

**MODELING ROUTE CHOICE BEHAVIOR
IN THE PRESENCE OF INFORMATION
USING CONCEPTS FROM FUZZY SET THEORY
AND APPROXIMATE REASONING**

by

TSIPPY LOTAN

B.Sc. Computer Sciences and Statistics
The Hebrew University of Jerusalem, 1982

M.A. Statistics and Operations Research
The Hebrew University of Jerusalem, 1984

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment for the Requirements for the Degree of

Doctor of Science in Transportation Systems

at the

Massachusetts Institute of Technology
August, 1992

© Massachusetts Institute of Technology, 1992.

Signature of Author _____
Department of Civil and Environmental Engineering
August 14, 1992

Certified by _____
Harris N. Koutsopoulos
Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by _____
Eduardo Kausel
Chairman, Departmental Committee on Graduate Studies
Department of Civil and Environmental Engineering

ARCHIVES

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

SEP 28 1992

LIBRARIES

MODELING ROUTE CHOICE BEHAVIOR
IN THE PRESENCE OF INFORMATION
USING CONCEPTS FROM FUZZY SET THEORY
AND APPROXIMATE REASONING

by

Tsippy Lotan

Submitted to the Department of Civil and Environmental Engineering
on August 14, 1992 in partial fulfillment of the
requirements for the Degree of Doctor of Science
in Transportation Systems

ABSTRACT

The route choice problem in general, and in the presence of information in particular, is not a very well understood problem in the transportation literature. Most route choice models are based on the random-utility discrete-choice paradigm and assume that users choose the alternative with maximum utility value among the different alternatives. Recently, with the increased interest in the area of Intelligent Vehicle/Highway Systems (IVHS) and specifically Advanced Traveller Information Systems (ATIS), limitations of existing models have become very apparent. ATIS provide drivers with real-time traffic information in order to facilitate better route choice decisions. Under these conditions drivers are required to incorporate verbal, visual and prescriptive information into their own perceptions and consequently make route choices.

In this thesis we propose three hypotheses regarding drivers behavior in the presence of information. The *simultaneous hypothesis* which assumes that drivers incorporate all the factors that affect their decision simultaneously (including traffic information). The *two-stage hypothesis* according to which the drivers first update their perceptions based on the new information that they have acquired or experienced, and subsequently they make route choice decisions based on their updated perceptions. Finally, the *default hypothesis* which assumes that there is an underlying behavioral pattern for each driver which serves as default behavior if no unusual or unexpected conditions occur. Information on unexpected conditions triggers reevaluation of the default behavior.

The two main elements required in order to implement the above hypotheses are models of drivers perceptions, and models of the decision process itself. We present a framework for modeling both perceptions and the route choice process in the presence of information using concepts from fuzzy sets theory, approximate reasoning, and fuzzy control.

Linguistic variables and possibility distributions are used to model perceptions of

both network attributes and the provided information. For modeling the decision process itself two main approaches, which use as inputs fuzzy perceptions, are suggested.

The first approach is based on the classical discrete choice paradigm appropriately extended to incorporate fuzzy attributes in the systematic utility component. The maximum likelihood function of the fuzzy utility model is formulated and used for the calibration of its parameters.

The second approach is based on concepts from approximate reasoning and fuzzy control. Linguistic rules of the form: "IF ... THEN ..." are used. The rules describe attitudes towards taking a specific route given (possibly vague) perceptions on network attributes. They are used as anchoring schemes for decisions, while the adjustments of the rules to changing conditions is done by an approximate reasoning mechanism. The use of the fuzzy, approximate reasoning methodology, facilitates a flexible rule interpretation by automatically deriving rules that are close to the original rules. Hence the suggested approach requires existence of a relatively small rule base (compared to other rule-based systems). All the adjusted rules are then applied simultaneously (each with the appropriate degree) and the process results in the final attractiveness of each alternative. From a behavioral point of view this approach may provide a more natural framework for modeling route choice decisions (especially when they are made under time pressure). The underlying assumption is that drivers have limited information processing capabilities and therefore use simple rules to make decisions. A rigorous approach, based on a mathematical programming formulation of the problem, has been developed for estimation of the parameters of the model.

We conclude by presenting results from a case study. The data for the case study was collected using a driver simulator that we have designed and implemented. Ten subjects participated in the experiment, and each performed 20 trips under various traffic conditions (congestion levels, incidents, etc). The data collected includes prior perceptions (based on interviews), observed traffic conditions while driving, the available pre-trip and en-route information, and the resulting choices made. A rule matrix was estimated both for each subject individually, and for the entire population. Preliminary results support our hypothesis that route choice behavior in the presence of information can be modeled by a small set of intuitive and reasonable rules. The results also compare favorably to results obtained by more traditional approaches.

Thesis Supervisor: Dr. Haris N. Koutsopoulos
Title: Assistant Professor of Civil and Environmental Engineering

ACKNOWLEDGEMENTS

To begin with, I would like to express my deep gratitude to my thesis supervisor, Haris Koutsopoulos, for his encouragement and support throughout the development of this dissertation. Haris has taught me many things: how to focus my thinking, to distinguish between the things that really matter and those that don't, and that I cannot solve all of the world's problems in one thesis. For all this, and for all the help and support he gave me beyond my thesis throughout my four years at MIT, I am truly grateful.

Moshe Ben Akiva's continuously constructive and useful criticism helped keep me "down to earth" in my work, and he was always there when I needed him. His support, confidence and advice were invaluable in every step of the way. The discussions I had with Nigel Wilson were always very interesting and helpful. His openness and breadth of knowledge helped me view my research from new and different angles, and provided original and important insights. His help in finalizing this thesis was very instrumental. Haruhiko Asada, my only source of "fuzziness" at MIT, displayed an unfailing interest in my work and provided many useful comments.

Lotfi Zadeh, the father of fuzzy set theory, was a real and admirable source of inspiration.

The financial support of the University Transportation Center program, the MIT Center for Transportation Studies, and the UPS Foundation, is deeply appreciated.

Outside my committee, many individuals have made indispensable contributions to this thesis, and I am grateful to all. In particular, I would like to mention Dinesh Gopinath, Andras Farkas, Frannie Humplick, Samer Madanat, Rabi Mishalani, Yoram and Alona Shiftan, and Rina Shneur. Thanks to all of you I had a marvelous time at MIT. Special thanks are due to Qi Yang, for his belief in the simulator and the endless time he devoted to it.

By her continuous love and dedication to our family, Yvonne Staples helped me maintain my peace of mind. Her hard work, kindness and devotion are greatly appreciated.

My parents, Sarah and Joseph Weiss have given me endless love, support and confidence not only throughout my graduate work, but during my entire lifetime. For their unequivocal love, for accepting the long separation, for being there every day, I will always be grateful.

My final thanks are for Jaron, Mor and Shannie. No one can imagine a better family. Shannie: thanks for taking the generals with me and helping to keep things in proportion. Mor: I will forever cherish your curiosity, understanding, love and cheerfulness. And Jaron: thank you above all for being you, and for all the other things. This thesis is dedicated to you.

TABLE OF CONTENTS	page
CHAPTER 1: INTRODUCTION	10
1.1 Motivation	10
1.2 Advanced Traveler Information Systems	11
1.3 Literature Review	12
1.4 Thesis Objectives	14
1.5 Thesis Outline	15
 CHAPTER 2: ROUTE CHOICE BEHAVIOR IN THE PRESENCE OF	
INFORMATION	16
2.1 Hypotheses on Models for Route Choice Behavior in the Presence of	
Information	18
2.1.1 The Simultaneous Model	19
2.1.2 The Two-Stage Model	19
2.1.3 The Default Model	19
2.2 Discussion of Models	21
 CHAPTER 3: MODELING PERCEPTIONS AND INFORMATION	24
3.1 Fuzzy Set Theory	24
3.2 Modeling Driver Perceptions	28
3.2.1 Travel Time Perceptions	28
3.2.2 Empirical Evidence	32
3.3 Modeling Information	36
3.4 Information Integration and Update of Perceptions	38
3.4.1 Information Integration Theory	39
3.4.2 Default Reasoning	40
 CHAPTER 4: FUZZY ROUTE CHOICE MODELS	47
4.1 Shortest Paths in Networks with Fuzzy Attributes	47
4.1.1 Addition and Comparison of Fuzzy Numbers	47
4.1.2 A New Consistent Comparison Method	56
4.2 Fuzzy Utility Models	64
4.2.1 Fuzzy Linear Regression - Literature Review	64
4.2.2 Utility Models with Fuzzy Parameters and Crisp Inputs	69
4.2.3 Utility Models with Fuzzy Attributes and an Error Term ..	73
 CHAPTER 5: APPROXIMATE REASONING ROUTE CHOICE MODELS ..	79
5.1 The General Decision Framework	79
5.2 Rules	81
5.2.1 Rules Structure	81
5.2.2 Rules Complexity	84
5.2.3 Model-Specific Rules	87
5.3 Input Propagation through Rule Premises	88
5.4 Rule Execution ..	90
5.5 Rule Combination	92

5.6	Defuzzification	95
5.6.1	General	95
5.6.2	A Random Utility Model (RUM) Defuzzification Scheme	96
5.7	Implementation and Calibration	97
5.7.1	Generation of Membership Functions	97
5.7.2	Rule Generation	98
5.7.2.1	Integer Linear-Programming Based Rule Calibration	98
5.7.2.2	Heuristic Approach for Rule Calibration	101
CHAPTER 6: CASE STUDY		106
6.1	Driving Simulator	106
6.2	Data Collection	109
6.3	The Approximate Reasoning Model Implementation	112
6.4	Calibration Algorithm	118
6.5	Disaggregate Analysis	119
6.5.1	Decisions at the Origin	119
6.5.1.1	Calibration of the Rule Matrix	119
6.5.1.2	Comparison with Optimal Results	125
6.5.1.3	Comparison with Random Utility Model	126
6.5.1.4	Prediction	127
6.5.1.5	Sensitivity Analysis	128
6.5.2	Diversion Decisions	133
6.5.2.1	Calibration of Rule Matrix	136
6.5.2.2	Comparison with Optimal Results	137
6.5.2.3	Prediction	137
6.6	Aggregate Analysis	138
6.6.1	Decisions at the origin	138
6.6.1.1	Calibration of Rule Matrix	138
6.6.1.2	Comparison with Optimal Results	140
6.6.1.3	Comparison with Random Utility Model	142
6.6.2	Diversion Decisions	142
6.7	Discussion	143
CHAPTER 7: SUMMARY AND CONCLUDING REMARKS		145
7.1	Summary	145
7.2	Contributions	147
7.3	Directions for Future Research	148
APPENDIX A: FUZZY SET THEORY - DEFINITIONS AND NOTATIONS		152
APPENDIX B: APPROXIMATE REASONING AND FUZZY CONTROL ..		154
APPENDIX C: MIT COMMUTING SURVEY		159

APPENDIX D: PRELIMINARY SESSION FOR CASE STUDY 166

References 167

List of Figures

	Page	
Fig. 1	Framework for Cognitive Task Analysis (Rasmussen, 1986)	17
Fig. 2	Information Integration Scenarios	22
Fig. 3	Travel Time as Fuzzy Set	25
Fig. 4	TrFN for Modeling Travel Times	29
Fig. 5	Travel Time Perceptions	31
Fig. 6	Updated Perceptions, Poss(AIC)	42
Fig. 7	Updated Perceptions, $0 < \text{Poss}(AIC) < 1$	44
Fig. 8	Addition of Two Fuzzy Numbers	48
Fig. 9	Extended Maximum Operation	50
Fig. 10	Methods of Comparing Fuzzy Numbers	52
Fig. 11	Distances Between Two Sets	59
Fig. 12	Comparison of Two Fuzzy Numbers	61
Fig. 13	Degree of Fit between Y_i^* and Y_i	66
Fig. 14	Possible Intersections of Two TsFN's	71
Fig. 15	Intersections With the Fuzzy Set V	76
Fig. 16	The Decision Framework	80
Fig. 17	LHS and RHS Designs	83
Fig. 18	Overlap Between Perceptions and Design Sets	89
Fig. 19	Attractiveness of an Alternative	93
Fig. 20	Rule Calibration	103
Fig. 21	Driving Simulator	107
Fig. 22	The Newton Network	110
Fig. 23	Travel Time Categories	114
Fig. 24	Travel Time Categories - Version II	132
Fig. 25	Diversion Categories	134
Fig. 26	Rule Weights	141
Fig. 27	Fuzzy Control Scheme	155
Fig. 28	Implication Schemes	157

List of Tables

		Page
Table 1	Frequency of TrFN Shapes, General Population: K=1138	34
Table 2	Frequency of TrFN Shapes, Unfamiliar Population: K=66	34
Table 3	TrFN Distribution for Travel Times	35
Table 4	Ranges of Possible Travel Times	36
Table 5	Travel Time Perceptions by Origin (Cambridge and Lexington)	36
Table 6	Comparing Three Fuzzy Numbers - Chen 's Method	55
Table 7	General Sample Information (totals)	111
Table 8	Initial Rule Matrix for Node 1	117
Table 9	Detailed Initial and First-Level results	120
Table 10	Improved Monotone Fit (node 1)	122
Table 11	Improved monotone fit, Second level (node 1)	124
Table 12	Optimal Results	126
Table 13	Results of Random Utility Models	127
Table 14	Results of Default Model	129
Table 15	Non-Symmetric LHS design	130
Table 16	Fit for Different Perception Modeling	131
Table 17	Initial Rule Matrix for Node 2	135
Table 18	Diversion Decisions at node 2	136
Table 19	Optimal Results - Node 2	137
Table 20	Global Rule Matrix for node 1	139

CHAPTER 1: INTRODUCTION

1.1 Motivation

Provision of real time information on traffic conditions to drivers has become technically possible due to emergence of new technologies and recent advances in existing technologies. There has been a lot of enthusiasm and hope that in-vehicle navigation and route guidance systems (Advanced Traveler Information Systems, ATIS), will become an integral part of the solution to the traffic congestion problem.

Many theoretical studies are underway to address design issues of such systems and their effectiveness (for example Jayakrishnan and Mahmassani, 1991, Koutsopoulos and Xu, 1992, Kaysi, 1992). The effectiveness of such systems and their potential for reducing congestion, depends heavily on drivers' reaction to the additional information that might be available to them. However, most studies make major simplifying assumptions regarding drivers' behavior such as drivers have complete information, infinite information processing capabilities, and are able to make optimal decisions, or that a certain compliance rate with the information provided is achieved.

Modeling route choice decisions in the presence of information can be extremely difficult. The model has to consider all the usual factors that affect travel decisions such as: travel time, travel distance, type of road, travel speed, weather conditions, personal preferences and more. In addition, the model has to incorporate drivers' acquired information and its processing. Furthermore, decisions have to be made in real-time, in a limited time period, and often while en-route where the driver is primarily occupied with the driving task. Therefore, availability of pre-trip and en-route traffic information to drivers, adds a new dimension to route choice decisions. It adds the following complicating factors to the decision process:

- Information forces route choice decisions to become inherently dynamic, since traffic information can change continuously and might be unpredictable.
- Information not only affects the attractiveness of the alternatives in the choice set, but it may also change the choice set itself. Information can force drivers to consider new alternatives in their choice sets on one hand, and can make existing alternatives infeasible on the other hand.
- Route choice decisions in the presence of information are made under time pressure. Furthermore, drivers have limited information processing capabilities.
- Information has to be considered in light of existing behavior patterns (when no traffic information is available), and could have various effects; it can support the existing behavior patterns, it can change them, or its impact may be unclear.

- Information can come from various sources and have different degrees of reliability.
- There is a learning process as drivers gain more experience with existing information systems. This affects the way drivers acquire, interpret, and react to information.

1.2 Advanced Traveler Information Systems

Existing information systems vary from pre-trip information, through roadside displays and broadcasting systems, to electronic in-vehicle route guidance systems (OECD 1988, Boyce 1988, Koutsopoulos and Lotan 1990). Traffic information helps users not only to widen their choice set by exposing them to alternatives which they were not aware of before, but also to avoid bottlenecks due to incidents, construction etc. The overall objective of Advanced Traveler Information Systems is to utilize the available network capacity more efficiently, and to reduce the stress associated with driving by reducing the uncertainty related to driving conditions.

The most important parameters of the information that is being provided to users are:

- **Type** - what type of information is being conveyed (general network description, congestion levels, navigation information, prescriptive, descriptive etc).
- **Spatial distribution and frequency** - on what parts of the network is information available and how often is it updated.
- **Temporal Characteristics** - does the information pertain to average traffic conditions (historic), current conditions or future projected conditions.

Kaysi (1992) provides a comprehensive review of frameworks and models for provision of real-time information to drivers. He categorizes the various types of driver information systems into:

- Pre-trip route planning systems.
- Autonomous navigation aids.
- Broadcasting systems and roadside variable message signs.
- Automatic route guidance systems.

Some of the major demonstration projects currently underway are (see French, 1990, and Kaysi, 1992, for more details):

- ALI-SCOUT and LISB in Germany are route guidance systems that use infrared transmitters and receivers to transfer navigation information between roadside

beacons and on-board displays in equipped vehicles.

- **AUTOGUIDE** in London provides route recommendations based on real-time information on area-wide traffic conditions. The recommendations are intended to guide drivers over the shortest route to their destinations.
- **AMTICS** in Japan uses roadside beacons to transmit real-time traffic information from a traffic control center to an in-vehicle navigation system. It does not recommend any specific route, rather displays road maps, and navigation and parking information.
- **TRAVTEK** in Florida determines the best route for a given destination (selected by the driver), and uses graphic displays as well as audio to provide instructions, traffic incidents and traffic congestion locations.

It is common that information is transmitted to drivers verbally (such as information provided by broadcasting traffic reports on the radio or through voice synthesizers on board), or visually (such as roadside displays or in-vehicle equipment). Important safety issues are associated with these systems (particularly in-vehicle devices) because they could distract the driver's attention (Sheridan, 1991a). No clear standard for provision of information has emerged from the current demonstration projects.

1.3 Literature Review

There is extensive literature on the route choice problem, and recently Bovy and Stern (1990), provided a comprehensive literature review. Their general framework presents the route choice process as synthesizing, through a *black box*, two sources of inputs: drivers characteristics and transport network characteristics. Studies of route choice behavior can then be categorized according to the degree of detail at which they explain the interior of the black box.

The general framework for making route choice decisions includes the following stages: first the driver forms a choice set which is composed of alternatives that are considered feasible. Next the driver evaluates the alternatives in the choice set, and finally chooses the most attractive alternative. However, as Bovy and Stern (1990) indicate, none of these stages is straightforward. The choice set formation stage is based on a factor-importance hierarchy according to which the driver first sifts through a fairly sizable set of available alternatives, and eliminates all that are not sufficiently attractive. Only after the choice set has been reduced significantly, does the driver make an in-depth evaluation of the remaining alternatives. The evaluation is done on the basis of a composite utility function which reflects the relative evaluation of all aspects, resulting in a preference rank order. The driver then chooses among the ranked alternatives according to a certain decision rule (the alternative with the highest utility, or the routine alternative provided it is not significantly worse than others, etc.).

Bovy and Stern (1990) suggest that the content of the black box (which corresponds to the decision process itself) is a complicated system of filters through which information is selected and transformed. They signify two types of filters:

- Perception filter through which the individual's cognition of alternatives and attributes' values are processed.
- Evaluation filter through which the perceptions are transformed into a desirability scale.

There are many factors that are used for the evaluation of each alternative and affect route choice behavior. Bovy and Stern (1990), provide a comprehensive discussion of the main attributes that influence route choice behavior. They categorize those factors into 4 major groups:

- Personal characteristics of the driver.
- Route attributes which relate to road characteristics, traffic conditions, and environmental considerations.
- Trip characteristics (e.g. trip purpose, mode, etc.).
- Other circumstances such as weather conditions, time of day, accidents, and more.

They report that a common finding in many studies is that travel time is usually the most important independent choice factor, although other factors can not be disregarded. Travel time often serves as a compound measure of other attributes, and Horowitz (1978) claims that "value of time spent in travel is a surrogate measure of the comfort, convenience, and reliability of the travel experience".

Most route choice models make two major assumptions:

1. The choice set from which a single alternative is chosen, is given.
2. Drivers are well informed about attribute values, and are capable of making an optimal selection based on those values.

Most existing route choice models fall under the category of random utility theory (Ben-Akiva and Lerman, 1985). In a random utility model it is assumed that U_{in} , the utility of alternative i for individual n , is given by: $U_{in}=V_{in}+\epsilon_{in}$, where V_{in} is the systematic component corresponding to some functional form of the attributes that affect the utility, and ϵ_{in} is the random error corresponding to unobserved attributes and taste variations, measurement errors, imperfect information, and instrumental variables. It is assumed that each individual attempts to maximize his/her utility. Thus the probability

that individual n will choose alternative i is given by:

$$\text{Prob}_n(i) = \text{Prob}(V_{in} + \epsilon_{in} \geq V_{jn} + \epsilon_{jn}) \quad \text{for all } j \neq i$$

Depending on the assumptions made about the functional form of V , and the distribution of the disturbance ϵ , various discrete choice models may be formulated.

Another modeling approach, is the *production-rule* systems approach. The premise underlying production rule systems is that choice behavior in a certain context (e.g. route choice) can be described by a series of "if-then" rules. If the condition in the "if" part is true, then the action of the "then" part follows (see for example Clark and Smith, 1985). This approach has shown the potential to model complicated cognitive processes, however, Bovy and Stern (1990), comment that more research is needed to make such approaches operational.

Teodorovic and Kikuchi (1990), used fuzzy inference techniques to address the standard route choice problem (without presence of information). However, their method is limited to a binary choice, and the methodology suggested can not easily be generalized to deal with multiple choices and attributes other than travel time.

Research on route choice behavior in the presence of information is still in its infancy. Most recently, motivated by the developments in the IVHS area outlined above, the problem of the effect of information on travel behavior in general, and route choice in particular, has received a lot of attention.

Ben-Akiva et al. (1991) propose a framework for modeling the process of drivers' information acquisition and travel behavior. Khattak et al. (1992) investigated commuters diversion propensity using data collected in the Chicago area, and evaluated the ways in which drivers use real-time information. Adler et al. (1992) suggest a framework based on conflict assessment and resolution theories to model individual en-route behavior in response to real time traffic information.

1.4 Thesis Objectives

The purpose of this thesis is to propose a new methodology for modeling route choice behavior in general, and in the presence of information in particular. The main objective is to provide natural and intuitive models that have the potential to capture realistic reasoning and thinking about the route choice decision. The main elements of the methodology are:

1. Modeling perceptions. Perceptions of systems' attributes and of the provided information are of great importance. Modeling of perceptions encompasses diverse problematic issues, including:

- Perceptions can not easily be associated with exact numerical measurements, and often they can only be expressed linguistically.
 - The same attribute's value can be perceived differently by different individuals.
 - The same "declared" perception can correspond to different actual attribute values.
 - Perceptions are being updated continuously.
2. Modeling the decision process. The decision process that accompanies route choice decisions is very complex. Most models make simplifying assumptions regarding this process (e.g. utility maximization, existence of complete information, choosing the optimal alternative and more). In this thesis we suggest a new modeling approach based on fuzzy rules which aims at modeling a more realistic decision making process.

1.5 Thesis Outline

The thesis is organized as follows: In Chapter 2 we present three hypotheses for route choice behavior in the presence of information. Chapter 3 presents approaches for modeling perceptions of network attributes and information using concepts from fuzzy set theory. It also suggests methods for integration of existing knowledge and available information. In Chapter 4 we present two fuzzy route choice models which can accommodate perceptions modeled by fuzzy sets: the fuzzy shortest path, and the fuzzy utility model. Chapter 5 presents a new methodology for route choice behavior in the presence of information: the approximate reasoning model which can be used to implement the three hypotheses developed in Chapter 2. In Chapter 6 the approximate reasoning model is implemented on a small data set that was collected using a driver simulator. Finally Chapter 7 summarizes the major findings and contributions of the thesis and suggests directions for future research.

CHAPTER 2: ROUTE CHOICE BEHAVIOR IN THE PRESENCE OF INFORMATION

A route choice decision can be viewed as an information processing operation in which personal characteristics, physical network attributes, prior experience, prior knowledge, current observations and information are processed together in order to come up with a final choice. An important characteristic of the information processing task involved in making a route choice decision is the need to aggregate different stimuli, data, signs and symptoms, indications, evidence and information in order to reach a final choice. Unfortunately a generic model for information processing, expressed independently of the specific system under consideration, does not exist. However, human behavioral models for judgement, decision and choice have received a lot of attention in general and in the psychology literature in particular (see for example Hammond et al., 1980, for a review of different approaches). The discussion that follows uses the framework for cognitive task analysis proposed by Rasmussen (1986) for information processing in human-machine interactions. This framework aims primarily at the control of a physical system in which the decision maker first *detects* the need for intervention, then *observes* important and relevant data, then analyzes the evidence available in order to *identify* the present state, and *evaluates* its consequences on the systems goals, and finally *chooses* a target state and a task procedure to achieve this desired target state. We use the above framework to describe the route choice process in the presence of information as an information processing task. The general scheme, as described by Rasmussen (1986), is demonstrated in Figure 1, in which rectangles correspond to information processing activities and circles to states of knowledge. The left-hand-side of the Figure corresponds to evaluating the state of the system by processing the data and the information available, and the right-hand-side corresponds to interpreting the identified state of the system in light of the relevant goals, and to making the appropriate choice. In a utility maximization framework, for example, the left-hand-side corresponds to data acquisition and coefficients estimation, and the right-hand-side corresponds to choosing the alternative which has maximal utility. The framework in Figure 1 can be used to describe other forms of decision processes, for example, rule-based systems which associate system conditions (left-hand-side) with choices (right-hand-side).

Figure 1 illustrates the various possible ways (paths) to get from the "activation" node (which defines the task to be performed) to the "execution" node (corresponding to completion of the task). For example, the path connecting the activation node directly to the execution node corresponds to a case of a captive user who does not really make a route choice decision but rather follows the same path whenever at the same decision point. On the other hand, a path that visits all the nodes in the Figure corresponds to a thorough information processing operation in which all the information processing activities are being carried out with no short-cuts or constraints on information processing capabilities. Typically we would expect experienced drivers who do not encounter unusual conditions to have shorter paths (measured in terms of number of links) than

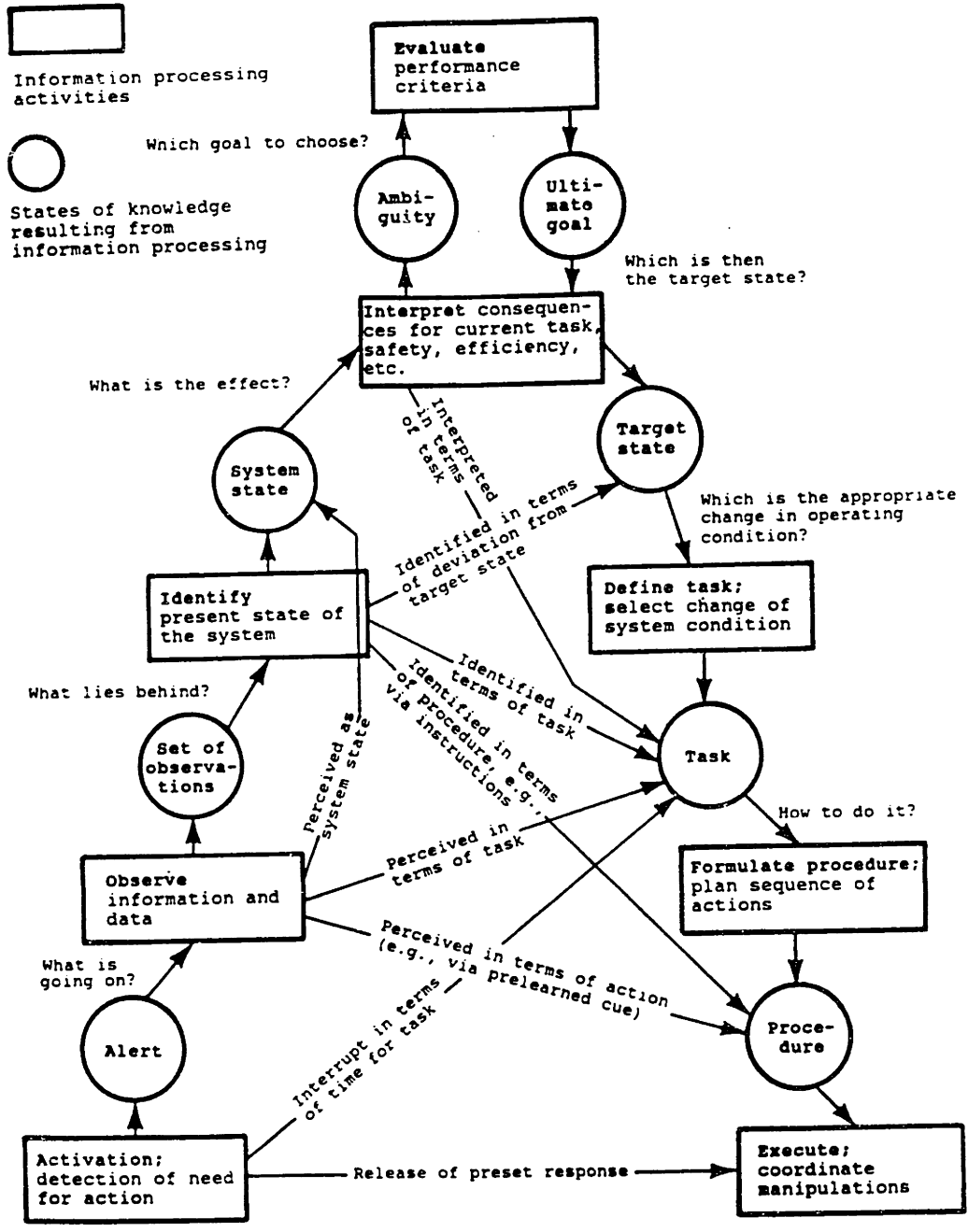


Figure 1: Framework for Cognitive Task Analysis (Rasmussen, 1986)

those of inexperienced drivers, or of drivers facing situations in which unusual conditions occur. Shortcuts from the left-hand-side of Figure 1 to its right-hand-side may also reflect the time pressure constraints, or limited information processing capabilities, or other biases that occur in information processing. Among those biases Hogarth (1980), mentions *selective perception* according to which people seek information consistent with their own views and when found utilize the *law of least resistance* (Rasmussen, 1986), meaning that as soon as an indication is found that a familiar, general search routine may be applied, it is chosen without considering a possibly more efficient procedure. There also seems to be a "point of no return" in the human's attention the moment he/she makes such a decision. The later behavior corresponds, for example, to a user who has decided to take path i as soon as an indication that path j is congested is encountered, without looking further into the state of path i (which could still be worse than path j). In Chapter 5 we discuss rules which associate conditions on an alternative i with attitudes towards taking alternative j for $j \neq i$. Other biases as explanations for shortcuts include the use of *heuristics* such as *rules of thumb*, and *anchoring and adjustments*.

A unique characteristic of route choice as a human information processing task is the necessity to perform this task at a low priority, since the first priority is the driving task. Rasmussen (1986), categorizes human behavior into three levels of performance; the *skill-based behavior* which represents sensorimotor performance without conscious control and corresponds in our case to the driving task (for experienced drivers); the *rule-based behavior* which is a sequence of sub-routines in a familiar work situation consciously controlled by a stored rule, or prepared by conscious problem solving and planning, and corresponds, in our case, to the choice process under familiar conditions; and the *knowledge based behavior* which corresponds to unfamiliar situations, in which a new behavior pattern has to be learned, and corresponds, in our case, for example, to establishing new behavioral rules that would accommodate new alternatives in the choice set, or reaction to information.

An added complication for representing route choice behavior as an information processing task is the possibility of the existence of several processes (modeled as directed paths in Figure 1) running either sequentially or in parallel. The existence of multi-processes corresponds to the existence of multiple considerations and attributes, or to the availability of multiple data sources. Thus, the decision process has to evaluate all existing processes together and derive the resulting choice which could then be viewed as a compromise among the various (possibly conflicting) processes.

2.1 Hypotheses on Models for Route Choice Behavior in the Presence of Information

A unique feature of the route choice problem we are addressing in this work is the existence of *dynamic* on-line current (or projected) traffic information. In this section we present three hypothesis relating to the way users respond to the existence of on-line information while making route choice decisions.

2.1.1 The Simultaneous Model

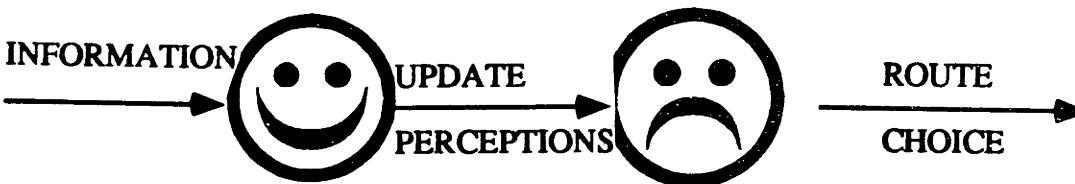
In this model all the inputs of the problem are fed *simultaneously* into the decision process. The inputs include existing knowledge, observation, new acquired information, and other user and network characteristics.

In the simultaneous model no differentiation is made between static-type data (relating for example, to prior perceptions and experience) and dynamic-type data (e.g. on-line information). This model does not make explicit assumptions regarding the nature of the information integration task which processes the new information. Possible implementations of the simultaneous model include:

- Random utility model (to be discussed in Section 4.2.3).
- Approximate reasoning model (to be discussed in Chapter 5).

2.1.2 The Two-Stage Model

This model assumes a two-stage process in which the output from the first stage serves as input to the second stage. At the first stage driver's perceptions are being updated based on the available information. At the second stage, a decision is made based on the updated perceptions. The updating of perceptions at the first phase is short-term, that is, perceptions are being updated for the current state of the system.



To implement the two-stage model, two separate models are needed: one for updating perceptions (to be discussed in section 3.4), and the other for route choice based on updated perceptions.

2.1.3 The Default Model

The default model is based on the premise that the driver has developed a route choice behavior that lies in the background of his/her mind, and that new information is incorporated in light of this existing "default" behavior. Default reasoning corresponds to the process of deriving conclusions based upon patterns of inference of the form: "in the absence of any information to the contrary, assume..". Reasoning patterns of this kind are required whenever conclusions must be drawn despite the absence of complete knowledge.

Common-sense reasoning appears to rely heavily upon the ability to use general rules subject to exceptions. Virtually none of the decisions one makes every day are made with complete certainty. The need to make default assumptions is frequently encountered in reasoning about situations that are not well specified. Default inferences are best viewed as beliefs which may well be modified or rejected by subsequent observations. Default reasoning is tentative in nature. Conclusions drawn from default reasoning can be wrong, but anyone using rules of thumb as common-sense information, is aware that the resulting conclusions would have to be abandoned if contradicted by further evidence.

Recently it has been noted that monotonic logic seems inadequate to capture the tentative nature of human reasoning (see for example, McDermott and Doyle, 1987). Since people's knowledge about the world is necessarily incomplete, there will always be times when they will be forced to draw conclusions based on an incomplete specification of the situation. Under such circumstances, assumptions are made (implicitly or explicitly) about the state of the unknown factors. Because these assumptions are not irrefutable, they may have to be withdrawn at some later time should new evidence prove them invalid. If this happens, the new evidence will prevent some assumptions from being valid, hence all conclusions which build upon those assumptions will no longer be derivable. This causes any system which attempts to reason consistently to exhibit **nonmonotonic** behavior.

The default model assumes a choice process in which at the basic level there exists underlying knowledge, based on prior experience and perceptions. On top of that, additional information may exist which is evaluated in light of the existing knowledge. Naturally default knowledge corresponds to the more static attributes of the problem, while the additional information gathered corresponds to the dynamic attributes. The default model assumes minimal information processing under "usual" conditions. Additional motivation for the default model comes from the conjecture that drivers often "drive without thinking", meaning they do not consciously make a route choice decision, rather they follow their **usual** pattern, and change it only if **unusual** conditions occur. This phenomenon is naturally modelled by default reasoning in which the defaults assume absence of unusual events. Furthermore, it is known that information has the (undesirable) effects of concentration and overreaction caused by extreme reaction to information, which is naturally expressed in the default model as departure from default behavior due to the existence of unusual conditions. Thus, the default model assumes minimal information processing under "usual" conditions. Detection of unusual conditions, though, triggers the need to re-evaluate the default behavior and to check whether it needs to be modified. The attitude towards information in this model is to evaluate whether it is "strong" enough to **modify** the default choice.

Default reasoning is especially attractive for the current state-of-the-art of existing information systems (demonstration projects phase), for which there is still no clear standard as to the best way to provide information, and in which drivers are not yet

familiar with the option of getting on-line information while driving. Thus drivers reaction to information could be based on their behavior in the absence of information, updated to accommodate the new information provided.

In the default model, information on current traffic conditions is used to make specific decisions on whether and how to change the default behavior. This model is consistent with our a priori expectation that drivers resort to heuristics, such as habits and rules of thumb in order to reduce the amount of information processing that is required to make a decision (especially in real time). Hogarth, (1980), calls such behavior anchoring and adjustment, in which an individual has some anchor behavior related to a familiar situation, and this familiar situation is adjusted to current conditions. Hence, according to this hypothesis, drivers may not re-evaluate all parameters involved in the decision process at each point in time, but rather make the necessary adjustments to accommodate current conditions. This behavior also reflects the fact that drivers have to make decisions in real time and have limited information processing capabilities, and thus try to minimize the amount of effort needed to process new information. Furthermore due to the dynamic nature of the problem, special conditions may dominate the decision. An accident report, for example, may cause drivers to divert even if the accident is minor. Thus anchoring and adjustment may be important characteristics of the decision making mechanism.

The default model is also attractive for dealing with prescriptive information systems (such as ALI-SCOUT for example), in which the information provided is the recommended path (or link). In such systems, the recommendation is compared with the original alternative and in case of disagreement between the two, a decision whether to ignore or respond to the information has to be made.

2.2 Discussion of Models

The differences among the three models are demonstrated in Figure 2. As in Figure 1, the left side of Figure 2 corresponds to system evaluation, and the right side to the choice process. When on-line information is available (thick lines), it is fed into different parts of the process depending on the specific model used. In the simultaneous model, current information together with existing perceptions are fed into the system evaluation phase and no direct interactions between prior perceptions and information exist. In the two-stage model, information serves as an input to the existing perception block. Interactions among prior perceptions, observations and information exist, resulting in updated perceptions which are fed into the continuation of the process. As for the default model, there are two processes: first there is a process which derives a choice based on existing perceptions (thin lines), and then, when on-line information is available, it is weighted either directly against the choice from the first phase, or against the decision process which leads to the default choice.

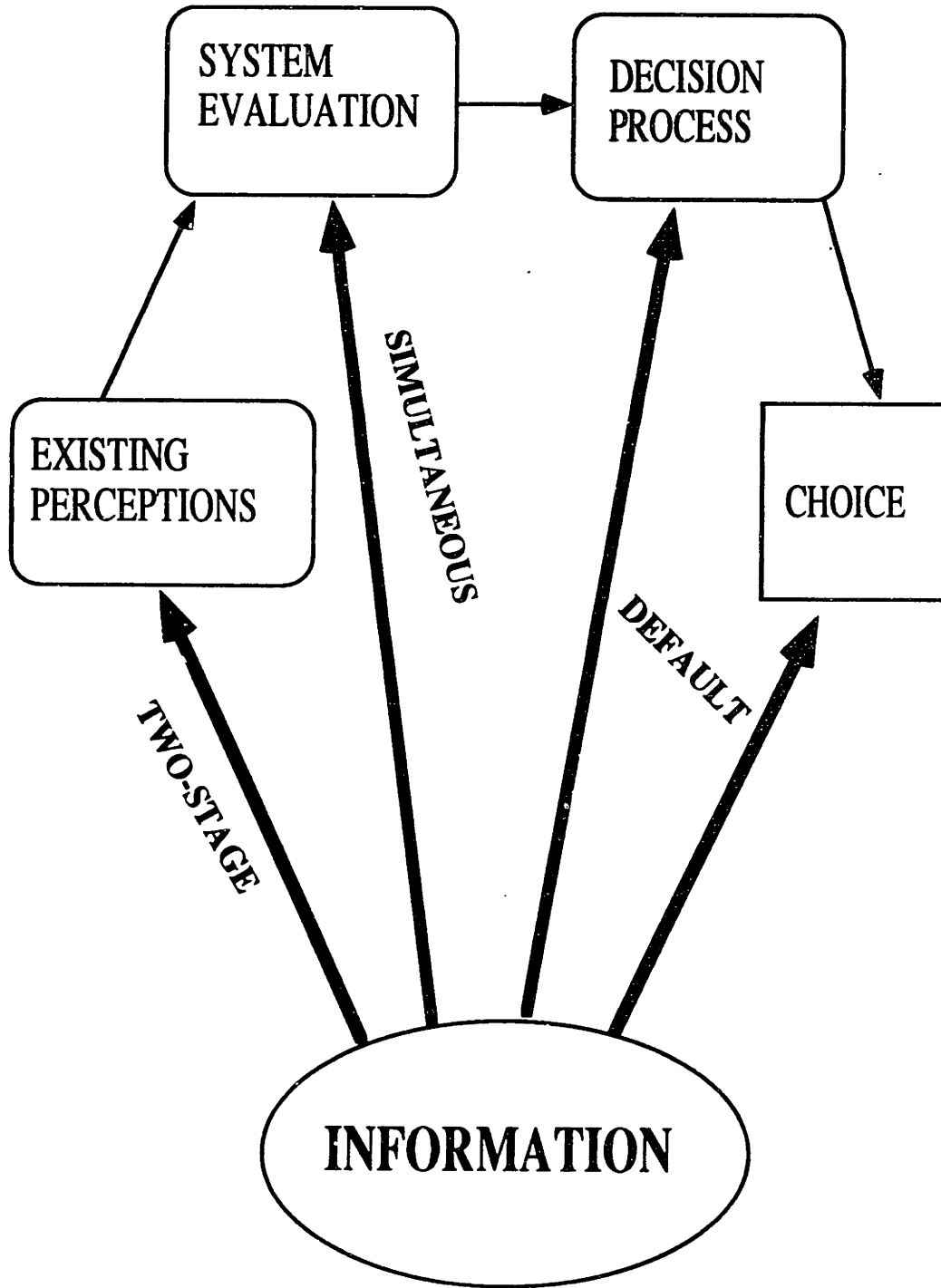


Figure 2: Information Integration Scenarios

The differences among the behavioral assumptions underlying the three models is that in the simultaneous model information is treated as any other attribute that could affect choice and no explicit integration of information is assumed, in the two-stage model information serves to update existing perceptions, and in the default model information is evaluated against the existing behavior pattern (choice).

The choice of model also depends, among other things, on the specific information system available and its reliability. For example, information from a reliable information system is more likely to be integrated at a more basic level (using the two-stage model for example), whereas information that is less reliable (or less available) is more likely to be integrated in light of existing behavior (utilizing the default model). The case in which drivers respond to information only if it reports extreme traffic conditions (e.g. accidents) can be handled more naturally using the default model. It has already been mentioned that for the current state-of-the-art of information systems, the default model, which is based on drivers behavior in the absence of information updated to accommodate the new information, is more appropriate. However, when information systems become an integral part of the driving task, a model which integrates information at a more basic level would be appropriate. The choice of model also depends on drivers personal characteristics. For example, more familiar drivers may do less information processing while driving (thus utilizing the default model, for example).

CHAPTER 3: MODELING PERCEPTIONS AND INFORMATION

In this Chapter we introduce the basic elements of fuzzy set theory, and use them to model perceptions and information. We also discuss methods for integrating information into existing knowledge.

3.1 Fuzzy Set Theory

Fuzzy set theory was developed by Zadeh (1965), as a tool to deal with problems that are characterized by uncertainty and vagueness. The development of fuzzy set theory was motivated by the observation that classical crisp sets are not natural, appropriate or useful in describing human behavior, and by the realization that "... as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes" Zadeh (1973). Zadeh claims that traditional techniques of systems analysis are not well suited for dealing with human systems, because they fail to take into account the existence of fuzziness in human perception and behavior. Thus in order to deal with such systems realistically, approaches which tolerate vagueness, imprecision and partial truth are needed. Zadeh suggests an approach which is based on the premise that the key elements in human thinking are not numbers but labels of fuzzy sets. According to his approach, the logic behind human reasoning is not the traditional binary valued logic, but a logic with fuzzy perceptions, fuzzy truth values, and fuzzy rules of inference.

