demonstrates the advantage of using CognAgent. This example does *not* intend to predict the traffic condition in the autonomous vehicle era but to demonstrate the advantage of using CognAgent as a modeling and analytical framework. A conclusion is then drawn.

## ***Literature Review***

Agent-based modeling and analysis (ABMS) has been studied and applied in a variety of fields. Helbing (2012) provides a general review of agent-based models and gives an assertion on the potential of the ABMS to a better understanding of social and economic systems. Macal and North (2005) review applications of agent-based modeling and simulation (ABMS) in a variety of fields from natural science to finance. Railsback et al. (2006) reviews and compares five general agent-based modeling frameworks such as NetLogo and MASON, and Railsback and Grimm (2011)

provides a practical introduction to agent-based (individual-based) modeling in general. The primer performed by Zhang et al. (2013) describes basic concepts, various methodologies, and recent progress in both general and transportation-specific applications. Typical modeling and analysis domains related to transportation systems include destination/mode/route choice, driving behaviors, traveler information update, land use and location choice, auto ownership, and activity.

Various cognition models have been used in human-involved systems analysis. Assumptions-Perceptions-Conclusions-Feelings-Behaviors (APCFB) model (Clawson, 1991) has been used in public and business administration to understand individual behaviors in organizations. Since the information flow within the model is clearly structured, insights can be drawn to propose prescriptions. A more homeomorphic approach is neural network models (Hertz, 2018) that seems becoming increasingly feasible for practice for the improved computational power and understanding of human cognition. Recently, quantum cognition modeling approach has been drawing attention. Although quantum mechanics (QM) is originated from microscopic physics, it is not so much about subatomic particles and fields as it is about the non-classical logical principles (Susskind and Friedman, 2014). A less rigorous analogy would be modern calculus – it was motivated by calculating the gravity of a large density-wise heterogeneous object, but it has become a general methodology with many applications out of physics. QM has been applied to fields such as human cognition, neuroscience, descriptive decision analysis, and finance, where the classical logic is challenging to explain and understand various phenomenon and paradoxes (Busemeyer and Bruza, 2012; Haven and Khrennikov, 2013). Some scientists and practitioners have been even searching for the underlying connection between quantum phenomena and consciousness. For example, the Orchestrated Objective Reduction (Orch-OR) hypothesizes that consciousness originated from a quantum process, objective reduction, that is orchestrated by microtubules (Hameroff and Penrose, 1996). Baker (1999) applies Quantum Cognition Models to the analysis of optic flow as a computational judgment model for a driver behavior study. Vitetta (2014) derives a



Quantum Utility Model from QM for modeling route choice behaviors. Busemeyers et al. (2009) replicated and extended interference of categorization phenomena found by Townsend et al. (2000). Franco (2009a) uses a simple quantum model to explain the conjunction fallacy proposed by Tversky and Kahneman (1983). Busemeyer and Bruza (2012) summarize their explorations on applying quantum probability theory to explain cognitive phenomena and provides their prospective outlook on quantum cognition and decision.

Agent-based models can be classified into the domain-specific and the cross-domain, based on the problem scale. The development of domain-specific platforms has been mostly driven by new policies and regulations (Johnston, 2004) and have been implemented in a broad range of applications. Jayarkishnan et al. (1994) provide an evaluation model that incorporates the driver response to information and the corresponding impact on the network performance into an integrated simulation framework that simulates individual vehicle movements as macroscopic flow principles. Ben-akiva et al. (1998) propose a real-time dynamic traffic assignment model that provides traffic prediction and travel guidance to influence driving and route choice. Davidsson et al. (2005) provide a survey of agent-based approaches to transportation and traffic management. Bhat et al. (2004) introduce an activity-travel pattern simulator (CEMDAP) that takes input from land use, demography, activity system, and transportation performance attributes to produce daily activity and travel patterns for each household individual. Sun and Kondyli (2010) refer to the TCP/IP protocol in computer network communications and develop an algorithm to consider competitive and cooperative lane-changing behaviors. Chong et al. (2011) use a neural network model for studying human driving behaviors. Ma et al. (2017) integrate the modified social force model with behavior decision to simulate vehicle turnings under movement constraints. Xu et al. (2011) encapsulate a reference point as endogenous into prospect-based user equilibrium formulation and apply to a congestion pricing study. Hasan et al. (2010) propose a heterogeneous behavioral model to understand household-level hurricane evaluation. Kaihara (2003) formulates

an agent-based supply chain model as a discrete resource allocation problem under a dynamic environment. Tumer and Agogino (2007) use an air traffic flow simulator developed at NASA to test a multi-agent algorithm for traffic flow management. Chen and Zhan (2008) examine the effectiveness of different evacuation strategies using agent-based simulation with diverse information availabilities and strategies. Sun and Wu (2014) propose a generic crowd model to improve the flexibility of incorporating heterogeneous agent behaviors. Fagnant and Kockelman (2014) describe an agent-based model for shared autonomous vehicle operations under a grid-based urban area without specifying the transportation network.

A large body of literature suggests the importance of incorporating uncertainty and reliability into transportation systems analysis to improve modeling result. Studies related to uncertainty and information can be classified based on the source of the information, the medium, and the recipient. Salvucci et al. (2006) utilize ACT-R cognitive architecture to demonstrate the feasibility of modeling drivers as humans rather than “rule-based robots.” Bogers et al. (2005) propose a conceptual modeling framework modeling impact of advanced travel information service, habit, and learning impacts on route choice. Hamdar et al. (2008) base the prospect theory to explore and evaluate a driving model that incorporates the stochastic character of the driver’s cognitive process and impact of risk. Ben-Elia et al. (2008) conduct an experimental study on the combined effect of information and experiment on route choice decisions. Gao et al. (2011) consider the cognitive cost and limited cognitive capacity in route choice decisions. Liu et al. (2002) incorporate heterogeneous perceived travel time uncertainty and perception error into dynamic traffic assignment.

Several attempts have been made in the field of traffic flow modeling for shared human-AV/ACV facilities. Van Arem et al. (2006) use a microscopic traffic simulator to study the impacts of cooperative adaptive cruise control for a highway-merging scenario with a different number of lanes. Levin and Boyles (2015) propose a framework to explore potential effects of heterogeneous

autonomous vehicle ownership on trip, model, and route choice by combining conventional methods such the four-step model and the multi-criterion discrete choice model. Also developed is a multi-class cell transmission model and a car-following model, in which flow-density relationship is modeled as a function of reaction time. In their model, autonomous vehicles are effectively modeled as human drivers but with different reaction times and following distance preferences.

Although cross-domain platforms, in a sense, can be viewed a combination of domain-specific applications, this integration is no trivial task and requires a vast effort for being able to track an arbitrarily-selected agent across different components within a simulation. TRANSIMS (Nagel et al., 1999) provides an integrated set of tools for conducting activity-based regional transportation system analysis with input data of network inventory, demography, land use, and decision-making rules and output as a link-level measure of effectiveness. POLARIS (Auld et al., 2016) provides a cutting-edge software development kit and its modeling and computational framework encompasses inter-connected and feedback-involved components (e.g., activity planning and re-planning). Within these components, persistent agents with learning capability are modeled with particular attention to the coding agility and computational scalability. Waddell et al. (2005) summarize the objectives and design of an open source platform, OPUS, that simulates land use, activity-based travel demand dynamics. Balmer et al. (2009) develop a microscopic traffic simulation tool MATSim-T as part of the research project, MATSim (an open source modeling framework for travel demand modeling, traffic simulation, and activity planning and re-planning). Salvini and Miller (2005) describe the development of an urban system microsimulation model with an intention to be sensitive to transportation, housing, and various urban policies. Arentze and Timmermans (2008) propose a theoretical framework to incorporate the dynamic interaction of social networks and activity-travel patterns in an integrated microscopic simulation. SimMobility (Ben-akiva, 2010) provides a “simulation platform with an integrated model of human and commercial activities, land use, transportation, environmental impacts, and energy use” as an

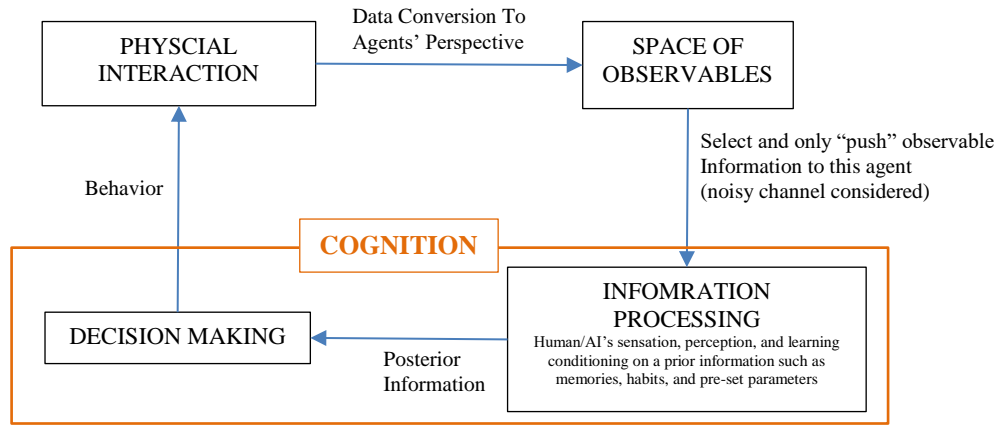
evaluation tool for information provision strategies for system users, operators, and managers in different time scales. Several paralleled evolvments of cross-domain agent-based models exist, and each has been putting effort on incorporating more domains while keeping its edge domains, while new types of modeling frameworks that attempt to cover unconventional domains, such as social networks and sharing economy, are emerging.

Existing ABMS in transportation will continue in the foreseeable future acting as an effective method to build model sensitive to various policies, management strategies, and emerging technologies but are also encountering challenges (Zheng et al., 2013). The improvement and standardization of population synthesis process are still in the phase of discussion and exploration (Muller and Axhausen, 2010). ABM calibration and its integration with its “parent” system are recognized as two of the greatest challenges facing microscopic simulation; other challenges include but not limited to computational scalability, data accessibility, justification of complete profile of assumptions on how agents make decisions (Picascia et al., 2016).

## ***Framework***

In contrast with typical behavioral models that only specify rules for decision and physical interaction, the proposed framework conceptually consists of a one-direction loop of information flow with three fundamental components: physical interaction, space of observables, and cognition, as shown in Figure 2.2. The cognition module contains two general submodules: information processing and decision making. Agents share one single “physical environment” to interact and the interaction at a given time step is then converted to the observable information back to agents. Whether a piece of information is observable depends on the agent’s location, medium/channel, and the type of sensors. The agent then perceives/processes this newly sensed information combining with memories, habits, emotions, personalities. If the agent is an AI entity, the algorithms defined by its manufacturer should be specified or assumed. This setup not only

improves model/code readability by standardizing heterogeneous agents' configuration but also facilitates the quality assurance of scripting and improves the efficiency of database management.



**Figure 2.2** The logical framework of CognAgent

Let us use an example to explain how CognAgent improves the tractability of information dynamics in a simulation. Suppose that an agent hears (acoustically senses) from a radio on a severe accident on a road segment ahead. This radio report is further confirmed by a dynamic message board information (visionarily senses) through foggy weather and hence shorter sight distance and a higher likelihood of misreading. This posterior information suggests a longer-than-usual travel time to go downtown. Though the driver's current route to the destination only shares a portion of the route to go downtown, the newly updated information is sufficient for him to change his route decision. Due to the foggy weather, he also decides to drive slower. Traffic operator in the local traffic management center has sensors (loop detectors at several point locations, cameras, and related emergency calls) at finite locations, and only these locations are observable to the operator. The operator perceives the available traffic information combining with software support and his professional judgment to recognize the traffic conditions and decides among alternatives. Examples of traffic operator interacting with the environment include sending emergency vehicles, changing traffic control timing, and putting up new content on dynamic message boards. The above processes are naturally fitted into the CognAgent framework.

An iteration can be either short-term, long-term, or mixed. For example, a driver may update a perceived traffic condition every 1-2 seconds but only update perceived alternative routes' travel times when non-recurrent events occur. In a model for land use-transportation interaction, a loop can be as long as multiple years. In an activity-based model with traffic assignment functionality, travelers may make adjustments to their schedules and route choices hourly while also plan for other decisions (e.g., location and auto ownership) on a longer timeframe. However, regardless of the loop length, all the agents share the same physical interaction module. The following subsections detail the three modules.

### **Physical Interaction**

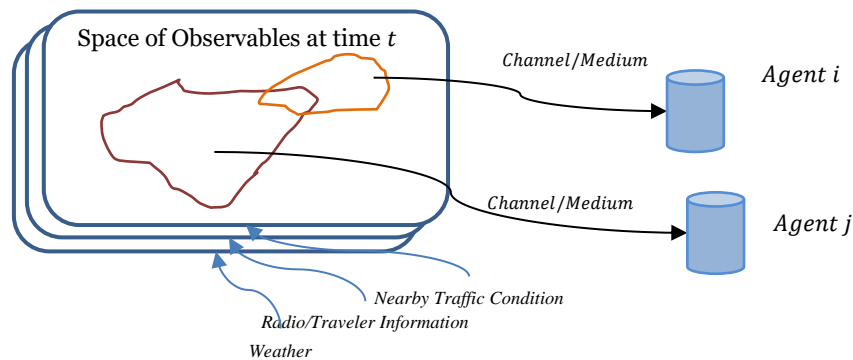
The module of physical interaction is the only module that all the agents share. Agents influence one another and the simulated environment through this module. Conventional discrete or agent-based models such as mesoscopic and microscopic traffic simulators are applicable. This module combines the behaviors from agents at a time step with the system state in the previous time step to form a new system state.

### **Space of Observables**

The module Space of Observable converts data obtained from physical interaction module to information observable by a given agent. As shown later in the case study, this seemingly redundant component greatly facilitates data management, programming extendibility, and information transmission in noisy media. An observer receives information from the Space of Observables through a channel which may or may not have noise. Examples of a noisy channel include rainy and foggy weather and signal interference in telecommunication.

The module determines whether an agent can observe certain information. Some of the factors are agent's location, sensor types, and sensor quality, noisy channel. The module can be seen as a manager organizing information and mapping them from the simulated "physical reality" to an

observer. One computational advantage is that agents do not need to determine whether he/she/it can observe a certain event him/her/itself. For example, when an event occurs, the traditional method would compute from agent's perspective, and each agent would determine whether to "use" the information about the event or not. But in CognAgent, the Space of Observables only assigns the information about the event to the relevant agents directly. In microscopic traffic simulation, data held by the Space of Observables component keeps being updated and resemble a master list of the information that has the potential to be observed by an agent. In an integrated model of activity scheduling and dynamic traffic assignment, if a traveler did not travel nor consult traffic information on a given day, travel time information is unavailable to him/her/it and. Figure 2.3 demonstrates how the module manages different types of information and only assigns information to the relevant agents.



**Figure 2.3** Conceptual Graph Demonstrating the Data Management Role of the "Space of Observable" Module in Human Driver-Only Traffic

### Cognition

Cognition module is formed by two submodules -- information-processing and the decision. Information-processing submodule obtains (senses) information from the space of observables and processes (perceives) information combining with prior knowledge. The goal of the information processing submodule is to provide *all* the inputs required for decision making.

Modern psychology and cognitive science recognize that there are, in effect, two modes of thinking (Evans, 2009). Mode 1 is intuitive and pattern-recognizing while Mode 2 is methodical and logical. Our brains usually “operate” on Mode 1, which is prone to be influenced by the memory, context, and emotion. Mode 2 is usually not triggered until unfamiliar, challenging, or stressing events emerge. However, Mode 2 might “exhaust” or “find” the problem too hard to justify the cost of attention that it reduces the effort to let the Mode 1 to use more heuristics. The cognition module in Figure 2.2 is a high-level abstraction, but it can be thought of as having two interacting procedures that share the two submodules, which one or how much each procedure is triggered depends on a variety of circumstances. The level of details of specifying how to model the two modes explicitly is out of the scope of this dissertation.

#### *Information Processing Submodule: Sensation, Perception, and Memory*

An observer receives information from the Space of Observables through the medium that may or may not be capacity-limited and noisy. Sensation represents stimuli from the environment (e.g., human body converts stimuli into electrochemical signals to transmit the information in the nervous system) and perception is a higher brain function about interpreting events and objects in the world (Mayer, 2011). The sensed information is combined with factors such as memories, habits, bias, belief, emotion, and influence to form perceived information with cognitive cost and temporal delay.

Perception is well-known as context- and category-dependent. For example, 17 min is just a number and does not mean anything to a traveler until it is perceived as travel time given a context with the specific origin, destination, type of the activity, and preferred arrival time.

Cognitive cost is associated with information acquisition, sensing, storing, perceiving, and learning. Cognitive disparity among agents can be classified into sensation-related (e.g., color blind and hearing loss) and perception-related (e.g., dyslexia for reading dynamic message board). A general agent-based traffic model should be capable of considering travelers/drivers with sensing



challenge (e.g., cataract or myopic eyes, hearing disorder, and Eustachian tube disorder) and sensory limitation/failure on, say, an autonomous vehicle. Common variables include but not limited to age, gender, education, memory, personality, value, beliefs, psychological and cognitive disorders and dysfunctions, social and cultural context. For autonomous/connected vehicles, factors include but not limited to processing speed, accuracy, and stability.

An agent's mental state evolves with newly obtained information that may have both short-term and long-term effects. An agent may receive a piece of information in two general situations. In Situation I, the perceived information is not in the working memory (short-time memory), and, therefore, the agent needs to load the context information from long-term memory to understand the newly obtained information. In Situation II, the agent obtains the information when he is making relevant decisions. Situation II is highly context-dependent and prone to mis-interpret the information.

Although any memory systems share similarities in encoding, storage, and retrieval, some functions of the memory system of human agents need special treatment to consider things like sensory memories, short-/long-term memories, and working memory. Consideration of characteristics such as automatic and effortful processing, serial position effect (e.g., tendency to recall best the first and the last in a short sequence of items or events) could potentially improve the predictive power of an agent-based model.

### *Decision Submodule*

Cognition module provides contextual and category-based information that an agent needs for deciding how to behave. When a modeler needs to consider the impact of uncertainty on a decision (e.g., route choice), the input requires a distribution of a percept (e.g., travel time) or a joint distribution of a set of percepts (e.g., travel time and monetary cost). This information is obtained from the information-processing submodule. When implementing conventional disaggregate travel demand models in this submodule, agents can make decisions based on information about the

choice attributes constrained by capacity (e.g., total number of seats in the vehicle), coupling (activities that require other agents), and destination (e.g., post office/bank/grocery opening hours, school time, work time, transit operation hours, etc.). Additionally, an agent can share information (as a decision) through physical interaction module with other agents. Note that an agent might decide not to behave and wait for more information before making decisions.

It is important to clarify which agent makes which decision. In passenger transportation, most mode and route decisions are made with/without the aid of traveler information providers. In a likely future, travelers may only decide destination, preferred arrival time, and certain preference (e.g., avoid toll and willing to share ride) and the vehicle/AI (and the server in the management center) will do the rest. In the current global freight systems, there already exists a level of discretion by algorithms to optimize, and humans such as shippers and port/warehouse queuing list managers usually have little need to decide unless there's major supply or demand changes, while truck drivers in minor route (e.g., distribution center to retail store, delivery and mailing service) have relatively high flexibility of route and sequence. Each agent imposes an adjustment on the physical interaction module by behaving (i.e., implementing the decision).

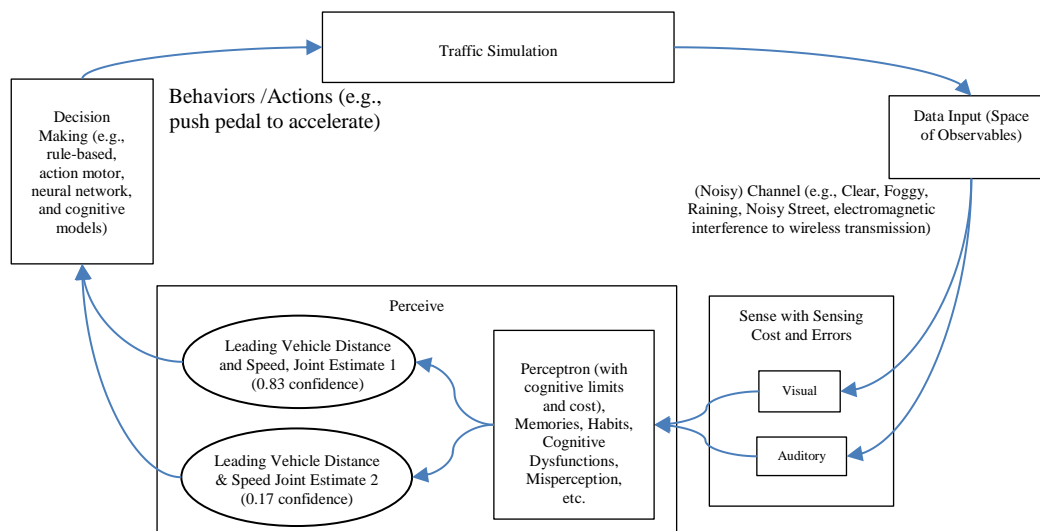
Although it has been a changing task to model creative solutions, this submodule could potentially incorporate this aspect. That is, even if a group of agents with the same cognitive characteristics and available information encounter the same situation, they might still form different choice set from which to choose. From a psychological perspective, the process of forming a choice set requires an iterative effort and, hence, interact heavily with the information-processing submodule.

A behavior results from an individual or joint decision. For example, the selected route might be a decision made by an autonomous vehicle with the preference decision of the traveler and the guidance of command from the traffic management center.

## Analytical Examples

### Human Driving Behavior

A human agent senses and perceives to form a new understanding of the surroundings when driving. Factors such as experience and habits also influence this understanding. Figure 2.4 shows an example of analyzing this agent's behavior cognitively. Conventional traffic simulator can be used in the physical interaction module. The module for the space of observables "translates" the results from the traffic simulator at each time step to information sensible by the agent. This module prepares the input for sensing and perceiving. When a driver perceives information such as distance from the leading vehicle under uncertainty (e.g., 0.3 confidence that the leading vehicle is 50ft away and 0.7 confidence 60ft away), his/her risk preference might influence how he/she learns information. The updated distribution is then sent to the decision-making submodule for the agent to behave and interact with the physical environment. Different risk levels might also trigger different perception and decision mechanism.



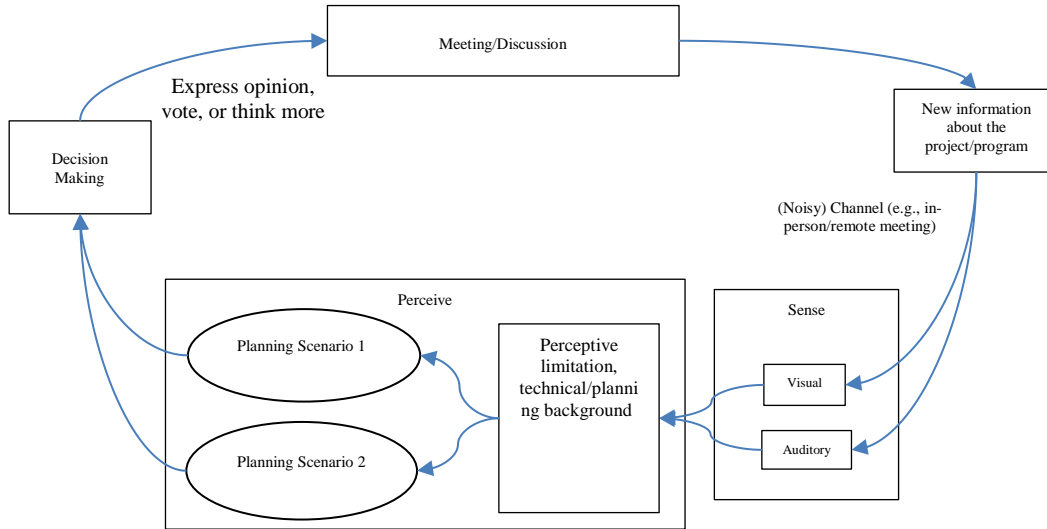
**Figure 2.4** Example of analysing driver behaviour under uncertainty based on CognAgent

The level of detail in modeling depends on the specific problem, scope, and information availability. Agents may or may not share the same information update and decision frequency.

However, it is important to clarify the underlying mechanism when simplifying the modeling approach. Cognitively, a driver estimates the distance from the leading vehicle by comparing the belief of the “actual” vehicle size with the perceived size of the leading vehicle to estimate speed by comparing the changing rate of the vehicle size. When perceived information arouses emotions such as stress, anxiety, and anger, the driver’s perception and decision style could change dramatically. A driver’s cognitive process may bypass the cortex (i.e., regular neural pathway via thalamus, sensory cortex, prefrontal cortex to amygdala) and directly send sensory input to the amygdala for an instant emotional reaction such as fear (LeDoux, 2003).

### **Multi-stakeholder planning decision**

Behaviors of different levels of public agencies, competing/collaborating private entities (e.g., consulting firms and transportation network companies), research institutes, political constituents, and policy and regulations are indispensable components for studying transportation systems. Under CognAgent, each relevant autonomous entity has its own corresponding space of observables and information processing scheme. For example, the collective cognitive character of a public agency is formed by the cognition of individuals within it. Some key factors to consider when studying the planning decision-making process are how the information is obtained, collective bias, standard procedure, organizational cultural, innovation level, criteria (objectives), and risk preference. Historical records on how these entities made decisions and how effective these decisions were implemented can be used for calibration. Figure 2.5 shows an example of the analytical framework, which has the same structure as in Figure 2.4.



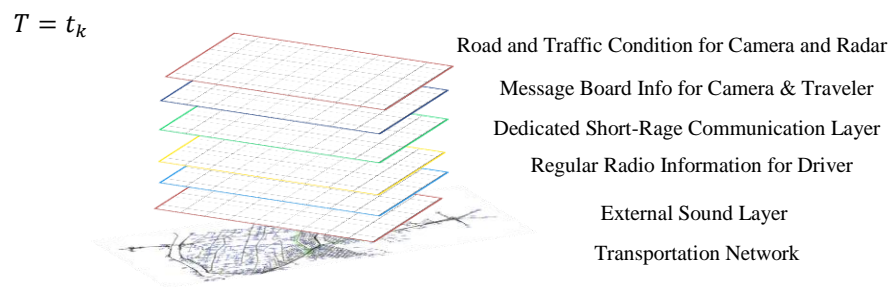
**Figure 2.5** Example of analysing a stakeholder meeting on infrastructure investment based on CognAgent

### ***Numerical Feasibility Study – Human-ACV Mixed Flows***

Modeling human drivers and intelligent vehicle technologies have no fundamental difference under CognAgent. V2X communication also features “sensing” and “perceiving” functions. An operating autonomous vehicle senses surroundings using sensors such as radar and camera and perceives using algorithms set by manufacturers, constrained by regulatory rules and industrial standards. Limited sensor resolution and measurement accuracy might render uncertainty. Travelers and their vehicles form joint decision (if semi-automated) based on the combined the result of the processed information and historical data (may or may not be mis-remembered/stored) to make decisions such as acceleration, deceleration, lane change, and route adjustment.

The concept of “space of observables” significantly simplifies the modeling effort in modeling and simulating heterogeneous ACVs when non-local information is significant. Figure 2.6 uses a set of raster layers to illustrate this concept. Each location has its corresponding observable which is managed in a system of layers. An agent at a certain location can *only* sense information from his/her/its corresponding “column” of cells. Whether the information is received depends on the condition of the observer in that cell at a certain time. For instance, an agent without radar or

with radar turned off cannot sense information from the radar layer. Therefore, the coding within the module Space of Observables is only in charge of retrieving and converting information from the module of physical interaction and need not to concern whether there is an agent at a certain location. Some layers are exogenous (e.g., radio layer) while some are endogenous (e.g., vehicle collision.) Some layers have highly heterogeneous information among cells (e.g., traffic condition along a roadway) while others have very similar information among cells (e.g., traffic radio and weather).



**Figure 2.6** Coding “space of observables” as a raster-like database.

## Background

A stretch of uninterrupted single-lane access-restricted straight roadway on level terrain is used to demonstrate the feasibility and advantage of the framework as a guidance of implementation in a traffic simulator. Although rules of ACV agents can be defined based on manufacturers, this case study categorizes agents regarding sensor type and risk preference. A general rule for ACV agents is to set objective to be staying in the middle of the leading and following vehicles by adjusting the acceleration/deceleration rate, while, in sparse traffic, ACV agents accelerate in their designed rates to the maximum safe speed until reaching the safe distance with the front vehicle. Scenarios with communication technology will be briefly discussed as an extension in the latter part of this section.

## Defining Agent Class

Two classes of agents, human and ACVs, are inherited from the same parent class, *CognAgent*, in which methods for sensation, *sense()*, and perception, *perceive()*, are specified correspondingly. The

algorithm Part I (pseudo code consistent with python 3.x syntax) defines how properties and methods are defined in the parent class. *Veh\_state* is a list that generates agent’s ID, original location, speed, acceleration; the list gets updated at each time step. The updated information is then stored in *lookup\_table* for each agent. The algorithm part I utilizes methods from algorithm part III, PISOO, which is an object instantiated from the class that holds two main functions – (1) storing locations, speeds, and accelerations of all vehicles at the current simulation time step; and (2) providing output of what information can be potentially sensed by the corresponding agent. As can see, the method *sense()* sequentially triggers the *perceive()* and *behave()* method of its own instantiation.

**ALGORITHM PART I: PARENT CLASS DEFINITION FOR HUMAN AGENT CLASS AND INTELLIGENT VEHICLE CLASS**

```

CLASS COGNAGENT:
    INITIALIZE (VEH_INITIAL_STATE):
        VEH_STATE = VEH_INITIAL_STATE
    METHOD SENSE(VEH_STATE.LOCATION, VEH_STATE.SENSORS):
        RETURN SENSEINFO = PISOO_OBJ.QUERY(VEH_STATE.LOCATION,
        VEH_PROPERTY.SENSORS)
    METHOD PERCEIVE(SENSEDINFO, PRIORINFO):
        RETURN PERCEIVEDINFO(SENSEDINFO, PRIORINFO)
    METHOD BEHAVE(PERCEIVEDINFO):
        RETURN VEH_STATE = BEHAVIOR(PERCEIVEDINFO)

```

Table 2.1 shows the type of information sensible by an agent. The camera is assumed to be used to sense distance, while radar can be used for both distance and speed. ACVs combine distance information from camera and radar with equal weights to form sensed information on distance (in foggy weather or at night, camera might have less weight). Human agent is assumed only able to use vision to sense and perceive the leading vehicle’s location and speed.

**Table 2.1** Sensing capability assumed in this paper on human and ACVs

Information available for an agent to sense at location <i>x</i> at time <i>t</i>	Human	ACV
Leading vehicle distance and speed	√	√
Following vehicle distance and speed		√
*Location and speed of immediate vehicles on the left & right lane(s)		√
*Location and speed of immediate vehicles on the left & right lane(s) within view range	√	√
DSRC within 1000ft (FHWA recommended range, 2016)		√

\*the example only considers single-lane plain terrain scenario.

Two-category perception is set on human drivers – the sensed information from the current time step and the sensed information 0.3-0.4 sec prior (uniformly distributed) to considering the sensory memory and error. The perception time is set to be 0.7-0.9 sec (uniformly distributed), and decision-to-behavior time is configured to be 1.0-1.3 sec (uniformly distributed). Therefore, the perception-reaction time (PRT) is 2.0-2.4 sec (uniformly distributed). Different types of delays are assumed independent with one another, and future study can incorporate their correlation. The perceived confidence for the current sensed information is randomly generated from the uniform distribution of 0.60 to 0.99 for each human agent. Newell’s car-following model is used for simulating human driving (Newell, 1961), the decision on acceleration/deceleration rate at each time step is set proportional to the speed difference to that of the leading vehicle at time step  $t - \tau$ . The reaction time,  $\tau$ , is set consistent with PRT and the sensitivity indicator,  $\lambda$ , is set uniformly distributed between 0.8 and 1.2 per second. The method *behave(.)* implements the decision by updating its physical location constrained by the location of the leading vehicle. The risk preference is considered for ACV and human agents by adjusting the weights for the favorable and unfavorable outcome. The adjusted weight is defined as

$$\pi_j = \pi(p_j) = \beta \cdot p_j^\alpha, \alpha > 0, \beta > 0 \quad (2.1)$$

for the relatively favorable outcome,  $j$ , while the relatively unfavorable outcome is  $1 - \pi(p_j)$ . In this case study, it is assumed that the lower the acceleration gap, the less favorable. When  $0 < \alpha < 1$ , the human agent penalizes more on the riskier outcome, and therefore, risk-averse; when  $\alpha > 1$ , risk-prone; when  $\alpha = 1$ , risk-neutral.  $\beta$  is a scaler set as unity in this case. The index,  $j$ , is ranked so that the closer to 1 the more preferred.

The ACV agent driving rule is inspired by a simple principle – each ACV agent attempts to remain in the middle of the leading and the following vehicle in congested traffic and accelerate with comfortable rate to either the maximum speed or reaching the allowable distance to the front



vehicle in uncongested traffic.  $x_{n+1} - x_n$  and  $x_n - x_{n-1}$  represent the distances to the leading and following car, respectively. The rule can be described as

$$a_t^n = a(s_t^{n-1}, s_t^n, s_t^{n+1}, \mathcal{M}_t^n) \quad (2.2)$$

$s_t^n$  is a state vector containing information of location, speed, and acceleration of vehicle  $n$  at time  $t$ . Following convention, human drivers are assumed to be only able to sense the location and speed of the front vehicle.  $\mathcal{M}_t^n$  is a perception map for the observer  $n$  at time  $t$  that converts the sensed to the perceived.  $\mathcal{M}_t^n$  is updated based on newly obtained information through a learning function  $f(\cdot)$ . That is,

$$\mathcal{M}_{t+\delta}^n = f(\mathcal{M}_t^n, s_t^{n-1}, s_t^n, s_t^{n+1}) \quad (2.3)$$

Learning rate is assumed consistent with the simulation time interval, and therefore, agents update their information every time step.

Detailed description for acceleration rate for each corresponding condition is given in Table 2.2. ACV agents adjust the physical location and speed, given traffic system state  $\mathbf{s}$ , via acceleration/deceleration which are  $a_{mild} = 3ft/s^2$ ,  $a_{max} = 9ft/s^2$ ,  $d_{mild} = -4ft/s^2$ , and  $d_{max} = -13.17ft/s^2$  depending on the prevailing local traffic condition. ACV reaction time is uniformly distributed between 0.01 and 0.015 sec.  $\epsilon$  is a perceptive threshold related to speed difference and linked with sensor quality and specific algorithm the manufacturer sets. When the ACV perceives the difference between the speed of itself ( $v$ ) and the speed of the leading vehicle ( $v_l$ ) to be over  $\epsilon$  or below  $-\epsilon$ , it considers this difference significant.  $e$  is set to be 9 ft/sec. ACV agents are set with 0.95 confidence on the sensed information at the current time step and 0.05 confidence on the historical processed result as a reflection of imperfect sensing and perceiving functions.

