# VIRTUAL MARKETS: THE APPLICATION OF AGENT-BASED MODELING TO MARKETING SCIENCE

by

Roger A. Parker

Submitted to the School of Marketing, Faculty of Business in fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

UNIVERSITY OF TECHNOLOGY SYDNEY

September 2010

# Certification of Authorship and Originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as fully acknowledged within the text. I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Production Note:
Signature removed prior to publication.

Author:________________________________________

Roger A. Parker, UTS ID #[redacted] 26 September 2010

## Dedication

To Anita.  Without her love and encouragement, this dissertation would not exist.  And, quite literally, neither would I.

## Abstract

Advances in any science ultimately depend on the creation of instruments that can create observations from which theories can be hypothesized and tested. This dissertation proposes that significant advances in marketing science can be realized with the engagement of the advanced computational science technique of agent-based modeling. To support this proposition, the methodology is examined from first principles to concrete implementation. The ontological and epistemological bases for agent-based modeling are developed, and the evolutionary science paradigm as it applies to marketing (for which the method is most useful), is reinforced with extensive analysis of evolved universal human behaviors, especially behavior relevant to marketing. The concept of the narrative framework is then posited. The primary property of the framework is the central role of choice as an expression of value and resource allocation. This framework then explicates the notion of virtual market, and a appropriate definition of agent derived. The computing requirements and skills needed to actually building a virtual market are also proscribed. Then a detailed, operational example of a virtual market is laid out. Called AirVM, it portrays the dynamics of the market for passenger air travel by simulating the product definition and ticket purchasing process for every passenger travelling on every regularly scheduled commercial flight in the world over a week time period – over 40,000,000 passengers flying on thousands of flights, offered by hundreds of carriers. The synthetic populations of passengers (customers) and airlines (sellers) have empirically-derived distributions of salient properties, called incidence distributions, which are described in detail with empirical data to support their formulation and parameter estimation. The computing logic and samples of the interface are presented, and the system critiqued using appropriate agent-based modeling criteria. The major contributions of the research are the verification of the ontological suitability of agent-based modeling to marketing science, the empirical confirmation of the evolutionary basis for marketing behavior, the conceptual structure for construction of agent-based models in market research, and the proscription of how to construct a virtual market, illustrated with a detailed example. There are also several contributions to the airline passenger industry that emerge from the work. Finally, the dissertation contributes another example of the application of the technique to the burgeoning literature on agent-based modeling.

## Acknowledgements

Of central importance in the conduct of this research are my supervisors from the UTS Faculty of Business. Paul Wang has had to suffer with the seemingly interminable delays as I worked at a full time job and meandered down myriad pathways leading to this final work. He is to be commended for his patience. Siggi Gudergan, who came late this particular party, has shown interest in the work since the beginning, and I am very pleased that he was able to formally participate. Louise Young, who has now moved on from UTS, was of invaluable assistance early on in the development of the concepts that are developed here. The other individual at UTS who deserves special recognition is Jordan Louviere, who first introduced me to Discrete Choice Modeling and had a foundational impact on the results of this effort.

The work reported in this thesis is, in part, the result of a research project into passenger behavior that I was responsible for while I was engaged as the Senior Marketing Scientist for the Marketing Department of Boeing Commercial Airplanes in 2002. I was privileged to assemble a research team of top class individual from around the world to work on better understanding why passengers bought airplane tickets. Many advances were made in our understanding of airline passenger behavior that are now being adopted by the broader airline community, and many of the contributions generated by that team contribute to and are discussed in this work. Cheri Jones, my manager at the outset at Boeing, had the foresight and courage to let me assemble and support the team. Richard Lonsdale (who taught me about airlines), Fred Ervin and Zhengjie (John) Zhang (who were programming masters) were my close Boeing colleagues during the entire adventure of the next six years.

Among those prominent in the effort were Jordan Louviere, who taught me choice theory, Richard Carson, then Chair of the Department of Economics at University of California San Diego, and Joel Watson, also at UCSD, who tutored me in economics and worked with me on the passenger OD demand model, and David Bunch, at University of California, Ervine, who estimated the first passenger itinerary choice model. From the University of Warsaw, Poland, the team from their Institute of Computational Mathematics led by Marek Niezgodka and with the help of Wojciech Wislicki, Jan Radomski, Kristoff Nowinski, and Andzrej Slodownik worked to developed advanced concepts of airline network analysis. DeAnn Julius, then on the

Monetary Advisory Board of the Bank of England and now the Chairwoman of the Royal Institute of Foreign Affairs in London, helped me understand the problems and difficulties of the modern airline industry, and how technology could be brought to bear on those problems. Moshe Ben-Akiva from MIT and his student Joan Walker, now at the University of California at Berkeley, contributed to the mixed logit model that is now the pag choice model described here. Frank Koppelmann, from Northwestern, and Laurie Garrow at Georgia Tech, and her student Dan Illiescu, worked with me on many aspects of the application of discrete choice modeling to the airline industry. Dan, in particular, developed his PhD dissertation on the ticket cancellation model discussed here based on my recommendations regarding approach and data.

Finally, I owe my two friends and business associates, Nick Lanyon in London and David Perroud of M1nd-Set SA in Geneva, a continuing debt of gratitude for the forbearance and encouragement during the process of writing the thesis itself over the last two years. They have contributed much in wisdom and experience, even though they would not admit it.

A final comment on the personal point of reference from which this analysis is created. I maintain, as explained in the body of the work, that all analyses are conditional on a reference narrative which sets the framework for the motivation, method and logic of the work. Most doctoral theses are scholarly examinations of topic of interest to their creator, often crafted as launching pads for their careers in the field of their choice. And they serve that purpose admirably, establishing the worth of the individual in the context of his or her reference narrative and that of others who share similar narratives. This is not exactly the case for me. Rather, this effort is a milepost in what has already been a rather long journey. As I write this I am in my mid-sixties, and have been working with computers and mathematical models of social phenomena for over four decades. I have been engaged in the development of models of such disparate subjects as consumer movement patterns in shopping districts, recreational area utilization, racial desegregation of public schools, vehicle reliability inspection programs, and (obviously) global air passenger demand. I have developed and implemented literally hundreds of consumer survey research designs and analyses. This long experience colors the approach and reasoning supporting this work, and it is my hope the reader finds that those hues enhance its value beyond simple research and reporting of results.

## Virtual Markets: The Application of Agent-Based Modeling to Marketing Science

# Table of Contents

# List of Figures

# List of Tables

# Chapter One:
# Overview

## 1.1: Introduction

**1.1.1:** All science depends on instruments. Instruments are devices that allow us to create observations. Observations are the fundamental substance of science. An experiment cannot be performed without the creation of observations to characterize the cause and effect relationships that form the hypothesis being tested by the experiment. This is no less true of marketing science. Marketing science needs instruments to gather data for the formulation and testing of hypotheses just as other sciences do, and without such instruments any claim to marketing science is at best conceptually weak and at worst intellectually specious. The subject of this dissertation is the definition, construction and application of a specific instrument for use in marketing science. This instrument is agent-based modeling and simulation using modern computing technology.

**1.1.2:** Over the past three decades, social science has begun to investigate the effective application of computers to problems hitherto resistant to the more classical scientific methods of the traditional "hard" sciences. I will argue in this dissertation that computers are more than handy data collection and retention vehicles, they also extend the human ability to explore and explain the inherently complex interrelationships that form societies and their attendant organizations. In particular, using this cognitive extension, it is possible to model operational markets of buyers and sellers interacting in realistic market settings, which, in my view, is essential to the advancement of the science which forms the underpinnings of market research.

**1.1.3:** Computational science is a powerful tool that has emerged from this frontier, and agent-based modeling is one of the more powerful and flexible implementations of the approach. My aim in this dissertation is to define and demonstrate how this tool can address certain fundamental problems in marketing science, and therefore in marketing as a whole, which have not been successfully addressed by the more traditional, classical methods. The purpose of science is to predict. Richard Feynman, the Noble laureate physicist, noted that the definition of science is " . . . really very simple. We make a guess about what will happen in the future. We then repeatedly watch what happens in the future, and if the guess proves correct often enough, we call it a scientific fact." (Feynman, 1965) If the tribal witch doctor can read his sacred bones with sufficient accuracy, he will keep his post, but if a challenger appears with different charms that produce better forecasts, he will soon be displaced. In this dissertation I argue the agent-based modeling is a better set of scientific charms. If that point can be successfully demonstrated, then the foundation is laid for the development of a substantial and significant battery of tools to attack longstanding, difficult market research problems.

**1.1.4:** The instruments used by a scientific endeavor condition and are conditioned by the science being developed. What can be observed, and therefore forecast, depends on what devices and methods are available for that observation. The advent of the microscope brought bacteria into view, and understanding the effects of bacteria on the body – either as essential to nutrition or vehicles of disease – came within reach of analysts. And since no model is ever completely accurate, no tool based on a model can be absolutely error-free, and any forecast derived using that tool will miss the mark by some degree. But there is constant pressure to develop models which lead to instruments with more acute vision, greater depth, and more accuracy, shedding light on things that before were invisible. This dissertation continues in that tradition, proposing a more sensitive instrument for the observation and forecasting of marketing techniques and results.

## 1.2: Agent-Based Modeling and the Current Marketing Science Toolset

**1.2.1:** Consider for a moment the toolset used by the practicing market research industry. As has been the case for the past half-century, the staple observation tool of the industry has been the survey, and an overwhelming proportion of the analysis applied to that data (analytics, as the field likes to call it) consists of banner-and-stub tables. Marketing as a field derives much of its set of research and analytical toolset from the social science fields of economics, sociology and psychology, and the person-to-person survey is a fundamental tool of these fields. Scanning the industry, probably the most common types of market research firms are those which specialize in consumer surveys. A typical listing of market research firms, say by the American Marketing Association, catalogs well over 2500 such companies in the United States. Recently, internet panels of consumers have grown in popularity, with the associated savings in data collection costs.[1]

**1.2.2:** Rarely do the survey results tables provide any statistical support, beyond a generic statement like "the margin of error is plus or minus four percent."[2] Often apparently strong effects are misrepresented and misunderstood. However, a table is relatively easy for a manager to understand, and she (wisely) leaves the statistical details to a more technically skilled associate – sometimes. Other times it's plainly simply misused. In the last twenty years or so, however, more advanced techniques have begun to appear in practice. Conjoint analysis has been widely applied in product feature trade-off analyses, in spite of its obvious theoretical drawbacks. Many of those deficiencies are effectively addressed with discrete choice modeling, which has a much more sound theoretical basis, being grounded in utility-maximization economic theory. Various other techniques approach product differentiation and consumer heterogeneity in other ways, hierarchical Bayes methods being increasingly applied, at least in the United States.[3]

**1.2.3:** In the academic community, as one might expect, more advanced tools and techniques are more likely to appear. Heirarchical Bayes, for example, arose out of work done at Ohio State University. Very often the market research academic explores the possible application of a tool from a related field to a marketing research problem. Discrete choice certainly has benefited from this technological transfer. And the methodology that is the subject of this dissertation falls into this technology transfer category. Agent-based modeling has been available for some two decades, but only recently has it moved into applied areas with any momentum. In the early development period, such models were wholly artificial, describing environments and agents that existed only as computer programs with little or no corresponding entities in reality.

---

[1] to the extent that it has been suggested that a professional "panel member" has emerged. It's been asserted that as many as 40% of internet panel members belong to more than one panel.

[2] which is, of course, patently absurd. A table contains discrete count data, and the "+/- 4%" applies to the proportion of the total that a particular cell contains. The confidence interval on a proportion clearly depends on sample size, but also depends on how close the proportion is to zero. Also, the confidence interval is symmetric only if the estimated proportion is 0.50.

[3] Note that I have not cited any references in this brief summary. Such references are identified in the chapter discussion.

Recently, however, ecology and biology have begun to create agent-based models that represent actual organisms or biological functions that exist in nature. Ecology is arguably the most advanced in this regard. And economics is moving away from utility optimization theory into so-called "behavioral economics" or "computational economics" as it starts to apply agent-based modeling principles to the understanding of human economic systems.

**1.2.4:** In a very real sense, one of the two primary goals of this dissertation is to move along the adoption of agent technology by the market research community, both academic and practicing. From the academic perspective, the methodology promises to create significant opportunities for advances in theoretical marketing science. The narrative framework – the apparatus used to characterize the agent behavior structure in a virtual market – is one such opportunity. But there is an intimate, even symbiotic, relationship between academic marketing science and practicing market research. That symbiosis is a significant part of this dissertation. It is not enough to invent and characterize a methodology for marketing science. It must be demonstrated with actual data in an actual market. This is the second of the primary goals. AirVM, the virtual market described in the course of this dissertation, is being introduced for applications in market research in the world's airline industry as this is being written. Much has been learned about agent-based models in marketing by the work attendant to that development and introduction. These lessons have shaped and tested the theoretical development in a number of ways.