A **fuzzy set** is a generalization of a crisp set which allows each element to belong to the set with a certain **degree of membership** μ , ($0 \leq \mu \leq 1$), where higher μ values represent higher degrees of set membership. The degree of membership $\mu(x)$ for a value x , represents the degree with which the value x belongs to the corresponding set. The concept of membership function allows the definition of sets with vague boundaries. In general, fuzzy sets enable us to model human oriented systems more realistically by allowing the use of **linguistic** descriptors, phrases, hedges and modifiers (Zadeh, 1973). Each set represents a linguistic label such as: "usual traffic conditions", "HIGH travel times", or "congested intersection".

To demonstrate the basic concepts of fuzzy set theory let us consider travel time as a linguistic variable, that is, a variable whose values are linguistic labels. Then modeling the statement: "travel time is very high" is illustrated in Figure 3. Each travel time, t , has a membership degree $\mu(t)$ which is interpreted as the degree to which the specific travel time t belongs to the set of "very high travel times". Thus t_1 belongs to the set of "very high travel times" with membership degree 0, t_2 has membership degree of $\mu(t_2)$, and t_3 has full membership in the set.

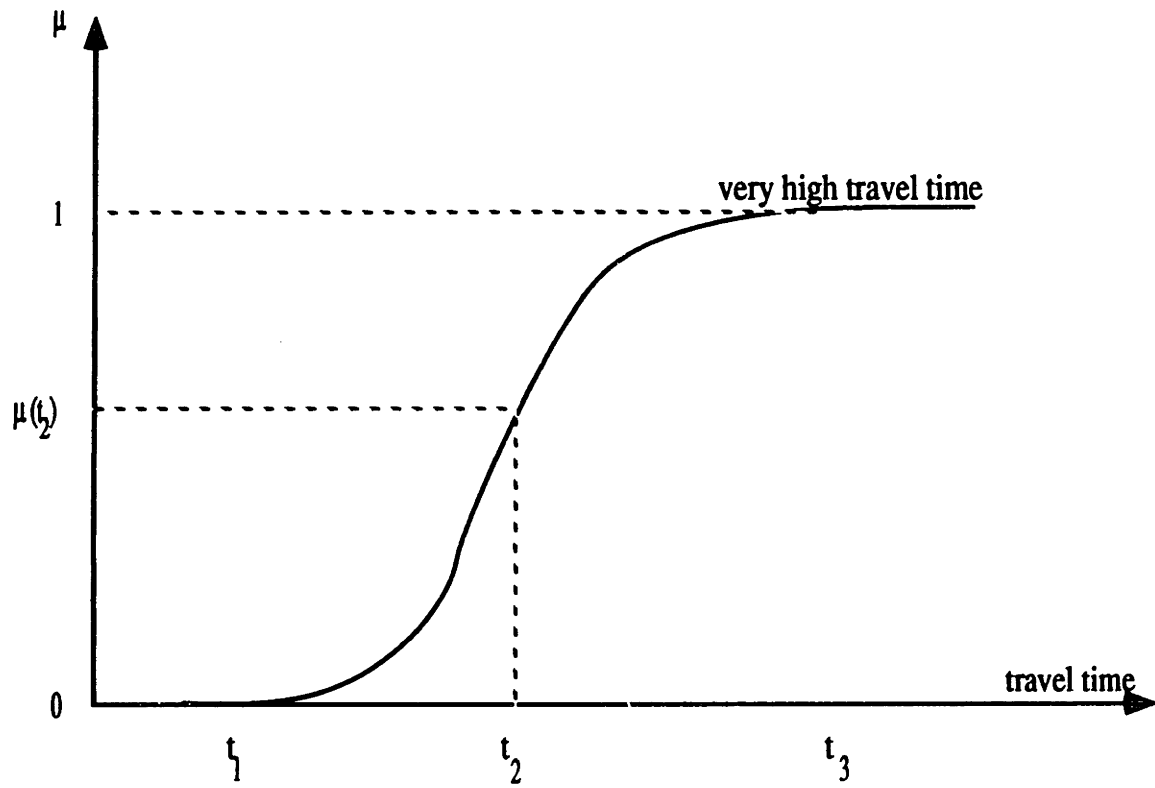


Figure 3: Travel Time as Fuzzy Set

Fuzzy sets are used as the basis for the **theory of possibility** (Zadeh, 1978). Possibility theory relates to the theory of fuzzy sets by defining the possibility distribution of a variable as a fuzzy restriction on the values that may be assigned to that variable. More specifically, if A is a fuzzy set defined on the universe $X=\{x\}$, characterized by the membership function μ_A , then a proposition of the form: "U is A", where U is a variable taking values in X, induces a possibility distribution Π_U which assigns the possibility of U taking a value x to $\mu_A(x)$. Thus a fuzzy variable is associated with a possibility distribution in a similar manner as a random variable is associated with a probability distribution. Zadeh claims, that the importance of possibility theory stems from the fact that much of the information on which human decisions are based is possibilistic rather probabilistic in nature. In particular, the inherent fuzziness of natural languages is possibilistic in origin. As an example, consider the statement: "travel time is very high", which induces a possibility distribution of travel time as given in Figure 3. The possibility degree of travel time being t_2 is given by $\mu(t_2)$, which is interpreted as the compatibility of t_2 with the concept labeled: "very high travel time". The definition of a possibility distribution, Π , implies that the degree of possibility may be any number in the interval $[0,1]$ rather than just 0 or 1. The continuity of possibility degrees relies on the existence of a variety of concepts related to possibility such as: "it is quite possible that..", "it is almost impossible that..", "there is a slight possibility that..".

An ever-lasting debate exists in the literature whether possibility is merely probability in disguise. Clearly the two theories are related: an impossible event is also improbable. However, high degree of possibility does not imply high degree of probability, nor does a low degree of probability imply a low degree of possibility. Yager (1979) suggests using possibility theory when information concerning the uncertainty is not strong enough to make probabilistic statements, or when information does not have the character of probability. He relates possibilistic information to **ordinal** theory of uncertainty, and probability theory to **cardinal** uncertainty. Zadeh (1978) views possibility theory as a means to deal with the **meaning** of information as opposed to the **measure** of information. Similarly Kosko (1992) claims that fuzziness describes **event ambiguity** and measures the degree to which an event occurs, not whether it occurs. An example relevant to our problem, which demonstrates the above differences, is the description of traffic conditions. The occurrence of the event "the intersection is congested" has a certain probability associated with it, while the exact interpretation of congested, or the degree to which the event occurs (or it is perceived to occur) is fuzzy. We believe that possibility should be used to model linguistic concepts, and to represent weaker forms of knowledge than probability, and in section 3.2 we will differentiate between the two theories in the context of our problem.

An important motivation for using fuzzy sets to model perceptions, is their ability to relate to linguistic terms. However, the association between fuzzy sets (represented by membership functions) and their linguistic labels is usually not clear-cut. The problem is two-folded:

- given a linguistic label, what is its corresponding membership function, and:
- given a membership function, what is its corresponding linguistic label.

The first issue relates to the more general issue of generation of membership functions, and the second to the area of linguistic approximation which is still an open research area. Follows is a brief general discussion of membership function generation, and then a discussion of the specific membership functions used for modeling travel time perceptions.

Membership Function Generation

Obtaining or estimating values of membership functions is problematic since by definition they correspond to vague concepts. Furthermore, they represent subjective evaluation and are context dependent. We briefly review here the most common methods for constructing membership functions (see Smithson, 1987, and Turksen, 1986, for further discussion).

The **normative approach** follows Zadeh's original ideas and states that imprecision conveyed by a linguistic variable is subjective and thus needs to be defined directly as a function. According to this approach membership functions have specific mathematical properties, thus the general distribution shape is assumed and parameters of the distribution are estimated from sample data.

The **experimental approach** on the other hand relies on the objectivity of an experimental procedure of measuring and estimates membership values from sample statistics. Under this approach the simplest scheme for generating membership functions is the **binary direct rating** method (or polling) in which subjects are asked whether a value x belongs to a set A . $\mu_A(x)$ is defined as the proportion of subjects who responded "yes". The criticism of this approach is the restriction of the response to yes/no which by itself contradicts the basic idea behind fuzzy sets. In **continuous direct rating**, on the other hand, selected elements are presented to a subject in a random order and the subject is asked to respond to the question: "how A is x " where A is a linguistic label of a set and x is the element. The subject's response is a value on the relevant scale and it is determined by a movement of an indicator along that scale. The same elements are presented a number of times and $\mu_A(x)$ is the appropriate statistic of the responses. In **Reverse Rating** the subject is asked to respond to the question: "identify an element that possesses α grade of membership in A ", where A is a label of a fuzzy set and $0 \leq \alpha \leq 1$.

Fuzzy Set Theory and Route Choice Behavior

Fuzzy sets can be used to model driver behavior with respect to route choice at two levels. At the first level fuzzy sets may provide a natural mechanism to model network attributes (and in the context of ATIS, information) as perceived by the drivers, and to relate to linguistic (and sometime abstract) concepts rather than to exact numerical

measures. And at the second level, fuzzy set theory may provide the foundation for modeling the route choice mechanism itself.

3.2 Modeling Driver Perceptions

Driver perceptions of network attributes may be inaccurate, erroneous or inexact due to uncertainty. Uncertainty, however, has two main sources: **randomness** and **vagueness**.

Randomness relates to uncertainty that is due to the nondeterministic nature of the problem. For example, (apparently) under similar conditions, a driver experiences different travel times on the same route on different days. A driver familiar with the route then, is able to derive an (approximate) distribution of travel times. Therefore, probability measures may be used to model the perceptions of the very familiar driver.

On the other hand uncertainty due to vagueness is mainly related to poor knowledge, no experience, or lack of familiarity with the network. Vague perceptions can not be modeled by probability distribution functions, since those require knowledge of frequencies or likelihood of occurrence, whereas a user who has unclear knowledge about a certain phenomenon can not associate reasonable probability measures with it. Fuzzy sets can be used to model driver perceptions, and incorporate vague knowledge. If, for example, a user is advised to use a path which he/she has never (or hardly ever) used before, then he/she can not associate a probability distribution function with travel times on that path. Instead the user may have some vague idea, based on map information or physical attributes of the path, of which travel times might be **possible** on that path.

In section 1.3 we briefly mentioned factors that affect route choice behavior. The most important factor seems to be travel time, and this factor will be used for demonstration purposes in our discussion. Another group of factors which does not appear there and which is crucial for the current study are factors relating to traffic information. Those factors include the type of information provided, its relevance, frequency, and spatial characteristics, as well as the reliability of the information source.

As discussed in Chapter 2, some of these elements have a dynamic nature, and thus their inclusion in any route choice model needs special considerations. In what follows we will focus on two major groups of factors for explaining route choice behavior: travel time perceptions, and traffic information. Other factors could be incorporated in a similar way.

3.2.1 Travel Time Perceptions

For modeling travel time perceptions, we define a fuzzy set A such that:

$A = \text{"travel times experienced along a certain facility"}$

To represent this fuzzy set we suggest using a trapezoidal fuzzy number (TFN) as illustrated in Figure 4. The definition of a fuzzy number appears in Appendix A.

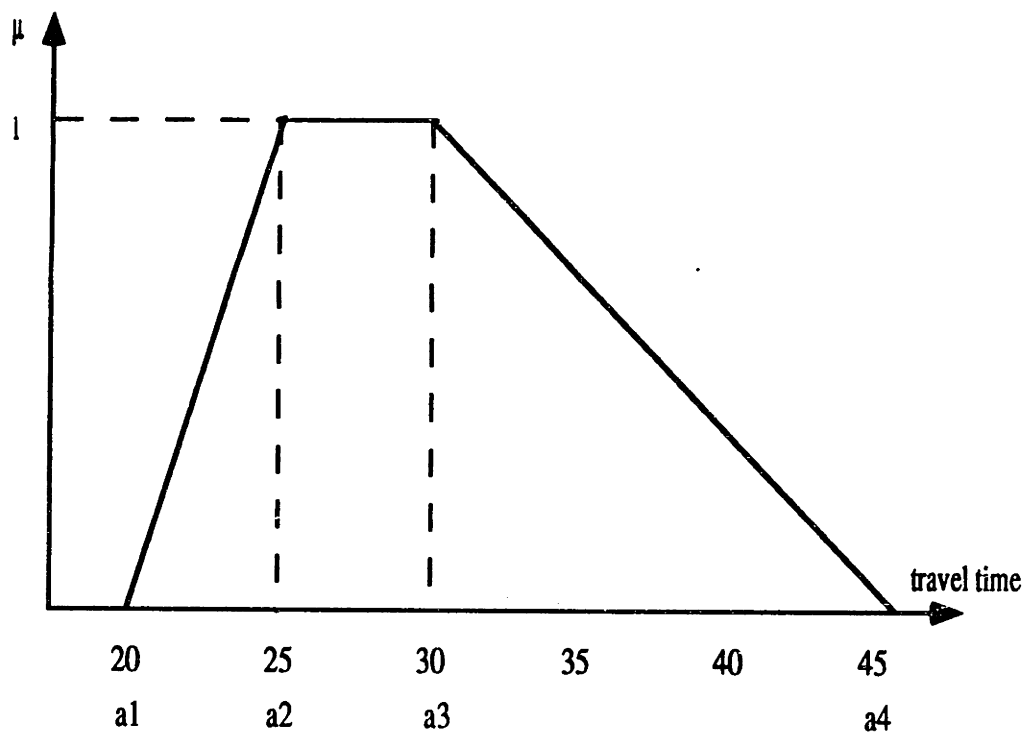


Figure 4: TrFN for Modeling Travel Times

The two main properties of fuzzy numbers, of interest in our case, are normality and unimodularity. Normality of a fuzzy set means that there is at least one value which has full membership in the set (i.e. there exists some x such that $\mu_A(x)=1$). It is a desirable property for the set "travel times experienced along a certain facility" since the attribute under consideration describes a phenomenon that must occur: a certain travel time will materialize. Unimodularity forces the shape of the set to have a single global peak, and monotonicity of membership values towards and/or from that peak. This prohibits cases in which $\mu(x)>\mu(y)$ and $\mu(y)<\mu(z)$, with $x<y<z$.

A T_FFN is determined by 4 points: a_1 , a_2 , a_3 , and a_4 , and is characterized by having a range, $[a_2, a_3]$, of very possible values (with membership degree of 1). It is worth mentioning that in most applications trapezoidal fuzzy numbers or triangular fuzzy numbers (which are a special case in which $a_2=a_3$) are commonly used (see for example Sugeno, 1985). We argue that a T_FFN is appropriate for our case since we expect to have a range of travel times that are very possible, corresponding to travel times that occur under "usual" conditions. The existence of this range is the result of the fact that even under similar traffic conditions, different travel times realizations may occur, and thus more than one travel time gets a membership degree of one. The extreme points of the T_FFN correspond to unusual conditions; a_1 and a_4 in Figure 4 represent the shortest and the longest travel times respectively that are thought to be possible.

In general, we assume that given some physical characteristics of a link (or facility), drivers determine an interval of possible travel times along that link: $[a_1, a_4]$. For each travel time value t , $t \in [a_1, a_4]$, a membership degree, $\mu_A(t)$, is associated, representing the degree of belief with which the driver thinks that travel time t will be experienced. Different familiarities with the link are modeled through both the shape and the range of the membership function. That is, users with poor knowledge can guess the range of possible values of travel time, but have no idea which of these values are more possible than others, while users with good knowledge have a better idea of what values of travel times are possible, and to what degree they are possible. Thus, the shape of the membership function is used to capture the degree of familiarity of users with a particular facility. The more diffused the shape - the higher the ambiguity and the vagueness. In Figure 5, for example, a membership function with rectangular shape corresponds to poor knowledge, meaning that the driver is unable to differentiate among the (long) range of all possible travel times. It is worth repeating the conceptual difference between possibility and probability in this context: the rectangular-shaped membership function in Figure 5 does not represent a uniform distribution according to which all travel times are equally likely to occur; rather it indicates that all travel times within the specified range could occur and the driver is unable to distinguish which values are more possible than others. A better knowledge is characterized by a less diffused (trapezoidal) possibility distribution, corresponding to the case in which a driver can specify a range of very possible travel times, and has knowledge on unusual travel times as well. Even better knowledge can be represented by a probability density function, and deterministic knowledge as illustrated in Figure 5.

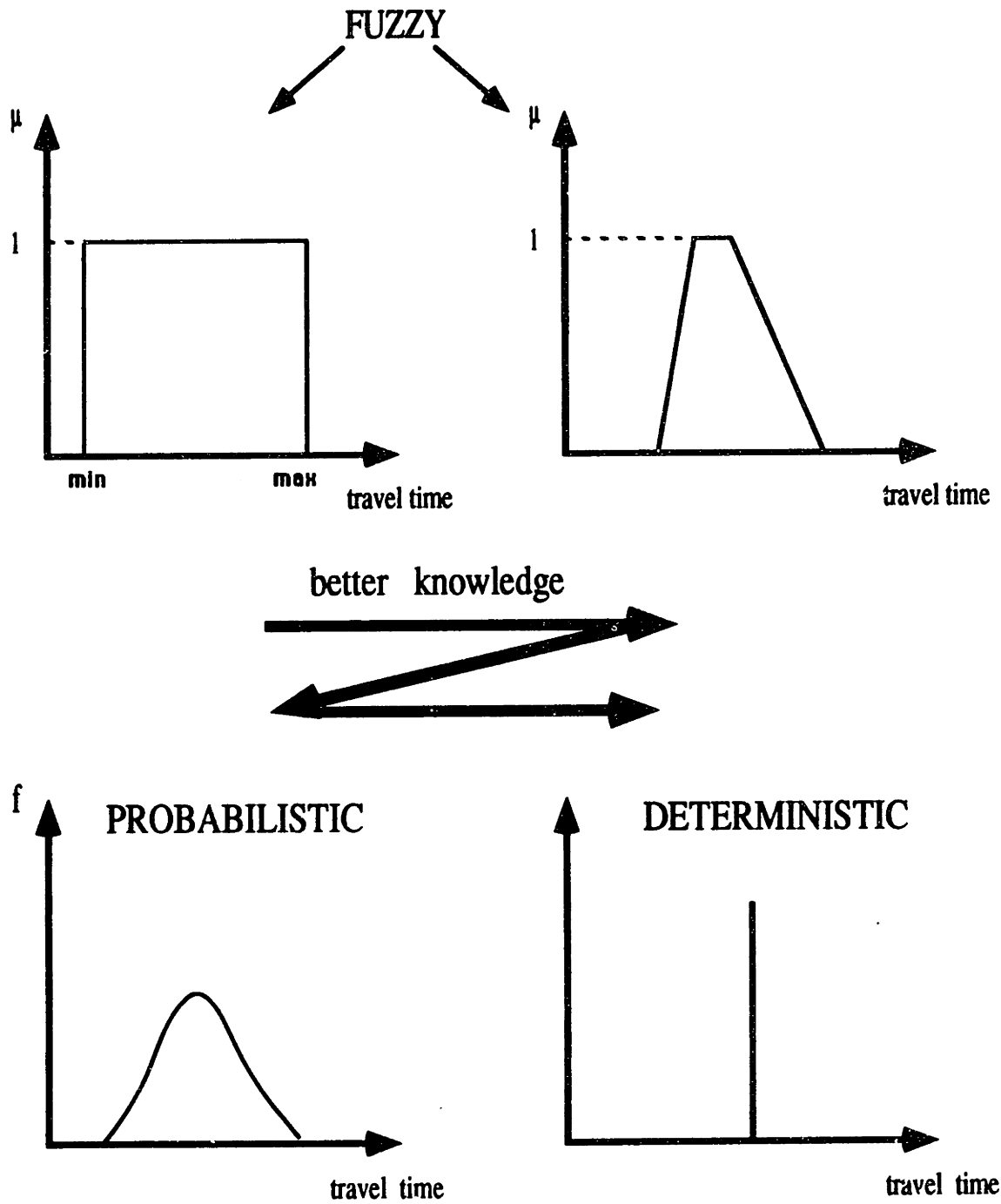


Figure 5: Travel Time Perceptions

The range and the shape of the possibility distribution also reflect driver and facility characteristics. A daring driver, for example, would consider a narrower range of possible travel times and the range would probably be located around shorter travel times than those of a conservative driver. In a similar fashion, different facilities may be characterized by various widths of the range of very possible travel times.

The interpretation of fuzzy sets presented above is also supported by the fact that uncertainty is related to ambiguity of evidence. For measuring ambiguity, Klir and Folger (1988) have generalized Hartley's information measure into a measure of non-specificity $U(\mu_A)$ associated with a fuzzy set A. Let μ_A be the membership function of the fuzzy number A, then its non-specificity measure, $U(\mu_A)$, is defined by:

$$U(\mu_A) = \int_0^1 \log_2 |A_\alpha| d\alpha \quad (3-1)$$

where $A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\}$ and $|A_\alpha|$ is the length of the interval A_α . This measure of uncertainty satisfies (among others) the following desirable properties:

- **Monotonicity** - for any pair of fuzzy numbers A and B with possibility distributions μ_A and μ_B with the same support, if $\mu_A(x) \leq \mu_B(x)$ for all x, then $U(\mu_A) \leq U(\mu_B)$.
- **Maximum** - among all possibility distributions of the same support, the possibility distribution with all elements having membership degree of 1 has the highest $U(\mu)$ value.
- **Minimum** - $U(\mu)=0$ iff exactly one member of μ has membership degree 1 and the rest - zero.

Thus the non-specificity measure is consistent with our hypothesis that the more diffused is the shape, the weaker (less specific) is the knowledge. A rectangle-shaped membership function, which corresponds to unfamiliar drivers who can not differentiate among the different possible travel times, has a maximum $U(\mu)$ value (over the given range). Corresponding to better knowledge, is a trapezoidal-shaped membership function which has a lower $U(\mu)$ value.

3.2.2 Empirical Evidence

So far we have argued that the shape of fuzzy sets used for modeling travel time perceptions is influenced by three main groups of factors: personal characteristics (such as socio-economic attributes, risk aversion, etc.), facility-related attributes (such as length, number of lanes, etc.), and familiarity with the facility. In this section we explore these assumptions using data collected in a commuter survey.

In May 1991 we conducted a survey among the MIT parking permit holders, concerning their home to work commute. The survey (see Appendix C) was composed of two parts. PART I related to the usual commute to MIT, and asked questions concerning usual routes, departure time, travel time, attitude towards traffic reports on the radio, and personal characteristics. PART II related to the day-to-day commute during a specific week, with emphasis on route choice, route switching, traffic information received on the radio, and actual traffic conditions. A total of approximately 1300 sticker holders responded to PART I of the survey. 1200 of them were drivers who regularly commute by car to MIT. Approximately 900 drivers responded to PART II. A total of 3800 trips were made by those drivers.

In this section we explore the response to questions 6, 8, and 9 in PART I of the survey as indicative of travel time perceptions.

6. *Think about a typical trip from your home to work. Assume "regular" traffic conditions, i.e. no extreme traffic delays, no major incidents and no weather related problems. Under these conditions, how long does it usually take you to get from your home to work? please specify a range (e.g. from 40 to 55 minutes): from _____ to _____ minutes.*
8. *What is the shortest time you have experienced during your home to work commute? _____ minutes*
9. *What is the longest time you have experienced during your home to work commute? _____ minutes*

Question number 6 relates to the range of "usual" travel times which we interpret as travel times that are very possible (with possibility degree of 1), and corresponds to the flat part of the T_FFN, that is the segment $[a_2, a_3]$ in Figure 4. Questions 8 and 9 ask about the extreme cases: what are the shortest and the longest possible travel times, and the responses correspond to a_1 and a_4 in Figure 4 respectively. It is important to note that questions 8 and 9 refer to extreme travel time values actually experienced by the driver, and thus could include two types of biases; on one hand extreme traffic conditions could be caused by unusual situations (such as major accident, or severe weather conditions) which are not perceived as possible to occur on a daily basis, and on the other hand it differentiates between experienced and unexperienced drivers since experienced drivers are more likely to have encountered more extreme values.

A total of 1138 commuters responded to all 3 questions. The frequency of the different T_FFN shapes appears in Table 1. As expected, nobody reported a single value as the "usual" travel time (i.e. for all cases we have: $a_2 \neq a_3$). This supports our choice of T_FFN (and not triangular fuzzy number).





order relations	frequency %	general shape
$a_1 < a_2 < a_3 < a_4$	74.0	
$a_1 = a_2; a_3 \neq a_4$	20.3	
$a_1 \neq a_2; a_3 = a_4$	3.7	
$a_1 = a_2; a_3 = a_4$	2.0	

Table 1: Frequency of T_{FN} Shapes, General Population: K=1138

The results presented in Table 1 are consistent with our interpretations of "usual" and "extreme" traffic conditions.

Effect of Familiarity on Travel Time Perceptions

To examine the effect of familiarity on travel time perceptions, the length of time at present address, and length of time at current job location (questions number 25 and 30 in PART I) were used as measures of familiarity with the network. In the survey 94% of the 1138 respondents both lived at their present address, and served at their current job location more than a year, so in general we can consider the overall population to be "familiar". This fact may explain the low percentage of drivers with rectangular shaped travel time perceptions (2%). Table 2 shows the distribution of travel time perceptions among the "unfamiliar" drivers (those who either lived at current address less (in the strong sense) than a year, or served at current job location less than a year).





order relations	frequency %	general shape
$a_1 < a_2 < a_3 < a_4$	62.2	
$a_1 = a_2; a_3 \neq a_4$	28.8	
$a_1 \neq a_2; a_3 = a_4$	4.5	
$a_1 = a_2; a_3 = a_4$	4.5	

Table 2: Frequency of Tr_{FN} Shapes, Unfamiliar Population: K=66

The results shown in Table 2 support our hypothesis that perceptions of unfamiliar drivers are less distinct than perceptions in the overall population; the case of $a_1 < a_2 < a_3 < a_4$ which corresponds to better knowledge is less frequent among the unfamiliar population, and the frequency of the rectangular shaped fuzzy set ($a_1 = a_2$ & $a_3 = a_4$) is larger than the

frequency in the overall population. Recall that the population in Table 2 is a subset of the population in Table 1.

Table 3 shows averages, standard deviations, and medians of the four points that constitute the T_{FN} distribution.

	Familiar Population K=1054				Unfamiliar Population K=66			
	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4
mean	23.96	29.24	39.97	69.45	24.23	28.65	39.67	60.67
std. dev.	13.65	15.23	17.95	46.00	14.83	17.32	21.17	40.37
median	20.00	25.00	40.00	60.00	20.00	22.50	35.00	50.00

Table 3: T_{FN} Distribution for Travel Times

The distributions show clear skewness towards high travel time values. Drivers experienced very high travel times (relative to the "usual" range) which are farther from the usual range than are the very low travel times. This phenomenon is indicated by a_1 being closer to a_2 than a_3 is to a_4 , and by the median being (almost) always less than the mean. This is explained by the fact that low travel times are bounded by the free flow travel time, whereas high travel times are not bounded from above. The familiar population exhibits a slightly smaller average a_1 value and a higher a_4 average value (but the differences are not statistically significant). This is explained by the fact that familiar commuters have more experience and thus they are more likely to encounter unusual conditions (both good and bad), but it could also be related to O/D characteristics. Note that typical commute time (as indicated by a_2 and a_3) does not differ significantly between the two populations.

To explore the effects of familiarity further, we define the ratio R by:

$$R = \frac{a_3 - a_2}{a_4 - a_1} \quad (3-2)$$

R serves as a measure of familiarity: the higher this measure - the less familiar is the commuter, since for familiar users we expect to have larger $[a_1, a_4]$ range and smaller $[a_2, a_3]$ range. And indeed the difference between the ratios in the two populations is statistically significant.

	Familiar Population K=1054			Unfamiliar Population K=66		
	a_3-a_2	a_4-a_1	R	a_3-a_2	a_4-a_1	R
mean	10.73	45.49	.32	11.02	45.03	.39
std. dev.	5.33	40.12	.19	5.54	39.66	.20
median	10	35	.29	10	30	.375

Table 4: Ranges of Possible Travel Times

Effects of Origin on Travel Time Perceptions

Effects of the specific facility used can be inferred from the origin information. From the survey we do not have exact origins specification, but we do have zip codes. We look at two extreme cases: drivers who live in Cambridge and have the same zip code as MIT (02139), and drivers who commute from Lexington (zip code 02173).

origin		a_1	a_2	a_3	a_4	a_3-a_2	a_4-a_1	R
Cambridge K=20	mean	9.10	11.9	18.75	34.6	6.85	25.50	.45
	std. dev.	6.52	9.94	12.78	50.04	5.92	48.83	.22
Lexington K=52	mean	24.77	30.15	41.72	66.02	11.56	41.25	.35
	std. dev.	5.09	5.89	7.21	30.76	4.74	29.15	.18

Table 5: Travel Time Perceptions by Origin (Cambridge and Lexington)

The Lexington commuters experience a longer commute, and their R value is significantly smaller than the R value of the Cambridge commuters. The ratio R measures how diffused is the shape of the set, and accordingly, the Lexington commuters have a less diffused shape, meaning that their $[a_2, a_3]$ interval is relatively more focused, and that their range interval, $[a_1, a_4]$ is relatively larger. A large $[a_1, a_4]$ range follows from the fact that in a long commute, extreme values are more likely to occur, while relatively small $[a_2, a_3]$ range indicates that commuters have a better idea which travel times are more possible than others.

3.3 Modeling Information

In general, information includes any type of knowledge acquired from the time the driving task to be performed was initiated until it was accomplished. In particular, we

focus on traffic information as the main source for dynamic knowledge, although other information sources, such as observation, can also be included.

Although there are several demonstration projects underway, there is still no clear standard with respect to the way traffic information should be provided. It is clear however, that for most of the possible information provision scenarios, information would be provided linguistically (e.g. congestion descriptions, accident information), or symbolically (e.g. map information, shortest path indication).

Fuzzy set theory provides an intuitive and natural way for modeling the information provided under the different scenarios. Fuzzy sets can be used to model vague concepts conveyed by the information system (e.g. "traffic is bad this morning"), ambiguous descriptions (e.g. "the bridge is congested"), recommendations (e.g. "you'll do slightly better on the other side of the river"), accident reports and more.

The three hypothesized models that were presented in section 2.1 handle existence of traffic information differently, and thus require different modeling of information:

- The simultaneous and the default models treat information as another source of input, and thus are quite flexible in handling different types of information.
- The two-stage model uses information to update perceptions at the first stage. Thus, a natural way to model information for this case is on the same scale as perceptions are modeled. Therefore this model is less flexible in relating to the various information provision scenarios.

Information can be absolute or relative. Absolute information pertains to information that is conveyed as objectively as possible, such as: queue lengths, closed lanes, detours, malfunctioning or defective traffic lights, etc. Relative information, on the other hand, relates to subjective estimates of traffic conditions, such as descriptions of congestion levels in comparison to some reference state. Given the type of information used (absolute or relative), there remains the issue of its units. For example, an accident report can be modeled in various ways: it can be reported linguistically (e.g. there is a major accident), or it can be translated into its travel times, delays, and queue consequences.

An important issue with respect to modeling information is that of perception: different drivers might perceive the same information in different ways depending on their specific O/D pair and their personal characteristics. On one extreme, information could be modeled as is (if possible), and personal characteristics could then be captured through the existing knowledge, or the decision process. On the other extreme, information is modeled as perceived by individual users.

3.4 Information Integration and Update of Perceptions

A crucial component in the information processing scheme is drivers ability to combine different pieces of data, signs and symbols, together with existing concepts, perceptions, experience and knowledge. Evidence aggregation has many manifestations in a route choice process. It pertains to combining a priori expectations with current observations, observation with information, a priori perceptions with current perceptions or information, and information from different sources. Before and while a route choice decision is made, a user acquires current traffic information. A route choice decision made after new information has been acquired, usually takes into consideration the new information along with existing knowledge. Thus information integration relates to combining two types of knowledge: *static* with *dynamic*. The static-type knowledge pertains to static system attributes such as network topology, and typical conditions. It also relates to existing long term perceptions and knowledge, based on prior experience and familiarity with the system. Dynamic-type knowledge on the other hand, pertains to specific current (or projected) conditions in the system which relate to the current choice (and not necessarily to other similar choices). The dynamic-type information can be acquired from several sources (e.g. observation, traffic reports).

An explicit model for updating perceptions is needed in order to implement the two-stage model, whereas in the simultaneous model no differentiation between static and dynamic attributes is made, and in the default model, the interaction between static and dynamic attributes is part of the decision process. In this section we focus on information integration for combining existing knowledge with information as it pertains to the first stage of the two-stage model. The problem of information integration can be phrased: how to combine the existing knowledge on travel time, "T is A" with the on-line (possibly incomplete) information, "T is C", where A and C are fuzzy sets corresponding to possible travel times as discussed before, and are modeled on the same scale.

The state of the art in the area of information integration is very limited and the difficulty in developing realistic models has been recognized by other researchers (Ben-Akiva and Kaysi, 1991). Furthermore the lack of any systematic data of drivers' perceptions before and after receiving information, makes the task of developing and testing even the simplest of models very difficult.

A priori factors that may influence the process of information integration include the strength of drivers prior perceptions, reliability of the information system, salience, relevance and quality of information, as well as driver and network characteristics. An appropriate information integration model should be able to capture the following aspects:

- Non-linear update between prior perception and real time information may exist. Extreme information may cause the updated perception to "jump", whereas mild information could have a negligible effect.

- Additional information does not always help the driver focus his/her perceptions, and cases in which the driver's perceptions after receiving the information are "fuzzier" than before are possible. Similarly, the updated perceptions can be "fuzzier" than the provided information.
- Information integration is a dynamic phenomenon. Updating of perceptions based on information received at time t is influenced by the information received at time $t-1$.

In the following sub-sections, we discuss two approaches for information integration: information integration theory (section 3.4.1), and default reasoning (section 3.4.2).

3.4.1 Information Integration Theory

Information integration theory was developed by Anderson (1981, 1982) and has since been implemented mostly in psychological and cognitive applications.

The underlying framework is as follows:

- A physical stimulus S is presented to a subject.
- The subject perceives this stimulus as having a psychological value s .
- Based on the psychological value, s , the subject establishes a response r to the stimulus.
- The perceived response, r , is translated into an observed physical response R .

In this framework S and R are observed while s and r are latent, and they are related through the following main functions:

Perception function:	$P(S) \rightarrow s$
Integration function:	$I(s) \rightarrow r$
Response function:	$M(r) \rightarrow R$

For our case:

S is the information provided,
 s is the perceived information,
 r is the updated perception, and
 R is the observed response, which can be either the reported updated perception, or the observed choice.

The efficacy of the approach lies heavily in the hypothesis that stimulus integration often obeys algebraic models. However, all models make the strong assumption that the response function is linear, that is R is a linear function of r . This

assumption is obviously not valid if R is the observed discrete choice. Thus use of the models requires that R will be the updated perceptions as reported by the subjects.

The most common models that have been proposed for information integration are summarized below:

- The linear model: $R_{ACx} = K_0 + w_A * s_A + w_C * s_C + \epsilon_{ACx}$
- The multiplicative model: $R_{ACx} = K_0 + s_A * s_C + \epsilon_{Ax}$
- The averaging model: $R_{ACx} = K_0 + w_A / (w_A + w_C) * s_A + w_C / (w_A + w_C) * s_C + \epsilon_{ACx}$

where: K_0 is a constant,
A is existing knowledge,
C is the information provided,
 w_A and w_C are weights of importance of stimuli A and C respectively,
 R_{ACx} is the observed response to stimuli A and C as reported by individual x, and
 ϵ_{ACx} is the error term

The linear type model for information integration ignores some of the important characteristics of the problem, mainly it assumes no interaction between existing knowledge and information. Of the three proposed models (linear, multiplicative, and averaging), the averaging model seems the most appropriate for our problem. In this model, the effect of each stimulus depends on what other stimuli it is combined with and thus the model is inherently non-linear. It provides a natural way to model conservatism on one hand (e.g. high w_A value, if A is the "usual" condition), and dominance of certain information pieces on the other (e.g. very high w_C value, if C is an accident report).

If direct observations on updated perceptions become available (by interviewing individual drivers with carefully designed questionnaires), then weight values can be estimated or expressed as functions of the characteristics of the system and the drivers.

3.4.2 Default Reasoning

The concepts behind default reasoning that were presented in Chapter 2 in the context of the default model, could be used for developing models to update perceptions. For updating perceptions after receiving information, we treat the existing knowledge as default knowledge, meaning that it applies as long as there is no strong evidence to change it, and when evidence which contradicts this default knowledge exists (from a relatively reliable source), the updated perceptions are more likely to be closer to the new evidence. Accordingly, travel time perceptions would have their "usual" values (depending on experience, time of day, etc) as long as no information on accidents or extremely unusual conditions exists.

Dubois and Prade (1988) suggested the following approach for combining different sources or pieces of information. Let n be the number of information pieces available, μ_i the evidence from source i , w_i the weight of importance of information i , and $\max_i w_i = 1$ (i.e. weights are normalized). Then the combined evaluation, μ , is given by:

$$\mu = \min_{1 \leq i \leq n} \max(\mu_i, 1 - w_i) \quad (3-3)$$

For the discussion that follows we will assume that information comes from the same source, and we will focus on integrating information into existing knowledge. Let's assume that the existing knowledge of travel time, T , is given by the fuzzy set A with membership function $\mu_A(\cdot)$. If no further information is available then $\mu_A(\cdot)$ is used to represent travel time perceptions. Now let us assume that information regarding travel time, T , is available, and it is modeled by the fuzzy set C with membership function $\mu_C(\cdot)$. The combination scheme given by equation (3-3) can now be applied by observing that there are two sources of evidence: existing knowledge and the provided information. Hence, the evidence of the first source, is given by the membership function of the set A , and similarly the evidence for the provided information is given by the membership function of the set C . We assume that the information comes from a relatively reliable source, and thus it has greater importance since it corresponds to actual conditions. Therefore we set: $w_C=1$. A natural weight for w_A , which captures the importance of existing (default) knowledge, is the degree of consistency of existing knowledge A with the information C , given by the possibility of A given C , that is:

$$w_A = \text{Poss}[A|C] = \max_x \min(\mu_A(x), \mu_C(x)) \quad (3-4)$$

Substituting $w_C=1$ and $w_A=\text{Poss}[A|C]$ in equation (3-3), the membership function of the updated perception, $C \cdot A$, becomes:

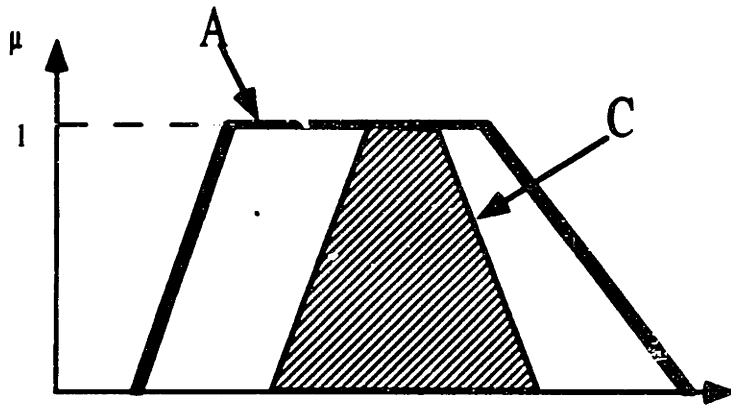
$$\mu_{C \cdot A}(v) = \min[\mu_C(v), \max(\mu_A(v), 1 - \text{Poss}[A|C])] \quad (3-5)$$

$\text{Poss}[A|C]$ measures the consistency between existing knowledge and the information. If $\text{Poss}[A|C]=0$ then A and C are inconsistent and the updated perception coincides with C and default knowledge is ignored. When $\text{Poss}[A|C]=1$, existing knowledge and information are consistent, and equation (3-5) collapses into:

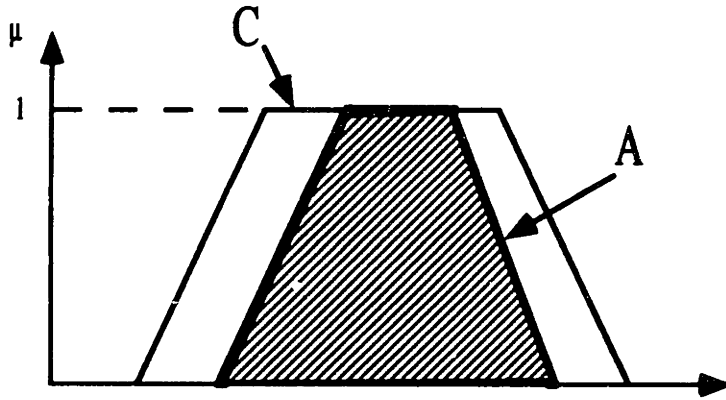
$$\mu_{C \cdot A}(v) = \min(\mu_C(v), \mu_A(v)) \quad (3-6)$$

In this case A and C are treated equally and the updated perception is the intersection of the two sets.

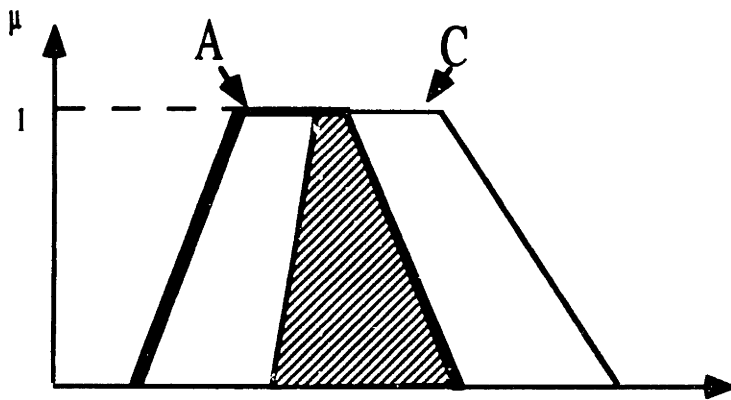
For the case of $\text{Poss}[A|C]=1$ there could be three types of overlap between A and C (as demonstrated in Figure 6):



(a)



(b)



(c)

Figure 6: Updated Perceptions, $\text{Poss}(A|C)=1$

1. $C \subset A$ meaning that the information is more specific than existing default knowledge as illustrated in Figure 6a. The updated perception, given by the shaded area set, coincides with C meaning that information has made existing knowledge more specific.
2. $A \subset C$ meaning that the information is less specific than existing knowledge (see Figure 6b). The updated perception, given by the shaded area, coincides with A meaning that the new information could neither change existing knowledge nor refine it.
3. $A \not\subset C$ and $C \not\subset A$, meaning that although A and C are consistent (their ranges of very possible travel times overlap), each has elements that are not common to the other as illustrated in Figure 6c. For this case the updated perception is taken to be the intersection of the two sets as illustrated by the shaded area in Figure 6c. For that case, information serves again (as in the first case) to refine existing knowledge.

Thus, when existing knowledge and information are consistent (i.e. $\text{Poss}[A | C]=1$), the updated knowledge is a more focused representation of either the prior knowledge (Figure 6a), or the information (Figure 6b), or both (Figure 6c). This result agrees with our expectation that information which is consistent with existing knowledge is used to refine it (if possible).

When $0 < \text{Poss}[A | C] < 1$, existing knowledge and information are consistent to some degree. To explain the interpretation of the integration scheme for this case, we consider the approach suggested by Dubois and Prade (1988) for modeling default sets. Their approach aims at dealing with situations in which sufficient information to determine whether an exceptional situation has occurred does not exist, and with cases in which the nature of some of the exceptions is unknown. Given a fuzzy set A, Dubois and Prade define a fuzzy default set A^λ with membership function:

$$\mu_{A^\lambda}(v) = \max(\mu_A(v), 1 - \lambda) \quad \lambda < 1 \quad \forall v \quad (3-7)$$

where A gathers the a priori, more plausible values on the phenomenon, and $1 - \lambda$ estimates to what extent it is possible that the phenomenon described by the fuzzy set A takes its value outside A (the values outside A are considered to be exceptions). $\lambda = 0$ corresponds to complete lack of knowledge, and thus results in a flat shaped membership function representing the worst case of knowledge as discussed in section 3.2, and $\lambda = 1$ corresponds to the case of no exceptions (A^λ coincides with A).