**Table 2.2** ACV Operation Rules (DSRC will be incorporated as an extension later)

Any DSRC Warning?	Distance Criteria		Acceleration/Deceleration ( $ft/sec^2$ ) for different speeds relationships			*Send DSRC Warning?
	$x_{n+1} - x_n$	$x_n - x_{n-1}$	$v_l - v > e$	$ v_l - v  < e$	$v_l - v < -\epsilon$	
No	$(0, s_{min}]$	$(0, s_{min}]$	$d_{mild}$	$d_{mild}$	$d_{max}$	N, N, Y
		$(s_{min}, s_{max}]$	$d_{mild}$	$d_{mild}$	$d_{max}$	N, N, Y
		$(s_{max}, \infty)$	$d_{mild}$	$d_{max}$	$d_{max}$	N, Y, Y
	$(s_{min}, s_{max}]$	$(0, s_{min}]$	$a_{max}$	$a_{mild}$	0	N, N, N

		$(s_{min}, s_{max}]$	$a_{mild}$	0	$d_{mild}$	N, N, N
		$(s_{max}, \infty)$	$a_{mild}$	0	$d_{max}$	N, N, Y
	$(s_{max}, \infty]$	$(0, s_{min}]$	$a_{max}$	$a_{max}$	0	N, N, N
		$(s_{min}, s_{max}]$	$a_{mild}$	$a_{mild}$	$a_{mild}$	N, N, N
		$(s_{max}, \infty)$	$a_{mild}$	$a_{mild}$	$a_{mild}$	N, N, N
Yes	$(0, s_{max}]$	-	$d_{mild}$	$d_{max}$	$d_{max}$	N, Y, Y
	$(s_{max}, \infty]$	-	$d_{mild}$	$d_{mild}$	$d_{max}$	N, N, Y

\*DSRC warning does not propagate in this case study and will leave as future extension.

The two inherited child classes have a similar initialization procedure. Non-sharing properties such as agent types and available sensors are defined and initialized based on the specific child class. Algorithm Part II is the pseudo code of class definition for ACV and human agents.

---

**ALGORITHM PART II: HUMAN AGENT AND ACV AGENT CLASS  
INHERITED FROM COGNAGENT**

---

**CLASS ACV(COGNAGENT):**

**INITIALIZE** (VEH\_INITIAL\_STATE):

PARENT.INITIALIZE (VEH\_INITIAL\_STATE)

VEH\_PROPERTY.AGENTTYPE = 'ACV'

VEH\_PROPERTY.SENSORS= ['CAMERA', 'RADAR', 'DSRC']

**METHOD UPDATE**(VEH\_STATE):

SENSE(VEH\_STATE.LOCATION, VEH\_PROPERTY.SENSORS)

RETURN PRIORINFO.APPEND(VEH\_STATE)

**CLASS HUMAN(COGNAGENT):**

**INITIALIZE** (VEH\_INITIAL\_STATE):

PARENT.INITIALIZE (VEH\_INITIAL\_STATE)

VEH\_PROPERTY.AGENTTYPE = 'HUMAN'

VEH\_PROPERTY.SENSORS= ['VISION']

**METHOD UPDATE**():

PRIORINFO.APPEND(VEH\_STATE)

RETURN PARENT.SENSE(VEH\_STATE.LOCATION,

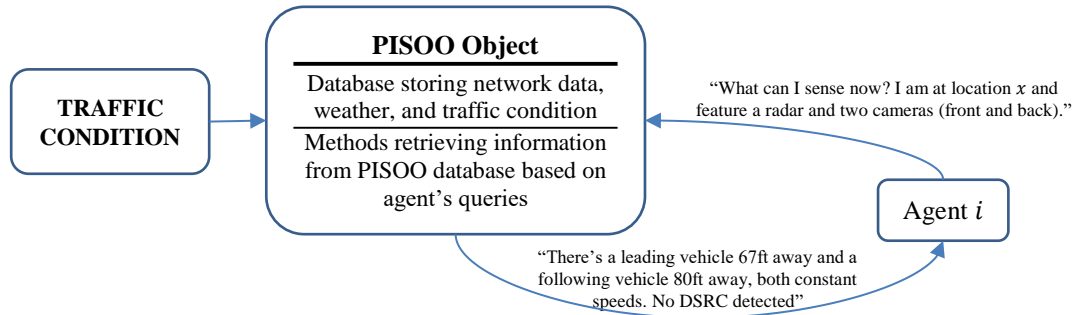
VEH\_PROPERTY.SENSORS)

---

### Defining PISOO Class

Methods related to the physical interaction and the space of observables are jointly embedded into one class, PISOO. When an agent senses information from a space of observables, it calls the method *updatePhysicalInteraction(.)* in a PISOO object, instantiated from its class. The PISOO object then searches *lookup\_database* for what this agent could potentially sense. Whether the agent can process that information depends on the type and the quality of the sensors the agent features and the perceptive function. *lookup\_database* stores all the vehicles' physical information (in this example: location, speed, acceleration, vehicle length) and network conditions (e.g., relative

location with other agents) at the current time step in a simulation. Figure 2.7 shows an instance of an ACV agent inquiring and receiving “customized” information from the PISOO object.



**Figure 2.7** Example of an agent “inquiring” the instantiation of PISOO at time  $t$ .

Similar to any simulation, finite time step could generate unrealistic traffic condition (e.g., the spacing between two vehicles is smaller than the length of the leading vehicle). Therefore, the algorithm also checks feasibility and constrains the update if necessary. This case study does not consider traffic accident, though there is no technical difficulty to incorporate it thanks to the flexibility of CognAgent. Algorithm Part III is the pseudo code of the class definition of PISOO based on Python 3.x syntax.

---

**ALGORITHM PART III: DEFINING PHYSICAL INTERACTION DATA BASED AND SPACE OF OBSERVABLE**

---

**CLASS PISOO:**

**INITIALIZE** (INITIAL\_TRAFFIC\_CONDITION):

LOOKUP\_DATABASE = INITIAL\_TRAFFIC\_CONDITION

**METHOD** UPDATEPHYSICALINTERACTION(TRAFFIC\_CONDITION AT  $T+1$ ):

FEASIBILITY = CHECKFEASIBILITY (TRAFFIC\_CONDITION AT  $T+1$ )

UPDATE LOOKUP\_DATABASE WITH TRAFFIC\_CONDITION\_AT\_ $T+1$ :

IF FEASIBILITY==1:

RETURN TRAFFIC\_CONDITION = UPDATE TRAFFIC\_CONDITION\_AT\_ $T+1$

ELSE:

FOR  $\forall$  AGENTS:

VEH[I].VEH\_STATE = CONSTRAINED TRAFFIC\_CONDITION\_AT\_TIME\_ $T+1$

RETURN TRAFFIC\_CONDITION = UPDATE CONSTRAINED

TRAFFIC\_CONDITION\_AT\_ $T+1$

**METHOD** QUERY(LOCATION, SENSORS):

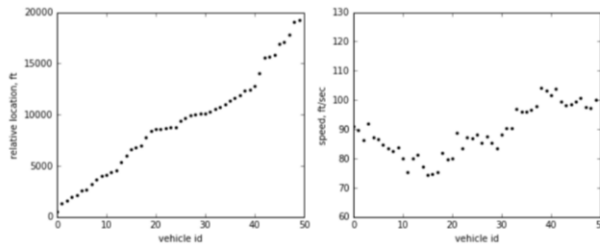
INFOAVAILABLE = LOOKUP\_DATABASE.LOOKUP(LOCATION)

RETURN INFORMATIONSENSABLE = INFOAVAILABLE.SENSORS(SENSORS)

---

## Agent Instantiation and Simulation

50 agents are instantiated from either the human class and the ACV class. The ratio of the two types of agents depends on the market penetration (MP) rate. Both classes are inherited from CognAgent class and yet have their own properties and methods. The location is randomly generated and tagged with sequential IDs while making sure minimum distance, 40ft, is maintained. The speed of the most leading vehicle is set at 100ft/s. The initial speeds of the following vehicles are randomly generated from a log-normal distribution with the average being the speed of the corresponding leading vehicle and the standard deviation being 23 ft/s. This initial setting is for testing the traffic performance for different MPs of AVCs in the mixed traffic. Figure 2.8 shows the initial (fixed) setting of the traffic to allow comparable sensitivity tests.



**Figure 2.8** Vehicle locations and speeds at  $t = 0$

When an agent is instantiated, the method, *Initialize()*, is called so that *veh\_initial\_state* is saved into the agent object as the agent's initial state (i.e., vehicle location, speed, and acceleration). The priorInfo in this example is set as the historical vehicle states. Information such as habits, bias, and (for human agents) can also be incorporated. An agent's cognition (sense, perceive, and behave) is triggered by calling the method *update()* in Algorithm Part II. Since most of the process is handled in the objects themselves, the simulation becomes simply calling the update method within each agent object. At each time step (0.01sec), Algorithm Part IV calls the PISOO object and each agent updates itself.

---

**ALGORITHM PART IV: SIMULATION**

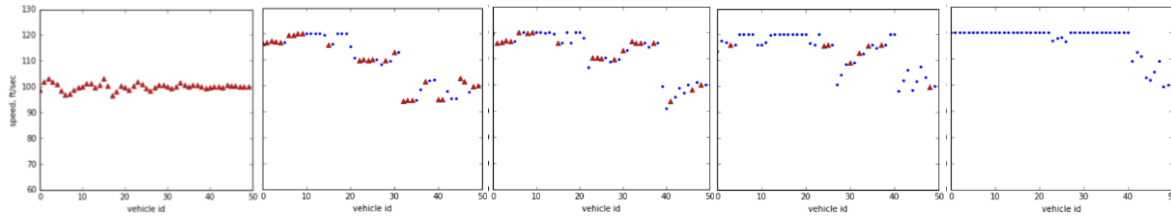
```

PISOO_OBJ = PISOO()
FOR T IN RANGE(0, TIMESPAN, STEPSIZE):
    FOR I IN VEH_ID:
        VEH[I].UPDATE()
    PISOO_OBJ.UPDATEPHYSICALINTERACTION()

```

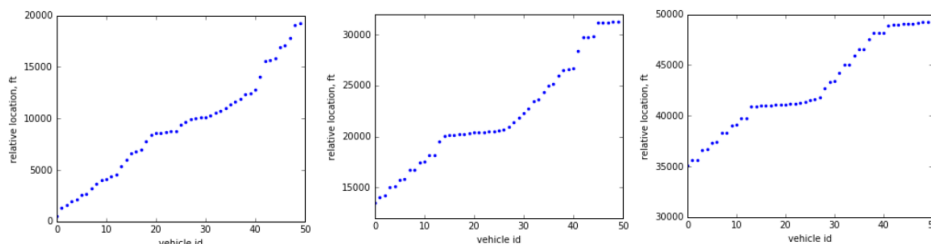
---

Figure 2.9 shows the sensitivity analysis of MP at different time steps. All the simulations share the same initial traffic condition by setting a fixed random seed.



**Figure 2.9** Y-axis (bottom to top): 0, 0.3, 0.5, 0.7, 1.0 MP Rate (red triangles are human drivers; blue dots are ACV). X-axis (left to right):  $t = 300s$ . The appendix shows the complete figure.

In the rule and parameter settings in this numerical example, even a low market penetration of autonomous vehicles will have a significant impact on the performance of the traffic flow. Although the human drivers tend to generate similar velocities at a faster rate (speed limit is set as 100ft/sec), they tend to waste the space when the leading vehicle is far. On the other hand, ACV tends to generate different brackets of speeds (speed limit is set as 120ft/sec) and tend to rapidly “catching up with the front vehicle.” The higher the MP of the ACVs, the faster it seems to converge the speed and produces “platoon” (the frontest vehicle, 50<sup>th</sup>, is set to 100ft/sec). In the scenario with full MP, a set of stable platoons are formed in the 300sec as shown in figure 2.10.

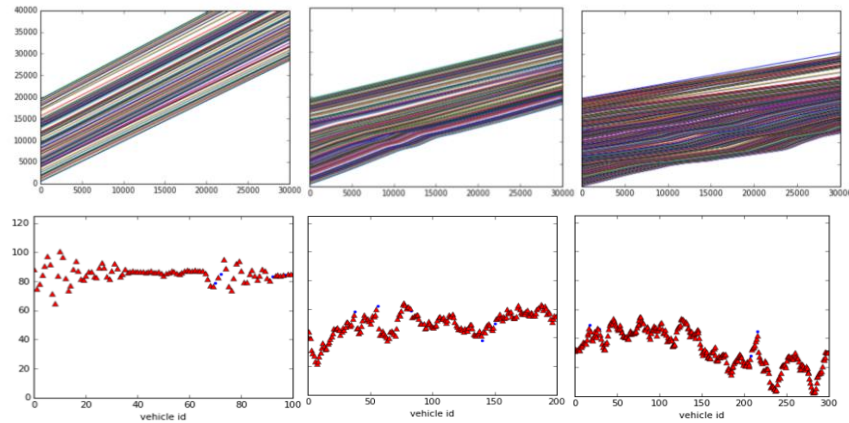


**Figure 2.10**  $t=0,120s,300s$  (y-axis are set to 20000ft to all the three plots for comparability)

Under this setup, the traffic flow with a high percentage of human agents quickly starts oscillating, while the traffic flow composed of a large percentage of AV agents tend to have smoother acceleration/deceleration and do not overreact to various traffic conditions. For more details, please see Appendix A.

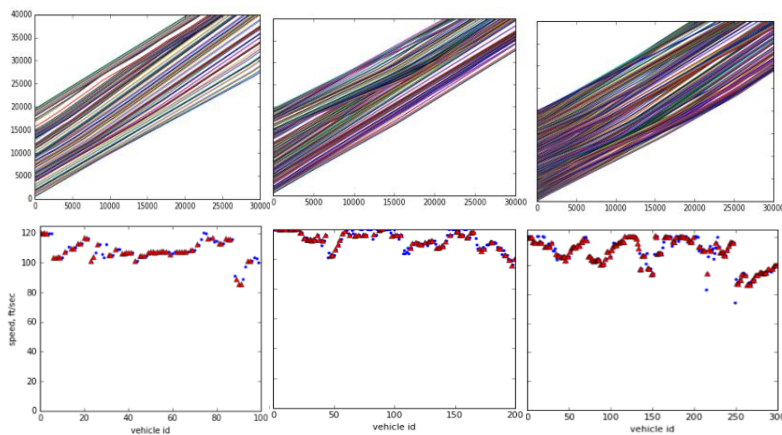
### Density, Market Penetration, and Risk Preference Sensitivity Test

The impact of various market penetration rates and traffic density are tested. Figure 2.11 shows the traffic flow at different density level when most agents are humans at different phases of the simulation. Shock waves and speed drop at 5 min appears when the density is heavy.

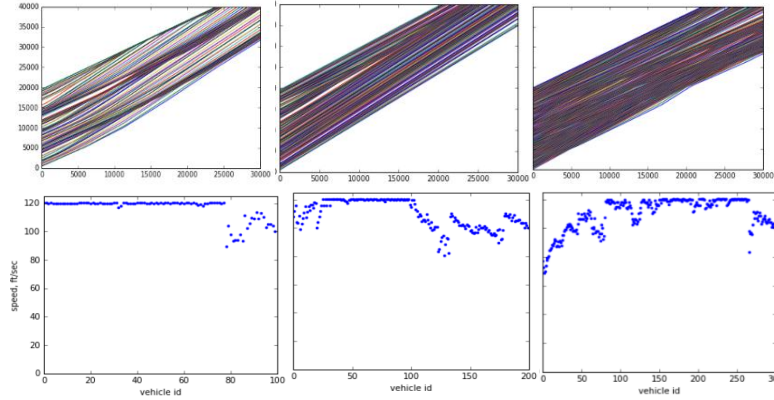


**Figure 2.11** The trajectories (upper) and speed profile (lower) while 5% MP and 5-min simulation horizon; number of vehicle is set as 100, 200, and 300, respectively.

Similarly, Figure 2.12 and Figure 2.13 shows the trajectory and speed profile when MP is 50% and 100%, respectively. The shocking waves are “observed” even when only half of the vehicles are AVs. Speed-wise, although the traffic stability decreases when traffic density rises, the traffic speed does not severely “cascade” When all the vehicles are AVs, vehicles quickly form platoons with little shocking waves. Speed-wise, the traffic stability is also improved and has a faster rate to reach steady traffic state.

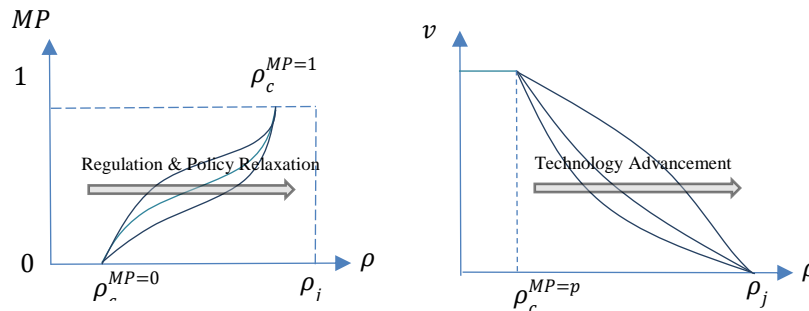


**Figure 2.12** 50% MP and 5-min simulation horizon; veh = 100, 200, 300

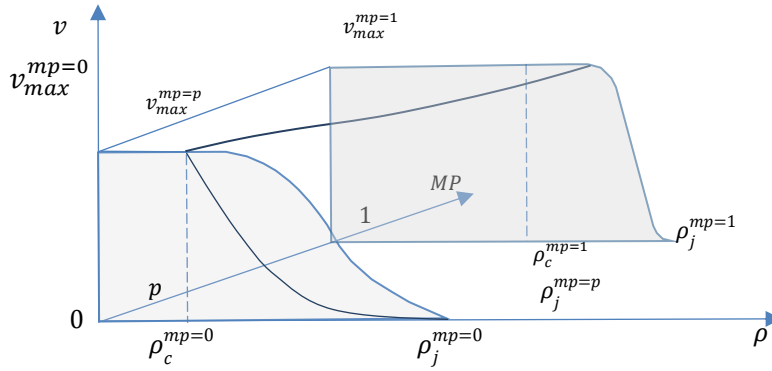


**Figure 2.13** 100% MP and 5-min simulation horizon; veh = 100, 200, 300

Three major observations from the test are: (1) with the MP rate increasing, both the limit speed of stability and the average speed in congested condition increases; (2) the improvement from MP=0 to MP=0.5 is significantly greater than that from MP=0.5 to MP=1.0; (3) given a non-zero MP, the speed drop is less significant in the part of the congested region closer to  $\rho_c$  than that of the further part. I generalize this trend as a proposal of a 3-dimensional density-speed relationship for future use of measuring traffic conditions and estimating FD in a mix-flow traffic with several parameters (Figure 2.14 and 2.15). Different MPs render different  $\rho_c$ , though the specific value depends on the driving rule and driving technology of the AVs. Traffic stability generally improves with the increased MP. Although autonomous vehicles have the potential to increase the maximum speed, the specific impact largely depends on regulation and public acceptance on the vehicle control parameters.



**Figure 2.14** Hypothetical  $v_c$  region in  $MP$ - $\rho_c$  space and hypothetical  $v$ - $\rho_c$  relationship based on the experimental results.



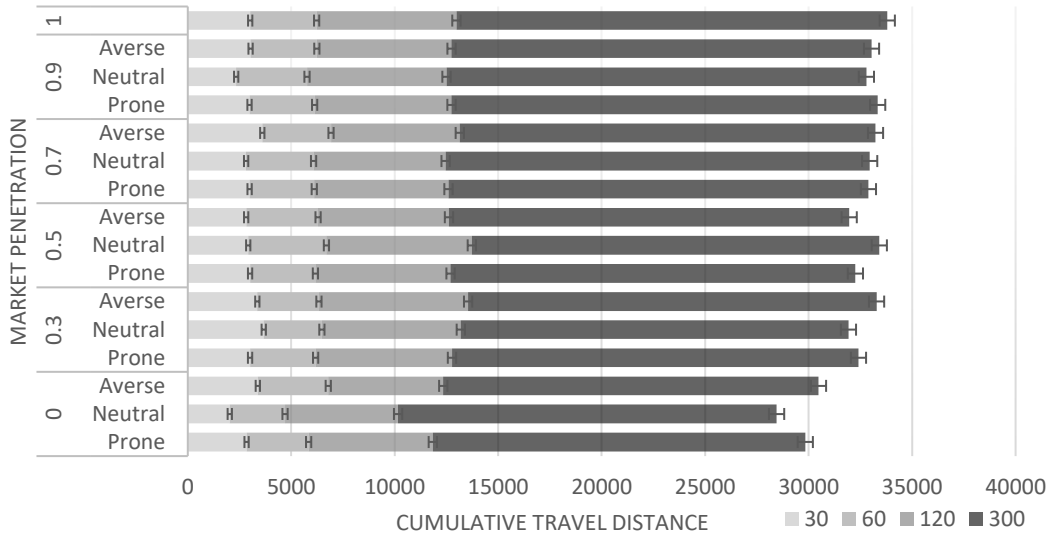
**Figure 2.15** Hypothetical MP- $v$ - $\rho$  relationship based on the experimental results.

Table 2.3 presents the travel distance (as a reflection of mobility impact) for different MP with heterogeneous risk-preference among human agents. Figure 2.16 shows the cumulative distance of total agent travel distance for different MP rates and human agents' risk preferences. Under this setup, scenarios with a higher percentage of human agents have a higher vehicle mile traveled (VMT) in the first 30-60 sec but, in a longer time horizon, AVs prevail. The risk preference has notable impact but the direction of impact replies on the specific traffic condition and MP.

**Table 2.3** Average travel distance (mobility) in 30, 60, 120, 300sec with different risk-preference ( $\alpha$ ) for human agents.

MP	$\alpha$ distribution			Travel distance marginal gain (ft)				Cumulative Travel Distance (ft)			
	.9	0	1.2	30sec	60sec	120sec	300sec	30sec	60sec	120sec	300sec
0	.1	.3	.6	2841.90	2998.74	6000.03	17996.90	2841.90	5840.64	11840.67	29837.57
	.6	.3	.1	2029.24	2670.08	5462.95	18290.07	2425.00	6087.09	12270.64	29726.91
	.2	.6	.2	3385.03	3404.08	5555.84	18132.29	2029.30	5378.33	10455.67	30088.87
.3	.1	.3	.6	3009.89	3167.45	6587.97	19642.93	3009.89	6177.34	12765.31	32408.24
	.6	.3	.1	3670.97	2809.51	6713.06	18730.95	3225.81	5333.37	12682.78	31668.50
	.2	.6	.2	3361.18	2975.87	7208.92	19737.65	3010.58	6452.00	12142.12	32360.85
.5	.1	.3	.6	3010.82	3156.42	6530.29	19563.58	3010.82	6167.24	12697.53	32261.11
	.6	.3	.1	2921.58	3778.35	7025.43	19689.42	2845.24	6632.97	12364.90	33048.45
	.2	.6	.2	2821.66	3474.96	6324.79	19339.24	2300.98	6868.57	12446.92	31771.62
.7	.1	.3	.6	2988.02	3124.24	6485.37	20292.39	2988.02	6112.26	12597.63	32890.02
	.6	.3	.1	2819.54	3259.12	6380.63	20491.17	2422.40	6034.19	12115.19	32447.75
	.2	.6	.2	3601.65	3320.33	6222.52	20088.01	3330.38	5489.78	12488.59	33100.22
.9	.1	.3	.6	2984.32	3146.42	6612.36	20593.26	2984.32	6130.74	12743.10	33336.36
	.6	.3	.1	2339.41	3422.56	6739.54	20289.60	3410.40	6200.22	12829.80	33581.62
	.2	.6	.2	3034.90	3202.70	6508.64	20292.75	2866.59	6179.64	13059.86	33243.79
1.0	-	-	-	3016.85	3209.62	6753.75	20819.98	3016.85	6226.47	12980.22	33800.20





**Figure 2.16** Market penetration and mobility in different market penetration rates (300 sec horizon) with different dominant risk preference.

### Highlighted Features

This section demonstrates three of the important capabilities of CognAgent, which differentiates itself from the conventional behavioral agent-based models. A model built on the proposed framework “collapses” into more behavioral agent-based model given certain assumptions on agent’s sensation, perception, and decision rules. In addition to the advantages of modeling agents from cognition’s perspective (rather than revealed behavior’s), applying the module “Space of Observable” improves computational efforts and post-simulation analysis capability since information source, channel/medium, and recipients are clearly specified and organized in one single place. One thing worth emphasizing is that CognAgent models information as the change of uncertainty and modeling risk becomes a “side-product” of modeling information dynamics when specifications about agents’ risk preferences are identified.

*Considering autonomous vehicles from different manufacturers, sensors, and processing algorithms*

The agent-based model in the example can produce individual vehicle’s trajectories by different manufacturers. Let us take an impact study of manufacture/model/sensor type composition as an example: Type A (heavily relies on computer vision/camera), Type B (relies on radar/Lidar), and

Type C (puts equal decision weights (and risk-averse) on both types of equipment). Under foggy weather, Type A is assumed having 0.71 sensing distance than that of Type B. Because of the risk-aversion setting in Type C, it uses the radar under foggy weather to determine if any immediate deceleration (i.e., being conservative in unfavorable situation) from the leading vehicle. Table 1.4 compares the trajectories of a fleet of vehicles equipped with different sensor (manufacture/model) types following varies distributions. How the sensed information is processed (perceived) and utilized for making decisions (behaving) remain unchanged. Since autonomous vehicles do not over-react, there is not significant traffic shocking wave, and the vehicles tend to form platoon groups.

**Table 2.4** System Performance Comparison Given Different Vehicle Type Composition over Type A, Type B, and Type C Autonomous Vehicles (Random Seed 3) during the simulation 0 to 30,000 steps (0.01sec/step) and in the road stretch from 0 to 35,000 ft. x-axis is the distance and y-axis is simulation step in the plots of vehicle trajectories.

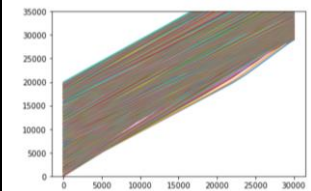
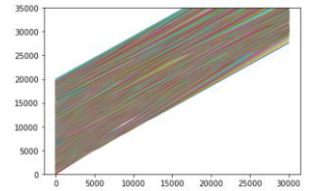
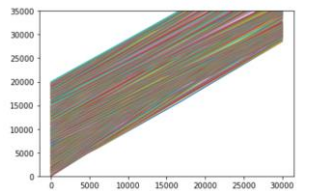
	(80%, 10%, 10%)	(10%, 80%, 10%)	(10%, 10%, 80%)
Trajectory Visualization			
VMT (ft.)	2.04e+4	2.07e+4	2.05e+4
VHT (sec.)	2.34e+2	2.43e+2	2.41e+2
VMT/VHT (ft./sec.)	84.7	85.1	84.9

Table 2.4 shows that, in the given foggy weather, Type A-dominated traffic produces lower VMT and lower average speed due to shorter sensing distance. Type B-dominated vehicle, on the other hand, produces the highest average speed.

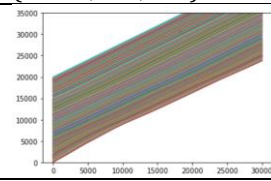
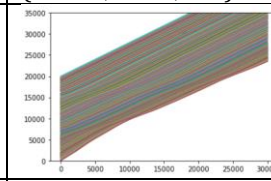
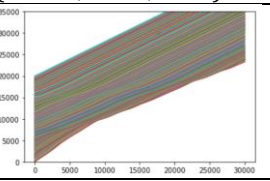
*Consideration of human agents with varying cognitive limitation for equity and ADA assessment*

Human’s cognitive limitation/bias/disability can be classified into two types: sensory and perceptive. Examples of sensory limitation/bias/disability are nearsightedness, cataract, hearing loss, while examples of perceptive limitation/bias/disability are short-term/working memory

dysfunction and varying levels of auditory and visual dyslexia. Modeling sensory dysfunction is handled by PISOO to determine whether to “push” certain information from the source to the agent. This section focuses on the perceptive limitation/bias/disability. The module Space of Observables “scans” the module Physical Interaction. The module Space of Observables then determines whether a certain condition can be sensed by each agent based on the medium (e.g., weather) and corresponding agent’s sensor capabilities. Since only perceiving capability is varied, there is no need to adjust any setting in the module Space of Observables.

Table 2.5 shows the trajectories of human-driving vehicle fleets. Note that this human driver is assumed having no sensing problem (i.e., same as other human drivers) and only having the perceptive challenge of understanding the suggested speed from the (dynamic) message board. This human agent only adjusts speed based on surrounding traffic situation and his/her driving preference.

**Table 2.5** The trajectories of the 400 simulated vehicles with and without severe dyslexia when he/she sense the message board’s speed recommendation (81 ft/sec (55 mph)) at the beginning of the simulation. In addition, 5% are randomly selected to miss the information

	(100%, 0%, 0%)	(92.5%, 0.5%, 7%)	(88.5%, 1.5%, 10%)
Trajectory Visualization			
VMT (ft.)	2.01e+4	1.99e+4	1.97e+4
VHT (sec.)	2.45e+2	2.46e+2	2.49e+2
VMT/VHT (ft./sec.)	82.0	80.9	79.1

### Feasibility and Advantage

This case study is conducted by treating the two seemingly different types of vehicles in the same object-oriented programming framework. Each agent, regardless of its type, is an instantiation of a parent class that includes attributes such as current state (location, speed, etc.) and methods of sensing, perceiving, and update. Sensing and perceiving are clearly separated to not only improve the potential of model sophistication (or resolution in terms of modeling cognitive functions) but

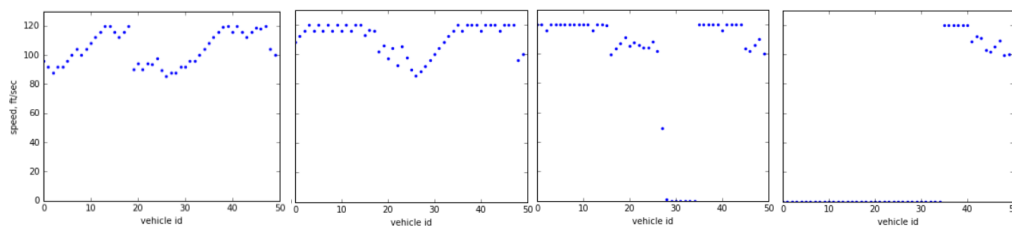
also model performance. The case study includes a simple scenario with 50 vehicles with a given random spacing and speed to test the impact of different market penetrations. A simulation results shows a nonlinear trend of less impact from shocking wave with the increase of AV market penetration – a shocking wave is nearly eliminated with when market penetration of AVs is around 30%. Intuitively, this can be punitively understood as that when 1 every 4 cars is AV, shocking wave that caused by human overreaction can be “cushioned” by AVs. Exploring further, a hypothetical 3-dimensional fundamental diagram (the 3rd dimension being market penetration) is proposed. This proposed fundamental diagram not only captures the increase of capacity and free flow speed in the “uncongested regime” but also the “cushion” benefits from increasing AVs market penetration that leads to less likelihood of speed drop with increase of traffic density.

### **Computational Efficiency**

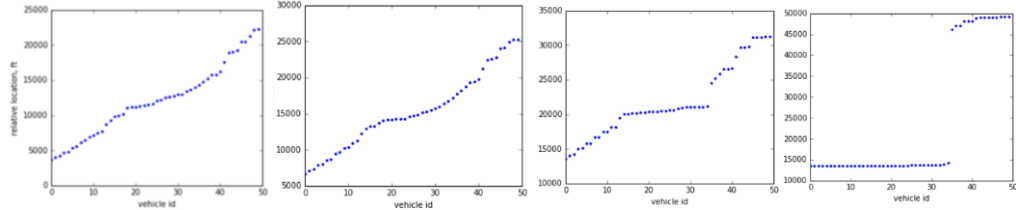
Since each vehicle only checks the state of the leading vehicle in the traditional car-following model, the computational benefit of having the module of space of observable is not significant. However, when agents can detect non-local information such as traffic radio or vehicle-to-vehicle communication, each agent “scans” the entire database during a simulation step without a manager (PISOO) to streamline the process. PISOO, the instantiation of the module Space of Observable, determines if a piece of information can be sensed, and this process iterates over every single agent during a time step. Though each agent may record, say, a deceleration event by itself, the same event was recorded multiple times by different agents that witnessed it. Another example is the foggy weather and limited cellular connection (for traffic information). Without PISOO, each agent has to either set the “detectable” range by itself or directly check the location, speed, and acceleration rate of the leading vehicle before determining if the information can be used for decision making. PISOO, however, “pushes” the information to agents and agents can only make decisions based on prior knowledge and newly obtained information “pushed” by PISOO.

Although it is conceptually clear how PISOO can streamline computation and data management, I demonstrate through a V2V example, in which the DSRC only serves as an emergency warning (without broadcasting from one vehicle to another). Suppose a scenario of a full market penetration, the 35<sup>th</sup> vehicle suddenly decelerates with 15ft/sec from 35<sup>th</sup> sec due to a mysterious reason. The vehicle sent a warning message to the vehicles within 1000ft behind (information propagation from one vehicle to another is not considered) as the DSRC coverage constraint (the U.S., Department of Transportation, 2016.) The space of observables for DSRC information is modeled as “continuous.” That is, each agent will inquiry the DSRC information layer and check if it is within the range of warning message if any. In a complex scenario with a significant amount of DSRC interactions and large network, it is recommended to set raster-based DSRC layer to reduce the computational effort. In addition to all the properties and methods in the previous object-oriented programming algorithms, physical interaction module stores method that gives two outputs -- whether there is DSRC warning at the current location and what the content is. The information from the layers of camera sensors, DSRC, and driver’s vision forms the sensed information as the input for perception.

Under this setting, the ACVs shows rapid response and safety improvement potential with DSRC application – the speed and location change over the time horizon is shown in Figure 2.17 and Figure 2.18, respectively. The DSRC warning is sent solely from the accident vehicle with 1000ft coverage. Allowing propagation but no overreaction of DSRC message could further meditate the risk.

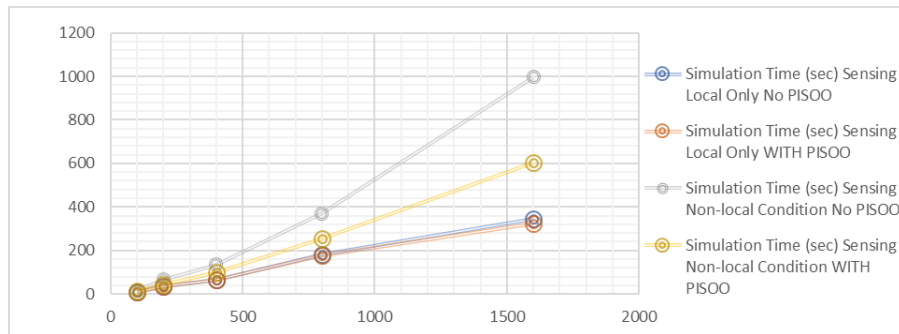


**Figure 2.17** Vehicle speed in the DSRC accident warning scenario (100% MP).  $t = 30s, 60s, 120s, 300s$



**Figure 2.18** Vehicle location in the DSRC accident warning scenario (100% MP).  $t = 30s, 60s, 120s, 300s$

Figure 2.19 shows the comparison of computational effort. When vehicle only senses the immediate adjacent (local) traffic, simulation time are similar, and PISOO does not show significant benefit. However, when each agent needs to process system-wide information (non-local), PISOO significantly improves the computational effort. The simulation tests are run on Computer configuration: Intel® Core™ i5-6200U CPU @ 2.30GHz 2.40 GHz, RAM 8.00GB (7.78 usable), 64-bit OS with a 64-based processor.