**1.2.5:** The creation of agent-based models for market research purposes is on the frontier of tool development for marketing science. This treatise explores the nature of these models, the attributes of the systems they are created to represent, and the kinds of issues that arise when building and applying them. Both theoretical and practical questions are addressed, and a working example of a real market is described. Now let me turn to the specific research program used here to create and implement this paradigm.

## 1.3: The Program for this Research

**1.3.1:** Simply put, I consider marketing science as the theoretical and instrumental basis for marketing research. I appreciate the distinction made by Shelby Hunt between *market* research and *marketing* research, where the former addresses questions of a company's specific marketing problems, while the latter is concerned with contributions to the general set of knowledge associated with marketing as a human activity. I assert that marketing science covers the application of science to both activities. The tools of theoretical physics apply equally well to the exploration of the origins of the universe and the design and construction of advanced electronics. Agent-based modeling fulfills the same dual role in marketing. Thus most of the early discussion in this thesis is directed toward application in marketing research, while the example described in the later chapters would qualify as market research.

**1.3.2:** The research problem posed by this challenge, and thus the structure of this discourse, is addressed with a five-step research program. The first step is epistemological. What can be "seen" using this method that is invisible otherwise, and are those things important to marketing? Modern computational capability, that is the availability of massive and inexpensive computing power, creates a new set of tools for the scientist. These new instruments extend the human power to observe and understand the world. But are they real? Can worlds that exist only as electric pulses in semi-conductor machines represent real things that happen in the real world? If so, what kinds of things can they represent, and is marketing an activity that meets the requisite criteria? So the first step in the research program is to establish that agent-based models, when properly built and applied, present valid representations of phenomena found in the real world. And to what kind of real-world phenomena do they apply, and does marketing fall into that class?

**1.3.3:** This is so, and this new capability benefits marketing science directly, if marketing science properly belongs in the class of scientific methods covered by the label "evolutionary science." The application of agent-based models to evolution has yielded unique and invaluable insights – such as emergence, path dependence, and sound stochastic forecasts – which have in turn led to significant advances in biology. As such it is better served by methods of computational and generative science than by other, more classical, scientific methods. Agent-based methods can represent and portray heterogeneous behavior and individual interaction dynamics substantially more accurately than aggregate statistical summaries. Current scholarly opinion puts marketing squarely into the evolutionary science domain. However, that opinion is based on rather elementary comparisons with biological evolution, and little in the way of empirical result supports the hypothesis. The second step in this research program is to pose and test this hypothesis directly. That is, can human marketing behavior be derived from the observed evolutionary attributes of human beings? If so, then two benefits accrue: first, models of human behavior that are necessary for agent model design, when validated in practice, can apply to many cultures and circumstances with only nominal, and known, modification; and second, the structure of the evolutionary heritage can point the way to a structure that supports agent design and construction. As will be seen in the discussion, the hypothesis is strongly supported by the evidence, and thus the two consequences result.

**1.3.4:** Marketing science is an evolutionary science because evolution is the basis for all human marketing behavior, and marketing science cannot avoid examination of human behaviors. While it is quite reasonable to study marketing in the context of inter-firm competitive dynamics, or as an organizational manifestation of an industrial or economic system, eventually the discussion must get down to buyer and seller, customer and supplier, or consumer and firm. These are human peculiarities, and an instrument that can capture and display these behaviors at the individual human level will provide, *a fortiori*, better forecasts of marketing results. So at some point what must be modeled is the behavior of the individual entities in the market. Moreover, the behavior of people in a marketing context contains features that are universal across all cultures and societies. The third step in the research program is then to define a framework for the portrayal and observation of human marketing behavior in sufficient depth to use it as a template for models of human agents in real markets. This research objective is met with the concept of the Narrative Framework.

**1.3.5:** But setting out the foundation and characteristics of a tool or instrument is more than an ontological or epistemological discourse. The models must be realized. The tool must actually be built. Agent-based models are, in fact, computer software objects that are written in real computing languages to run on actual machines. There is technique, skill, and a degree of art required to carry out that construction. And because these models extend the human ability to understand the phenomena they portray, how skillfully they are built directly affects how useful they will be. This practical side of the application of agent-based models to marketing science I refer to as *virtual markets*. So the fourth research objective is to identify and characterize the features and properties of virtual markets, and how are they built.

**1.3.6:** Finally, to validate the underlying conceptualization and technical implementation of a virtual market, one must actually be built. This is the fifth step in the research program. The dissertation provides a detailed description of a working virtual market now being developed for+ the airline passenger industry, called AirVM. Here marketing agents – buyers and sellers – and choice protocols are defined and calibrated with operational data. Several aspects of virtual markets are explicitly specified, including synthetic populations with attendant incidence probability distributions, client user (avatar) agents, and the application of the virtual market as an instrument in a variety of ways such as for more accurate demand estimation. However, the reader should not assume that value of AirVM as a commercial product has been validated by its inclusion in this thesis. It is discussed here as an example of the process of construction of agent-based models. Agent-based models represent complex processes and require equally complex validation procedures, and that validation process for AirVM is still underway as of this writing.

**1.3.7:** In summary, the result of a five-step research program is described in this dissertation. First, the scientific credentials of agent-based modeling are validated, and the kind of science it can be applied to – namely evolutionary science – specified. Second, that marketing science properly belongs in the domain of evolutionary science is established, with the additional benefits of wider generality for the method. Third, a framework for determining what to model with the agents in order to validly represent human marketing behavior is derived and empirically validated. Fourth, the definition of an agent-based model for marketing research – a virtual market – is offered, which is necessary for the valid construction of a model. And fifth, an actual virtual market is built. These five activities interconnect to present a foundation for the development and application of evolutionary science and agent-based modeling to marketing science.

## 1.4: Contributions to the Literature of this Research

**1.4.1:** There are four major contributions to the literature of marketing science resulting from this work. In addition, there are several minor contributions to the literature surrounding passenger air transport planning and analysis that emerge as a product of the research, and this leads to a fifth, somewhat unexpected general contribution. The most important, in my view, is the demonstration that agent-based models are a feasible and useful instrument in the array of tools available to both the academic and practicing market researcher. Real markets can be modeled using agent-based models, and new discoveries and insights gained by that application, as well as increased forecast accuracy and reliability for practitioners and the management they support.

**1.4.2:** A second contribution is the confirmation of the empirical connection between marketing science and evolutionary science, thus setting the foundation for a sophisticated paradigm that can be applied to marketing science to address problems of heterogeneity and dynamics that have been intractable before.

**1.4.3:** A third contribution is the formulation of the narrative framework, which both empirically and deductively supports the role of *choice* as a basic, indivisible "atom" of marketing science. Moreover, it lays the groundwork for quantitative advances in market portrayal that include what have normally been considered as qualitative endeavors, such as ethnology. It also offers a structure on which behavior outside of what is normally termed "rational" can be described and included in the analysis of market behavior, thus freeing the marketing research from the bounds of utility maximization as the only grounds for validly modeling consumer behavior.

**1.4.4:** The final contribution is the definition of the needed properties and realization in practice of the notion of virtual market. What data is needed, how it can be acquired, and how it all fits together is demonstrated by this concept. Further, the operational and practical aspects of building a virtual market are demonstrated by the construction of an actual, working virtual market simulation.

**1.4.5:** Within the AirVM virtual market, a number of unique contributions to the air passenger industry emerged from the research. They are, to the best of my knowledge, contributions to the literature in this field of endeavor, but are secondary to the general purposes of marketing research. Among these discoveries are: 1) the mixed-logit model of passenger itinerary choice; 2) the traveling group size distributions; 3) the ticketing curve compound, non-homogeneous Poisson stochastic process; 4) a highly efficient sequential sampling method of origin-destination data collection; 5) the imputation method of origin-destination demand estimation and calibration; 6) an advanced form of dynamic flight itinerary generation; and 7) a global market simulation at the individual flight level which encompasses all the relevant dynamics of passenger ticket purchase. These results are of interest, of course, to the airline industry as they struggle to create a viable and stable business model for the future.

**1.4.6:** One other result ranks, in my opinion, as a valuable contribution to the literature, but not in marketing or the airline industry. AirVM demonstrates the value of an agent-based model as an instrument which reveals otherwise hidden characteristics of the phenomenon being studied. That agent-based models can do this has been long known. But to my knowledge this is one of the first cases where agent technology has been applied to study an actual, dynamic human social structure such as a market. Such applications are being published with more and more regularity, and AirVM is a substantial contribution to that growing literature. As more and more results of this kind are developed, the value of the technique and its future application opportunities are significantly enhanced.

## 1.5: Organization of the Dissertation

**1.5.1:** This dissertation contains eight chapters. Of course this introduction and summary is the first. Chapter 2 reviews and critiques the literature supporting the validity of agent-based modeling as a tool for scientific research, the evolutionary science paradigm that the technique supports, and argues that human marketing behavior fits the requirements of evolutionary science, and hence can be validly represented with agent-based modeling. I suggest that this representation of the scientific basis of marketing is more effective and holds more promise than other, more traditional formulations. The foundations of modern evolutionary biology as they contribute to the explication of marketing are examined, thus linking the discussion to current thinking in the development of theoretical marketing science. This supports the position that agent-based modeling is a sufficient tool for analysis in evolutionary science. The discussion offers what I believe to be a compelling case for the central role agent-based modeling can assume in marketing science, even to the point of standing as a fundamental reorientation of how the field is approached. An underlying assertion is that the success of the application of agent-based modeling to marketing science rests on the creation of reasonably accurate models of the behavior of human agents in a marketing context. I must, therefore, delineate what constitutes a "reasonably accurate model." In keeping with the generative nature of agent-based models, I restrict my attention to human behavior that has been rather convincingly tied to human evolution. In the name of parsimony, not to mention generality, I also want to offer a case that marketing behavior is universal. For if marketing arises only from purely cultural factors, then culture must be the focus of study for the development of human agents, and while there is nothing inherently wrong with this, it implies that every culture will have its own unique and separate conceptualizations and realizations, and hence theoretical constructs, which describe what it refers to as marketing. This reality would make the task of creating efficacious agent-based models of human marketing behavior significantly more onerous. Moreover, it is important here to reiterate a fundamental fact of evolution often lost in the popular view. Evolution is *not* the survival of the fittest, but rather the survival of the *least unfit*. Thus behavior that is less than optimal – indeed far less – can be discerned throughout human culture. The admission of less than optimal traits into the constellation of species variation creates opportunities for a huge variety of behavioral patterns that seem to serve no purpose, but likewise exact no survival price. That variety, I argue, is a fundamental property of the marketing behavior exhibited by humans in all societies and, therefore, is the stuff of which any marketing science must be crafted.

**1.5.2:** Chapter 3 describes the narrative framework that supports the modeling of human marketing behavior with agent-based computer simulations. I argue that human beings, as a result of evolution, have an extraordinary ability to construct and utilize patterns. When these patterns meet certain criteria, namely when they are "time-framed structuring patterns of cause-and-effect," then such patterns are the basic apparatus on which concepts of value and resource rest, and out of which human choices are derived. These important characteristics can be derived from narrative structure: 1) how values come about and are expressed in human behavior, (such as in marketing), 2) specifically, how the act of choice-making is the common and central feature of narratives that support market behavior, and 3) how the narrative owner engages resources to affect the outcome of a situation to comply with the requirements of a narrative. The chapter includes an examination of how the narrative framework fits in with empirical observations of universal human behavior in general, and marketing behavior in particular. Since choice is the fundamental unit of the narrative framework, a sampling of choice protocols is offered. These protocols are classified into four broad categories: 1) rational methods, specifically concepts of bounded rationality, 2) heuristics, which are quick and easy (but often very inaccurate) rules-of-thumb, 3) social network protocols, which rely on communications between individuals to make choices, and 4) biases, which are significant errors often found in choice-making.

**1.5.3:** With a framework in hand, Chapter 4 then sets out the definition and important constructs for a virtual market – my term for an agent-based market simulation. A formal, if working, definition of agent is offered, consisting of four essential components – perceptor, ratiocinator, actor and state vector. Using these definitions, then, the notion of virtual market, as an agent-based simulation model of a specific market, including its customers and suppliers, is made precise. Agents as (simulation) client representatives – avatars – are also described. The notions of synthetic populations and incidence distributions are introduced to specify how agent-models are designed and built.