Updated perceptions, as given by equation (3-5), can then be viewed as the intersection of the information C with the default set A^λ , given by equation (3-7) with $\lambda = \text{Poss}[A | C]$. That is, since A is not completely consistent with the information C, it is generalized to accommodate exceptions, where $1 - \lambda$ is the possibility of exceptions. We demonstrate the approach in Figure 7 by examining the two possible cases:

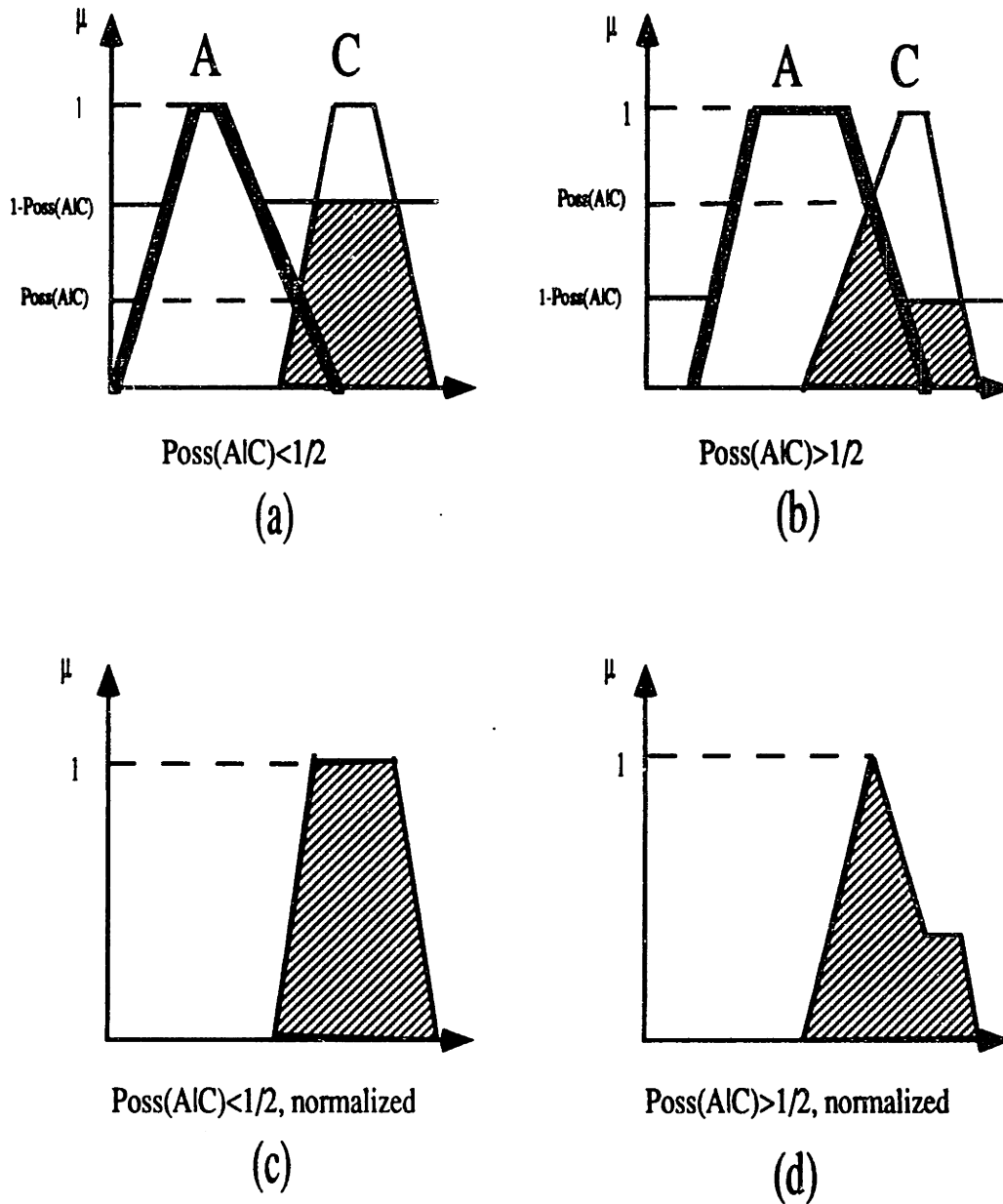


Figure 7: Updated Perceptions, $0 < \text{Poss}(A|C) < 1$

1. $0 < \text{Poss}[A | C] \leq 1/2$: Existing knowledge, A, is quite inconsistent with the information, C, and the possibility of exception, $\lambda = 1 - \text{Poss}[A | C]$, is greater than $1/2$, meaning that the default set, A^λ , is very defused, and the information, C, dominates the updated perception (represented by the shaded area in Figure 7a).
2. $1/2 < \text{Poss}[A | C] < 1$: In this case existing knowledge is quite consistent with the information, and thus the possibility of exception is considerably low (less than $1/2$). The updated perception is then composed of all elements which are common to A and C (have positive membership degree in both sets), and elements belonging to the information set, C, truncated by the possibility of exception, (as illustrated by the shaded area in Figure 7b).

When $0 < \text{Poss}[A | C] < 1$ updated perceptions are always a truncated version of C, and thus are not normalized (i.e. there is no x such that $\mu(x) = 1$). However, following the case analysis above, it is clear that the height of the updated perceptions is always greater or equal to $1/2$. Updated perceptions could be normalized by dividing the membership function of $C * A$ by the height of $C * A$. That is the membership function of the normalized set $C * A$ is given by:

$$\mu_{C * A}^{\text{norm}}(x) = \frac{\mu_{C * A}(x)}{\max_y(\mu_{C * A}(y))} \quad (3-8)$$

Dubois and Prade (1988) indicate that this normalization means that the actual value looked for definitely belongs to $C * A$. This is consistent with our previous discussion about travel time perceptions (at least one value is certain to occur and thus its membership degree is 1). The normalized versions of the updated perceptions in Figures 7a and 7b appear in Figures 7c and 7d respectively, and result in fuzzy numbers (as defined in Appendix A). For the case of $0 < \text{Poss}[A | C] \leq 1/2$, the normalized updated perception is always a superset of the information C, and thus less specific than the information (according to the specificity measure of equation (3-1)). This behavior is consistent with our a priori expectations regarding the requirements from an integration scheme; the fact that information overlaps with prior knowledge only to a small degree (less than $1/2$), results in updated perceptions which are a "fuzzier" version of the provided information.

In summary, the suggested combination scheme (as given by equation (3-5)), shows that default knowledge is used to refine the available information when the two are consistent, it is ignored when they are totally inconsistent, and it serves to modify the information when quite consistent with it.

If the information source is not very reliable, then we could change the assumption underlying equation (3-5) that the new information "T is C" has maximal importance, into giving the maximal importance to the default knowledge, and using $\text{Poss}[C | A]$ as degree of consistency of the new information. The updated perception then is given by:

$$\mu_{C \cdot A}(v) = \min[\mu_A(v), \max(\mu_C(v), 1 - \text{Poss}[C|A])] \quad (3-9)$$

The problem with this approach is that it would bias the integration towards information that is consistent with default perceptions, even though the most useful information is typically information concerning incidents and unusual events which is by definition inconsistent with default expectations. Such information would be ignored if we utilize the integration scheme given by equation (3-9): if $\text{Poss}[C|A]=0$, for example, (meaning that information is completely different from existing knowledge), then the updated perception, $C \cdot A$, coincides with prior knowledge A , and the information C is completely ignored.

A better approach to deal with unreliable information sources would be to let the information provided be subject to "exceptions", where an exception in this context corresponds to the possibility that the information is incorrect, and hence encountering values outside the inferred possibility distribution is possible. To implement this, let ρ be the reliability of the information and $1-\rho$ the possibility of encountering values different from those provided. Then the membership function of the new information "T is C" is given by:

$$\mu_{C^\rho}(v) = \max(\mu_C(v), 1 - \rho) \quad \rho < 1 \quad \forall v \quad (3-10)$$

The information integration formula, given by equation (3-5), can now be used with the set C^ρ instead of C :

$$\mu_{C^\rho \cdot A}(v) = \min[\mu_{C^\rho}(v), \max(\mu_A(v), 1 - \text{Poss}[A|C^\rho])] \quad (3-11)$$

When $\rho=1$, the information source is reliable and equation (3-11) is equivalent to equation (3-5). If, on the other hand, $\rho=0$, corresponding to completely unreliable information, then the new information is ignored, and the resulting perception coincides with the default knowledge μ_A . For the case of $0 < \rho < 1$ the behavior is very similar to the case of reliable information (where the updated perception was always a subset of C), but for the new set C^ρ . For example, if $\text{Poss}[A|C^\rho]=0$ then the updated perception is given by the set C^ρ .

Another possible weight for prior perception (instead of $\text{Poss}[A|C]$) is the conditional certainty of A with respect to C , $\text{Cert}[A|C]$, as defined in Appendix A. Since $\text{Cert}[A|C] \leq \text{Poss}[A|C]$ for all A and C , utilizing the certainty measure for weighing prior perceptions, results in combination schemes that favor more the information and less existing knowledge.

CHAPTER 4: FUZZY ROUTE CHOICE MODELS

Route choice behavior in the presence of information, as described in chapter 2, has two main components: drivers' perceptions of attributes of the system, and the route choice mechanism. In this chapter we adapt the principles of shortest path and utility maximization for the solution of route choice problems with fuzzy perceptions.

4.1 Shortest Paths in Networks with Fuzzy Attributes

The major assumption underlying the route choice model we present in this section is that drivers, based on their perceptions of travel times on links of the network, follow the **shortest path** to their destination. Hence we develop algorithms for determining the shortest path in a network with fuzzy travel times (costs) associated with its links. Obviously, the assumption that drivers follow the shortest path is a very strong one, as it assumes that drivers are well informed (i.e. know the costs on all links), have unlimited information processing capabilities, and make optimal decisions. Still, the fuzzy shortest path is of interest both for comparison with more realistic decision processes, and from the point of view of information systems that provide shortest path recommendations. It could also be that the fuzzy shortest path is a more realistic decision process than traditional shortest path behavior, because it corresponds to estimating and comparing paths' lengths based on linguistic, symbolic, or possibilistic evaluations (rather than exact numerical measures).

In what follows we assume that all the fuzzy numbers we deal with have positive values.

4.1.1 Addition and Comparison of Fuzzy Numbers

Two operations are of importance in the shortest path determination: addition and comparison of fuzzy numbers.

Addition

Addition of two fuzzy numbers uses the extension principle as defined in Appendix A; the sum of the fuzzy numbers A and B is given by the fuzzy number C defined by the following membership function:

$$\mu_C(z) = \max_{z=x+y} \min(\mu_A(x), \mu_B(y)) \quad (4-1)$$

To illustrate how the above definition of the sum of two fuzzy numbers applies in practice, let us introduce the following notation: let $A_L(\alpha)$ and $A_R(\alpha)$ denote the smallest and the largest real numbers respectively that have membership degree α in the fuzzy number A, that is: $\mu(A_L(\alpha)) = \mu(A_R(\alpha)) = \alpha$ and $A_L(\alpha) \leq A_R(\alpha)$. For the case of a distribution that is vertically truncated, we define $A_L(\alpha)$ and $A_R(\alpha)$ to be the left and

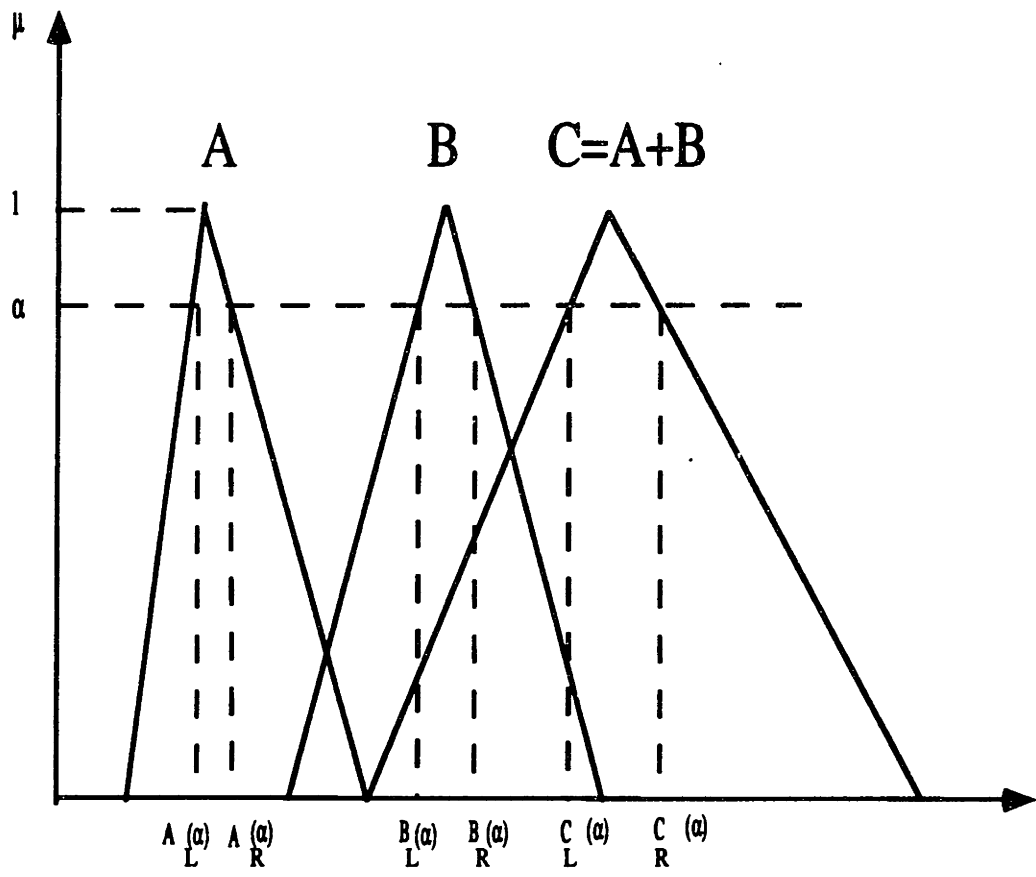


Figure 8: Addition of Two Fuzzy Numbers

right real values that correspond to the two points of intersection of the membership function with a horizontal line at level α . The fuzzy number C corresponding to the sum of A and B is then given by:

$$C_L(\alpha) = A_L(\alpha) + B_L(\alpha) \quad \text{for all } \alpha \in [0,1] \quad (4-2)$$

$$C_R(\alpha) = A_R(\alpha) + B_R(\alpha) \quad \text{for all } \alpha \in [0,1] \quad (4-3)$$

Figure 8 demonstrates graphically the addition of two triangular fuzzy numbers (TFN) A and B.

Comparison

Comparison of fuzzy numbers lacks a "golden rule". It is defined in the literature in many ways, depending on the particular application (see for example Bortolan and Degani, 1985, and McCahon and Lee, 1990, for reviews of available methods). Unfortunately, to the best of our knowledge, from the methods available in the literature on comparison of fuzzy numbers none is appropriate in the context of our problem. We proceed by giving a brief review of existing comparison methods, especially those useful in the context of the fuzzy shortest path problem. In section 4.1.2 we propose a new method which is more appropriate for determining shortest paths in a transportation network.

Freeling (1980) proposed the use of the **extended maximum** operation to compare two fuzzy numbers. The extended maximum of two fuzzy numbers A and B, $e\sim\max$, is defined (following the extension principle as defined in Appendix A) by the membership function:

$$\mu_{e\sim\max}(z) = \max_{z=\max(x,y)} \min(\mu_A(x), \mu_B(y)) \quad (4-4)$$

or equivalently:

$$e\sim\max_L(\alpha) = \max(A_L(\alpha), B_L(\alpha)) \quad \text{for all } \alpha \in [0,1] \quad (4-5)$$

$$e\sim\max_R(\alpha) = \max(A_R(\alpha), B_R(\alpha)) \quad \text{for all } \alpha \in [0,1] \quad (4-6)$$

The extended minimum of two fuzzy numbers A and B, $e\sim\min$, is defined in a similar fashion:

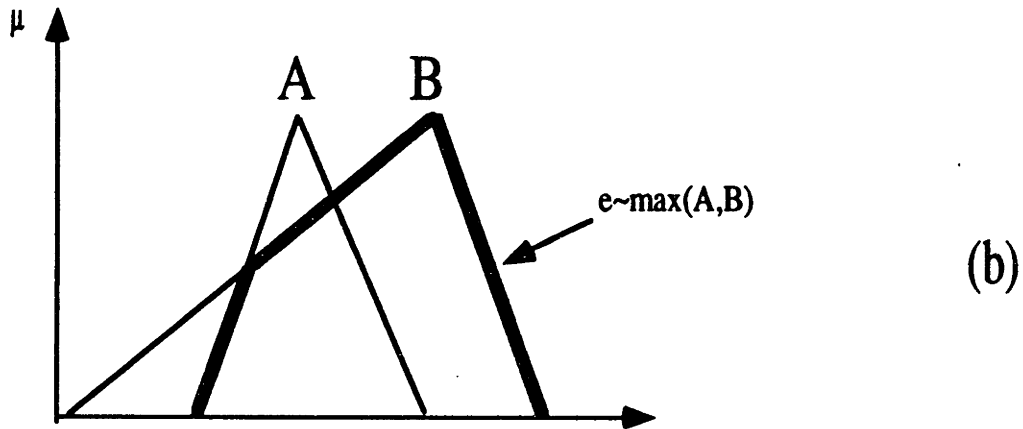
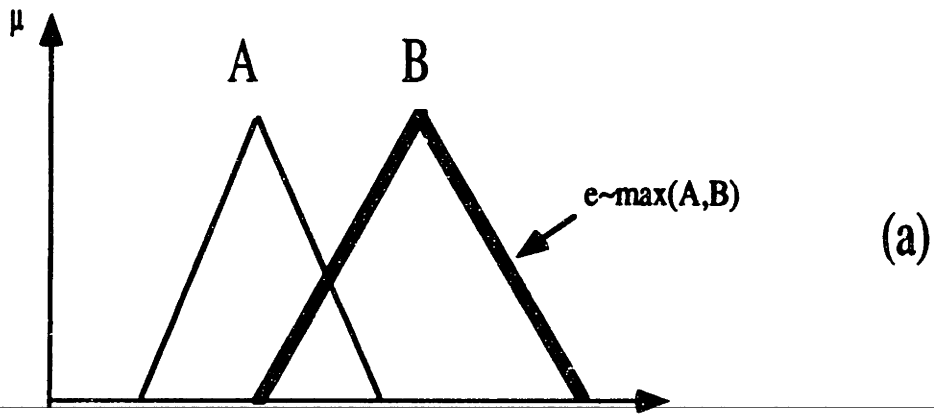


Figure 9: Extended Maximum Operation

$$\mu_{e\text{-min}}(z) = \max_{z=\min(x,y)} \min(\mu_A(x), \mu_B(y)) \quad (4-7)$$

or equivalently:

$$e\text{-min}_L(\alpha) = \min(A_L(\alpha), B_L(\alpha)) \quad \text{for all } \alpha \in [0,1] \quad (4-8)$$

$$e\text{-min}_R(\alpha) = \min(A_R(\alpha), B_R(\alpha)) \quad \text{for all } \alpha \in [0,1] \quad (4-9)$$

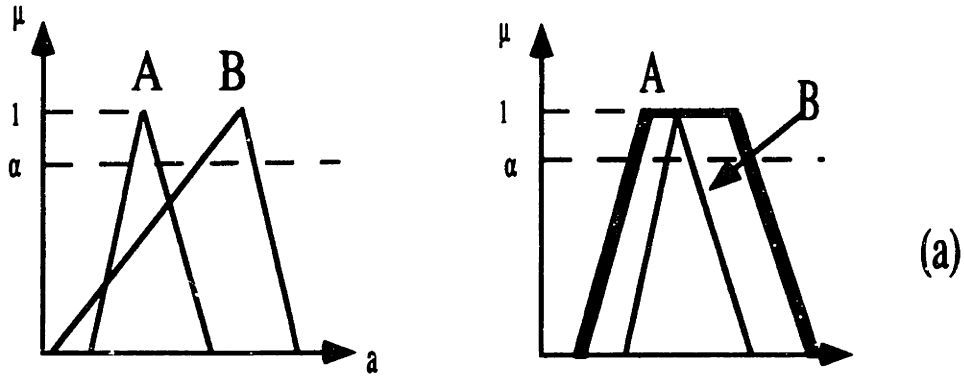
The fuzzy number which coincides with the extended maximum set is chosen as the bigger of the two (or similarly, the fuzzy number which coincides with the extended minimum set is chosen as the smaller). Figure 9 demonstrates this approach. In case 9a the set B coincides with the extended maximum set (and A with the extended minimum), and thus A is considered to be the smaller of the two. However in case 9b no set coincides with the extended maximum set and the method fails to provide a crisp answer. Complete overlap with the extended maximum set is a strong requirement, and typically does not occur.

Another method is based on the use of α -cuts. An α -cut of a fuzzy number is the set of values whose degree of membership is at least α , for $0 \leq \alpha \leq 1$ (as defined in Appendix A). For high enough values of α it is possible, in some cases, to get a set whose α -cut lies completely to the left of the other set and for that case this set would be the smaller of the two. Consider for example Figure 10a; on the left A is considered to be smaller than B for the given α level, whereas on the right, the order relation between A and B is unclear. Therefore the method is not generally applicable.

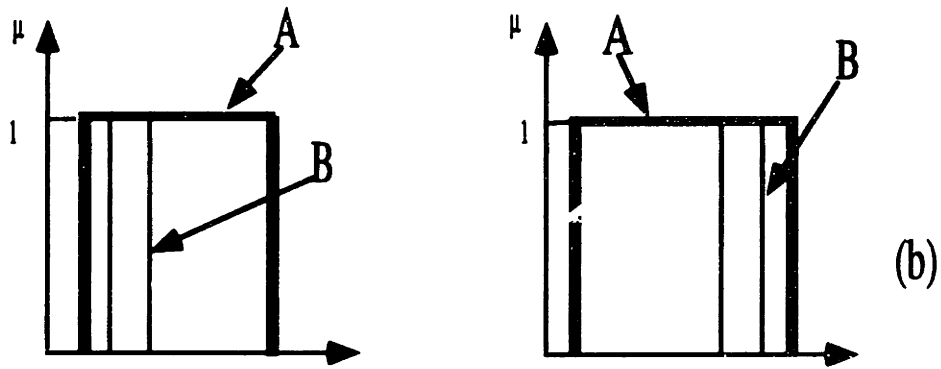
Dubois and Prade (1983) suggest a method of comparing fuzzy numbers which is based on four indices. For comparing two fuzzy numbers A and B, the indices are determined based on the following relations:

- The relative location of large values of A with respect to small values of B.
- The relative location of large values of A with respect to large values of B.
- The relative location of small values of A with respect to small values of B.
- The relative location of small values of A with respect to large values of B.

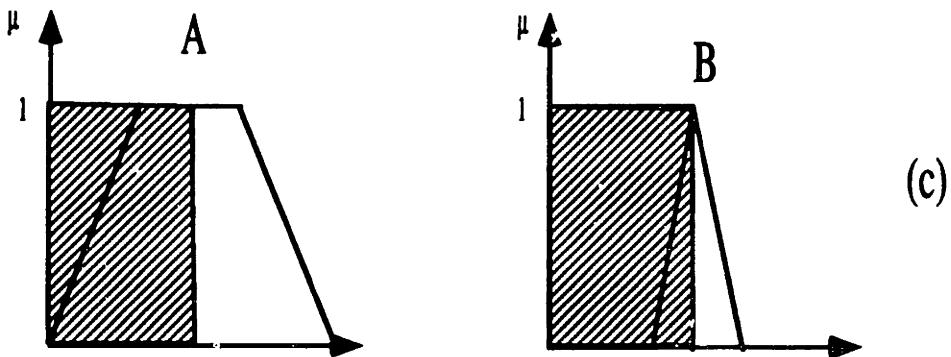
The four indices are used then either simultaneously or sequentially to get an idea how the two numbers relate to one another (total dominance, larger on the left, smaller on the right etc). However since the comparisons are based on relative locations of high and low values and not on magnitudes of differences, counter intuitive results might occur. For example, in Figure 10b the results of comparing A and B would be identical for both cases. This result is inconsistent with our interpretation of fuzzy numbers representing link travel times. According to this interpretation in the case on the left of Figure 10b we would be inclined to designate B as the smaller of the two, whereas in the case on the



The α -cut method



Dubois and Prade's method



Yager's method

Figure 10: Methods of Comparing Fuzzy Numbers

right, A would probably be the smaller.

Yager (1981) proposed a ranking function which maps each fuzzy set into a scalar in order to get an order relation among different fuzzy sets. His ranking function is based on the mean line of a fuzzy set, defined as the collection of the mid-points for each membership level. The ranking function is equal to the area between the vertical axis and the mean line. For our problem this method is inappropriate since it suffers from the common drawbacks of dealing with averages, particularly that it ignores the shape of the possibility distribution. The two fuzzy numbers in Figure 10c, for example, have the same mean line and thus will get the same ranking, however for our case they represent different levels of familiarity with the facility whose travel time they model.

In a recent paper, McCahon and Lee (1990) present the "proportion of the optimum" method for comparing fuzzy numbers. Their method is hierarchical and is based on two indices: MP(A) and mp(A), which measure the degree of agreement between the fuzzy number A and the extended maximum and minimum respectively, and are given by:

$$MP(A) = \frac{\int_x \min[\mu_{e^{-\max(A,B)}}(x), \mu_A(x)]dx}{\int_x \mu_A(x)dx} \quad (4-10)$$

and:

$$mp(A) = \frac{\int_x \min[\mu_{e^{-\min(A,B)}}(x), \mu_A(x)]dx}{\int_x \mu_A(x)dx} \quad (4-11)$$

They compare their method with 8 other comparison methods using several criteria (robustness, accuracy, flexibility, and ease of use), and conclude that it satisfies all the criteria. However, their method is evaluated based on 5 examples, and no global properties are proved. As it turns out, the proportion of optimum method does not always satisfy the transitivity property. Consider the following example: A, B, and C are trapezoidal fuzzy numbers each defined by four parameters:

$$A = (14,16,16,30)$$

$$B = (10,18,22,28)$$

$$C = (1,26,26,26)$$

The suggested comparison results in: $A < B$, $B < C$, but $A > C$, which obviously violates the transitivity property.

Chen (1985), suggested a very appealing and intuitive approach for ranking fuzzy numbers using their intersections with their maximizing set M, and their minimizing set G. Let A_1, \dots, A_n be n fuzzy numbers to be ranked. The membership functions of M and G are given by:

$$\mu_M(x) = \begin{cases} [(x-x_{\min})/(x_{\max}-x_{\min})]^k & x_{\min} \leq x \leq x_{\max} \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_G(x) = \begin{cases} [(x-x_{\max})/(x_{\min}-x_{\max})]^k & x_{\min} \leq x \leq x_{\max} \\ 0 & \text{otherwise} \end{cases}$$

where $x_{\min} = \min\{x \mid \mu_{A_i}(x) > 0, i=1, \dots, n\}$ and $x_{\max} = \max\{x \mid \mu_{A_i}(x) > 0, i=1, \dots, n\}$ for all the sets A_i that participate in the comparison.

The author defines the right index of each set A_i , $U_M(A_i)$, by:

$$U_M(A_i) = \max_x \min(\mu_M(x), \mu_{A_i}(x)) \quad \text{for } i=1,2,\dots,n \quad (4-12)$$

and the left index of each set A_i , $U_G(A_i)$ is defined by:

$$U_G(A_i) = \max_x \min(\mu_G(x), \mu_{A_i}(x)) \quad \text{for } i=1,2,\dots,n \quad (4-13)$$

The total index of each fuzzy number A_i is then given by:

$$U_T(A_i) = \frac{U_M(A_i) + (1-U_G(A_i))}{2} \quad \text{for } i=1,2,\dots,n \quad (4-14)$$

and it serves to order the fuzzy numbers. However, the definition of the minimizing and maximizing functions, depends on the specific numbers to be ranked (in particular on x_{\min} and x_{\max}), and thus inconsistencies and counter intuitive cases could occur. Consider, for example, the three trapezoidal fuzzy numbers given by:

$$\begin{aligned} A &= (12,12,12,24) \\ B &= (7,14,16,19) \\ C &= (2,17,17,17) \end{aligned}$$

Table 6 shows results of comparing different combinations of the fuzzy numbers A, B, and C.

fuzzy #	x_{\min}	x_{\max}	$U_r(A)$	$U_r(B)$	$U_r(C)$	result
A,B	7	24	.440	.446		A<B
B,C	2	19		.675	.676	B<C
A,C	2	24	.550		.544	A>C
A,B,C	2	24	.550	.547	.544	A>B>C

Table 6: Comparing Three Fuzzy Numbers - Chen's Method

It can be easily verified that the comparison is not transitive, and the result of ranking the three numbers contradicts the pairwise comparisons. Thus it can not be applied in a consistent way.

The above methods of comparing fuzzy numbers are often employed in algorithms suggested in the literature for determining the shortest path in a network with fuzzy costs associated with its links. However, given their inability to satisfy basic requirements, may result in counter-intuitive results.

Dubois and Prade, (1978), adapt Ford's and Floyd's algorithm for the solution of the shortest path problem with fuzzy costs. They suggest using the sum operation (as defined in section 4.1.1) for addition of costs on links and the extended minimum for performing the minimum operation required by the algorithm. That way they identify a fuzzy number which represents the optimal path length. However this length usually does not correspond to any path (because of the extended minimum definition). To identify the shortest path they define a path criticality index to be the height of the intersection of the fuzzy number representing the path length with the fuzzy number representing the length of the optimal path length. The path with the highest criticality index is identified as the shortest. However, as Chanas (1987) pointed out, the algorithm is equivalent to solving a deterministic shortest path with arc lengths equal to the mode of the fuzzy number (the mode is the value which has the highest degree of membership in the set).

To determine the fuzzy shortest path, Chanas and Kamburowski (1985) first enumerate all paths and evaluate their lengths, then they proceed as follows. For two fuzzy numbers A and B they define a preference relation $\mu^s(A,B) = \max\{\mu(A,B) - \mu(B,A), 0\}$, interpreted as the degree to which A is strictly less than B. $\mu(A,B)$ is a measure of the degree to which the relation $A \leq B$ is true. The solution for the shortest path problem then depends on the exact definition of the relation $\mu(A,B)$. Chanas (1987) suggests two definitions of $\mu(A,B)$. One corresponds to solving a deterministic shortest path with arc lengths equal to the mode of the fuzzy numbers (representing link costs). The other corresponds to solving the shortest path problem with arc lengths equal to the expectation of the fuzzy numbers given by:

$$E(A) = \int_0^1 \frac{A_L(\alpha) + A_R(\alpha)}{2} d\alpha \quad (4-15)$$

where $A_L(\alpha)$ and $A_R(\alpha)$ are defined as before.

The major weakness of this algorithm is the averaging effect of the mode and expectation of a fuzzy number which does not take into account the shape of the distribution which is critical in our analysis.

In a recent paper Delgado et al., (1990), deal with optimization problems on fuzzy graphs. They define a fuzzy graph as a graph whose nodes and arcs are fuzzy sets. An α -cut of a fuzzy graph is defined as the sub graph which contains all nodes and arcs whose degree of membership in the graph is at least α . Since the node and arc sets are finite, there exists a finite number of different α -cuts. For each distinct α -cut level they solve the shortest path problem, and the set of solutions for all the distinct α -levels constitutes a **fuzzy solution** to the problem. The shortest path problem for each α -level is solved for two special cases: the case in which time is expressed linguistically, and the case in which times are not normalized fuzzy numbers. Of particular interest is the determination of the shortest path when times on links are expressed linguistically. The authors suggest a fuzzy relation matrix which is used to obtain path lengths from link lengths. However, the given fuzzy relation matrix is order-dependent and thus the aggregation procedure is inconsistent (the length of a path consisting for example of three links with labels "low" "null" and "highest" is "moderately low" if traversed from left to right, and "very low" when it is traversed from right to left). When times on links are fuzzy numbers, instead of labels, Delgado et al. use classical shortest path algorithms to find the minimum path. The method they propose for comparison of fuzzy numbers is hierarchical in the sense that it first determines the shorter fuzzy set based on the mode of the membership function. If this fails, the spread of the membership function is taken into account.

4.1.2 A New Consistent Comparison Method

The various methods that were described in the previous section often fail to provide reasonable results for all possible intersections among fuzzy numbers. Partially the difficulty in comparing fuzzy numbers emanates from the different interpretations associated with them.

A comparison method should satisfy the following requirements in order to be consistently useful for any application. This is true in particular for the problem addressed in this thesis.

- A comparison method should be sensitive to the specific range and shape of the membership functions.

- A comparison method should satisfy transitivity, that is if $A < B$ and $B < C$ then $A < C$.
- The order relation between A and B that result from the comparison should not depend on the existence of another fuzzy number, C, in the comparison.

None of the comparison methods described in the previous section satisfy all three requirements. We propose a new method for ranking fuzzy numbers which takes into consideration relative location, magnitude of differences and shape of the fuzzy numbers involved. The method provides consistent results and satisfies all the three requirements stated above. Consequently, the suggested comparison method can be used in algorithms for the deterministic shortest path problem to find the minimum path length in a network with fuzzy costs.

To facilitate the demonstration we present the approach for comparing two fuzzy numbers. In comparing two fuzzy numbers A and B, we consider distances between them and their extended minimum (denoted by $e\sim\min$) and extended maximum ($e\sim\max$).

Extending Dubois and Prade's (1983) ideas about relations between A and B, in order to claim that A is smaller than B we expect to have:

- large distance between A and the extended maximum and large distance between B and the extended minimum.
- small distance between A and the extended minimum and small distance between B and the extended maximum.

Kaufmann and Gupta (1985) define the distance between two fuzzy numbers A and B, $d(A,B)$, by:

$$d(A,B) = d_L(A,B) + d_R(A,B) \quad (4-16)$$

where $d_L(A,B)$ is the distance from the left given by:

$$d_L(A,B) = \int_0^1 |A_L(\alpha) - B_L(\alpha)| d\alpha \quad (4-17)$$

and $d_R(A,B)$ is distance from the right given by:

$$d_R(A,B) = \int_0^1 |A_R(\alpha) - B_R(\alpha)| d\alpha \quad (4-18)$$

The shaded areas in Figures 11a and 11b illustrate the distance from the left and from the right respectively for the two sets A and B.

In order to compare two fuzzy numbers A and B we define indices I_A and I_B such that:

$$I_A = w_l \cdot d_L(A, e\sim\min) - w_h \cdot d_R(A, e\sim\max) \quad (4-19)$$

$$I_B = w_l \cdot d_L(B, e\sim\min) - w_h \cdot d_R(B, e\sim\max) \quad (4-20)$$

where w_l and w_h are weights.

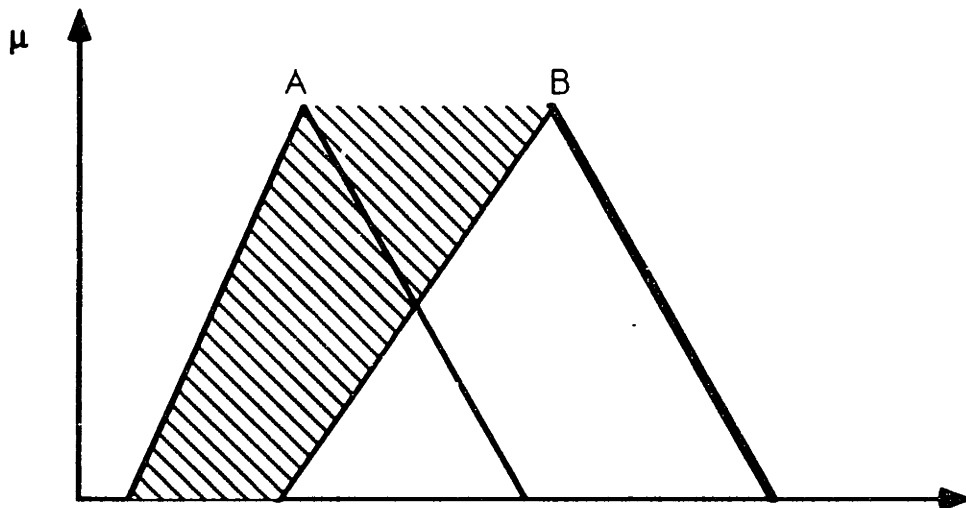
Note that $d_L(A, e\sim\min)$ is a measure of the degree to which A is bigger than the extended minimum. Thus, for A to be less than B we require that this measure is as small as possible. On the other hand, $d_R(A, e\sim\max)$ is a measure of the degree to which A is less than the extended maximum and it should be as large as possible (to claim that A is less than B). Therefore we use these indices to rank the fuzzy numbers and we define:

$$A < B \leftrightarrow I_A < I_B \quad (4-21)$$

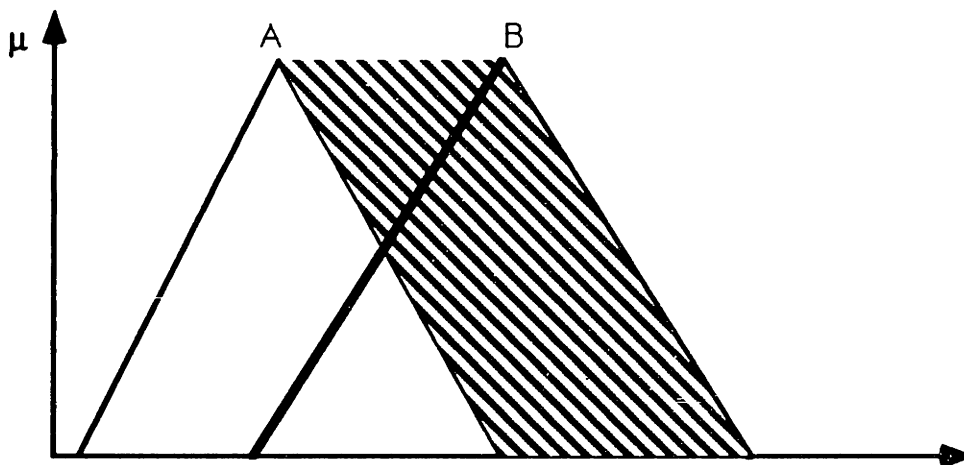
We use the weights w_l and w_h (associated with low and high values) to distinguish between the importance that users associate with possible low and high values respectively. That way we can, for example, model users who are risk averse (by assigning larger values to w_h), or risk prone. Naturally, w_l will accompany distance to the left whereas w_h would be associated with distance to the right.

Using the definitions of distance to the left and right we obtain:

$$\begin{aligned} d_L(A, e\sim\min) &= \int_0^1 |(A_L(\alpha) - e\sim\min_L(\alpha))| d\alpha = \\ &= \int_0^1 (A_L(\alpha) - \min(A_L(\alpha), B_L(\alpha))) d\alpha = \int_0^1 \max(0, A_L(\alpha) - B_L(\alpha)) d\alpha \end{aligned} \quad (4-22)$$



(a) Left Distance



(b) Right Distance

Figure 11: Distances Between Two Sets

and similarly:

$$d_R(A, e\sim\max) = \int_0^1 |(e\sim\max_R(\alpha) - A_R(\alpha))| d\alpha = \int_0^1 \max(0, B_R(\alpha) - A_R(\alpha)) d\alpha \quad (4-23)$$

Therefore the condition $I_A < I_B$ can be expressed as:

$$\begin{aligned} w_l \int_0^1 [\max(0, A_L(\alpha) - B_L(\alpha)) - \max(0, B_L(\alpha) - A_L(\alpha))] d\alpha < & \quad (4-24) \\ < w_h \int_0^1 [\max(0, B_R(\alpha) - A_R(\alpha)) - \max(0, A_R(\alpha) - B_R(\alpha))] d\alpha \end{aligned}$$

But $\max(0, x) - \max(0, -x) = x$, therefore:

$$A < B \quad \Leftrightarrow \quad w_l \int_0^1 [A_L(\alpha) - B_L(\alpha)] d\alpha < w_h \int_0^1 [B_R(\alpha) - A_R(\alpha)] d\alpha \quad (4-25)$$

The last inequality has an intuitive interpretation: A is less than B if the distance to the left between A and B is big when A_L is located to the left of B_L , and small when A_L is located to the right of B_L . Similarly A is less than B when the distance to the right between A and B is big when B_R is located to the right of A_R , and small when A_R is located to the right B_R .

We demonstrate the comparison method on three examples in Figure 12. Case (a) is the trivial case for which most comparison methods would result in A being less than B, and indeed our weighted-distance comparison method indicates that:

$$A < B \quad \Leftrightarrow \quad -w_h \cdot [\text{area(II)} + \text{area(III)}] < w_l \cdot [\text{area(I)} + \text{area(III)}] \quad (4-26)$$

where the Roman letters correspond to the areas marked accordingly in Figure 12a. Obviously, the condition is always true and thus A is indeed always less than B.

In Figure 12b the set B is contained in A. In this case the decision on which is the smallest depends on areas I and II and the weights w_l and w_h . Applying equation (4-25) we get that:

$$A < B \quad \Leftrightarrow \quad w_h \cdot \text{area(II)} < w_l \cdot \text{area(I)} \quad (4-27)$$

which means that in order for A to be less than B, the area in A that corresponds to lower A values (area I) has to be greater than the area in A which corresponds to higher values

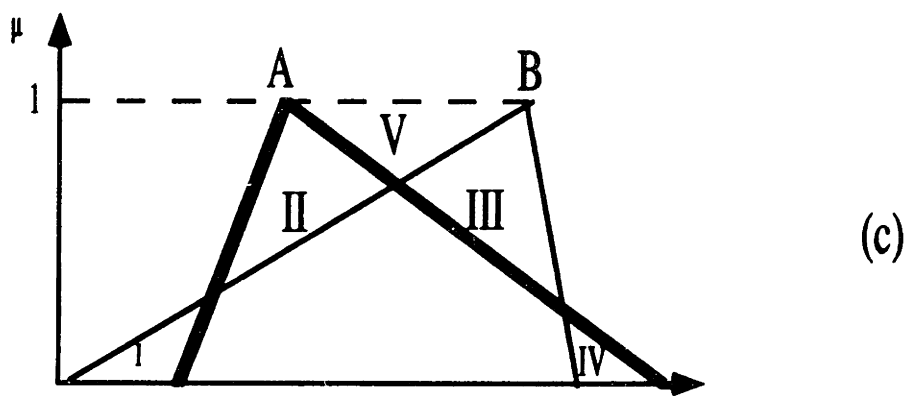
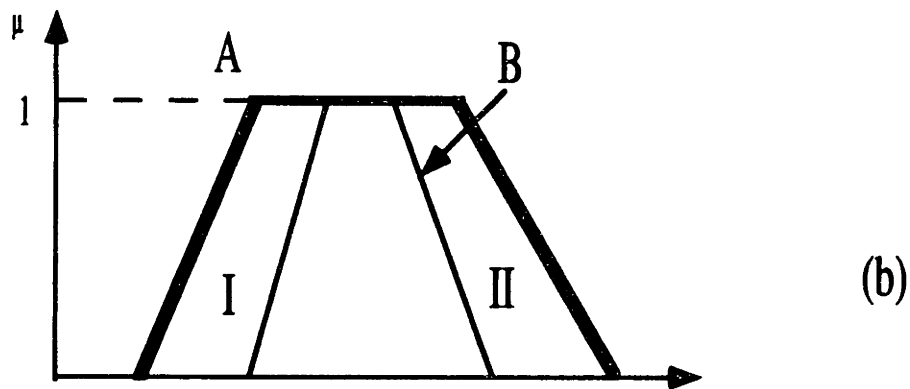
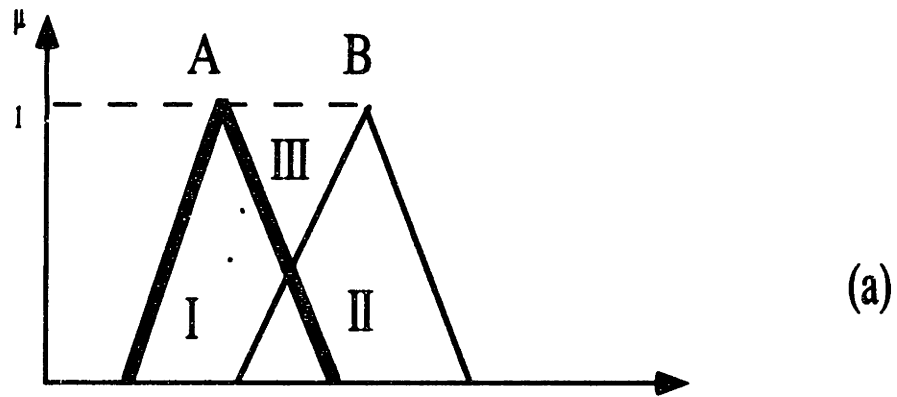


Figure 12: Comparison of Two Fuzzy Numbers

(area II), multiplied by the appropriate weights.