**Figure 2.19** Computational effort comparison among different modeling needs and numbers of agents

## Conclusion

The paper proposes a unifying framework to model and analyze complex transportation phenomena in a cognitively consistent manner. Comparing to models focusing on approximating revealed behaviors, CognAgent adopts a homeomorphic perspective and explicitly considers heterogeneous agents' sensation, perception, and decision making. The framework also enhances modeling flexibility and generality, simplifies programming practice, reduces data management efforts, and improves results' interpretability. A model guided by the framework is sensitive to agent's sensor type, information availability, and information processing methods, and therefore,

capable of considering broader policy scenarios such as equity issue among travelers with visual and auditory impairment.

Modeling uncertainty and modeling information are equivalent under CognAgent. The uncertainty for a certain percept (e.g., travel time) is the state of lacking the information on this percept, and information can be quantified by measuring the change of the perceived uncertainty. To model uncertainty and information, the source of information, the channel/medium, and the recipient needs be clearly identified. Any statement on uncertainty and information that does not clarify these three core components is meaningless. The distribution of category-based outcomes generated from the perception sub-module is fed into the decision-making sub-module to produce decision under uncertainty/risk. Heterogeneous agents interact with one another through behaving in the physical interaction module.

The numerical example demonstrates the framework's flexibility and computational efficiency and the potential of considering broader and more fine-grained environmental and policy scenarios.

## CHAPTER 3 QUANTIFYING INFORMATION AS CHANGE OF PERCEIVED UNCERTAINTY

*"Information is the resolution of uncertainty."*

*–Claude Shannon*

### ***Background and Literature Review***

A abstract statement such as “information is the resolution of uncertainty” can be explained through a concrete example. If one can differentiate all six possible outcomes of a dice, this person obtains more information than a person who can only differentiate odd (i.e., 1, 3, 5) from even outcomes (i.e., 2, 4, 6). Another way to put it is that a person who expects one of the six equally-likely outcomes would obtain more information from a rolling dice than a person who expects one of the two equally likely outcomes from the same rolling dice. In effect, information can be quantified as the change in uncertainty perceived by an observer and therefore can share the same unit of measurements, such as bit, nat, and qubit. Indeed, uncertainty can be quantified as the amount of information needed to reduce uncertainty to zero. A common example includes that one needs 1 bit of information to know the result of a flipping-coin experiment when the prior is non-informative (i.e., fair game).

In general economics and transportation economics, uncertainty and information are commonly modeled through their revealed effect. For example, expected utility theory and prospect theory are two classic methods to describe decision under risk or uncertainty. Chorus et al. (2013) present a formulation of information acquisition that uses search theory to evaluate the value of information and then proposed in a subsequent paper a discrete-choice model of traveler response to information. Levinson (2003) studies the value of advanced traveler information systems for route choice. Using other measures such as variance- and percentile-related methods to



approximate uncertainty and change of uncertainty are theoretically less sound but demonstrate strong practical adoption and empirical value.

Common information types in transportation modeling make implicit assumptions on information sources and transmission media such as mobile application, dynamic message boards, radio, and weather. Xuan and Kanafani (2014) compare various aggregate analysis methodologies for evaluating accident messages from changeable message signs (CMS). The phase of analyzing or modeling the cognition characteristics of the travelers is skipped, and the correlation between driver's diversion behavior and the types of messages are estimated. Xiong et al. (2015) go one step further and model traveler's en-route diversion behavior based on Bayesian rules, combining it with stated-preference (SP) driving simulator data and recalibrating it using Bluetooth-based field data. Gao et al. (2016) summarize the related study achievements from a survey conducted in the Beijing area with a specially designed questionnaire that takes into account the services of traffic information provision. Kopitch and Saphores (2011) study whether Changeable Message Signs about congestion, incidents, roadway zones, speed limits, and Amber alert would influence the number of secondary incidents. Chorus et al. (2009) present a formal model of travelers' compliance with "personalized" and "non-personalized" advice that is based on predicted travel time. Chorus et al. (2007) propose a theoretical model of travel information by integrating notions of Bayesian update and a regret-based framework of travel choice for studying the use and effects of transit information among car drivers. Gao et al. (2011) explicitly considered the information acquisition constrained by "cognitive cost." Bruijin and Leijten (2007) proposed the concept of contested knowledge and its effect on traveler decision making. Jha et al. (1988) adopt a Bayesian perception update model, for day-to-day travel choice dynamics to incorporate information provision, but this approach lacks theoretical support from cognitive science since it essentially approximates perceived travel time uncertainty through travel time variation. Mitsakis et al. (2015) propose an integrated framework that incorporates real-time ATIS in a large-scale dynamic traffic

assignment model and applies to the network of Thessaloniki, Greece. In addition to conventional Bayesian-based and Markov Chain information update, a quantum-based approach has been also discussed. For example, Asano et al. (2012) update probability based on linear algebra, the von Neumann–Lüders projection postulate, Born’s rule, and the quantum representation of the state space of a composite system by the tensor product to generalize the classical Bayesian inference.

The proposed paradigm is consistent with the current consensus in information theory, communication, and cognitive science. Under a strict definition of uncertainty proposed and proved by Claude Shannon (1949), three fundamental components of any constructive discussion on information have to be explicitly identified – source, medium/channel, and recipient. When a recipient is a human, his or her sensation and perception are critical in understanding how much he or she learns from the newly observed information. Norwich (1993) studies how information affects sensed and perceived uncertainty within a consistent framework. In this chapter, I quantify the information that effectively changed the perceived travel time distribution, not the amount of information received by sense organs.

In practice, the proposed paradigm can be first applied to incorporate information provision scheme in travel demand modeling. Then the result can be an input for evaluating advanced traffic information and management systems (Jayarkshnan, 1993) and incorporating uncertainty and reliability into traffic simulation and transportation planning models (Liu, 2004). Similar principles can be extended to measure system uncertainty when observers are entities such as system operators, planners, and decision-makers. It is worth emphasizing that uncertainty measure is a subjective *ex ante* concept and may or may not be “accurate.” Ben-Elia, et al. (2013) investigate the impact of various types of traveler information accuracy on route choice via designed experiment and show valuable insights, though the difference between actual objective information accuracy (objective) and the perceived subjective information accuracy was not entirely clarified.

The rest of the paper presents the general methodology, which is followed by a numerical example that quantifies information provision to evaluate the effectiveness of various alternatives of a dynamic message board.

## **Methodology**

Suppose a traveler  $n$  is in mental state  $\Phi_{t,l}^n$  right before perceives travel time information at time  $t$  at physical location  $l$ .  $\Phi_{t,l}^n$  is an abstract construct and contains all the necessary information relevant to the study. Supposing that a piece of information is sensed, the traveler updates the sensed distribution and constructs the perceived distribution correspondingly. The category settings are flexible, but here it is postulated to be related to preferred arrival time (PAT), traveler's characteristics, frequency, and trip purpose. There are multiple methods to update perceived information and the decision of which method to use depends on the assumptions of the modelers and data availability. Once the new perceived distribution is formed, information entropy can be calculated and compared with the prior information entropy. In this chapter, we use Shannon's definition, though Von Neuman's quantum information theory may be more general by not assuming there is no "interference" among each cognitive category/group. The amount of effective information provided by the traveler information systems is the change of the information entropy. Using Shannon's information entropy definition, we have,

$$Q(\Theta|\Phi_{t,l_t}^n) = \Delta H_{t+\tau,l_{t+\tau}}^n = H_{t+\tau,l_{t+\tau}}^n - H_{t,l_t}^n = \sum_{j \in J|\Phi_{t+\tau,l_{t+\tau}}^n} p_{j,t+\tau}^n \log_b \frac{1}{p_{j,t+\tau}^n} - \sum_{j \in J|\Phi_{t,l_t}^n} p_{j,t}^n \log_b \frac{1}{p_{j,t}^n} \quad (3.1)$$

where,  $Q(\cdot)$  is a function quantifying information,  $\Theta$ , obtained by observer  $n$  and is calculated by measuring the change of information entropy with a base of  $b$  (when  $b = 2$ , the unit would be Bit).  $\Phi_{t,l_t}^n$  and  $\Phi_{t+\tau,l_{t+\tau}}^n$  represent the observer's mental state before and after the observation. Since the location of  $n$  might change after  $\tau$ , the location is denoted as a function of time. Perceived category set  $J$  is determined based on the perceptive categorization assumption, given the observer's mental

state. In this chapter, category set is set unchanged to simplify the study but the formulation can be flexible when  $J$  is a function of  $\Theta$  and  $\Phi_{t,t}^n$ . That is, the newly observable information given the mental state when the observer is perceiving  $\Theta$  might change the category set. It is possible that the quantified information is positive, unchanged, or negative (when the observer becomes even more confused), depending on the combined effect of prior information and the information obtained. Here the external information transmission channel/medium is not discussed for simplification purpose. However, the common effects of noisy channel/medium are the longer sight distance and misperception such as mistakenly reading 17 min as 12 min.

Some might notice that using the change of the uncertainty to quantify information might render zero as long as the perceived distribution is updated “symmetrically.” This can be responded in two points. First, “effective information” refers to the information that causes the perceived uncertainty to change. When the prior entropy and the posterior entropy are the same, it really implies that the traveler’s perceived uncertainty level is unchanged. For example, before receiving the information, a traveler might think a travel time distribution of a certain route to be {0.8: 16min; 0.2: 25 min} while after receiving the information, the traveler’s becomes {0.8: 16min; 0.2: 25 min}. In both cases, the traveler has low uncertainty, though in the first distribution the traveler is quite certain he/she will arrive in 16 min while in the second distribution the traveler is quite certain he/she will arrive in 25 min. Second, dynamically, when the interval is small, the “process” of the transition is captured during the process and the reason the overall change of uncertainty is zero is that the traveler first experiences a reduction of uncertainty but later he/she receives conflicting information that causes confusion (and therefore, increased uncertainty), and vice versa. The two distributions, though same in entropy, may differ in expected value, which can be captured by various diversion measures such as Kullback-Leibler divergence.

Why is it so important to incorporate grouping effect of human perception in the proposed method? In other words, why it is emphasized that only a limited number of discrete events

*simultaneously* can be perceived due to limited brain capacity? Let us express the entropy formulation in the terms of probability density as follows.

$$-\sum_j p(x_j)\Delta x \log(p(x_j)\Delta x) = -\sum_j p(x_j) \log p(x_j) - \left(\sum_j p(x_j)\Delta x\right) \log \Delta x \quad (3.2)$$

where we use  $x$  to represent one possible outcome. But now, when we limit the number of perceived outcomes, we have

$$H = -\int_{\mathbb{X}} p(x) \log p(x) dx - \lim_{\Delta x \rightarrow 0} \log \Delta x \quad (3.3)$$

where  $p(x)$  is the density distribution of  $x \in \mathbb{X}$  when an observer is able to perceive infinitely small differences. The second term makes the whole formulation approaching infinite, which violates the limited capacity of perception. Therefore, it is necessary to transform travel time into categories such as “arriving late,” “arriving on-time,” relative to the preferred arrival time (PAT).

Which scheme of information update (learning) to use has significant influence on the overall results of the quantified information. The prior distribution can be obtained by historical information or a general consensus from the modeling experts about travelers’ perceptive characteristics. The posterior distribution is obtained based on certain learning mechanisms given each observer’s mental state. Comparing different schemes of information update is out of the scope of this dissertation. But it is important to point out the two important categories of the schemes that are widely used in practice – one is based on weighted experience; the other is Bayesian. The two methods are essentially equivalent (i.e., information can be treated as weighted experience or “fictitious previous observations”) as shown by Poirier (1995), but they in practice vary greatly in parameters, computer coding, and computational efficiency/feasibility.

The first type estimates the distribution by attaching various weights to historical experiences and information. That is, experiences are treated as data points with various weights in the formation, shown below,

$$\psi_{t+\tau}^n(TT) = \psi_{t+\tau}^n(TT | \mathbf{tt}_t^n, \psi_t^n(TT)) \quad (3.4)$$

where,  $\psi_{t+\tau}^n(\cdot)$  is the perceived travel time distribution by  $n$  at time  $t + \tau$ .  $\mathbf{tt}_t^n$  is the newly perceived travel time, an instantiation of  $\Theta$  in this example.  $\mathbf{tt}_t^n$  contains link-based travel time perceived by  $n$ .

For a Bayesian update, the posterior can be formulated as below:

$$\psi_{t+\tau}^n(TT) \propto \psi_t^n(TT) \cdot f_t^n(tt) \quad (3.5)$$

where  $f_t^n(tt)$  is the perceived distribution of travel time by  $n$  without associating or comparing with ex-ante travel time distribution. Traveler information updates traveler's prior travel time distribution and this process can be justified, in a way, by treating the information as an equivalent "fictitious sample" and through the "likelihood principle" that the information in the "samples" is stored in the likelihood function. Here  $TT$  is just a general travel time and it can be segment-level, route-level, or network-level, depending on the specific context.

By assuming the unimodal perceived travel time distribution, a point estimate can be obtained by searching for the "peak" of the posterior distribution. This can be formulated as

$$\operatorname{argmax}_{tt_1, tt_2, \dots, tt_L} \psi_t^n(tt_1, tt_2, \dots, tt_L) \quad (3.6)$$

Subject to:

$$tt_{m_1} + tt_{m_2} + \dots + tt_{m_k} + \epsilon_n = \theta_m, \forall \theta_m \in \Theta \quad (3.7)$$

$$\int_0^{+\infty} \psi_t^n(TT) dTT = 1 \quad (3.8)$$

Here,  $\epsilon_n$  is a random variable and depends on the perceptive characteristics of the observer. If an observer fully believes the information,  $\epsilon_n = 0$ , or else,  $\epsilon$  can be a random variable following a given distribution. Link  $m_2, \dots, m_k$  are  $k$  links whose travel time information get updated using the inference from the provided information  $\Theta$ . The inferred link travel time forms route  $m$ 's travel time  $\theta_m$ . Equation (8) says to guaranty that the perceived travel time distribution is normalizable.

It is possible that travelers/shippers use not only historical information but also predicted information. Modelers can assume that the traveler's predicted distribution is a map of historical

information, currently given traveler information, and some characteristics of the traveler him/herself. In the following example, a non-collaborative traffic network is used where individuals do not share information with one another nor predict travel time – the only two influencers on the perceived travel time distribution is the historical information and newly perceived information.

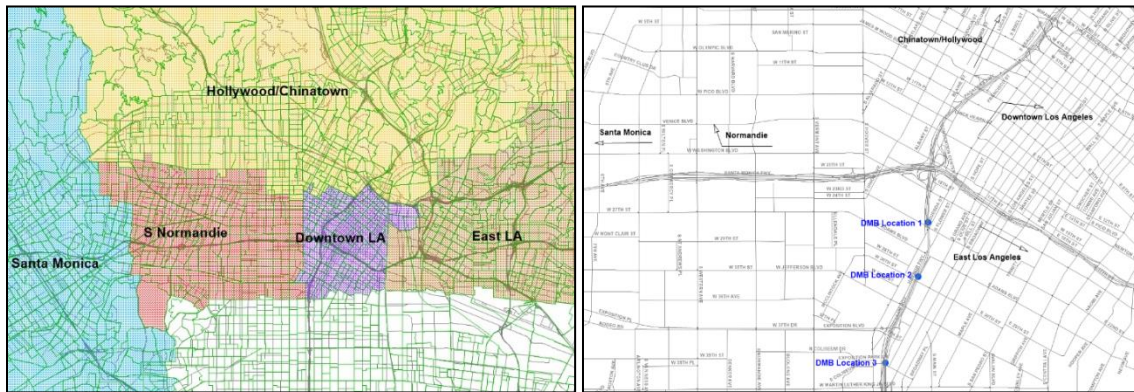
### ***Numerical Example***

Suppose that we are interested in the amount of information that a proposed dynamic message board (DMB) can provide. This proposed project has 12 alternatives (3 variations on location, shown in Figure 3.1, and 4 content variations) on I-110 northbound near-downtown Los Angeles, California. Users are classified into six types to be consistent with the Southern California Association of Governments' 2016 RTP model Year 2016 Scenario 3 Setting 7 -- drive alone (DA), 2-person occupancy-vehicle (HOV2), 3-r-more-person occupancy vehicle (HOV3+), light-duty truck (LT), medium-duty truck (MT), and heavy-duty truck (HT). Also identified are the information on trip destinations classified into 6 user classes – Santa Monica, Normandie, Hollywood/Chinatown, Downtown Los Angeles, and East Los Angeles, shown in Figure 3.1. Preferred arrival times (PATs) are used for setting categories and assumed to be 1.2 of the duration of historical average travel. A sounder estimation can be obtained departure time choice or activity-based model, if available.

Listed below are some fictitious instances:

- Traveler #1: DA, going downtown, 3-category perception (arriving on-time, late, early), regular commuter
- Traveler #2: HOV2, going downtown, 2-category perception (arriving before PAT, arriving after PAT), irregular commuter
- Traveler #3: DA, going to Hollywood, 2-category, first-time user
- Traveler #4: MT, going to East Los Angeles, 3-category, regular commuter

The idea of quantifying network-wide information provision comes from measuring the amount of information received by each individual that belongs to his/her user class and aggregating individuals based on the known composition of user classes. The demand data is obtained from SCAG to obtain the vehicle type and trip direction split and use PeMS data to obtain the dynamic profile of the traffic which is used to distribute the hourly traffic flow into 5-min intervals at the study location. The observed dynamic link speeds are also obtained PeMS. We are interested in measuring the effectiveness of information provision in different scenarios regarding the content and location. The sight distance is unchanged in the candidate locations.



**Figure 3.1** Destination Groups in Subarea Analysis (Left) and the study network (right). The blue nodes on the left indicates the candidate location of the proposed DMB.

The destination distributions and traffic composition of the 6 user classes are obtained from select/critical link analysis on SCAG RTP16 model Year 2016 Scenario 3 setting 7. The results for AM peak are shown in Table 3.1.

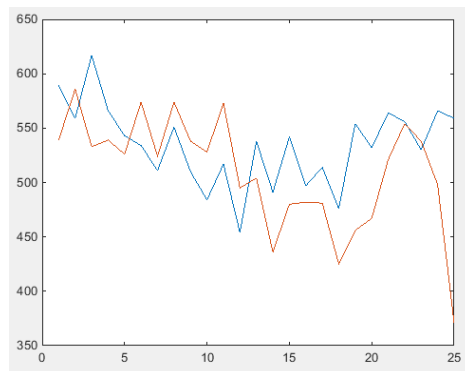
**Table 3.1** Trip destination split percentage at the study location

AM Peak	Santa Monica (D1)	Normandie Ave (D2)	Hollywood (D3)	Downtown LA (D4)	East LA (D5)
DA	0.2%	5.3%	56.2%	22.7%	15.6%
HOV2	0.6%	5.9%	56.4%	17.2%	19.9%
HOV3+	0.5%	10.2%	53.7%	21.4%	14.2%
LT	1.0%	8.9%	57.8%	18.2%	14.2%
MT	1.1%	8.9%	57.4%	19.9%	12.8%
HT	0.8%	9.9%	58.8%	18.2%	12.4%
Total	0.3%	6.4%	55.8%	21.8%	15.7%



The total of 27,018 trips are shown from the select link analysis and the split factor of the six user classes are 68.31%, 9.67%, 19.05%, 0.84%, 0.68%, and 1.45%, respectively. The destination split is used as the initial solution for the calibration process in TransModeler. This destination split is further adjusted along with 5-min interval to fit the PeMS count data.

The dynamic traffic counts on workdays of March 1-28, 2017, 7:00-8:00 am was collected from detector stations of the Performance Evaluation and Measurement System, PeMS. Loop detector's data was only used when reliability indicator was over 80%. The removed "unhealthy" data was linearly interpolated using the adjacent days. March 29 (Wednesday) is the study day that the DMB information provision strategy is tested on. Figure 3.2 shows an example of the fluctuation of the aggregate 5-min interval flow in a selected period. Due to lack of further temporal details of the user class and destination split, the splitting factors in Table 3.1 are used for all the study time intervals.



**Figure 3.2** Dynamic Flow Rate Sample (veh/5-min) from 6:30am – 8:30am, Monday March 9, 2017 (Red) and Thursday March 23, 2017 (Blue)

Based on the methodology described in the previous section, four concrete steps to measure uncertainty were developed as: (1) specify the study period, user type (traffic composition), and DMB content and location; (2) calculate entropy at time step  $t$  prior to perceiving new travel time information at that time step; (3) update perceived travel time distribution based on the prior distribution and the newly perceived information. (4) calculating change of uncertainty for each user and aggregate.

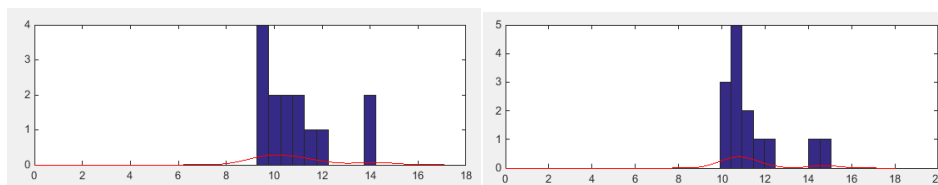
Fitted joint distribution for each user is developed based on historical link travel times. Here I assume a kernel distribution conditioned on non-negativity. When assuming unimodal, a point estimate is approximated using a modified line search. Note that the estimate is selected based on the maximum value of probability density rather than the expected value to be consistent with typical Bayesian point estimation approach. For demonstration, information update follows a Bayesian scheme, though other methods can also be incorporated. There was no particular cognitive consideration using this method. Note that when the provided information is not direct, only the relevant information is updated. For example, when a traveler going to Santa Monica direct perceives the travel time information about the travel time to downtown, the information is conditioning only on the relevant links with the link travel time correlation considered. Perceived correlation is considered in the update.

Below is an example for three travelers from 3 different user classes (DA, HOV2, MT), who arrive at the study location within the same time interval, with the given information of 12 min and 14 min for route 1 (to downtown LA). The Drive-Alone (DA) class of drivers are assumed to commute every Monday-Thursday, the HOV2 drivers are assumed to commute every Monday-Wednesday, and each DMB message is set equivalent to 1.5 experiences in information update, and the MT is assumed to commute every Monday-Friday.

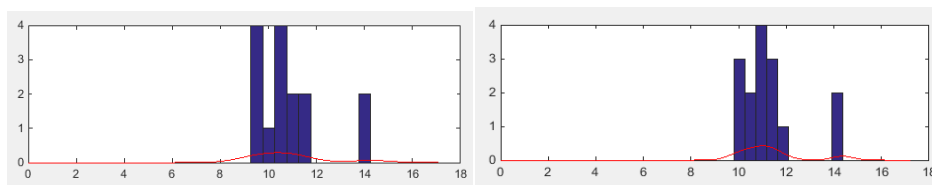
Again, the user classes are set without considering the trip/activity purpose even though it is convenient to incorporate it when such data becomes available. The study can be extended in the future to study the possibility that trucks set the categories based on factors such as fleet operators' preference, business and service types, nature and value of the freight inventory if appropriate freight demand model can be obtained. The categorization in this examples uses  $\alpha^n \cdot \bar{t}t$ , where  $\alpha^n = 1.2, \forall n$  is a nonnegative coefficient assigned to traveler  $n$  which is typically assumed to be greater or equal to 1.

Travelers might also “opt out,” ignore, or simply miss the information. This can be considered in the information update procedure, by either converting information into a weighted equivalent experience (when weight is 0, traveler/truck driver does not consider this information) or by utilizing models and methods from cognitive science. Here the information is assumed simple enough that the observers can fully understand it with no mistakes or misperception (i.e., no cognitive cost). For modeling phenomenon of noisy channel/medium (e.g., foggy weather or visual/hearing impairment), one can treat a constant travel time information as a random variable with stochastic error term or refer to formal cognitive models in this issue.

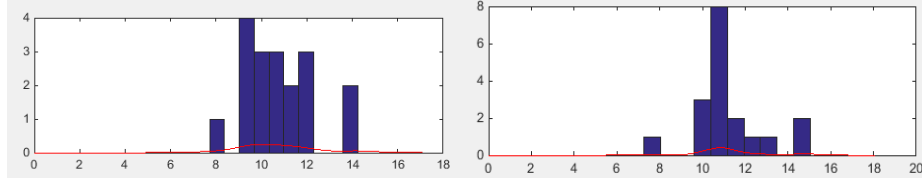
Figure 3.3, 3.4, and 3.5 show examples of the empirical travel time distributions from the freeway observations (blue histogram) at a segment near the proposed DMB location and the fitted distributions. The fitted distributions are used as perceived prior distribution for travelers. The x-axis shows travel times and the y-axis shows the frequency (count). The distributions for a selected time interval are also shown. Figure 3.6 shows the dynamic profile of four selected user class segmented based on their destination in the study period when the DMB shows travel time to downtown.



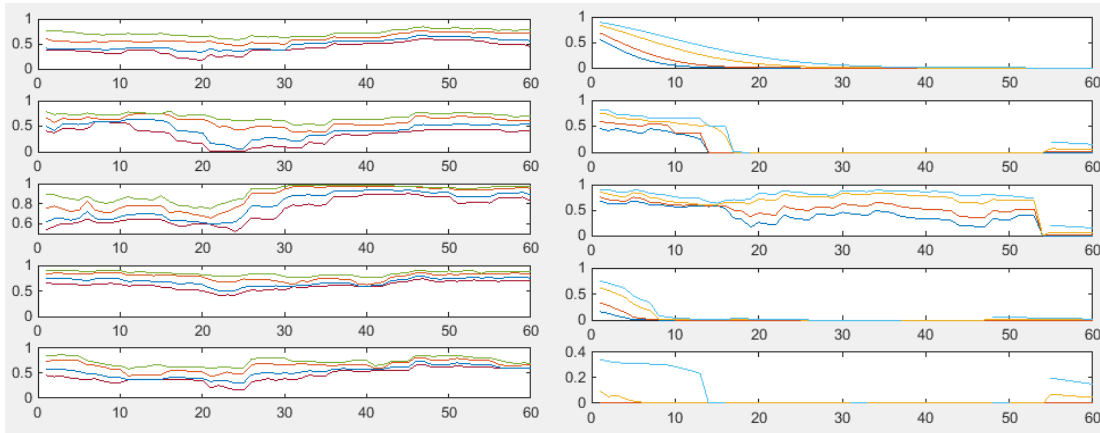
**Figure 3.3** The histograms show Santa Monica Direction (Drive-Alone (DA)) before and after the information provision. The red line is the fitted kernel distribution based on experience and newly provided information.



**Figure 3.4** The histograms show 2-Passeger High Occupancy Vehicles (HOV2) before and after the information provision. The red line is the fitted kernel distribution based on experience and newly provided information.



**Figure 3.5** The histograms show Median-Duty Truck (MT). The red line is the fitted kernel distribution based on experience and newly provided information.



**Figure 3.6** Comparing Information Entropy (unit: bits) without (left) and with (right) providing Information through Dynamic Message Board (5-minute Interval). The five rows are for the traffic to the five destinations (D1-D5). The four selected user classes are: DA (Green), HOV2 (Orange), HOV3 (Blue), and MT (Red).

Through visualization, we can have a general understanding of the effectiveness of the DMV content on different user classes dynamically. For example, the content has low effectiveness around 7:20-7:50 for travelers going to Chinatown/Hollywood direction. This seems counterintuitive because routes to Santa Monica or Normandie direction share fewer links with the route to downtown LA. One explanation is that the perceived travel time on rest of the links of these two directions is more correlated with the travel time to downtown than with travel times on links to Chinatown/Hollywood direction. Therefore, travel time information to downtown provides more information to Santa Monica and Normandie direction than to Chinatown/Hollywood direction.

By calculating the change of uncertainty for users in different classes, the amount of information provided by the dynamic message board can be calculated. Table 3.2 shows the change of information entropy for each user class and each time interval on the study day. The content

shows the predicted travel time from the location of the DMB to downtown Los Angeles. The total information provision is calculated over all six user classes.

**Table 3.2** Change of Information Entropy (Uncertainty) for 15-min time interval and each user class classified by vehicle type and destination (D1-D5).

		Time Interval $\tau_{1-14}$	Time Interval $\tau_{15-29}$	Time Interval $\tau_{30-44}$	Time Interval $\tau_{45-59}$	Total Information (Bits)
D1	DA	-0.1885	0.2974	0.3894	0.5884	6211.983
	HOV2	-0.2706	0.3916	0.4946	0.6584	1333.63
	HOV3+	-0.2333	0.4058	0.5826	0.7624	1883.496
	LT	-0.1323	0.3474	0.6407	0.8319	148.4674
	MT	-0.1323	0.3474	0.6047	0.8319	169.942
	HT	-0.1323	0.3474	0.6047	0.8319	685.6373
D2	DA	-0.0476	0.3650	0.0686	0.4189	6000.786
	HOV2	-0.0919	0.5673	0.2024	0.5301	1656.712
	HOV3+	-0.0778	0.3966	0.3836	0.6977	2293.446
	LT	-0.0296	0.2976	0.0543	0.0021	36.654
	MT	-0.0296	0.2976	0.0323	0.0075	40.51377
	HT	-0.0296	0.2976	0.0127	0.0082	152.6086
D3	DA	-0.1885	0.2974	0.3894	0.5884	38533.89
	HOV2	-0.2706	0.3916	0.4946	0.6584	7440.07
	HOV3+	-0.2333	0.4058	0.5826	0.7624	9895.657
	LT	-0.1323	0.3474	0.6047	0.8319	763.6152
	MT	-0.1323	0.3474	0.6047	0.8319	893.9846
	HT	-0.1323	0.3474	0.6047	0.8319	3606.874
D4	DA	0.4976	0.5725	0.5791	0.7317	48556.28
	HOV2	0.4024	0.6786	0.5920	0.7822	9149.612
	HOV3+	0.2114	0.8139	0.6224	0.8518	10945.11
	LT	0.1487	0.8493	0.7608	0.0044	548.0299
	MT	0.1487	0.8493	0.7608	0.0105	643.7723
	HT	0.1487	0.8493	0.7608	0.0005	2582.939
D5	DA	0.4565	0.3765	0.3649	0.6625	86440.72
	HOV2	0.5786	0.4121	0.4942	0.7324	18940.81
	HOV3+	0.6407	0.5447	0.6707	0.8136	26933.74
	LT	0.5108	0.6191	0.7340	0.0021	1336.483
	MT	0.5018	0.6191	0.7340	0.0072	1561.039
	HT	0.5018	0.6191	0.7340	0.0003	6275.593
Total Information (Bits)		40246	86073	72538	96805	295662.1

Table 3.2 shows that the DMB provides a total of 29566.21 Bits (around 288.73KB) between 7-8am. Information about the non-recurrent congestion is a positive contributor to the reduction of the perceived uncertainty. These are considered an advantage of the methodology used in this

chapter because the information provision considers observers' perception rather than only the traffic conditions. It is worth noticing that traveler might perceive more uncertainty, though the updated distribution (posterior) might be better reflecting the actual traffic situation. In another word, traveler information could raise a traveler's perceived uncertainty, though an updated (more "flat") distribution reflects better the actual traffic conditions. For example, when a traveler thinks a certain route will take him 5 minutes for certain, a piece of information showing a prediction of 10 minutes may cause the traveler feels confused if he/she puts similar weights on his belief and the new formation. This situation could increase the traveler's perceived uncertainty, though new perceived travel time distribution might reflect the actual travel time better. Separating the role of information as the change of uncertainty from the concepts such as information accuracy and reliability provides potentials to explicitly incorporate criteria in existing decision support systems without overlapping consideration. Also, since the distribution is updated differently in different user classes, different degrees of misperception among users can also be considered.

The result of installing DMB at the Location 2 has already shown previously. Now, let's test different content alternatives on all the candidate locations. Four candidate contents are:

1. Travel time to Downtown Los Angeles
2. Travel time to Santa Monica
3. Travel time to Downtown Los Angeles from 7:00-7:29 and travel time to Santa Monica from 7:30-7:59
4. Travel time to Santa Monica from 7:00-7:29 and travel time to Downtown Los Angeles from 7:30-7:59

Table 3.3 summarizes information provision from the DMB installed at one of three candidate locations identified previously.

**Table 3.3** Summary of effective information provision (Kilobytes) for different DMB installation locations and different strategies.

DMB Location	Content 1	Content 2	Content 3	Content 4
Link 1	291.12	228.70	<b>307.01</b>	191.12
Link 2	288.73	225.12	303.97	189.66
Link 3	286.53	222.45	298.74	183.93

Though link 1, 2, and 3 are close, it is noticeable that earlier information (i.e., Location 3) provision tends to be more effective. This minor margin of advantage accumulates, and if work days have similar dynamic traffic composition and flow, the benefit could become significant. However, a tradeoff should be considered since the earlier the information provided the lower the accuracy it could be. Content 1 provides more information than content 2 might be caused by a higher proportion of travelers going downtown than going to Santa Monica direction. On the other hand, different contents result in significant change regarding the change of distribution. Two main factors are suspected to contribute to the outperformance of using content 3. First, most commuter trips to Santa Monica direction might have preferred arrival time being between 7:30 am and 9:00 am. Therefore, a higher proportion of travelers going to Santa Monica pass through the study link in the first half of the study hour. Second, commuters to downtown might have similar preferred arrival time between 7:30 am and 8:00 am. But since it only takes around 15-20 min to arrive downtown from the DMB location, a higher proportion of travelers going to downtown pass through the study link in the second half of the study hour.

The above further leads our interest to the role of trip purpose, trip frequency, and departure time in the effectiveness of information provision. When travelers are not familiar with the travel time on the selected route, his/her perceived travel time might be “flatter,” and therefore, any information provision might generate substantial uncertainty change, and hence, uncertainty reduction. Although travel frequency can be obtained in the future from an activity-based model (SCAG is scheduled to release around 2020), this chapter assumes all the travelers are regular commuters.

Due to the lack of information on PAT and related information on setting cognitive categories, I assume that a traveler's preferred arrival time is 1.2 times larger than the perceived average travel time at each location and time point. In this case study, I only set two categories, though it is possible to set a different number of categories for various users based on their cognitive characteristics (e.g., brain capacity, etc.), the purpose of the trip, etc. For example, for a senior, he/she might perceive fewer categories (especially for unimportant trips) while a mid-age for purpose of commute might perceive more. However, it should be noted that human has a limited cognitive capacity and can only *simultaneously* perceive limited number of cognitive categories or possible outcomes.

In addition to the cognitive category settings, parameters such as how users are classified, how historical data are used, and how trip purposes are defined could also have an influence on the final results and conclusions. A more comprehensive analysis is needed for more efficient/effective resources allocation to improve modeling quality. Analysis for the sensitivity of change of parameter settings would be helpful for enhancing the efficiency of project budget allocation.