**1.5.4:** For all practical purposes, agents-based models are computer simulations. Therefore, I present a discussion of how an agent-based simulation is implemented on general computing architectures, including an examination of the types of simulation, especially discrete event and stochastic simulation, and general questions of programming, testing and validation. In addition, important computing concepts on which the implementation of agent-based models rest are presented, such as object-oriented programming (OOP), parallel processing, and messaging architecture. Agents are independent entities, and as such can be programmed as objects in an OOP structure. The autonomy and interdependence properties of agents can be easily represented using messaging architecture. This technique is ideal for agent program development, where one agent object can communicate with other agents and with its environment through posting and responding to messages residing in a queue.

**1.5.5:** Chapter 5 applies the virtual market constructs to a specific market – that of passenger tickets in the global airline industry. Called the Airline Virtual Market (AirVM for short), it has been introduced to the passenger airline industry as a market research tool, and is currently being calibrated and validated by several organizations in that market. To begin, the nature and workings of the market for airline passenger tickets is presented in substantial detail. Like any market, there are terms of art and ranges of folklore that need to be explained to those not familiar with the practice. A glossary is included as an appendix to assist those unfamiliar with the terminology of the industry. Following that background, the three important narrative structures of the simulation, those of the airline passenger, of the air carrier, and the ticket sales and distribution systems used in the industry are set out, which provides sufficient detail to build a conceptual overview of the AirVM simulation to give context to the details that follow.

**1.5.6:** The world's airline network is composed of the set of itineraries that can be flown between all city pairs with commercial service in the world. This network is completely connected. That is, a ticket can be purchased for travel between any two cities on earth that are served by commercial air service. It also means that the availability of tickets on any given itinerary is unpredictably connected to the availability of tickets on other itineraries. Because flights are shared among many itineraries, and have finite capacity, when a leg is full it makes all itineraries for all city pairs that use that leg no longer available for additional passengers. The filling of airplanes occurs over time as departure approaches, and the sale of tickets is a random process, so the removal of itineraries from the network is also a random process. It is this dynamic variability in leg availability that makes the airline market simulation so amenable to agent-based modeling, while also so intractable to other forms of analysis. The revenue of an airline is inextricably intertwined with the dynamics of ticket sales and the revenue management behavior of its competition, with fare classes opening and closing unpredictably throughout the ticketing process as a result of ticketing levels in parts of the network that are remote from the particular leg at hand. AirVM represents all city pairs in the world. Each simulation run represents an allocation of every passenger in the world to every regularly scheduled flight in the world across a time period of one calendar week.

**1.5.7:** Chapter 6 then discusses the properties of the synthetic populations for the virtual market, and describes in mathematical detail the several incident distributions necessary to generate the respective synthetic populations. Several sections in this chapter are given over to the descriptions of aspects of the passenger agents, or pags, since the behavior of the passenger agent set is the primary focus of the AirVM. This includes a description of the pag itinerary mixed logit discrete choice model, stochastic processes which represent ticketing rates over time as departure approaches, travel party sizes, and the distributions of ideal departure and arrival times.

**1.5.8:** Attention now turns to the actual simulation computer program itself in Chapter 7. I first describe its current computing architecture, input data requirements, operating characteristics (speed, memory requirements, etc.) and the nature of the user perspectives and portrayals. The simulation executes in two phases. The first is initialization where the synthetic populations for that simulation run are created. These pags are assigned basic characteristics in accordance with the stochastic processes that represent those characteristics. The initialization also defines the revenue management protocols for the arasags, and the airline itinerary networks for the dsags. The actual execution of a simulation occurs after initialization. The execution sequence is roughly as follows. 1) Selection of the pags, determination of the size of the travel party, trip purpose, and the setting of a booking date and time for that simulation cycle. 2) Posting time-of-day messages, randomly selecting the order in which synthetic populations will be polled, and managing pag responses within a synthetic population. 3) If a pag responds to a timing message, the ticketing message process is initiated and continues to run asynchronously until completed. 4) At the end of each repetition of the 127-day cycle[4], results data that has been specified by the simulation user is written to a results data base. The cycles can be repeated as many times as desired. It is possible to program changes in the network (added flights, new pricing, etc.) between cycles. The execution time of a simulation depends on the number of processors available. With 4 processors, 27 million pags (equaling approximately 42,000,000 passengers), 1028 arasags (airline agents), and one dsag (ticket distribution agent), a typical simulation execution cycle requires approximately one hour of real time. The user operation of AirVM is illustrated with a variety of snapshots of the screens used to control the simulation's execution and view the results. After illustrating how it works, how the simulation can be applied by various clients – airlines, airports, regulatory authorities, the financial community, aircraft manufacturers, and industry data supplier – is discussed. These include the exploration of the effects of service changes on air travel demand in a broad variety of ways – anticipated loads, potential revenue and profitability, airport traffic volumes, carbon footprint effects, to name but a few. The validation and calibration analyses that are currently underway to establish the accuracy of AirVM are also delineated. But I must reiterate the warning offered earlier: The discussion of AirVM in this thesis in no way establishes its empirical validity. That question is still be examined, as I will discuss in due course.

**1.5.9:** The final chapter, Chapter 8, summarized the results of the research. The degree to which the research program and objectives have been met are examined. Details are

[4] Why this is 127 days has to do with ticket purchase behavior, and is explained in detail in Chapter 7.

offered of the contributions to the literature are spelled out. And areas of further research suggested by the research program results are offered.

## 1.6: Perspective

**1.6.1:** To the greatest extent possible, in each of the important arguments of this dissertation I have attempted to identify first principles and build the analysis from there, citing appropriate authorities and/or direct evidence to support the inferences. At times the reader may feel this takes the discussion rather far from marketing, into areas of philosophy or mathematics not usually encountered in marketing science literature. But I feel strongly that one should take great care in formulating one's arguments in a field like marketing, the scientific heritage of which is young and still emerging. In my opinion, such care has not been taken in other, similar scientific applications in the business domain, at substantial cost to society and individuals. As a specific example, I offer the current state of "financial engineering" which has fostered the development of "advanced" concepts using probability theory for the creation of instruments like derivatives and credit default swaps. These investment instruments are based on models which are ultimately grounded on application of the Central Limit Theorem of probability theory. But the assumptions of the Central Limit Theorem are not met by the phenomena being modeled, and so the resulting instruments are *prima facie* invalid. The cost of society of this inadequacy is now measured in several trillions of dollars. This is neither science nor ignorance – it is stupidity. So I beg the reader to bear with my pedantry for the sake of demonstrable validity or invalidity, as the case may be. (To the extent that continuity of the presentation is not affected, much of the detail of this fundamental material is reserved for the appendices.)

**1.6.2:** Regarding the format of the text, as is evidenced by this chapter, I use a numbered paragraph structure. This is efficient for internal representation and referencing, while, in my opinion, creates a minimum of distraction for the reader. A paragraph number is usually in and **X.Y.Z** format where **X** is the chapter number, **Y** the section within the chapter, and **Z** the paragraph number. In the rare cases where a paragraph needs sub-paragraphs, a fourth position is used. The only exception to this scheme is in Chapter 7, where italics are used in paragraph number formats to indicate corresponding flow chart tasks. Figures and tables are numbered consecutively within chapters, such as Figure 1.1 or Table 2.2. The appendices found at the end of the main text are identified with letters A, B, C, etc. I have also elected to write in the first person when I describe those portions of the work that is wholly mine, to distinguish it from the many, many contributions to the effort by others. By thereby signifying areas that are my sole responsibility, however, I remove any concern from my colleagues that they bear any fault for the errors, inconsistencies and inaccuracies that are bound to be found in this text.

**1.6.3:** The author of this work is Senior Scientist and Partner in the firm of Virtual M1nds, SA, of Vevey, Switzerland, and AirVM is a commercial product owned by Virtual M1nds and currently in use within the airline industry. However, who uses it, how it is applied by specific organizations to various problems, and the technical details of its application in these multiple contexts are generally protected by nondisclosure agreements or other intellectual property vehicles. Similarly, the calibrated values of the empirical parameter of many of the models used in these applications are trade secrets or otherwise not in the public domain. These are legal agreements that restrict access to some details of the workings of AirVM, and they must be respected in this discussion. Nonetheless, the concepts on which AirVM is based, the stochastic processes that are engaged, and other important conceptual structures that are applied to the virtual market simulation come from published works or activities of my own carried out in the preparation of this thesis. However, the reader will occasionally see notations to the effect that a certain parameter value, for example, cannot be represented as used by a particular customer. Also, some AirVM contributors are identified in this dissertation, but for others this would be a violation of the arrangement between the customer and Virtual M1nds. Only data and results from identified sources, specifically the Airline Reporting Corporation and The Boeing Company, are used in the discussion.

**1.6.4:** Finally, patent protection for several of the unique aspects of AirVM has been applied for under the intellectual property laws of the United States. The United States was chosen for this filing because its intellectual property law extends patent coverage to business processes, including computer software and systems, to a greater extent than that of other countries. This approach, as opposed to trade secret or other methods of intellectual property protection, was chosen to allow full disclosure and peer review of the concepts and methods created and implemented as part of the development process of AirVM.

# Chapter Two:

# Agent-Based Models and Marketing Science

## 2.1: Introduction

**2.1.1:** The topics of this chapter are the first two steps in the research program. First, is agent-based modeling a legitimate form of scientific analysis, and if so what kind of scientific paradigms is it suitable for? Second, does marketing fit the scientific structures for which agent-based modeling is suited? If these two propositions can be answered in the affirmative, then I can assert that this approach to representing the scientific basis of marketing is more effective and holds more promise than other, more traditional formulations. Thus I must outline the ontological and epistemological justifications for the use of agent-based modeling. It will be asserted that the method is particularly appropriate for the treatment of evolutionary sciences, and therefore the foundations of modern evolutionary biology as they contribute to the explication of marketing must be examined, thus linking the discussion to current thinking in the development of theoretical marketing science.

**2.1.2:** Before these claims can be proposed and defended, they must be made clear. I begin with a brief discussion of modeling and simulation, setting a context for a definition of agent-based modeling, both to build some intuition regarding the subject and to set some important concepts necessary to proceed. For reasons that will become evident, agent-based modeling is usually classified as a form of computational science, a conceptualization that will be reinforced using the arguments of Humphreys (2004). Humphreys makes a compelling case that computational science is a fundamental and unprecedented improvement in the set of instruments available to mankind for scientific exploration, and agent-based modeling, as a subset of computational science, substantially contributes to this emerging human extension.

**2.1.3:** I will then delineate the outlines of evolutionary science. I will naturally draw upon several scholars from the field of biology for direction, primarily Ernst Mayr

(2001, 2004), and Steven jay Gould (2002). This will reinforce my position that agent-based modeling is a sufficient, if not necessary, tool for analysis in evolutionary science. This position is also maintained by Shelby Hunt (2001). Hunt's conception of marketing science as an evolutionary science is based on economic principles and concepts delineated by Hogdson (1993). In making that connection, I agree with Hunt, but it is necessary to challenge him in some significant respects, since his conceptualization seems significantly weak in some areas. In sum, the discussion will offer what I believe to be a compelling case for the central role agent-based modeling can assume in marketing science, even to the point of offering a fundamental reorientation of how the field is approached.

**2.1.4:** Attention then will turn to the relationship between human evolution and human behavior. Agent-based models in marketing must represent, as faithfully as resources will allow, the actual behavior human beings engage in when they are participating in marketing activities. I contend that the existence of universal human behaviors supports the position that observed behavior is grounded in evolutionary history, and scientific tools which can recognize that heritage are more powerful than those which do not. Moreover, if all human cultures contain a set of human universal behaviors, then a structure which portrays these universals can be applied to any marketing context. That is, there is more nature than nurture in human behavior, and agent design is not uniquely dependent on culture.

**2.1.5:** Surely not all observed behavioral attributes can be traced back to evolutionary origins, so what of the behavior of people that is important to marketing activity? Cialdini (2001) offers a description of seven behavior patterns that are characteristic of people when found in a marketing context. I will summarize his taxonomy, not so much because I intend to model each characteristic, but more to show that these elementary behaviors can be traced to our evolution, and therefore comply with the requirements of evolutionary science, and thus can be expected to be tractable with an agent-based tool.

## 2.2: Modeling and Simulation

**2.2.1:** Human beings largely think in terms of mental models of the world around them and their relationship to it (Kunda, 1999). We formulate concepts and ideas and link them together to represent the way we think the world works in some significant regard, and use those representations to make decisions on our future actions. These models can range from simple statements of assumed cause and effect – "if I step out in front of a moving bus, there's a good chance I could be seriously hurt" – through physical scale models of buildings or vehicles in their design stages through mathematical representations of complex social or physical systems. Common to models of whatever composition or subject is that they are *abstractions*, and therefore *simplifications*, of reality, retaining what is believed to be salient features of the problem at hand.