For the two sets in Figure 12c, we get that:

$$A < B \quad \leftrightarrow \quad w_l \text{area(I)} + w_h \text{area(IV)} < w_l \text{area(II)} + w_h \text{area(III)} + (w_l + w_h) \cdot \text{(V)} \quad (4-28)$$

An intuitive explanation for the above inequality follows from the fact that areas II and III are "good" areas for claiming that A is less than B since they correspond to lower possible A values and higher possible B values; similarly areas I and IV are "bad" areas since they correspond to small B values and high A values; finally area V, although it does not under any possibility distribution, it is a measure of the potential smaller A values on one hand and potential larger B values on the other hand and thus is weighted by both w_l and w_h .

So far we have shown that the suggested method fulfills the first requirement, namely, it takes into account the exact shape and range of the sets being compared. Next we prove that transitivity, the second requirement, holds.

$$A < B \quad \leftrightarrow \quad w_l \int_0^1 [A_L(\alpha) - B_L(\alpha)] d\alpha < w_h \int_0^1 [B_R(\alpha) - A_R(\alpha)] d\alpha \quad (4-29)$$

$$B < C \quad \leftrightarrow \quad w_l \int_0^1 [B_L(\alpha) - C_L(\alpha)] d\alpha < w_h \int_0^1 [C_R(\alpha) - B_R(\alpha)] d\alpha \quad (4-30)$$

Summing (4-29) and (4-30) and simplifying we obtain that:

$$w_l \int_0^1 [A_L(\alpha) - C_L(\alpha)] d\alpha < w_h \int_0^1 [C_R(\alpha) - A_R(\alpha)] d\alpha \quad (4-31)$$

which implies $A < C$, thus the proposed comparison method satisfies the transitivity property.

The suggested weighted-distance comparison method satisfies the third requirement as well, that is, it gives consistent results, which are independent of the specific number of fuzzy numbers involved in the comparison. As a result, it can be used to rank n fuzzy numbers (in a way consistent with pairwise comparisons).

Let A^1, \dots, A^n be n fuzzy numbers. We define the extended minimum and maximum functions of them, $\underline{e\sim\min}$ and $\underline{e\sim\max}$, by:

$$\begin{aligned} \underline{e\sim\min}_L(\alpha) &= \{\min A_L^i(\alpha) \mid 1 \leq i \leq n\} ; & \underline{e\sim\min}_R(\alpha) &= \{\min A_R^i(\alpha) \mid 1 \leq i \leq n\} \\ \underline{e\sim\max}_L(\alpha) &= \{\max A_L^i(\alpha) \mid 1 \leq i \leq n\} ; & \underline{e\sim\max}_R(\alpha) &= \{\max A_R^i(\alpha) \mid 1 \leq i \leq n\} \end{aligned}$$

for every α level, $0 \leq \alpha \leq 1$.

For every fuzzy number A^i , $1 \leq i \leq n$, we define the index:

$$I_{A^i}^n = w_l \cdot d_L(A^i, \underline{e\sim\min}) - w_h \cdot d_R(A^i, \underline{e\sim\max}) \quad (4-32)$$

We claim that $A^1 < \dots < A^i < \dots < A^n$ iff $I_{A^1}^n < \dots < I_{A^i}^n < \dots < I_{A^n}^n$. We will show that if $A^i < A^j$, based on $I_{A^i}^n$ and $I_{A^j}^n$, then it also holds that $A^i < A^j$ using I_{A^i} and I_{A^j} (i.e. the indices based on the extended min and max of A^i and A^j).

The index $I_{A^i}^n$ is given by:

$$\begin{aligned} I_{A^i}^n &= w_l \int_0^1 [A_L^i(\alpha) - \underline{e\sim\min}_L(\alpha)] d\alpha - w_h \int_0^1 [\underline{e\sim\max}_R(\alpha) - A_R^i(\alpha)] d\alpha = \quad (4-33) \\ &= w_l \int_0^1 [A_L^i(\alpha) - \min_k(A_L^k(\alpha))] d\alpha - w_h \int_0^1 [\max_k(A_R^k(\alpha)) - A_R^i(\alpha)] d\alpha \end{aligned}$$

$A^i < A^j$ iff $I_{A^i}^n < I_{A^j}^n$, that is:

$$\begin{aligned} &w_l \int_0^1 [A_L^i(\alpha) - \min_k(A_L^k(\alpha))] d\alpha - w_h \int_0^1 [\max_k(A_R^k(\alpha)) - A_R^i(\alpha)] d\alpha < \quad (4-34) \\ &< w_l \int_0^1 [A_L^j(\alpha) - \min_k(A_L^k(\alpha))] d\alpha - w_h \int_0^1 [\max_k(A_R^k(\alpha)) - A_R^j(\alpha)] d\alpha \end{aligned}$$

which is equivalent to:

$$w_l \int_0^1 [A_L^i(\alpha) - A_L^j(\alpha)] d\alpha < w_h \int_0^1 [A_R^j(\alpha) - A_R^i(\alpha)] d\alpha \quad (4-35)$$

Equation (4-35) however, is equivalent to equation (4-25), and thus it implies that $A^i < A^j$ when using I_{A^i} and I_{A^j} (defined for two fuzzy numbers only).

We have shown that the new comparison method is transitive and consistent, thus most shortest path algorithms (for example the algorithms described in Ahuja et al.,

1988), can be applied, when appropriately modified for additions and comparisons, for the solution of the shortest path problem in networks with fuzzy costs. These algorithms correctly identify the shortest path (i.e. the path that has the lowest total fuzzy cost among all paths).

4.2 Fuzzy Utility Models

As discussed in section 1.3, discrete choice models based on the utility maximization principle are the most commonly used models for solving the route choice problem. They are based on the assumption that the (latent) utility of each alternative in the choice set can be expressed as a linear function of the main attributes that influence the choice. In this section we extend the classical utility model to incorporate fuzzy attributes. We begin by reviewing methods for the fuzzy linear regression problem since some of these ideas are used for the development of the fuzzy utility models presented in this thesis.

4.2.1 Fuzzy Linear Regression - Literature Review

Surprisingly, there is not much work which deals with fuzzy linear regression. Among the papers found in the literature, there seems to be unanimity on using the fuzzy linear regression method suggested by Tanaka et al. (1982). This approach deals with a linear model with crisp inputs and fuzzy coefficients. All existing models assume symmetric triangular fuzzy numbers (T_fFN). The membership function of a T_fFN, A, is given by:

$$\mu_A(x) = \begin{cases} 1 - |\gamma - x|/c & \gamma - c \leq x \leq \gamma + c \\ 0 & \text{otherwise} \end{cases}$$

that is, a fuzzy set centered at γ with left and right spreads equal to c , which models the concept of "approximately γ ".

Fuzzy Linear Regression - Tanaka et al.

Tanaka et al. (1982), suggested the following regression model:

$$Y_i = A_1 x_{i1} + \dots + A_n x_{in} \quad (4-36)$$

where: A_p 's are T_fFN's represented by two parameters: (γ_p, c_p) $p=1, \dots, n$
 Y_i 's are T_fFN's represented by two parameters: (y_i, e_i) $i=1, \dots, K$, and
 x_{ip} are scalars $i=1, \dots, K$, $p=1, \dots, n$.

In traditional regression models, the error term: $\epsilon_i = y_i - y_i^*$, defined as the deviation between the observed value y_i , and the estimated value y_i^* , is regarded as random variable

with zero mean, and the objective function is to minimize a function of the error term of all K observations. In fuzzy linear regression, on the other hand, deviations of the estimated value from observed value are attributed to the fuzziness of the system. Moreover, a fuzzy error term can be estimated by the constant of the regression equation (i.e. the fuzzy coefficient corresponding to the case in which $x_{i0}=1$ for all i).

Thus, estimation of the fuzzy coefficients A_p is based on the minimization of their vagueness (where vagueness is defined by c_p , the spread of the T_rFN A_p , for $p=1, \dots, n$), subject to the constraint that a certain degree of fit between observed and predicted output values is achieved. Hence, the objective function minimizes the sum of the individual spreads:

$$\min \sum_{p=1}^n c_p \quad (4-37)$$

In order to formulate the constraints, a measure of the degree of fit between the observed fuzzy output, $Y_i = (y_i, e_i)$, and the estimated output, Y_i^* , is required, where:

$$Y_i^* = A_1^* x_{i1} + \dots + A_n^* x_{in} \quad (4-38)$$

Tanaka suggested measuring this degree of fit by a containment relation between the estimated and observed outputs. The degree of fit of the estimated output Y_i^* with the observed output Y_i , is measured by an index h_i , which maximizes h subject to the constraint: $Y_i^h \subset Y_i^{*h}$, where:

$$Y_i^h = \{ y \mid \mu_{Y_i}(y) \geq h \} \quad (4-39)$$

$$Y_i^{*h} = \{ y \mid \mu_{Y_i^*}(y) \geq h \} \quad (4-40)$$

An example of the determination of h_i is illustrated in Figure 13.

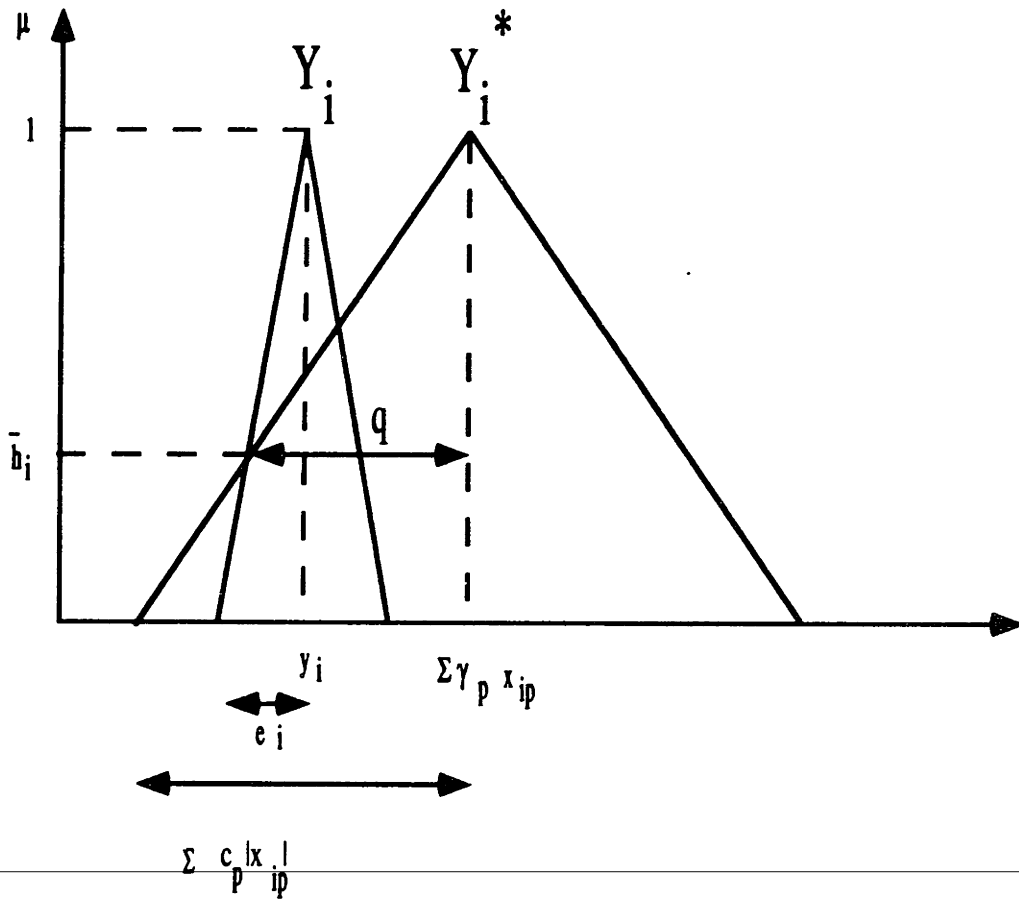


Figure 13: Degree of Fit Between Y_i^* and Y_i

From Figure 13 we can see that:

$$\frac{1}{\sum_{p=1}^n c_p |x_{ip}|} = \frac{1-\bar{h}_i}{q} \quad (4-41)$$

where:

$$q = |y_i - \sum_{p=1}^n \gamma_p x_{ip}| + e_i(1-\bar{h}_i) \quad (4-42)$$

hence:

$$\bar{h}_i = 1 - \frac{|y_i - \sum_{p=1}^n \gamma_p x_{ip}|}{\sum_{p=1}^n c_p |x_{ip}| - e_i} \quad (4-43)$$

Consequently, the goodness of fit constraints are expressed as:

$$\bar{h}_i \geq H \quad \text{for } i=1, \dots, K \quad (4-44)$$

where H is the minimum accepted goodness of fit.

Using equation (4-43), the constraints (4-44) now become:

$$\sum_{p=1}^n \gamma_p x_{ip} + (1-H) \sum_{p=1}^n c_p |x_{ip}| \geq y_i + (1-H)e_i \quad (4-45)$$

and:

$$-\sum_{p=1}^n \gamma_p x_{ip} + (1-H) \sum_{p=1}^n c_p |x_{ip}| \geq -y_i + (1-H)e_i \quad (4-46)$$

So the problem can be formulated as a linear programming problem (LP) with $2n$ variables (the vector γ and the vector c), and $2K$ constraints (for K observed choices).

Maximum Likelihood Estimation - Diamond's Approach

Diamond (1988), modified Tanaka et al.'s approach to allow for stochastic variations, and suggested two models that include a random fuzzy error term:

$$(i) \quad Y_i = B_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in} + E_i$$

$$(ii) \quad Y_i = B_0 + B_1 x_{i1} + \dots + B_n x_{in} + E_i$$

where capital letters correspond to fuzzy sets. In the first model, the output, the inputs and the error term are fuzzy while the coefficients are crisp, and in the second model the

inputs are crisp, while the output, the coefficients and the error term are fuzzy.

As did Tanaka et al., Diamond assumes symmetric triangular fuzzy numbers (T_fFN). The assumption about the random term, E_i , is that it is also a T_fFN, defined by the two parameters: (γ_i, c_i) , where its endpoints: $\gamma_i - c_i$ and $\gamma_i + c_i$ are drawn from a uniform distribution.

The major criticism of the suggested approach, which in our opinion makes it inappropriate for various applications, is summarized below:

- Superficial ability to handle stochastic variation; the method eliminates the stochastic nature of the error term by assuming a uniform distribution that leads to a constant valued likelihood function. The uniformity assumption for the error term is problem specific. However, in most cases a non-truncated error term distribution seems more reasonable.
- Generalization to other error terms is very difficult.
- The estimated parameters are level 2 and 3 fuzzy numbers (which correspond to higher levels of uncertainty regarding the membership functions), whose interpretation is not clear.

Furthermore, extension of the model to solve the discrete choice problem is not at all straightforward.

After looking into the existing literature on fuzzy linear regression, it is clear why it has not gained more attention. All models make very limiting assumptions regarding the fuzzy sets involved: they all assume symmetric triangular fuzzy numbers.

Tanaka et al.'s approach is very appealing from a computational point of view; transforming an estimation problem into a linear programming problem is very attractive. However, an issue which is not addressed in their paper is the (pre-determined) value of H . There is a tradeoff between H and the fuzziness of the parameters. A possible approach to address the problem would be to treat H as a variable, and solve the problem as a multi-criteria LP: minimize the sum of the spreads and maximize H . The applicability of Tanaka et al.'s approach to the route choice problem is limited. The main attraction for the use of fuzzy sets for the problem arises from the idea of fuzzy perceptions, thus treating the inputs as crisp, and the coefficients as fuzzy is not useful.

Diamond's approach assumes a special error term distribution, adds non-intuitive levels of fuzziness, and can not be easily extended to discrete choice problems.

4.2.2 Utility Models with Fuzzy Parameters and Crisp Inputs

The concepts presented in the previous section can be applied for solving discrete choice problems in general, and route choice problems in particular. In this section we extend Tanaka's approach such that it can be used for discrete choice problems.

Given a choice set with m alternatives, the utility of alternative i is given by:

$$U_i = A_1 x_{i1} + \dots + A_n x_{in} \quad (4-47)$$

where: U_i is the fuzzy utility of alternative i ,
 A_p 's are fuzzy sets representing the importance of attribute p in the overall utility, and
 x_{ip} are scalars corresponding to the value of attribute p for alternative i .

In the fuzzy utility model, x_{ip} 's are given inputs, the fuzzy coefficients A_p 's are the parameters to be estimated, and U_i are fuzzy latent utilities (which, unlike the regression model, are not known or observed). For simplicity we omit the index corresponding to a specific individual.

Following the utility maximization principle, we assume that alternative j is chosen if $U_j \geq U_i$ for all the alternatives i in the choice set ($j \neq i$). The problem is then to estimate the fuzzy coefficients A_p such that the agreement between observed and predicted choices is maximized.

The underlying behavioral assumption of this model is that the inputs relating to the important attributes of the utility (e.g. travel time, travel distance) are crisp, but their contribution to the overall utility is modeled by fuzzy sets, and as a result the overall utility is also represented by a fuzzy set. Consequently, the uncertainty in the problem is attributed to the fact that the importance of the different attributes can not be modeled deterministically, or even probabilistically (using a random coefficient model), and thus is modeled by a fuzzy set. Individuals have the "correct" perceptions, however, their weights (and importance) in the decision process are not clear. Thus coefficients may be fuzzy when the utility is assumed to be fuzzy, and exact measurements (x_{ip}) are available, or when the nature of the contribution of specific factors to the overall utility is not completely clear.

We assume that the coefficients A_p are T_rFNs (as in Tanaka's approach), represented by two parameters: (γ_p, c_p) $p=1, \dots, n$. From equation (4-47) it follows that the utility U_i is also a T_rFN defined by the two parameters: (U_i^m, C_i) where U_i^m is the center point given by:

$$U_i^m = \sum_{p=1}^n \gamma_p x_{ip} \quad (4-48)$$

and C_i is the spread given by:

$$C_i = \sum_{p=1}^n c_p |x_{ip}| \quad (4-49)$$

If alternative j was chosen, then we want U_j to be bigger than U_i for all $i \neq j$. The measure of fit used by Tanaka et al. is not appropriate for this case since it tries to maximize the overlap between Y_i^* and Y_i , whereas for discrete choice we are interested in having the two sets, U_i and U_j , as far apart (in the right order) as possible, representing the fact that when one alternative is chosen over another, the utility of this alternative is better than the utility of the other.

The discrete choice problem can then be formulated as follows:

$$\min \sum_{p=1}^p c_p \quad (4-50)$$

$$\text{such that:} \quad U_{oc(k)}^k \geq U_i^k \quad \text{for all } i \neq oc(k) \quad k=1, \dots, K \quad (4-51)$$

where $oc(k)$ is the alternative chosen at the k^{th} trip, and U_i^k is the utility of alternative i for trip k .

For comparing the utilities, which are represented by fuzzy numbers, we suggest using the comparison method proposed in section 4.1.2, which other than having desirable properties such as transitivity and applicability to multiple-sets comparisons, results in linear constraints for the discrete choice problem. The following claim translates the set of constraints given by equation (4-51) to linear constraints (assuming that the weights w_l and w_h are known).

Claim 1:

Given U_i and U_j T_fNs defined by: $U_i=(U_i^m, C_i)$, $U_j=(U_j^m, C_j)$, $U_i < U_j$ iff:

$$w_l[(U_j^m - U_i^m) + (U_j^m - C_j) - (U_i^m - C_i)] > w_h[(U_i^m - U_j^m) + (U_i^m + C_i) - (U_j^m + C_j)] \quad (4-52)$$

where w_l and w_h are weights associated with low and high values respectively.

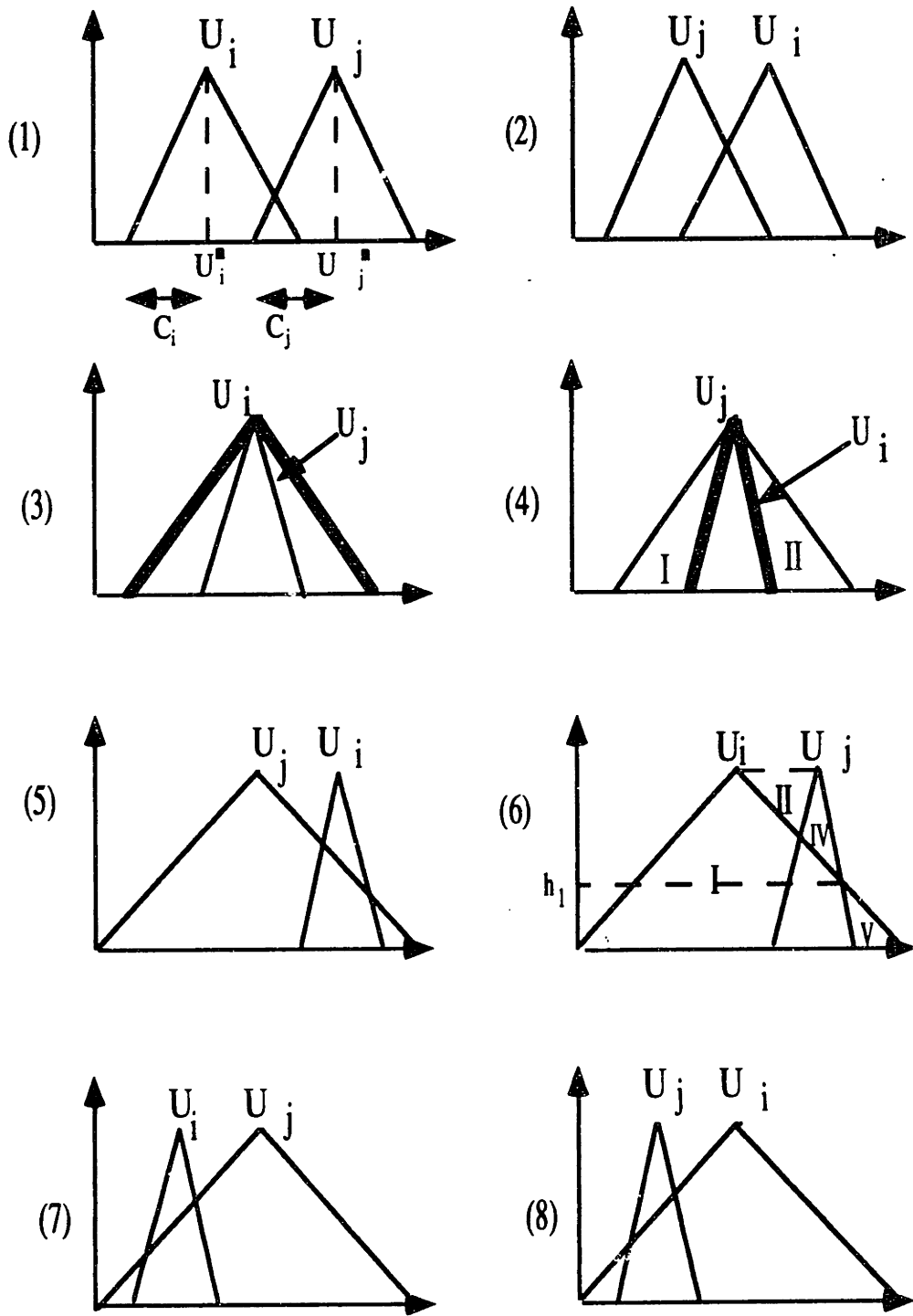


Figure 14: Possible Intersections of Two TsFN's

Proof:

There are 8 possible intersections between two T_jFN's as it is shown in Figure 14. Cases 1 and 2 are trivial (U_i is smaller in case 1 and U_j is smaller in case 2). We will prove the claim for cases 4 and 6. The other cases can be proved in a similar way.

Case 4:

$$d_L(U_i, e\sim\min) = d_R(U_i, e\sim\max) = 0 \quad \rightarrow \quad I_{U_j} = 0 \quad (4-53)$$

$$d_L(U_i, e\sim\min) = \frac{1}{2}[(U_i^m - C_i) - (U_j^m - C_j)] = \frac{1}{2}(C_j - C_i) \quad (4-54)$$

$$d_R(U_i, e\sim\max) = \frac{1}{2}[(U_j^m + C_j) - (U_i^m + C_i)] = \frac{1}{2}(C_j - C_i) \quad (4-55)$$

Graphically, as shown in Figure 14, $d_L(U_i, e\sim\min)$ is equal to area I, and $d_R(U_i, e\sim\max)$ to area II.

The condition $I_{U_i} < I_{U_j}$ is thus equivalent to: $w_i(C_j - C_i) < w_h(C_j - C_i)$ which is the equivalent to equation (4-52) (after substituting $U_j^m = U_i^m$). For this case (case #4), $C_j > C_i$ and thus:

$$U_i < U_j \quad \leftrightarrow \quad w_i < w_h \quad (4-56)$$

meaning that since areas I and II are equal, alternative j will be chosen if w_h is larger (i.e. the individual is risk averse).

Case 6:

$$d_L(U_i, e\sim\min) = 0 \quad (4-57)$$

$$d_R(U_i, e\sim\max) = \frac{1}{2}(U_j^m - U_i^m)(1 - h_1) \quad (4-58)$$

$$d_L(U_j, e\sim\min) = \frac{1}{2}[(U_j^m - U_i^m) + (U_j^m - C_j) - (U_i^m - C_i)] \quad (4-59)$$

$$d_R(U_j, e\sim\max) = \frac{1}{2}[(U_i^m + C_i) - (U_j^m + C_j)]h_1 \quad (4-60)$$

where:

$$h_1 = \frac{(U_1^m - U_j^m + C_1 - C_j)}{(C_1 - C_j)} \quad (4-61)$$

Graphically, as demonstrated in Figure 14, $d_R(U_i, e \sim \max)$ is equal to areas II and IV, $d_L(U_j, e \sim \min)$ to area I and II, and $d_R(U_j, e \sim \max)$ to area V.

Thus $U_i < U_j$ iff:

$$w_l[2U_j^m - 2U_i^m - C_j + C_l] > w_h \{ [U_i^m + C_1 - (U_j^m + C_j)]h_1 - (U_j^m - U_i^m)(1-h_1) \} \quad (4-62)$$

which is equivalent to equation (4-52). ■

Thus, for binary choice problems there would be one linear constraint for each observed choice. In general, for multinomial choice, with a given choice set of size m , there would be $(m-1)$ linear constraints for each choice, thus the total number of constraints will be $2K(m-1)$ where K is the number of observed choices.

4.2.3 Utility Models with Fuzzy Attributes and an Error Term

In this section we develop a discrete choice model based on fuzzy (or crisp) attribute values, crisp coefficients, and a stochastic error term. This model fits our problem better than the model presented in the previous section since it allows for fuzzy inputs (which correspond to perceptions of network attributes). Furthermore, it includes a random error term to account for the non-deterministic nature of the problem. In this way, both aspects of uncertainty that were introduced in section 3.2 are being captured: vagueness through allowing attribute perceptions to be fuzzy, and randomness through inclusion of a random error term.

We assume again that a choice set of size m is given. The utility of alternative i in the choice set is given by:

$$U_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in} + \epsilon_i \quad (4-63)$$

where X_{ip} 's are fuzzy sets, β_p 's are scalars to be estimated, and ϵ_i is an error term. For simplicity of notation, we omit the index corresponding to a specific individual. The utility, U_i , is best described as a hybrid number (see Kaufmann and Gupta, 1985), that is, a fuzzy number that shifts according to some probability density function. The systematic component of the utility, V_i , is fuzzy since perceptions of attributes (e.g. travel time) are fuzzy. The error term adds noise and serves to model inherent randomness plus other omissions as discussed in Ben-Akiva and Lerman (1985).

We assume that attribute perceptions are given by trapezoidal fuzzy numbers, T_rFN, (rather than symmetric triangular fuzzy numbers). A T_rFN is determined by four parameters: location parameter γ and three spread parameters: c_1 , c_2 and c_3 as illustrated in Figure 15. The systematic component of the utility of alternative i , V_i , is also a T_rFN.

For a binary choice between alternatives i and j the probability that a certain individual will choose alternative j is given by:

$$\text{Prob}(j) = \text{Prob}(U_i \leq U_j) = \text{Prob}(V_i + \epsilon_i \leq V_j + \epsilon_j) = \text{Prob}(\epsilon_i - \epsilon_j \leq V_j - V_i) = \text{Prob}(\epsilon \leq V) \quad (4-64)$$

where V is T_rFN and ϵ is a random variable with known probability density function.

The probability of choosing alternative j can be also written as:

$$\text{Prob}(j) = \text{Prob}(\epsilon \leq V) = \int \delta(\epsilon \leq V | \epsilon = \epsilon_k) f(\epsilon_k) d\epsilon_k \quad (4-65)$$

where:
$$\delta(\epsilon \leq V | \epsilon = \epsilon_k) = \begin{cases} 1 & \text{if } \epsilon_k < V \\ 0 & \text{otherwise} \end{cases}$$

and $\epsilon_k \in \mathfrak{R}$.

Claim 2:

Using the weighted-distance comparison method of equation (4-25), for a T_rFN V defined by (left) location parameter γ and three spread parameters: c_1 , c_2 and c_3 , a random variable ϵ , and a scalar ϵ_k , we have that:

$$\delta(\epsilon \leq V | \epsilon = \epsilon_k) = \begin{cases} 1 & \text{if } \epsilon_k < x \\ 0 & \text{otherwise} \end{cases}$$

where:

$$x = \gamma + \frac{1}{2}c_1 + \frac{1}{2}w_h(c_1 + 2c_2 + c_3) \quad (4-66)$$

Proof:

We use the comparison method proposed in section 4.1, to determine whether ϵ_k , a realization of ϵ , is less than V . This is a special case of comparing two fuzzy sets (since for this case one of the sets is deterministic). The comparison method is consistent in the sense that if $\epsilon_p \leq V$ for a specific ϵ_p value, then $\epsilon_q \leq V$ for all $\epsilon_q \leq \epsilon_p$. Thus, there exists a threshold value, x , $x \in \text{Support}(V)$, such that for all $\epsilon_p \leq x$, $\epsilon_p \leq V$, and for all $\epsilon_q \geq x$, $\epsilon_q \geq V$.

To find x we have to examine the three cases demonstrated in Figure 15.

Case (a):

Applying equation (4-25) results in:

$$\epsilon_k < V \quad \Leftrightarrow \quad w_1 \cdot \text{area(I)} < (w_1 + w_h) \cdot \text{area(III)} + w_h \cdot \text{area(II)} \quad (4-67)$$

where the roman letters correspond to the appropriate areas in Figure 15a as determined by the threshold value x . Using the fact that $w_1 + w_h = 1$, we get that:

$$\frac{1}{2}w_1(x-\gamma)h_1 \leq \frac{1}{2}(\gamma+c_1-x)(1-h_1) + w_h[\frac{1}{2}(1+h_1)(\gamma+c_1-x)+c_2+\frac{1}{2}c_3] \quad (4-68)$$

where:

$$h_1 = \frac{x-\gamma}{c_1} \quad (4-69)$$

For x , the threshold value, we want equation (4-67) to hold as equality, hence:

$$x = \gamma + \frac{1}{2}c_1 + \frac{1}{2}w_h(c_1+2c_2+c_3) \quad (4-70)$$

Case (b):

$$\epsilon_k < V \quad \Leftrightarrow \quad w_1 \cdot \text{area(I)} < w_h \cdot \text{area(II)} \quad (4-71)$$

where areas I and II are determined by x , as demonstrated in Figure 15b.

hence:

$$w_1(\frac{1}{2}c_1+x-\gamma') = w_h(\frac{1}{2}c_3+\gamma'+c_2-x) \quad (4-72)$$

and hence:

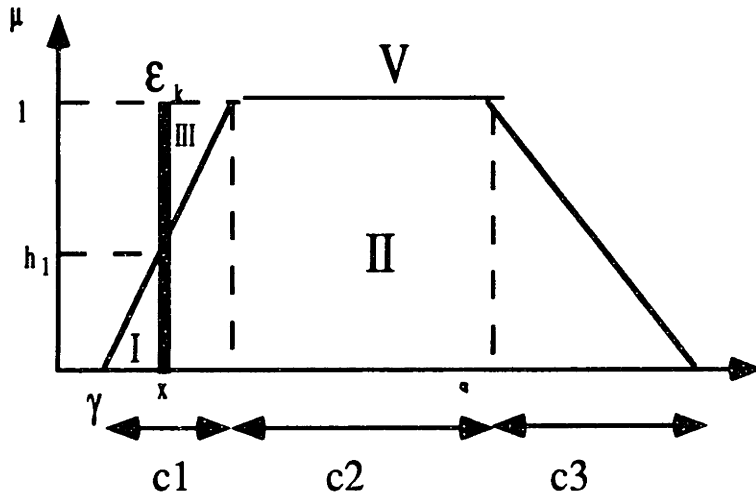
$$x = \gamma' + w_h(\frac{1}{2}c_3+c_2) - \frac{1}{2}w_1c_1 \quad (4-73)$$

Substituting $\gamma' = \gamma + c_1$, and $w_1 = 1 - w_h$, we get that equation (4-73) is equivalent to equation (4-66).

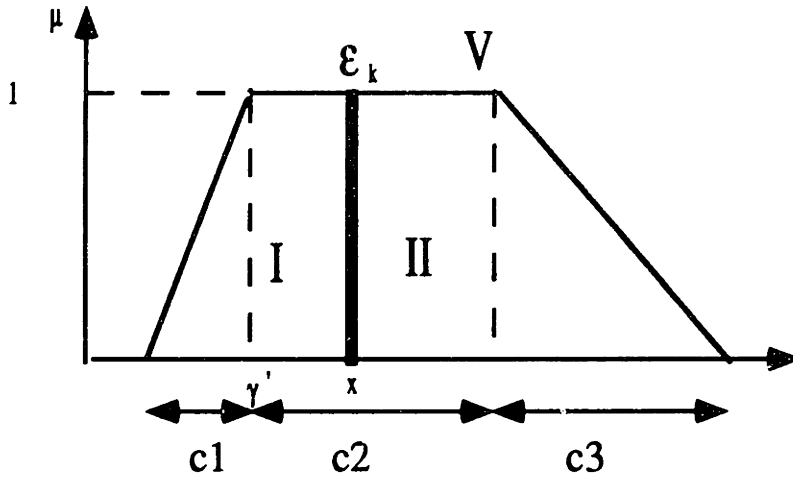
case (c):

$$\epsilon_k < V \quad \Leftrightarrow \quad w_h \cdot \text{area(II)} > (w_1 + w_h) \cdot \text{area(III)} + w_1 \cdot \text{area(I)} \quad (4-74)$$

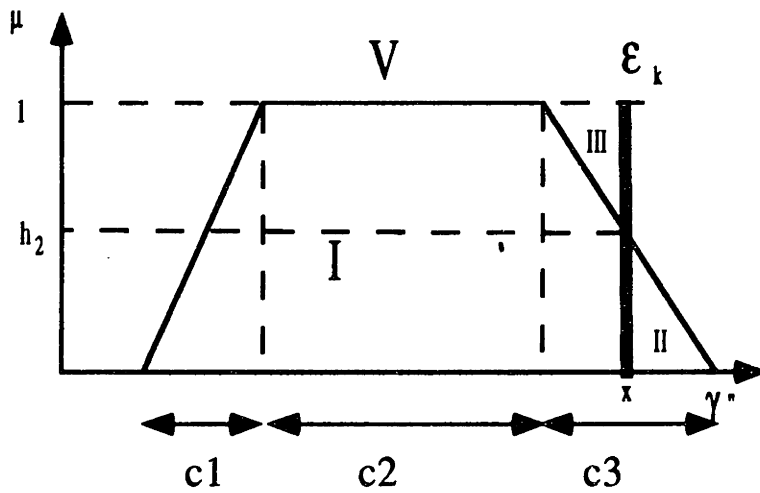
where areas I, II, and III are determined by x , as demonstrated in Figure 15c.



(a)



(b)



(c)

Figure 15: Intersections with the Fuzzy Set V

hence:

$$\frac{1}{2}w_h(\gamma''-x)h_2 > \frac{1}{2}(x-\gamma''+c_3)(1-h_2) + w_1[\frac{1}{2}c_1+c_2+\frac{1}{2}(1+h_2)(x-\gamma''+c_3)] \quad (4-75)$$

where:

$$h_2 = \frac{\gamma-x}{c_3} \quad (4-76)$$

hence:

$$x = \gamma'' - \frac{1}{2}c_3 - \frac{1}{2}w_1(c_1+2c_2+c_3) \quad (4-77)$$

Substituting $\gamma'' = \gamma + c_1+c_2+c_3$, and $w_1 = 1 - w_h$, we get that equation (4-77) is equivalent to equation (4-66). ■

Corollary:

The probability of choosing alternative j over i is given by:

$$\text{Prob}(j) = \text{Prob}(\epsilon \leq V) = \int_{-\infty}^x f(u) du \quad (4-78)$$

where f is the probability density function of ϵ , and x is given by equation (4-66).

The Estimation Procedure

Without loss of generality, we assume that β_p 's, the coefficients to be estimated for the utility function (as they appear in equation (4-63)) are non-negative. If, for example, we expect that some attribute k has negative effect on the utility (e.g. travel time), then we take the opposite image of the attribute value X_{ip} . The opposite image of a T_pFN A defined by: $A = (a_1, a_2, a_3, a_4)$ is given by (Kaufmann and Gupta, 1985 p. 70):

$$A^{ops} = (-a_4, -a_3, -a_2, -a_1)$$

For two T_pFN's A and B such that:

$$A = (a_1, a_2, a_3, a_4)$$

$$B = (b_1, b_2, b_3, b_4)$$

the result of adding A and B is a T_pFN given by (see section 4.1):

$$A + B = (a_1, a_2, a_3, a_4) + (b_1, b_2, b_3, b_4) = (a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4) \quad (4-79)$$

Subtraction of B from A would again result in a T_rFN given by (Kaufmann and Gupta, 1985 p. 70):

$$A - B = (a_1, a_2, a_3, a_4) - (b_1, b_2, b_3, b_4) = (a_1 - b_4, a_2 - b_3, a_3 - b_2, a_4 - b_1) \quad (4-80)$$

Let γ_{ip} be the left point of the T_rFN X_{ip} , and c^1_{ip} , c^2_{ip} , and c^3_{ip} its left, middle, and right spreads respectively. Thus, it follows that the left-most point, γ , and the spreads, C_1 , C_2 , and C_3 of the T_rFN, $V = V_j - V_i$, are given by:

$$\gamma = \sum_p \beta_p [(\gamma_{jp} - \gamma_{ip}) - (c^1_{ip} + c^2_{ip} + c^3_{ip})] \quad (4-81)$$

$$C_1 = \sum_p \beta_p (c^1_{ip} + c^3_{ip}) \quad (4-82)$$

$$C_2 = \sum_p \beta_p (c^2_{ip} + c^2_{ip}) \quad (4-83)$$

$$C_3 = \sum_p \beta_p (c^3_{ip} + c^1_{ip}) \quad (4-84)$$

If w_h and w_l are known, then for a given set of K observed choices, and for a given pdf for ϵ (typically normal or logistic), we can write the likelihood function as:

$$L^*(\beta_1, \dots, \beta_p) = \prod_{k=1}^K \text{Prob}(i)^{y_{ik}} \text{Prob}(j)^{y_{jk}} \quad (4-85)$$

where:
$$y_{ik} = \begin{cases} 1 & \text{if the } k^{\text{th}} \text{ observed choice is } i \\ 0 & \text{if the } k^{\text{th}} \text{ observed choice is } j \end{cases}$$

and:

$$\text{Prob}(j) = \text{Prob}(\epsilon \leq V) = \int_{-\infty}^{\gamma + \frac{1}{2}C_1 + \frac{1}{2}w_h(C_1 + 2C_2 + C_3)} f(u) du \quad (4-86)$$

$$\text{Prob}(i) = 1 - \text{Prob}(j) \quad (4-87)$$

where γ and C_1 , C_2 , and C_3 are given by equations (4-81), (4-82), (4-83) and (4-84). Thus, maximum likelihood estimation techniques can be used to estimate the parameters β_1, \dots, β_n (see Ben-Akiva and Lerman, 1985, for the general methodology). If w_l and w_h are unknown, then the estimation becomes non-linear.

CHAPTER 5: APPROXIMATE REASONING ROUTE CHOICE MODELS

5.1 The General Decision Framework

The approximate reasoning models described in this Chapter utilize concepts from approximate reasoning and fuzzy control. Appendix B describes briefly the general approximate reasoning and fuzzy control frameworks. Figure 16 illustrates the general scheme of the decision process adapted for the route choice problem as a three phase process.

The Approximate Reasoning Phase:

Current inputs, A^* , are applied to a set of N rules, and the appropriate outcome vectors $(B_i^*(1), \dots, B_i^*(m))$ are derived, $i=1, \dots, N$ (m is the size of the choice set).

The Internal Presentation Phase:

All rules that were fired to a positive extent are **combined** simultaneously into one vector: $(B^*(1), \dots, B^*(m))$, which serves as an internal presentation of the attractiveness of the alternatives in the choice set.

The Defuzzification Phase:

The attractiveness of the alternatives in the choice set are compared, and the best alternative is chosen.

This general framework is very attractive for modeling route choice decisions made in the presence of information. The on-line information and its parameters (e.g. reliability, relevance, and salience), the time pressure for making decisions, and possible new alternatives which may have to be considered, impose additional constraints into the decision process. The approach we propose for modeling this decision process has the potential to account for some of the biases that exist in making choices which require huge information processing capabilities (as discussed in Chapter 2). It is well suited for dealing with missing data and unavailable information, and thus it has the potential to model realistically route choice decisions made en-route while new information is being acquired dynamically.

For the purpose of this work, we again assume that perceived travel time and traffic information are the most important factors in making route choices, but the ideas presented in what follows can also be applied when other factors are considered.

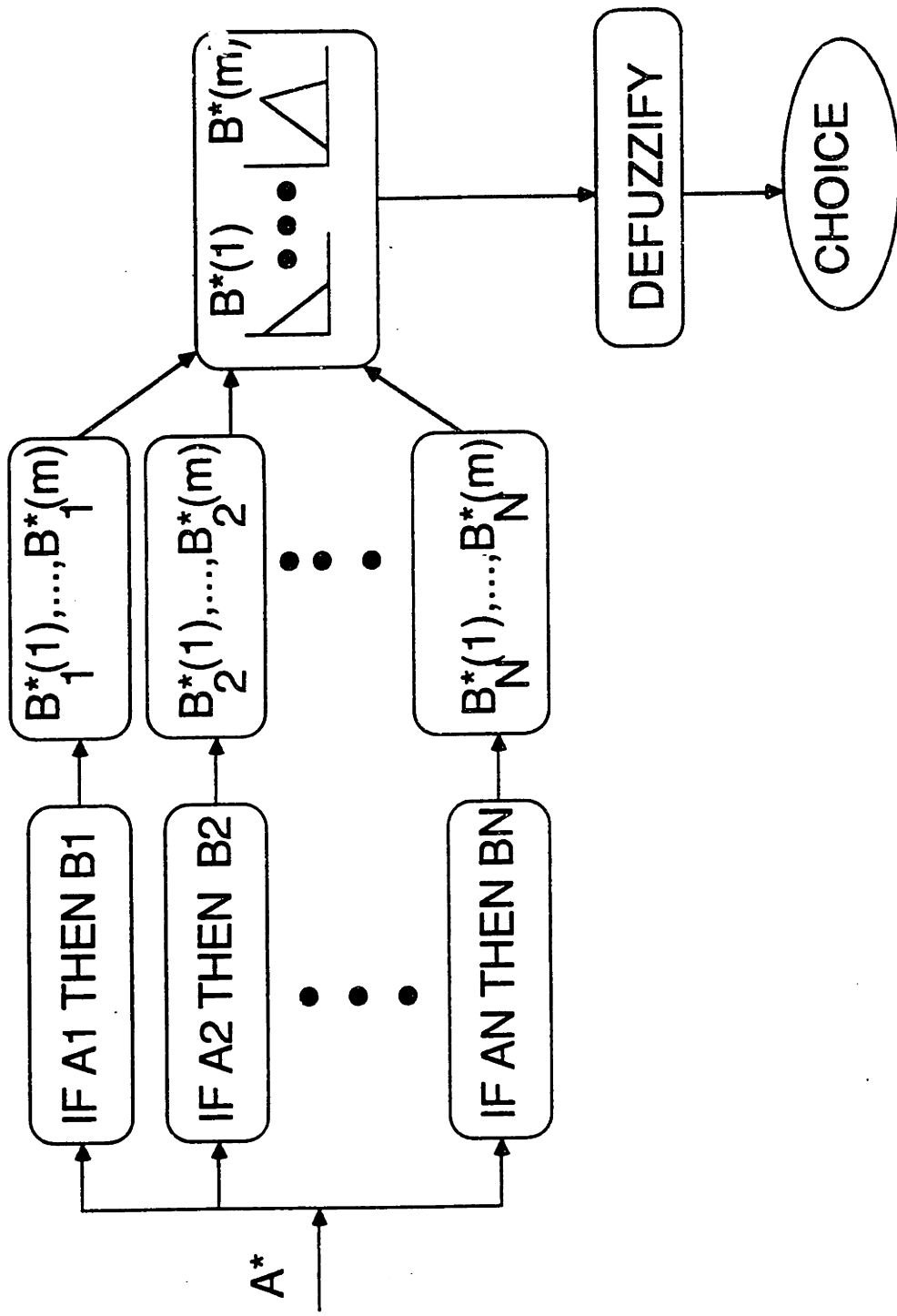


Figure 16: The Decision Framework

5.2 Rules

The building blocks of the decision process are rules of the form: if A_i then B_i , which associate the state of the system with choice-related attitudes and preferences. The use of rules resembles other rule-based systems in which decisions are related to specific input conditions (e.g. expert systems). However, the condition and the consequence part of the rules can include fuzzy labels, and thus the rules become more general and intuitive, especially when modeling human thinking. Examples of such typical reasonings are:

- "It's Monday, I'd better take Route 2".
- "If Storrow Drive is very bad - I'll probably take Memorial Drive".
- "I'll never take the Mass. Pike in rush hour again".
- "Yesterday Memorial Drive was terrible, I'd better not take it today".
- "It looks like an accident , I am going to switch quickly to Route 9".
- "The radio report said that 128 is usual, but it looks bad to me, I'd better take Highland Avenue".
- "Bumper-to-bumper on The Mass. Pike. - what can I do, that's my only alternative..."