The change of uncertainty captures how uncertain about travel time travelers perceive before and after the information provision. When traveler changes the travel time distribution from  $\{0.3, 0.7\}$  to  $\{0.7, 0.3\}$ , this method generates zero entropy since the traveler experiences no uncertainty change, though the information is provided to change the traveler's perceived average travel time. If analysts or decision makers are interested in capturing this part of the information, formulation (2.1) can be replaced by methods such as Kullback-Leibler divergence.

One step further, information might even be negative when the received information conflicts with the observable's previous belief and causes the observable to feel more confused. Also, information provision quantification procedure shows sensitivity to which destination's travel time is shown. For instance, there's a significant change of perceived uncertainty by the newly provided information for users with downtown as the destination (DMB shows travel time



information on traveling to downtown as well); however, does not vary much for users going to other destinations. This can be understood using conditional entropy, and the more different of two routes of going to different destinations, the less sharing information the DMB can provide.

Therefore, which destination's travel time to show as content makes a difference. It could be a valuable practice to change information type based on traffic composition. For example, suppose it is given that the majority of traffic flows is going to downtown in morning peak hour, the predictive travel time shown on DMB should be given with destination as downtown while at night, same location, the predictive travel time on DMB should be given with destination as Santa Monica, to maximize information provision. However, it is also worth noticing that maximizing information provision might favor certain segments of a market and, therefore, might raise equity concerns.

Although information update in this example follows a Bayesian method, more specific models that consider specific cognitive characteristics can be incorporated thanks to the individual-based information quantification process. For example, some observers might have sensing challenge (e.g., myopia and cataract) or perceptive challenge (e.g., dyslexia), and they will update information in different manners, if update at all.

## ***Conclusion***

This chapter proposes to use the change of uncertainty as information quantification measure. That is, information is quantified by the difference between the entropy before and after the information provision. This proposal applies to measure perceived information on travel time and can be extended to more general settings when three fundamental elements are specified: information source and content, noisy channel/medium, and sensation and perception of the information recipient. The proposed method is cognition-based and consistent with the information theory.

One major application is to study the network-wide impact of information provision and evaluate the effectiveness of various information provision strategies. In addition to demonstrating

the feasibility, the numerical example shows that the effectiveness of information provision is sensitive to the dynamic content as well as the location. Specifically, to which destination travel time information shows on a dynamic message board should be dependent on the destination distribution of the dynamic traffic composition. When most traffic in the first half hour is heading to destination A while the second half hour is heading to destination B, the content should be adjusted correspondingly. I also recognize that maximizing information provision is not the ultimate purpose and should only be used as one criterion in the decision-making process. When the objective of providing information is to, say, “nudge” the traffic pattern closer to system optimum or improve information equity, the criterion of selecting from alternatives would be different.

Studying the impact of information provision on mode and route decisions can be an immediate extension of the proposed paradigm. A further step can be quantifying the effects of information provision on the of activity schedules and other travelers’ perceived uncertainty under multi-medium channel. Also of interest are the equity impact of information provision on people with sensory or perceptive impairment. Other factors that worth exploring may include age, gender, mental health, and personality.

## CHAPTER 4 ELASTIC SURPRISE THEORY FOR DECISION UNDER RISK

*“My colleagues, they study artificial intelligence; me, I study natural stupidity.”*

*–Amos Tversky*

### ***Introduction***

Conventional methods have made significant accomplishments in predicting human preference. These methods usually tend to be behavior-based, meaning they highly depend on arithmetic operations to match the observables while emphasizing less on their cognitive foundation and implication. The aim of modeling behaviors under risk in a more fundamental level makes it necessary to develop a more homeomorphic approach while maintaining the connection with the conventional paramorphic approaches and their practical edge.

In this chapter, I follow the convention in decision science field and classify decision under risk as a type of decision under uncertainty where probabilities are given or known, following conventional terminology. Perceived risk is determined by the simultaneous existence of possible events perceived by an observer. Suppose that a traveler perceives linear utility function for a route’s travel time to be  $\alpha$  min with full confidence. The question now is: does the traveler perceive the same event under 80% confidence be bigger, equal, or smaller? Realizing the difference between the *ex ante* and *ex post* utility of a payoff event leads to our interest in measuring the gap, which, in this chapter, is explained by the Elastic Surprise (ES).

Major contributions of this chapter are as follows. First, the proposed method unifies conventional methods in modeling decision under risk by specifying their underlying cognitive assumptions, so that modelers would have a consistent guidance of which model to choose. The relations among some common methods such as the Expected Utility Theory (EUT), the Cumulative Prospect Theory (CPT), and the Mean-Variance Method (MV) are explained under one single

framework. Non-compensatory methods such as the Elimination-By-Aspects (Tversky, 1972) is also applicable when ES is considered as an attribute. Second, the proposed method directly connects with the Information Theory (IT). In fact, the utility of a prospect equates the corresponding perceived information entropy when an observer has no preference towards the perceived possible outcomes. Appendix C verifies and discuss that information entropy is a unique measure of uncertainty to satisfy three basic conditions – continuity, monotonicity, and compound equivalency if probabilities are represented in reals. For strict proof, see Shannon (1949). Third, that EUT generates paradoxes is in part explained by the indifferentiation between a possible outcome's utility under risk and utility under certainty, and a revised version (REUT) is proposed to quantify the discrepancy between these two utility functions. REUT, unlike the original EUT, avoids the paradox caused by rescaling and convexity dependency between utility assessed by the probability-equivalency approach and the certainty-equivalency approach. Fourth, the proposed differentiation between risk and misperception improves a model's interpretability. The paper proposes to deal with misperception and risk in a relatively separate manner. It is worth emphasizing that the proposed method is not creating additional issues. Instead, it only brings explicit some used-to-be implicit assumptions on observers' cognition, and the framework "collapses" into conventional methods when certain assumptions about cognition are made.

Decision under risk in decision science community commonly refers to a decision context where probabilities are given, while decision under uncertainty is a broader case where probability may or may not be given. However, semantically speaking, decision under risk should involve at least one unfavorable outcome. Indeed, when a decision involves many possible outcomes but all of them are perceived favorable, it is hardly a decision under risk (Fischhoff, et al., 1984). This chapter puts efforts on consistency with the conventional definition of risk and of uncertainty in decision science; however, due to a different definition on uncertainty in IT, the term, uncertainty, may also involve specific probabilities in a discussion related to entropy.

The next two sections review related literature and propose additional challenges using conventional methods. Then the key concept of this chapter, Elastic Surprise (ES), is defined. How ES is incorporated into a utility function is also discussed. The logarithmic form of ES is studied as a relatively independent section for its normative nature and the relationship with entropy. Following the discussion on the logarithmic form of ES is the analysis of the cognitive and behavioral implications from using certain ES functional forms. They are oriented to the applications and findings of using ES to study various existing models including EUT, MV method, and CPT. The numerical study provides a concrete example of incorporating ES into the existing method (CPT) on studying traveler route choice under risk.

### ***Literature Review***

A large body of literature has been from different fields has been proposing and assessing various EU and non-EU frameworks. Wakker (2010) provides a compressive survey on the connections among conventional methods and the evolution from EU to CPT. De Palma et al. (2008) provides a review on using conventional additive utility model and categorizes them into the EU and non-EU framework for further discussion in discrete choice in static and dynamic contexts for travel behavior analysis. Non-additive decision-making models include but not limited to elimination-by-aspects (e.g., Tversky, 1972) and decision tree/random forest method in machine learning (e.g., Myles et al., 2004).

EUT was initially proposed by Bernoulli in 1738 as a normative model and was developed by von Neumann and Morgenstern (1945) to model human decision under risk and uncertainty. Therefore, it comes as no surprise that when studied closely, challenges arise. Researchers such as Allais (1953a) and Ellsberg (1963) propose paradoxes and questioned its underlying axioms. Methods such as prospect theory (PT, Kahneman and Tversky, 1979), anticipated utility (Quiggin, 1982), and dual theory (Yaari, 1987) can be viewed as expected utility methods with subjective

value functions and subjective weights. One well-known issue is the violation of stochastic dominance when some researchers attempted to incorporate subjective probabilities/weights as a nonlinear transformation of given probabilities (Handa, 1977). Camerer (1989) tests several generalized utility theories and finds that prospect theory can explain most of the data through experiments. The CPT, proposed by Tversky and Kahneman in 1992, advances PT by introducing Quiggin's rank dependence (Quiggin, 1982) to resolve the issue of violating stochastic dominance while keeps concepts such as framing effect, loss aversion, and likelihood insensitivity.

Despite that experimental violations of betweenness axiom and nonlinearity in perceived probability are observed, many theorists are still reluctant to forsake them as a key behavioral foundation (Camerer and Ho, 1994). Although the rank-dependent weighting function resolves a theoretical problem of stochastic dominance, the weighting function is only improved to be sensitive to the rank rather than the continuous change of prospect(s) (Tversky and Kahneman, 1992). Becker and Sarin (1987) propose the Lottery Dependent Expected Utility (LDEU) through a specific parameter sensitive to a given lottery but not the specific location on that lottery. Daniels and Keller (1990) compare the LDEU and the EUT and found an improved prediction in choices among risky options but not so in using probability or certainty equivalent indifference judgments. Another direction is through studies on quantifying surprise for understanding the nonlinear characteristics of the weighting function. Atkinson (1957) proposes to capture hedonic intensity through a linear function of  $(1 - p)$ . Brandstätter et al. (2002) extend this construal as a surprise function and propose to account for the nonlinearity of the weighting function using the notion of elation and disappointment, and yet, it is not addressed how elation (disappointment) emerges in a choice with only unfavorable (favorable) outcomes.

A large body of evidence suggests the significant impact of risk and reliability in human decisions in transportation systems (e.g., Small et al., 2005; Brownstone and Small, 2005). This concern leads to efforts on capturing risk preference on decisions in transportation systems such as

route and departure time choice by travelers and shippers. Noland and Small (1995) examine the impact of travel time variability on departure time choice. Ramos et al. (2014) provide a review of the developments in EUT, PT, and Regret Theory (RT) for modeling traveler behavior under uncertain travel time. Chow et al. (2010) use a genetic algorithm to estimate CPT parameters for selection of high-occupancy-vehicle lane. Lo et al. (2005) use MV-based method to capture risk preference. Boyles et al. (2010) account for the impact of day-to-day travel time variation on route decision using the MV-based model in determining congestion pricing strategy.

### ***More on Scale and Convexity Paradox in Existing Methods***

The literature review in the previous section mentions some commonly recognized paradoxes and problems against existing methods, especially EUT. This section, more paradoxes, and challenges are proposed for existing methods, including the heavy dependence on the curvature of the value function and inseparability between misperception and risk aversion in the cumulative prospect theory (CPT).

First, there is a caveat in using the convexity of a utility function to identify risk preference. In EUT, risk aversion is equivalent to the concavity of the utility function ( $u'' < 0$ ). However, suppose that we conducted a survey with choices under certainty and estimated a convex utility function. Now, if we do another survey about decisions under risk, should we adjust this convex function that we estimated in the first place if the two surveys reveal different utility function? Which one is correct? Besides, measure unit matters in the convexity the utility function revealed by a survey. For instance, suppose we use two utility scales (one in  $\mathbb{T}$ , the other in  $\sqrt[3]{\mathbb{T}}$ ) for the payoff  $x$ . When using the first scale, we obtain  $u = x^2$ ; while using the other scale, we obtain  $u = x^{\frac{2}{3}}$ . One is strictly convex, and the other is strictly concave. Therefore, the convexity of utility function itself does not capture the essence of risk preference. CPT in this aspect clearly differentiates loss aversion and risk aversion.

Although methods such as CPT can be used to describe some of the paradoxes and approach issues such as framing effect and risk aversion, it made no difference than methods such as EUT in principle due to the indifferentiation between misperception and risk (it is noted that the risk loss is captured partially by concave value function and partially by the nonlinear cumulative weighting function). Suppose that a choice maker was framed in a way that the post-misperceived probabilities and reference point become that in the Allais paradox. The same issue arises as does in EUT.

Last, a common misunderstanding is that a choice maker based on the CPT overestimates small probabilities. This is not necessarily the case. Suppose a choice maker has a decision that has three possible outcomes: -50, -20, 30, relative to the reference 0, with the probability 0.20, 0.10, and 0.70. According to the CPT, the weights of the three outcomes would become 0.26, 0.07, and 0.53 base on the estimates by Tversky and Kahneman (1992). That is, the second outcome, though has small probability, is underestimated rather than overestimated. Situations like this suggest that either this overestimation bias is not universally the case, or there could be additional factors underneath the revealed weighting function. In this chapter, I in effect propose that there are two underlying factors that contribute to the revealed nonlinear decision weights – probability misperception and elastic surprise.

### ***Elastic Surprise (ES)***

In this section, I introduce a few formal notations and then introduce the concept of Elastic Surprise (ES). I denote state space as  $\mathcal{S}$  for states of nature where  $s_i \in \mathcal{S}$  is the  $i$ th state. All the elements within  $\mathcal{S}$  are perceived as exhaustive and mutually exclusive. A choice maker has no impact, whatsoever, on the probabilities of any states to occur. Any subsets of  $\mathcal{S}$  are called events, commonly notated as  $E_\alpha$ , where  $\alpha$  is the index of a specific event. When  $E_\alpha$  and  $E_\gamma$  are complementary,  $E_\gamma$  is equivalent to  $E_\alpha^c$  and  $E_\alpha$  is equivalent to  $E_\gamma^c$ . If  $E_\alpha \succ$  (or  $\succcurlyeq$ )  $E_\gamma$ , we say  $E_\alpha$  is



preferable (or weakly preferable) to  $E_\gamma$ . If  $E_\alpha <$  (or  $\leq$ )  $E_\gamma$ , We say  $E_\gamma$  is preferable (or weakly preferable) to  $E_\alpha$ .

A Prospect is a function that maps an event to an outcome (Wakker, 2011). A prospect represents a course of actions in the prospect set  $\Lambda$  in a decision context and maps all possible states to real numbers. Note that a prospect contains more meaning than a choice. For example, a person might choose an option A, but the prospect associated with option A not only contains the meaning of choosing A but also contains a correspondence of all possible outcomes. If prospect  $x \in \Lambda$  is presented as a tuple  $(x_1, x_2, \dots, x_i, \dots, x_I)$  and prospect  $y \in \Lambda$  is presented as  $(y_1, y_2, \dots, y_i, \dots, y_I)$ , where  $i \in \mathbb{I}$  is an index of a possible event in the set  $\mathbb{I}$ ,  $I = \text{Card}(\mathbb{I})$ , we can write preference  $(x_1, x_2, \dots, x_i, \dots, x_I) \succcurlyeq$  (or  $\succ$ )  $(y_1, y_2, \dots, y_i, \dots, y_I)$  as  $x \succcurlyeq$  (or  $\succ$ )  $y$ . When we want to explicitly clarify the associated events, we can write the tuple as  $(E_1: x_1, E_2: x_2, \dots, E_i: x_i, \dots, E_I: x_I)$ . The appendix I shows that, for multiple prospects under consideration, one can always find a number  $n$  such that all the prospects have the same  $n$  events set up by allowing empty sets. The function  $V(\cdot)$  maps a prospect to a real number so that the relative preference can be compared numerically among prospects.  $V(\cdot)$  is often called certainty equivalent (CE) under the EUT.

A lottery,  $L$ , is a complete set of feasible combination of probabilities of all possible outcome  $(p_1, \dots, p_j)$  with  $p_j \geq 0 \forall j \in \mathbb{J}$ ,  $J = \text{Card}(\mathbb{J})$ , and  $\sum_j p_j = 1$ , where  $p_j$  is presented probability of outcome  $j$  to occur. We use  $\phi_j$  to map  $p_j$  to the corresponding post-misperceived probability.  $\phi_j$  has specific cognitive interpretation and is different than the decision weight,  $\pi_j$ , used in the prospect theory.

Function  $u(x, E_i)$  maps a specific prospect  $x$ , event  $E_i$ , and  $E_i$ 's associated decision weight to a real number so that the numerical relationship among all possible combinations represents relative preferences.  $u(x, E_i)$  is different than  $x_i$  since  $x_i$  might not reflect the perceived value of the event  $i$  for the particular choice maker. For example,  $x_1 = \$10$  and  $x_2 = \$20$ , the choice maker does not necessarily value  $E_2$  twice as much as value  $E_1$  given  $x$ . This chapter discusses cases where the

probabilities of events are given or known by the choice maker, and therefore,  $u(x_i, E_i)$  and  $u(x_i, \phi_i)$  mean the same. When a choice maker puts decision weight on each possible outcome with their corresponding probabilities,  $u(x_i, \phi_i)$  equates  $u(x_i, p_i)$ . When  $\phi_i = 1, \phi_j = 0, \forall j \neq i$ , the decision is perceived to be made under certainty for event  $i$ . We keep the convention by naming this utility Bernoulli utility and denote it  $u(x_i, 1)$ , or simply  $u_i$ . As can see, the *ex ante* utility,  $u(x_i, \phi_i)$ , and *ex post* utility,  $u(x_i, 1)$  are clearly differentiated. We denote the joint utility  $U = U(x, E_1 E_2 \dots E_j \dots E_j)$  to represent the utility of prospect  $x$  before the choice maker knows which event occurs. When no confusion occurs, we write  $U(x, E_1 E_2 \dots E_j \dots E_j)$  as  $U(x)$  or  $U$ .

Elastic Surprise (ES) of an event measures the amplitude of “feeling surprised” when this event happens. Concretely, when perceived probability is unity, there perceives no surprise. On the other hand, if the perceived probability approaches 0, the surprise shall generally increase. Instead of weighing the “sure-to-happen” utility of the perceived outcomes  $u(x_j, 1)$ , we can incorporate the utility of the ES to explain the gap between  $u(x_j, 1)$  and  $u(x_j, \phi_i)$ . Indeed, perceived utilities with and without certainty are different even if referring to the same event. To simplify the notation, the utility of a possible outcome is always anchored on the reference,  $u_0 = 0$ , so that  $u(x_j, 1) - u_0 = u(x_j, 1)$ .

We denote  $S(\phi_j)$  as the ES function.  $S(\phi_j)$  or, simply  $S_j$ , generally decreases over  $(0,1]$ . Indeed, when an event is perceived as very likely ( $\phi_j$  approaches 1), the observer will not be so surprised when it happens. Similarly, when the observer perceives the outcome nearly impossible ( $\phi_j$  approaches 0), the observer will be very surprised when it happens. We can also associate ES function with more specific emotions or with neural stimulation, but we leave them as two interpretations among many. In non-compensatory choice models, ES can be treated as an attribute for evaluating an outcome. The paper focuses on the additive utility method.

Let an operator  $\beta_j$  associated with the outcome  $j$  act on the ES function associated with  $E_j$  to have  $u^s(\phi_j)$ , so  $S(\phi_j)$  can be scaled to that of  $u(x, \phi_j)$ . That is,

$$u^s(\phi_j) = \beta_j \cdot S(\phi_j) \quad (4.1)$$

One can then obtain the total perceived utility of a prospect. Note that an ES function is fixed for all outcomes to simplify the analysis, though this assumption can be relaxed. The specification of  $\beta_j$  can be obtained through stated and revealed preference data, and a numerical example is shown in a later section.

By the proposition of additivity and linearity under independence axiom, one can form the additive utility with the expectation form as

$$U = \sum_{j \in \mathbb{J}} \phi_j U_j = \sum_{j \in \mathbb{J}} \phi_j u(x_j, \phi_j), \forall j \in \mathbb{J} \quad (4.2)$$

where

$$U_j = u(x_j, \phi_j) = u(x_j, 1) + u^s(\phi_j) = u_j + u_j^s. \quad (4.3)$$

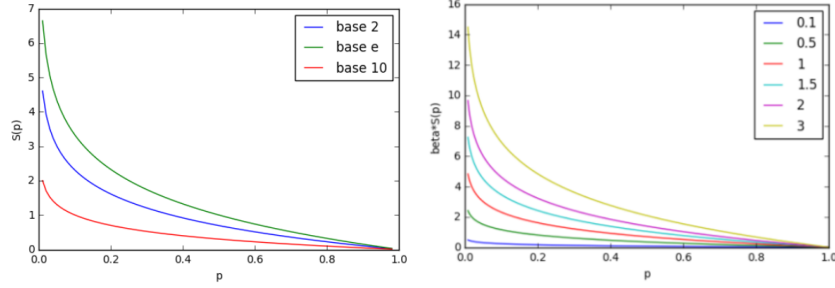
$U$  is the ex-ante utility of  $x$  on  $L$ . When no confusion arises,  $u_j^s$  is denoted as  $u^s(\phi_j)$ .

### ***Logarithmic ES Function and Information Entropy***

This section discusses a specific form of ES function, logarithmic ES (L-ES) function and explain why this is a natural benchmark for studying other ES functions. Probabilities are not transformed to the misperceived ones to allow our focus on the ES function itself. L-ES has some intriguing and cognitively implicative properties that will be further discussed in Section 6. Let L-ES function be

$$S(p_j) = \log_b \frac{1}{p_j} + C_j \quad (4.4)$$

where  $b$  is the base that captures the scale of the L-ES function, and  $C_j$  is a constant.  $C_j = 0$  since when  $p_j = 1$  the L-ES function is 0. The L-ES with different bases and scalers are plotted in Figure 4.1.



**Figure 4.1** The left graph shows the logarithm of  $1/p$  where  $p = (0,1]$  with base 2, e, and 10. The right graph shows different natural logarithm of  $1/p$  by varying the scaler

Expanding  $U$  linearly with L-ES for prospect  $x$  and combining (1),

$$\sum_j p_j (u_j + u_j^s) = E(u) + E(u^s) \quad (4.5)$$

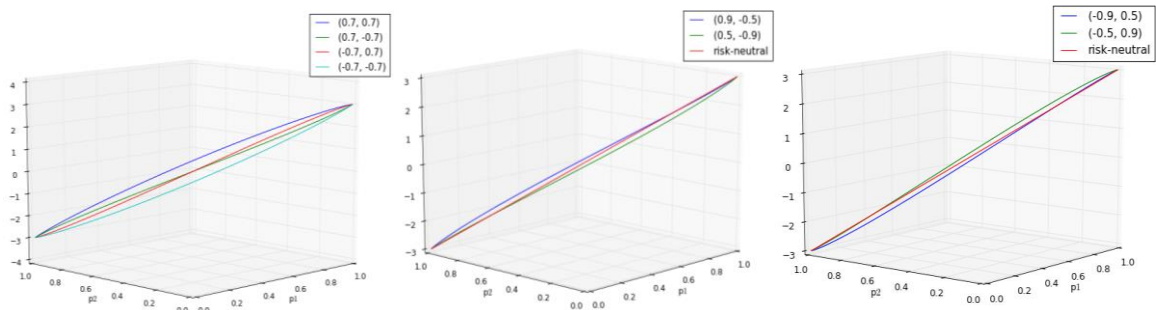
where  $E(u) = \sum_j p_j u_j$  and  $E(u^s) = \sum_j p_j u_j^s = \sum_j p_j \beta_j \log_b \frac{1}{p_j}, \forall j \in \mathbb{J}$ .

How does the formulation capture risk preference? To answer this question, let's study the effect of  $\beta_j$  on  $u_j^s$  and  $U_j$ . Suppose the "sure-to-happen" utility  $u_j$  is measurable. If  $U_j$  is larger than  $u_j$ , the choice maker overestimates the actual gain or underestimate the actual loss of  $x_j$ . We say that this choice maker is risk-prone for  $x_j$ . If the revealed utility for this outcome,  $U_j$ , is larger than  $u_j$ , the choice maker underestimates the actual gain or overestimates the actual loss of  $x_j$ , and we say that this choice maker is risk-averse for  $x_j$ . If the revealed utility for this outcome,  $U_j$ , equates  $u_j$ , the choice maker perceives the actual gain or loss of the  $x_j$ , and we say the choice maker risk-neutral for  $x_j$ . We can describe the above statement mathematically by let

$$U_j = u_j + \beta_j \cdot \log_b \frac{1}{p_j}, \quad \beta_j = \begin{cases} k_1, & x_j > x_0 \\ 0, & x_j \sim x_0 \\ k_2, & x_j < x_0 \end{cases} \quad (4.6)$$

where  $x_0$  is the reference (or anchor) and  $k_j \in \mathbb{R}, j \in \mathbb{J}$ . When  $x_j$  is favorable ( $x_j > x_0$ ),  $k_1 > 0$  renders risk-prone and  $k_1 < 0$  renders risk-averse for  $x_j$ . On the other hand, when  $x_j$  is unfavorable ( $x_j < x_0$ ),  $k_2 > 0$  renders risk-prone and  $k_2 < 0$  renders risk-averse for  $x_j$ . When people have different risk preferences for different possible outcomes, the overall utility,  $U$ , with respect to the expectation of  $u_j$  (Bernoulli utilities), depends on the relationship of  $k_1$  and  $k_2$  and

specific location on  $L$ . this complication is illustrated in Figure 4.2 with different combinations of  $k_1$  and  $k_2$  in a gamble ( $p_1: 3, p_2: -3$ ). When  $k_1 > 0$  and  $k_2 > 0$ ,  $U$  is systematically larger than  $E(u)$ . This suggests that the choice maker perceives uncertainty as a positive factor, and therefore, reveals as risk-prone. When  $k_1 < 0$  and  $k_2 < 0$ ,  $U$  is systematically lower than  $E(u)$ . This suggests that the choice maker perceives uncertainty as a negative factor, and therefore, reveals as risk-prone. When  $k_1 \cdot k_2 < 0$ , the formulation reveals risk-prone on one side while risk-averse on the other, depending on the “dominant”  $k_j$  at the specific location of  $L$ .



**Figure 4.2** The left shows the effect of different combinations of  $k_1$  and  $k_2$  ( $k_1 \cdot k_2 > 0$ ) on  $U$ ; the middle shows the effect of different combinations of  $k_1 > 0$  and  $k_2 < 0$  on  $U$ ; The right shows the effect different combinations of  $k_1 < 0$  and  $k_2 > 0$  on  $U$

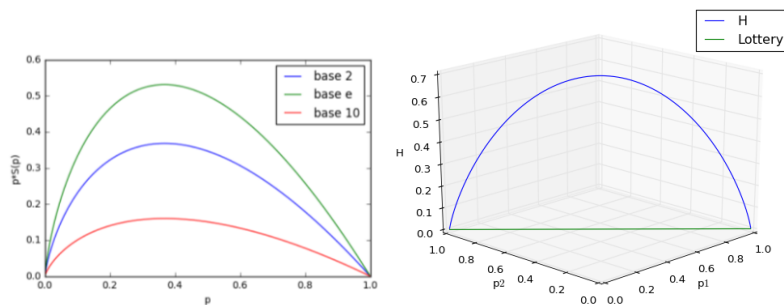
Note that the two types of risk preference (one for  $U$  and the other for  $U_j$ ) should be clearly specified before any meaningful discussion. For example, an observer might be risk-prone to the favorable outcomes (i.e., overestimating its potential gain) and risk-prone to the unfavorable outcomes (i.e., overestimating its potential loss) and the overall revealed choice behavior might be risk-prone, risk-averse, or risk-neutral. This suggests that ES can be used to consider more complex situations where choice maker reveals different risk preference on different locations on a lottery. This flexibility will be shown in the case studies.

Now, let’s explore a scenario where all outcomes are perceived to be favorable or unfavorable comparing to the subjective reference. That is,  $\beta_1 = \beta_2 = \dots = \beta_j = \dots = \beta_j = \beta$ . Let’s take it into (4.5) to have

$$U = E(u) + \beta \cdot H \tag{4.7}$$

where  $H = \sum_{j=1}^J p_j \log_b \frac{1}{p_j}$ ,  $\forall j \in \mathbb{J}$  is known as Shannon's information entropy.

Let's pause it here. (4.7) shows that the information entropy can be viewed as the expectation of L-ES and the utility of a prospect can be interpreted as the combination of the expected Bernoulli utility and the utility associated with the risk when the observer has no preference over the possible outcomes. In other words, the expectation of L-ES can be used to quantify perceived uncertainty when  $U$  only considers the elastic surprise. Appendix II proves that when three basic conditions are required, information entropy is the unique uncertainty measure. L-ES becomes, therefore, a natural benchmark to determine whether a given ES function overestimates or underestimates uncertainty. Figure 4.3 plots the multiplication of the probability and the ES function, which can be interpreted as the perceived uncertainty on  $L$ . When probability approaches zero or one, the uncertainty approaches zero because the observer perceives the outcome as a sure-thing. Uncertainty is maximized when the observer perceives the outcomes equally likely. Due to the constraint  $\sum_j p_j = 1$ , the curve for  $p \cdot \log_b \left( \frac{1}{p} \right)$  loses a degree of freedom and is skewed towards the y-axis.



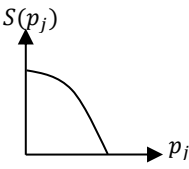
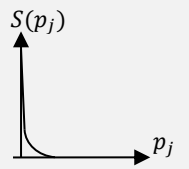
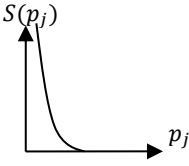
**Figure 4.3** The left shows  $p \cdot S(p)$ . The right shows entropy  $H$  over a bi-outcome lottery

There are two implications from this analysis. First, when uncertainty is not perceived consistently with L-ES function, the choice maker fails to evaluate uncertainty objectively if the three basic conditions in Appendix II are required and, therefore, over/under-estimating risk is unavoidable. Second, the amount of information received by an observer can be measured by the

change of the perceived entropy. Therefore, L-ES function is a natural pivot to study other types of ES functions.

Table 4.1, as a demonstration, shows other candidate ES functions and their corresponding impact on  $U$ . We use the constant  $\beta$  so that we can focus on the effect of different forms of ES functions.

**Table 4.1** Selected surprise functions and their corresponding linear transformation to utility

$S(\cdot)$	General Shape	$U$	Property
$1 - p_j^n$		$E(u) + \beta \sum_{j=1}^J p_j (1 - p_j^n)$ $= E(u_j) + \beta (1 - p_1^{n+1} - \dots - p_J^{n+1})$	Since $p_j^n \leq p_j$ when $n \geq 1$ , $U$ is formed by a “hyper-sphere” around the hyperplane $E(u_j)$ on $L$ . When $n = 0$ , $U$ is risk neutral. When $n = 1$ , $U$ measures uncertainty using Gini-Simpson index. When $n \rightarrow +\infty$ , $U$ takes risk as a constant
$\left(\frac{1}{p_j}\right)^m - 1$		$E(u) + \beta \sum_{j=1}^J p_j \left( \left(\frac{1}{p_j}\right)^m - 1 \right)$	When $m = 1$ , the second term becomes $\beta \cdot (J - 1)$ . This uncertainty measurement is only related to the number of perceived outcomes. When $m > 0$ , it is essentially ES function $1 - p_j^n$ when $n < 0$ .
$\tan(\pi(1 - p_j))$		$E(u) + \beta \sum_{j=1}^J p_j \tan(\pi(1 - p_j))$	This is a legit ES because $S'(p) < 0$ and $\lim_{p \rightarrow 0^+} p \cdot \tan(\pi(1 - p)) = 0$ .

## Properties and Cognitive Implication

This section discusses some interesting properties and revealed implications when limiting the perceived probability in an ES function to zero or increasing the number of perceived possible outcomes. Then, we discuss a potential theoretical issue critical to many theorists.

### Limit at $0^+$

We have discussed the property of ES, in which the closer the  $p_j$  to unity the more certain the observer perceives  $E_j$  to occur (i.e., low uncertainty); similarly, the closer the  $p_j$  to 0, the more certain the observer perceives  $E_j$  to not to occur (i.e., low uncertainty as well). In the example of limiting property of  $p \log_b \frac{1}{p}$  where  $p$  approaches 0, “ $0 \cdot \infty$ ” condition emerges. The rate of

approaching 0 and the rate of approaching  $+\infty$  can be compared for  $\log_b \frac{1}{p}$ ,  $b > 0$  using the L'Hospital's Rule. Since  $\lim_{p \rightarrow 0^+} p \log_b \frac{1}{p} = \lim_{p \rightarrow 0^+} \frac{p'}{\left(\log_b\left(\frac{1}{p}\right)\right)^{-1}} \rightarrow 0$ ,  $b > 0$ , the form of expectation of ES is closed.

Recognizing that human only perceives and evaluates a limited number of discrete events simultaneously due to brain's capacity (Miller, 1956) necessitates the consideration of grouping effect. An appropriate homeomorphic model should be capable of capturing the situation where an increasing number of events leads to a heavier burden on perception. Unfortunately, this is commonly omitted in the existing methods. In fact, many methods treat the capability of considering continuous random events as an advantage. In contrast, the conventionally implicit assumption on infinite power of perception becomes explicit using the ES-based method. Take L-ES as an example and expand (4.5) as

$$-\sum_j p(\alpha_j) \Delta\alpha \log(p(\alpha_j) \Delta\alpha) = -\sum_j p(\alpha_j) \log(p(\alpha_j)) - \left(\sum_j p(\alpha_j) \Delta\alpha\right) \log \Delta\alpha \quad (4.8)$$

and when limit the number of perceived outcomes,

$$H = - \int_{\alpha \in \mathbb{R}} \log(f(\alpha)) dF(\alpha) - \lim_{\Delta\alpha \rightarrow 0} \log \Delta\alpha \quad (4.9)$$

where  $f(\alpha)$  and  $F(\alpha)$  are the probability density function and the cumulative density function, respectively, of the payoff  $\alpha$  when an observer could perceive infinitely number of events simultaneously. The second term of (4.9) causes the whole formulation to approach infinity, which violates the limited nature of cognitive capability. In other words, the weights associated with the ES increase nonlinearly with the number of perceived outcomes. Attenuating weights might not reflect the weights a choice maker assigns, and therefore, the number of perceivable events should be carefully determined.

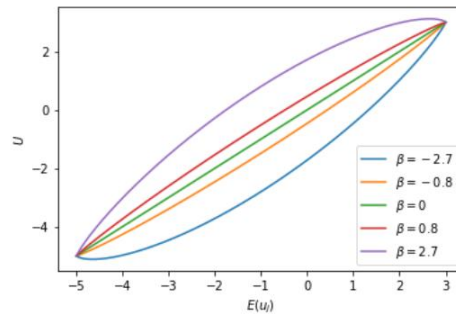
I have made similar discussion in the assumptive caveat on an observer's cognition in Chapter 3 (formulation (3.2) and (3.3)). It is important to notice that this seemingly repetitive work



comes from two different foundation and starting point. In Chapter 3, the assumption of observer’s cognition when using Shannon’s information entropy to quantify perceived information is discussed, while, in this chapter, the same conclusion is reached when studying the underlying cognitive assumption on L-ES function.

### Rationality, Stochastic Dominance, and Trade-off Consistency

Figure 4.4 shows the impact of different  $\beta$  on  $U$ . Since,  $\beta$  is a constant for the two possible outcomes in  $(p_1: -5, p_2: 3)$ , regardless of the preference on different outcomes, the overall convexity is not influenced by the location on  $L$ . That is, when  $\beta > 0$ , the curve  $U$  is concave, and when  $\beta < 0$ , the curve  $U$  is convex. When  $\beta = 0$ , the perceiver is risk-neutral.