**2.2.2:** I am only concerned with models that are represented mathematically, either in the form of one or more equations or as a computer program. Most marketers and marketing researchers will be familiar with some form of simulation as applied to marketing. Consider the simple "What if?" scenario analysis applied to the results from a conjoint analysis, or more elaborate "war games" in which competing teams of managers (or MBA students) develop and implement strategies in an interactive fashion. As a concept, simulation has almost as many meanings as applications but a central, underlying principle is the representation of one process or set of processes with another, usually simpler in some useful sense, process or set of processes. The Monopoly® board game, for example, simplifies a rather complex economic system into a limited set of transactions. Part of the fun of playing the game derives from the degree to which the outcomes (wealth accumulation and bankruptcy) resemble the outcomes from the real world process that is being simulated.

**2.2.3:** An important concept inherent in this process orientation to simulation is *time*. Processes can be viewed as actions and reactions of components *over time*. This temporal property is an essential characteristic of process-representative simulation, and differentiates the application of simulation as an analytic and scientific tool from many other approaches, such as deductive logic or statistical inference. The central role of time is also an integral part of agent-based model simulations. With an express representation of time, the *dynamics* of a system can be explicitly studied.

**2.2.4:** There are compelling reasons to use simulation to study a wide range of problems. In many situations the scale of the process under study prohibits any other approach. In astronomy and astrophysics, the universe is not available for experimental manipulation, but a simulation of important aspects of it are. Similarly, explorations of cultural development or species evolution often cannot be carried out in a physical laboratory environment, whereas a simulation permits hypothesis testing and inference on a reasonable time scale. In many circumstances, the phenomenon under investigation cannot be ethically or safely subjected to experimentation. The study of disease epidemics, social intolerance and military tactics are obvious examples. Finally, some systems are so complex that traditional experimental science seems hopeless as a research approach. Among these systems are ecological dynamics, evolutionary economics, and the subject of interest here – marketing.

**2.2.5:** Figure 2.1 lays out an "ethnology" of mathematical modeling in general and simulation modeling in particular, starting with the invention of the calculus in the 1700's and continuing up to the present day. This diagram is adapted from Gilbert and Troitszch (1999). The two broad categories of stochastic and deterministic simulation models are indicated by the shaded ovals. The bold face labels define the mathematical contexts of the various modeling formalizations, while their genealogy is spelled out with the lines. As illustrated in this diagram, agent-based simulation belongs to the class of stochastic simulations and descended from a form of simulation called *discrete-event* simulations. Discrete event simulations are (stochastic) simulations that attempt to mimic the behavior of the discrete parts of a system and their interaction. As indicated in Figure 2.1, a number of other varieties of simulation have evolved from the discrete event form.

**2.2.6:** Also contributing to the heritage of agent simulation are game theory, artificial intelligence and cellular automata models. Some of the early applications of agent-based models were in the area of rational game play. Axelrod's (1997) experiments with the prisoners dilemma game is the classic example, and is cited by virtually every practitioner of agent-modeling somewhere in their writings. Artificial intelligence models a key component of the development of robotic systems, (Wooldridge, 2002, is a chief proponent). Cellular models are an extension of the work by Wolfram on cellular automata. In fact, Wolfram argues that essentially all modeling can be cast in this form (Wolfram, 1994). That may be true, but then, in the abstract, all models can exist in an infinity of forms.

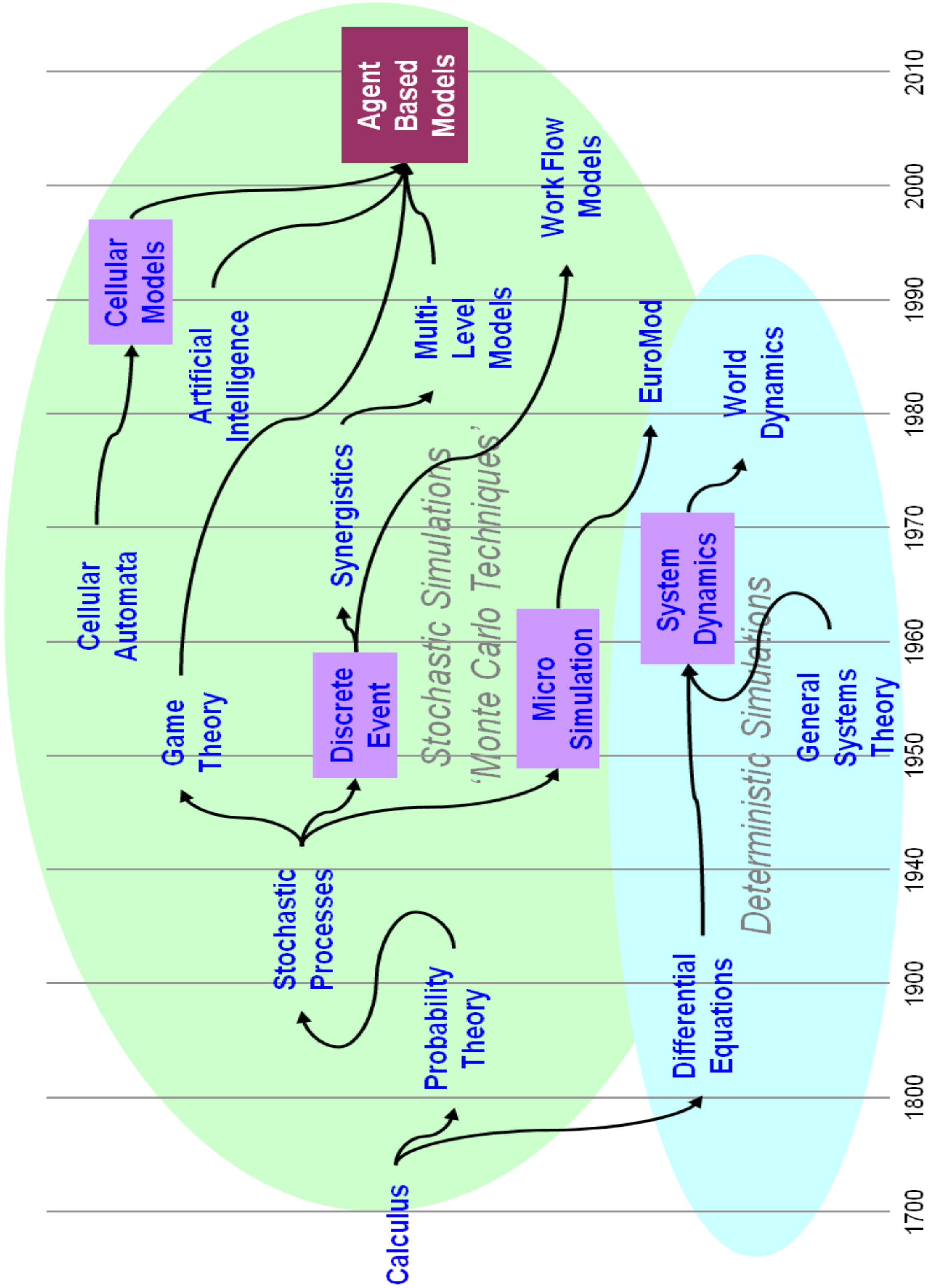



**Figure 2.1: The Ethnology of Mathematical Modeling and Simulation.**

**2.2.7:** If the behavior modeled by the discrete event simulation contains stochastic elements, as it usually does, then the simulation can be run repeatedly using random numbers generated by the relevant probability distributions, and the distributional characteristics of the resulting dynamics portrayed. In fact, for a system of even moderate complexity, this is the only way such dynamics can be validly studied, since the ability to express dynamics through ordinary or stochastic partial differential equations can quickly exceed any reasonable closed form.

## 2.3: Reductive and Structural Models

**2.3.1:** Mathematical models can be classified into two forms: *reductive* models and *structural* models. In a reductive model, a set of data is reduced to one or more equations or similar relationships that represent the data, but in a simpler, more parsimonious form. There is some loss of accuracy, which is the price of greater understandability. Linear regression is a ready example, where a set of pairs (or *n*-tuples) of data is replaced with an equation representing the relationship between the data values of one element of the set, called the dependent variable, as a linear combination of some collection of the other (*n-1*)-tuples, called independent variables. Structural models, on the other hand, are intended to be representations that not only reproduce patterns of observed data, but also characterize the process, or structure, by which the variables represented by the data relate to one another. Reductive models are not required to completely replicate the process by which the observed data is generated. They need only reproduce the observed results in a more economical and parsimonious way. Structural models focus on the way in which the observed values come about.

**2.3.2:** A good deal of applied mathematics and a substantial part of modern statistics is aimed at easing the problem of the generation of reductive models. I mentioned linear regression, but there are literally thousands of other applied techniques; Fourier series (Jackson, 1941), time series analysis (Grenander and Rosenblatt, 1952), discrete choice modeling (e. g. Ben Akiva and Lerman, 1985, Louviere *et al.*, 2000), and proportional hazard models (Cox and Oakes, 1984) to name only a very few. Indeed, the so-called Stone-Weierstrass theorem in classical abstract analysis (Hewitt and Stromberg, 1965, pp 94-98) describes a set of general conditions under which an arbitrary data set can be approximated to any desired degree of accuracy by a broad class of simpler functional forms. And reductive models are powerful and extremely useful tools. In econometrics, for example, reductive models are widely used for forecasting (Auffhammer and Carson, 2008, Parker, Cenesizoglu, and Carson, 2005). For a thorough treatment of the structures and formal aspects of mathematical models, see Chang and Keisler (1973).

**2.3.3:** But in the "harder" sciences like physics, reductive models give way to structural models. Newton's laws of motion, or Einstein's theories of relativity, are mathematical constructs which purport to not only represent the results of data sets arising from observations of natural phenomena, but also how the objects in the natural world interact with each other to generate the observed data. That is, the models don't merely represent the observations, but also describe the *process* by which the observations come about. As such, they carry more explanatory weight than reductive descriptions. They are more likely to be valid beyond the range of the initial observed data sets that lead to their formulation. Also, they have repeatedly been shown to be robust across varied data sets and when connected to other models which represent related systems.

**2.3.4:** Simulation is one of the more powerful structural modeling techniques. A simulation is a model which, by design, represents the relationships between entities in a system. In particular, a simulation captures the dynamics of the relationships between components of a system, reflecting how changes in one component create changes and affect the response of other components. Humphreys (2004, pp 108-109), in fact, defines a simulation as a structural model that explicitly includes time as a dimension, so that the dynamics between the variables described in the model can be appropriately portrayed.[5] The increasingly wide spread use of simulation also is tied to the development of computing power.

**2.3.5:** The literature on simulation is very broad. Thompson (2000) gives a general introduction, while Gilbert and Troitsch (1999) discuss modeling in the social sciences, discussing the genealogy of various modeling approaches leading up to simulation. Note that the use of the word simulation here as a modeling technique should not be confused with the phrase when applied to certain methods of finding solutions to equations. For example, computing the volume of a complex solid in a multi-

[5] He uses time as the dimension carrying the dynamics, primarily because of its unidirectionality. Other dimensions, such as spatial coordinates, could also be used, but they don't necessarily have this unidirectional property.

dimensional space, such as the area under a multi-dimensional normal probability distribution, can be done by generating a very large number of points in the relevant space and determining the ratio of those in the solid versus those not in the solid. This technique is sometimes referred to as simulation, as in Train's *Discrete Choice Methods with Simulation* (Train, 2003).

## 2.4: A Definition of Agent-Based Modeling

**2.4.1:** Agent-based modeling uses cheap computing power to represent the interactions of a large number of relatively simple entities called *agents*. Agents are computer programs (more precisely, *objects* in computer programs) which are designed to duplicate the important and salient behavior of entities observed in the real world. The literature uses a number of terms for the agent-based modeling, reflecting both its many origins and diverse applications. These include agent-based models (e. g. Bonabeau, 2002, Epstein and Axtell (1996) and others too numerous to mention), agent-driven models (Parker *et al.*, 2004), agent-based simulation (Hales *et al.*, 2003), more recently generative science (Epstein, 2006), and, in ecology, individual based modeling (Grimm and Railsback, 2005).

**2.4.2:** Joshua Epstein can be legitimately considered as one of the founding fathers of agent-based modeling. Epstein and Axtell's groundbreaking (and fun!) work, *Growing Artificial Societies: Social Science from the Bottom Up*, (Epstein and Axtell, 1996), is one of the first treatises that brought the full power of agent-based modeling to a social science context. Epstein has emerged as a leading figure in applications of the methodology in the social sciences, which includes economics (and, in my view, marketing). Others approach the topic from different perspectives, such as Wooldridge (2002) for robotics or Kennedy and Eberhart (2001) for so-called "swarm intelligence."