These reasonings (or associations) correspond, for example, to directed paths from the left-hand-side of Figure 1 to its right-hand-side, and the different short-cuts correspond to different behavior patterns, and different information processing levels. In this section we talk more formally about rules to model such reasoning, and we narrow down into more specific rule structures. However, it should be kept in mind that the motivation for the use of rules (as discussed in Chapter 2) stems from the fact that they correspond to a natural and reasonable approach for modeling human thinking.

5.2.1 Rules Structure

To model the decision process we use rules; rule i has the general form: "if A_i then B_i ", where the left-hand-side (LHS) of the rule is represented by the statement A_i , and the right-hand-side (RHS) by the statement B_i . Depending on the model used (simultaneous, two-stage, or default), the LHS of a rule deals with the various possible traffic conditions, traffic information, and other relevant data associated with the alternative paths. The right-hand-side is choice related, but does not correspond directly to choice, rather it serves to model attitudes and preferences with respect to the alternatives in the choice set. In general A_i and B_i are multi-dimensional vectors defined as:

$$\begin{aligned} A_i &= (A_i^1, \dots, A_i^m) & i=1, \dots, N \\ B_i &= (B_i^1, \dots, B_i^m) & i=1, \dots, N \end{aligned}$$

where m is the number of alternatives in the choice set, A_i^j is the j 'th component of the

LHS of the i 'th rule and corresponds to the state of alternative j . Similarly, B_i^j deals with the attractiveness of choosing alternative j in view of the information conveyed by the vector A_i . For example:

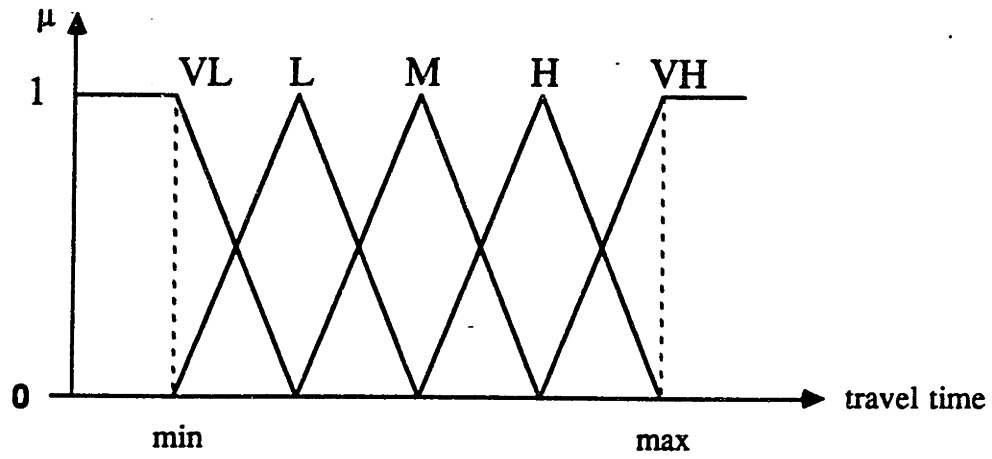
- if travel time on path p is very low then I'll take it
- if path p is much worse than usual and path q is usual then I'll probably take path q
- if there is an accident on path p then I'll definitely not take path p and I'll probably take path q .

We do not expect to have all m components of A_i and B_i filled for every rule i , since we do not expect that drivers make such complex multi-dimensional judgements when making a route choice. Rather, our hypothesis is that the final choice is based on a combination of many **simple** considerations each of which is modeled by a relatively "simple rule". Rules may belong to two general categories:

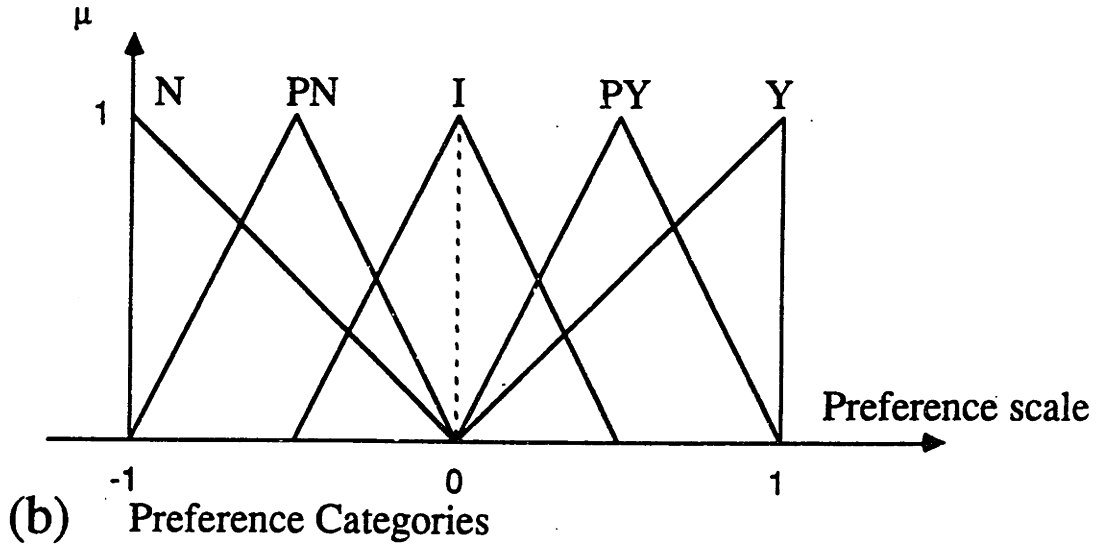
1. rules dealing with perceived travel times (or other measures of attractiveness),
2. rules dealing with traffic information.

The LHS of the first group of rules characterizes a given performance measure (e.g. travel time) according to fuzzy labels. For example, travel time on a path can be categorized into one of the following five fuzzy sets: Very Low, Low, Medium, High and Very High travel times, as shown in Figure 17a. Thus for this case: $A_i^j \in \{VL, L, M, H, VH\}$. The specific design is influenced by the length of the interval $[\min, \max]$, the shape of the membership function, and the amount of overlap between adjacent sets (see section 5.7.1 for discussion of generation of membership functions). The underlying design sets in Figure 17a do not correspond necessarily to perceptions (we do not expect perceptions to be symmetric), rather they serve as a conceptual scale for evaluation.

Rules dealing with traffic information are determined by the available traffic information, and depend on the type of information system. These rules are applicable to the simultaneous and the default model, and hence their LHS directly relates to the inputs given by the information system. For example, for an information system in the form of radio traffic reports on major facilities and known bottlenecks, the LHS would relate to traffic conditions on those facilities (provided that they overlap with paths in the choice set). On the other hand, the LHS of rules relating to an information system which provides path recommendations, will consist of the specific path recommendation, e.g. "if path P_j is recommended then..."



(a) Travel Time Categories



(b) Preference Categories

Figure 17: LHS and RHS Designs

When categorizing travel times, we have to define the underlying scale; the scale can be O/D related on which all the m alternatives are placed, or it can be alternative specific. The advantages of an O/D scale are generality of the rules and more natural comparison of alternatives (for example, "Very High" does not depend on the specific alternative under consideration). On the other hand, if ranges of travel times on the various alternatives differ significantly, the 5 suggested travel time categories could not provide fine enough characterization of a given perceived travel time if they are O/D related. Constructing the initial rule matrix should be consistent with the scale used.

The **right-hand-side** of the rules, the "then" part, corresponds to aspects of the final decision. The RHS serves as an intermediate step in the decision process and corresponds to the stage at which attractiveness (or utility) of each alternative is evaluated based on the input. The **multi-dimensionality** of the RHS representation captures the fact that even if the LHS of a rule relates to a specific alternative j , it could also affect perceptions of the attractiveness of another alternative k . Thus $B_i = (B_i^1, \dots, B_i^m)$, where B_i^j is the attractiveness of alternative j in view of the knowledge $A_i = (A_i^1, \dots, A_i^m)$. B_i^j is measured on a scale ranging from -1 to 1, with -1 corresponding to the case of complete aversion to taking alternative j , 1 corresponding to the case of choosing alternative j without reservations, and 0 corresponding to the indifference point. We define five fuzzy sets representing the driver's attitude towards taking an alternative:

N corresponds to "I will Not take this alternative",
 PN corresponds to "I will Probably Not take this alternative",
 I corresponds to "I am Indifferent with respect to taking that alternative",
 PY corresponds to "I will Probably take this alternative",
 and Y corresponds to "I will take this alternative".

Figure 17b demonstrates the design used for the preference sets on the given scale. According to this design positive preferences are located in the $[0,1]$ interval, and negative preferences are located in the $[-1,0]$ interval. The indifference attitude gives equal weight to positive and negative values, and the N and Y attitudes cover all the relevant preference intervals with clear inclination towards the extreme values of -1 and 1. Note that all 5 sets have equal area, a fact that simplifies the mathematical formulation.

5.2.2 Rules Complexity

The general rules structure, described in the previous section, permits great flexibility in generating the rules. In this section we categorize the rules into two levels of complexity, and discuss the behavioral assumptions underlying both levels.

First-Level Rules

We call "first-level" rules, rules corresponding to the "rule-based behavior"

discussed in Chapter 2. Rasmussen (1986), suggested that this behavior corresponds to rules controlling a sequence of sub-routines in a familiar work situation. The rules have been previously derived empirically, or communicated from other persons' know-how as an instruction or cook-book recipe. The performance at this level is goal-oriented but very often the goal is not explicitly formulated.

At the first level the rules are **simple**. This simplicity is consistent with the characterization of human behavior by Simon (1969), who claims that: " a man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity of the environment in which he finds himself ". In the proposed approach this simplicity is represented by simple and straightforward rules which correspond to common-sense knowledge and intuitive behavior. The complexity of the overall decision process is then captured by the existence of multiple (possibly conflicting) rules which are being processed simultaneously. Use of simple rules is also very attractive in light of previous discussion (in Chapter 2) about individuals using heuristics and rules of thumb to reduce the amount of information processing needed when making decisions; it is natural and intuitive to assume, for example, that under time pressure, drivers can not process rules with a high level of sophistication.

Consequently, we define the first-level rules to have **one dimensional LHS and RHS**. First-level rules capture, as explained before, common-sense behavior, for example, on the same path low travel times are preferred to high travel times, and paths with "bad" traffic conditions (e.g. paths with accidents) have low attractiveness. First-level rules are organized in a block matrix of the form:

LHS ₁	LHS ₂	...	LHS _m	RHS ₁	RHS ₂	...	RHS _m
x				y			
x				y			
	x				y		
	x				y		
			x				y
			x				y

where $x \in \{VL, L, M, H, VH\}$ or x is the available traffic information, and $y \in \{N, PN, I, PY, Y\}$.

From a behavioral point of view first-level-rule type of behavior is analog to a

random utility model in which the utility of alternative i is comprised of a group of factors and considerations regarding only the state of alternative i , and no explicit interactions are allowed. The intuitive appropriateness of the model is then determined by the coefficients having the expected sign (+/-) and a reasonable (relative) magnitude.

Second Level Rules

At the second level, allowance for more complicated behavior is made, corresponding to the knowledge-based behavior that was discussed in Chapter 2. At this level performance is goal-controlled as it has to be able to deal with unfamiliar situations. The rules of this level can capture more complex reasoning and interactions, implicit preferences, as well as apparently irrational and non-intuitive behavior. Rules at the second level are characterized by:

- (a) Multi-dimensional LHS
- (b) Multi-dimensional RHS
- (c) Multi-dimensional LHS and RHS

Interactions of type (a) correspond to the case in which the driver considers simultaneously traffic conditions on several alternatives, and concludes about the attractiveness of a certain path. Interactions of type (b) correspond to the case in which traffic conditions on one path also affect the attractiveness of other paths. Interactions of type (c) typically correspond to complex information processing which could exceed the limited human information processing capabilities. Furthermore, their computational complexity is enormous. Hence, for the discussion that follows, we concentrate on interactions of types (a) and (b).

The interactions of type (a) and (b) above have a similar flavor: travel time on alternative i affects the attractiveness of another alternative j . But the strength of the effect differs; type-(b) interactions are stronger since the condition part in the LHS is less restricted. And indeed interactions of type (b) correspond to a higher level of knowledge since limited information has multi-dimensional consequences, whereas in interactions of type (a) multi-dimensional information results in a single dimension preference. Furthermore, interactions of type (a) are less attractive because they relate more to the choice itself rather than to modeling the internal representation of the attractiveness of a single alternative which eventually leads to a choice. Hence, we assume that, given traffic conditions on several alternatives, a user can come up with the resulting choice, but not necessarily with the attractiveness of a specific alternative. The decision process that we try to capture is the process in which information is translated into attractiveness and then into choice, whereas it seems that in (a) information is translated directly into choice.

The structure of the rules (first and/or second level) is correlated with the amount of experience that the driver has. Experienced drivers may utilize more second-level rules

since good knowledge and understanding are needed in order to deduce how traffic conditions on alternative i will affect the attractiveness of alternative j. Inexperienced drivers, on the other hand, can deduce only the simple rules that relate to each alternative separately, and are unaware of existence of interactions or implicit relations among alternatives.

Thus if we allow first level rules, and type (b) interactions, we can capture the two extremes regarding user's experience and sophistication, and allow enough flexibility in between to model a realistic decision process. Our approach would be to start with a set of rules at the first level, and to use observed choices to:

- Validate and modify first-level rules.
- Generate interactions of type (b).

5.2.3 Model-Specific Rules

The general rule structure discussed in the previous two sections depends on the specific model used. In this section we present a general rule structure for the three models that were introduced in section 2.1.

The Two-Stage Model: In the two-stage model, the LHS of the rules relates to perceived travel times. Perceived travel times are updated based on prior knowledge and information that was available at an earlier stage. The number of initial rules is of order $m \cdot k$, where m is the number of alternatives in the choice set, and k is the number of categories of travel time. A typical rule has the form: "if travel time on path j is very low then I'll probably take path j ".

The Simultaneous Model: In the simultaneous model, perceived travel times, and information compose the LHS of the rules. However, due to the underlying assumptions of the model, there are no direct interactions among all those factors, meaning that there is a set of rules dealing with perceived travel times, and a separate set of rules related to information (as discussed in section 5.2.1). The number of initial rules is of order $m \cdot (k+r)$, where k and m are as before, and r is the number of categories of the provided information.

The Default model: In the default model, as in the simultaneous model, there are two groups of rules: rules related to prior knowledge, and rules related to information. However, unlike the simultaneous model, the rules pertaining to prior knowledge are being treated as default rules, and hence they are being applied if no exception occurs. The default reasoning logic that was explained in section 3.4.2 is used to determine the extent to which an exception has occurred, this process is explained in more detail in section 5.5.

5.3 Input Propagation through Rule Premises

The core of the information processing part of the decision process lies in the **approximate reasoning** scheme which allows flexible rule interpretation, as well as rule adjustments (as explained in Appendix B). This phase distinguishes the suggested model from traditional rule-based models (e.g expert systems). Flexible rule interpretation is achieved by allowing rule premises to be partially true. The extent to which a rule premise is fulfilled is determined by the degree with which the given inputs match the rule premise, and that determines the extent to which the rule will be fired. In this section we describe how the match between inputs and rule premises is determined.

The inputs to the decision process are values of the attributes under consideration. For our case this includes perceived travel times (based on observation and/or prior experience), and traffic information on the various alternatives. The inputs feeding the decision process differ among the three different models that were proposed in section 2.1. For the simultaneous model the inputs include prior perceptions, information, current observations and possibly other factors. For the two-stage model the inputs for the second stage include updated perceptions of the relevant attributes (assuming that those perceptions were updated using information at the first stage). As for the default model, its input is default behavior and its relevant inputs, and information. Thus in general, the input A^* is a vector: $A^*=(A^{1^*}, \dots, A^{m^*})$ where m is the number of alternatives in the choice set. For rules dealing with perceived travel times, A^{i^*} is the travel time, as actually perceived by the driver, on alternative i . Perceived travel times will typically be modeled as fuzzy sets following the discussion in section 3.2.1. For rules dealing with traffic information, A^{j^*} corresponds to the information conveyed on traffic conditions on alternative j , and depends of the type of information system available.

The perceived travel time on alternative j , A^{j^*} , does not necessarily coincide with one of the travel time categories described in section 5.2.1 and illustrated in Figure 17a; rather it is likely to have some overlap with more than one of the underlying design sets. The relationship between the underlying design sets of the LHS of the rules and the input to the rules is illustrated in Figure 18 where the thick lined membership functions corresponds to the input A^* . In Figure 18a the input is deterministic (corresponding for example, to exact prediction of travel time), and in case (b) the input corresponds to the fuzzy set labeled "travel times experienced on alternative j " as discussed in Chapter 3. For illustration purposes lets look at a single alternative, and let $\mu_{A^*}(\cdot)$ be the membership function of the input A^* on that alternative. Let α_{VL} , α_L , α_M , α_H , and α_{VH} be the amount of overlap between the input A^* and the sets "Very Low", "Low", "Medium", "High", and "Very High" travel times on that alternative respectively.

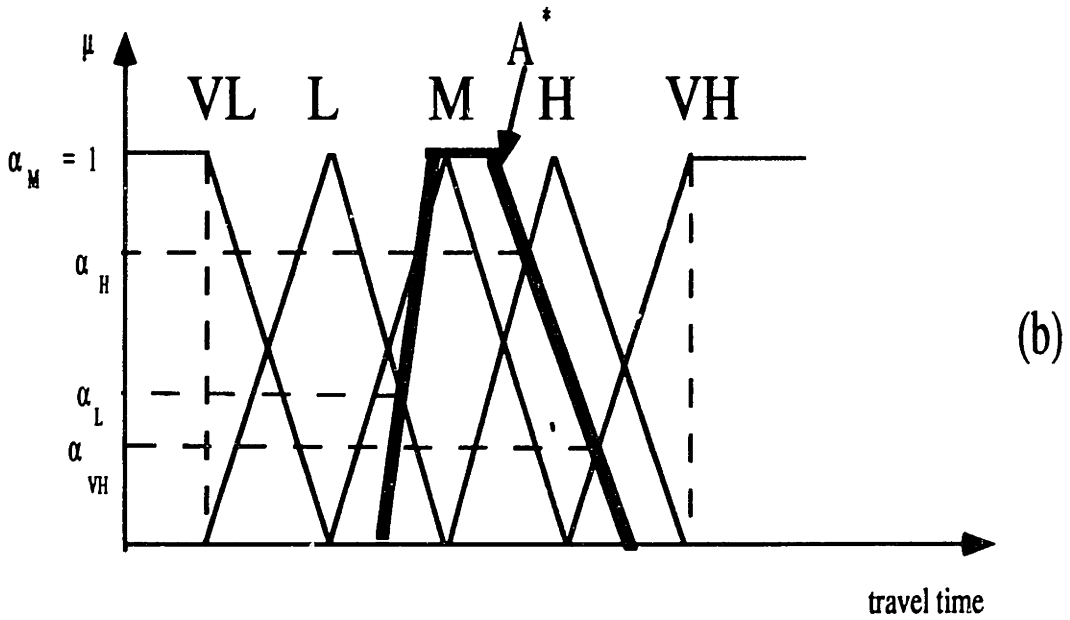
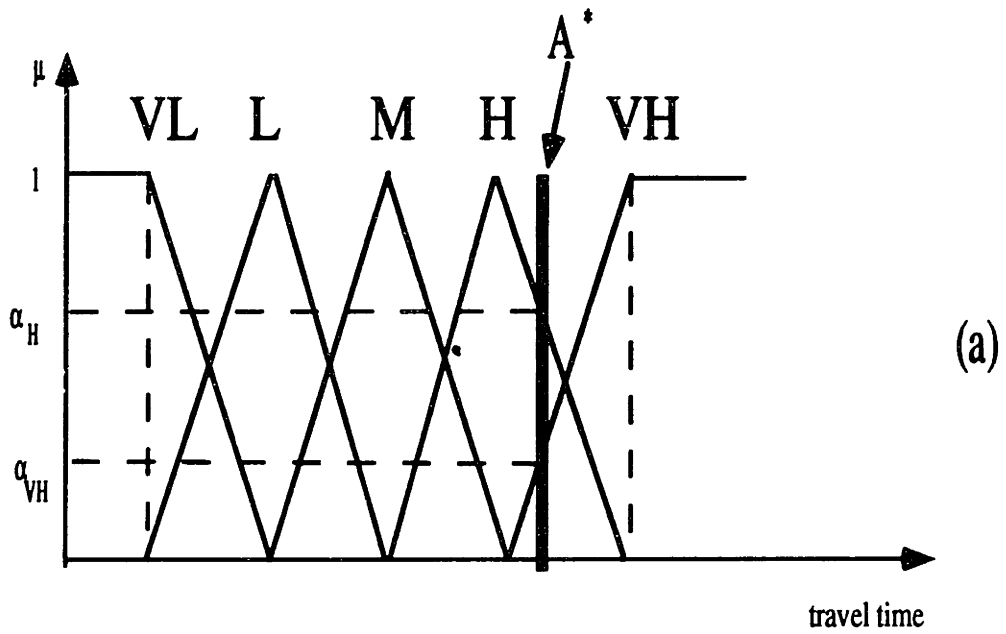


Figure 18: Overlap Between Perceptions and Design Sets

α_i , the degree of overlap between A^* and the fuzzy set represented by the membership function μ_i , is given by the max-min composition:

$$\alpha_i = \max_x \min(\mu_{A^*}(x), \mu_i(x)) \quad (5-1)$$

The max-min composition corresponds to the highest membership degree among all elements that are common to both sets. α can get values in the interval [0,1]; $\alpha=1$ means that there is overlap between the flat section of the T_rFN and the peak of the design set, and $\alpha=0$ corresponds to the case where there are no common elements to the two sets. The deterministic travel time in Figure 18a, belongs with degree α_{vH} to the set "very high travel time", and with degree α_H to the set "high travel time". Thus we can categorize this travel time as being somewhere between "very high" and "high", and since $\alpha_H > \alpha_{vH}$, we can conclude that it is closer to "high" travel times than to "very high" travel times. In a similar way, we can categorize linguistically a fuzzy set corresponding to travel time perceptions, such as the fuzzy set A^* , demonstrated by the thick line in Figure 18b. The perceived travel time, A^* , is "very low" with a degree $\alpha_{vL}=0$, "low" with degree α_L , "medium" with degree α_M , "high" with degree α_H , and "very high" with degree α_{vH} .

It is easy to verify that given the underlying design sets of Figure 18 and the values of α_{vL} , α_L , α_M , α_H , and α_{vH} , the input set, A^* , is uniquely defined. This is a very desirable property for the presentation of perceptions, and an important guideline in determining the underlying design sets.

The fact that a certain input agrees with different design sets, reflects the fact that travel time can not necessarily be exactly associated with *one* pre-determined category, rather it can be looked at as belonging to different categories. The various *degrees* associated with the overlap of a given input with the different design categories, reflect partial truth values, and will be utilized in executing the relevant rules as explained in the next section.

5.4 Rule Execution

In this section we discuss how rules whose premises have positive overlap with the given inputs are executed (or fired). Rule execution is based on the approximate reasoning scheme described in Appendix B, and results in modified rule consequences as illustrated in Figure 16.

This process is consistent with the description of human performance models that can be found in the literature. Rasmussen (1986), for example, asserts that: "the efficiency of human cognitive processes seems to depend upon an extensive use of model transformations together with a simultaneous updating of the mental models in all categories with new input information, an updating which may be performed below the level of conscious attention and control". We use the approximate reasoning scheme,

described in Appendix B, to account for model transformations and unconscious updates. Thus the approximate reasoning phase deals with rule interpretations, and serves to adjust the rule-base to current inputs, and to allow adaptable rule structure. It agrees well with information processing under pressure and imperfect and limited information processing capabilities.

The main issue regarding rule execution, is which implication scheme to use to model the relation "if A then B". In Appendix B we presented the two most commonly used implications and discussed the way they influence the resulting RHS set B^* (Mamdani's implication, and Kosko's correlation-product encoding). In general, implication schemes can be categorized into two broad families:

- Implication schemes which result in $B^* \subseteq B$
- Implication schemes which result in $B^* \supseteq B$

Mamdani's implication and the correlation-product scheme implication belong to the first group. The decision of which family to use is context dependent, although we have not yet found any application of fuzzy control which uses the second family of implication schemes.

For our case, the set B and its transformation B^* correspond to the attractiveness of a certain alternative (as represented by a fuzzy set). To assess the appropriateness of an implication scheme, we examine the effect on attractiveness when attractiveness becomes vaguer: does it become more or less diffused? Let us look, for example, at the case $B = \text{"I will probably take alternative j"}$. In order to make this statement vaguer we could have a set B^* which would be either a subset or a superset of the original set B. A subset interpretation seems more appropriate for this case since it would infer that we can not guarantee that "I will probably take alternative j" will be definite ($\mu=1$) for some value on the possible scale, whereas a superset B^* will guarantee that the attitude "I will probably take alternative i" will definitely take place for some interval on the relevant scale. When modeling travel time, in section 3.2, we insisted on having normalized fuzzy sets since at least one travel time was certain to be realized. However, the phenomenon that we are modeling on the RHS of the rules is different; it does not relate to event occurrence, rather it relates to attitudes and preferences, and thus the normalization requirement is not desirable. Having a set B^* that is a subset of the original set B would imply that the preference modeled by B^* is weaker and thus gets lower membership values, whereas a superset would imply stronger membership values for that preference. Thus we use implication schemes for which $B^* \subseteq B$. Among them we use the correlation-product encoding scheme (as given by equation (B-6)) which preserves the shape of the original set B, and in addition it has some desirable mathematical properties (which will be used in the ILP formulation for rule calibration, discussed in section 5.7.2.1).

In general the input A^* is m dimensional, and thus α of equation (B-5) has to be changed accordingly. Let $A^* = (A^{1*}, \dots, A^{m*})$, and let the LHS of the i'th rule be

$A_i=(A_i^1, \dots, A_i^m)$. Each element A_i^j of A_i^* is matched against the corresponding element A_i^j from A_i in order to obtain α_i^j , the overall degree with which the i 'th rule is being fulfilled:

$$\alpha_i = \min_j(\alpha_i^j) \quad (5-2)$$

where:

$$\alpha_i^j = \max_x \min(\mu_{A_i^j}(x), \mu_{A_i^j}(x)) \quad (5-3)$$

The minimum operation used in equation (5-2) captures the fact that all the condition parts of rule i (the elements of the vector A_i) have to be fulfilled simultaneously. Thus α_i is the degree with which the condition: " A_i^1 is A_i^1 , and ... and A_i^m is A_i^m " is true.

α_i is interpreted as the strength of the i 'th rule (given the current inputs), and thus it serves as the degree with which the i 'th rule is being "fired" or executed. It is clear that more than one rule can have $\alpha_i > 0$, thus more than one rule can contribute to the final decision. Figure 19a shows an example in which 4 rules were fired, each to a different extent. Thus the attractiveness of the alternative, is the result of the 4 relevant rule consequences. It can be seen that in general the alternative has a positive attractiveness since the "PY" and "Y" RHS values were fired to the largest degrees. The non-linear nature of the process can also be seen in Figure 19a; a change in the degree with which a certain rule is fired, will cause an increase or decrease in the amount by which the corresponding RHS is fired. As a result the relevant rule can dominate the overall attractiveness or have negligible effect.

5.5 Rule Combination

All rules whose premises have non-empty overlap with current inputs (i.e. rules i with $\alpha_i > 0$), are being fired, each with a different degree (α_i). The result is N (or less) B_i^* vectors, $B_i^*=(B_i^{1*}, \dots, B_i^{m*})$ as illustrated in Figure 16, each of which corresponds to the attractiveness of the m alternatives. Consequently, the attractiveness of an alternative is composed of results of several rules, each contributing to a different aspect of the final decision as shown in Figure 19a. Suppose that there are N rules "if A_i then B_i ". After applying the input A^* to all the rules, we obtain up to N vectors of fuzzy sets B_i^* , $B_i^*=\{B_i^{1*}, \dots, B_i^{m*}\}$. For each alternative j , we have to combine the individual B_i^{j*} 's over all the rules i into a set B^j .

Combing the B_i^{j*} 's is a critical part of modeling the decision process and thus requires careful consideration. The aggregation scheme used to combine B_i^* should reflect the nature of the problem under consideration and the interpretation of the fuzzy sets involved. The aggregation corresponds to the process in which different pieces of information and their resultant implications are combined.

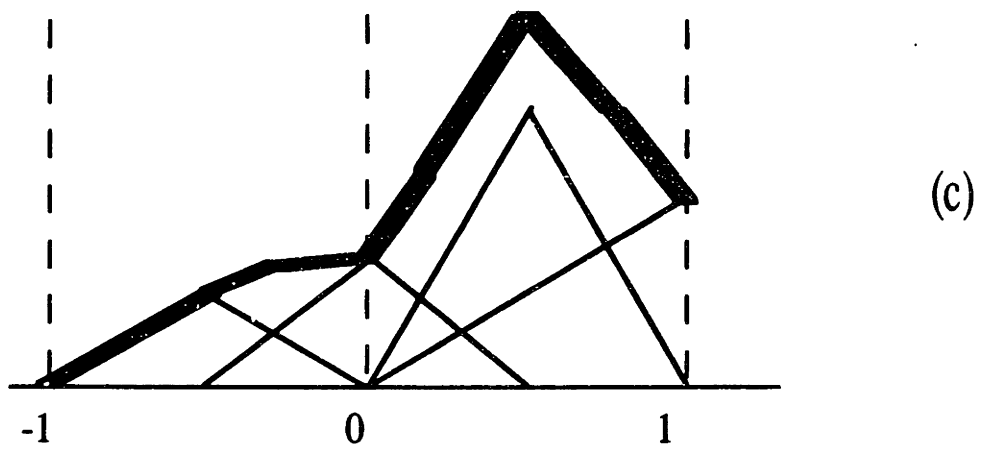
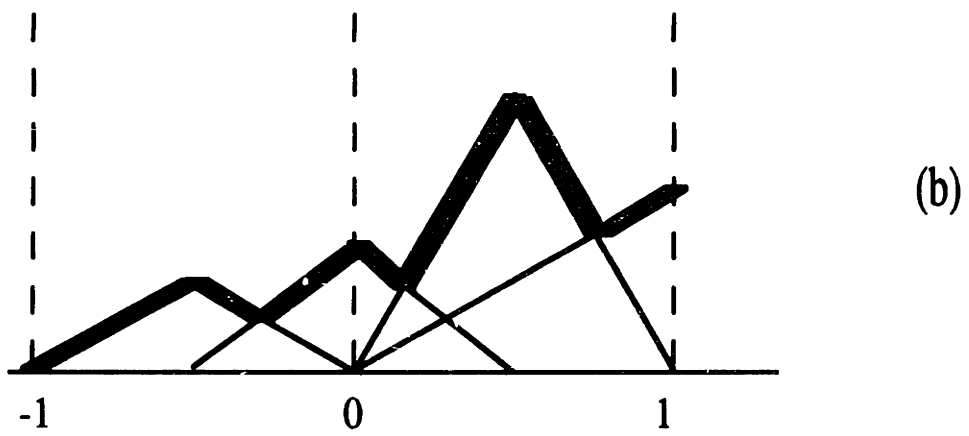
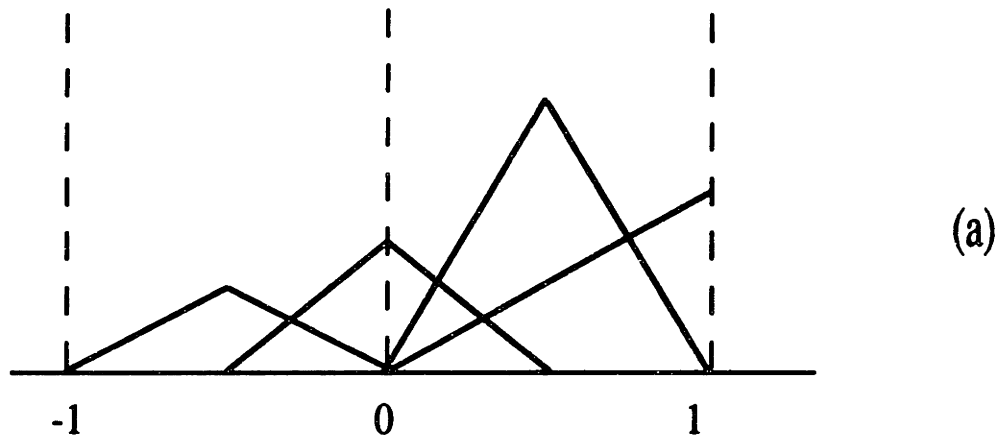


Figure 19: Attractiveness of an Alternative

The fuzzy control framework provides a mechanism to combine the inferred sets B_i^{j*} given by equation (B-7) in Appendix B, which corresponds to taking the union of all B_i^{j*} 's. This approach is motivated by the fact that $\mu_{B_i^{j*}}(y)$ dominates $\mu_{B_k^{j*}}(y)$ if $\mu_{B_i^{j*}}(y) \geq \mu_{B_k^{j*}}(y)$ since it represents a higher possibility for the value y to occur. The idea of dominance, however, is not appropriate for the process we model; The set B^{j*} is the final attractiveness of alternative j , obtained by application of different rules, where each rule corresponds to a different aspect of alternative j being attractive or not. Thus taking the union, as suggested by equation (B-7), and as demonstrated in Figure 19b by the thick line, would correspond to loss of information. In order to make sure that all the evidence contributing to the attractiveness of alternative j is considered (and not only the strongest evidence), we use the following aggregation scheme (Kosko, 1992):

$$\mu_{B^{j*}}(y) = \sum_{i=1}^N \mu_{B_i^{j*}}(y) \quad (5-4)$$

This aggregation scheme considers the contribution of all rules (and the shapes of their membership functions) to the attractiveness of each alternative, and thus it is more appropriate. It is demonstrated in Figure 19c by the thick line.

A more general version of this aggregation scheme is given by:

$$\mu_{B^{j*}}(y) = \sum_{i=1}^N w_i \mu_{B_i^{j*}}(y) \quad (5-5)$$

where w_i is the weight of rule i . Rule weights were suggested by Kosko (1992), to model credibility, frequency or strength of rules. For our case rule weights can have two major applications: to weigh default rules in the default model, and to model reliability of the provided information the simultaneous and the default models. The rule weights that we use are always less than 1, thus we can view them as a way to discount the degree with which a rule is fired by changing α_i into $\alpha_i w_i$.

In the default model we distinguish between two groups of rules: default rules which correspond to routine behavior under usual circumstances, and information rules, relating to the provided traffic information. Default rules are fired only to the extent to which usual conditions have occurred. This is done by weighing the default rules by weights corresponding to the consistency between usual conditions, and actual conditions (as reported by the information system). If conditions are usual, default rules are fired to the full extent. However, when unusual conditions are reported, the associated information rules dominate the decision, while default rules are fired to a lesser extent. For a default rule, "if A_i then B_i ", where A_i relates to perceptions of usual conditions, we propose to use: $\text{Poss}[A^* | A^{**}]$ or $\text{Cert}[A^* | A^{**}]$ (as defined in Appendix A) as w_i , the weight for rule i , where A^* represents travel time perceptions, and A^{**} is the information. Thus, default rules are being fired only to the extent with which they relate to perceptions that are consistent with information, while they are ignored when they are not consistent.

If no information exists then they are fired to the full extent (depending of course on the agreement between A_i and A^*). If for example, w_i is given by $\text{Poss}[A^* | A^{**}]$, and $\text{Poss}[A^* | A^{**}] = 0$, then prior perceptions do not have any overlap with the provided information, and the relevant default rule is ignored. However, if $\text{Poss}[A^* | A^{**}] = \gamma$, $0 < \gamma < 1$, then perceptions and information overlap to the extent γ , and thus the strength of the i 'th rule is decreased by $1 - \gamma$, the amount by which perceptions and information are not consistent.

Rule weights can also be used to model the reliability of the information provided. If the information source is reliable the corresponding rules are fired as determined by the appropriate α_i , whereas unreliable information source will result in discounting the strength of the fired rule.

5.6 Defuzzification

5.6.1 General

The defuzzification phase deals with translating the combined RHSs, B^* 's, into a control action, or as in our case into choice. In fuzzy control, the defuzzification scheme most often used is the center of gravity method given by equation (B-8) in Appendix B. The final control action, z , usually corresponds to the level of operating a certain control device such as speed, temperature, flow etc. If several devices are operated then the output is multi-dimensional, and each action to be taken has a corresponding membership function. The membership functions are defuzzified separately and independently and result in operating instructions to be executed by the corresponding device.

However, the route choice problem is a discrete choice problem, hence even though the output is multi-dimensional, given the attractiveness of each of the m alternatives, the model should finally provide a single discrete choice. That is, the defuzzification scheme has to result in a discrete choice based on m fuzzy sets (and not one). There are several possibilities to address this problem. Two main approaches are described below:

- Compare m fuzzy sets using some comparison method, and choose the alternative whose attractiveness is the highest.
- Defuzzify each of the m membership functions, μ_j , separately into its centroid: z_j , and:
 - i. choose the alternative with the highest centroid, or
 - ii. treat the centroids as the systematic components of a random utility model.

There are different behavioral assumptions underlying the two approaches; in the first approach the final choice is based on the result of comparing m fuzzy sets, and in the second on comparing m scalars that are either deterministic or subject to random noise. For the first approach there is the issue of which is the appropriate comparison

method (which by itself imbeds behavioral assumptions), and the result is deterministic in the sense that the same user will make the same decision under identical conditions. We will focus on the second approach, which is very appealing for modeling the final choice since it can incorporate noise in the choice process.

5.6.2 A Random Utility Model (RUM) Defuzzification Scheme

Let z^j be the center of gravity of the fuzzy set B^j as given by equation (B-8) in Appendix B. B^j corresponds to the combined attractiveness of alternative j , and z^j to its defuzzification into a scalar. Thus, z^j can be viewed in a deterministic framework as the attractiveness of alternative j , and in a probabilistic framework as the systematic component of the utility of alternative j in a random utility model. That is, the utility of alternative j for individual n is given by: $Z_{jn} = z_n^j + \epsilon_{jn}$, where ϵ_{jn} is the random component. The underlying behavioral assumptions of this model are similar to those of the general RUM (Ben-Akiva and Lerman, 1985).

We call the model: $Z_{jn} = z_n^j + \epsilon_{jn}$ ARRUM (Approximate Reasoning Random Utility Model). ARRUM can be viewed as a combination of two processes:

- The systematic component, z_n^j , modeled by an approximate reasoning process, which associates the available (possibly fuzzy) inputs into attractiveness of the various alternatives in a non-linear fashion. This process models the underlying decision process as a combination of "if-then" considerations which are processed together to come up with a final attractiveness for each alternative. Thus, the attractiveness of each alternative is the result of simple and rational considerations as well as multi-causalities, interactions among alternatives, and possibly irrational behavior.
- The random component, ϵ_{jn} , corresponds to random noise in human behavior. This random noise accounts for, empirically observed, inconsistent behavior. The same individual, for example, may choose different alternatives under (hypothetical) identical conditions. It also captures unexplained behavioral factors in the approximate reasoning process (e.g. missing rules or parts of rules).

In summary, both the traditional RUM and the suggested ARRUM model are random utility models in which the utility of each alternative i in the choice set is given by: $U_{jn} = V_{jn} + \epsilon_{jn}$, and the underlying assumption is that individuals try to choose the alternative with maximal utility. The difference lies in the underlying assumptions about the structure of the systematic component V_{jn} . Most existing RUM models assume a linear form for the systematic component (i.e. $V_{jn} = \beta' X_{jn}$). The suggested model, on the other hand, assumes an approximate reasoning process resulting in V_{jn} .

5.7 Implementation and Calibration

The implementation and calibration of the models depends on the nature of the specification: aggregate or disaggregate. The ultimate objective for practical use, prediction, and planning is aggregate modeling. However, disaggregate modeling is important by itself when implementing and testing new methodologies, and recently has gained additional motivation from detailed simulation models in which the behavior of individual drivers is being simulated. We assume that individual characteristics are captured by the different membership functions used to model perceptions as was discussed in chapter 3. With respect to rules, the models presented in this chapter can be implemented both for disaggregate and aggregate modeling. In disaggregate modeling we assume that each individual has its own set of rules that correspond to its own behavior pattern. For aggregate modeling we assume that a homogeneous group of users has a common set of rules, and the individual differences are being accounted for by the different individual perceptions that serve as inputs to the rules.

There are two major groups of variables in the approximate reasoning model: membership functions, and rules. As was discussed in section 3.2.1, travel time perceptions as modeled by membership functions are individual-specific and are influenced by the specific individual characteristics. Thus in implementing the models we maintain individual-specific membership functions in order to capture the differences among individuals.

The rule matrix can be individual-specific, or global (the same rules matrix for different individuals). Ultimately we believe that one rule matrix can be common to several individuals whereas the differences among individuals are captured through the different perceptions as modeled by the individual-specific membership functions. We do not think that different individuals have identical rules, however, because of differing perceptions, different rules will be fired to different extent for similar circumstances.

5.7.1 Generation of Membership Functions

In the proposed model there are two types of membership functions. The first corresponds to driver's perceptions of system attributes and information, and the second to design aspects of the model. The first type is the input to the model (A^*), and the second corresponds to rules' premises (A_i).

Modeling driver's perceptions of travel times and the generation of the corresponding membership functions was discussed in Section 3.2. In most applications of fuzzy set theory TFN's (triangular fuzzy numbers) or TrFN's (trapezoidal fuzzy numbers) are used, thus limiting the membership function estimation to 3 or 4 parameters. It is assumed that the shape of the membership function captures many of the individual characteristics with respect to the way the driver perceives the system, as well as personal attributes, preferences and familiarity with the system. Currently it seems that personal

interviews or surveys are needed in order to generate the membership functions involved. For further applications, we hope that future research will provide mechanisms to generate membership functions which would capture personal and system characteristics.

The second type of membership function is used to divide the feasible range of a given attribute into labels. These labels constitute the LHS and the RHS of the rules. An example appears in Figure 17a where the range of possible travel times is divided into 5 categories ranging from Very Low to Very High, and each set is modeled as a Triangular Fuzzy Number with 50% overlap between neighboring sets. The membership functions of the labels presented in Figures 17a and 17b do not correspond to actual perceptions and serve as a standard base on which different drivers perceptions could be compared.