**Figure 4.4** Relationship of  $U$  and  $E(u)$  on the two-outcome lottery with a range of  $\beta$  that captures risk-preference with the potential trade-off inconsistency.

For normative models founded on rationality axiom,  $\beta$  should not render the case of  $\min\{u_1, u_2\} \leq E(u) + E(u^s)$  – how could one ever prefer losing \$5 for sure than only having 50% chance of losing \$5 and another 50% chance of gaining \$3? This type of constraint is sometimes referred to as first-order stochastic dominance – moving probability mass to the less favorable outcome should strictly worsen this prospect. However, the condition  $\min\{u_1, u_2\} \leq E(u) + \beta \cdot H$  requires  $\beta$  to infinitely approach 0. This is because  $H$  “drags down”  $u_1$  more severely than the effect of “pulling up” by  $u(x_2)$  when  $p_1$  approaches 1 regardless of how close the negative  $\beta$  is to 0. Similarly,  $H$  “pulls up”  $u_2$  more severely than the effect of “dragging down” by  $u(x_2)$  when  $p_2$  approaches 1 regardless of how close the positive  $\beta$  to 0. The assumption that an observer can perceive infinite surprise in L-ES function form is the culprit. Three of the possible ways to avoid

this issue are (1) choosing reasonable ES that contains certain upper threshold, (2) using nonlinear scaler  $\beta$  to compensate the effect caused by large ES, or (3) a combination of (1) and (2). I do not argue that this normative assumption must be true in a descriptive model, but it is important to be aware that the form and scale of an ES function could have a significant implication on behavioral assumptions of a descriptive model.

Trade-off consistency in a broad sense refers to a decision where improving an outcome in a prospect improves the favorability of this prospect. In a simple scenario where  $\beta_1 = \beta_2 = \dots = \beta_j = \dots = \beta_j = \beta$ , this consistency is automatically satisfied as long as  $u_j$  strictly increases with  $x_j$ . In a case where  $\beta$  and ES are outcome-specific, trade-off consistency needs to be closely examined to avoid violating the trade-off consistency (if this consistency is assumed necessary.)

### ***EUT and Its Revision***

Under the EUT, utility on  $L$  and utility on sure-amount are termed as von-Neumann-Morgenstern (v.N-M) expected utility and Bernoulli utility, respectively. The v.N-M utility is the expectation of sure-amount utilities. In other words, the EUT assumes that a choice maker values utility of  $j$  under certainty the same as the utility of  $j$  under uncertainty.

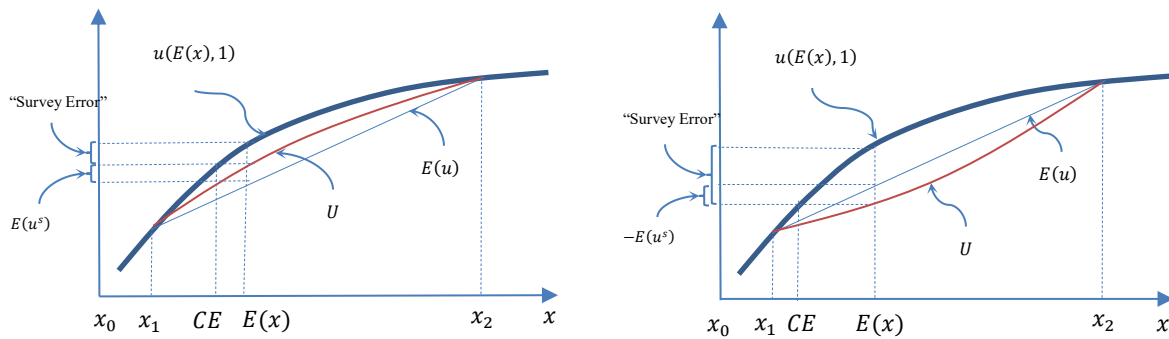
The EUT determines the risk preference by comparing their expected (Bernoulli) utility with the utility of certainty equivalent (CE). Historical and newly proposed drawbacks of using this relation to capture risk preference are reviewed in Section 2 and 3. I revise the EUT and name it Revised Expected Utility Theory (REUT) by differentiating  $U_j$  and  $u_j$ , so that risk preference is determined based on the relationship between  $U$  and  $E(u_j)$ . For instance, in  $(p_1x_1p_2x_2)$ , if  $U < p_1u_1 + p_2u_2$ , the choice maker is risk-averse. The gap between  $U$  and  $E(u)$  is explained by  $E(u^s)$ . In other words,  $u_j$  is “freed” from the mission of capturing the risk preference.

With  $u_j$ ,  $E(u)$ ,  $U_j$ , and  $U$  differentiated, one can use two different criteria for determining risk preference: the expected payoff and the expected utility. The relationship is summarized in Table 4.2.

**Table 4.2** Risk preference regarding expected payoff and expected utility in REUT.

	Regarding utility of expected payoff	Regarding expected Bernoulli utility
Risk-Prone	$U > u(E(x), 1)$	$U > E(u)$
Risk-Neutral	$U = u(E(x), 1)$	$U = E(u)$
Risk-Averse	$U < u(E(x), 1)$	$U < E(u)$

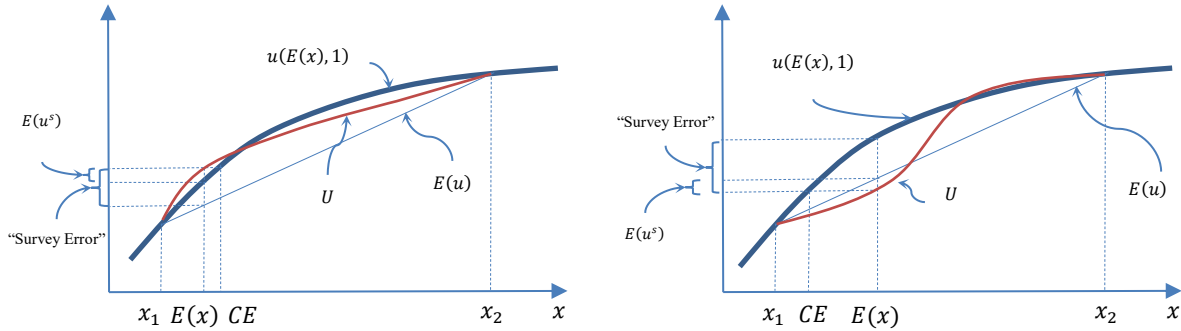
Figure 4.5 illustrates the concepts by comparing  $u_j$ ,  $E(u)$ ,  $U_j$ , and  $U$  in different scenarios.  $E(x) = \sum_{i=1,2} p_i x_i$  is the expectation of the pay-offs for the prospect  $x$ . Also shown is their relationship with CE and the survey “error” when using the Bernoulli utility,  $u(E(x), 1)$ , to capture the effect of  $U = u(E(x), p)$ . Although rare, it is indeed possible for  $U$  to overlap with  $u(E(x), 1)$ . In this case, the gap between  $u(E(x), 1)$  and  $E(u(x, 1))$  is the the same as that between  $U$  and  $E(u(x, 1))$ . In fact, this is the only case where REUT and EUT are equivalent.



**Figure 4.5** In both cases, the choice maker behaves risk-averse regarding the utility of expected pay-off since  $U < u(E(x), 1)$ . The left behaves risk-prone and the right behaves risk-averse, regarding the expected utility since  $U > E(u(x), 1)$ .

By differentiating  $U$  and  $u(E(x), 1)$ , modeling choice maker’s risk preference has greater flexibility. Figure 3.6 shows two more complex cases where risk preference changes on  $L$ . In terms of expected utility, the choice maker on left behaves risk-prone since  $U > u(E(x), 1)$ . In terms of the utility of expected payoffs, however, it behaves risk-prone when close to  $x_1$  and risk-averse when

close to  $x_2$ . The choice maker on right behaves risk-prone in terms of expected utility since  $U > u(E(x), 1)$ . In terms of the utility of expected payoffs, however, it behaves risk-prone when close to  $x_1$  and risk-averse when close to  $x_2$ .



**Figure 4.6** Examples of changing risk preference on  $L$ .

Not only the differentiation of  $u_j$ ,  $E(u)$ ,  $U_j$ , and  $U$  resolves conventional issues in EUT and paradoxes caused by the dependence of the curvature of  $u(x)$ , it also allows the change of risk preference over a given  $L$ . Although it is possible that  $U$  “overlaps” with  $u(E(x), 1)$  or  $E(u)$ , the risk-prone (risk-averse) behavior is caused by the relative relationship between  $u(CE, 1)$  and  $U$ , not the convexity of  $u(E(x), 1)$ .

Note that the introduction of ES modifies the meaning of some commonly used inequalities in EUT. For example, the Jensen’s inequality,  $\int u(\alpha)dF(\alpha) \leq u(\int \alpha dF(\alpha))$ , would only suggest the concavity of the Bernoulli utility function and no more no less. Even if  $\int u(\alpha)dF(\alpha) \leq u(\int \alpha dF(\alpha), 1)$ ,  $U$  could still be greater than  $u(\int \alpha dF(\alpha), 1)$ , and therefore, risk-prone regarding the expected payoff. This way, the change of utility unit and scale does not influence the risk preference in REUT.

### ***Reference Dependency and Relationship with CPT***

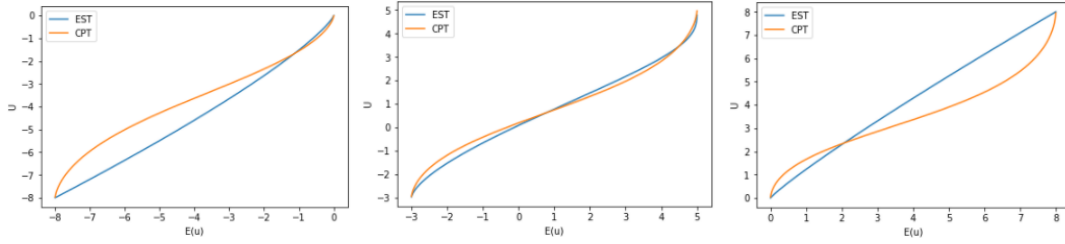
This section studies the effect of reference/framing conditions on ES and the overall preference. I assume no misperception in this section. First, suppose  $(p_1x_1p_2x_2)$  and ES is defined as

$$S(p_j; k_j) = \left( \frac{1}{p^{k_j}} - 1 \right), \quad k_j = \begin{cases} k^+, & x_j \geq x_0 \\ k^-, & x_j < x_0 \end{cases} \quad (4.10)$$

The corresponding Bernoulli utility  $u(x_j, 1)$  is used to scale the ES so that

$$u(x_j, p_j) = u(x_j, 1) \cdot (1 + S(p_j; k_j)) \quad (4.11)$$

Note that the scaler for the ES in (14) is proportional to the corresponding outcome's Bernoulli utility. The test of a range of reference from  $x_0 = -3$  to 5 is plotted in Figure 4.7 for scenario (0, 8), (-3, 5), and (-8, 0) with parameter  $k^+$  and  $k^-$  to be (-, 0.1), (1.9, 1.1), (0.1, -), respectively. The result from CPT for comparison is also plotted.



**Figure 4.7** Comparing the effect of reference-dependence using proposed method with that in CPT. X-axis is the expected payoff.

Although using ES within utility function generates similar results to that from CPT when reference is set to have both favorable and unfavorable events, there is a significant difference when the reference approaches -3 or 5. That is, certain framing effect might lead the ES function generates distinct results than that from CPT. This discrepancy leads to a curiosity of what kind of ES function is “embedded” in CPT.

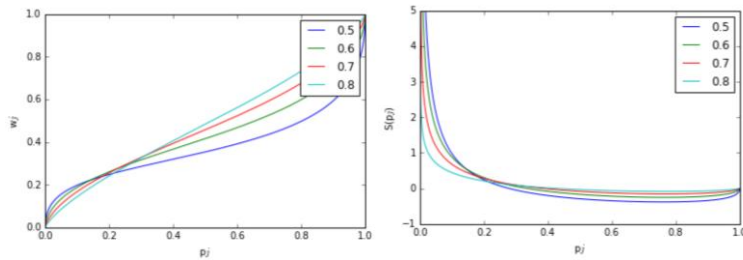
When  $\pi(p_j) \cdot v_j$  in CPT is numerically the same as  $\phi(p_j) \cdot u(x_j, p_j)$ , there exists a numerical equivalency. Keeping the value function as,

$$v(x) = \begin{cases} x^{\alpha^+}, & x \geq 0 \\ -\lambda(-x)^{\alpha^-}, & x < 0 \end{cases} \quad (4.12)$$

By equating  $\pi$  and  $p \cdot (1 + S(p))$ ,

$$S(p) = \frac{p^{\delta-1}}{(p^\delta + (1-p)^\delta)^{1/\delta}} - 1 \quad (4.13)$$

The weighting function with various  $\delta$  and the corresponding ES function with and without probability misperception is plotted in Figure 3.8.



**Figure 4.8** Left: CPT weighting function; Right: equivalent ES function for  $\delta$

Figure 4.8 shows that negative surprise occurs when there exists a probability misperception ( $\delta \neq 1$ ). How to interpret the positive surprise and negative surprise? On the other hand, this clearly verifies the so-called fourfold pattern of risk attitudes, which can be explained in CPT by the interplay of over/underweighting of small probabilities and convexity of value functions: “when gains have moderate probabilities and losses have low probabilities, choice maker behaves as risk-averse because the losses tend to be overestimated (positive region in the right plot of Figure 4.8) and the gains tend to be underestimated (negative region in the right plot of Figure 4.8); when losses have moderate probabilities and gains have small probabilities, choice maker behaves as risk-prone because the gains tend to be overestimated and the losses tend to be underestimated.” (Scholten and Read, 2014). What’s more, a choice maker might overestimate gain (loss) and underestimate loss (gain) and reveals risk-seeking, risk-neutral, or risk-aversion when the risk preference over favorable (unfavorable) event has a larger overall effect.

On the other hand, we know that zero surprise suggests that the observer would experience no surprise when the corresponding outcome happens; one way to interpret the positive and the negative surprise is through emotions: the observer would tend to take the outcome with moderate or large probability “for granted” and take the outcome with low probability as “doomed-to-happen.”

The positive and negative surprise can also be explained the effect of probability misperception. Indeed, by adjusting  $k$ , we may also achieve the same numerical property when  $S(\cdot)$  is strictly decreasing. For example, when  $\phi(p) = w(p) \cdot p^k$ ,  $S(p)$  is simply (10) and non-negative. Therefore,  $\phi(\cdot)$  and  $S(\cdot)$  could have more than one possible combination to generate the same overall weighting function. This inseparability between probability misperception and ES will be further discussed in Section 11.

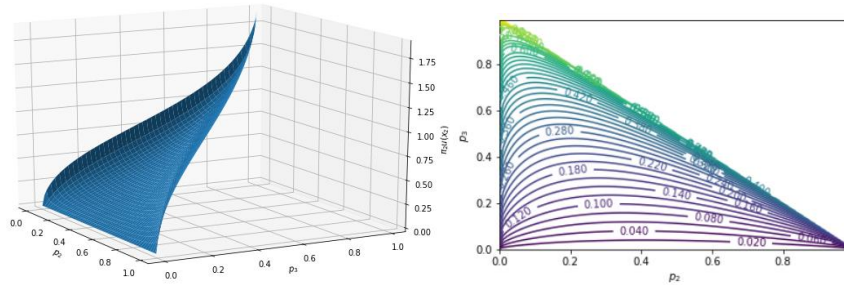
Similar dissection applies to the situation with more than two events. In  $(E_1x_1E_2x_2E_3x_3)$  and  $x_3 > x_2 > x_0 > x_1$ , we have

$$\pi_2^+ = w^+(p_2 + p_3) - w^+(p_3) \quad (4.14)$$

in which,

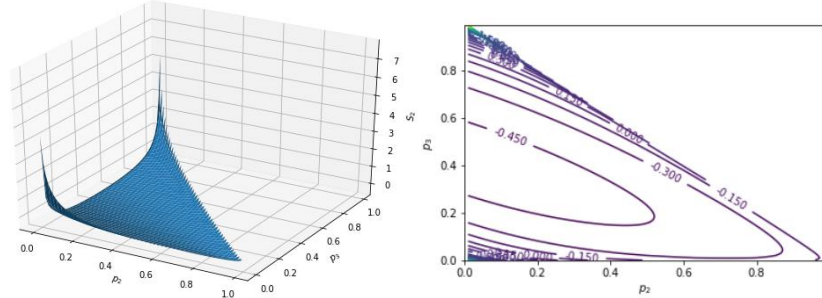
$$w^+\left(\sum_{i \geq j} p_i; \delta\right) = \frac{(\sum_{i \geq j} p_i)^\delta}{\left((\sum_{i \geq j} p_i)^\delta + (1 - \sum_{i \geq j} p_i)^\delta\right)^{1/\delta}} \quad (4.15)$$

Figure 4.9 shows the surface and the contour of the surprise function of  $\pi_2^+ u(x_2)$  over  $p_2$  and  $p_3$ , where  $u(x_2) = 2$  is the Bernoulli utility of payoff  $x_2$  with the preference  $x_0 = 0$ .



**Figure 4.9** Left:  $\pi_2^+$  (z-axis) in a three-outcome decision, where  $X_1 \leq 0 \leq X_2 \leq X_3$ ; Right: the contour of  $\pi_2^+$

Let's define the Compound Weighting (CW) Function as the CPT weighting function consisted of ES function that that  $\pi = p(S + 1)$ .  $S_2 = \frac{\pi_2^+}{p_2} - 1$  is plotted in Figure 4.10. Note that  $S_2$  contains an interaction between  $E_2$  and  $E_3$ . In other words, the decision weight for  $E_2$  considers both the probability for  $E_2$  and that for  $E_3$ .

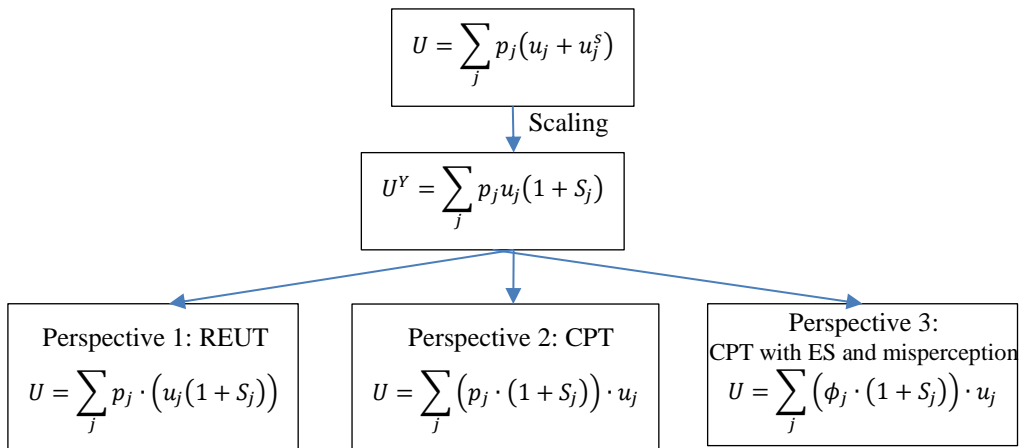


**Figure 4.10**  $S_2$  over  $p_2$  and  $p_3$  derived from a 3-outcome decision based on the Cumulative Prospect Theory

One might argue that instead of CW function from  $p$ , why not  $\phi$ ? Indeed, if  $w$  is treated as the misperceived cumulative probability, the ES could be a function of the post-misperceived probability rather than the pre-misperceived one. The only reason that  $p$  is used in this section rather than  $\phi$  is to isolate the potential nonlinear effect of the misperception to allow the focus on the effect of ES.

### Three Perspectives

The connection of ES and categorical perception with EUT and CPT has been discussed in the previous section. With the revision on reference and addition of appropriate ES function on REUT makes it essentially equivalent to CPT in principle. Figure 4.11 illustrates this claim.



**Figure 4.11** The Revised EUT (REUT) and CPT can be derived from incorporating ES differently.



Note that  $\Psi_j = \phi_j \cdot (1 + S_j)$  in Perspective 2 and Perspective 3 should be differentiated from  $\pi_j$  in CPT, though the overall effect may be the same.  $\pi_j$  in CPT has no explicit cognitive construal, while  $\Psi_j$  are specifically referring to the combined effect of the (probability) misperception and the ES. Despite the explicitness of  $\Psi_j$ , differentiating the effect of the ES and that of  $\phi_j$  encounters the identification issue since only the revealed effect of  $\Psi_j$  is observable through experiment, which requires assumptions on the cognition in practice.

### ***ES in Mean-Variance (MV) Method***

What is the underlying ES function in the Mean-Variance method? For its popularity in finance and traveler choice modeling, I choose to discuss MV method as an example to demonstrate the unifying nature of ES. In MV, uncertainty is captured by historical variation measured by variance  $Var(TT)$ .

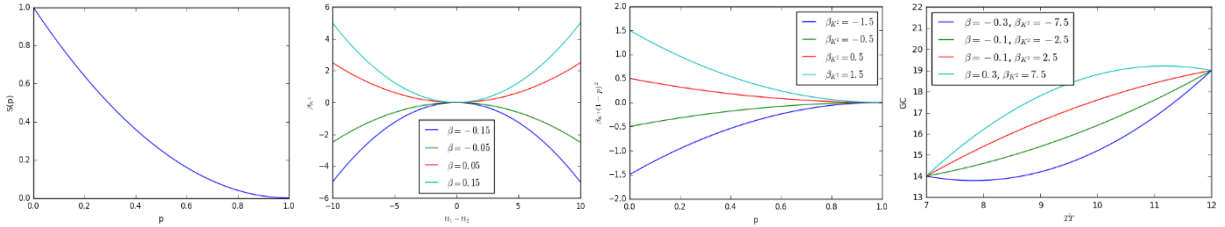
Suppose a general cost function  $GC$  is defined as

$$\bar{t}t + \beta \cdot Var(TT) = \bar{t}t + \beta \cdot \sum_{i=1,2} p_i (tt_i - \bar{t}t)^2 \quad (4.16)$$

where  $tt_1$  and  $tt_2$  are possible realizations of the travel time random variable  $TT$  with distribution  $(p_1: tt_1; p_2: tt_2)$  and perceived mean travel time  $\bar{t}t$ ;  $\beta$  is a scaler. Since  $\sum_{i=1,2} p_i (tt_i - \bar{t}t)^2$  can be written as  $p_1(p_2 tt_1 - p_2 tt_2)^2 + p_2(p_1 tt_2 - p_1 tt_1)^2$ , (4.16) can be decomposed as

$$p_1[tt_1 + \beta_{K^2} p_2^2] + p_2[tt_2 + \beta_{K^2} p_1^2] = \sum_{i=1,2} p_i [tt_i + \beta_{K^2} S(p_i)] \quad (4.17)$$

where  $\beta_{K^2} = \beta \cdot K^2 = \beta \cdot (tt_1 - tt_2)^2$ , a constant in the given decision context. Since  $p_1 + p_2 = 1$ , we have  $S(p) = (1 - p)^2$ . Figure 4.12 shows the implicitly embedded ES function, the relationship between  $K$  and  $\beta_{K^2}$ , the relationship between  $p$  and  $\beta_{K^2}(1 - p)^2$ , and the relationship between  $\bar{T}T$  and  $GC$  for different scalars given that  $tt_1 = 7min$  and  $tt_2 = 12min$ .



**Figure 4.12** (a) The implicitly embedded ES function. (b)  $K$  versus  $\beta_{K^2}$ . (c)  $p$  versus  $\beta_{K^2}(1 - p)^2$ . (d) The  $\overline{TT}$  versus GC given different scalers  $\beta$ .

The scaler,  $\beta$ , needs to be tuned to avoid violating the rationality assumption which has been rarely examined in historical MV model estimation processes. For instance, if  $j$  is the most preferable event in the prospect  $x$ , any location on the lottery should be worse (in terms of  $U$ ) than the general cost/disutility of  $j$  under certainty. In (d), the case of  $\beta = -3, \beta_{K^2} = -7.5$  and the case of  $\beta = 0.3, \beta_{K^2} = 7.5$  apparently violates the rationality assumption. Note that one key characteristic of a MV-based model is that the parameter used to convert surprise to the same scale with the general cost is a function of  $tt_2 - tt_1$ ; therefore, it is possible to determine  $\beta$  to obtain a model applicable to a range of decision contexts without violating the rationality assumption. This role of  $\beta_{K^2}$  is similar to using scaler sensitive to the magnitude of  $u_j$ .

A similar analytical approach can be applied to multiple-outcome cases and related approaches such as mean-deviation models and quantile-index models, though the transformed formulation might not be closed.

### ***Empirical Study on Route Choice under Risk***

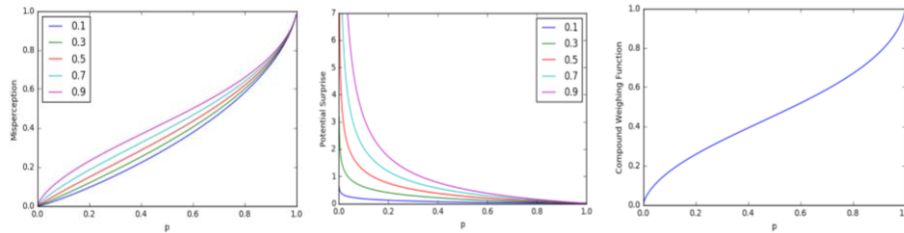
Recalled that EST connects with CPT since the weighting function can be understood as an ES function that has both a positive and a negative region or that has one (original) monotonic ES function affected by a probability misperception. This section uses an empirical study to demonstrate how to use a strictly decreasing ES function in conjunction with CPT to improve model interpretability and predictions. This section adjusts CPT to explicitly incorporate misperception

and ES. We also estimate the parameters based on a stated-preference survey and examine the impact of adjusting CPT to incorporate ES on capturing revealed behavior under travel time risk.

The survey was conducted among 91 participants (67 undergraduates and 24 graduates) from the University of California, Irvine. The participants were randomly grouped into two for reducing the test length. ES function  $S(p; k^-) = \frac{1}{p^{k^-}} - 1$ , is selected for its convenient form of the CW function,  $w^*$ . That is,

$$w^* = w \cdot (1 + S(p; k^-)) = \frac{p^{\delta_{up}^-}}{(p^{\delta_{down}^-} + (1-p)^{\delta_{down}^-})^{1/\delta_{down}^-}} \quad (4.18)$$

where  $\delta_{down} = \delta$  and  $\delta_{up} = \delta_{down} - k^-$ . Note that  $w_j$  has the same form as does in CPT but  $\delta$  only captures misperception. Consistent with CPT, the superscript + and - denotes favorable and unfavorable outcomes, respectively. Since the reference is set as 0, positive travel time is perceived as a loss. Let's define the ratio of splitting parameter for the ES from  $\delta_{up}^-$  as  $\theta$  and define  $\delta_{down}^- = \theta \cdot \delta_{up}^-$  and  $k^- = -(1 - \theta) \cdot \delta_{down}^-$ .  $w^*$  with various  $\theta$  is plotted in Figure 4.13 with fixed  $\delta_{up}^-$  as 0.69 to be comparable with CPT.

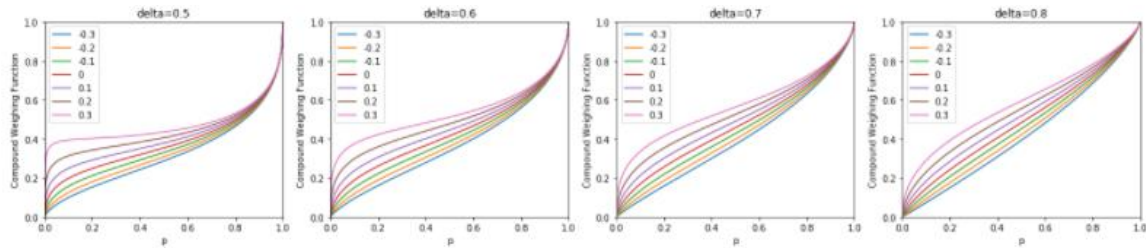


**Figure 4.13** Effect of different splitting factor  $\theta$  on the shape of misperception  $\pi$ , ES Function, and the Compound Weighting Function  $\pi^*$

Since varying  $\theta$  has no impact on the CW function, one immediate question is what portion of  $\delta + k^-$  “belongs to” misperception and what portion “belongs to” the ES? Since only the revealed decision weight is observable, the effect caused by misperception and that by ES are inseparable. However, theorists who are hesitant to forsake some properties in the CPT can set  $\delta^- = \delta_{up}^- = \delta_{down}^-$  so that (18) becomes

$$w^* = \frac{p^{\delta^- - k^-}}{(p^{\delta^-} + (1-p)^{\delta^-})^{1/\delta^-}} \quad (4.19)$$

Figure 4.14 shows the effect of  $k^-$  on different  $\delta^-$ s. As expected, the formulation is equivalent to CPT when  $k^- = 0$ . Scenarios of  $k < 0$  are included since the scaler for ES may not be the same as the Bernoulli utility with respect to the reference.



**Figure 4.14** Effect of  $k$  on  $\pi^*$  given different  $\delta^-$ .

The value function follows the loss side of (4.12), and the reference point is set to be the event with the smaller travel time. In addition to travel time, factors such as the importance of activity, departure time, and comfort level may also have impacts on how a traveler evaluates alternatives. More specifically, a participant would, say, depart 15 min earlier than the start of the activity if he knows the travel time to work could be following distribution (15min: 0.7, 25min: 0.3), while another participant would depart 25 min earlier with the same perceived distribution. The first participant is risk-prone since the penalty of being 25min is low (e.g., grocery shopping) while the second participant is risk-averse since the penalty of being late is significant (e.g., work). To mitigate this type of ambiguity, the survey participants were told that all the answers should be based on their typical experience of commuting. Table 4.3 shows the expected and median travel time in the survey for varied decision context. Table 4.4 summarizes the mean travel time equivalent and sample standard deviation from the survey.

**Table 4.3** Mean and median travel time and event probabilities in the survey (in minutes)

Event	Expected travel time for various probabilities of the second event (not given to participants)					Median of travel time certainty equivalent given various probabilities of the second event				
	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
15 25	16.00	17.50	20.00	22.50	24.00	16	17	19	20	22
15 30	16.50	18.75	22.50	26.25	28.50	17	19	22	24	27

30	45	31.50	33.75	37.50	41.25	43.50	32	33	35	35	44
30	50	32.00	35.00	40.00	45.00	48.00	32	34	39	41	48
40	50	41.00	42.50	45.00	47.50	49.00	43	44	44	45	48
40	60	42.00	45.00	50.00	55.00	58.00	44	45	47	52	57

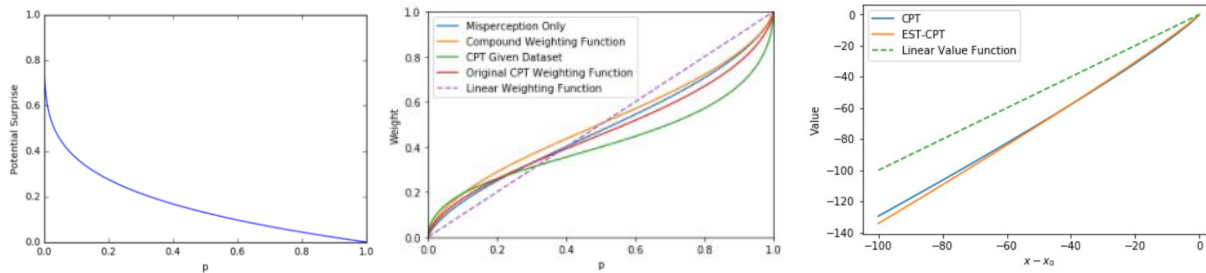
**Table 4.4** Mean travel time and standard deviation (sample) from the survey (in minutes)

Event		Mean of travel time equivalent for various probabilities of the second event					Sample standard deviation of travel time certainty equivalent for various probabilities of the second event				
		0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
15	25	15.90	16.34	19.05	18.54	23.90	0.96	0.92	0.74	1.22	0.46
15	30	16.66	18.83	22.64	23.60	27.66	0.30	0.48	2.18	1.65	0.81
30	45	31.79	33.97	36.03	38.35	42.79	0.57	1.01	1.17	1.46	0.57
30	50	31.25	32.50	39.39	43.27	46.25	2.06	0.36	0.40	0.66	2.06
40	50	42.40	44.65	43.71	46.23	44.40	0.42	1.21	0.89	1.58	1.42
40	60	43.22	45.44	47.26	53.73	43.22	1.31	1.01	1.47	1.18	0.91

Parameter estimation uses the generalized reduced gradient algorithm.  $\delta^- = 0.79$ ,  $k^- = 0.09$ ,  $\alpha^- = 0.92$ , and  $\lambda = 1.94$  with a mean squared error (MSE) 2.89. Without incorporating  $k^{-1}$ , the MSE is 3.03. During the estimation,  $\frac{n-1}{n-p_k} \approx 0.034$  and  $\frac{n-1}{n-p_c} \approx 0.022$ , where  $n$  is the sample size,  $p_c$  is the number of parameters in the model without  $k^-$  and  $p_k$  is the number of parameters in the model with positive  $k^-$ , respectively. Therefore, the Adjusted  $R^2$  for the model with and without incorporating  $k^{-1}$  are 0.767 and 0.784, respectively. No significant outliers are identified based on Cook's distance. The CW function is verified to be monotonically increasing under this specification. Although the goodness-of-fit analysis is not significantly improved, the model with the consideration of elastic surprise (i.e., consider misperception and uncertainty in a relatively separate manner) endows more explicit cognitive interpretation to the results. It would be worthwhile to explore other forms of ES function.

The estimate reveals the change of risk preference over  $L$  remains some similarity to that in the original CPT ( $\delta^- = 0.69$ ,  $k^- = 0$ ,  $\alpha^- = 0.88$ ,  $\lambda = 2.25$ ) derived from a monetary survey. However, the participants in this example, averagely, reveal less significant probability misperceptions because of the higher  $\alpha^-$ . On the other hand, the value function is less concave ( $\alpha^-$  is closer than unity) because part of the concavity is considered by the ES. The estimated the ES function,

misperception, and CW function are plotted in Figure 4.15 and a comparison is made with that from the original CPT. Note that a moderate probability is also shown underweighted but not as dramatic as that in CPT.



**Figure 4.15** Left: Estimated ES Function. Middle: Weighting function for misperception and compound effect of both misperception and ES. Right: new value function comparing to the original one (without considering ES)

In this study, the reference is set as the event with the smaller travel time. However, it is likely that the reference is set differently for different travelers and decision contexts (e.g. the type and importance of the activity to participate in). Mixed-type decision scenario increases the number of parameters, and I leave the study of different anchor settings to future work. Although  $k^-$  in the estimation is only sensitive to the favorability (relative to the anchor) of each event, the degree of favorability of events can also be considered by setting  $k^-$  sensitive to the specific rank in prospects with two or more outcomes.