**2.4.3:** Disappointingly, Epstein never sets out a succinct definition of an agent-based model *per se*, but rather puts forth a set of attributes that a computational model will have if it is properly considered as agent-based.[6] By implication, agents are computational objects, although there is nothing essential in the concept of agent that requires the existence or application of computers. So consider an agent as a computer program that represents, to a greater or lesser degree of realism, the behavior of some

[6] He is not alone in this reluctance. The field is perhaps too young for an exclusive and exhaustive definition, so property specification seems the next best thing.

entity of interest, such as a consumer. The properties recited by Epstein (2006, p. 51-52) necessary to qualify such a program as an agent are these:

**2.4.3.1:** *Heterogeneity*: Agents are characterized by attributes of relevance to the phenomena being modeled. For example, they may have mass, position (in space), wealth, strength, agility, life expectancy, or any of a number of properties. By *heterogeneity* is meant that different individual agents have different levels of these distinct and defining characteristics. This is in contrast to the normal, macroeconomic assumption of homogeneity across consumer segments or among population members. Thus they cannot be replaced by "representative agents," as is assumed for example in the argument of Anderson, de Palma and Thisse (1992). Not only can the agents differ in terms of defining features, they can also differ in their how they interact with other agents. How they communicate, respond to communications from others, and interpret "language" all can vary from agent to agent. This heterogeneity is in stark contrast to many of the assumptions made in most marketing research investigations. Being able to manage – indeed capitalize on – this property of agent-based modeling forms a compelling argument favoring its application to marketing science.

**2.4.3.2:** *Autonomy*: Agents are not subject to central, top-down control. That is, to at least some degree they are assumed to act independently of any other agent or force. Of course they interact and change (in response to the behavior of other agents or conditions in their environment), and a centralized control may develop as a product of that interaction, but no hierarchical structure is set in place at the definition stage of an agent-based model.

**2.4.3.3:** *Explicit Space*: Agents exist in a known and explicit *environment*, which has a specified, sufficiently complete topology[7]. The environment can contain, of course, other agents, or entities which are not agents (that is, which have no "behavior" as such). This environment can exist in ordinary one, two or three-dimensional space, in an *n*-dimensional lattice, in a network with various structures of nodes and edges, or in a more obscure topological construct, such as a three-dimensional continuous space replete with singularities, or 11-dimensional

[7] Epstein offers no definition of what a "sufficiently complete" topology is. It would seem that any topology would do, since other than a notion of nearness, nothing else is required. Surely a metric space would be sufficient. See Hocking and Young (1961, pp 9-12) for an introduction to the mathematical theory of topology.

abstract space such as posed for string theory. As Epstein points out, however, "The main *desideratum* is that the notion of 'local' be well posed." (Epstein, 2006, p. 6) The specification of the environment within which the agents operate is one of the critical features of an agent-based model, sometimes not accorded the attention it demands. Indeed, I find Epstein's treatment of this topic quite elementary.

**2.4.3.4:** *Local Interaction*: Agents interact with other agents and entities which are "neighboring" in the Explicit Space. Further, the nature and expression of these interactions are completely specified by the agent-based model. Indeed, Epstein argues (Epstein, 2006, p 36-37) that the specification of an agent is incomplete if its interactions with other agents and entities in its environment are not completely specified.

**2.4.3.5:** *Bounded Rationality*: Epstein asserts that agents must have some volition and decision-making capability (i. e. ratiocination), but that this capacity has limitations. The limitations are along at least two dimensions – bounded information and bounded computation. Coupled with heterogeneity, this also implies that a) the actual bounds along either dimension may differ from agent to agent, b) those bounds can differ at different points in the Explicit Space, and c) they can change in response to Local Interaction. A number of scholars have contributed to the examination of bounded rationality, among them Simon (1957, 1996, 1997), Tversky (1972), Gigerenzer (2002), Gigerenzer and Selten (2001), and Lupia *et al.* (2000), and Rubenstein (1998). It should be noted that the inclusion of rationality (however conceived, bounded or otherwise) among the criteria which characterize agent-based models differentiate them from the class computer models referred to as *micro simulations*, as, for example, of subatomic physics models, cosmological models of the universe or near-term weather forecasting models. Louviere (2001) notes the empirical inconsistency of consumer choice, which can be considered another form of bounded rationality.

**2.4.3.6:** *Non-Equilibrium Dynamics*: Epstein writes "Non-equilibrium dynamics are of central concern to agent modelers, as are large-scale transitions, 'tipping phenomena,' and the emergence of macroscopic regularity from decentralized local interaction. These are in sharp distinction to equilibrium existence theorems and comparative statics …" (Epstein, 2006, p 52.) This is a crucial *raison d'etre* for Epstein's case – the challenge to the poor success historically demonstrated by classical economics, a situation that can be substantially improved with the application of agent-based modeling. Thus he focuses on economic behavior that may not necessarily lead to a Nash equilibrium, or at least cannot come to such a condition in a reasonable amount of human time. See Epstein and Hammond (2001) for a specific example. Marketing science can only welcome this view. In a world which is in economic equilibrium there can be no marketing. By definition, an equilibrium state exists when there is no advantage of one product over another. All consumers are as satisfied as they can be given the current constellation of conditions. Thus, from the standpoint of explaining marketing phenomena, the exploration of non-equilibrium conditions is essential.

**2.4.4:** There are some aspects of agent modeling not discussed by Epstein that will become important as agent-based modeling moves into the marketing science mainstream. *Adaptation* is the ability of an agent to modify its behavior in response to changes in its surrounding environment, which obviously includes the behavior of other agents in that environment. Adaptive agents can, at some elemental level, "understand" what is happening around them and take steps to change their response to environmental stimuli given that level of understanding. For in-depth discussions, see Holland (1992, 1995, 1998). Adaptivity can come in a wide variety of forms. Another way an agent can develop and utilize adaptive capability is through *learning*. Modeling learning behavior is one of the more active areas of agent research. Deeper than just simple adaptation is what Holland (1992, p. 197) terms *second-order* adaptation. This is the situation where the agent not only is adaptive, but *knows* it is adaptive and can modify its adaptation capability to meet its changing needs.

**2.4.5:** The general field of agent-based modeling continues to explode. Since it ultimately relies on computing power, interest in the field is directly proportional to the increasing availability of inexpensive computational capability. Applications range from robotics to artificial societies to evolutionary economics to computational biology. As a new tool, efforts are underway to apply the technique in a number of disciplines. Tesfatsion (2009) maintains a web site which maps out as well as any we have found the full scope of what constitutes agent-based modeling[8]. As might be expected, all kinds of theories and applications are being put forth as scholars from a variety of

[8] Although it had not been updated since 19 January, 2009, when viewed in late May of 2009.

disciplines try out the new methods and techniques to see if this tool set can address and successfully solve longstanding issues in their fields. An examination of the scope of the literature gives some insight into the breadth of the field. Some of the more general and abstract, and therefore more widely applicable, discussions include: Kennedy and Eberhart (2001), who describe “swarm intelligence,” or the emergent results of the actions of a large number of independent and simple entities to solve otherwise intractable problems. Wooldridge (2002) discusses agent theory as the basis for robotics. Rubenstein (2004) lays out a mathematically formal description of agent bounded rationality. Axelrod (1997), offers a description of wide-ranging applications of agent modeling to social systems theory. Gilbert and Troitzsch (1999) examine simulation in the social sciences and fits agent simulation into that context. Epstein and Axtell (1996) create challenging, almost cute, agent simulations that demonstrate how complex social phenomena, including markets, can arise from simple agent behavior

**2.4.6:** These works rely to greater or lesser degrees on fundamental characteristics captured in the central themes of Holland (1992, 1995, 1998), regarding emergence and complex adaptive systems, Minsky (1985), Minsky and Papert (1969) on the human brain as a complex system, and Wolfram (1994) in his fundamental work in cellular automata, a specific form of agent model with perhaps deep implications on the nature of scientific endeavor. Approachable discussions of emergence, complexity and chaos can also be found in Waldorp (1992) and Johnson (2001). Herbert Simon created the concept of bounded rationality with his notion of ‘satisficing” (Simon, 1957) and gives a succinct account of the implications of these ideas to complex systems analysis in his later work (Simon, 1996). Examples of recent work in agent-based modeling in the social sciences include Conte and Paolucci (2001) on social learning and Sun and Naveh (2004) addressing organization theory. Other examples indicating the breadth of the subject include Andersson and Andersson (2009), Bone, Hey and Suckling (2007), Oluyomi, Karunasekera and Sterling (2007), and Greiner and Flaschel (2009). The field is clearly a young one, rich with promise, and the output of work in the field is truly stupefying.

**2.4.7:** The literature on the uses of agent modeling in economics, especially for the analysis of market behavior, is also evolving rapidly. Epstein and Hammond (2001) and Arthur (2006) illustrate how agent-based modeling can address economic systems where it can be shown that no equilibrium condition can exist. At a more practical level, Luna (2000) discusses the application to economic issues of SWARM, a software agent modeling capability developed by the Santa Fe Institute. Deguchi (2004) examines the problem from a formal perspective, and offers a comprehensive presentation of classic economic issues in the agent modeling context, and Bernard (1999) describes an adaptive agent-based analysis of urban policy development (specifically rent control). Valente (1999) offers a complete programming system to implement economic agent simulations. Lude and Tesfatsion (2006) have published a significant compendium of computational economics literature.

**2.4.8:** However, I have not yet found specific and substantive works in the economics literature that moves easily into the marketing world. In fact, citations of the development or application of agent modeling in the marketing science literature are sparse. A search of the sources up to the end of 2008 reveals no example of an application of the methods to an actual, real world marketing problem. There have been a small number of discussions of how the techniques *might* be used in marketing research, however. Parker and Bakken (2005, 2007) describe agent modeling in general terms, review methodological and theoretical issues, and discuss how agent simulation might be used to attack certain marketing problems, with a couple of hypothetical examples. Articles by Goldenberg, Libai and Muller (2004) and Lusch and Tay (2004) also discuss potential applications in marketing, but again offer no actual examples. Garcia (2005) suggests the use of agent-modeling techniques for problems of market diffusion and resource allocation, but presents no empirical demonstrations. Parker and Lonsdale (2007) present early work supporting AirVM, an agent-based model discussed in depth in later chapters of this dissertation, as does Parker and Perroud (2008), and Parker (2008). More recent works are now entering the field, however, so the list above is by no means exhaustive.

## 2.5: The Epistemological Basis of Agent-Based Models

**2.5.1:** The epistemological support for agent-based modeling is quite straightforward, but should be explicated nonetheless to assuage the sensibilities of those less familiar, but no less worried, about the scientific basis for a method for which the terminology of many of its practitioners (Epstein and Axtell, 1996, in particular), makes extensive use of the word *artificial*. Agent-based models exist as computer programs, often very large and complex ones. But any computer program that actually executes without error is

Turing-computable, and any Turing-computable algorithm (or realization of that algorithm) is tautologically a true well-formed formula (or wff) (Hughes and Cresswell, 1968, Kleene, 1971). In fact, any computer program is absolutely equivalent to some partial recursive function (Hodel, 1995). Now certainly the wff represented by an agent-based model can be long and complex, probably unintelligible to most people. But it is a wff nonetheless, and if proved true is therefore a proved mathematical theorem. In fact, due to the way agent-based models are designed and implemented (from the bottom up, as it were) the observed macro behavior (emergent behavior) of an agent-based model is proved by construction.[9]

**2.5.2:** Thus, in a precise mathematical sense, the large-scale, system properties of an agent-based model are deduced from the micro-specification of the agents it is based on, and so the properties of the agents are *sufficient* for the macro-behavior of the model, but not *necessary*.[10] This means that if large scale behavior is demonstrated from the interactions over time of agents with given properties and interaction structures, then it can be safely asserted that those micro-behaviors can cause the macro result (as in the analysis of wealth distributions in Epstein and Axtell, 1996, pp 32-42). But it does not mean that there are not other agent proscriptions – indeed other mathematical models having no concept of agency – that will also lead to the same macro-level phenomena. Epstein (2006, p. 58) points out (somewhat optimistically) that in the face of competing agent-based models that lead to the same observed macro-behavior, the one which is most appropriate becomes a well-defined empirical question, to be decided by experimentation and observation.[11]

**2.5.3:** Finally, I need to consider the problem of uncertainty: Is the biological-social-economic world inherently stochastic? My answer is unequivocally 'yes.' Quantum forces, if modern physics is true, generate random churning at the sub-atomic level. But that is only apparent at the extremely small scale. The effect smoothes as a more natural human scale is reached. But the random effect that lies below the observable fabric can

[9] For the reader who's experience has taken him far away from his college mathematics training, recall that a theorem can be proved a number of ways, including by *construction* (building the theorem from other true statements), by *induction* (a somewhat different form of construction), and by *contradiction* (showing that if the theorem is not true, then some other known true theorem cannot be true).
[10] Again, for the mathematically rusty, $A$ is termed sufficient for $B$ if $A$ implies $B$, and $A$ is necessary for $B$ if $B$ implies $A$. And, in the opinion of some, "necessary and sufficient" is somehow more elegant than "if and only if."
[11] Epstein comments sardonically that it is more often than not the case that only one agent-based model available, much less two or more, so the issue of competing models is a luxury. (Epstein, 2006, p. 54)

create results that are clearly apparent at the larger scale. For example, fundamental nuclear uncertainty may trigger a radioactive particle decay that knocks an electron out of its normal shell, precipitating a DNA mutation, which in time triggers an evolutionary effect that eventually changes the overall fitness of a species. Alternatively, a nearby star could supernova, at its root a sub-atomic phenomenon governed by quantum physics, which leads to a sharp increase in interstellar radiation and thereby substantial, and sudden, changes in the earth environment. It would not be unexpected to see a number of species sharply affected by such an event. (Science fiction relies on such scenarios for many of its dramatic settings.)