5.7.2 Rule Generation

The rules are the building blocks of the approximate reasoning models presented in this chapter, and thus their generation needs careful consideration. In this thesis we assume that membership functions are obtained first, and rules are subsequently calibrated. An interesting future research activity would be to develop approaches for the calibration of fuzzy inputs and rules simultaneously.

In this section two approaches for rule generation are suggested. The first approach is an optimization approach which formulates the rule generation problem as an integer linear programming problem (section 5.7.2.1), and the second approach is heuristic in nature (section 5.7.2.2).

5.7.2.1 Integer Linear-Programming Based Rule Calibration

In this section we formulate the approximate reasoning model as an integer linear programming (ILP) problem. Typically we can not expect to have closed form representation and solutions for the models, but the special case considered in this section obeys all the requirements that we have specified in the previous sections, and thus presents a very attractive implementation of the model.

We assume that membership functions are given. If there are major faults with the way membership functions are generated, we expect to detect them through a poor fit with observed choices. Thus the rules are the unknowns of the problem. We start with an initial rule matrix, and use an improvement procedure to fill it, or test its appropriateness. Following the discussion in section 5.2, we treat the LHS structure of the rules as given, and treat the RHS vectors as variables. How many of the RHS components are variables depends on the degree of rule complexity that is being implemented (first or second level rules). Thus in general, the variables are the possible RHS entries, which can take exactly one of the following values: Y,PY,IPN,N.

For implementing the approximate reasoning model, we use the product-correlation implication scheme as given by equation (B-6) in Appendix B for rule execution, and the summation rule combination scheme given by equation (5-4). For this case Kosko (1992) showed that there is a closed form representation for the centroid of the combined set B^* which is given by:

$$z = \frac{\sum_{i=1}^N \alpha_i V_i S_i}{\sum_{i=1}^N \alpha_i S_i} \quad (5-6)$$

where α_i is the degree to which the i 'th rule was fired, V_i is the centroid of the fuzzy set corresponding to the RHS entry of rule i , and S_i is the area of this set (if $\sum_{i=1}^N \alpha_i S_i = 0$ z is equal to 0). If all the possible RHS outcomes have equal area (such as in the case of Figure 17b), then we have:

$$z = \frac{\sum_{i=1}^N \alpha_i V_i}{\sum_{i=1}^N \alpha_i} \quad (5-7)$$

Notation and Definitions

Inputs:

- K :** number of observed choices (index k)
- N :** number of rules (index i)
- m :** number of alternatives (index j)
- P :** number of possible RHS outcomes (index p)
- $\alpha_i(k)$:** degree with which rule i is fired for the k^{th} choice, (as calculated by equation (5-2))
- oc(k):** the k^{th} observed choice
- V_p :** the centroid value for outcome p
- M:** a very big number

Variables:

$Z^j(k)$ = centroid of alternative j for trip k

$$C_{ip}^j = \begin{cases} 1 & \text{if RHS}_i^j = p \\ 0 & \text{otherwise} \end{cases}$$

$$y_k = \begin{cases} 1 & \text{if correct choice was made for trip } k \\ 0 & \text{otherwise} \end{cases}$$

Formulation:

$$\max \sum_{k=1}^K y_k \quad (5-8)$$

subject to:

$$Z^j(k) = \frac{\sum_{i=1}^N \alpha_i(k) \sum_{p=1}^P C_{ip}^j V_p}{\sum_{i=1}^N \alpha_i(k)} \quad \forall j,k \quad (5-9)$$

$$\sum_{p=1}^P C_{ip}^j = 1 \quad \forall i,j \quad (5-10)$$

$$Z^{oc(k)}(k) \geq Z^j(k) - \epsilon_k \quad \forall j \neq oc(k) \quad \forall k \quad (5-11)$$

$$\epsilon_k \leq (1 - y_k)M \quad \forall k \quad (5-12)$$

$$\epsilon_k \geq 0 \quad \forall k \quad (5-13)$$

$$C_{ip}^j \in \{0,1\} \quad \forall i,j,p \quad (5-14)$$

$$y_k \in \{0,1\} \quad \forall k \quad (5-15)$$

The objective function as given by (5-8) maximizes the number of trips that are correctly predicted by the model. The set of constraints given by (5-9) defines the centroids $Z^j(k)$ as a function of the variables of the problem by utilizing equations (5-6) and (5-7). The set of constraints given by (5-10) guarantees that exactly one RHS entry is chosen. The group of constraints (5-11) corresponds to the deterministic defuzzification scheme in which the alternative which corresponds to the highest centroid value is chosen. It requires that the centroid of the chosen alternative (as observed) will be greater than the centroids of all other alternatives. However, since it is not clear

whether a feasible solution exists, we allow this group of constraints to be violated by amount ϵ_k for each trip k . If the correct choice was made, we want ϵ_k to be zero and y_k to be equal to one. This is guaranteed by the set of constraints (5-12); the objective function drives y_k to be equal to one whenever possible, and if y_k is one then ϵ_k must equal zero. Constraints (5-14) and (5-15) are the integrality constraints.

The number of variables for first level rules are $N \cdot P$ (there is a single RHS entry for each rule), and $N \cdot P \cdot m$ for second level rules.

Another group of constraints that can be added to the formulation are monotonicity constraints. Monotonicity constraints require that preference towards taking a specific alternative obeys direct weak monotonicity with respect to traffic conditions on that alternative. For example, they do not allow the following mapping (concerning the same alternative):

$$\begin{array}{lcl} \text{VL} & \rightarrow & \text{PN} \\ \text{H} & \rightarrow & \text{PY} \end{array}$$

which implies that if travel time is Very Low the alternative will probably not be chosen, whereas if travel time is High it will probably be chosen. This consistency requirement is somewhat analog to the "right" sign expectation in random utility models (we expect the sign of the travel time coefficient to be negative). This requirement can be imposed by the constraints:

$$\sum_{p=1}^P C_{i'p}^j V_p \leq \sum_{p=1}^P C_{i''p}^j V_p \quad (5-16)$$

whenever $LHS_{i'}^j$ is worse than $LHS_{i''}^j$.

Typically there are no closed-form representations for fuzzy control applications, hence the formulation presented in this section is very appealing. Furthermore, the specific design used for the RHS (all categories having equal area), results in a linear (integer) formulation which is also very appealing from a computational point of view.

5.7.2.2 Heuristic Approach for Rule Calibration

In this section we assume that we have an initial rule matrix, which is used as a basis for changes to improve the predictive power of the model. In the extreme case the calibration procedure can be done on an "empty" rule matrix, and in that case all the parameters of the rules are generated.

The calibration of the approximate reasoning model can be done in two phases; in the first phase the process is evaluated and problems are detected; and in the second phase an improvement procedure tries to eliminate the problems detected in the first phase, and generate missing parameters. The second phase can take place without the first, but it seems that the first phase can make the second phase computationally more

efficient.

For the first phase, an initial set of rules may be formed by techniques similar to the ones used for knowledge acquisition in *expert systems*. The experts in this case are the drivers themselves. Once an initial set of rules has been established, it can be tested, updated and expanded so that the accuracy of the model is improved. The first phase is demonstrated in Figure 20; given an initial rule matrix we implement the choice process, compare its outcome with the observed choice, and use the result of this comparison to "reward" good rules (rules that supported a correct choice) and "punish" bad rules (which supported a wrong choice). We compare each predicted choice $Z(k)$ with the observed choice $oc(k)$. Based on the outcome of this comparison we construct a weight vector $W=(W_1, \dots, W_N)$ which assigns a weight W_i to each rule i . Each rule i is being "fired" to a degree $\alpha_i(k)$ for the k^{th} observed choice. Thus rule i contributes $\alpha_i(k)$ to W_i if the k^{th} observed choice coincides with the predicted choice, and $-\alpha_i(k)$ otherwise. The weight of rule i based on all observations is then given by:

$$W_i = \frac{\sum_{j=1}^k \alpha_i(j) \delta(j) - \sum_{j=1}^k \alpha_i(j) (1 - \delta(j))}{\sum_{j=1}^k F_{Ri}(j)} \quad \text{for } i=1, \dots, N \quad (5-17)$$

$$\text{where: } \delta(k) = \begin{cases} 1 & \text{if } Z(k)=oc(k) \\ 0 & \text{if } Z(k) \neq oc(k) \end{cases}$$

$$F_{Ri}(k) = \begin{cases} 1 & \text{if } \alpha_i(k) > 0 \\ 0 & \text{otherwise} \end{cases}$$

We take the effective average (by dividing by the number of cases in which each rule was actually fired to some degree) since we may have "good" rules that are fired very rarely (such as rules dealing with special events or incidents). We expect to have all weights W_i positive and of about the same magnitude, meaning that existing values are consistently good. Rules with weights that are significantly lower than the other weights indicate some problems, possibly in the rules themselves, the relevant membership functions, or some combination of the above.

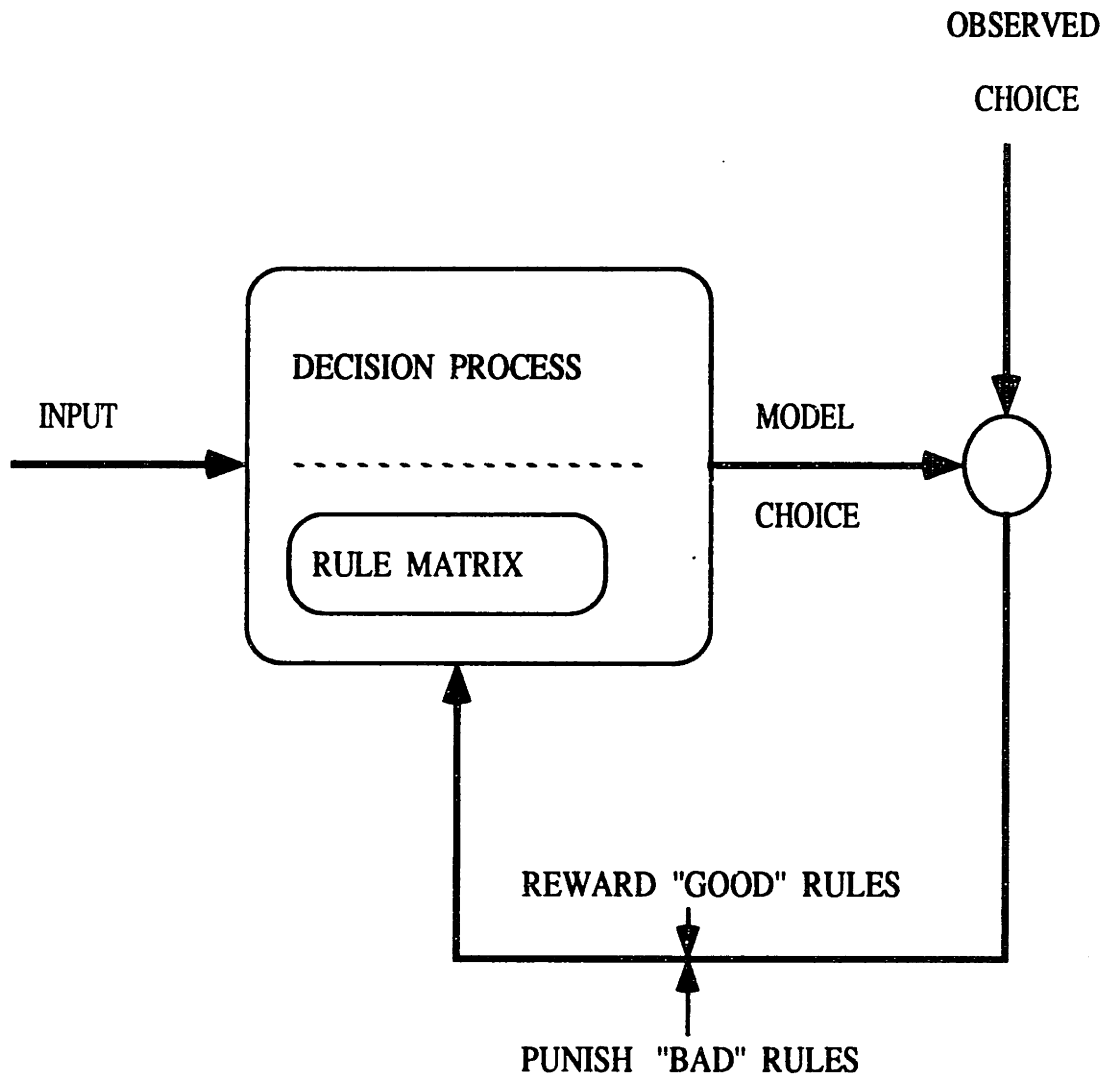


Figure 20: Rule Calibration

At the second phase, depending on the defuzzification scheme used, we try to maximize the fit between the model choices and the actual choices. For example, if we implement the ARRUM with probabilistic defuzzification scheme, we maximize the likelihood of the actual choices. For demonstration purposes let's look at a binary choice between alternatives q and r. Assuming a probit-type model, the probability that alternative q is chosen for the kth trip is given by:

$$P_q(k) = \Phi\left(\frac{Z^q(k) - Z^r(k)}{\sigma}\right) \quad (5-18)$$

where $\Phi(\)$ denotes the standardized cumulative normal distribution, and:

$$\sigma^2 = \sigma_q^2 + \sigma_r^2 - 2\sigma_{qr} \quad (5-19)$$

where σ_q^2 and σ_r^2 are the variances of the error terms associated with the utility of alternatives q and r respectively.

Thus, the likelihood function is given by:

$$L^* = \prod_{k=1}^K P_q(k)^{y_q} P_r(k)^{y_r} \quad (5-20)$$

where: $y_{jk} = \begin{cases} 1 & \text{alternative j was chosen for the k}^{\text{th}} \text{ trip} \\ 0 & \text{otherwise} \end{cases}$

The centroids $Z^j(k)$ depend on the specific design used. For example, the model which uses product correlation encoding, and the aggregation scheme described in 5.5, has a closed form for the centroids as given by equation (5-7).

The maximization is complicated by the fact that most parameters involved are discrete, and thus the likelihood function can not be differentiated. However, we can take advantage of the discreteness of the problem in the following way; the discrete parameters correspond to missing entries in the rules, and can take a **finite** (and small) number of values according to the pre-determined design. Missing LHS entries (for a specific alternative) belong to the group {VL,L,M,H,VH}, and missing RHS entries belong to the group {N,PN,I,PY,Y} for which the centroids V_i^j and the areas S_i^j are known. Thus the optimization problem is to maximize the likelihood function (or log of the likelihood function) as given by equation (5-20) subject to the constraints that exactly one possible value from the relevant group will be chosen, and that S_i^j will have a non-zero value only for the chosen V_i^j entry.

An attractive way to solve the above described optimization problem is to use a **branch-and-bound (B&B)** procedure. A B&B algorithm is appropriate for solving combinatorial problems with few integer variables, especially when their possible values are finite, since in that case the branching phase does not involve large computational requirements. Moreover, if the first phase of the calibration process is done properly, we can obtain good bounds, since the optimization treats as variables only components of rules that were identified as "problematic". An inherent difficulty of any B&B algorithm is to come up with a good branching rule. In our case, we might have some idea which values are more likely than others from the problem structure; for example if the LHS of a specific rule conveys "bad" information on the state of a certain alternative i , then the attractiveness of other alternatives $j, j \neq i$, will tend to increase. Thus, branching on more possible values might give better bounds and help in eliminating branches, and branching on less possible values might eliminate them at an earlier stage. The other inherent difficulty is the bounding phase. For our case, we can generate bounds from existing rule matrix results and from the fact that every $Z^j(k)$ is bounded from below by -1 and from above by $+1$.

CHAPTER 6: CASE STUDY

The data needed for the calibration, and refinement of the approaches presented in Chapters 4 and 5 should represent a diary for drivers and their behavior under various scenarios of traffic information availability. The various demonstration projects will hopefully provide this sort of data in the near future. However, in the absence of any such data currently, we use data from a driver simulator designed and implemented for this purpose.

6.1 Driving Simulator

There is evidence that 'real-world' behavior is related to behavior in computer-simulated environments (Clark and Smith, 1985), thus computer simulators can be used:

- to simulate real world decision making environments and to record the behavior of human subjects interacting with this simulated environment;
- to aid in calibrating models of the decision making behavior;
- to permit simulations of decision making behavior in a large variety of contexts.

Hence, driving simulators for advanced driver information systems (see for example Bonsall, 1991) provide means to collect data on driver behavior (route choice, reaction to information, etc.) which can be used together with interviews and direct observations on route choice, to calibrate models similar to the ones presented in this thesis. Driving simulators facilitate controlled experiments in a realistic environment; information, for example, can be provided on one link at a time during a simulated trip, so that the response is clearly associated with the changing condition of that link.

In order to collect the data necessary for implementing our route choice models, we have designed and implemented a two-dimensional driving simulator (see Figure 21) which has the following characteristics:

- Travel times on links are sampled from a normal distribution and are updated during the trip.
- Link congestion levels are indicated by colors according to the following categories:

FREE FLOW	→	GRAY
LIGHT TRAFFIC	→	YELLOW
USUAL TRAFFIC	→	GREEN
HEAVY TRAFFIC	→	BLUE
BUMPER TO BUMPER	→	MAGENTA

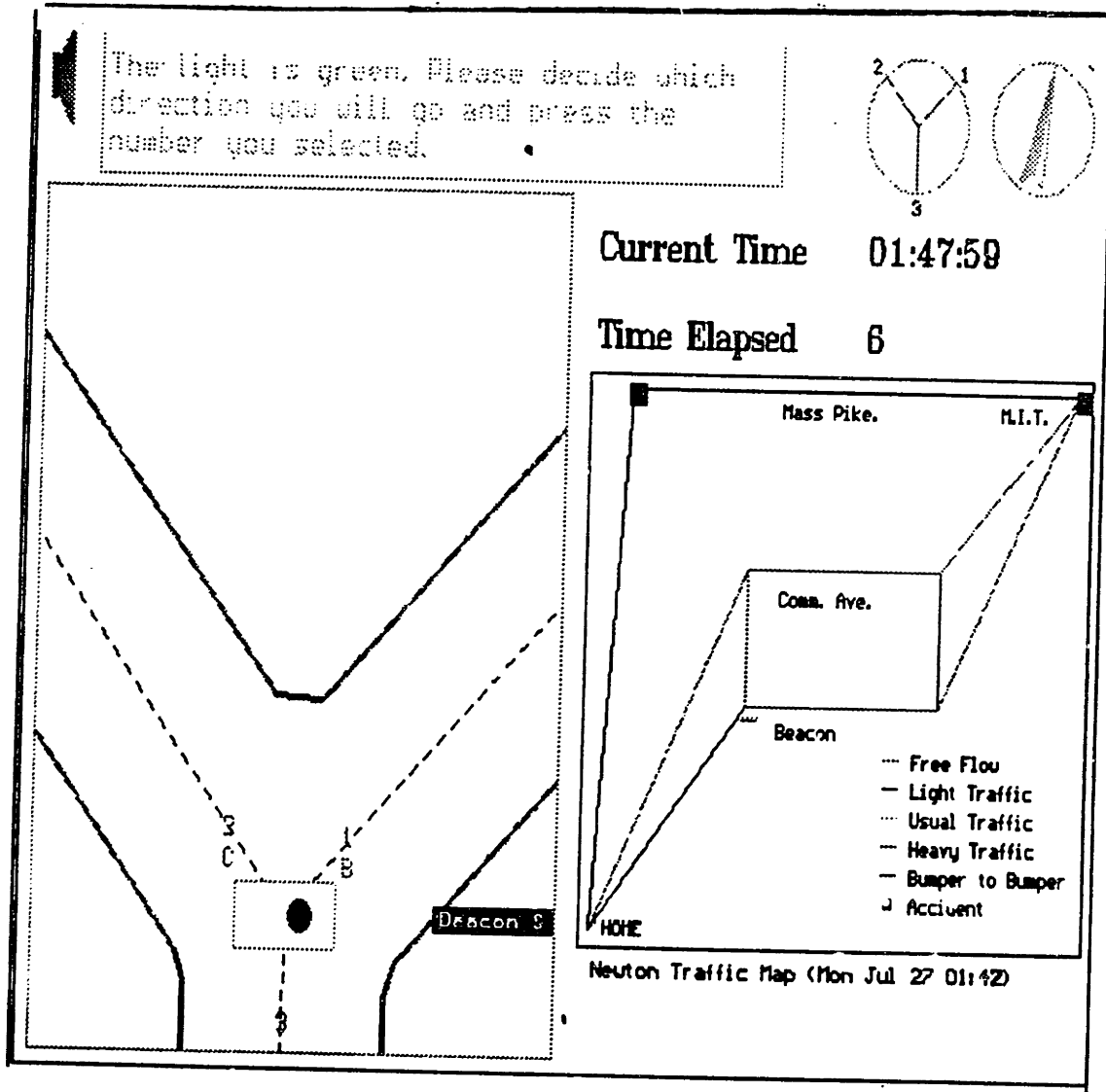


Figure 21: Driving Simulator

- Accidents appear as magenta-colored squares on nodes. Their location is randomly distributed among the nodes (excluding the origin and the destination nodes), and their starting time and duration are randomly sampled from a uniform distribution with appropriate parameters.
- The screen is composed of two major windows (see Figure 21):
 - The right window, the information window, corresponds to the provided information, and includes (other than the map), link congestion levels (indicated by link colors), and indications of accident locations.
 - The left window, the observation window, corresponds to the driving itself and is dynamic in nature; at each intersection it freezes and the driver has to make a choice. Once a choice has been made, the car starts moving along the chosen link. While "driving" the driver has to accomplish a very simple task: to keep a randomly-moving ball within the car frame without hitting its edges. This simple requirement corresponds to the driving task, which is the primary task the driver is engaged in, and forces the driver to make decisions and to process all the available information while "driving" (Sheridan, 1991b). While "driving" the driver sees his/her car moving along the chosen link, and can observe only the part of the network that includes links adjacent to the current location of the car. At the same time, on the information window, the driver can see his/her car moving along the overall network map.
- The general direction of the destination, compared to the driver's current location, is being shown continuously and is being updated according to the direction traveled.
- While traveling on a link, a sound that is associated with the congestion level, is being heard.
- Current time, and elapsed time since the trip has started are continuously shown and updated.
- Time advances in real-time at decision nodes while waiting for a decision, and faster while moving along links.
- After the completion of each trip, the subject gets general information concerning the trip; total travel time (in minutes), total decision time at nodes (in seconds), a score of how well the subject did compared to the shortest path, and a safety score which indicates how well the subject performed the driving task.

6.2 Data Collection

The Network

The experiment was conducted among drivers commuting on the "Newton Network", the commuting corridor between south-west Newton and MIT. The network, as presented to the drivers, appears in Figure 22. It includes three major alternatives:

Beacon Street,
Commonwealth Avenue, and
The Mass. Pike.

Paths that do not match one of the pre-specified 3 alternatives are considered diversions. For example: if a subject traveled from node 1 to 2 and then to 3, it is considered a diversion from Beacon to Comm. Ave. In order to simplify the network and limit the possible choices, all links are one-way directed towards MIT, except links 2↔3 and 4↔5.

Sample

Ten subjects participated in the case study. Most subjects were familiar with the Newton network, and were identified from the responses to a survey that we conducted in May 1991 (the survey that appears in Appendix C). The sample size was ten people of whom eight commute regularly from Newton to MIT, and the other two are familiar with the network.

Interview

We conducted a short interview before the subjects actually drove the simulator. The purpose of this preliminary session was to learn more about the individual choice-set of each subject, and to associate the alternatives that the subject actually knows and uses, with the given network. Thus for different subjects, the "Beacon" option can represent different but similar alternatives, which use Beacon street for some portion of the trip. In addition, we asked about the favorite alternative, about travel time perceptions (as discussed in section 3.2.1) on the three major alternatives, and the percentage of time spent on each link. A copy of the preliminary session appears in Appendix D.

Driving the Simulator

Each subject performed a total of 22 trips from the home node to MIT. The first two were considered to be practice trips, thus a total of 200 trips were performed. Traffic scenarios for the 20 trips varied randomly according to the following design parameters:

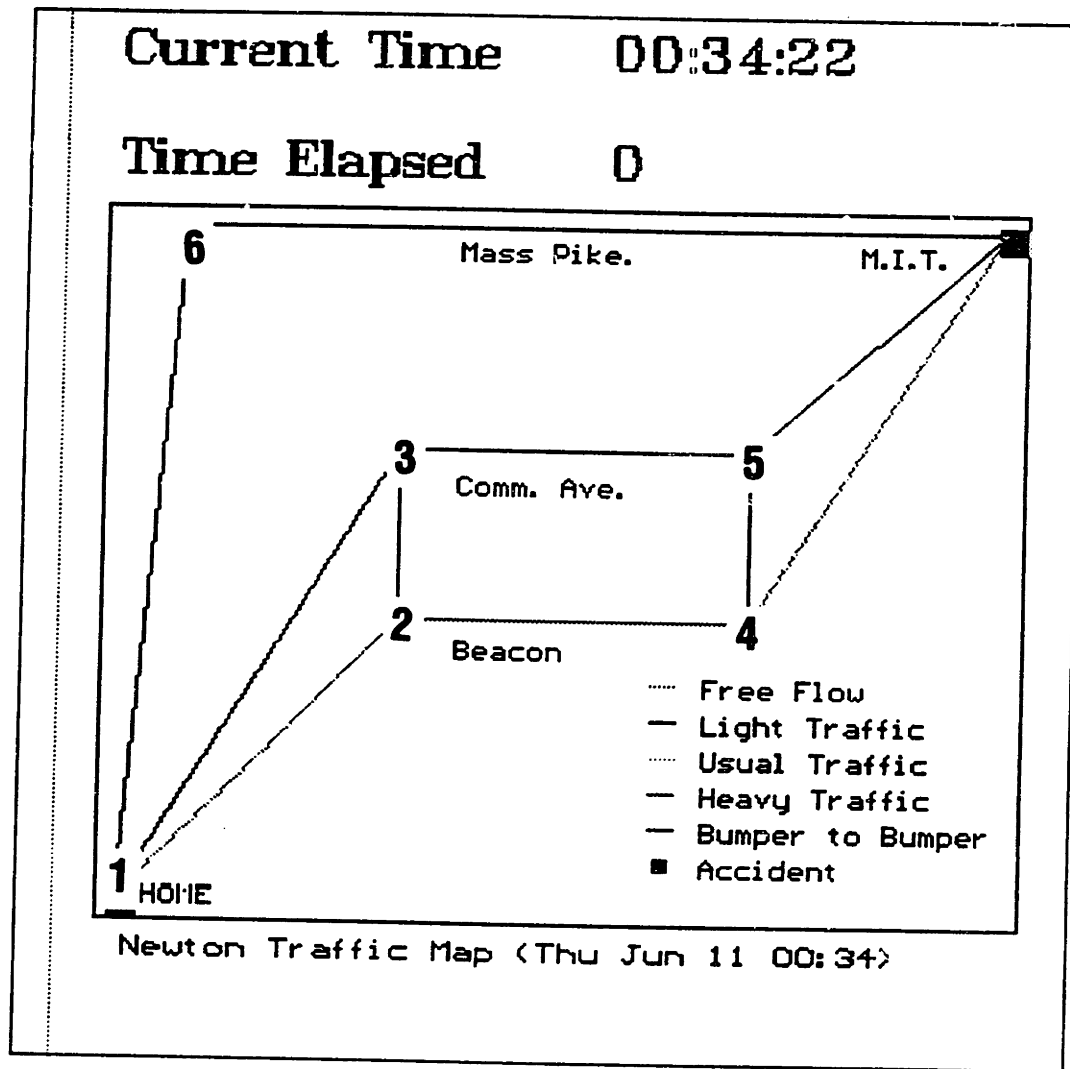


Figure 22: The Newton Network

- congestion levels
- accidents
- information availability

More specifically, in 4 trips no information was available (the information window did not appear), 11 trips had node accidents (on at most one node), and in 50% of the trips link accidents (on one link) occurred. Congestion levels appeared in all trips as link colors, and varied from 0.8*mean-travel-time to 1.8*mean-travel-time. Information was always reliable, where the reliability of the information could be inferred by the match between observed link colors (at the observation window), and provided link colors (in the information window). Overall, traffic conditions in the network were worse than typical traffic conditions in order to focus on conditions in which information is important, and route choice behavior has to be re-evaluated.

While subjects were driving the simulator, the following data were collected: at each decision point all link colors (information and observation), shown accidents, decision time, and route choices. Table 7 summarizes the choices of the participants during the experiments.

subject	favorite alternative	chosen alternative			diversions	unfinished trips
		Beacon	Comm.	Mass.		
1	Comm.	5	6	1	7	1
2	Comm.	2	16		2	
3	Mass.	4	3	11	1	1
4	Beacon	4	4	5	7	
5	Beacon	5	4		9	2
6	Beacon	10	2		8	
7	Beacon	5	10		5	
8	Beacon	9	4	3	4	
9	Mass	1		17	2	
10	Beacon	11	5	1	3	
total		56	54	38	48	4

Table 7: General Sample Information (totals)

The favorite alternative corresponds to the alternative most often chosen, as declared in the preliminary session. Unfinished trips are trips that were not completed because the subject was delayed in traffic due to an accident. Diversions are choices that do not correspond to the pre-defined 3 alternatives. They include diversions between the Beacon alternative and Comm. Ave., (no diversions are allowed once the Mass. Pike has been chosen).

6.3 The Approximate Reasoning Model Implementation

In order to implement the approximate reasoning model that was presented in Chapter 5, we need to specify all the underlying design parameters, and a method to derive the inputs from the available data. Since this case study is the first attempt at implementing the approximate reasoning model, we focussed on a relatively simple version of the algorithms involved. We focus on modeling individual behavior (disaggregate analysis) as a starting point, and then get a general feeling of the behavior of the participants as a group (aggregate analysis).

Input Variables

The inputs to the approximate reasoning models (A^* in Figure 16), are observed traffic conditions and the information available. Thus they belong to two major groups: observation and information. Another possible group of inputs is a priori perceptions. The specific scenarios used for data collection in this case study, simulated traffic conditions which are much worse than typical traffic conditions in the network, and moreover, observation was always available. Thus, it was found that the subjects, in general, followed the information and observation rather than their own perceptions since it related to situations with which they were not familiar and was always reliable.

Observation inputs are the current traffic conditions (conveyed by link colors), as can be seen from the current node in the observation window (the driving window), and thus relate to congestion levels on links adjacent to the current node. Inputs relating to the information available are of two types: congestion levels on all the links in the network (as displayed in the information window), and accidents shown in the network. The uniqueness of the input presentation in the AR model presented in Chapter 5, lies in its being symbolic and linguistic (i.e. colors that correspond to linguistic categories such as "heavy traffic", and accident indications).

We need to define the membership functions corresponding to each of these input groups. Since information that is related to travel times and congestion levels is conveyed symbolically in the simulator (by link colors and the appropriate translation), we need to model travel time perceptions according to the provided categories (free flow, light traffic, usual traffic, heavy traffic, and bumper-to-bumper).

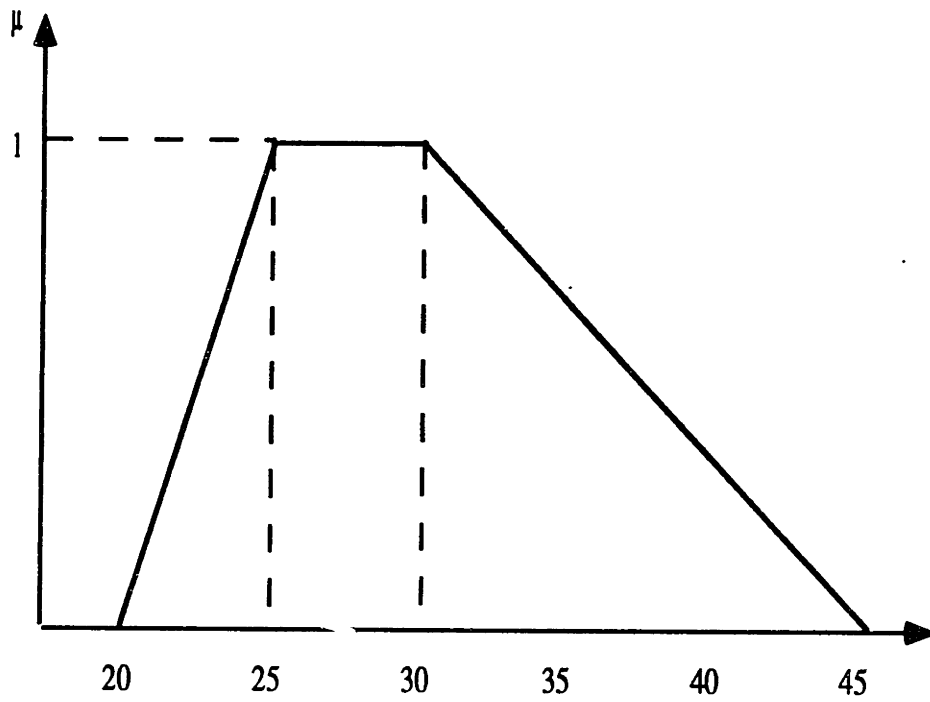
For modeling the fuzzy set corresponding to "travel times experienced along a

given path" we use, following the discussion in section 3.2.1, the responses to the questions in the preliminary session concerning the range of most possible travel times, and the minimum and maximum possible travel times. For example, the fuzzy set in Figure 23a corresponds to a person who specified that usually it takes 25-30 minutes to travel on that specific path. The shortest possible commute time is 20 minutes, and the longest is 45 minutes. As discussed in section 3.2.1, the shape of the membership function captures a lot of the individual and the facility characteristics, and these should be maintained when determining the fuzzy sets corresponding to the specific travel time categories.

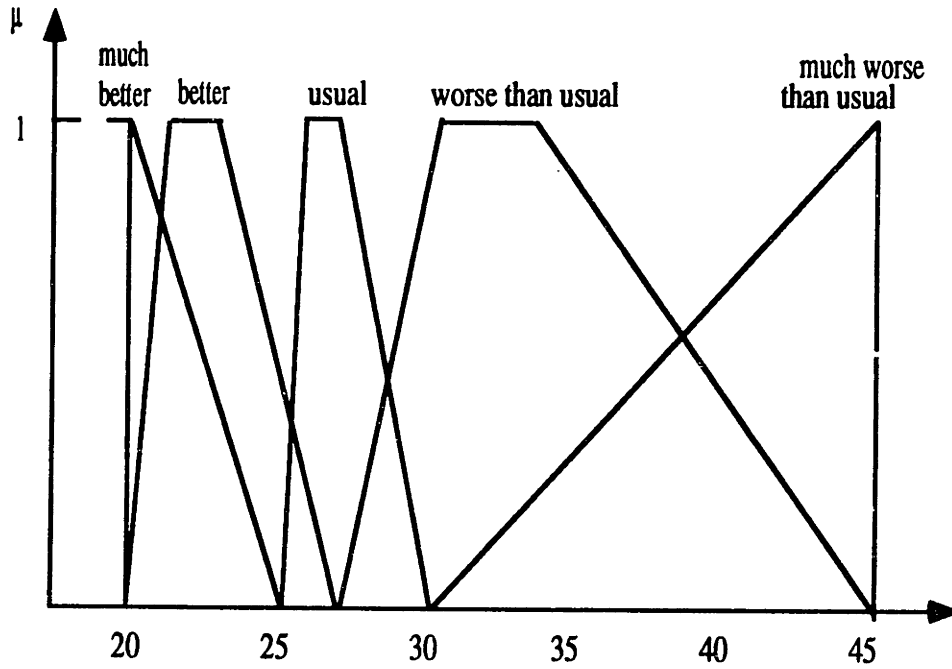
For determining the ranges and the shapes of the 5 travel time categories for each alternative, we use the following guidelines (see Figure 23b):

- The range of the fuzzy set corresponding to "usual" travel times is defined to be the range which the user stated (during the preliminary session) as the range of usual travel times, that is: $[a_2, a_3]$.
- The shape of the fuzzy sets "light traffic" and "heavy traffic" over a given range, is proportional to the shape of the T_rFN that defines "possible travel times" as appears in Figure 23a.
- The extreme sets of "free flow" and "bumper-to-bumper" do not overlap the set of "usual" travel times.
- Each travel time in the range $[a_1, a_4]$ belongs to two different sets with positive membership degree (except possibly points in which a set starts or ends).
- The set of "light traffic" and the set of "heavy traffic" meet where the flat portion of the "usual" travel times ends.
- The minimum possible travel time (a_1) belongs to the set of "free flow" with membership degree of 1, and similarly the maximum possible travel time (a_4) has full membership in the set of "bumper-to-bumper".
- The extreme sets ("free flow" and "bumper-to-bumper") are Triangular Fuzzy Numbers.

The above guidelines uniquely determine the T_rFNs corresponding to the 5 categories of travel times. Figure 23b illustrates the 5 travel times categories corresponding to the general path perception of Figure 23a. As it can be seen, the resulting design is not symmetric due to the fact that the underlying perception set (in Figure 23a) is clearly skewed towards high travel time values, which is often the case with travel time perceptions, as was explained in section 3.2.1. The suggested design preserves the general shape of Figure 23a, except for the extreme sets that are clearly biased towards the extreme values. The resulting membership functions, as shown in Figure 23b, represent a categorization of the feasible range, as stated by the driver, into travel time perceptions. It is motivated by two major principles: preserving the general shape of the declared overall path perceptions, and treating the range of travel times that occur frequently as the range of "usual" travel times. Sensitivity to this suggested design is explored in section 6.5.1.5.



(a)



(b)

Figure 23: Travel Time Categories

Given the above path perceptions (for each path in the choice set), link perceptions are derived proportionally by taking the corresponding link percentage (as reported by the subject) of the overall path. Thus for each link we have defined 5 fuzzy sets corresponding to the 5 possible colors observed or provided on that link. For deriving A^{M*} , the information provided on the Mass. Pike, for example, we add the two fuzzy sets corresponding to the given colors on the two links that compose the Mass. Pike path.

As discussed before, there are three types of inputs (observation, information on link congestion levels, and accident indication). The observation inputs rely on the observed link condition as conveyed by the link color in the observation window. For the unobserved links in the path, we assume that they have the same color, that is if the first link on Beacon street indicates "heavy traffic", it would certainly also bias the perception of Beacon street towards heavy. Furthermore, link traffic conditions for the entire network are always sampled according to the same severity (captured by the factor which multiplies the mean travel time), so in general we can expect, that if traffic is "bad" (high factor) then most of the links will have dark colors. This design captures the correlation among link travel times in the network. However, special conditions that may affect only a few links, such as accidents, are modeled separately. Hence, the input to the accident rules is binary: either there is an accident on a path, or there is not.

For the information rules, the input is given by the summation of the link conditions (as conveyed in the information window) that correspond to the various alternatives.

LHS & RHS DESIGNS

The LHS design is symmetric as was discussed in section 5.7.1 and illustrated in Figure 17a. It divides the range of possible travel times into 5 symmetric categories: Very Low, Low, Medium, High, and Very High travel times. As it was discussed in section 5.7.1, those categories serve as underlying design sets on which different perceptions can be compared. For determining the minimum and maximum values (min and max in Figure 17a), we take the lowest and the highest possible travel time on all three alternatives in the choice set. The advantage of this approach is that it facilitates sets that are O/D related, and a direct comparison among the different alternatives is possible. The RHS follows the design of Figure 17b, with equal areas for all 5 sets.

Rule Matrices

The structure of the rules follows the discussion in section 5.2.1. There are three groups of rules: observation rules, information rules, and accident rules. The structure of the observation and information rules is similar. The structure of the rule matrix follows the discussion of rule complexity in section 5.2.2, that is the LHS is one dimensional, and the RHS is one dimensional for first-level rules, and possibly multi-dimensional for second level rules. Since the LHS is one-dimensional, it determines the

size of the rule matrix. The number of observation and information rules is equal to the product of the number of alternatives by the number of categories, and the number of accident rules is equal to the number of alternatives in the choice set. For the initial rule matrix we define the rules to be the intuitive rules resulting from the trivial mapping between the 5 LHS and RHS categories:

VL	→	Y
L	→	PY
M	→	I
H	→	PN
VH	→	N

The initial rules concerning accidents are again intuitive: "if there is an accident on path j then I will not take path j". Table 8 shows the initial rule matrix for decisions made at the home node (node 1).

Decision Variables

The decision variables are the RHS entries of the rule matrix. For the first -level rule matrix, their number is equal to the number of rules, while in the second-level design, the number of the design variables is equal to the product of the number of rules by the number of alternatives. The decision variables are discrete and can take one of 5 possible outcomes (Y/PY/I/PN/N). Their initial values are determined by the trivial mapping discussed above. The only constraint imposed on the RHS values is that the preference towards taking a specific alternative satisfies weak monotonicity with respect to traffic conditions on that alternative. For example, we do not allow the following mapping (concerning the same alternative):

VL	→	PN
H	→	PY

which implies that if travel time is Very Low the alternative will probably not be chosen, whereas if travel time is High it will probably be chosen. As discussed in section 5.7.2.1, this consistency requirement is somewhat analogous to the correct sign expectation in random utility models (we expect the sign of the travel time coefficient to be negative).

Other parameters of the model can also be treated as variables (i.e. membership function values), however in this case study we only treat the RHS entries as variables, and check the sensitivity of the results to other parameter values (such as membership function values).

rule #	LHS - Beacon St	LHS-Comm. Ave.	LHS-Main Pike	RHS-Beacon St	RHS-Comm. Ave.	RHS-Main Pike
1	VL			Y		
2	L			PY		
3	M			I		
4	H			PN		
5	VH			N		
6		VL			Y	
7		L			PY	
8		M			I	
9		H			PN	
10		VH			N	
11			VL			Y
12			L			PY
13			M			I
14			H			PN
15			VH			N
16	VL			Y		
17	L			PY		
18	M			I		
19	H			PN		
20	VH			N		
21		VL			Y	
22		L			PY	
23		M			I	
24		H			PN	
25		VH			N	
26			VL			Y
27			L			PY
28			M			I
29			H			PN
30			VH			N
31	accident			N		
32		accident			N	
33			accident			N

Table 8: Initial Rule Matrix for Node 1

The Model

We implemented the specific version of the approximate reasoning model that was described in Chapter 5. For rule execution we used the correlation-product encoding scheme given by equation (B-6) in Appendix B, and for rule combination we used the weighted combination scheme of equation (5-5). For defuzzification we used the simple defuzzification scheme according to which the alternative with the highest centroid value is chosen. That is, the chosen alternative for trip k is the alternative p such that:

$$Z^p(k) = \max_j Z^j(k)$$

where $Z^j(k)$ is the centroid of the set B^j (the attractiveness of alternative j) for trip k .

6.4 Calibration Algorithm

In order to calibrate the model, we implemented a simple heuristic based on the general guidelines for calibration that were discussed in section 5.7.2.2 and illustrated in Figure 20. The calibration algorithm is described by the following steps:

0. Calculate model choices based on the given rule matrix.
1. Calculate rule weights by using equation (5-17).
2. Choose rule i with the lowest weight.
3. For the chosen rule i , try all possible RHS outcomes (Y/PY/I/PN/N), calculating the fit for each possible outcome.
4. Choose the outcome with the highest fit and update the rule matrix
5. Mark rule i as examined; stop if all rules have been examined.
6. Go to step 0.

Step 3 of the algorithm depends on whether it is being applied to first or second level rule improvements. For the first-level, only the RHS column corresponding to the alternative under consideration is checked, as for the second-level rules, all the RHS entries in the chosen rule are checked sequentially. Since this algorithm considers each rule separately, the resulting rules may violate the monotonicity requirement discussed before, and thus monotonicity should be forced on its results, as will be demonstrated in the next section.

6.5 Disaggregate Analysis

6.5.1 Decisions at the Origin

Decisions made at the origin (node 1, or HOME) correspond to choosing one of three alternatives: Beacon street, Commonwealth Avenue, and the Mass. Pike. These decisions are the most crucial for the given network. A user who chooses the Mass. Pike, for example, can not change to another alternative at all. Users who choose Beacon or Comm. Ave. can still divert, but at a higher cost. Choices at the origin can also be viewed as pre-trip decisions, and require a large amount of information processing to evaluate all the alternatives. Furthermore, for decisions made at the origin we have the largest number of observations (20 choices for each of the 10 subjects) since each subject started all the trips at the home origin. Thus choices at the origin will be thoroughly explored.