## Conclusion

This chapter proposes a unifying approach for incorporating risk preference by considering the different types of Elastic Surprise functions and their scaling factors of possible outcomes. This approach bridges the existing methods such as Information Theory, Expected Utility Theory, Cumulative Prospect Theory, and Mean-Variance Theory, and non-additive attribute-based theories. Categorical perception and cognitive limitation becomes a natural derivative with certain behavioral foundations such as stochastic dominance and trade-off consistency.

Utilizing ES functional shows high flexibility and some common existing methods such as EUT and CPT can be treated as special cases with certain explicit assumptions on choice makers' cognitive characteristics. Therefore, the proposed method can be used as a guide for understanding the assumptions the used to be hidden. The difference between misperception and ES is discussed and the identification issue in estimation is recognized. Due to the common confusion of misperception and perceived uncertainty, conventional methods typically rely on transforming probability and Bernoulli utility to match with revealed choice behavior. This confusion leads to biasedness in estimation and result interpretation. This chapter demystifies this confusion using ES to explain paradoxes and issues.

The empirical study demonstrates the application of incorporating ES in route choice modeling that improves data fitting and result interpretability. The result reveals a lottery-sensitive preference and hypothesizes that this partially explains the inconsistency among existing research about traveler's risk preference. In addition to extending the application to decisions under ambiguity and uncertainty, a specific formulation with strict proof to connect with behavior foundations sensitive to decision context shall further strengthen the proposed method.

## **CHAPTER 5**

# **MULTICLASS, MULTICRITERIA DYNAMIC TRAFFIC ASSIGNMENT WITH PATH-DEPENDENT LINK COST AND ENTROPY-BASED RISK PREFERENCE**

*“We can’t control systems or figure them out. But we can dance with them!”*

*– Donella Meadows*

### ***Introduction***

Dynamic Traffic Assignment (DTA), in short, is a process of allocating traffic onto a time-space network with given physical constraints, usually aimed at dynamic user equilibrium (DUE) or dynamic social optimum (DSO). DTA has a broad range of applications in conventional transportation systems operation and planning as well as in understanding the impact of new strategies and emerging technologies. Peeta and Ziliaskopoulos (2001) briefly review the foundations of DTA, the status at the time, the challenges, and the opportunities. They also categorize existing methods into those based on mathematical programming, optimal control, variational-inequalities, and simulation. The bi-level optimization method proposed by Jayakrishnan et al. (1995) perhaps does not fall in any of those categories as it analytically embeds a simulation within a strict mathematical programming formulation. Jayakrishnan et al. (1994) proposes a static traffic assignment procedure using gradient project method and proves greater convergence efficiency. This method was further developed by Yang and Jayakrishnan (2012) that combines the concept of a static user equilibrium formulation transformed from Beckman’s formulation with a traffic simulator to practically achieve DUE with a single-user class and path-independent link cost(s). DTA models typically capture traffic dynamics utilizing flow-cost (macroscopic), density-speed (mesoscopic), or driving behaviors (microscopic) relationships.

Transportation system users have diverse attributes, behaviors, and impacts on the system. Yang and Huang (2012) identify two situations for naming a traffic assignment problem multi-class



- heterogeneous driving behaviors and heterogeneous route choice. Heterogeneous route choice behaviors and driving behaviors are typically considered with random-parameter models (Washington et al. 2010) and stochasticity-incorporated traffic simulators (Lu et al. 2010), respectively. A system user could evaluate alternatives based on multiple objectives/criteria. For example, a traveler could be interested in saving time, money, and mileage all at the same time. When one criterion cannot be further improved without worsening the other, trade-offs have to be made. One common modeling approach is to use additive utility-based compensatory methods (with/without random term) and to assume decision makers as disutility minimizers.

Transportation systems commonly have significant variability and uncertainty. Liu et al. (2004) classify the sources of this stochasticity to be either from the demand side or the supply side. Several researchers have found that average travel time and monetary cost are not sufficient to explain travelers' decision. Risk preference contributes to travelers' departure time, mode, and route decisions as well (de Palma and Picard, 2005). Therefore, capturing multi-criteria decision-making with risk-preference in dynamic traffic assignment could further improve modeling results. Travel time uncertainty/reliability has been commonly incorporated into decision models either explicitly or implicitly. Explicit methods commonly use measures of variability (*ex-post*) such as variance, standard deviation, and percentiles, to approximate the effect caused by uncertainty/reliability (*ex-ante*), while implicit methods typically transform utility and probability without directly quantifying uncertainty/reliability itself. Ramos et al. (2014) review the state-of-the-art of applying utility theory, prospect theory, and regret theory to investigate travelers' behavior under travel time uncertainty.

Emphasizing on practicality, this chapter proposes a density-based formulation for multi-class multi-criteria dynamic user equilibrium with path-dependent link costs (MMDUE-MP), following it with a Stochastic Quasi-Gradient Projection (SQGP) solution scheme that uses a traffic simulator. The criteria considered in the discussion and the case-study are travel time (link-

additive), monetary cost (non-additive), and travel time uncertainty (path-dependent). An information entropy-based uncertainty measure is proposed due to concerns on using conventional measures such as variability and reliability. The case study shows stochastic but efficient convergence, demonstrates the ability of SQGP to bypass local optima and exemplifies the significant impact of using both path-independent and path-dependent link cost to forecast traffic pattern and toll revenue. Modeling travel time uncertainty as path-dependent link cost rather than non-additive cost also allows strict objective function and tractable analytical form.

The following sections review relevant literature and clarify some basic concepts and terminologies important to this chapter. Then the formulation (traffic density-based) and a simulation-based solution (based on GP algorithm) are introduced. A case study and conclusions immediately follow.

## ***Literature Review***

There have been many attempts on solving bi-/multi-criteria shortest path problem with costs that are not sensitive to users' paths (Climaco and Martins, 1982; Mote et al., 1991; Martins, 1984; Skriver and Andersen, 2000). Chen et al. (2011) categorize travel costs into link-based, origin-based, and path-based, and they propose a solution algorithm for finding reliability-based shortest paths. However, the path travel time variance is calculated based on the summation of link travel time variances without considering typical correlation among adjacent links.

Multi-class multi-criteria static and dynamic traffic user equilibrium with *only path-independent cost* have been studied along with the development of bi-/multi-criteria shortest path algorithms. Yang et al. (2004) extend the Beckman's static traffic user equilibrium formulation to incorporate heterogeneous users and multiple criteria. Tan et al. (2014) examine the Pareto efficiency of various reliability-based traffic equilibria and the risk-taking behavior of travelers. Bliemer et al. (2003) propose a quasi-variational inequality formulation for multi-class dynamic

traffic assignment. Zhang et al. (2013) use a Probit-based model to consider the perception error for heterogeneous users' two possible criteria (average travel time and toll) for making route choice decisions. Huang et al. (2007) use a Logit-based model to capture the stochastic traffic equilibrium as well.

There have been pioneering research efforts that attempted to incorporate perceived reliability into dynamic traffic assignment with heterogeneous users. Jiang et al. (2011) propose an algorithm for minimizing a general cost gap function, combining previous research on bi-criterion traffic assignment, inequality formulations and solutions, path-restricted multi-class DUE solution methods, and the Column Generation-based DUE algorithm. They even apply it to a metropolitan network to demonstrate the feasibility. However, several points are worth attention. First, it might be more appropriate to name the standard deviation of travel times (regressed using average travel time and travel distance) as an indicator of travel time *unreliability* rather than *reliability*. Second, reliability-based measures (even if scaled by travel distance) commonly have a high correlation with travel time, and therefore, their estimation could be inefficient and even cause multicollinearity. Third, such a reliability measure does not consider specific route conditions such as the local street and freeway portions that a path goes through, which could be a contributing factor to the relatively low regression performance indices. Fourth, since the cost function uses travel time reliability with a positive coefficient, the formulation does not capture the travelers' common risk-prone behavior. Although some studies showed that a constant risk coefficient would work, this could be because survey participants adjusted their departure time to maintain some degree of reliability. Fifth, since network skimming is performed *after* the network reaches path-restricted DUE at a given outer-loop iteration, it could be less likely to find a new path, and therefore, the solution could be easily "trapped" in a local optimum, despite the oscillations being less. Last, the process does not "drop" any paths that are, in later iterations, found to be unlikely to be used, and therefore, the data storage becomes increasingly burdensome. Previous papers that

consider uncertainty and reliability lack the consideration of travelers' preferred arrival times (PAT) and the activity chains.

As shown by Carey (1992), regular homogeneous-user single-criteria DTA is nonconvex. Adding heterogeneous users and different types of perceived cost only further complicate the convexity of the objective function. Therefore, it is necessary to develop a method that can bypass local minima and improve convergence.

### ***Important Concepts***

Before moving on to the multi-class multi-criteria DUE formulation, several concepts are worth being clarified.

#### **Path-dependent Link Cost**

The word “multi-criteria” in this chapter means two folds. First, the general cost is a combination of two or more criteria (e.g., travel time, monetary cost, and travel time uncertainty). Second, some link costs are path-independent (e.g., average travel time and zone-based toll), while some are path-dependent (variance of path travel time with high correlated link travel times). Though the result of summing path-dependent link costs resembles “non-additive path cost,” it is different, in that the impedance of each link can be explicitly stated by conditioning it on the given path, which yields a strict objective function in a tractable analytical form.

Here is an example of path-dependent link cost. Suppose that links 1, 2, 3 in the network as in Figure 5.1 have average travel times of 5.0, 4.2, and 9.0 (all eastbound), and that the travel times

on the three links have a covariance matrix  $\begin{bmatrix} 2 & -0.5 & 0.3 \\ -0.5 & 1.5 & -0.2 \\ 0.3 & -0.2 & 3 \end{bmatrix}$ . The average travel time over L1

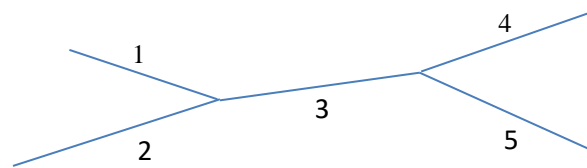
and L3 would be  $\overline{TT}_{L1+L3} = 5 + 9 = 14$  and on L2 and L3 would be  $\overline{TT}_{L2+L3} = 4.2 + 9 = 13.2$ . For

variance,  $\sigma_{L1}^2 = 2$ ,  $\sigma_{L2}^2 = 1.5$ , while  $\sigma_{L1+L3}^2 = \sigma_{L1}^2 + \sigma_{L3}^2 + 2\sigma_{L1,L3} = 2 + 3 + 2 \cdot 0.3 = 5.6$ ,  $\sigma_{L2+L3}^2 =$

$1.5 + 3 - 2 \cdot 0.2 = 4.3$ , which means the increase of variance for travelers from L1 is 3.6, while it is

2.8for travelers from L2. The path-dependency in this example is due to the travel time covariance among links. Now, suppose that link 3 is a tolled link. A correct optimization of the toll for reducing travel time variability for both requires information not only on the toll link itself for different users but also on the' predecessor links (Link 1 and 2) and successor links (Link 4 and 5) on the users' own paths. One immediate implication is that studying the willingness-to-pay for travel time reliability improvement should consider more than just users' travel time distribution on the toll segments.

Similarly, the eastbound variance-based cost on L4 (or L5) depends on whether the traveler comes through L1 and L3 or through L2 and L3. Note that two travelers might not experience the same travel times on L3 when they entered link 1 and 2 at the same time due to different link travel times.



**Figure 5.1** Example of path-dependent link-additive cost

### **Uncertainty, Reliability, Risk, and Variability**

Definitions of uncertainty, reliability, risk, and variability have not been clearly set in the field of transportation modeling, and they are often used interchangeably. In this chapter, the state of uncertainty associated with a set of perceived alternatives is strictly quantified, as per Information Theory, as the amount of information needed to eliminate the current perceived uncertainty for a percept (e.g., travel time) or a set of percepts (e.g., as occurs while driving on a road on a foggy day). Possible interpretations of a percept (or a set of percepts) must be *simultaneously* in the observer's mind for him/her to perceive uncertainty. Reliability is defined as the perceived probability of favorable outcome(s), and risk is defined as the expected value of the unfavorable outcome(s). Different from reliability and risk, variability is *ex-post*. For instance, a bus service that alternates

between 10 minutes and 30 minutes in headway based on a schedule, has high variability but no uncertainty for a regular commuter, if there is no non-recurring delay.

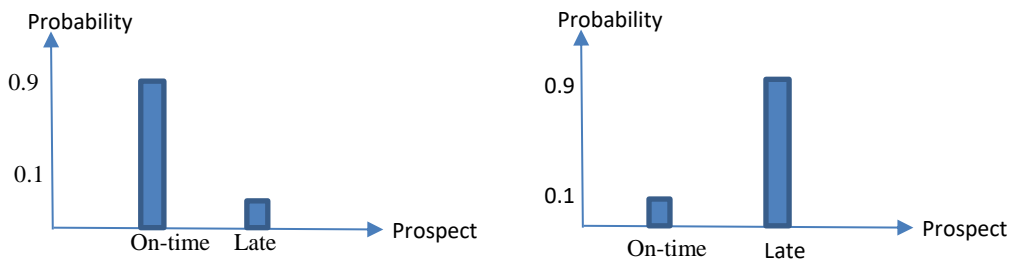
Suppose a traveler believes {On-time: 0.4; Late: 0.6} his/her perceived reliability to be proportional to the probability of being on time (favorable outcome), and his/her perceived risk to be proportional to the expected travel time of being late (unfavorable outcome). However, this probability distribution has the same uncertainty as that in {On-time: 0.6; Late: 0.4}. If an observer perceives these two distributions to have different uncertainty, the outcome setting does not match that of the observer and should be adjusted by the modeler. Using measures such as variance and standard deviation essentially assumes that the observer can differentiate between any two events (no matter how similar they are) and can store an infinite number of events *simultaneously* in his/her/its brain (i.e., infinite brain capacity). Understanding how a traveler perceives has indeed been a challenge, but it does not mean that this issue should be avoided if improving the modeling of users' behavior is the goal. Table 5.1 shows the definition of variability, uncertainty, reliability, and risk used in this chapter, along with their common measures. Note that using strict definitions of these four terms does not generate any new issue – it only makes explicit the assumptions that used to be implicit and assists constructive discussion among modelers.

**Table 5.1** Clarifying definitions of variability, uncertainty, reliability, and risk.

	<i>Ex-ante</i> or <i>Ex-post?</i>	Interpretation	Common Measure
Variability	<i>Ex-post</i>	The quality of being subject to variation or lacking uniformity	Variance (Standard Deviation) of the sample or the population, Percentile, diversity index, etc.
Uncertainty	<i>Ex-ante</i>	A state of doubt of which perceived outcome will happen	Entropy
Reliability	<i>Ex-ante</i>	The degree of confidence about the favorably-perceived outcome to happen	Probability of favorably-perceived outcome
Risk	<i>Ex-ante</i>	Positively related to both the confidence and the disutility of the unfavorably-perceived outcome(s)	Expected cost/disutility of the unfavorably-perceived outcome(s).

### ***Entropy-based Measure of Perceived Uncertainty***

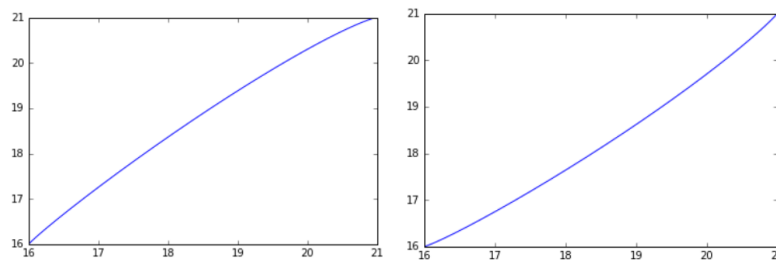
How do we measure uncertainty? There have been many attempts to measure uncertainty and reliability and to incorporate them into general cost functions. This chapter uses the strict definition of uncertainty as in Information Theory, in that it is quantified by the amount of information needed for an observer to be certain about a percept or compound percepts. Claude Shannon (1949) names  $H = \sum_j p_j \log\left(\frac{1}{p_j}\right)$  as information entropy (where  $p_j$  is the probability of the category  $j$  to happen) and mathematically demonstrates that if an uncertainty measure that satisfies three basic conditions,  $H$  is the *only* appropriate measure. In the case of a traveler, information for the traveler’s percept (e.g., travel time) can be measured by the change or divergence of the perceived uncertainty for that percept. This chapter focuses on travel time uncertainty, though it can be extended to multiple percepts (e.g., the perceived uncertainty of the composite space of travel time, monetary cost, and comfort level). Using uncertainty rather than reliability-/variability-based measures can also capture risk-preference conveniently without issues of correlation with average travel time. Figure 5.2 shows an example, where the traveler perceives *low* uncertainty in both cases – on the left, he/she is quite certain to be *on-time*, while on the right, he/she is quite certain to be *late*. The traveler prefers the left over the right due to its lower average travel time. Thus, unlike using reliability or variation as the criterion in the cost function, common high correlation with average travel time can be avoided and the Value-of-Uncertainty (VOU) can be estimated efficiently without bias. Note that some modelers categorize decision with (subjective) probability as decision under risk. I refer to this as “under uncertainty” instead of “under risk” to be consistent with the convention of Information Theory.



**Figure 5.2** Example of a traveler’s category-based perceived travel time distribution

In a case where a traveler uses two criteria (travel time and travel time uncertainty), one can obtain the general cost (or disutility),  $GC = \overline{TT} + \beta \cdot H$ , by assuming utility additivity.  $\beta$  can be set to  $\lambda \cdot \overline{TT}$  to scale the entropy to the same unit of the general cost, where  $\lambda$  is a constant for the traveler. In Chapter 4, this entropy-based general cost formulation is viewed as a special case of a disutility function with elastic surprise considered for each possible outcome.

In Figure 5.3, the x-axis represents the average travel time  $\overline{TT}$  (minute), while the general cost (minute) is  $\overline{TT} \cdot (1 + \beta \cdot H)$  on the y-axis. The left graph shows that when  $\beta > 0$ , the traveler makes a risk-averse decision. The right side shows that when  $\beta < 0$ , the traveler makes a risk-prone decision. This is because when  $\beta > 0$  ( $\beta < 0$ ), the traveler values uncertainty as a positive (negative) factor. When  $\beta = 0$ , the traveler is risk-neutral and the formulation “collapses” into just average travel time. Uncertainties in both cases are maximized at  $\overline{TT} = 18.5 \text{ min}$ . Note that this example only measures the cost/disutility change caused by uncertainty. Other factors such as misperception could also affect how a decision maker weighs the cost/disutility of an alternative.

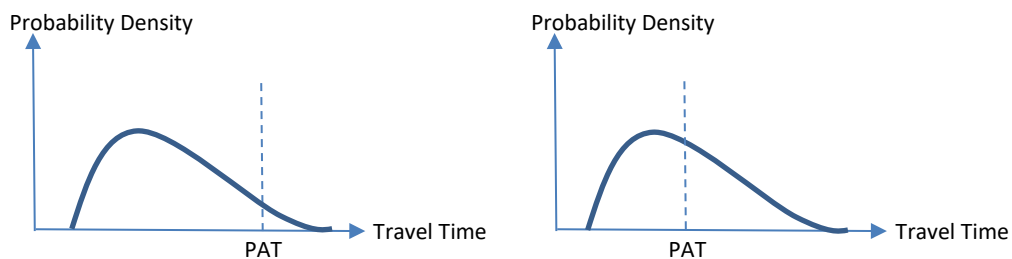


**Figure 5.3** Travel Time (x-axis) vs. General Cost with Risk Preference (y-axis). The left is risk-averse and the right risk-prone.

Factors that influence how travelers set the cognitive categories need more study. In this chapter, the Preferred Arrive Time (PAT) of the traveler’s activity is used as the main factor. This not only eases the modeling effort and post-analysis interpretation but also creates a natural interface between uncertainty-incorporated DTA and Activity-based models (ABM). Travel time distribution is approximated using empirical evidence. Other factors that may influence the cost of the perceived uncertainty, such as departure time and the weights in the general cost function, are



assumed to be user-specific and exogenous. Figure 5.4 shows the importance of PAT to perceived uncertainty: a traveler perceives the same travel time distribution (i.e. the posterior distribution after updating the memory with newly observed information) in both cases, but the traveler on the right perceives higher uncertainty than the one on the left due to their different PATs. As might be expected, the perceived uncertainty could still have a high impact on the left-side scenario if this traveler imposes a greater weight on the perceived uncertainty in his/her/its “general cost function.”



**Figure 5.4** Influence of preferred arrival time and travel time distribution on perceived uncertainty.

One possibility that is worth further investigation is the use of the logarithmic form of the term  $\log\left(\frac{1}{p_j}\right)$  in the entropy-based uncertainty measure. This term measures the reduction of the amount of perceived uncertainty (for a given percept) for an observer if the outcome  $j$  happens. However, perceived magnitude might not be in a logarithmic manner, especially when the corresponding outcome is perceived with small probability. That is,  $\log\left(\frac{1}{p_j}\right)$  approaching infinity when  $p_j$  approaches zero may be unrealistic. Therefore, a modified form might be needed to better capture the observer’s cognition.

It is important to realize that travelers typically “adjust” their travel time risk (or reliability) by making decisions on departure time, mode, and route together to reduce perceived disutility. The weight associated with uncertainty may vary depending on the importance of the activity of the traveler. A traveler might also leave earlier than the optimal departure time for the activity due to the time constraints on the following activities. Therefore, departure time, trip purpose, and even

the schedule of the entire day should be incorporated in future work to fully capture the effect of perceived uncertainty in a traveler's decision.

### ***Single-Class Single-Criterion DUE Formulation***

This section first introduces some notations that are used in a single-class *density-based* formulation with only path-independent link cost (average travel time). The following list contains notations that will be used in the rest of the chapter.

- $a$ : directed link  $a \in A$ , where  $A$  is the set of directional links in the time-dependent network,  $G$ , under study.
- $L_a$ : the length of link  $a$ .
- $x_a$ : a point location  $x$  distance away from the entrance of link  $a$
- $t$ : moment  $t \in T$ , where  $T$  is the temporal planning horizon
- $\tau$ : departure time moment  $\tau$
- $\rho_{x_a,t}$ : density at location  $x_a$  at time  $t$ , so that  $\rho_{x_a,t} dx$  is the number of vehicles at segment  $[x, x + dx]$  at time moment  $t$ .
- $\rho_{rs\tau}$ : density at the origin  $r$  departing at time moment  $\tau$  and going to destination  $s$ .
- $\rho_{rs\tau k}$ : density at the origin  $r$  departing at time moment  $\tau$  going to destination  $s$  and using path  $k$  in the path set  $K_{rs\tau}$ .  $\rho_{rs\tau k} dx$  is the amount of traffic generated at the origin from  $r$  to  $s$  through  $k$  during  $[\tau, \tau + d\tau]$ . Figure 5.5 shows the trajectory of  $\rho_{rs\tau k} dx$  in the time-space.
- $c_{x_a,t}(\rho)$ : cost rate at location  $x$  at time  $t$ .  $C_{x_a,t}(\rho) = c_{x_a,t}(\rho) dx dt$  is the cost (travel impedance) when traffic passes through space-time  $(x_a, t)$ . When it considers the nonlocal traffic condition when it is a function for systemwide state vector  $\boldsymbol{\rho}$ .
- $g(\cdot)$ : traffic generation (dissipation) rate (unit: vehicle/time/length).

$\delta_{rstk}^{x_a t}$ : the proportion of traffic departures from  $r$  to  $s$  through  $k$  during  $\tau$  and at space time  $(x_a, t)$

An objective function along with the constraints can be formulated as

$$\min Z = \int_t \sum_a \int_0^{L_a} \int_0^{\rho_{x_a, t}} c_{x_a, t}(\omega) d\omega dx dt \quad (5.1)$$

Subject to:

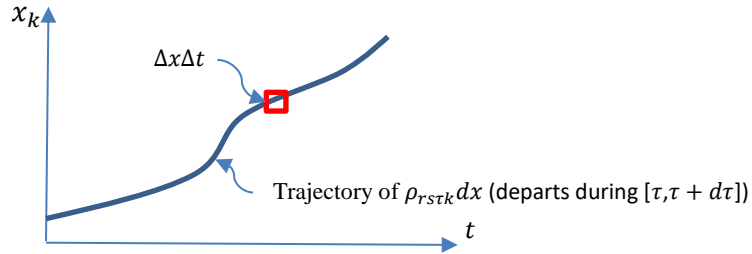
$$\sum_{k \in K_{rst}} \rho_{rstk} = \rho_{rst}, \forall r, s, k, \tau \quad (5.2)$$

$$\rho_{x_a, t} = \sum_{rstk} \rho_{rstk} \delta_{rstk}^{x_a t}, \text{ where } \delta_{rstk}^{x_a t} = [0, 1], \forall a, x_a, t \quad (5.3)$$

$$\sum_{x_a} \delta_{rstk}^{x_a t} = 1, \forall t, r, s, \tau, k \quad (5.4)$$

$$\rho_{x_a, t} \in [0, \rho_{x_a, t}^{jam}], \forall x_a, t \quad (5.5)$$

$$\frac{\partial \rho_{x_a, t}}{\partial t} + \frac{\partial (\rho_{x_a, t} v_{x_a, t}(\rho_{x_a, t}))}{\partial x_a} = g(x_a, t) \quad (5.6)$$



**Figure 5.5** The cost of the trajectory of an infinitely small portion of traffic load passing through  $dxdt$  is  $c(\rho_{x,t})dxdt$

Since the constraint (5.2) refers to density,  $\delta_{rstk}^{x_a t}$  is not binary but continuous between 0 and 1. However, when referring to the same “chunk” of traffic, all the  $\delta_{rstk}^{x_a t}$  that associated with this “chunk” of traffic at a given time step should always add up to 1 (i.e., the constraint (5.4).) For example, suppose there are five vehicles appeared on the starting road segment of a route at time  $t$ ,

at time  $t + \Delta t$ , three vehicles are only the road segment  $L_1$  and two vehicles are on the road segment  $L_2$ , then  $\delta_{rstk}^{L_1(t+\Delta t)} = 0.6$  and  $\delta_{rstk}^{L_2(t+\Delta t)} = 0.4$ .

Other constraints on traffic characteristics such as state continuity and the maximum of non-negative speed are also critical but can be inherently considered in a traffic simulator, and are therefore not specified in the formulation. Figure 5.5 shows an example of a trajectory of a small “chunk” of traffic on path  $k$  at post-mile  $x_k$ . When this “chunk” is infinitely small, an integration of density over the whole time-space would be the traffic itself.

Supposing that  $c(\cdot)$  is continuous in  $[0, +\infty]$ , monotonically increasing, and first-order differentiable, we would have

$$\frac{\partial Z}{\partial \rho_{rstk}} = \frac{dZ}{d\rho_{x_a t}} \cdot \frac{\partial \rho_{x_a t}}{\partial \rho_{rstk}} = C_{rstk} \quad (5.6)$$

The second-order derivative of  $Z$  on  $\rho_{rstk}$  is

$$\frac{\partial^2 Z}{\partial \rho_{(rstk)_i} \partial \rho_{(rstk)_j}} = \int_t \sum_a^{L_a} \int_0^{\rho_{x_a t}} \frac{\partial c_{x_a t}(\rho_{x_a t})}{\partial \rho_{x_a t}} \delta_{(rstk)_i}^{x_a t} \delta_{(rstk)_j}^{x_a t} dx dt, \forall i, j \quad (5.7)$$

where  $(rstk)_i$  is the  $i_{th}$   $rstk$  combination. The formulation (1) has unique optimum when the Hessian matrix with element  $(i, j)$  as  $\frac{\partial^2 Z}{\partial \rho_{(rstk)_i} \partial \rho_{(rstk)_j}}$  is greater than zero.

Given the convex cost function, one can set up the Lagrangian for the problem as  $L = Z + \sum_{rst} \lambda_{rst} (\rho_{rst} - \sum_{k \in K_{rst}} \rho_{rstk})$  and  $\frac{\partial L}{\partial \rho_{rstk}} = \frac{\partial Z}{\partial \rho_{rstk}} - \lambda_{rst} = C_{rstk} - \lambda_{rst}, \forall r, s, \tau$ . The corresponding Karush-Kuhn-Tucker conditions, at optimum where  $\rho_{rstk} = \rho_{rstk}^*$ , become: if  $\rho_{rstk}^* = 0$ , then

$\frac{\partial L}{\partial \rho_{rstk}} \geq 0$ ; if  $\rho_{rstk}^* \geq 0$ , then  $\frac{\partial L}{\partial \rho_{rstk}} = 0$ . This is equivalent to:

$$C_{rstk} > \lambda_{rst}^*, \forall r, s, \tau, k, \text{ when } \rho_{rstk}^* = 0,$$

$$C_{rstk} = \lambda_{rst}^*, \forall r, s, \tau, k, \text{ when } \rho_{rstk}^* > 0.$$

Therefore, the solution satisfies Wardrop's first principle in the dynamic case. Although  $c(\cdot)$ , as a function of density, is convex, the Hessian matrix might not be strictly positive. Therefore, the formulation might not be strictly convex, and uniqueness can of course not be guaranteed.

The speed-density relationship to use depends on the modeler's assumptions. A simple continuum relationship ( $v = v(\rho)$ ), a higher order continuum relationship, or a selected traffic simulator can be used as alternatives for establishing this relationship. Also dependent on the modeler's assumption is the "First-In First-Out (FIFO)" requirement. Although typical DTA practice prefers a model to satisfy FIFO, the formulation is adjustable to reflect better actual traffic conditions. In a case where the cost is not only a function of the local density but also density in other locations (e.g., drivers or autonomous/connected vehicles determine their speed/acceleration based on non-local traffic conditions),  $C(\cdot)$  can be a function of  $\rho$ , where  $\rho$  is a density state vector that stores density information (available to this observer) throughout the network under study.

### ***Multi-Class Multi-Criteria Extension With Path-Dependent Link Cost***

Since perceived travel time uncertainty can be considered as a path-dependent link cost, the single-criterion single-class formulation described given previously can be extended.

$$\min Z = \int_t \sum_a \int_0^{\rho_{x_a,t}} \left[ \int_0^{\rho_{x_a,t}} \bar{t}_{x_a,t}(\omega) d\omega + \sum_{\forall rstk m} \rho_{rstkm} \delta_{rstk}^{x_a,t} (\pi_{C,\overline{TT}}^m \cdot c_{x_a,t}^{Toll} + \pi_{H,\overline{TT}}^m \cdot h_{x_a,t}^{rstkm}) \right] dxdt \quad (5.8)$$

$h$  is the "entropy rate" so that  $H_{xt}^{rstkm} = \int_t \int_x h_{xt}^{rstkm}(\tilde{x}, \tilde{t}) d\tilde{x} d\tilde{t}$ . User class  $m \in M_{krst}$  or  $M$  depends

on whether  $m$  considers path-dependent link costs.  $\pi_{C,\overline{TT}}^m = \frac{\beta_C^m}{\beta_{\overline{TT}}^m}$  and  $\pi_{H,\overline{TT}}^m = \frac{\beta_H^m}{\beta_{\overline{TT}}^m}$ , and they scale

monetary cost and entropy cost to that of travel time.  $\beta_{\overline{TT}}^m$ ,  $\beta_C^m$ , and  $\beta_H^m$  are the coefficients in the

linear general cost function. Constraints are similar to that in the single-class case except that each

constraint should be user class-based. One can prove that the optimal solution of  $Z$  satisfies DUE.

DSO can be obtained easily by replacing the cost functions with their first-order derivatives (i.e. their marginal cost functions).

Travel time is modeled as an *endogenous* variable in the formulation, while the monetary and uncertainty costs are modeled to be exogenous. In cases where there is a relationship between endogenous and exogenous variables (e.g., adaptive tolling scheme), an iterative process can adjust the exogenous variables based on the feedback from the resultant endogenous variables.

### ***Stochastic Gradient Project Based Solution***

Combining constraint set (10) and (11), we can obtain a transformed and dimensionally-reduced objective function that includes the traffic densities on links of only the non-shortest paths (non-SPs) by recognizing that  $\rho_{x_{a,t}} = \sum_{r\tau\tilde{k}m} (\rho_{r\tau\tilde{k}m} \delta_{r\tau\tilde{k}}^{x_{a,t}}) + \sum_{r\tau m} \rho_{r\tau\bar{k}m} \delta_{r\tau\bar{k}}^{x_{a,t}}$  and  $\rho_{r\tau\tilde{k}m} = \rho_{r\tau m} - \sum_{\tilde{k} \in \tilde{K}_{r\tau m}} \rho_{r\tau\tilde{k}m}$ , where  $\tilde{k} \in \tilde{K}_{r\tau}$  is one of the non-SP and  $\bar{k} = \bar{K}_{r\tau m}$  is the shortest path (SP) given  $r\tau m$ .

Transforming the original objective function  $Z(\boldsymbol{\rho}_{x_{a,t}})$  to  $\tilde{Z}(\boldsymbol{\rho}_{r\tau\tilde{k}m})$ , the first-order derivatives (Jacobian) with respect to density  $\rho_{r\tau\tilde{k}m}$  becomes

$$\frac{\partial \tilde{Z}}{\partial \rho_{r\tau\tilde{k}m}} = \frac{\partial \tilde{Z}}{\partial \rho_{x_{a,t}}} \cdot \frac{\partial \rho_{x_{a,t}}}{\partial \rho_{r\tau\tilde{k}m}} = GC_{r\tau\tilde{k}m} - GC_{r\tau\bar{k}m}, \forall r, s, \tilde{k} \in \tilde{K}_{r\tau}, \bar{k} = \bar{K}_{r\tau m} \quad (5.9)$$

Let  $S$  be the second-order derivative (Hessian) of the objective function. The element,  $S_{(r\tau\tilde{k}m)_i(r\tau\tilde{k}m)_j}$ , therefore, is

$$\frac{\partial^2 \tilde{Z}}{\partial \rho_{(r\tau\tilde{k}m)_i} \partial \rho_{(r\tau\tilde{k}m)_j}} = \int_t \sum_a \int_0^{L_a} \frac{tt_a(\rho_{x_{a,t}})}{\partial \rho_{x_{a,t}}} \left( \delta_{(r\tau\tilde{k})_i}^{x_{a,t}} - \delta_{(r\tau\bar{k})_i}^{x_{a,t}} \right) \left( \delta_{(r\tau\tilde{k})_j}^{x_{a,t}} - \delta_{(r\tau\bar{k})_j}^{x_{a,t}} \right) dxdt, \forall i, j \in I \quad (5.10)$$

Set  $I$  stores all the possible  $r\tau\tilde{k}m$  combinations. According to (5.10), we can calculate the element  $(i, j)$  of the Hessian matrix by summing all the  $\Theta_{i,j}^{x_{a,t}} \cdot \frac{tt_a(\rho_{x_{a,t}})}{\partial \rho_{x_{a,t}}}$ , where  $\Theta_{i,j}^{x_{a,t}} = \left( \delta_{(r\tau\tilde{k})_i}^{x_{a,t}} - \right.$

$\delta_{(rst\tilde{k})_i}^{x_a,t} \left( \delta_{(rst\tilde{k})_j}^{x_a,t} - \delta_{(rst\bar{k})_j}^{x_a,t} \right)$  equals to -1, 0, or 1 depending on the relationship of  $(rst\tilde{k})_i$ ,  $(rst\bar{k})_i$ ,  $(rst\tilde{k})_j$ , and  $(rst\bar{k})_j$  with  $(x_a, t)$ . These relationships are shown in Table 5.2.