**2.5.4:** This raises the important issue of what is referred to in the complex adaptive system literature as *path dependency*. Causal chains – one cause leading to a given effect, which yields another effect, and so on – are the threads that form the fabric of the world, or at least our understanding of it. In most real world observed conditions, there are in theory an uncountable infinite number of causal chains that can be described.[12] It is inevitable that when I observe a causal chain, I always pick it up in the middle, in some sense. There were events that led to the state I find when I first encounter the object of study, and I do not know what they were, so it is epistemologically impossible to lay out, in complete detail, the event structure that created the configuration of events I see unfolding.[13] Thus, in the general case, I must also ascribe a stochastic property to all phenomena which contain sufficiently rich causal chains. This means I cannot forecast exactly the future state of a causal chain. There is no Laplacian Daemon able to discern the past and the future in complete detail because of its superior computing capability. The best I can do is assert a probability distribution structure for any useful set of paths. Agent-based modeling is an instrument that allows me to explore that distributional structure.

**2.5.5:** Agent-based models are expressions of computational science, and so are subject to the constraints and opportunities computational science creates. These are important not only for practical reasons (regarding computer hardware requirements, execution speed issues, programming hardware and software architectures, and so on) but also

[12] A given chain only needs a single time-interval component (e. g. when something occurs) to imply that an uncountable number of subintervals are available for alternative paths, since time is a real number and any interval of time is an everywhere dense set, which has uncountable infinite members.

[13] Controlled experiments try to disconnect events from unknown causal chain effects. Unfortunately, such are rare in the social science fields.

from both an ontological and epistemological perspective. Consider the Parker and Perroud (2008) discussion on an agent-based model of airline passenger behavior which routinely simulates the actions of tens of millions of passengers, effectively simultaneously. This is so large that under no conceivable human-based computational structure could an equivalent analysis be undertaken. Thus, in a very substantial sense, the agent-based model is an extension of human observational capability because it is able to dramatically increase the amount of empirically valid data. Each run of the agent-based passenger model would be impossible without the assistance of substantial computing capability which didn't even exist two decades ago! That is, each *observation* would be impossible without the computers. This and many other aspects of computational science are explored in depth by Paul Humphreys in his work *Extending Ourselves: Computational Science, Empiricism, and Scientific Method* (Humphreys, 2004). Appendix A offers a detailed examination of Humphrey's approach, and clearly delineates the value of computational science as an instrument in scientific endeavor. This, of course, substantiates the application of agent-based modeling in any appropriate scientific domain.

## 2.6: Evolutionary Science

**2.6.1:** So modeling is logically sound and is an efficacious instrument for the conduct of science. How, then, can this adroit and awesome instrumentation be brought to bear on issues of marketing science? Essentially, I will pick up where Shelby Hunt left off in his discussion on the ontology of marketing science. He asserts that marketing science is an evolutionary science. What is encompassed by this concept, then, becomes important to the advancement of the theory of marketing science. But to understand evolutionary science one must have some familiarity with evolution, for while there is a distinction to be made between evolutionary science and biological evolution, the latter is the premier example of the former. Thus, while it seems difficult to ascribe a central role of biology to marketing (notwithstanding the frequent use of beautiful women or handsome men in marketing campaigns), the outlines of evolutionary biology must be understood as part of a compelling argument that marketing is an evolutionary science.

**2.6.2:** As a guide to understanding the workings of evolution, I will follow the work of Ernst Mayr. Mayr was an evolutionary biologist, and for him evolution meant Darwinian evolution as first laid out in *The Origin of Species* (Darwin, 1859). Arguably peerless among practicing biologists, Mayr was Emeritus Professor of Biology at Harvard, and produced nearly three quarters of a century of biological research. As he approached his 100th birthday in 2004 he published a delightful text entitled *What is Biology?* (Mayr, 2004) which summarized his views on the nature of the biological sciences in general and the nature of evolution as a scientific concept in particular. This, and another recent work, *This is Evolution* (Mayr, 2001), form a succinct explication of the nature and structure of evolution and evolutionary science. Another more in-depth and detailed exploration can be found in the work of a colleague of Mayr at Harvard, Stephen Jay Gould, (Gould, 2002).[14] In this discussion I will leave out the biochemical foundations of genetics that underlie modern biology, but I will delve into some of the fundamental logic that has emerged in evolutionary theory since Darwin.

**2.6.3:** Most people are somewhat aware of the general ideas of evolution. A species *evolves* as a result of changes in their genetic make-up that pass from generation to generation, subtly improving the ability of the species to survive in the context of the environment it finds itself in. This ability improvement is encapsulated in perhaps the most cited principle of Darwin's theory of evolution – "survival of the fittest." At the core of this position is the central role of survival. Biological evolution is measured on a scale of *species survival* – stronger, better equipped individuals do better than weaker, more poorly prepared ones. Survival is measured almost universally in biology by the number of offspring of an individual that survive (Sober and Wilson, 1998). Survive what? This is generally taken to mean survive the environment in which the individual finds itself. That environment, of course, can include other individuals like itself (of the same species) or completely different (of other species), as well as non-biological hazards like fire and weather. And, survival of an individual doesn't count unless it lives long enough to produce offspring. Thus the theory of biological evolution is concerned with the ability of individuals to survive in their current environment long enough to generate offspring.[15]

[14] Gould and Mayr were contemporaries at Harvard. Gould died in 2002 at age 60, Mayr in 2005 at age 100. They both published major discussions of evolutionary science in the first four years of the new millennium. Perhaps as a measure of the value of age, Mayr's texts were two small, succinct volumes, each less than 300 pages in length. Gould's *The Structure of Evolutionary Theory* is a huge book, over 1400 pages long, and anything but succinct. They all sell for roughly the same price.

[15] There are many other measures of fitness that could be applied. However, for this exposition, the measure based on number of offspring is sufficient to illustrate the important concepts.

**2.6.4:** The phrase "survival of the fittest" did not originate with Darwin, but was coined by Herbert Spencer, a contemporary of Darwin (Quammen, 2006, pp 179-180). In fact, Darwin did not believe it, and indeed it is not a valid inference from either classical or neoclassical Darwinian theory. More precisely, Darwin created the concept of the *survival of the least unfit*. The point is effectively illustrated by this bit of humor: If you are in a group being chased through the forrest by a bear, to survive there is no need to be the fastest individual, just not the slowest. This means that a successful population of species contains many, many individuals who have less than optimal fitness, but are quite capable of surviving and thriving.

**2.6.5:** If survival is dependent on the environment in which the individual is embedded, then changes in that environment can have fundamental effects on survival. The biologist assumes that the environment is always changing and thus the conditions under which one member of a species survives in one instant may not be the same as that same individual (or its offspring) may encounter at a later instant. Thus an individual well suited to survival at one time may not be so well off at another, depending on changes in the environment. From this constant change comes the notion of *biological adaptation.* An individual that is more likely to survive in a given environmental configuration is said to be better adapted than one which is less likely to survive. In terms of the progeny measure, the better adapted individual thus will produce, on average, more offspring than the less adapted individual. In biology it is generally assumed that individuals do not change their survivability across the course of their lives, but rather their offspring, due to inherent biochemical processes associated with procreation, can be better (or worse) at survival than those from which they originate. It is this idea of survival by adaptation that is at the core of modern evolutionary biology.

**2.6.6:** Mayr (2001) asserts that three conditions must be met in order for a (biological) system to be considered evolutional.

> **2.6.6.1:** The first is *population***.** There must be a collection of individuals. A single individual is not enough. There is no particular minimum number of individuals required, just that there be more than one, if reproduction is asexual, or more than two if two-parent sexual reproduction is needed. In common parlance, however, the word "population" implies a rather larger group rather than just a pair or few.

**2.6.6.2:** Indeed, this presumption of a larger number is reinforced by the second condition. This is *variation*. There must be variation among the individuals in the population along one or more *traits* of interest. By "trait of interest" I mean some aspect of the individual's behavior, physical capabilities or mental acuity that is relevant to the individual's survival. And the concept of variation requires that different individuals in the population possess these traits to various degrees of strength, proficiency, efficiency, or efficacy. Of course, there is very likely substantially more variation within the population if the population is bigger, supporting the idea that the population should be bigger rather than smaller. In fact, to the extent that adaptation is carried between generations, the bigger the population, the more flexible the adaptation, and the more quickly it can be engaged to respond to environmental changes. This is why infectious diseases like influenza can only be combated with the constant development of countermeasures. A new antidote changes the environment of the biological agent, but there are so many of them, and they reproduce so rapidly, that there is a very high likelihood some members of the species are resistant to the antidote, and their offspring will flourish.

**2.6.6.3:** Finally, Mayr requires *heritability*. There must be a way in which the traits of an individual are passed on to its progeny; the definition of heritability. Further, not only must trait values be passed on to offspring, but there must be a mechanism for creating variation in the offspring trait values beyond what the siring individual(s) contribute. This last requirement is referred to as *heritability variation*.

Summarizing briefly, there must be a sufficiently varied population (with respect to the traits that promote survivability), and that variation must be constantly renewed. Without such continual variation, a species will not survive a capricious environment where a change can put an otherwise marvelous survivor at high risk.

**2.6.7:** An important distinction Mayr emphasizes is the focus on *population variation* as the basic unit in the analysis of evolution, in contrast to the "typical" individual, as held by what he refers to as the "typologists." Succinctly:

> "The assumptions of population thinking are diametrically opposed to those of the typologist. The populationist stresses the uniqueness of everything in the organic world. What is true for the human species – that no two individuals are alike – is equally true to all other animals and

> plants. Indeed, even the same individual changes continuously throughout its lifetime when placed into different environments. All organisms and organic phenomena are composed of various features and can be described collectivity only in discrete terms. Individuals, or any kind of organic entities, form populations of units. We can determine the arithmetic mean and the statistics of variation. Averages are merely statistical abstractions, only the individuals of which the population is composed are reality. The ultimate conclusions of the population thinker and of the typologist are precisely the opposite. For the typologist the type (*edos*) is real and the variation an illusion, while for the populationist the type (*average*) is an abstraction and only the variation is real. No two ways of looking at nature could be more different." (Mayr, 2001, in quoting himself from 1959, page 84)

The implications for the study of marketing science could be no more dramatic. Rather than follow the Aristotelian process of classification and taxonomy, which is implicit in every attempt at market segmentation, product classification, or choice characterization, this view point argues the need to adopt a perspective where the behavior of individual customers as part of a larger population of consumers is the focus of attention. Agent-based modeling is precisely the kind of tool needed to enable that perspective. I don't need to model in detail the consumer behavior of specific, living human beings, but I do, and can, model heterogeneous individual entities – agents – that mimic that individual behavior in important and relevant ways. With the creation of populations of such individual agents, I can capture and explore the response of markets (populations) to the changes in the product offerings (environment) introduced by producers.