6.5.1.1 Calibration of the Rule Matrix

Tables 9 gives detailed cross tabulation of observed choices and choices predicted by the model using the initial rule matrix, and using the rule matrix resulting from the suggested first-level improvements (without monotone fit). The rows in correspond to observed choices of each of the three alternatives. The columns correspond to choices predicted by the model, where the initial fit is based on the model with the initial rule matrix (as given in Table 8), and the first-level improvement results are based on the improved rule matrix (after applying the first-level improvement algorithm to each subject separately). For example, subject #2 chose Beacon Street 3 times and Comm. Ave. 17 times. Both models (the one based on the initial rule matrix and the one after the first-level improvement) predicted correctly all the Comm. Ave. choices, and 2 out of the 3 Beacon choices (the incorrectly predicted trip was "Comm. Ave." instead of "Beacon").

model choice →	Initial Model			First-Level Improvement		
obs. choice ↓	Beacon St.	Comm. Ave.	Mass. Pk.	Beacon St.	Comm. Ave.	Mass. Pk.
sub. #1	initial fit: 30%			improved-fit: 75%		
Beacon	0	5	4	5	4	0
Comm.	0	5	5	1	9	0
Mass.	0	0	1	0	0	1
sub. #2	initial fit: 95%			improved-fit: 95%		
Beacon	2	1	0	2	1	0
Comm.	0	17	0	0	17	0
Mass.	0	0	0	0	0	0
sub. #3	initial fit: 75%			improved-fit: 80%		
Beacon	2	0	3	3	0	2
Comm.	0	2	1	0	2	1
Mass.	1	0	11	1	0	11
sub. #4	initial fit: 45%			improved-fit: 70%		
Beacon	8	0	0	8	0	0
Comm.	6	1	0	1	6	0
Mass.	5	0	0	5	0	0
sub. #5	initial fit: 65%			improved-fit: 80%		
Beacon	10	4	0	13	1	0
Comm.	3	3	0	3	3	0
Mass.	0	0	0	0	0	0

Table 9: Detailed Initial and First-Level Results

model choice →	Initial Model			First-Level Improvement		
obs. choice ↓	Beacon St.	Comm. Ave.	Mass. Pk.	Beacon St.	Comm. Ave.	Mass. Pk.
sub. #6	initial fit: 25%			improved-fit: 95%		
Beacon	3	15	0	16	0	0
Comm.	0	2	0	1	1	0
Mass.	0	0	0	0	0	0
sub. #7	initial fit: 60%			improved-fit: 90%		
Beacon	2	8	0	8	2	0
Comm.	0	10	0	0	10	0
Mass.	0	0	0	0	0	0
sub. #8	initial fit: 60%			improved-fit: 85%		
Beacon	8	4	1	13	0	0
Comm.	1	3	0	2	2	0
Mass.	0	2	1	1	0	2
sub. #9	initial fit: 85%			improved-fit: 85%		
Beacon	0	0	1	0	0	1
Comm.	0	0	2	0	0	2
Mass.	0	0	17	0	0	17
sub #10	initial fit: 80%			improved-fit: 80%		
Beacon	10	4	0	10	4	0
Comm.	0	5	0	0	5	0
Mass.	0	0	1	0	0	1

Table 9: Detailed Initial and First-Level Results (cont.)

Note that the initial fit, without calibration, estimation or optimization, is quite good (the average fit is 62%). This supports our hypothesis regarding the intuitive and simple reasoning process associated with route choice behavior: a very simple and intuitive initial rule matrix provides a good initial fit. The first-level calibration improves the average fit with a value of 83.5%. As mentioned before, the fit after applying the first-level improvement is not necessarily monotone. After forcing monotonicity (as will be demonstrated shortly), we end up with the results presented in Table 10 (with average fit of 80.5%). Note that the initial fit and its average of 62% corresponds to a **global** fit in the sense that the same rule-matrix (the initial matrix) applies to all 10 subjects, whereas the improved first-level and the resulting monotone rule matrices are individual-specific.

subject	initial fit %	monotone fit %
1	30	55
2	95	95
3	75	80
4	45	60
5	65	80
6	25	95
7	60	90
8	60	85
9	85	85
10	80	80
average	62	80.5

Table 10: Improved Monotone Fit (node 1)

In order to provide some insight on how the calibration process works, we examine more closely subject #6 (with an initial fit of 25% and an improved fit of 95%), and subject #4 (with an initial fit of 45% and an improved fit of 60%).

The changes recommended by the first-level improvement algorithm for subject #6 were as follows:

<u>rule #</u>	<u>LHS alternative</u>	<u>initial RHS entry</u>	<u>recommended RHS entry</u>
2	Beacon St;	PY	Y
3	Beacon St;	PY	Y
7	Comm. Ave;	PY	I

The recommended changes indicate a stronger preference for Beacon Street, and a slightly lower preference for Comm. Ave., and all of the changes satisfy the monotonicity requirements. A closer look at the rule weights for the initial fit and the fit after the three recommended changes, reveals that indeed they have improved significantly. For example, the initial weight of rule #2 was -0.473 (which was the second-worst rule weight), and after the changes it improved to 0.818 (the second-best rule weight). The average of the rule weights increased from -0.056 to 0.27. It is also interesting to observe improvements in rule weights of rules that were not changed. For example, rule # 12 (which says: "if Mass. Pk. is L, then PY"), initially got the worst rule weight (-0.5), however it was not recommended for change because neither the model nor observed choices picked the Mass. Pike, thus making the Mass. Pike less attractive would not improve the fit. Nevertheless, when other rules were changed, the weight of rule #12 increased to 0.9 (the best weight), since this rule now was part of correct choices.

The recommended changes for subject #4 were:

<u>rule #</u>	<u>LHS alternative</u>	<u>initial RHS entry</u>	<u>recommended RHS entry</u>
9	Comm. Ave;	PN	Y
17	Beacon St;	PY	N
21	Comm. Ave;	Y	I

The changes indicate higher preference for Comm. Ave. and lower preference to Beacon Street. However, none of them satisfies the monotonicity requirement. In order to achieve monotonicity we either change other rules (that were not recommended for change) so that the overall rule matrix satisfies the monotonicity requirement, or change the recommended rules such that they will be monotone with other rules. The best fit achieved after forcing monotonicity was 60%, and it was achieved by the following changes (compared to the initial rule matrix):

<u>rule #</u>	<u>LHS alternative</u>	<u>initial RHS entry</u>	<u>monotone RHS entry</u>
17	Beacon St;	PY	N
18	Beacon St;	I	N
19	Beacon St;	PN	N

Thus the monotone fit was achieved by adjusting the changes suggested to rule #17, and ignoring the other suggested changes. A closer look at that rule reveals that the improvement achieved due to its change was indeed the largest (from 40% to 60%), and

its weight was the lowest. Moreover, the optimal monotone fit for that subject is 60% (as will be presented in section 6.5.1.2).

Second-Level Rules

When the heuristic was applied with second-level rule improvements, the following results were obtained:

subject	initial fit %	monotone fit %
1	30	60 ↑
2	95	95
3	75	85 ↑
4	45	60
5	65	80
6	25	95
7	60	90
8	60	90
9	85	85
10	80	80
average	62	82 ↑

Table 11: Improved Monotone Fit, Second-Level (node 1)

The symbol ↑ corresponds to a result that is higher than the first-level updates, and no symbol means the same fit. Overall, the second-level rules perform slightly better than the first-level updates.

Second level improvements take advantage of their ability to model interactions among the three alternatives. For example, as was mentioned before, subject #6 had a very low weight for rule #12, but the rule did not change in the first level because it could relate only to the Mass. Pike option. However the second-level improvements suggest changing it to:

rule #	LHS - Beacon St.	LHS- Comm. Av.	LHS- Mass. Pike	RHS- Beacon St.	RHS- Comm. Av.	RHS-Mass. Pike
12			L	Y	I	PY

Indicating that even though the Mass. Pike is Low, Beacon Street is preferred, and indeed a look into the choices of subject #6, shows that the Mass. Pike was never chosen, whereas Beacon Street was chosen 18 times out of 20.

6.5.1.2 Comparison with Optimal Results

As was suggested in Chapter 5, a branch and bound algorithm has the potential to optimize the solution efficiently (find the RHS entries that maximizes the number of correctly predicted choices). For the purposes of this case study we used explicit enumeration. The results of this optimization will be used to assess the performance of the heuristic method presented in the previous section.

The optimization procedure had three groups of constraints:

1. Only first-level rules are allowed;
2. Monotonicity constraints;
3. Observation rules are equal to information rules.

The last constraint was imposed for computational reasons. Hence the performance of the generated rule matrix may not be the absolute best. However, given that the traffic conditions simulated in most of the trips are worse than usual, and that the information source is assumed reliable, we expect that observation and information rules would be similar. Therefore we do not expect the effect of this additional constraint to be of significance. The results obtained from this explicit enumeration are summarized in Table 12. The symbol \uparrow corresponds to a higher fit than the monotone fit obtained by the heuristic calibration algorithm (as was summarized in Table 10), and the symbol \downarrow corresponds to a lower fit. A lower fit (such as for subject #8) is possible because of the third constraint imposed in the optimization (i.e. identical rules for observations and information). The optimal results are not significantly better than the results obtained by the heuristic calibration algorithm and the monotone fit. This observation supports our hypothesis that the decision process is simple and intuitive, such that a trivial initial matrix and an intuitive and simple improvement procedure can yield results that are quite close to optimum.

subject	fit %
1	65 ↑
2	95
3	85 ↑
4	60
5	80
6	95
7	95 ↑
8	80 ↓
9	100 ↑
10	85 ↑
average	84

Table 12: Optimal Results

6.5.1.3 Comparison with Random Utility Model

In order to get an idea how the results presented so far compare with the widely used traditional random utility models, we estimated two random utility models in which the explanatory variables are: observation, information, and accidents. We used the multinomial logit model according to which the probability that individual n will choose alternative i is given by:

$$\text{Prob}_n(i) = \frac{e^{V_n}}{\sum_j e^{V_j}} \quad (6-1)$$

where V_i is the systematic component of the utility of alternative i for individual (or trip) n . We estimated two alternative specifications:

$$\begin{aligned} \text{model 1:} & \quad V_i = \beta_i + \beta_{\text{inf}} X_{i,\text{inf}} + \beta_{\text{acci}} \text{ACC}_i \\ \text{model 2:} & \quad V_i = \beta_i + \beta_u X_{i,u} + \beta_{\text{inf}} X_{i,\text{inf}} + \beta_{\text{acci}} \text{ACC}_i \end{aligned}$$

where $X_{i,u}$ is the travel time on alternative i , $X_{i,\text{inf}}$ is the travel time information on alternative i , and $\text{ACC}_i = 1$ if there was an accident on alternative i and 0 otherwise. In

order to obtain discrete values for the explanatory variables, we used the center of gravity of the T_FNs that represent perceptions on observation and information. The resulting fit (%) for each individual is given in Table 13.

subject	Model 1	Model 2
1	55 ↓	60
2	90 ↓	90 ↓
3	70 ↓	70 ↓
4	45 ↓	50 ↓
5	75 ↓	85 ↑
6	95	95
7	70 ↓	70 ↓
8	70 ↓	75 ↓
10	80	90 ↑
average	72.2	76.11

Table 13: Results of Random Utility Models

The symbol ↓ corresponds to results that are lower than the ones obtained by the approximate reasoning model (as appear in Table 11), and ↑ corresponds to higher values. The random utility model did not converge for subject #9. The fit of both logit specifications is generally worse than the fit obtained by the approximate reasoning model. The above comparison however is only intended to provide some preliminary evidence regarding the appropriateness of the suggested framework. Further work is necessary in order to compare it to alternative approaches and draw definite conclusions. Due to the small sample most of the coefficients are not significantly different from zero, and do not always have the expected sign (negative sign for travel time observation and information).

6.5.1.4 Prediction

In the results that we have presented so far we used the available data for both calibration of the models and predictions. A better (less biased) validation of the models can be achieved by using for prediction purposes, data that have not been used for the estimation. Although such data is not readily available, one subject of the 10 (subject # 8) did perform 60 additional trips. The prediction results for the additional trips with the initial matrix, and the matrices derived by first and second level improvements (for that

subject) are as follows:

<u>rule-matrix</u>	<u>fit of 1st session % (K=20)</u>	<u>fit of prediction % (K=60)</u>
initial rule matrix	60	60
first-level calibration	85	80
second-level calibration	90	80

For comparison, we also predicted the choices made by subject #8 using the results of the random utility model. For the initial set of data the derived coefficients for subject #8 were:

β_1	β_2	β_u	β_{inf}	β_{acc1}	β_{acc2}
1.38 (1.74)	-.287 (-.33)	-.59 (-2.08)	-.15 (-.53)	-9.79 (-.15)	-7.4 (-.136)

where β_1 and β_2 are the constants for the Beacon St. and Comm. Ave. alternatives respectively, β_u is the coefficient associated with observation of travel time, β_{inf} is the coefficient of the provided information value, and β_{acc1} is the coefficient of accident i. These coefficients were used to calculate the probabilities of choosing each alternative for the additional 60 trips (according to equation (6-1)). It was found that whereas the fit obtained by these coefficients for the initial data set was 75%, the fit for the new data set was only 65%. Thus we see, that for this case, the approximate reasoning provided a better prediction tool.

6.5.1.5 Sensitivity Analysis

In this section we check the sensitivity of the results to some of the underlying design parameters of the model.

Sensitivity to Type of Model

So far we have been implementing the simultaneous model in which all the rules are being fired to the extent with which the given input matches the LHS of the rule. Next we examine the default model for which the degree with which the rules relating to observation rules (rules 1-15 in Table 8) are fired is discounted as was explained in section 5.5. The discounting factor for a rule whose LHS deals with alternative j, is equal to $\text{Poss}[A^{j*} | A^{j**}]$ (as defined in Appendix A) where A^{j*} is the observed travel time on alternative j, and A^{j**} is the information regarding travel time on alternative j. The results of the initial fit and the monotone fit after first and second level improvements are given in Table 14.

subject	initial fit %	1st-level fit %	2nd-level fit %
1	30	55	60
2	95	95	95
3	75	80	85
4	45	60	70 ↑
5	65	80	80
6	25	95	95
7	60	90	95 ↑
8	60	85	90
9	85	85	85
10	80	80	80
average	62	80.5	83.5

Table 14: Results of Default model

The results are identical to the results of the simultaneous model with two cases that have a higher fit. The higher fit was achieved due to the different order in which rules are picked (based on their weight) by the improvement procedure, rather than the different model structure (when the rule matrices that correspond to this higher fit were applied using the simultaneous model, the results were identical). Note also that the fit for subject #7 is higher than the optimal fit of 60%, because the optimal results were achieved under the constraint that the observation and information rules are the same. This comparison shows that the simultaneous model and the default model give identical results. This could be explained by the nature of the data collected by the simulator. In the simulator, observation and information are highly correlated, since the observation on links adjacent to the current decision node coincides with the information on those links (they always have the same colors). Moreover, link congestion levels are sampled according to the same congestion factor, thus if the observed links are highly congested (e.g. bumper-to-bumper), it is also assumed that other links along the path are highly congested. Furthermore, most congestion scenarios are worse than the typical congestion levels in that network, and thus users are likely to forget their a priori perceptions and adapt to the conditions portrayed by the simulator. A more realistic implementation of the default model could assume "usual" traffic condition for the observation rules (possibly by no observation colors in the observation window), and for this kind of implementation we expect that the default model would differ from the simultaneous model.

Sensitivity to LHS Design

The results presented so far are based on the symmetric LHS design as appears in Figure 17a. Another possible LHS design is a relative design according to which the LHS categories are: much better than usual, better than usual, usual, worse than usual, and much worse than usual. We implemented the approximate reasoning model using these relative categories as LHS entries, with the corresponding membership functions derived according to the design of Figure 23b. The results of the initial fit and the monotone fit after first and second level improvements are given in Table 15.

subject	initial fit %	1st-level fit %	2nd-level fit %
1	35 ↑	50 ↓	60
2	20 ↓	90 ↓	90 ↓
3	70 ↓	75 ↓	70 ↓
4	35 ↓	50 ↓	65 ↑
5	45 ↓	50 ↓	70 ↓
6	10 ↓	90 ↓	95
7	25 ↓	80 ↓	75 ↓
8	35 ↓	75 ↓	75 ↓
9	20 ↓	90 ↑	100 ↑
10	20 ↓	70 ↓	85 ↑
average	31.5	72	78.5

Table 15: Non-symmetric LHS Design

Most of the results are significantly lower than the results obtained by the symmetric LHS design of Tables 10 and 11. The main reason for the lower fit could be the fact that for the non-symmetric LHS design more rules are being fired and to a higher average degree, but with a lower average rule weight. This reflects the fact that for the non-symmetric design the differentiation among rules decreases and thus their consequences can not dominate each other, whereas in the symmetric design less rules are being fired, and typically few rules dominate the final outcome, reflecting a more realistic decision process.

Sensitivity to Modeling of Perceptions

So far, modeling of travel time perceptions was done according to the design of Figure 23b. In this section we examine the sensitivity of the results to this design. The most characteristic feature of this design is that it is highly non-symmetric (as dictated by the stated path perceptions). We tested how the approximate reasoning model performs on a more symmetric design. The new design, demonstrated in Figure 24, has equal ranges for the three categories of: light traffic, usual traffic, and heavy traffic, and over those ranges the distribution of the T_rFN is determined proportionally to the stated relevant path perception. The design in Figure 24 for example, corresponds to a driver who declared the range [25,30] as the range of most possible travel times, and 20 and 45 minutes as the minimum and maximum possible travel times respectively. The extreme categories of free flow and bumper-to-bumper are modeled as truncated T_rFNs. The ranges of the 5 categories in Figure 24 are identical to the ranges of the symmetric LHS design as shown in Figure 17a. The results of implementing the approximate reasoning model, with the design of Figure 24 for modeling the input perceptions, are given in Table 16.

subject	initial fit %	1st-level fit %	2nd-level fit %
1	30	50 ↓	55 ↓
2	95	95	100 ↑
3	70 ↓	80	90 ↑
4	45	45 ↓	45 ↓
5	65	75 ↓	80
6	30 ↑	95	90 ↓
7	65 ↑	90	95 ↑
8	75 ↑	75 ↓	85 ↓
9	85	85	85
10	70 ↓	85 ↑	80
average	63	77.5	80.5

Table 16: Fit of Different Perception Modeling

The results do not show a clear trend of being better or worse than the initial design, and in general are not significantly different. Thus we conclude that the model is not very sensitive to the design used as long as it reflects the underlying path perceptions.

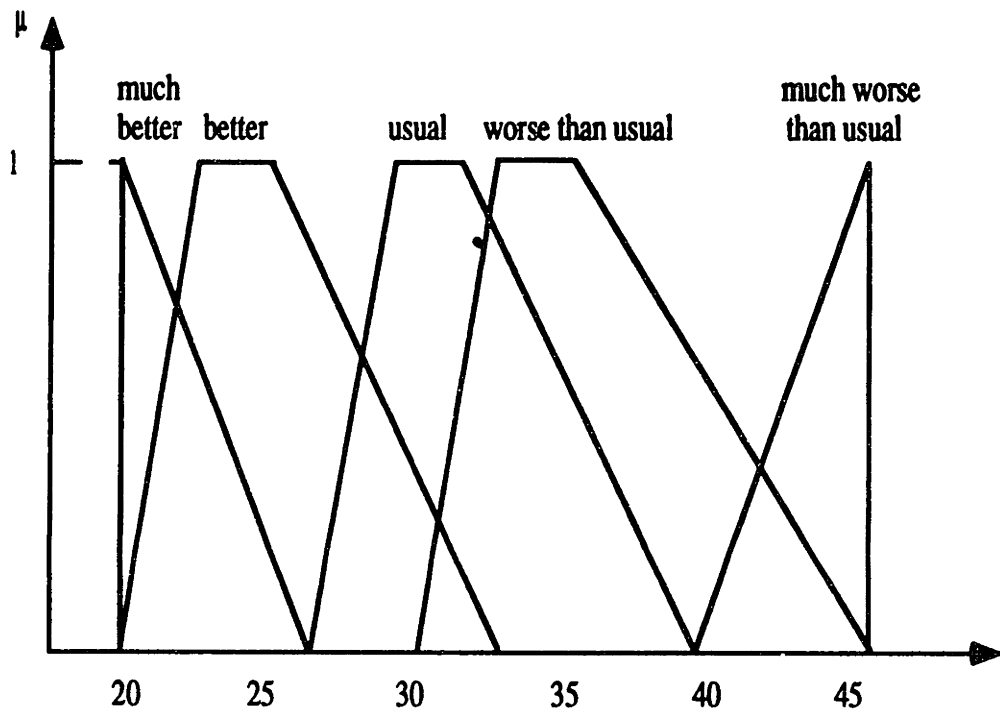


Figure 24: Travel Time Categories - Version II

6.5.2 Diversion Decisions

En-route decisions are made after the trip has started, and in general pertain to decisions to divert from the planned route due to observed or acquired information. In the Newton network (of Figure 22) diversion decisions could occur at nodes 2, 3, 4, and 5. We focus on diversion decisions at nodes 2 and 3 because the link 4→5 (which corresponds to St. Paul street) is often considered as part of the Beacon Street alternative, whereas link 2→3 is not very well known.

The total number of trips in which a diversion at node 2 occurred (continued from node 2 to 3 rather than to 4) was 13 out of 100, and the total number of diversions at node 3 (trips in which the link 3→2 was chosen) was 6 out of 76. We will model diversion decisions at node 2 because of the 6 diversions at node 3, 1 was a mistake, and 4 were triggered by an accident at node 5.

A decision whether or not to divert is a binary decision. Furthermore, the two alternatives are complementary (if diversion is the attractive alternative then no-diversion automatically becomes unfavorable). Thus, the RHS of the rules is one-dimensional, with attractiveness of the diversion alternative modeled by the following sets:

D	-	Divert
PD	-	Probably Divert
I	-	Indifferent
PND	-	Probably Not Divert
ND	-	Not Divert

The design of the fuzzy sets that correspond to the above categories is similar to the RHS design of Figure 17b, and it is illustrated in Figure 25.

The LHS is two-dimensional and corresponds to traffic conditions on the pre-planned path and on the diversion alternative. There are again three groups of rules: observation, information, and accidents. The initial rule matrix corresponds to the trivial mapping (if the pre-planned route is good - do not divert, if the alternative is good - divert) shown in Table 17.

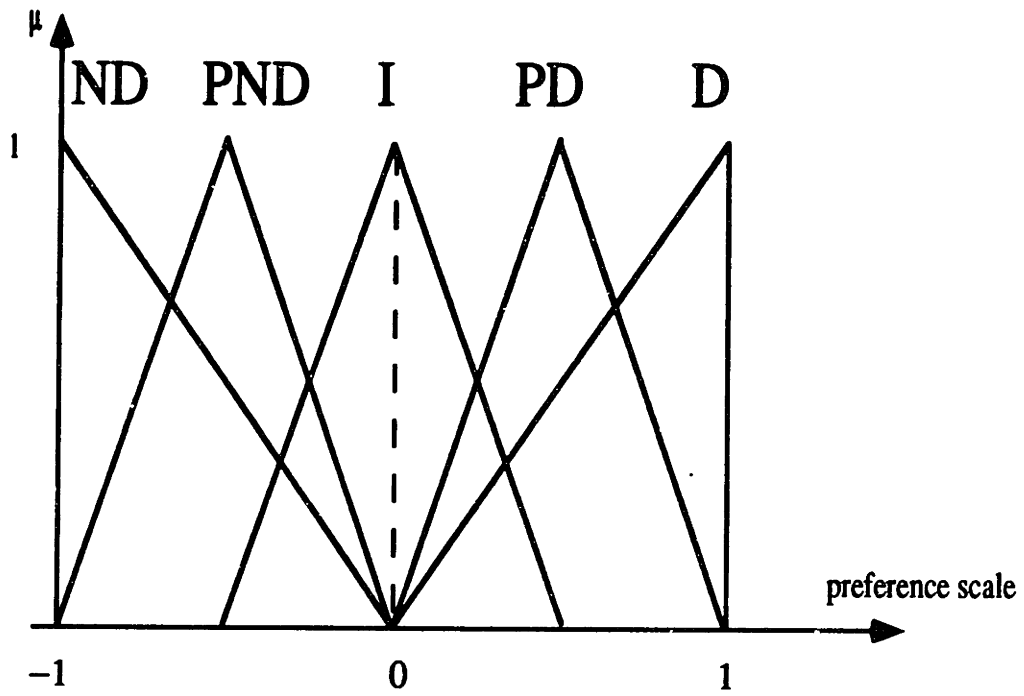


Figure 25: Diversion Categories

rule #	pre-planned alternative	diversion alternative	RHS
1	VL		ND
2	L		PND
3	M		I
4	H		PD
5	VH		D
6		VL	D
7		L	PD
8		M	I
9		H	PND
10		VH	ND
11	VL		ND
12	L		PND
13	M		I
14	H		PD
15	VH		D
16		VL	D
17		L	PD
18		M	I
19		H	PND
20		VH	ND
21	accident		D
22		accident	ND

Table 17: Initial Rule Matrix for Node 2

Note that the initial rule matrix gives equal preference to the pre-planned and to the diversion alternatives, whereas in reality we expect drivers to have a stronger preference for the pre-planned alternative. We expect the calibration algorithm to detect this trend (if it really exists).

6.5.2.1 Calibration of Rule Matrix

The results of applying the first-level improvement algorithm are given in Table 18 (note that for binary diversion decisions there are no second-level rules).

subject	initial fit %	improved fit %	monotone fit %	# of trips	# of diversions	diversions correctly predicted %
1	91.66	91.66	91.66	12	2	50
2	100	100	100	3	1	100
3	100	100	100	5	-	
4	77.77	77.77	77.77	9	2	0
5	85.7	100	100	14	2	0
6	88.88	88.88	88.88	18	3	33.3
7	90	100	100	10	2	50
8	84.6	92.3	92.3	13	1	0
9	0	100	50	2	-	
10	92.85	100	100	14	1	0
weighted average:	88	94	93	$\Sigma = 100$		

Table 18: Diversion Decisions at node 2

Note that only 100 out the 200 trips passed through node 2. The rule matrix, calibrated by applying the same heuristic approach, provides a very good fit between predicted and actual choices. However, due to the small number of diversions, the success in predicting diversions is only average.

Some of the changes observed between the initial and the calibrated rule matrix are:

<u>Subject #</u>	<u>rule #</u>	<u>existing entry</u>	<u>recommended entry</u>
7	18	I	ND
8	14	PD	ND
9	3	I	ND

The nature of the few recommended changes is to give higher preference to the "Not-Divert" option, which supports our a priori expectation that drivers favor their pre-planned

alternative. But even without these changes, the initial rule matrix provided a very good fit. This is explained by the input values to the rules, which are higher when the divert option is exercised, causing rules which deal with good traffic conditions on the diversion option to fire only to a low degree.

6.5.2.2 Comparison with Optimal Results

The enumeration-based optimization with 10 decision variables (again assuming monotonicity and equal rules for the observation and the information groups) yields the following results:

subject	optimal fit %
1	91.6
2	100
3	100
4	88.9
5	92.8
6	94.4
7	100
8	100
9	100
10	100
weighted average:	96

Table 19: Optimal Results - Node 2

These results present a fit that is not significantly better than the fit obtained by the improvement procedure (as presented in Table 18), and supports again the good performance of the improvement procedure.

6.5.2.3 Prediction

We used the data available for subject #8 who performed 60 additional trips, of which 32 passed through node 2, for prediction purposes. For those additional 32 trips, the initial rule matrix resulted in a fit of 90.6%, the improved rule matrix resulted in a fit of

93.75%, and the optimal matrix for subject #8 resulted in a fit of 96.9%. These prediction results are very good, and in fact are even slightly better than the results for the initial set of data from which the improved rule matrices were derived.

6.6 Aggregate Analysis

In aggregate analysis we try to fit the same rule matrix to all subjects. The underlying assumption is that there exists a global rule matrix for all subjects and that individual differences are accounted for through the different perceptions which serve as inputs to the rules. In the aggregate analysis we should keep in mind that even though most of the subjects are familiar with the network, and live in that area, they do not live at the same origin, and thus their choice sets are not identical. For example, for some the option of the Mass. Pike is not realistic, whereas for others the Mass. Pike is the most favorable alternative. However, we expect the different input perceptions to induce higher costs for alternatives that are not really feasible, and thus to cause rules which favor those alternatives not to fire (or to fire to a low extent). The initial rule matrices presented in Tables 8 and 17 provide examples of global rule matrices since the same matrices are being applied to all 10 subjects.

6.6.1 Decisions at the origin

At the origin (node 1) subjects chose one of three alternatives: Beacon Street, Comm. Ave., or the Mass. Pike. Although these alternatives are not exactly the same for all 10 subjects, we treated them as the same in the analysis that follows.

6.6.1.1 Calibration of Rule Matrix

The initial fit using the rule matrix of Table 8 was 62%. The monotone global fit resulting from first-level improvements was 72.5%, and included the following changes to the initial rule matrix:

<u>rule #</u>	<u>initial RHS entry</u>	<u>monotone RHS entry</u>	<u>relevant path</u>
1	Y	PY	Beacon
3	I	PY	Beacon
5	N	PN	Beacon
6	Y	PY	Comm. Ave.
21	Y	PY	Comm. Ave.
29	I	PN	Mass. Pike

The nature of the changes indicates a stronger preference for Beacon Street, and somewhat lower preferences to Comm. Ave, and the Mass. Pike.

The monotone rule matrix resulting from second-level improvements is more complex (compared with the initial rule matrix) and it is given in Table 20.

rule #	LHS - Beacon St.	LHS-Comm. Ave.	LHS-Mass. Pike	RHS-Beacon St.	RHS-Comm. Ave.	RHS-Mass. Pike
1	VL			Y	I	
2	L			PY		
3	M			I		Y
4	H			PN		
5	VH			PN	Y	
6		VL			Y	
7		L			PY	
8		M			I	
9		H			PN	
10		VH		N	N	
11			VL			Y
12			L			PY
13			M		Y	I
14			H	PY	PN	PN
15			VH		PY	N
16	VL			Y		PN
17	L			PY		
18	M			I		
19	H			PN		
20	VH			N	Y	
21		VL			Y	
22		L			PY	
23		M			I	PN
24		H		Y	PN	
25		VH			N	
26			VL			Y
27			L			PY
28			M			I
29			H	PY		PN
30			VH	PN	Y	N
31	accident			N		
32		accident		PY	N	
33			accident	Y		N

Table 20: Global Rule Matrix for Node 1

The rule matrix given in Table 20 results in a global fit of 75.5%. It is interesting to note that the most important changes correspond to the addition of interactions among alternatives rather than changing existing entries (as done in the first-level updates). The interactions include:

- if Beacon Street is "bad" then Comm. Ave. becomes more attractive (rules 5, 20),
- if the Mass. Pike is "bad" then Comm. Ave. becomes more attractive (rules 13, 15, 30).

A better understanding of the importance of specific rules can be gained by looking at their weights as calculated by equation (5-17). Figure 26 illustrates the rule weights calculated for the rules in Table 20. As expected, the three accident rules (rules #31, 32, and 33) are good rules and have the highest weights. Rules # 32 and 33 even get additional entries such that Beacon Street is the favorable alternative when there is an accident on Comm. Ave. or the Mass. Pike. The next 6 highest weights correspond to the rules which map "L" to "PY" (rules #27, 22, 12, 17, 7, and 2). These rules not only have very high weights, but they also maintain their original format. The dual rules (with symmetric membership functions) which map "H" to "PN" do not get such high weights and are often recommended to be changed. This reveals a reasoning logic according to which if an alternative is "quite good" it is likely to be chosen, whereas if an alternative is "quite bad" it does not necessarily mean that it will not to be chosen. This supports our expectation that a "usual" or "favorable" path may be chosen even if its congestion level is quite bad. Additional support for this is the fact that the group of rules which have the lowest weights is the group which maps "VH" to "N". This trend supports our initial hypothesis that the reasoning behind the choice process is not linear; good traffic conditions indicate high attractiveness, but poor conditions do not necessarily indicate similar-in-magnitude low attractiveness.

6.6.1.2 Comparison with Optimal Results

As we did in the disaggregate analysis, we used explicit enumeration to calculate the optimal fit in order to evaluate the solution obtained by the improvement procedure. The enumeration had the same three groups of constraints that were described in section 6.5.1.2. The optimal fit achieved was 71.5% which is lower than the fit achieved by the improvement procedure (which was 72.5 for first-level rules). The lower fit resulted from forcing the observation rules and the information rules to be identical. This constraint has affected the results only in the aggregate analysis, implying that in order to use one rule matrix for several subjects we need the flexibility provided by being able to differentiate between the observation and the information rules.

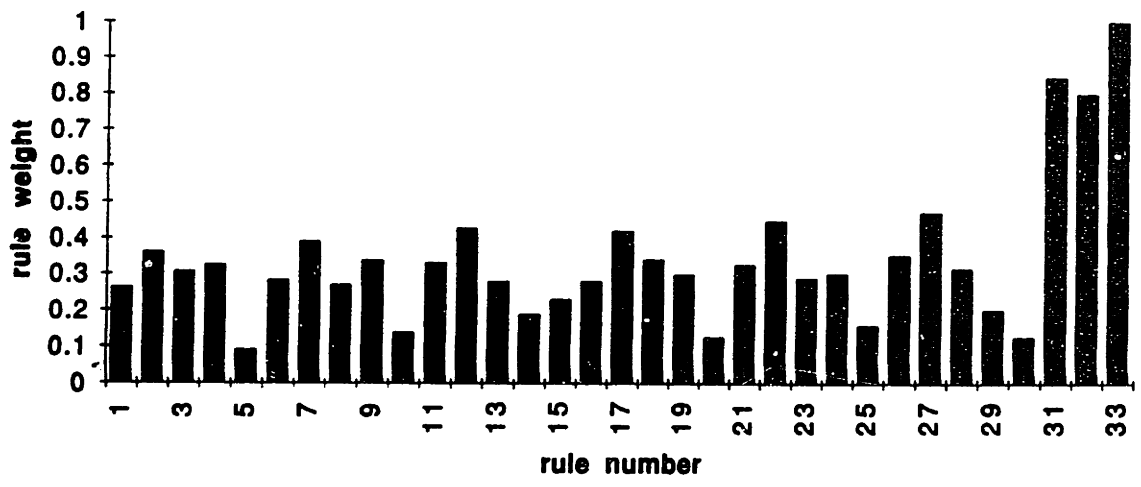


Figure 26: Rule Weights

6.6.1.3 Comparison with Random Utility Model

We estimated the two random utility models that were presented in section 6.5.1.3 using all the 200 observations with the following results:

	β_1	β_2	β_u	β_{inf}	β_{acc1}	β_{acc2}	β_{acc3}
model 1	.86 (3.8)	.466 (1.99)		-.18 (-6.66)	-10.66 (-.22)	-1.73 (-1.61)	-8.48 (-.13)
model 2	.86 (3.8)	.44 (1.87)	-.022 (-.46)	-.16 (-3.25)	-10.67 (-.22)	-1.73 (-1.61)	-8.48 (-.13)

where β_1 and β_2 are the constants for Beacon St. and Comm. Ave. respectively, β_u is the coefficient associated with observation of travel time, β_{inf} is the coefficient of the provided information value, and β_{acci} is the coefficient of accident i .

	fit %	L(0)	L(β)	ρ^2
model 1	72.5	-219.72	-164.04	.226
model 2	73.0	-219.72	-193.94	.085

The fit obtained by the random utility models is comparable to the fit obtained by the approximate reasoning model with first level improvement (72.5%), but is slightly lower than the fit obtained by second-level improvements (75.5%). It is also be remembered that the approximate reasoning results have been calibrated based on a heuristic approach, whereas the results from the random utility model are based on global optimization. Thus we can conclude that the approximate reasoning model provides a better fit. Based on our observation about the non-linearity of travel time (that was explored in section 6.6.1.1), we can recommend a possible improvement to RUM by treating travel time perceptions in categories rather than as a continuous variable, thus allowing low travel times and high travel times to have non-linear effects on the utility. However, a more systematic comparison of the models suggested in this thesis and other more traditional approaches should be the subject of further research.

6.6.2 Diversion Decisions

The results obtained for diversion decisions using the initial rule matrix (as given by Table 17) provide a global fit of 88%, which is a very good initial fit. The improvement procedure does not provide any improvement to that fit, and the global optimization provides a fit of 89%. Thus we can conclude that the initial rule matrix

provides a very good global fit.

6.7 Discussion

The driving simulator turned out to be a very good tool to collect data for testing and calibration of route choice models. The subjects who used the driving simulator were entertained on one hand, and on the other hand could associate the trips made in the simulator with their usual commuting behavior. The full potential of the simulator has not been explored in this case study, and future use of it can include other scenarios such as information on recommended paths, and provision of information from changeable message signs. Another group of possible scenarios relates to the reliability of the provided information, since in the current version information was always reliable (as reflected by the same link colors in the observation and the information windows).

When analyzing the results of this case study we have to keep in mind some of the inherent biases that are related to the sample used and the data collection procedure. Although all subjects were chosen such that they were familiar with the Newton network, they lived in different neighborhoods at the origin node, and thus their choice sets were not identical. This fact is important when evaluating the aggregate models. Another bias relates to the traffic conditions sampled. In order to study drivers reactions to information, the traffic scenarios that were presented to the subjects corresponded to worse-than-usual traffic conditions, because we assumed that under usual traffic conditions, drivers follow their usual routes. Thus, subjects had a clear bias to trust the information, not only because it was always reliable, but also because it conveyed information on traffic conditions with which they were not familiar (the Newton network is typically not congested). Therefore the role of a priori perceptions was downplayed. As a result the simultaneous and the default models (which is based on a priori perceptions), performed in a similar way.

Keeping those biases in mind, the results obtained are very supportive of the concepts that motivated the development of the approximate reasoning model. Furthermore, the favorable comparison with the well-established random utility model lends credence to the approximate reasoning model.

An important finding was the good fit that was achieved by the *initial* rule matrices, which correspond to the most trivial reasoning (i.e. "good" traffic conditions are attractive, and "bad" traffic conditions are not attractive). It supports our hypothesis that the final choice is composed of many *simple* considerations.

The results obtained also support our hypothesis regarding the non-linearity of the reasoning process. It was found out that the mapping between congestion levels and attractiveness of an alternative is non-linear: good congestion levels indicate high attractiveness, but poor congestion levels do not necessarily imply a similar in magnitude reduction in attractiveness.

The importance of interactions among alternatives was also explored, and it was found that only few interactions exist, and typically they relate to the attractiveness of alternative j when traffic conditions on alternative i are very bad ($i \neq j$).

Finally, we would like to comment here on the highly heuristic nature of the improvement procedure used and the way monotonicity is forced. Thus occasionally we obtained contradictory results which depend on the order in which rules were picked. Still, the very simple heuristic that we used, performed well and resulted in solutions that were quite close to the optimal solutions. This again strengthens our hypothesis about the intuitive nature and simplicity of the overall choice process. However, there is a strong need for a more rigorous and efficient optimization algorithm (such as implementation of the ILP formulation suggested in section 5.7.2.1, or development of the branch-and-bound algorithm as suggested in section 5.7.2.2) in order to obtain optimal results.

CHAPTER 7: SUMMARY AND CONCLUDING REMARKS

7.1 Summary

The objective of the thesis was to develop new approaches for modeling route choice behavior in the presence of information. We proposed three hypotheses regarding the way drivers react and interact with traffic information systems:

- The simultaneous model, in which drivers simultaneously consider all the factors that affect their route choice (including traffic information).
- The two-stage model according to which the drivers first update their perceptions using the available information, and at the second stage make a route choice decision based on their updated perceptions.
- The default model, according to which default behavior exists, and is followed unless there is a good reason to divert from it. Information is considered in light of this default behavior, and plays an active role in the decision process only if it differs from usual or expected conditions.

Models for route choice behavior, for all three hypotheses, can be characterized by two main components: perceptions of the system's attributes, and a decision mechanism based on given perceptions.

A new modeling approach for modeling perceptions and information was introduced. It uses concepts from fuzzy set theory for modeling aspects relating to vagueness of perceptions, and to the linguistic and symbolic nature of perceptions and information. Fuzzy sets were used to model various aspects of travel time perceptions. In Chapter 3 we used fuzzy sets to model the degree of belief with which each travel time value is thought to be possible (along a certain facility). Later, in Chapters 5 and 6, we used fuzzy sets to model the underlying travel time concepts and perceptions. On an absolute scale, we modeled travel time concepts according to the categories: Very Low, Low, Medium, High, and Very High. These categories were then used to represent the multi-dimensionality of human perceptions, by utilizing the idea of partial membership in a set. Consequently, a certain travel time value could belong with different degrees to several of the given travel time categories. We also suggested modeling travel time on a subjective scale according to the categories: Much Better than Usual, Better than Usual, Usual, Worse than Usual, and Much Worse than Usual. These categories were then used to model the inputs to the approximate reasoning model. For modeling information, we focused on modeling information related to travel time, and thus used the same ideas described before. We also suggested an approach for integrating information into current perceptions, when information and existing knowledge are modeled on the same scale.

In Chapters 4 and 5 we developed models for decision processes that can incorporate fuzzy perceptions and linguistic information as inputs. Chapter 4 adapts models based on the principles of shortest path, and utility maximization to handle fuzzy perceptions as inputs. We derived a new method for comparing fuzzy numbers that recognizes the specific interpretation of the fuzzy sets involved, is transitive, and provides consistent results. This comparison method is used for solving the shortest path problem with fuzzy costs on links. We expanded the fuzzy linear regression model with fuzzy coefficients (and crisp inputs) to discrete choice model with fuzzy utilities associated with each alternative. We also developed a discrete choice model for fuzzy utilities with fuzzy inputs as explanatory variables, and random error term, and formulated the corresponding maximum likelihood function.

A new approach for modeling the route choice decision process, the approximate reasoning model, is presented in Chapter 5. It is based on decision rules, but unlike traditional rule-based systems, it offers flexible rule interpretation. This is achieved both by using rules that relate to linguistic terms, and by allowing rule premises to be partially true. The approximate reasoning mechanism is used to derive updated rule consequences, which are added simultaneously into a combined attractiveness of each alternative. The final choice is then presented by choosing the most attractive alternative in the choice set either deterministically or in a random utility framework. The underlying behavioral assumption of the model is that the route choice process can be viewed as a combination of many simple considerations (rules) that are applied together, each contributing to a different aspect of the final decision.

We showed how the three hypotheses can be implemented using various models; the shortest path model for implementing the two-stage model using default reasoning (as suggested in section 3.4.2) for the first stage; the random utility model with fuzzy attributes, and the approximate reasoning model, can be applied for the three hypotheses. In the case study we implemented the simultaneous and the default models. Due to the specific nature of the data collected, the two models provided very similar results. We suspect that when data that corresponds to commuting behavior under more "typical" conditions (rather than "worse than usual" scenarios) is available, the default model would out-perform the simultaneous model.

For collecting data to test, validate, and calibrate models for route choice behavior (such as the models presented in this thesis), two approaches were used. The first belongs to the revealed preferences approach, and was based on conducting a commuter survey, which corresponded to specific commuting behavior in the presence of radio traffic reports. The second belongs to the stated preference approach and included design and implementation of a driving simulator that collects data under various information provision scenarios. It simulates pre-trip and en-route decisions while interacting with a traffic information system. In the simulator the driver has to perform a simple task corresponding to actual driving, while advancing along the network map from a given origin to a pre-defined destination. Information is provided by colors and sounds that

correspond to different congestion levels on links, accident indication, shortest path indication, and changeable message signs. Current time and duration of the trip are continuously displayed. The simulator collects all the available data that was displayed to the driver, and the corresponding decisions at intersections.

The driving simulator was used to collect data for implementation of the approximate reasoning model. The small case study presented in Chapter 6 shows that the approximate reasoning model has potential, and supports our hypotheses regarding the intuitive and simple nature of the decision process. The underlying simplicity of human reasoning concerning route choice was demonstrated in the case study by the fact the initial rule matrices, based on trivial reasoning provided a very reasonable fit. This agrees with our prior expectation that: "a man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity of the environment in which he finds himself" (Simon, 1969). Additional support for the underlying simplicity, and thus the appropriateness of the approximate reasoning model, was obtained from the good performance of the simple improvement heuristic used. The case study also supported our prior expectation regarding the non-linearity of human thinking. The results obtained in the case study show that the mapping from travel time perceptions to the attractiveness of the alternatives is non-linear.

With respect to the three hypotheses suggested for route choice behavior in the presence of information (the simultaneous, the two-stage, and the default models), we can not, at this point, determine which model is more appropriate for a given scenario.