**Table 5.2** Relationship of  $(rst\tilde{k})_i$ ,  $(rst\bar{k})_i$ ,  $(rst\tilde{k})_j$ , and  $(rst\bar{k})_j$  to  $(x_a, t)$  for deciding  $\Theta_{i,j}^{x_a,t}$  according to (5.10)

	$(rst\tilde{k})_i$	$(rst\bar{k})_i$	$(rst\tilde{k})_j$	$(rst\bar{k})_j$	$\Theta_{i,j}^{x_a,t}$
On $(x_a, t)$ ?	Yes	No	Yes	No	1
	No	Yes	No	Yes	
	Yes	No	No	Yes	-1
	No	Yes	Yes	No	
The rest 12 relationships					0

Summarizing (5.10) and Table 5.2, the element,  $S_{(rst\tilde{k}m)_i(rst\tilde{k}m)_j}$ , or  $S_{ij}$ , in the second-order derivative (Hessian) of the objective function, is determined by the addition of positive or negative marginal link travel time conditioning on the following conditions: the coefficient is 1 when the link is on  $(rst\tilde{k})_i$  and  $(rst\tilde{k})_j$  but not on their corresponding shortest path(s); the coefficient is also 1 when the link is on  $(rst\tilde{k})_i$  and  $(rst\bar{k})_j$  but not on their corresponding non-shortest path(s); the coefficient is -1 when the link is on  $(rst\tilde{k})_i$  and  $(rst\bar{k})_j$  but on neither  $(rst\tilde{k})_i$  nor  $(rst\tilde{k})_j$ ; the coefficient is -1 when the link is on  $(rst\bar{k})_i$  and  $(rst\tilde{k})_j$  but on neither  $(rst\tilde{k})_i$  nor  $(rst\bar{k})_j$ ; the coefficient is 0 in any other case.  $S_{ii}$  (or  $S_{\tilde{k}}$ ) is  $i$ th diagonal element of the Hessian matrix. Note that two routes have to be both spatially and temporally on a location to be considered overlapping at this location.

One special case is when  $i=j$  (i.e.,  $S_{ii} = \int_t \sum_a \int_0^{L_a} \frac{tt_a(\rho_{x_a,t})}{\partial \rho_{x_a,t}} \left( \delta_{(rst\tilde{k})_i}^{x_a,t} - \delta_{(rst\bar{k})_i}^{x_a,t} \right)^2$ ), so that  $\Theta_{i,i}^{x_a,t} = 1$  when either  $(rst\tilde{k})_i$  or  $(rst\bar{k})_i$  is on  $(x_a, t)$  but not both; otherwise,  $\Theta_{i,i}^{x_a,t} = 0$ .

Since the Hessian matrix might not be strictly positive, the objective function is not necessarily convex. However, one can still utilize the gradient with certain mechanisms to search

for the global optimum without being “trapped” into the local one. As can see, existing methods that use (5.9) as the main objective runs the risk of “converging” at a reflection point.

Projection is the step in which the path loads that are negative is made zero after the gradient-based update, with the vehicles then placed in the corresponding SPs to satisfy non-negative demand constraints. In mesoscopic/macrosopic simulation-based methods, however, traffic is formed with individual vehicles, and therefore, the non-negativity constraint for “traffic load” is automatically satisfied. I propose a method that stochastically switches vehicles from their non-SPs to their corresponding SPs. Suppose there are  $N_{\tilde{k}}^n$  vehicles on the non-SP given  $r, s, \tau, m$ , in the iteration  $n$ .  $p_{\tilde{k} \rightarrow \bar{k}}^n$  is the proportion to be switched from the non-SP,  $\tilde{k}$ , to the SP,  $\bar{k}$ . The “traffic load” of the next iteration can be estimated as  $N_{\bar{k}}^{n+1} = N_{\tilde{k}}^n - p_{\tilde{k} \rightarrow \bar{k}}^n \cdot N_{\tilde{k}}^n = N_{\tilde{k}}^n - \frac{\alpha^n}{S_{\tilde{k}}^n} (GC_{\tilde{k}^n} - GC_{\bar{k}^n})$ .

Therefore,

$$p_{\tilde{k} \rightarrow \bar{k}}^n = \frac{\alpha^n}{S_{\tilde{k}}^n \cdot N_{\tilde{k}}^n} (GC_{\tilde{k}^n} - GC_{\bar{k}^n}) \quad (5.11)$$

Since  $0 \leq p_{\tilde{k} \rightarrow \bar{k}}^n \leq 1$ , the non-negativity constraint,  $\max\left\{0, N_{\tilde{k}}^n - \frac{\alpha^n}{S_{\tilde{k}}^n} (GC_{\tilde{k}^n} - GC_{\bar{k}^n})\right\}$ , for a traffic simulator is almost trivial.  $p_{\tilde{k} \rightarrow \bar{k}}^n$  can be interpreted as the probability of adjusting a vehicle’s current non-SP to the SP in a simulation.

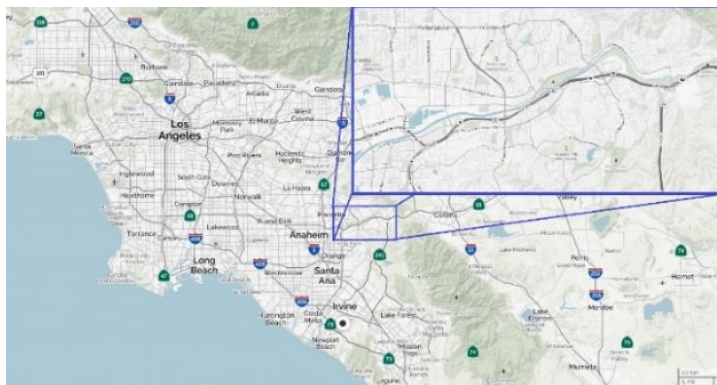
An example might be helpful to understand the derived result above intuitively. Suppose there are two rooms of the same size. Three people are in Room A and two in Room B. Although Room A is more crowded than Room B ( $GC_A > GC_B$ ), none in room A wants to go to Room B unilaterally due to the high elasticity of the crowd level of the two rooms. That is, if a person in Room A moves to Room B, Room B will then become more crowded ( $GC_A < GC_B$ ).

## ***Case Study***

The State Route 91 (SR-91) is a major east-west freeway located in Southern California (Figure 5.6). It currently runs from Vermont Ave. in Gardena (just west of the interchange with the I-110), east

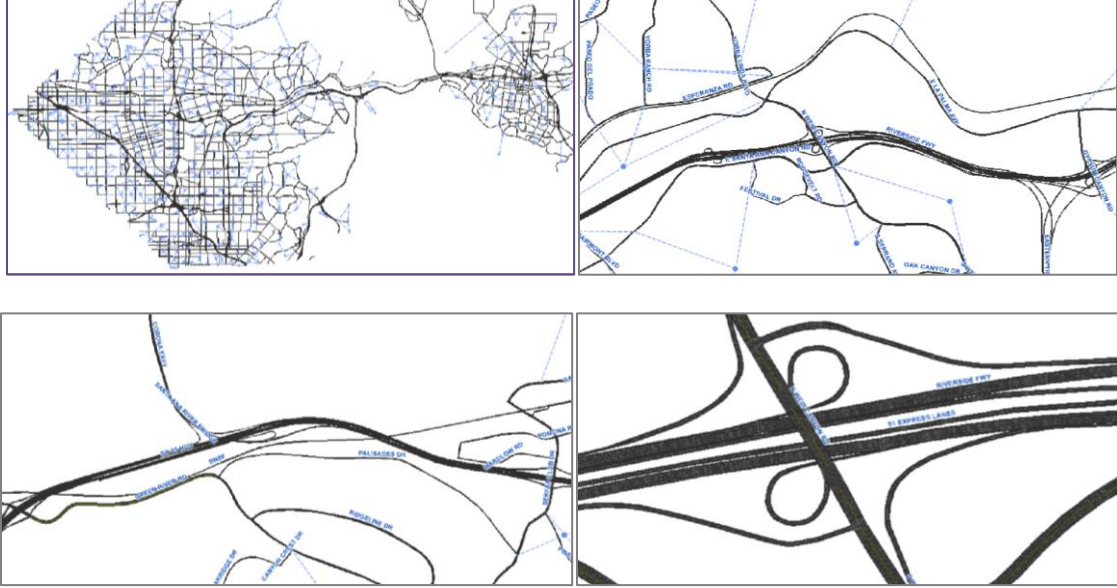


to City of Riverside at the junction with SR-60), and I-215. The segment from the interchange with SR-55 to the Riverside County line is the 91 Express lanes – a ten-mile (16km) high-occupancy toll road. All the tolls are collected using an open-road tolling system, with each vehicle required to carry a FasTrak transponder.



**Figure 5.6** Relative location of the SR-91 HOT Facility to City of Los Angeles and City of Irvine. (Source: MapQuest.com)

The network is built based on data from the Southern California Association of Governments' RTP 2016 model (scenario 3) and the Inventory Database of the California Department of Transportation (Caltrans). For the study area, the static OD matrices for different user classes are obtained based on the identification of 17 gateways, 286 internal zones, and 3 special generators (John Wayne Airport and two amusement parks) of the Southern California Association of Governments' RTP 2016 model. To obtain the time-dependent OD demand, the resultant AM peak period OD demand is split based on 5-min PeMS data from 6:45-8:15 am, April 12 (Tuesday), 2016. The ramp metering data is obtained from the Caltrans. The toll scheme was obtained from the SR-91 Express Lanes official website (on April 12, 2016). The SR-91 Express Lanes system uses a variable pricing scheme based on the time of day. 15-min "warm-up" and "cool-down" simulations are performed before and after the study period, and the time-dependent OD matrices are estimated to capture the count at 5-min aggregation level between 7-8am.

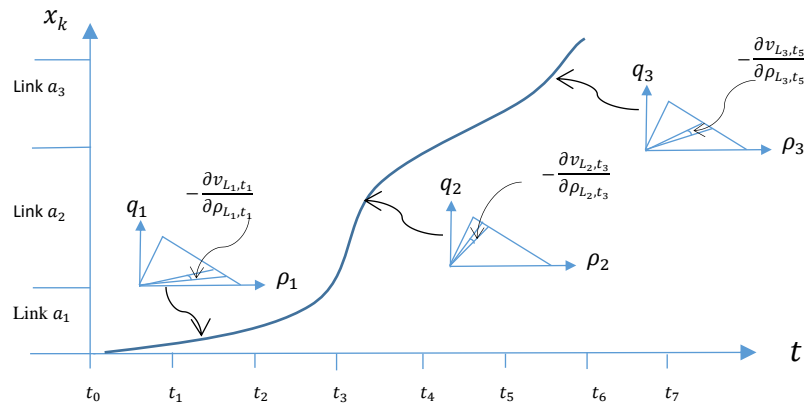


**Figure 5.7** The Model Network. The upper left shows the entire model network; the upper right shows the west end of the SR91 HOT facility; the lower left shows the east end of the SR91 HOT facility; The lower right shows the interchange area of the SR91 and N Weir Canyon Road.

Vehicles are classified based on whether travelers possess real-time traffic information, occupancies of the vehicles (Drive Alone (DA) and HOV), and serving purposes (passenger vehicles and trucks). Informed-HOV and Uninformed-HOV have 60% to be risk-prone ( $\beta_H = 1.25$ ), 25% risk-averse ( $\beta_H = -2.12$ ), and 15% risk-neutral ( $\beta_H = 0$ ). Informed-DA and Uninformed-DA have 80% to be risk-prone, 15% risk-averse, and 5% risk-neutral. Informed-Truck and Uninformed-Truck have 20% to be risk-prone, 70% risk-averse, and 10% risk-neutral. Value-of-Time (VOT) of \$11.5/hour is used as  $\beta_c$  for passenger and \$24.5/hour to convert toll to time. A linear additive general cost function is used. In this case study, the category is set based on two prospects: arriving by or before the PAT and arriving later than the PAT. Due to lack of PAT information, the PAT is set as  $\alpha^m \cdot \overline{TT}, \forall m$ . In this case study,  $\alpha^m = \alpha = 1.2, \forall m$ .

Finding the first-order derivative of the objective function (the cost gap between non-SP and the SP) is relatively convenient to obtain, and yet obtaining the “second-order derivative” requires recognizing that a vehicle’s “path” is in time-space. With this understanding, one can obtain the marginal path cost in every position along the trajectory by evaluating the marginal

change of the cost from a marginal increase of the density without considering the overlapping segments with the corresponding SP. In this case study, the second-order derivative at link-level is approximated using Triangular Fundamental Diagram (TFD), though it is also feasible (though more complex) to incorporate higher-order speed-density relationships. Within the congested regime of the TFD (at a given link and time interval), the marginal cost for the vehicles that depart in time interval  $\tau$  from  $r$  to  $s$  through path  $\tilde{k}$  can be approximated by evaluating the speed changes in terms of the marginal density changes. Within the uncongested regime of the TFD, the cost is not sensitive to the change of the density, as the speed is constant. The second-order derivative then can be obtained by summing the marginal costs on links that are either on the non-shortest trajectory or the shortest trajectory but not on both. An example is illustrated in Figure 5.8.



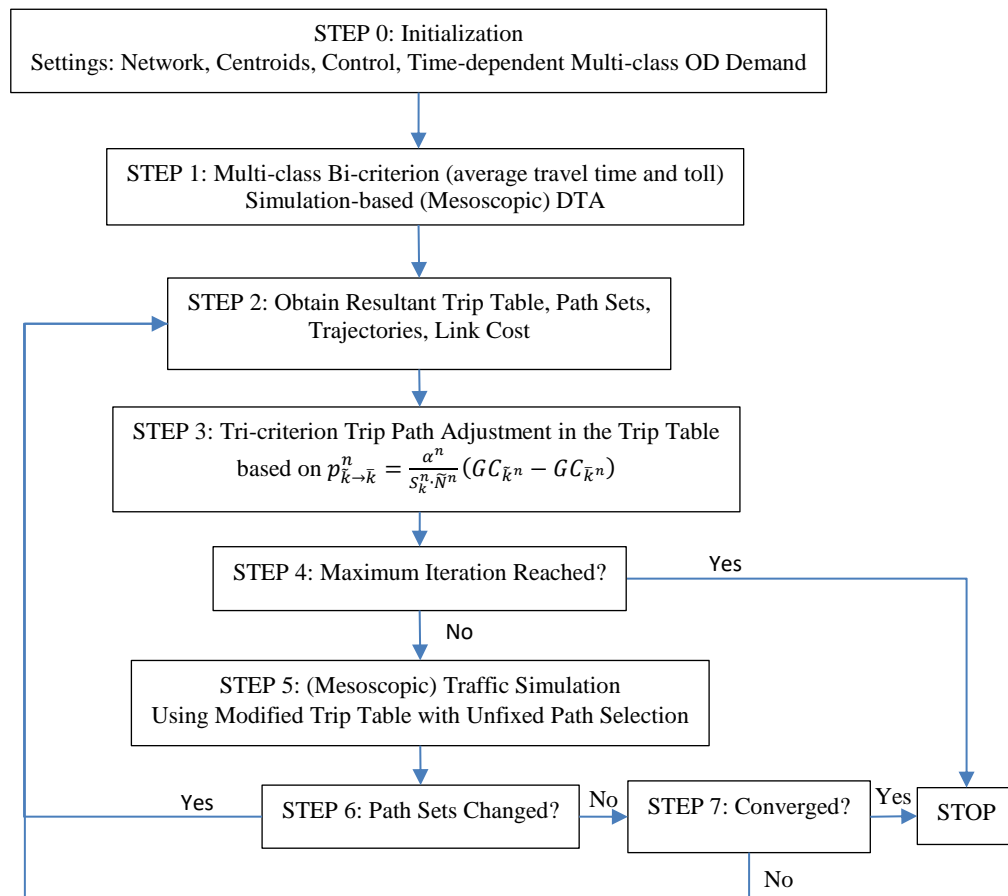
**Figure 5.8** Approximating the (partial) second-order derivative of a traveler's general cost along the trajectory.

The TFD parameters in the case study are set based on the calibration result from a previous work (Dervisoglu, 2015). The capacities of the freeway links and local road links are 2200veh/hr/lane and 1300 veh/hr/ln, respectively. The discontinuity caused by local traffic signal and ramp metering is not considered in this case study.  $\frac{\rho_c}{\rho_j}$  is set to be 0.19 ( $\rho_c$  is the density when the traffic flow reaches capacity and  $\rho_j$  is the jam density). For a fundamental diagram that considers lane-changing, refer to Jin (2010). Trajectory data from simulations are used to obtain

densities on links or sub-links at given time intervals. To maintain computational efficiency, the Hessian matrix is approximated with its diagonal elements.

The process of “switching trips” in this case study is essentially a “trip table adjustment” procedure with the aid of the trajectory data. To prevent the algorithm from being “trapped” at local optimum, the step size is set relatively large in the first iterations than later ones. A brief flowchart is shown in Figure 5.9. A sample from the trip table is shown in Table 5.3. A Path ID will be adjusted to that of its corresponding shortest path ID if the vehicle is selected.

The simulation-based mesoscopic dynamic assignment was performed on TransModeler 4.0 on a machine with Intel Core™ i7-4770 CPU with 3.40GHz, 8.00GB 64-bit operating system.



**Figure 5.9** Flow-chart of the GP solution adaptive to the case study

**Table 5.3** Sample trip table used in the algorithm

Vehicle ID	OriID	DesID	Vehicle Type	$\beta_C$	...	Path ID	Departure Time	Expected Arrival Time	Actual Arrival Time	Toll (\$)
10375	3	7	SOV	3.71	...	6	36549	36649	36651	--
10376	3	6	HOV2	2.44	...	12	36552	36601	36603	--
10377	2	5	HOV3+	2.39	...	4	36555	36597	36588	1.5
10378	1	7	SOV	3.71	...	15	36561	36613	36616	2.4
...	...	...	...	...	...	...	...	...	...	...

## Results Discussion

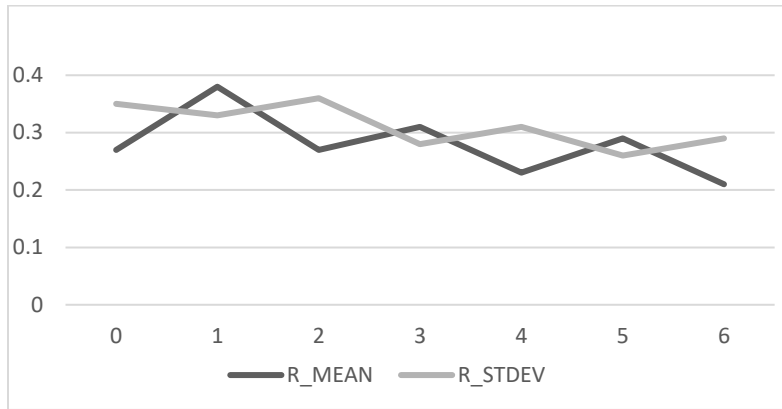
Two user-oriented convergence criteria are used in this case study: the average and the standard deviation of the general cost gap between the cost of the current path and the corresponding SP.

That is,  $R_{MEAN} = \frac{GC - GC_{\bar{k}}}{GC_{\bar{k}}}$  and  $R_{STDEV} = \frac{\sqrt{Var(GC - GC_{\bar{k}})}}{GC_{\bar{k}}}$  in the planning horizon (temporal and spatial),

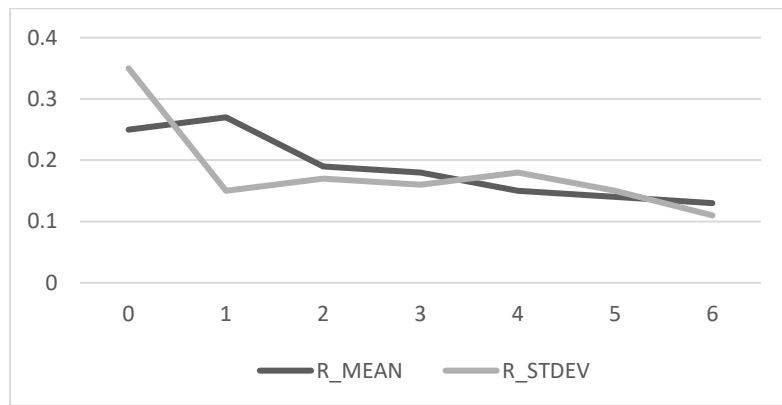
where  $GC$  is treated as a random variable, each realization is that of each user, and  $GC_{\bar{k}}$  is the corresponding SP. The convergence criteria is based on both  $R_{MEAN}$  and  $R_{STDEV}$  being low. Figure 5.10 shows  $R_{MEAN}$  and  $R_{STDEV}$  at each iteration when only consider the first order derivative of the objective function (0.3 as the step size). Figure 5.11 shows  $R_{MEAN}$  and  $R_{STDEV}$  at each iteration when considering both the first order derivative and the second order derivative (only the diagonal elements of the Hessian matrix) of the objective function. The results show a superior convergence tendency when the second order derivative is considered. Notice that the consideration of second-order derivative does not cause significant computational effort for the following two reasons. First, the most computationally-heavy tasks (network skimming and path storage) are already performed for calculating the first-order derivative. Second, the case study approximates  $S_{ii}$  at the link-level when the users entering the study link – the traffic density is calculated as the total vehicles on the study link (when the users entering the link) divided by the multiplication of the total link length and the number of lanes.

Setting the criteria  $R_{STDEV}$  is for avoiding the occasion where a high variation exists in different travelers' general cost gap between the SPs and the non-SPs. The stochasticity of

projecting vehicles, simulating driving behaviors, and trading-off between reducing “average gap” and “gap variation” in the simulation all contribute to the non-monotonic convergence.



**Figure 5.10** Convergence test for both the  $R_{MEAN}$  and the  $R_{STDEV}$  when only considers the first order derivative of the objective function.

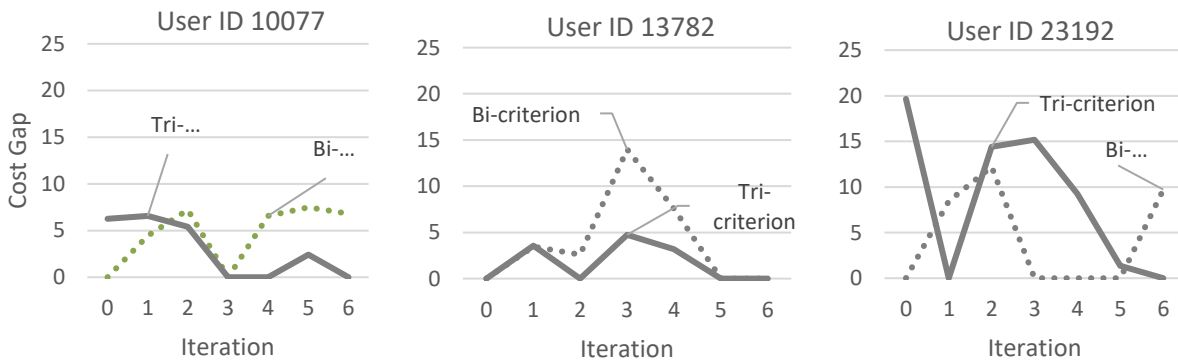


**Figure 5.11** Convergence test for both the  $R_{MEAN}$  and the  $R_{STDEV}$  when consider the first-order and the second-order derivative of the objective function.

Toll revenue forecast under a bi-criterion DUE is \$3142.2002, while \$2867.8243 it is for a tri-criterion DTA that incorporates travel time uncertainty (1.5% HOV facility usage violation is assumed, and neither fines nor enforcement effectiveness is considered). A set of different combinations of coefficients are used, and they produce similar convergence effects; this could be caused by the relatively small size of the network used.

Figure 5.12 shows the “route adjustment” over iterations for three randomly-selected users from different classes, occupancies, OD pairs, departure times, and preferred arrival times (their attributes are shown in Table 5.4). The x-axis is the sequence of iteration while the y-axis the gap

for both the bi-criterion and tri-criterion cost gaps. Bi-criterion cost gaps generally increase over the iterations while tri-criterion cost gaps decrease. However, it is possible that the SP based on the bi-criterion and on the tri-criterion are the same (user 13782). Risk-averse users (e.g., user 10077) unsurprisingly tend to use HOT lanes since average travel time reduction and uncertainty reduction can be achieved simultaneously by the increase of monetary cost. On the other hand, analyzing and predicting route choice is relatively more challenging when the objectives are conflicting, i.e., reducing average travel time and reducing uncertainty simultaneously. This is the case for risk-prone users (e.g., user 13782 and user 23192) because their decisions are highly related to the probability of arriving earlier than the PAT in the alternative route (i.e., HOT lanes). Records on the cost gap and path switching over the iterations are shown in Table 5.5.



**Figure 5.12** Bi-criterion and tri-criterion gap over iterations for user 10077, 13782, and 23192

**Table 5.4** User attributes for 10077, 13782, and 23192

User ID	Origin	Destination	Dep Time	User Class	Vehicle Type	$\beta_{TT}$	$\beta_C$	$\beta_H$	Risk preference
10077	Entrance 1	Exist 6	7:03	2	SOV	1	3.70	1.51	Averse
13782	Entrance 1	Exist 6	7:18	2	HOV2	1	2.44	-1.43	Prone
23192	Entrance 4	Exist 6	7:47	1	SOV	1	3.70	-1.68	Prone

**Table 5.5** Path switch over the iterations for user 10077, 13782, and 23192

User 10077 Iteration	Bi-criterion	Tri-criterion	Current Path ID	SP-bi ID	SP-tri ID	Switch?
0	0.00	6.26	7	7	6	No
1	4.45	6.57	7	6	6	Yes
2	7.21	5.39	6	5	7	Yes
3	0.00	0.00	7	7	7	No
4	6.58	0.00	7	6	7	No

	5	7.49	2.44	7	5	6	Yes
	6	6.81	0.00	6	5	6	--
<u>User 13782</u>							
Iteration	Bi-criterion	Tri-criterion	Current Path ID	SP-bi ID	SP-tri ID	Switch?	
0	0.00	0.00	6	6	6	No	
1	3.47	3.55	6	5	7	Yes	
2	2.61	0.00	7	6	7	No	
3	13.99	4.74	7	6	6	No	
4	7.58	3.20	7	6	6	Yes	
5	0.00	0.00	6	6	6	No	
6	0.00	0.00	6	6	6	--	
<u>User 23192</u>							
Iteration	Bi-criterion	Tri-criterion	Current Path ID	SP-bi ID	SP-tri ID	Switch?	
0	0.00	19.65	11	11	15	Yes	
1	8.45	0.00	15	11	15	No	
2	12.21	14.39	15	14	14	Yes	
3	0.00	15.17	14	14	11	No	
4	0.00	9.25	14	14	15	No	
5	0.00	1.37	14	14	15	Yes	
6	9.72	0.00	15	14	15	--	

Although travel time uncertainty and toll are treated as exogenous, it can be accommodated as endogenous variables by setting an “outer loop” to feed the forecast traffic pattern back to the exogenous variables in the “inner loop.” Variability measures can be easily incorporated similarly to entropy-based methods if variability is preferred as uncertainty proxy.

Modeling heterogeneous users with discrete random distribution has its advantage. First, It is more convenient to capture the correlation among different parameters in large scale simulation, secondly, users are indeed discrete in reality.

To consider the stochasticity of the parameters in the simulation (including the driving behavior parameters) and obtain higher confidence on the assignment results, a batch run is recommended. Step size  $\alpha^n$  in the gradient projection algorithm could be set flexible so as to avoid local optima. One can evaluate DSO conditions by replacing users’ general cost functions with their marginal general cost functions. Although it is a common practice to use linear forms for the perceived general cost, there could be more realistic forms that capture the interaction among different criteria.



## ***Conclusion and Future Direction***

This chapter proposes an entropy-based method to consider the perception of uncertainty and provides a formulation based on traffic density for multi-class dynamic user equilibrium with path-dependent link costs and risk preference. A GP-based solution procedure is tailored for utilizing a microscopic traffic simulator that provides traffic densities via vehicle trajectory outputs. The case study demonstrates the convergence efficiency of using a stochastic quasi-gradient projection method. The “path-switching behavior” for each user over interactions is set in a way to efficiently approaching DUE, and therefore, it does not necessarily contain a behavioral foundation. The scheme considers additive, non-additive, and path-dependent link costs in the dynamic assignment process, even while using approximated Hessian matrices.

The gradient-based solution explains that using “gap function” is its special case and using “gap function” as the main objective leads to the risk of “converging” at a reflection point. Therefore, the second-order derivative of the objective function should be considered. Since the gradients apply to individual vehicles stochastically, the error is “noisier” than that of the regular gradient projection. However, this stochasticity gives the algorithm the potential to “escape” from local minima, especially when computing in a batch mode. Implementing adjustable step size further strengthens the algorithm’s ability in avoiding local optima.

The GP-based solution scheme has two computational advantages over other existing methods: (1) There is no “hidden loop” in each step -- the algorithm only requires reaching bi-criterion DUE in the initialization phase and the rest of the procedures are only one-loop simulations, during which travelers are set to have various abilities and information availability to adjust their path. (2) The algorithm “drops” very unlikely path(s) in each iteration helps save memory and speed up the computing speed. These findings are consistent with that in a computational study examining several static path-based traffic assignment algorithms (23). A future study in a larger non-corridor network will be useful to verify the algorithm’s ability to

generate realistic path sets. The case study intentionally selected a “corridor-shaped” network so that paths can be enumerated to check whether the result contains reasonable path sets.

Several important points are worth emphasizing: (1) variation, uncertainty, risk, and reliability are different, and should not be used interchangeably. (2) A tolling scheme that attempts to consider the reduction of impact on variation and uncertainty change should consider different users’ whole paths (especially predecessor links) rather than only the tolling segments. (3) Uncertainty and reliability are highly related to travelers’ and shippers’ departure time choices and preferred arrival times/time windows, depending on the attitude toward and the importance of each activity and time budget. (4) Despite the goal of reaching dynamic user equilibrium, it does not postulate that there exists any form of DUE in reality. The SQGP-based MDUE-MP solution only gives a suggestion of how the traffic pattern could be given certain prior information and initial condition. The system-optimal versions of the scheme with marginal costs have more relevance in practical control in traffic networks.

There are at least four important future research directions: development of an efficient and effective path-cost-based shortest path algorithm, effective calibration based on both aggregate (e.g., detector station data) and disaggregate data (e.g., sentiment and physical trajectory data) for large networks, research on how to set cognitive categories for travelers and shippers, and an integration of activity-based models (ABM) and dynamic traffic assignment (DTA) through the category-based uncertainty modeling method proposed in this chapter.

## CHAPTER 6 CONCLUSION

*“Governing a country is like frying a small fish. You spoil it with too much poking.”*

*– Lao Zi*

The quote from Lao Zi’s Tao De Ching points out a critical principle about how to deal with complex transportation systems. Modern transportation systems have developed to such a complexity that hardly anyone is able to claim they understand how exactly the systems are functioning due to limited cognitive capacity and effective information availability (for both human and machine.) The ignorance of ignorance leads to the danger that one uses “gut feelings” or “common sense” when trying to improve the systems. For instance, a decision maker might not be aware that he/she does not know enough about the systems to make decisions. Indeed, in practice, it is not uncommon for a decision maker to use a model to only “verify” what he/she thinks, and when the model result does not match with what he/she anticipates, the model must be wrong. On the other hand, although a humbler attitude towards a “bottom-up” modeling approach to support the decision is indeed essential, more valid and theoretically sound assumptions are also desired.

The proposed homeomorphic framework is essentially a feedback system that explicitly considers information sources, channels/media, and recipients with heterogeneous sensation, perception, learning, and decision-making schemes. It not only targets its modeling output to match the reality but also its underlying processes. These recipients learn from the newly-acquired information and incorporate it with prior information to form posterior information for deciding and behaving. Different recipients interact through behaviors, and a new loop begins. Since any information transmission within systems can be modeled in this fashion, modeling autonomous entities such as humans and artificial intelligence becomes unified. The case study is shown how the framework can be used as a guide for object-oriented programming when considering mixed-flow traffic of human drivers and autonomous/connected vehicles.

The cognition-based modeling framework, CognAgent, demonstrates a flexible and robust methodology for researchers and modelers interested in modeling scenarios that involves information dynamics. Four fundamental components within the framework are physical interaction, space of observables, information processing (for sensing, perceiving, and learning), and decision making. An agent's behavior generated from its decision-making submodule interacts with other agents' behaviors in the physical interaction module to form a feedback loop. The size of time step depends on the modeling scale and problem need. The case studies from Chapter 2 shows how the framework helps improve interpretability and computational efficiency of an agent-based model through explicating the underlying cognitive mechanisms rather than only focusing on approximating the revealed behaviors. This transition from a paramorphic to a homeomorphic approach makes explicit the underlying assumptions about how behaviors are made and fosters the development of the rest of the dissertation.

The framework is followed by two major paradigms as further substantiation of its cognition module. The first paradigm (Chapter 3) proceeds in the idea of separating sensation and perception in the cognitive process of judgment so that the perceived information can be quantified as the change of perceived uncertainty. The case study shows that the location, content, and the change of content over time of the travel time information provision has a non-negligible impact on the change of travelers' perceived travel time uncertainty (and, hence, route decisions) in urban traffic networks. What's more, it is suggested that activity schedules and trip purposes also influence how travelers learn and utilize the provided information.

The second paradigm (Chapter 4) can be viewed as a specification of the Decision-Making submodule of the framework. The core concept, Elastic Surprise (ES), bridges existing theories for describing decisions under risk and information theory. ES also has a concrete cognitive implication, which makes it ideal to incorporate into existing cognition-based models. Applying ES to CPT improves the goodness of fit and has explicit cognitive assumptions about a decision maker.

Probability misperception and ES are identified as two distinct but interrelated factors, the combination of which leads to the nonlinearity of weighing potential outcomes. A specific deviation of considering ES in a utility-based decision model under risk bridges this paradigm with the information entropy; when an ES function is assumed to be logarithmic and a decision maker assigns the same weight to the perceived possible outcomes, the overall utility of a choice is proportional to the information entropy. This, in a way, justifies the theory since when a rational observer only perceives the level of chaos of a system and has no preference among the possible outcomes, entropy *should* be equivalent to the utility (or disutility) of this level of chaos. Since a decision maker may not use logarithmic surprise function, a paradox arises. On the other hand, it also implies that a system operator/engineer/planner can avoid bias and inconsistent decisions by (implicitly) evaluating uncertainty using a non-logarithmic ES measure. The case study on the perceived certainty equivalent of travel time shows that jointly using CPT and EST could improve data fitting and feature more specific cognitive interpretation.