**2.6.8:** Based on the observations noted in **2.6.4**, another important implication of modern evolutionary biology is that there is nothing optimal whatsoever about the process of evolution. The members of the species that are alive within the context allowed by a particular environment at a given point in time do not necessarily consist of the "best" (in a survival sense) individuals, but only those good enough to have remained alive long enough to be counted. (You just need to run faster than the slowest individual being chased by the bear, not faster than everyone else in the group.) It is a very common misconception, without a doubt brought on by the misinterpretation by Spencer of "survival of the *fittest*" as inferring that only the strongest, or smartest, or more agile, or quickest, will rise to the top of the evolutionary heap as time moves on. Perhaps this concept is a residue of the homocentric nature of 19$^{th}$ and early 20$^{th}$ century science. Much of the controversy surrounding the theory of evolution can be traced back to the debate between science and religion, especially those components of the religious community that hold a more biblically-literal view of the creation of man. Throughout its history, biologists have tried to reconcile the evolutionary view of man as a species with the religious view of his uniqueness. Thus, if the fittest survive, and God defined fitness, then superior human would emerge from evolution, and evolution itself would be a tool of God. That evolution concerns the least unfit, and not the fittest *per se*, makes that position more difficult to support.[16]

**2.6.9:** A corollary to this is that evolution does not necessarily proceed from "lower" to "higher" species. While natural history seems to favor that perspective, it is more an artifact of the changes in the natural environment and the endless modifications to natural species as adaptation to those changes emerge. Evolutionary adaptations that no longer confer a survival advantage do not necessarily then disappear. The human appendix is a ready example. That is, a characteristic that evolved in response to some environmental configuration doesn't necessarily fade away (unless it is environmentally deleterious to the survival of the individual) just because the environment that led to its appearance is no longer present. The attribute will simply hang around, contributing nothing and probably costing nothing.[17] Moreover, it is quite possible for a species to maintain essential stasis over very long periods of time when environmental changes are slow over that same time period. There being no significant challenges to the survival skills of members of the current population, variation in succeeding generations will not confer any substantial advantages, and so members of the population which fail to carry those changes will not suffer any debilitation, and will not be selected out of the population. This is probably the current situation with respect to the evolution of human intelligence. It seems the contribution of intelligence to biological survival is fully met by the current population levels (the history of the 20th century not withstanding).[18]

[16]This debate has not subsided in many quarters. A component of the Protestant Christian faith in America, commonly called the evangelists, had recently gained considerable influence in America's national politics. Indeed, serious suggestions were put forward to amend the US Constitution to make the literal reading of the Bible the force of law (the so-called "creationist" view). It reminds me of the Indiana State Legislature's attempt in the 1880's to legislate the value of the irrational number $\pi$, as in the formula for the area of a circle, $A = \pi r^2$, to be 7/3, exactly.

[17] Such a feature, however, may serve a different, beneficial purpose in a changed environment, something Gould termed *exaptation* (Gould, 2000, pp 1229-1234).

[18] Jacques Berzun, in his *From Dawn to Decadence* (Barzun, 2000), offers a comprehensive history of western culture from the Renaissance to the present. His discussion makes a compelling, if not scientific, case that there has been little change in human intellectual capacity over that time period. Five hundred

**2.6.10:** When considering science, laws like Newton's First Law or the Second Law of Thermodynamics readily come to mind. But, as Mayr points out (2001, pp. 227-8), laws in evolution are different than laws in physics. While laws in the physical sciences are asserted to hold through all time and spatial span (at least up to known transformations that define the modern, relativistic concepts of the physical universe), laws of evolution, on the other hand, seem to be effective only when viewed in a local time and place, and clearly vary as evolution proceeds. To the extent that marketing is more consistent with the theory of evolution such temporal and spatial constraints on the "laws" implicit in a science of marketing also must obtain. Moreover, in the same way that evolution cannot be accurately described with the tools of the classical physical sciences, including the mathematics which is the expression of many of those tools, so it may well be that the evolutionary paradigm is a better fit to the true nature of marketing science. I contend that we are far from exhausting the efforts to adapt evolutionary thinking to the issues central to any reasonable conceptualization of a science of marketing. And one of the very promising tools, both conceptual and mathematical, to emerge in the development of evolutionary theory is agent-based modeling. This alone justifies a closer, more careful examination of the potential role of agent modeling in marketing science.

## 2.7: Biological Evolution as a Paradigm for Marketing Science

**2.7.1:** The view that marketing science is an evolutionary science is put forth by Shelby Hunt (Hunt, 2000). Hunt based his conclusion on the work of Geoffrey Hodgson, who has fashioned a thorough and tightly reasoned analysis of the general theory of economics as an evolutionary science (Hodgson, 1993). Hodgson traces the attempts to apply the biological evolution metaphor through several important economics scholars, starting with early authorities Karl Marx (1973) and Frederick Engels (1964), Herbert Spencer (1890, 1892), Alfred Marshal (1949/1890) and Carl Menger (1990/1892, 1963/1883, 1981/1871), and then moving to important later authorities, namely Thorsten Veblen (1990/1914, 1990/1919), Joseph Schumpeter (1954, 1976) and Friedrick Hayek (1988, 1978, 1967). Hodgson then discusses the difficulties and opportunities associated with the application of the biological evolution paradigm to

---

years is about 25 human generations, and evolutionary changes take much less time than that to appear if they are important to species survival.

economics. His thoughts on the state of the science, if you will, shed light on the nature and magnitude of the problem as it relates to marketing science, as well.

> "It is not a straightforward matter to build up a science with the use of the biological metaphor. It would be unwise to pillage biology uncritically of its ideas. Even if such a reckless approach was generally viable, biology itself does not present a consensus of ideas or methods: it is developing rapidly and contains a diversity of sometimes conflicting views. Consequently, any attempt to examine the utility of the biological metaphor must take a careful look at the internal controversies within biology itself.
>
> Furthermore, the reconstruction of economics is far from being an easy matter, and it cannot be achieved in a single book, nor by one person. The task here is more modest: to clear away the undergrowth of dangers and misconceptions, and to till the soil of fundamental concepts, so that the seeds of a future evolutionary economics may be germinated." (Hodgson, 1973, pp 34-35.)

Thus, to the extent that marketing science must recognize and accommodate economic theory (and surely it must), this relative infancy of an evolutionary perspective in economics dictates that no better than the same level of development can be found in its application to marketing science. The value of the evolution paradigm has been established, but there are decades of work ahead to extract the salient pieces, modify the concepts to comply with the empirical reality, and invent new constructs that better explain and predict marketing behavior. And biology has offered other models to the marketing community, including the widely-used Bass diffusion model (Bass, 1969).

**2.7.2:** Hunt (2000) adopts Hodgson's proposal that the unit of evolution in marketing is the institution. He puts forth a theory of marketing which he refers to as Resource-Advantage (R-A). He argues that R-A Theory assumes that consumers possess imperfect information and utility maximization constrained by moral codes. (Hunt, 2000, pp 109-122). He proposes the "Hunt-Vitell" theory of ethics (Hunt, 2000, p 120), but does not explore in any depth the origins, structure or any key characteristics of this moral code and how they relate to the evolutionary properties of marketing science. Deeper examination of this ethical philosophy is unnecessary for this discussion, unfortunately, since it has been superseded by more recent evolutionary-based theories of ethics, such as that proposed by Shermer in *The Science of Good and Evil*, (Shermer, 2004).

**2.7.3:** Hunt asserts that R-A theory is evolutionary (Hunt, 2000, pp 148-152). Using Hodgson's elementary taxonomy, (Hodgson, 1993, pp 43-46), he argues that it is a phylogenetic, non-consummatory, evolutionary theory of competition. (By philogenetic is meant that individual members of a species are the fundamental unit of evolution, and by non-consummatory is meant that it progresses forever, without end.[19]) Hunt asserts that there are two units of selection: firms and resources. He doesn't feel that brand or consumer behavior are heritable except insofar as they are resources. He also notes that, because environmentally acquired resources can be passed on, then R-A theory is Lamarkian.[20] That is, capabilities acquired from the environment can become heritable. He also comments on the importance of path dependency in evolutionary biology, (and it was noted in **2.5.4** that such dependency can be explored by agent-based modeling). He also subscribes to Hodgson's notion of the nature of agents in the economic system:

> ".. the economic system involves agents, not mere particles who mutually interact with each other according to Newtonian laws. Economic agents have knowledge and purpose. For this reason it is not possible to separate completely the observer from the object of observation. Observation and knowledge of the economic system reacts upon the system itself." (Hunt, 2000, p 23).

The analogy with quantum theory and the Heisenberg Uncertainty Principle is obvious. Because humans act with purpose, and this purpose transcends mere efficiency, "the explanations of economic phenomena require reference to intentions, not simply stochastic outcomes of mechanical cause and effect."

**2.7.4:** Finally, Hunt shares Hodgson's position that the unit of inheritance in the evolutionary economic system is the set of written policies and procedures created by an institution to define its operation and safeguard its future (Hunt, 2000, pp. 252-254). But this definition is too weak. Anyone with experience in even a moderately large organization can cite example after example of how written policies and procedures are modified on the fly, circumvented in "special" situations, ignored by individuals with

[19] Hunt makes the error of asserting that evolution is always improving the status of the evolving creature; that is, moving from more basic, less sophisticated members of the species to higher, more complex, and therefore better, members. As has been pointed out in **2.6.8**, there is no such inherent "improvement" in evolution.

[20] Genetic traits that are acquired during the life of an individual, usually from exposure to the environment, and are passed on to offspring are referred to as Lamarkian after Jean-Baptiste Lamark (1744-1829), a French naturalist who, among other significant contributions, first used the word "biology."

sufficient power to suffer no significant consequences for such overrule, and, at best, form only the framework under which organized groups function.

**2.7.5:** Thus it must be concluded that the Hunt/Hodgson explication of marketing science as evolutionary science is not yet complete. As noted above, the problem is a complex one, and much work remains to be done. But there is strong support for the concept that the principles and tools of evolutionary science can be beneficially applied to marketing science. This further enhances the value of agent-based modeling to marketing. In fact, agent-based modeling seems to be an effective tool for parsing the complexity of marketing into manageable pieces, for exploring the efficacy of various variation creation and maintenance mechanisms, and for the formulation and testing of a host of hypotheses as the theory supporting marketing science grows.

**2.7.6:** Another concept that would seem to hold promise in constructing marketing science in the paradigm of evolutionary science is the idea of *meme*. Coined (at least in the modern era) by Dawkins (1976), a meme is a "unit of idea" expressing in some fashion a concept or thought that replicates itself, with modifications, as it moves throughout a culture. Brodie (1996) and Blackmore (1999) explore the notion of memes and their gene-like qualities in depth, and on its face the proposal that there exists a cultural, thought-based entity that would fill the role of the gene in the evolution of human culture – the spread of ideas, learning, art, music and so forth – is quite appealing. (Predating that, by several decades, is the pioneering work of Marvin Minsky (1985, 1969, 1968) which also suggests trails into the problem.) However, the concept of meme seems not to have taken hold. A central journal – *Memetics* – ceased publication in 2005, and its web site has gone dormant. The circumstances causing its dormancy are beyond the scope of this discourse, but it might be due in part to a certain amount of over-generality of the notion of meme. Mayr (2004, pp 143-54), for example, feels the term is no more than a synonym for concept (asserting that Dawkins liked the word because of its phonetic similarity to 'gene'). My reading of Brodie and Blackmore left me wondering what a meme actually was, how one could be uniquely identified, and what are the mechanisms might for variation and heritability.

**2.7.7:** The mechanics of biological evolution are important if how evolution works is to be fully understood. Any biology text can give an overview of the chemistry involved, the nature of DNA and the amino acids it is built from, how the chromosomes split and recombine, and so forth. I will not repeat this elementary biology here. However, as an example of how evolution works, (and as a very simple example of agent-based modeling) it is valuable to have some familiarity with a technique called the *genetic algorithm* which is a computational method that mimics biological evolution. It is commonly used to solve various kinds of engineering optimization problems, although it is beginning to appear in social agent-based modeling. Appendix B offers a simplified discussion of genetic algorithms that the interested reader may find beneficial regarding this aspect of evolution.

## 2.8: The Evolution of Human Behavior

**2.8.1:** Marketing is a human behavior. Therefore it is to a significant degree a product of our evolutionary history. If that is the case, then as one observes human society across cultures one should repeatedly encounter common behavioral attributes, since the environment in which humans have evolved have most essential elements in common. Moreover, one can then assert a behavioral universality that would support hypotheses that could be tested across cultures. Finally, if such universality can be supported, then the construction of agents that replicate important marketing behavioral characteristics can be expected to be applicable in a wide range of contexts.

**2.8.2:** A thorough cataloging of human behavior patterns in cultures around the world has been assembled by the anthropology community. The University of Illinois at Urbana-Champaign maintains the Human Relations Area Files (HRAF), an organized and indexed compilation of every reported ethnographic study of human culture (UIUC, 2009). Many scholars have used this resource in recent years to undertake cross-cultural studies of human behavior. Brown (1991) has compiled a list of some 373 patterns of behavior that have been recorded in every human society that has been studied – anywhere in the world, large or small, old or modern. These universal behavioral traits can be mined for those relevant to marketing and marketing science; those characteristics of human culture are most important in describing, understanding and predicting the course of markets and how products can be expected to fare when introduced into those most human of complexes.