7.2 Contributions

The thesis provides a new direction for modeling route choice behavior in general, and in the presence of information in particular. It is based on the premise that human perceptions are more naturally modeled using linguistic terms, and that human behavior can be more realistically modeled using flexible linguistic rules.

In Chapter 1 the route choice process was presented as a black box through which the inputs are filtered, resulting in a choice. Two kinds of filters were mentioned: a perception filter through which knowledge is processed, and an evaluation filter through which perceptions are transformed into a choice. Route choice models are categorized according to the degree of detail with which they explain the interior of the black box. The contribution of this thesis to the explanation of the black box embraces both dimensions: perception and evaluation. For modeling perceptions, we proposed models based on fuzzy set theory, which have the potential to capture linguistic perceptions, as well as weak forms of uncertainty, (and individual and system characteristics). For the decision process, we proposed the ARRUM model, which is a behavioral model based on (fuzzy) linguistic rules that are processed simultaneously in a flexible and adaptable fashion.

We also suggested three hypotheses on general route choice behavior in the presence of information (the simultaneous, the two-stage, and the default models). We implemented the simultaneous and the default models in a case study. However, due to the nature of the data collected, conclusions about the relative performance of the hypotheses are limited at this point.

From the modeling point of view, we extended the shortest path algorithm to be able to handle inputs which are modeled by fuzzy sets. A contribution to the theory of fuzzy sets is the derivation of a consistent comparison method of fuzzy numbers. The comparison suggested takes into account the shape of the fuzzy numbers compared, is transitive, and can be used for ranking. We also introduced fuzziness into random utility models. The linear regression model with fuzzy coefficients was extended to a discrete choice model, which estimates the fuzzy coefficients of the utilities. Furthermore, we developed a utility model with fuzzy attributes and an error term which is more appropriate for the problem addressed in this thesis.

In this thesis we used the fuzzy control framework for modeling human behavior rather than machine behavior. Hence our models are characterized by having fuzzy inputs feeding the decision rules (and not exact numerical values). This distinguishes human oriented systems from machine and processes control, (the different perception inputs trigger the appropriate rules). When inputs are fuzzy, more rules are typically fired, which corresponds to more complex decision processes. Hence we suggested modeling the complexity of human decision processes as a combination of many simple processes. From a theoretical point of view, we expanded the fuzzy control framework to deal with multi-dimensional outputs in which all dimensions relate to the same decision. Consequently we developed defuzzification schemes which result in discrete choice. A defuzzification scheme that is especially attractive for the route choice problem, is the random utility defuzzification, which adds a random noise term to the attractiveness of each alternative, to account for the inherent randomness of the decision process.

Most rule-based systems assume that the rules are given, either by extraction from experts knowledge, or by utilizing properties of the processes under control. However, for human behavior modeling, each person is an expert, and there exists an infinite number of possible rules. Thus we suggested consistent methods for rule generation and calibration. The methods suggested vary from a rigorous optimization (formulated as an ILP problem), to a simple and intuitive heuristic which seems to perform well.

7.3 Directions for Future Research

The research presented in this thesis could be extended in various directions. In general, the modeling approaches proposed in this thesis need to be tested, validated and calibrated using (preferably real) data in order to make them applicable and useful. In this section we briefly discuss directions for future research.

Relaxation of assumptions

Throughout the thesis we focused on route choice decisions in isolated decision points (intersections). However, a route choice decision can be viewed as a sequence of decisions at all decision points that exist from the origin to the destination. Choices at two successive decision points along a path could be correlated. It would be interesting to see whether the same (or slightly modified) rule matrix can be used for such decisions.

We also assumed that at each decision point the choice set is given. It seems that in the dynamic environment created by the existence of on-line information, a more flexible determination of choice sets is required. The approximate reasoning model might provide such flexibility by adjusting (automatically) the appropriate rules for new alternatives suggested by the information system.

Travel time and traffic information were assumed to be the most important factors that affect route choices. This assumption could be easily relaxed to incorporate other factors, and empirical work is necessary to examine the importance and role of other factors.

Empirical work and testing

The three hypotheses suggested for route choice behavior in the presence of information (the simultaneous, the two-stage, and the default models), should be further evaluated and tested. We expect that for different scenarios of knowledge and experience - different hypothesis would be appropriate. We speculate that for the current state of the art regarding information systems, the default model is more appropriate since drivers are not yet familiar with the new information systems, and thus have their typical behavior pattern which is subject to changes only if a good enough reason for change exists. When information systems become more widely used, it is likely that the two-stage or the simultaneous models, which incorporate information at a more basic level would be appropriate.

In this thesis we expanded the shortest path algorithm to handle fuzzy costs on links using a consistent comparison method that we have developed. Although shortest path behavior is not realistic, it could be that "fuzzy" shortest path behavior is more realistic. Thus an implementation of the shortest path algorithm in networks with fuzzy costs could be attractive. Furthermore, costs on links could correspond to updated perceptions (after utilizing the default reasoning update mechanism of section 3.4.2), and thus implementation of the shortest path would correspond to application of the two-stage model.

An interesting application would be the implementation of the random utility model with fuzzy attributes and random error term (as was suggested in section 4.2.3). It is of special interest to implement the fuzzy RUM since its comparison to traditional

RUM is very straightforward. Estimation of the weights w_1 and w_h is also of importance both from estimation point of view, and for providing insight into drivers' attitudes towards risk.

Several approaches for modeling route choice behavior in the presence of information were introduced in this thesis. It would be interesting to compare the various approaches on the same set of data according to a common performance measure (such as percent of correctly predicted choices). It would also be interesting to compare performance with other traditional route choice models (such as random utility model). The comparison between the approximate reasoning model and the random utility model provides such an example, although a more rigorous comparison is necessary. Based on the results of such comparisons, conclusions on how to improve the various models could be drawn.

Model development and extensions

The comparison method provides consistent results when comparing fuzzy numbers. An interesting extension would be to generalize the comparison such that it could handle general fuzzy sets.

The approximate reasoning model, as suggested in Chapter 5, could be extended in numerous directions. Following the discussion on rule structure and complexity (in sections 5.2.1 and 5.2.2), we concentrated on first-level rules, and second-level rules with RHS interactions. Other rule structures could be tested and compared. For calibrating the rule matrix in the approximate reasoning model we used a very simple heuristic. Other more rigorous estimation techniques that could be used are the ILP formulation (as suggested in section 5.7.2.1), and the branch and bound approach (as suggested in section 5.7.2.2) which needs to be developed further to accommodate the unique characteristics of the model. For calibrating the approximate reasoning model, we determined the membership functions as inputs, and treated the rules as decision variables. A combined estimation of membership functions and rules could be a very interesting direction to explore, both from a conceptual and a mathematical point of view. We implemented the default reasoning hypothesis using the approximate reasoning model by utilizing a very simple approach: weigh the existing knowledge rules by their consistency with the provided information. Other schemes for implementation of the default model should be explored as well.

Development of empirical methods for data collection

The effort for collecting data on route choice behavior in the presence of information should be continued for both revealed and stated preferences data. On the revealed preferences front, hopefully the "real data" (which corresponds to actual choices under the various information provision scenarios) will be available in the near future.

Until such data is available, efforts should be concentrated on collecting revealed preference data through surveys and interviews, including sequences of choices, and panel data collection. As for the stated preferences front, we believe that driver simulators, such as the one described in Chapter 6, have good potential to realistically simulate route choice decisions under various information provision scenarios, with a fully controlled environment. Use of the simulator that we have designed and implemented should be extended to include other information provision scenarios (e.g. recommended shortest path, changeable message signs), and different degrees of information reliability (through degree of match between observed and provided traffic conditions).

Issues related to route choice behavior in the presence of information

We proposed using fuzzy sets for modeling perceptions, and suggested utilizing trapezoidal fuzzy numbers (based on simple interview) to represent the corresponding membership functions. More thorough investigation of other forms of membership functions, and methods for generating membership functions are needed.

Reliability of the information provided is an important issue which requires further research. In most of our discussions we assumed that the information comes from a relatively reliable source. In sections 3.4.2 and 5.5 we discussed how to adapt the models to unreliable information sources, however no estimation procedure was suggested. Development of models which can estimate the reliability of the information source are needed in order to be able to handle unreliable information or information with unknown reliability.

As drivers gain experience with the network and the use of existing information systems, their long term perceptions and behavior pattern may change. In this thesis we focused on short-term decisions based on current conditions and existing knowledge, and ignored the learning processes associated with daily interaction with such new systems. Models that include learning associated with the interaction with existing information systems, formation of perceptions regarding their reliability, and the resulting behavior pattern, should be the topic of future research.

Finally, we emphasized the human aspects of the decision process by utilizing concepts from fuzzy set theory and approximate reasoning. If such models are indeed appropriate for the problem addressed in this thesis, as the results obtained so far indicate, then other human behavior decision processes might also benefit from models similar to the ones presented in this thesis.

APPENDIX A: FUZZY SET THEORY - DEFINITIONS AND NOTATIONS

- X is a set of objects called the *universe*, whose elements are denoted by x .
- A *fuzzy set* A is defined by a *membership function* $\mu_A: X \rightarrow [0,1]$, where $\mu_A(x)$ is the grade of membership of x in A , the higher $\mu_A(x)$ - the more x belongs to A .
- The *support* of a fuzzy set A is the subset: $\text{Support}(A) = \{x \in X \mid \mu_A(x) > 0\}$.
- The *height* of a fuzzy set A is given by: $\text{Hgt}(A) = \max_{x \in X} \mu_A(x)$.
- A_α , the α -*cut* of a fuzzy set A is given by: $A_\alpha = \{x \in X \mid \mu_A(x) > \alpha\}$.
- A *fuzzy number* F is defined on \mathbb{R} such that:
 - (a) F is normal, i.e. there exists $x \in \mathbb{R}$ such that $\mu_F(x) = 1$
 - (b) F is unimodal, i.e. for all $\lambda \in [0,1]$ and for all $x, y \in \mathbb{R}$:

$$\mu_F(\lambda x + (1-\lambda)y) \geq \min(\mu_F(x), \mu_F(y))$$
 - (c) μ_F is piecewise continuous
 - (d) F has a bounded support.
- A *fuzzy relation* is a fuzzy set defined on the Cartesian product of crisp sets X_1, \dots, X_n , where tuples (x_1, \dots, x_n) may have various degrees of membership within the relation. The membership grade is usually represented by a real number in the closed interval $[0,1]$, and indicates the strength of the relation present between the elements of the tuple.
- A *t-norm* is a function of two arguments: $t: [0,1] \times [0,1] \rightarrow [0,1]$ such that for all $x, y, z, w \in [0,1]$:
 - i) it is non-decreasing in each argument: for $x \leq y, w \leq z$:

$$t(x,w) \leq t(y,z)$$
 - ii) it is commutative: $t(x,y) = t(y,x)$
 - iii) it is associative: $t(t(x,y),z) = t(x,t(y,z))$
 - iv) it satisfies the set of boundary conditions: $t(x,0) = 0, t(x,1) = x$
- *The extension principle* allows performing an operation f between two fuzzy numbers A and B results in a third fuzzy number C which is defined by the following membership function (see Pedrycz 1989 for general conditions):

$$\mu_C(z) = \max_{z=f(x,y)} t[\mu_A(x), \mu_B(y)]$$

where $z \in \mathbb{R}$ and t is a t -norm. Usually the t -norm chosen is the min operation, that is: $t(x, y) = \min(x, y)$.

- Fuzzy sets A and B are called *equal* if $\mu_A(x) = \mu_B(x)$ for all $x \in X$.
- The *possibility distribution function* of U induced by the proposition: "U is A", is a function $\Pi_U: X \rightarrow [0,1]$ which is equal to μ_A .
- The *possibility of A* is defined by:

$$\text{Poss}(A) = \max_{x \in X} \mu_A(x) \quad (\text{A-1})$$

- The *possibility of A with respect to B*, $\text{Poss}(A | B)$ is defined as:

$$\text{Poss}(A | B) = \max_{x \in X} \min(\mu_A(x), \mu_B(x)) \quad (\text{A-2})$$

- The *certainty (or necessity) of A with respect to B*, $\text{Cert}(A | B)$ is defined as:

$$\text{Cert}(A | B) = \min_{x \in X} \max(\mu_A(x), 1 - \mu_B(x)) \quad (\text{A-3})$$

APPENDIX B: APPROXIMATE REASONING AND FUZZY CONTROL

Fuzzy Control has been used successfully in various control applications (see for example Sugeno, 1985, for general applications, and Yasunobu and Miyamoto, 1985, for automatic control of train operation). In this Appendix we describe briefly the approach, more details can be found in Pedrycz (1989), De Mantaras (1990), and Kosko (1992).

The general fuzzy control scheme is illustrated in Figure 27. The building blocks of the decision process are rules of the form: "if condition is A_i then do B_i ", where A_i and B_i are typically labels of fuzzy sets (e.g. "if temperature is low then decrease fuel rate slightly"). Let the condition part of a rule be called left-hand-side and be denoted by LHS, and the consequence part, right-hand-side and denoted by RHS. Rule-based systems usually require existence of a huge set of rules in order to account for all possible states of the system. However, in the fuzzy control framework the number of rules required is reduced significantly by employing an approximate reasoning scheme. According to this scheme, each rule is treated as a family of rules, and thus rules that are "close" to it can be derived automatically.

Boolean logic assumes the following inference scheme (modus ponens):

$$\frac{\begin{array}{l} \text{if } X \text{ is } A, \text{ then } Y \text{ is } B \\ X \text{ is } A \end{array}}{Y \text{ is } B}$$

That is, if the premise of the rule is true, then the consequence is also true. In generalized modus ponens (Zadeh, 1975), A and B could be labels of fuzzy sets, and the rules are being interpreted with more flexibility by allowing the premise to be partially true as illustrated below:

$$\frac{\begin{array}{l} \text{if } X \text{ is } A, \text{ then } Y \text{ is } B \\ X \text{ is } A^* \end{array}}{Y \text{ is } B^*}$$

That is, the rule applies even though its premise is only partially true, and the consequence is changed accordingly. The condition part is being fulfilled only to a certain extent (the extent with which A^* is close to A), and as a result the consequence part, B, is applied to the appropriate extent, resulting in the outcome B^* . The membership of the set B^* is given by:

$$\mu_{B^*}(y) = \max_x t(\mu_{A^*}(x), I_1[\mu_A(x), \mu_B(y)]) \quad \forall y \quad (\text{B-1})$$

where t is the t-norm (as defined in Appendix A) and I_1 is an implication operator to be discussed below.

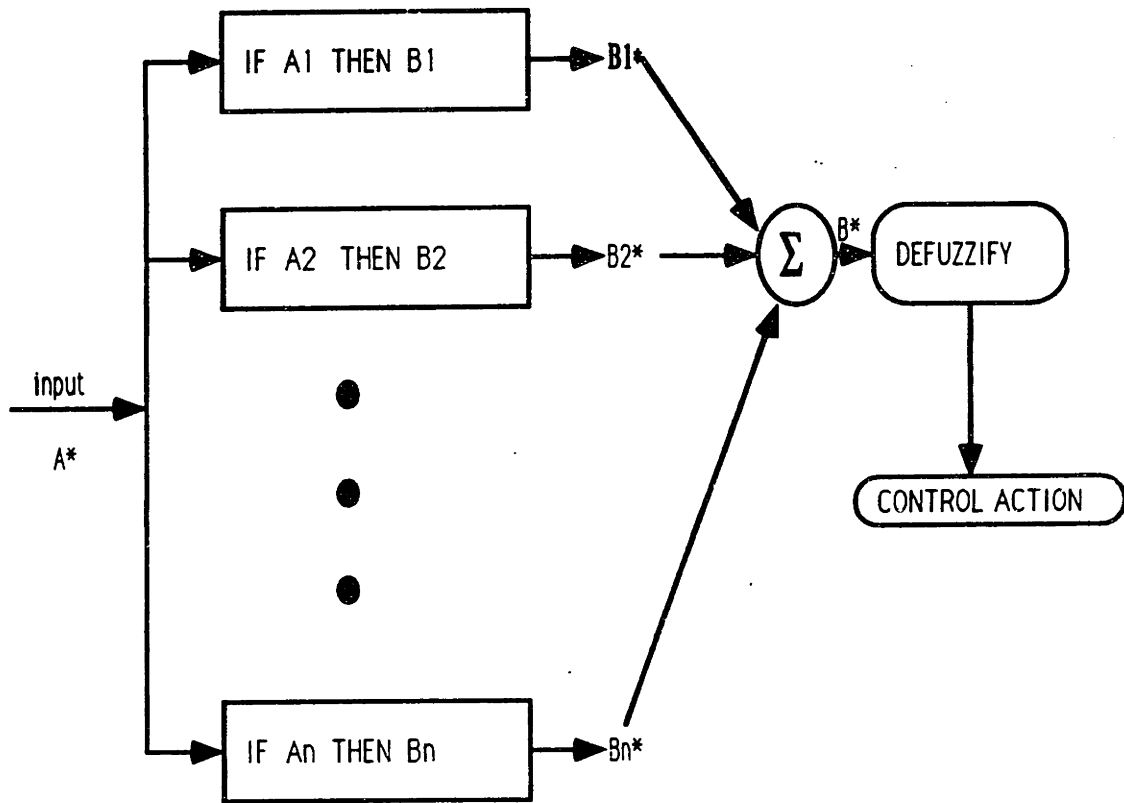


Figure 27: Fuzzy Control Scheme

Equation (B-1) can be written more intuitively as:

$$B^* = A^* \circ A \rightarrow B \quad (B-2)$$

where \circ corresponds to the max-composition of a t-norm. The representation in equation (B-2) shows that the new consequence set, B^* , is the result of the composition of the set A^* with the implication "if A then B". For implementing equation (B-2) the t-norm and the implication operation I_t have to be defined. In most fuzzy control applications the t-norm used is the minimum operation, that is: $t(a,b)=\min(a,b)$. As for the implication operator, various alternatives exist. The most popular implication scheme used is the one given by Mamdani (1977) according to which:

$$I_T(\mu_A(x), \mu_B(y)) = \min(\mu_A(x), \mu_B(y)) \quad (B-3)$$

Substituting the min t-norm and Mamdani's implication operator in equation (B-1) we get:

$$\mu_{B^*}(y) = \min(\alpha, \mu_B(y)) \quad (B-4)$$

where α is the degree with which the sets A and A^* overlap given by their max-min composition, and it is a measure of the degree with which the rule $A \rightarrow B$ is fired for a specific A^* set:

$$\alpha = \max_{x \in X} \min(\mu_{A^*}(x), \mu_A(x)) \quad (B-5)$$

This implication operator has the desired property that if $A^*=A$ then $B^*=B$. Figure 28a illustrates the application of Mamdani's operator. When the input A^* is deterministic, that is:

$$\mu_{A^*}(x) = \begin{cases} 1 & x=x_0 \\ 0 & x \neq x_0 \end{cases}$$

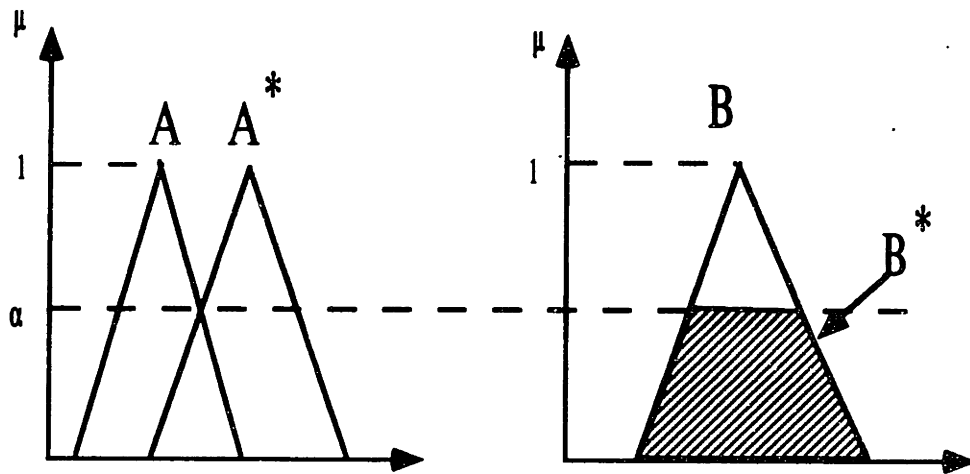
then $\alpha = \mu_A(x_0)$.

Kosko (1992), suggested another inference scheme, called the correlation-product encoding scheme, according to which the membership function of B^* is given by:

$$\mu_{B^*}(y) = \alpha \cdot \mu_B(y) \quad (B-6)$$

This scheme, has the property (as with Mamdani's implication), that if $A^*=A$ then $B^*=B$. In addition it has the advantage that it preserves the shape of the membership function of the set B. Figure 28b demonstrates the application of the correlation-product encoding scheme.

(a) Mamdani's Implication



(b) Product-Correlation Implication

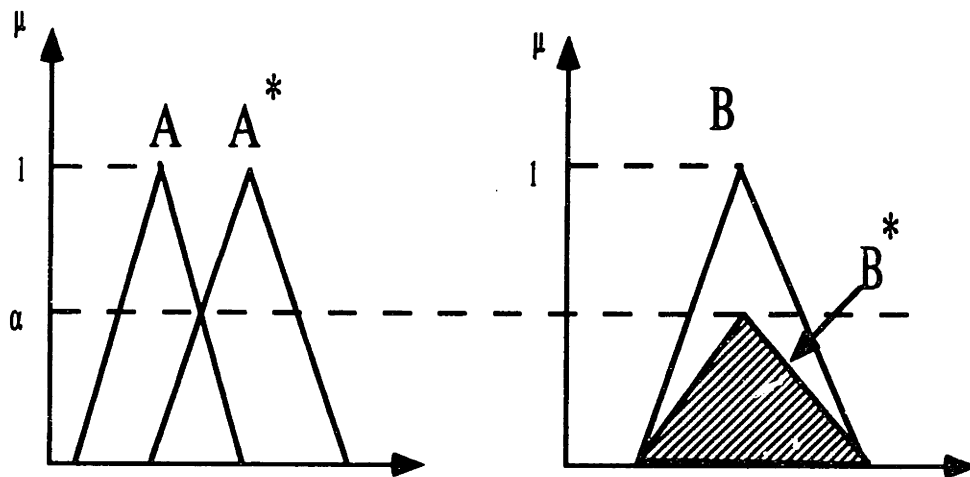


Figure 28: Implication Schemes

In traditional rule-based expert systems, only one rule is being 'fired' at a time, and the rule that is being fired is a rule whose LHS condition is fulfilled (with a conflict resolution strategy to pick one rule whenever more than one rule can be fired). However, in the fuzzy control decision framework a more flexible rule interpretation is utilized (allowing degrees of truth to LHS conditions), thus more than one (updated) rule consequence can be fired. Therefore, depending on the inputs, more than one rule may contribute to the final decision. Let α_i be the degree with which the input A^* overlaps with the i 'th rule condition, A_i , as given by equation (B-5). α_i dictates to what extent the i 'th rule consequence, B_i , is being fired, and results in B_i^* . In order to achieve the final output, all B_i^* 's (resulting from the various values with which the input A^* overlaps) are combined into one set B^* , and later defuzzified into control discrete action. In most fuzzy control applications which use Mamdani's implication, the membership function of B^* is given by (see e.g. Pedrycz , 1989):

$$\mu_{B^*}(y) = \max_{1 \leq i \leq N} \mu_{B_i^*}(y) \quad (B-7)$$

which is graphically equivalent to the union of all $\mu_{B_i^*}$. The use of the maximum operator over all the membership functions $\mu_{B_i^*}$ is motivated by the fact that the RHS action is supposed to be consistent with all the different rules that were fired, and the rule that was fired with the highest degree thus dominates the membership value for the specific value.

For defuzzifying the single combined fuzzy set B^* into a control action, the scheme most often used in fuzzy control applications is the center of gravity method given by:

$$z = \frac{\int y \mu_{B^*}(y) dy}{\int \mu_{B^*}(y) dy} \quad (B-8)$$

where z is the discrete value for the control action. The centroid is unique and uses all the information of the output set B^* . Kreinovich (1991) provides further mathematical justifications for using the center of gravity defuzzification scheme.

A Survey of Your Home to M.I.T Commute

Center for Transportation Studies
Massachusetts Institute of Technology

May 1991

Attached is a questionnaire prepared by a research team in the Center for Transportation Studies at MIT, in cooperation with the MIT Planning Office. This questionnaire is part of an on going research at MIT in the area of Intelligent Vehicle Highway Systems (IVHS), which aim at reducing congestion by providing on-line and user specific information to commuters using a variety of technologies. It is designed to provide transportation planners with a better understanding of the routes you follow on your daily commute, your preferences when it comes to choosing those routes, your reliance on traffic reports, and your parking needs at MIT.

You are kindly requested to fill out this questionnaire. By filling out this questionnaire you will help us design systems which will provide drivers with relevant, useful, and reliable information. It will also allow the MIT Planning Office to be more responsive to your preferences and needs in the future. All responses are *strictly confidential*.

How to fill out this questionnaire:

The questionnaire consists of two parts; **Part I** asks you about your usual commute to MIT, and **Part II** about your specific commute during the week of May 6 to May 10. Please fill out both parts and return them in the envelope provided by May 16, 1991.

Please *feel free* to write comments in the margins wherever appropriate.

Thank you in advance for your time and effort.

PART I: Your Usual Commute to MIT

1. In a typical 5 day work week, **how many times** do you use each of the following modes to commute to work:

__ Drive Alone __ Carpool Driver __ Carpool Passenger
__ Public Transportation __ other: _____

2. In a typical 5 day work week, how many days do you come to MIT? ____ days

3. Do you come to MIT on weekends? often occasionally never

FOR THE REST OF THIS SURVEY WE ARE INTERESTED IN DRIVERS BEHAVIOR AND CHOICES. IF YOU NEVER DRIVE TO WORK, PLEASE DO NOT FILL THE REST OF THE SURVEY. THANK YOU FOR YOUR WILLINGNESS TO PARTICIPATE.

4. When driving to MIT:

a. What time do you usually leave home? hour ___ min ___ a.m. p.m.

b. What time do you usually arrive at MIT? hour ___ min ___ a.m. p.m.

5. How much flexibility do you have in choosing the time you arrive at work on a daily basis?

- none up to 15 minutes 16 - 30 minutes
 31 - 60 minutes more than an hour

6. Think about a typical car trip from your home to work. Assume "regular" traffic conditions, i.e. no extreme traffic delays, no major incidents and no weather related problems. Under these conditions, how long does it usually take you to drive from your home to work? Please specify a range (e.g. from 40 to 55 minutes): from ___ to ___ minutes.

7. How often does your driving time to work exceed the range you specified in question 6?

- very often (more than once a week) often (approx. once a week)
 occasionally (approx. twice a month) rarely (approx. once a month)
 very rarely (less than once a month)

8. What is the shortest driving time you have ever experienced during your home to work commute? ___ minutes

9. What is the longest driving time you have ever experienced during your home to work commute? ___ minutes

10. In a typical 5 day work week, how many significantly different routes from home to MIT do you use? ___ routes (by "significantly different" we mean routes which almost do not overlap, for example: Mass. Pike and Route 9, or 93 and Morrissey Blvd.)

11. Please describe below your most frequently used route to MIT by indicating the major streets, highways, and bridges that compose the route: _____

12. Do you usually make stops on your way to MIT? no → proceed to question 14

yes, total duration of stops is approx. ___ minutes

13. What is the purpose of your stops? drop a passenger pick a passenger
 eat run errands fill gas other: _____

14. Where do you usually park your car?

- at an MIT parking lot on street
 on street at a meter other: _____

15. To which MIT parking facility do you have a sticker? _____

16. How long does it usually take you to get from your parked car to your MIT destination? ___ minutes

17. What time do you usually leave MIT? hour ___ min ___ a.m. p.m.

Your Attitudes and Preferences

18. On a scale of 1 to 5, where 1 indicates "strongly disagree" and 5 indicates "strongly agree", indicate your level of agreement with the following statements by checking the appropriate box:

	strongly disagree	1	2	3	4	5	strongly agree	not relevant
I am very familiar with at least 2 significantly different routes to work								
I often change my planned route while driving								
I like discovering new routes								
I am willing to try new routes to avoid traffic delays								
I always listen to radio traffic reports								
I usually follow the recommendations of radio traffic reports								
Radio traffic reports are usually reliable								
When traffic reports are different from my own observation, I ignore them								
I often change my route after listening to radio traffic reports								
I trust my own judgement more than traffic reports								
Traffic reports do not provide relevant information								
I am willing to pay in order to get more useful traffic information								

19. On a scale of 1 to 5 where 1 indicates "not important at all" and 5 indicates "very important", indicate the importance of the following factors in choosing your route to work:

	not important at all	1	2	3	4	5	very important
Time of day							
Commute time							
Habit							
Time spent stopped in traffic							
Number of traffic lights							
Traffic reports							
Risk of delay							
Weather							

ABOUT YOURSELF

The information requested in this section relates to your personal and household data. We need this information to better understand how personal and family characteristics affect commuting choices. All information collected will remain strictly confidential.

20. Sex: Male Female
21. Marital status: Married Unmarried
22. What is your age group?
 Less than 20 years 20 - 29 years 30 - 39 years
 40 - 49 years 50 - 64 years 65 or older
23. What is the highest level of education you have completed?
 High school or less Some College
 Graduated College Post graduate work
24. What is your home Zip Code? _____
25. How long have you lived at your present home address? ____ years
26. Do you own or rent your dwelling unit?
 own rent
27. How many persons including yourself live in your household? _____ persons
28. What is the total number of automobiles owned by your household? ____ automobiles
29. What is your household's approximate yearly income from all sources (before taxes)?
 Less than \$20,000 \$20,000 - \$40,000 \$40,000 - \$60,000
 \$60,000 - \$80,000 \$80,000 - \$100,000 More than \$100,000
30. How long have you worked at your present job location? ____ years
31. Which of the following categories best describes your position?
 Undergraduate Student Graduate Student Academic Staff
 Tenured Faculty Non-Tenured Faculty Administrative Staff
 Service Support Staff Research Staff
 Other: _____

General

32. Write any other comments related to your daily commute to work, the way you choose and follow routes, your attitude towards traffic reports, and your parking needs (optional):

PART II: YOUR COMMUTE TODAY

THIS PART OF THE QUESTIONNAIRE RELATES TO YOUR DRIVING BEHAVIOR DURING A SPECIFIC WEEK. IT CONTAINS 5 IDENTICAL SECTIONS, ONE FOR EACH DAY FROM MAY 6 TO MAY 10. PLEASE FILL OUT EACH SECTION AFTER YOU HAVE COMPLETED YOUR COMMUTE TO MIT FOR THAT DAY.

EVEN IF YOU MISS FILLING OUT THE SURVEY FOR ONE DAY DUE TO ANY REASON, PLEASE FILL OUT THE INFORMATION FOR THE SUBSEQUENT DAYS.

Would you be willing to respond to a follow-up questionnaire about your driving behavior during the next year?

If so, please fill out the following:

Name: _____

MIT address: _____

Participants will also receive a summary of our findings.

If you wish to remain anonymous, please take the time to fill out PART II anyway.

All responses are strictly confidential.

This part is designed to monitor your daily commute from home to MIT. We need to know whether the route you followed to work everyday was influenced by traffic information you listened to, by unusual traffic conditions you encountered on your way to work, or by commuting experience on the previous day. We are also interested in diversion decisions you made while on your way to MIT.

YOUR COMMUTE FOR MONDAY, MAY 6, 1991

- (1) How did you get to MIT today? Drive Alone Carpool Driver
 Carpool Passenger Public Transportation other: _____

IF YOU DID NOT DRIVE YOURSELF TO WORK TODAY, PLEASE IGNORE THE QUESTIONS FOR TODAY. CONTINUE TOMORROW WITH THE NEXT SECTION.

- (2) Did you receive traffic information before you left home today?

yes, from radio yes, from TV
 no → proceed to question 5

- (3) Did the information that you received before leaving home influence your route choice for today?

a lot somewhat very little not at all

- (4) What did the information you received indicate about traffic conditions on the route you decided to take?

much worse than usual worse than usual usual traffic conditions
 better than usual much better than usual I don't remember
 no information other: _____

- (5) Once you started your trip, what were the traffic conditions you observed at the beginning of your trip?

much worse than usual worse than usual usual traffic conditions
 better than usual much better than usual other: _____

- (6) While driving, did you receive any information about the route you were following?

yes, which radio station? _____ no → proceed to question 8

- (7) What did the information indicate about traffic conditions on the route you were following?

much worse than usual worse than usual usual traffic conditions
 better than usual much better than usual I don't remember
 no information other: _____

- (8) After you started your trip to work, was there a way to switch to another route that will take you to your destination?

yes no → proceed to question 14

- (9) While driving, did you switch from the route you were following ?

yes no → proceed to question 14

- (10) Before you switched routes, did you get any radio information about the route you switched to?

yes no → proceed to question 12

- (11) What did the information indicate about traffic conditions on the route you switched to?

much worse than usual worse than usual usual traffic conditions
 better than usual much better than usual I don't remember
 no information other: _____

(12) Based on your observation, what were traffic conditions on the route you switched to?

- much worse than usual worse than usual usual traffic conditions
 better than usual much better than usual other: _____

(13) Why did you switch? check all boxes that apply:

- radio traffic reports your own observation on traffic conditions
 forced detour other (please specify): _____

(14) Did you make any stops on your way? yes, total duration of stops was ___ minutes

no → proceed to question 16

(15) What is the purpose of your stops?

- drop a passenger pick a passenger
 eat run errands fill gas other: _____

(16) When did you arrive at your MIT destination? hour ___ min ___ a.m. p.m.

(17) Please describe the route you took today to MIT by indicating the major streets, highways, and bridges that compose the route: _____

About Your Trip Today

(18) On a scale of 1 to 5, where 1 indicates "strongly disagree" and 5 indicates "strongly agree", indicate your level of agreement with the following statements:

	strongly disagree					strongly agree	not relevant
	1	2	3	4	5		
Traffic conditions today were better than usual							
Traffic information received today was useful							
My commute time today was worse than usual							
I am satisfied with my route choice today							
I could have saved at least 5 min. had I taken another route							
I could have saved at least 5 min. had I gotten relevant information							
When I made my decision to switch routes today, I was confident that I would be better off							

(19) Write any other comments related to your trip to work today (optional):

Appendix D

A PRELIMINARY SESSION

- (1) What is your home address? _____
- (2) How long have you lived at your present home address? ____ years
- (3) How long have you worked at your present job location? ____ years
- (4) How much flexibility do you have in choosing the time you arrive at work on a daily basis?
 none up to 15 minutes 16 - 30 minutes
 31 - 60 minutes more than an hour
- (5) How many significantly different alternatives to get from home to MIT do you know?
 ____ alternatives
- (6) How many significantly different alternatives to get from home to MIT do you use in a typical 5 day week? ____ alternatives
- (7) Please describe your most often used alternative in as much detail as possible by indicating names of streets, highways and bridges that compose it.: _____

- (8) In a typical 5 day work week, how many times do you choose the alternative described in question (7)? ____ times

The attached map describes three alternatives to get from Newton to M.I.T.: using Beacon Street, using Commonwealth Avenue, and using the Mass. Pike. For each of the three alternatives please indicate what is the range of most possible travel times, and what is the **shortest** and the **longest** possible travel times on that alternative.

	range of "usual" travel times (min)	shortest possible travel time (min)	longest possible travel time (min)
Beacon St.	from ____ to ____		
Comm. Ave.	from ____ to ____		
Mass. Pike	from ____ to ____		

Thank you for your willingness to participate in this survey.

References

Adler J. L., Recker W. W., and McNally M. G. "A Conflict Model and Interactive Simulator (FASTCARS) for Predicting Enroute Assessment and Adjustment Behavior in Response to Real-Time Traffic Condition Information", Paper presented at the 71st TRB Annual Meeting, Washington D.C. 1992.

Ahuja R. K., Magnanti T. L., and Orlin J. B. "Network Flows", Sloan School of Management, MIT, 1988.

Anderson N. H. "Foundations of Information Integration Theory" Academic Press, 1981.

Anderson N. H. "Methods of Information Integration Theory" Academic Press, 1982.

Ben-Akiva M., De Palma A., and Kaysi I. "Dynamic Network Models and Driver Information Systems", Transportation Research 25A 5, pp. 251-266, 1991.

Ben-Akiva M. and Kaysi I., "Drivers Information Processing in the Context of Advanced Driver Information Systems (ADIS)", Paper presented at the 70th Annual TRB Meeting, Washington D.C. 1991.

Ben-Akiva M. and Lerman S. "Discrete Choice Analysis, Theory and Application to Travel Demand", MIT Press, 1985.

Bonsall P. W. "Using an Interactive Route Choice Simulator to Investigate Drivers' Compliance with Route Guidance Advice", Transportation Research Board, 70th Annual Meeting, Washington D.C. 1991.

Bortolan G. and Degani R. "A Review of Some Methods for Ranking Fuzzy Subsets", Fuzzy sets and systems, 15, pp. 1-19, 1985.

Bovy P.H. and Stern E. "Route Choice: Wayfinding in Transport Networks", Kluwer Academic Publishers, 1990.

Boyce D. E. "Route Guidance Systems for Improving Urban Travel and Location Choices", Transportation Research-A vol 22A, No. 4, pp. 275-281, 1988.

Chanas S., "Fuzzy Optimization in Networks" in Optimization Models Using Fuzzy Sets and Possibility Theory, ed. by Kacprzyk J. and Orlovski S. P. Reidel, Dordrecht, The Netherlands, pp. 303-327, 1987.

Chanas S. and Kamburowski J. "The Fuzzy Shortest Route Problem", Proc. Polish Symp. Interval and Fuzzy Math, Poznan, pp. 35-41, 1985.

Chen S. "Ranking Fuzzy Numbers with Maximizing Set and Minimizing Set", Fuzzy Sets and Systems, 15, pp. 113-129, 1985.

Clark W.A.V. and Smith T.R. "Production System Models of Residential Search Behavior: A Comparison of Behavior in Computer-Simulated and Real-World Environments", Environment and Planning A, 17(4), pp. 555-568, 1985.

Delgado M., J. L. Verdegay and M.A. Vila, "On Valuation and Optimization Problems in Fuzzy Graphs: A General Approach and Some Particular Cases", ORSA Journal of Computing, vol 2 no 1, pp. 74-83, Winter 1990.

De Mantaras R.L. "Approximate Reasoning Models", John Wiley & Sons Inc., 1990.

Diamond P. "Some Maximum Likelihood Estimators for the Fuzzy Linear Model", in Uncertainty and Intelligent Systems, 1988.

Dubois D. and H. Prade, "Shortest Path Algorithms with Fuzzy Data" (in French), R.A.I.R.O. Op. Res. 12 pp. 214-227, 1978.

Dubois D. and H. Prade, "Ranking Fuzzy Numbers in the Setting of Possibility Theory". Information Sciences, 30, pp. 183-224, 1983.

Dubois D. and Prade H. "Default Reasoning and Possibility Theory", Artificial Intelligence, 35, pp. 243-257, 1988.

Freeling A. N. S., "Fuzzy Sets and Decision Analysis". IEEE Transactions on Systems, Man, and Cybernetics, vol SMC-10, no 7, pp. 341-355, 1980.

French R.L. "Intelligent Vehicle/Highway Systems in action", ITE Journal, pp. 23-31, November 1990.

Hammond K.R., McClelland, and Mumpower J. "Human Judgement and Decision Making", New York: Hemisphere Publishing, Frederick A. Prager, 1980.

Hogarth R. M. "Judgement and Choice. The Psychology of Decisions", John Wiley & Sons, 1980.

Horowitz A. J. "The Subjective Value of the Time Spent in Travel", Transportation Research vol 12, pp. 385-393, 1978.

- Jayakrishnan R. and Mahmassani S. "A Dynamic Modeling Framework of Real-Time Guidance Systems in General Urban Traffic Networks", in Proc. of the Second International Conference on Applications of Advanced Technologies in Transportation Engineering, ed. by Stephanedes Y.J. and Sinha K.C., 1991.
- Khattak A.J., Koppelman F.S., and Schofer J.L. "Stated Preferences for Investigating Commuters' Diversion Propensity", Paper presented at the 71st TRB Annual Meeting, Washington D.C. 1992.
- Kaysi I. A. "Framework and Models for the Provision of Real-Time Driver Information", Ph.D. Dissertation, Dept. of Civil Engineering, MIT, 1992.
- Kaufmann A. and Gupta M. M. "Introduction to Fuzzy Arithmetic", Van Nostrand Reinhold Company, 1985.
- Klir G. J. and Folger T. A. "Fuzzy Sets, Uncertainty and Information", Prentice Hall, 1988.
- Kosko B. "Neural Networks and Fuzzy Systems. A dynamical Systems Approach to Machine Intelligence", Prentice Hall, 1992.
- Koutsopoulos H. N. and Lotan T., "Motorist Information Systems and Recurrent Congestion: A Sensitivity Analysis of Expected Benefits", Transportation Research Record, 1281, pp. 148-159, 1990.
- Koutsopoulos H. N. and Xu H. "An Information Discounting Routing Strategy for Advanced Driver Information Systems (ADIS)", Paper presented at the 71st TRB Annual Meeting, Washington D.C. 1992.
- Kreinovich V. and Quintana C. "What Procedure to Use While designing a Fuzzy Control? Towards Mathematical Foundations of Fuzzy Control". in Working Notes of the 1st International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems, College Station, TX. Nov 1991.
- McCahon C. S. and Lee E. S. "Comparing Fuzzy Numbers: The Proportion of the Optimum Method", International Journal of Approximate Reasoning, 4, pp. 159-181, 1990
- McDermott D. and Doyle J. "Nonmonotonic Logic I", in Readings in nonmonotonic reasoning, ed. by Ginsberg M. L., 1987.
- Mamdani E. H. "Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis", IEEE Transactions on Computers, vol C-26, no. 12, pp. 1182-1191, 1977.

- OECD. Route Guidance and In-Car Communication Systems. Paris, 1988.
- Pedrycz W. "Fuzzy Control and Fuzzy Systems", John Wiley & Sons Inc., 1989.
- Rasmussen J. "Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering", North-Holland Series in System Science and Engineering, vol 12, 1986.
- Sheridan T. B., "Human Factors of Driver-Vehicle Interaction in the IVHS Environment", Progress Report to Office of Crash Avoidance Research, National Highway Traffic Safety Administration, US department of Transportation, 1991a.
- Sheridan T. B., Personal Communication, 1991b.
- Simon H. A. "The Sciences if the Artificial", MIT Press, 1969.
- Smithson M., "Fuzzy Set Analysis for Behavioral and Social Sciences", Springer-Verlag, 1987.
- Sugeno M. "Industrial Applications of Fuzzy Control", Elsevier Science Publishers, 1985.
- Tanaka H., S. Uejima and K. Asai "Linear Regression Analysis with Fuzzy Model", IEEE Transactions on Systems, Man and Cybernetics, vol SMC-12, no 6, pp. 903-907, 1982.
- Teodorovic D. and Kikuchi S. "Transportation Route Choice Model Using Fuzzy Inference Technique", Paper presented at the First International Symposium on: "Uncertainty Modeling and Analysis: Fuzzy Reasoning, Probabilistic Models, and Risk Management", University of Maryland, College Park, 1990.
- Turksen I.B., "Measurements of Membership Functions". In Applications of Fuzzy Sets Theory in Human Factors, ed by W.Karwowski and A. Mital. 1986.
- Yager R. R., "Possibilistic Decision making". IEEE Transactions on Systems, Man, and Cybernetics, vol. smc-9, no 7, pp. 388-392, 1979.
- Yager R. R., "A Procedure for Ordering Fuzzy Subsets of the Unit Interval", Information Science, 24, pp. 143-161, 1981.
- Yasunobu S. and Miyamoto S. "Automatic Train Operation System by Predictive Fuzzy Control", in Industrial Applications of Fuzzy Control, etd. by Sugeno M. Elsevier Science Publishers, 1985.

Zadeh L.A., "Fuzzy Sets", Information and Control, vol 8, pp. 338-353, 1965.

Zadeh L.A., "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes". IEEE Trans. Systems, Man, and Cybernetics SMC-3, pp. 28-44, 1973.

Zadeh L.A., "The Concept of Linguistic Variable and its Application to Approximate Reasoning I, II and III", Information Sciences, 8, pp.19-249, 301-357, 9, pp. 43-80, 1975.

Zadeh L.A., "Fuzzy Sets as a Basis for the Theory of Possibility", Fuzzy Sets and Systems, 1, pp.3-28, 1978.