Chapter 5 applies the ideas from previous chapters to studying complex transportation network dynamics, in which heterogeneous travelers make route decisions based on multiple criteria under risk. I propose to use path-dependent link cost instead of conventional non-additive cost to allow a multi-criteria multi-class density-based formulation under perceived travel time uncertainty for dynamic user equilibrium. The case study is demonstrated through the equilibrium is approximated through a Stochastic Gradient Projection algorithm. The results show that the toll pricing study should consider not only travel time variability on toll segments but entire paths of the travelers. The calibration of the general entropy-based cost function for route choice uses PeMS data (i.e., the aggregated revealed preference data), while the calibration of the utility function for travel time equivalency in Chapter 4 is performed on EST-Augmented CPT function using survey data. Although the two models are derived from the same theory and individually checked for their

reasonability, closer studies could be conducted on their relationship by taking into account the effect of survey mode and instrument.

There are two perspectives to intuitively understand the stochastic gradient projection-based algorithm proposed to solve the multi-criteria multi-class dynamic user equilibrium. First, the algorithm provides an optimization process for a nonlinear objective function with nonlinear constraints. Second, different from regular agent-based models that have known decision rules, the algorithm for this agent-based model has the known goal state of a system but unknown decision rules. The proposed algorithm, in a way, can be thought of as a procedure to find this unknown decision rule so that agents behave *as if* they collectively achieve this system performance as efficiently as possible.

Since decision makers are an integral part of the transportation systems, they should be part of the modeling practice to describe and predict the systems' behaviors better. This dissertation shows that modeling complex systems at a more fundamental level about humans (e.g., travelers, planners, traffic operators, and shippers) and emerging autonomous entities can be achieved systematically and coherently. Transportation systems have constantly been "interrupted" by emerging business models and technologies, and the modelers and analysts in both practice and academia have been prone to either modify (sometimes dramatically) the modeling framework or adding one extension over another. "Modeling based on the problem" indeed has its advantage, but it also causes a model to lack extendability and adaptability to new business models, travel options, and emerging technologies. The proposed framework studies the revealed behavior from the underlying cognitive mechanics has greater adaptable since a new scenario is considered by "adjusting the parameters."

The dissertation can be extended to multiple directions. The generality of the framework can be further tested in more scenarios and cognitive models. The paradigms can be applied to scenarios that involve cooperative games such as ridesharing and joint decision among

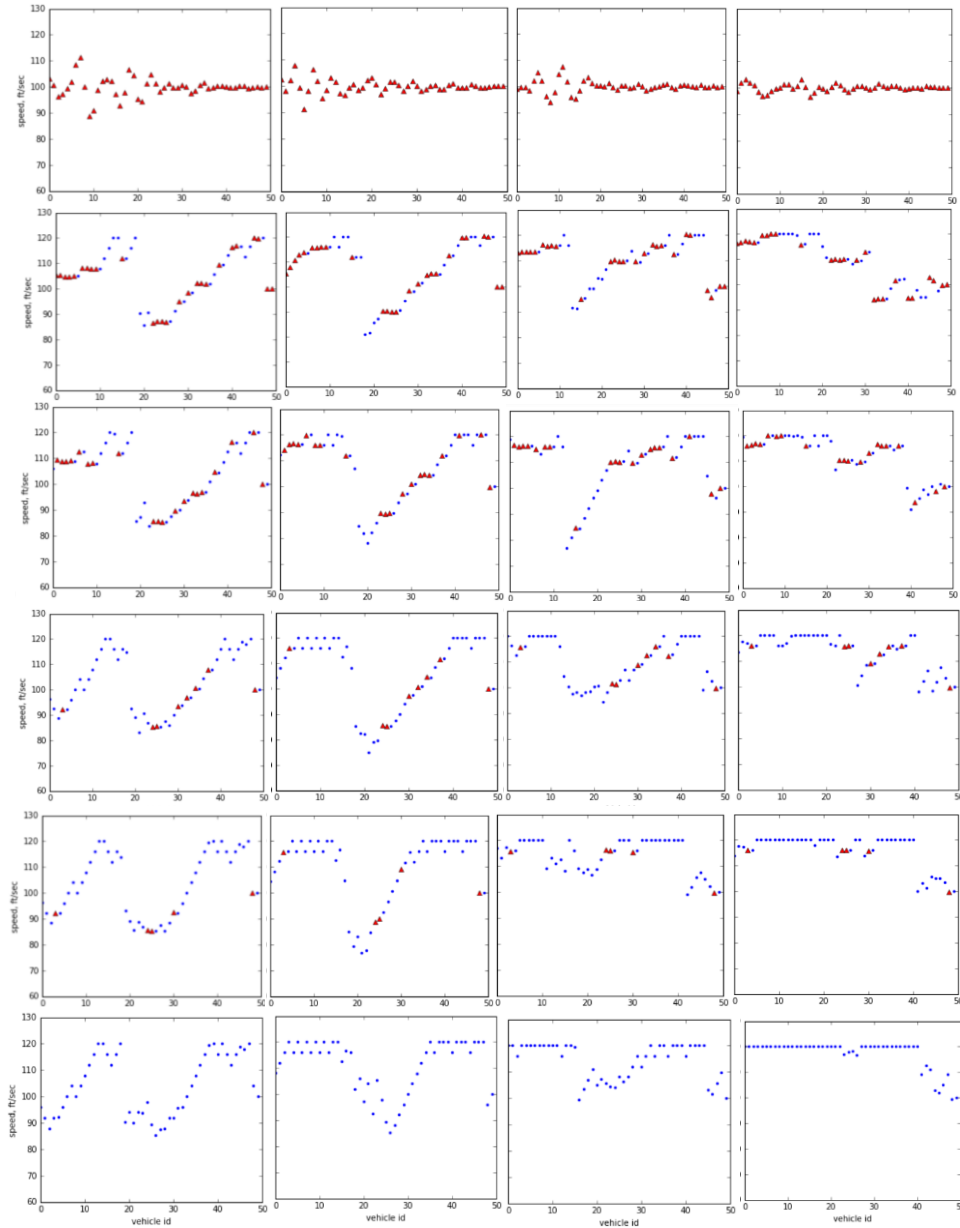
autonomous vehicles and human drivers. The proposed ES-based method can also be studied on the relationship with more recently emerging methods in cognitive and decision science. Modeling decision and behaviors of planners, traffic operators, policymakers, and other entities that heavily influence the network performance but rarely considered in conventional transportation system models might be also of interest.

With the improving understanding of human cognition and computational advancement in big data processing, the shift from the behavior-based methods to cognition-based methods has become promising. This shift may become increasingly useful when considering a society full of emerging technologies, new business models, and seemingly nuanced adjustment of choice context and regulatory strategies. Cognition-based models provide a promising modeling and analytical option.

# APPENDICES

## Appendix A

As a complement of Figure 1.9, the graphs below show traffic condition at different time step given different MP. Y-axis: 0, 0.3, 0.5, 0.7, 0.9, 1.0 MP Rate (red triangles are human drivers; blue dots are ACV). X-axis (left to right):  $t = 30s, 60s, 120s, 300s$





## Appendix B

This appendix proves that that when there perceive  $m$  states of nature in a decision, one can always set up the same  $n$  ( $n \leq m$ ) events in each prospect under consideration.

Proof. Since the states of nature can be represented in a set  $\{s_1, s_2, \dots, s_m\}$ . Let prospect  $x$  contain  $l$  events and prospect  $y$  contain  $k$  events. The  $l$  events in  $x$  and  $k$  events in  $y$  can always be partitioned in a way that when a state occurs in the  $j_{th}$  event in  $x$ , it would also happen in the same event set if  $y$  was chosen, allowing empty event set.  $\square$

## Appendix C

This appendix verify and discusses that information entropy is the unique uncertainty measure that satisfies three basic conditions in real number realm. For strict proof, see Shannon (1949)

Proof. Suppose that there exists at least one uncertainty measure such that  $f: \delta \rightarrow \mathbb{R}^+$ , where  $\delta$  is  $J$ -dimension vector and each element  $\delta_j \in [0,1], j \in \mathbb{J}, Card(\mathbb{J}) = J$ .  $\mathbb{J}$  is the set for all the perceived possible states of the observable.  $\mathbf{P} = (p_1, p_2, \dots, p_j, \dots, p_J)$  are the probability set for each possible state  $j$ . When the probability of one state is unity,  $f$  equates 0 for certainty.  $f$  shall be always greater than 0 in any other cases. The following are three basic conditions for the uncertainty measure  $f$  to satisfy:

- (1)  $f(P)$  is a continuous function and  $p_j \in [0,1]$  and  $\sum_j p_j = 1, \forall j \in \mathbb{J}$
- (2)  $f(\frac{1}{J}, \frac{1}{J}, \dots, \frac{1}{J})$  is a monotonically increasing function of  $J$
- (3)  $f(P) = f(P_A) + \sum p_{A_m} f(P_{B|A_m})$ , in which,  $C$  is the original experiment with probability distribution  $P$ , and  $A$  and  $B$  are two sequential measurements of the experiment  $C$ .  $P_A = (p_{A_1}, p_{A_2}, \dots, p_{A_m}, \dots, p_{A_M})$  and  $P_{B|A_m} = (p_{B_1|A_m}, p_{B_2|A_m}, \dots, p_{B_n|A_m}, \dots, p_{B_N|A_m})$ , respectively.

The condition (3) means that the uncertainty of a system should be the same regardless of the sequence of measuring it. Since  $A$  and  $B$  are two sequential steps that form the same distribution of  $C$ , the tensor product of  $P_A$  and  $P_{B|A_m}$  (i.e.,  $P_A \otimes P_{B|A_m}$ ) have the same support of that of  $P_C$  with element of  $P_A \otimes P_{B|A}$  to be  $p_{A_m} \cdot p_{B_n|A_m}, m, n \in \mathbb{N}^+$ . One can always find  $m$  and  $n$  so that  $n$  is not a function of  $A_m$  (see Appendix I for proof). Let's write  $P$  as matrix format and let the element  $(m, n)$  to be  $p_{m,n} = p_{A_m B_n}$ , and the expected value for the uncertainty of  $B$  given  $A_m$  would be

$$f(P_{B|A_m}) = \sum_m p_{A_m} \cdot f(P_{B|A_m})$$

And therefore

$$f(P) = f(P_A) + \sum_m p_{A_m} \cdot f(P_{B|A_m})$$

Now consider a particular case where both events in  $A$  and events in  $B|A_m$  are equally likely (i.e.,  $p = p_j, \forall j \in J$  and  $p_{A,k} = p_A, \forall k$ ), so that  $P_A = (p_{A_1}, p_{A_2}, \dots, p_{A_m}, \dots, p_{A_M}) = (p_A, p_A, \dots, p_A)$ , and  $P_{B|A_m} = (p_{B_1|A_m}, p_{B_2|A_m}, \dots, p_{B_n|A_m}, \dots, p_{B_N|A_m}) = (p_{B|A_m}, p_{B|A_m}, \dots, p_{B|A_m})$ . This also means that  $P = (p_1, p_2, \dots, p_{MN}) = (p_{MN}, p_{MN}, \dots, p_{MN})$ . Without loss of generality, one can also write  $P_A$  and  $P_{B|A_m}$  as a tuple  $(\frac{1}{M}, \frac{1}{M}, \dots, \frac{1}{M})$  and a tuple  $(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N})$ , respectively. Plug the tuples into the formulation for  $f(P)$ , we have

$$f(P) = f(P_A) + M \cdot \frac{1}{M} \cdot f(P_{B|A_m}) = f(P_A) + f(P_{B|A_m})$$

$$\text{where } P = \left( \underbrace{\frac{1}{MN}, \frac{1}{MN}, \dots, \frac{1}{MN}}_{MN} \right), P_A = \left( \underbrace{\frac{1}{M}, \frac{1}{M}, \dots, \frac{1}{M}}_M \right), P_{B|A_m} = \left( \underbrace{\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}}_N \right).$$

Let  $g(x) = f\left(\underbrace{\frac{1}{x}, \frac{1}{x}, \dots, \frac{1}{x}}_x\right)$ . Since the equation above shows that  $g(MN) = g(M) + g(N)$ , and

Aczél and Daróczy (1977) proves that, when  $g(\cdot)$  is continuous and  $x$  is a positive real number, the only solution for  $g(x)$  is  $c \cdot \log_b(x)$  for some constant  $c$ . Therefore, we have

$$g(MN) = f\left(\underbrace{\frac{1}{MN}, \frac{1}{MN}, \dots, \frac{1}{MN}}_{MN}\right) = c \cdot \log_b MN$$

where  $b > 0$  and  $c$  is an arbitrary real constant. One can verify by plugging it back to the formulation in condition (3) so that

$$c \cdot \log_b MN = c \cdot \log_b M + c \cdot \log_b N = c \cdot f\left(\underbrace{\frac{1}{M}, \frac{1}{M}, \dots, \frac{1}{M}}_M\right) + c \cdot f\left(\underbrace{\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}}_N\right)$$

Since the function form  $c \cdot \log(x)$  is unique in satisfying a special case of the condition (3), this is the only candidate that has the potential to satisfy condition (3). Generalizing  $\log_b MN$  to  $\log_b Q$  by not assuming  $P_A$  is uniformly distributed ( $P_{B|A_m}$  keeps being uniformly distributed with  $n$  events).  $Q = \sum_m N_m$ , where  $N_m$  is the number of equally likely event given the event  $A_m$ . we have

$$f(P_A) = c \cdot \log_b Q - \sum_m p_{A_m} \cdot c \cdot \log_b N_m = c \cdot \sum_m p_{A_m} \cdot \log_b \frac{Q}{N_m}$$

Since  $\frac{N_m}{Q} = p_{A_m}$ ,

$$f(P_A) = c \cdot \sum_m p_{A_m} \cdot \log_b \frac{1}{p_{A_m}}$$

The scaling factor can be removed without loss of generality to have

$$f(P) = \sum_w p_w \cdot \log_b \frac{1}{p_w} = - \sum_w p_w \cdot \log_b p_w, \forall w \in \{1, 2, \dots, mn\}$$

Condition (1) and (2) are also easily tested since the formulation is a basic combination of elementary functions. The first derivative of  $f(P)$  is,

$$\frac{\partial f(P)}{\partial p_w} = \log_b \frac{1}{p_w} + \frac{p_w^2}{\ln b} > 0, p_w \in [0, 1]$$

Therefore  $f(P)$  is not only continuous but also strictly monotonically increasing. Any other uncertainty measure with probabilities being in real number space is bound to produce paradoxes.

## REFERENCES

- Abdel-Aty, M. A., Kitamura, R., & Jovanis, P. P. (1997). Using stated preference data for studying the effect of advanced traffic information on drivers' route choice. *Transportation Research Part C: Emerging Technologies*, 5(1), 39-50.
- Aczél, J., & Daróczy, Z. (1975). On measures of information and their characterizations. *New York*.
- Allais, M. (1953a). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 503-546.
- Arentze, T., & Timmermans, H. (2008). Social networks, social interactions, and activity-travel behavior: a framework for microsimulation. *Environment and Planning B: Planning and Design*, 35(6), 1012-1027.
- Asano, M., Basieva, I., Khrennikov, A., Ohya, M., & Tanaka, Y. (2012). Quantum-like generalization of the Bayesian updating scheme for objective and subjective mental uncertainties. *Journal of Mathematical Psychology*, 56(3), 166-175.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological review*, 64(6p1), 359.
- Auld, J., Hope, M., Ley, H., Sokolov, V., Xu, B., & Zhang, K. (2016). POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies*, 64, 101-116.
- Baker, R. G. (1999). On the quantum mechanics of optic flow and its application to driving in uncertain environments<sup>1</sup>. *Transportation research part F: traffic psychology and behaviour*, 2(1), 27-53.
- Balmer, M., Rieser, M., Meister, K., Charypar, D., Lefebvre, N., & Nagel, K. (2009). MATSim-T: Architecture and simulation times. In *Multi-agent systems for traffic and transportation engineering* (pp. 57-78). IGI Global.
- Becker, J. L., & Sarin, R. K. (1987). Lottery dependent utility. *Management Science*, 33(11), 1367-1382.
- Ben-Akiva, M. E. (2010). Smart-future urban mobility. *JOURNEYS*, 30.
- Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H., & Mishalani, R. (1998, February). DynaMIT: a simulation-based system for traffic prediction. In *DACCORS short term forecasting workshop, The Netherlands*.
- Ben-Elia, E., Di Pace, R., Bifulco, G. N., & Shiftan, Y. (2013). The impact of travel information's accuracy on route-choice. *Transportation Research Part C: Emerging Technologies*, 26, 146-159.
- Ben-Elia, E., Erev, I., & Shiftan, Y. (2008). The combined effect of information and experience on drivers' route-choice behavior. *Transportation*, 35(2), 165-177.
- Bhat, C., Guo, J., Srinivasan, S., & Sivakumar, A. (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record: Journal of the Transportation Research Board*, (1894), 57-66.
- Bliemer, M. C., & Bovy, P. H. (2003). Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem. *Transportation Research Part B: Methodological*, 37(6), 501-519.
- Blumenschein, K., Blomquist, G. C., Johannesson, M., Horn, N., & Freeman, P. (2008). Eliciting willingness to pay without bias: evidence from a field experiment. *The Economic Journal*, 118(525), 114-137.

- Bogers, E., Viti, F., & Hoogendoorn, S. (2005). Joint modeling of advanced travel information service, habit, and learning impacts on route choice by laboratory simulator experiments. *Transportation Research Record: Journal of the Transportation Research Board*, (1926), 189-197.
- Boyles, S. D., Kockelman, K. M., & Waller, S. T. (2010). Congestion pricing under operational, supply-side uncertainty. *Transportation Research Part C: Emerging Technologies*, 18(4), 519-535.
- Brandstätter, E., Kühberger, A., & Schneider, F. (2002). A cognitive-emotional account of the shape of the probability weighting function. *Journal of Behavioral Decision Making*, 15(2), 79-100.
- Brownstone, D., & Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39(4), 279-293.
- Brownstone, D., & Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39(4), 279-293.
- Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.
- Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). Empirical comparison of Markov and quantum models of decision making. *Journal of Mathematical Psychology*, 53(5), 423-433.
- Camerer, C. F. (1989). An experimental test of several generalized utility theories. *Journal of Risk and uncertainty*, 2(1), 61-104.
- Camerer, C. F., & Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of risk and uncertainty*, 8(2), 167-196.
- Carey, M. (1992). Nonconvexity of the dynamic traffic assignment problem. *Transportation Research Part B: Methodological*, 26(2), 127-133.
- Chen, A., Lee, D. H., & Jayakrishnan, R. (2002). Computational study of state-of-the-art path-based traffic assignment algorithms. *Mathematics and computers in simulation*, 59(6), 509-518.
- Chen, B. Y., Lam, W. H., Sumalee, A., & Shao, H. (2011). An efficient solution algorithm for solving multi-class reliability-based traffic assignment problem. *Mathematical and Computer Modelling*, 54(5), 1428-1439.
- Chen, X., & Zhan, F. B. (2008). Agent-based modelling and simulation of urban evacuation: relative effectiveness of simultaneous and staged evacuation strategies. *Journal of the Operational Research Society*, 59(1), 25-33.
- Chong, L., Abbas, M., & Medina, A. (2011). Simulation of driver behavior with agent-based back-propagation neural network. *Transportation Research Record: Journal of the Transportation Research Board*, (2249), 44-51.
- Chorus, C. G., Arentze, T. A., & Timmermans, H. J. (2009). Traveler compliance with advice: a Bayesian utilitarian perspective. *Transportation Research Part E: Logistics and Transportation Review*, 45(3), 486-500.
- Chorus, C. G., Molin, E. J., Van Wee, B., Arentze, T. A., & Timmermans, H. J. (2006). Responses to transit information among car-drivers: regret-based models and simulations. *Transportation Planning and Technology*, 29(4), 249-271.

- Chorus, C. G., Walker, J. L., & Ben-Akiva, M. (2013). A joint model of travel information acquisition and response to received messages. *Transportation Research Part C: Emerging Technologies*, 26, 61-77.
- Chow, J., Lee, G., & Yang, I. (2010). Genetic algorithm to estimate cumulative prospect theory parameters for selection of high-occupancy-vehicle lane. *Transportation Research Record: Journal of the Transportation Research Board*, (2157), 71-77.
- Clawson, J. G. (1991). *Why People Behave the Way they Do*. Darden Case No. UVA-0B-0183
- Climaco, J. C. N., & Martins, E. Q. V. (1982). A bicriterion shortest path algorithm. *European Journal of Operational Research*, 11(4), 399-404.
- Daniels, R. L., & Keller, L. R. (1990). An experimental evaluation of the descriptive validity of lottery-dependent utility theory. *Journal of Risk and uncertainty*, 3(2), 115-134.
- Davidsson, P., Henesey, L., Ramstedt, L., Törnquist, J., & Wernstedt, F. (2005). An analysis of agent-based approaches to transport logistics. *Transportation Research part C: emerging technologies*, 13(4), 255-271.
- De Bruijn, H., & Leijten, M. (2007). Megaprojects and contested information. *Transportation Planning and Technology*, 30(1), 49-69.
- de Palma, A., & Picard, N. (2005). Route choice decision under travel time uncertainty. *Transportation Research Part A: Policy and Practice*, 39(4), 295-324.
- de Palma, A., Ben-Akiva, M., Brownstone, D., Holt, C., Magnac, T., McFadden, D., ... & Walker, J. (2008). Risk, uncertainty and discrete choice models. *Marketing Letters*, 19(3-4), 269-285.
- Dervisoglu, G., Gomes, G., Kwon, J., Horowitz, R., & Varaiya, P. (2009, January). Automatic calibration of the fundamental diagram and empirical observations on capacity. In *Transportation Research Board 88th Annual Meeting* (Vol. 15).
- Dogterom, N., Ettema, D., & Dijst, M. (2018). Behavioural effects of a tradable driving credit scheme: Results of an online stated adaptation experiment in the Netherlands. *Transportation Research Part A: Policy and Practice*, 107, 52-64.
- Dolnicar, S. (2013). Asking good survey questions. *Journal of Travel Research*, 52(5), 551-574.
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, 643-669.
- Evans, J. S. B., & Frankish, K. E. (2009). *In two minds: Dual processes and beyond*. Oxford University Press.
- Fagnant, D. J., & Kockelman, K. M. (2014). The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, 40, 1-13.
- Fischhoff, B., Watson, S. R., & Hope, C. (1984). Defining risk. *Policy sciences*, 17(2), 123-139.
- Fujii, S., & Gärling, T. (2003). Application of attitude theory for improved predictive accuracy of stated preference methods in travel demand analysis. *Transportation Research Part A: Policy and Practice*, 37(4), 389-402.
- Gao, J., Zhang, F., Sun, L., & Li, H. (2016). A study of traveler behavior under traffic information-provided conditions in the Beijing area. *Transportation Planning and Technology*, 39(8), 768-778.

- Gao, S., Frejinger, E., & Ben-Akiva, M. (2011). Cognitive cost in route choice with real-time information: an exploratory analysis. *Transportation Research Part A: Policy and Practice*, 45(9), 916-926.
- Gao, S., Frejinger, E., & Ben-Akiva, M. (2011). Cognitive cost in route choice with real-time information: an exploratory analysis. *Transportation Research Part A: Policy and Practice*, 45(9), 916-926.
- Guo, X., & Liu, H. X. (2011). Bounded rationality and irreversible network change. *Transportation Research Part B: Methodological*, 45(10), 1606-1618.
- Hamdar, S., Treiber, M., Mahmassani, H., & Kesting, A. (2008). Modeling driver behavior as sequential risk-taking task. *Transportation Research Record: Journal of the Transportation Research Board*, (2088), 208-217.
- Hameroff, S., & Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and computers in simulation*, 40(3-4), 453-480.
- Handa, J. (1977). Risk, probabilities, and a new theory of cardinal utility. *Journal of Political Economy*, 85(1), 97-122.
- Hasan, S., Ukkusuri, S., Gladwin, H., & Murray-Tuite, P. (2010). Behavioral model to understand household-level hurricane evacuation decision making. *Journal of Transportation Engineering*, 137(5), 341-348.
- Haven, E., & Khrennikov, A. (2013). *Quantum social science*. Cambridge University Press.
- Helbing, D. (2012). Agent-based modeling. In *Social self-organization* (pp. 25-70). Springer Berlin
- Hertz, J. A. (2018). *Introduction to the theory of neural computation*. CRC Press.
- Huang, H. J., & Li, Z. C. (2007). A multiclass, multicriteria logit-based traffic equilibrium assignment model under ATIS. *European Journal of Operational Research*, 176(3), 1464-1477.
- Janis, I. L. (1993). Decision making under stress. *Handbook of stress: theoretical and clinical aspects*, 56-74.
- Jayakrishnan, R., Mahmassani, H. S., & Hu, T. Y. (1994). An evaluation tool for advanced traffic information and management systems in urban networks. *Transportation Research Part C: Emerging Technologies*, 2(3), 129-147.
- Jayakrishnan, R., Tsai, W. K., & Chen, A. (1995). A dynamic traffic assignment model with traffic-flow relationships. *Transportation Research Part C: Emerging Technologies*, 3(1), 51-72.
- Jayakrishnan, R., Tsai, W. T., Prashker, J. N., & Rajadhyaksha, S. (1994). A faster path-based algorithm for traffic assignment. *University of California Transportation Center*.
- Jha, M., Madanat, S., & Peeta, S. (1998). Perception updating and day-to-day travel choice dynamics in traffic networks with information provision. *Transportation Research Part C: Emerging Technologies*, 6(3), 189-212.
- Jiang, L., Mahmassani, H., & Zhang, K. (2011). Congestion pricing, heterogeneous users, and travel time reliability: Multicriterion dynamic user equilibrium model and efficient implementation for large-scale networks. *Transportation Research Record: Journal of the Transportation Research Board*, (2254), 58-67.
- Jin, W. L. (2010). A kinematic wave theory of lane-changing traffic flow. *Transportation research part B: methodological*, 44(8), 1001-1021.)

- Johnston, R. A. (2004). CHAPTER 5 The Urban Transportation Planning Process. *The geography of urban transportation*, 115.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.
- Kaihara, T. (2003). Multi-agent based supply chain modeling with dynamic environment. *International Journal of Production Economics*, 85(2), 263-269.
- Kopitch, L., & Saphores, J. D. M. (2011). *Assessing Effectiveness of Changeable Message Signs on Secondary Crashes* (No. 11-4270).
- LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and molecular neurobiology*, 23(4-5), 727-738.
- Levin, M. W., & Boyles, S. D. (2016). A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transportation Research Part C: Emerging Technologies*, 62, 103-116.
- Levinson, D. (2003). The value of advanced traveler information systems for route choice. *Transportation Research Part C: Emerging Technologies*, 11(1), 75-87.
- Liu, H. X., Recker, W., & Chen, A. (2004). Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A: Policy and Practice*, 38(6), 435-453.
- Liu, H. X., Recker, W., & Chen, A. (2004). Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A: Policy and Practice*, 38(6), 435-453.
- Lo, H. K., Luo, X. W., & Siu, B. W. (2006). Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. *Transportation Research Part B: Methodological*, 40(9), 792-806.
- Lu, L., Yun, T., Li, L., Su, Y., & Yao, D. (2010). A comparison of phase transitions produced by PARAMICS, TransModeler, and Vissim. *IEEE Intelligent Transportation Systems Magazine*, 2(3), 19-24.
- Lu, C. C., Mahmassani, H. S., & Zhou, X. (2009). Equivalent gap function-based reformulation and solution algorithm for the dynamic user equilibrium problem. *Transportation Research Part B: Methodological*, 43(3), 345-364.
- Macal, C. M., & North, M. J. (2005, December). Tutorial on agent-based modeling and simulation. In *Proceedings of the 37th conference on Winter simulation* (pp. 2-15). Winter Simulation Conference.
- Martins, E. Q. V. (1984). On a multicriteria shortest path problem. *European Journal of Operational Research*, 16(2), 236-245.
- McNeil, B. J., Pauker, S. G., Sox Jr, H. C., & Tversky, A. (1982). 230 On the Elicitation of Preferences for Alternative Therapies. *Preference, Belief, and Similarity*, 583.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2), 81.



- Mitsakis, E., Grau, J. M. S., Stamos, I., & Aifadopoulou, G. (2015). An integrated framework for embedding large-scale dynamic traffic assignment models in advanced traveler information systems. *Transportation Planning and Technology*, 38(8), 866-877.
- Mote, J., Murthy, I., & Olson, D. L. (1991). A parametric approach to solving bicriterion shortest path problems. *European Journal of Operational Research*, 53(1), 81-92.
- Müller, K., & Axhausen, K. W. (2010). *Population synthesis for microsimulation: State of the art*. ETH Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau (IVT).
- Myers, D. G. (2004). *Exploring psychology*. Macmillan.
- Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics*, 18(6), 275-285.
- Nagel, K., Beckman, R. J., & Barrett, C. L. (1999, September). TRANSIMS for urban planning. In *6th International Conference on Computers in Urban Planning and Urban Management, Venice, Italy*.
- Neumann, L. J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Newell, G. F. (1961). Nonlinear effects in the dynamics of car following. *Operations research*, 9(2), 209-229.
- Noland, R. B., & Small, K. A. (1995). *Travel-time uncertainty, departure time choice, and the cost of the morning commute*. Institute of Transportation Studies, University of California, Irvine.
- Norwich, K. H. (1993). *Information, sensation, and perception* (p. 247). San Diego: Academic Press.
- Peeta, S., & Ziliaskopoulos, A. K. (2001). Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1(3-4), 233-265.
- Perez, P., Banos, A., & Pettit, C. (2016, May). Agent-Based Modelling for Urban Planning Current Limitations and Future Trends. In *International Workshop on Agent Based Modelling of Urban Systems* (pp. 60-69). Springer, Cham.
- Poirier, D. J. (1995). *Intermediate statistics and econometrics: a comparative approach*. MIT Press.
- preferences. *Econometrica: Journal of the Econometric Society*, 1221-1239...
- Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4), 323-343.
- Railsback, S. F., & Grimm, V. (2011). *Agent-based and individual-based modeling: a practical introduction*. Princeton university press.
- Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9), 609-623.
- Ramos, G. M., Daamen, W., & Hoogendoorn, S. (2014). A state-of-the-art review: developments in utility theory, prospect theory and regret theory to investigate travellers' behaviour in situations involving travel time uncertainty. *Transport Reviews*, 34(1), 46-67.
- Razo, M., & Gao, S. (2013). A rank-dependent expected utility model for strategic route choice with stated preference data. *Transportation Research Part C: Emerging Technologies*, 27, 117-130.

- Salvini, P., & Miller, E. J. (2005). ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, 5(2), 217-234.
- Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), 362-380.
- Scholten, M., & Read, D. (2014). Prospect theory and the “forgotten” fourfold pattern of risk preferences. *Journal of Risk and Uncertainty*, 48(1), 67-83.
- Shannon, C. E. (1949). A mathematical theory of communication. *Bell System Technical Journal*, 30:47-51
- Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of personality and Social Psychology*, 39(2), 211.
- Skriver, A. J., & Andersen, K. A. (2000). A label correcting approach for solving bicriterion shortest-path problems. *Computers & Operations Research*, 27(6), 507-524.
- Small, K. A., Winston, C., & Yan, J. (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4), 1367-1382.
- Stopher, P. R., Wilmot, C. G., Stecher, C., & Alsnih, R. (2006). Household travel surveys: proposed standards and guidelines. In *Travel survey methods: Quality and future directions* (pp. 19-74). Emerald Group Publishing Limited.
- Stopher, P., FitzGerald, C., & Xu, M. (2007). Assessing the accuracy of the Sydney Household Travel with GPS. *Transportation*, 34(6), 723-741.
- Sun, D. J., & Kondyli, A. (2010). Modeling Vehicle Interactions during Lane-Changing Behavior on Arterial Streets. *Computer-Aided Civil and Infrastructure Engineering*, 25(8), 557-571.
- Sun, Q., & Wu, S. (2014). A Configurable Agent-Based Crowd Model with Generic Behavior Effect Representation Mechanism. *Computer-Aided Civil and Infrastructure Engineering*, 29(7), 531-545.
- Susskind, L., & Friedman, A. (2014). *Quantum mechanics: the theoretical minimum*. Basic Books (AZ).
- Tan, Z., Yang, H., & Guo, R. (2014). Pareto efficiency of reliability-based traffic equilibria and risk-taking behavior of travelers. *Transportation research part B: methodological*, 66, 16-31.
- Townsend, J. T., Silva, K. M., Spencer-Smith, J., & Wenger, M. J. (2000). Exploring the relations between categorization and decision making with regard to realistic face stimuli. *Pragmatics & Cognition*, 8(1), 83-105.
- Tumer, K., & Agogino, A. (2007, May). Distributed agent-based air traffic flow management. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems* (p. 255). ACM.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, 79(4), 281.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157), 1124-1131.
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of business*, S251-S278.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297-323.

- United States Department of Transportation, 2016, *Connected Vehicles and Cybersecurity Fact Sheet*.  
[http://www.its.dot.gov/factsheets/pdf/cybersecurity\\_factsheet.pdf](http://www.its.dot.gov/factsheets/pdf/cybersecurity_factsheet.pdf)
- Van Arem, B., Van Driel, C. J., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, 7(4), 429-436.
- Von Neumann, J., & Morgenstern, O. (1945). Theory of games and economic behavior. *Bull. Amer. Math. Soc*, 51(7), 498-504.
- Waddell, P., Ševčíková, H., Socha, D., Miller, E., & Nagel, K. (2005). Opus: An open platform for urban simulation. In *Computers in urban planning and urban management conference, London*.
- Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge university press.
- Washington, S. P., Karlaftis, M. G., & Mannering, F. (2010). Random-Parameter Models (Chapter 16). *Statistical and econometric methods for transportation data analysis*. CRC press.
- Wolf, J., Bricka, S., Ashby, T., & Gorugantua, C. (2004). Advances in the application of GPS to household travel surveys. In *National Household Travel Survey Conference, Washington DC*.
- Xiong, C., Chen, X., He, X., Lin, X., & Zhang, L. (2015). Agent-based en-route diversion: Dynamic behavioral responses and network performance represented by Macroscopic Fundamental Diagrams. *Transportation Research Part C: Emerging Technologies*.
- Xu, H., Lou, Y., Yin, Y., & Zhou, J. (2011). A prospect-based user equilibrium model with endogenous reference points and its application in congestion pricing. *Transportation Research Part B: Methodological*, 45(2).
- Xuan, Y. E., & Kanafani, A. (2014). Evaluation of the effectiveness of accident information on freeway changeable message signs: A comparison of empirical methodologies. *Transportation research part C: emerging technologies*, 48, 158-171.
- Yang, H., & Huang, H. J. (2004). The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. *Transportation Research Part B: Methodological*, 38(1), 1-15.
- Yang, I., & Jayakrishnan, R. (2012). Gradient projection method for simulation-based dynamic traffic assignment. *Transportation Research Record: Journal of the Transportation Research Board*, (2284), 70-80.
- Zhang, K., Mahmassani, H., Lu, C. C.. (2013) Dynamic pricing, heterogeneous users and perception error: Probit-based bi-criterion dynamic stochastic user equilibrium assignment. *Transportation Research Part C: Emerging Technologies* 27, 189-204.  
 Online publication date: 1-Feb-2013.
- Zheng, H., Son, Y. J., Chiu, Y. C., Head, L., Feng, Y., Xi, H., ... & Hickman, M. (2013). *A Primer for Agent-Based Simulation and Modeling in Transportation Applications* (No. FHWA-HRT-13-054).