**2.8.3:** Brown focused on the hypothesis that a number of features of human culture would be found in all human experience, regardless of time, place, or history. When originally published, his views were sharply opposed to the prevailing anthropological wisdom. At the heart of the controversy was the nature-nurture debate, which still circulates actively today. Brown spends considerable space refuting the concept of cultural relativism, which holds that human cultures are vastly different with limitless variety.[21] Further, culture completely determines human behavior, and therefore there can be no human universals. (For example, by taking a contrary view, Brown contradicted the great anthropologist Margaret Mead.[22]) This position is important for this analysis because, if there were no human traits that were independent of culture, then the problem of simulating human behavior with agents in a marketing context becomes extensively more difficult. In that case every culture would have a unique heritage and historical path, making generalization very difficult. Brown's refutation of the relativistic view of human behavior is therefore valuable to the arguments justifying agent modeling in marketing science. If I cannot characterize human behavior in some reasonably perspicacious and parsimonious way, the task of defining human marketing agents will be significantly more onerous.

**2.8.4:** Brown's fundamental justification for his thesis is that human behavior is grounded in and conditioned by human evolution. The cultural artifacts contained in the list of human universals have emerged because of the mechanisms of the human brain that have evolved create the necessary physiological capability. Moreover, the fact that they are *universals* argues strongly in favor of their being of evolutionary origin. How else could all human cultures display such characteristics? One non-evolutionary hypothesis is that all human culture sprang from a single source – an Atlantis or an extraterrestrial race, perhaps. But there is no archeological or anthropological data to support this view. Another explanation might be the actions of a Supreme Being. But I reject this stance due to its lack of falsifiability.[23] It should be noted that while Brown dismisses extreme cultural relativism, he accepts that there is substantial relativism in human culture, and much of human behavior reflects the effects of the cultural circumstances within which the individual exists, so therefore cultural influences cannot be ignored in the design of agents which represent humans in a marketing context. If human universals are derived from biological forces, and if culture is the result of the

[21] Cultural relativism *per se* became prominent in post-World War II sociological research largely in an attempt to refute the naïve "survival of the fittest" philosophy exemplified by the German Nazi era.
[22] Specifically by pointing out that adolescents in Samoa indeed led stressful lives, just like teenagers everywhere else in the world
[23] In fact, besides having wandered into the heart of the nature-nurture debate, the various propositions that one could outline as the explanations for the universals themselves constitute a set of narratives. This hints at some of the discussion to follow.

operation of complex adaptive systems, then the wide variety of observed characteristic among cultures undoubtedly is the result of path dependency.

**2.8.5:** For additional support, Brown calls on Benjamin Lee Whorf, an anthropologist who wrote in the first third of the 20th century (Carroll 1956), making the argument that all human thought is conditioned by, indeed even exists because of, the human use of language. This position implies, for instance, that if the language of a culture did not contain a word or words for time, then that culture would have no concept of time.[24] It also ties the realization of human culture and behavior intimately to the study of human language, and thus linguistics, semantics, semiotics, communications, drama, fiction, and finally narrative. Brown also points out that a mind complex enough to calculate weighted summation utility function values (much less non-linear utility functions) probably could never evolve (see Cosmides and Tooby, 1987, for a more detailed discussion). This is because human beings, as all life, evolve in response to specific environmental constellations of factors, not generalized, abstract sets of circumstances. This is an example of the "survival of the least unfit" concept as discussed earlier. Indeed, Duckworth (2009) explores another hypothesis arising of the least unfit position especially important to humans: That higher levels of behavioral flexibility, such as is observed in humans, leads to evolutionary retardation. Such species are able to adapt their behavior in response to environmental pressure, rather than wait for the physiological changes required for traditional adaptation.

**2.8.6:** From Brown's synopsis, I can identify at least 76 human universals that seem related to why humans engage in and respond to marketing. The actual number is somewhat open to interpretation, since the classification was done by a single individual – me. Brown presents his collections of universals in a narrative form, in the mode of ethnographic studies common to the field of anthropology. He does not offer an itemized, numbered list of universal traits, but crafts anecdotes of how all people live and interact. Thus, if he wants one, a reader of Brown has to extract a list for himself.[25] The following is what I have extracted from Brown (1991, pp. 130-141). I have organized the items into roughly similar families for ease of comprehension and

[24] Whorf had studied the Hopi Indians of the American southwest, who seemed to have a significantly different linguistic structure describing time.

[25] Shermer lists 202 human universals that he feels relate to the evolution of human ethics and morals. His list can be found in Shermer (2004) pp. 285-292.

discussion, but the categories have no particular value beyond that. Indeed, some traits could easily fall into multiple classifications.

**2.8.6.1:** *Culture*:

1. Humans are aware of culture, and behaviors and thought processes are spread among individuals and from generation to generation by learning.
2. We are aware of our apparent uniqueness among known living things in having culture, although it is not strictly true.
3. All human cultures use speech, often ceremonial or otherwise distinguished from ordinary, day-to-day usage, to tell stories that explain how things came to be or what they will become.
4. All human cultures share a common external world and internal existence.

**2.8.6.2:** *Language*:

5. All human cultures have language. We use language to think about and discuss among each other both our internal states and the states of the external world around them.
6. Language is not a perfect representation of a human's views or thoughts. There are discrepancies between what we think, what we say and what we do.
7. Because the use of language is not a literal description of the world and our place in it, we must, and do, distinguish between the world as it actually is and the world as we conceptualize it.
8. Humans use language to organize, respond to and manipulate other humans.
9. An important means of verbal manipulation is gossip.
10. Language is used to misinform as well as inform.
11. All humans lie and mislead others, some to a greater extent than most.
12. All humans understand what a lie is and have methods of detecting it when it occurs.
13. Those who are more proficient in the use of language are of higher status, in the view of others, and are more capable of manipulating the behavior of others.
14. Language is highly symbolic in structure, and even though the exact sounds and words used to denote a given entity or concept may differ, rather arbitrarily, from one group to another, all languages contain a basic set of common features. Thus different cultures can communicate.
15. All languages have a similar structure in that they contain nouns, verbs, and possessive forms.
16. A few concepts are found in all languages, such as *face*, *black* and *white*, *male* and *female*.
17. Male and female gender is always distinguished in languages, creating a sexual content in all human language.
18. Our sexual terminology is dualistic – male and female. Other terms can portray different sexual contexts, but only in reference to this basic duality.

19. All languages describe family and kinship relationships in terms directly related to procreation, such as father, mother, daughter, etc.
20. All languages express the concept of time, and units thereof.[26]
21. All language contains groups of contrasting terms that could be expressed in three ways. For example, we could contrast *good* with *bad* by saying *good or bad*, *good or not good*, or *not bad or bad*. But, interestingly, the third case never occurs in ordinary language.
22. Language changes over time, adopting new words and letting others fall into disuse. Language structure changes more slowly, but does change nonetheless.
23. All languages make extensive use of metaphors.

**2.8.6.3:** *Analysis*:

24. Humans learn by trial and error, by which is meant if some event or sequence of events does not occur as expected, our *expectations* change, not the perceived events or sequence of events.
25. Humans can measure and express the size of things.
26. All human cultures use language to describe physical properties –speed, motion, dimensionality – and actions, such as giving, lending, and affecting other things and other people.
27. All human languages have terms which identify and discuss parts of the body, internal states such as emotions or thoughts, behavior, the physical world, the weather, tools that are available, and space.
28. An essential feature of our ability to recognize, learn and understand the world is that we can identify distinctions and name them; that is, we can create taxonomies.
29. The basic form of distinction is that of binary discriminations, the basis for all categorizations (a thing either is, or is not, in a specified category, or has, or has not, a particular property).
30. All human cultures can also classify things that do not necessarily fall into discrete categories by ordering them along some kind of continuum.
31. All human cultures use basic logic constructions such as "not," "and," "same," "if … then," "equivalent," and "opposite."
32. Humans have the ability to deduce, from a variety of even subtle clues, the current state and future condition of things, often very inaccurately.

**2.8.6.4:** *Recognition of mind*:

33. All individuals have a distinct concept of person and of themselves separate from others.
34. We can easily and intuitively get into the minds of others and imagine what they are thinking and feeling.
35. We know that other people are like us have inner thoughts, make plans, and make decisions and choice.
36. We are able to think not only of our own relationships with others, but of the relationships between others and themselves.
37. All of us also mask or modify expressions to mislead and confuse.
38. All of us communicate non-verbally, especially with facial expressions.

[26] He thus contradicts Whorf in **2.8.5** above.

39. We respond to sexual attraction.
40. We use reciprocal exchanges in all aspects of our lives. This includes its negative forms, such as retaliation and revenge.
41. We are able to distinguish between normal and abnormal mental states.
42. Humans have ways of making themselves feel better, including the use of stimulants, narcotics, and intoxicants.

**2.8.6.5:** *Tools:*

43. We make and use far more tools than any other animal, and use tools to make other tools.
44. We know how to use fire.
45. All cultures have means of shelter from the elements.

**2.8.6.6:** *Group association:*

46. Most of us live part or all of our lives in groups.
47. We have a sense of territory associated with the groups to which we belong, and are well adapted to the environment associated with that territory.
48. There is a distinct sense of "us" and "them" between the groups to which we belong and other groups.
49. We judge members of other groups in terms of those qualities found in our own.
50. Marriage, in the sense of a well understood protocol for access to females of child-bearing age, exists in all cultures.
51. We have a pattern of socialization in the sense that child-rearing responsibilities are shared among adults, who are obliged to teach the young.
52. Children learn from adults by copying and mimicry.
53. The existence of roles and social structure is implied by kinship, sex and age statuses found in all cultures.
54. Prestige is not equally distributed among group members.
55. An ethical dualism exists between the groups to which we belong and other groups.
56. We have leaders, but they are never completely powerful.
57. There is never a complete democracy nor an absolute autocracy. Thus there is always an oligarchy.

**2.8.6.7:** *Economic behavior:*

58. All societies have economies – systems for barter or trade.
59. We engage in trade, defined as exchange of goods that is not based on future expected reciprocal behavior.
60. All societies have divisions of labor, and customs of cooperative labor.
61. We are all materialists to some extent, and distinguish who owns what.
62. Members of groups are not economically equal, in that some have more materiel resources than others.
63. We are envious.
64. We give gifts.

**2.8.6.8:** *Ethics, morals, art and metaphysics:*

65. We all have a definition of, and can distinguish, right from wrong
66. All cultures have standards of sexual modesty
67. We all believe things that are clearly and demonstrably false.
68. We all practice magic, especially in trying to prolong our lives or attract others in a sexual context.
69. All humans attempt to predict and to plan the future.
70. All cultures have theories of fortune and misfortune.
71. We attempt to control the weather.
72. We all have a coherent world view, in that we understand the world as a unitary whole regardless of the sensory mode by which it is experienced.
73. Our worldview plays into our mythology and concepts of the supernatural.
74. All cultures have rituals, especially to demarcate changes in states, and all cultures mourn the dead.
75. All cultures have aesthetic standards and preferences.
76. All cultures have music, poetry, play and story-telling.

**2.8.5**: This is a reasonably broad list of traits. Appendix C explores the nature of these traits in more detail. And I am asserting that they arose as a product of human evolution. But how? Perhaps a brief consideration of the trait of altruism can shed some light on how a complex, and *prima facie* counter-survival, behavior can emerge.

**2.8.6:** Evolution, as it is often simplistically defined, is the survival of the more fit over the less fit. Traits which improve the survival of a species persist, those that do not fade away. But it is very hard to see the evolutionary value of altruism. The biological definition of altruism is behavior that *reduces* the chance of survival (Sober and Wilson, 1998, p. 17). The common social meaning of the word – the granting of a kindness to another for no immediate return – alludes to this inherently anti-survival characteristic. But altruism persists in spite of its evolutionary unsuitability. Mayr (2001), Brown (1991), Shermer (2004, 2008) and even Darwin (1859) offer substantial evidence that altruism is a basic human universal. Further, it is difficult to see how altruism is just an unused artifact of evolution left over from a prior era, or is just there because of evolutionary background. That is, is an *exaptation* (see the discussion in Gould, 2002). It would seem to be downright dangerous to the future of the species. But how did it come about, and how does it continue, if it confers no adaptive advantage?

**2.8.7:** It is not surprising that the continued existence of altruism has caused substantial consternation among biologists when it comes to understanding human evolution. One line of thought, that fostered by Dawkins (1976), the author of *The Selfish Gene*, requires that altruism cannot really exist, and that it only arises in the context of "reciprocal altruism," where a kind deed done at one point in time serves the purpose of helping to insure a reciprocal kindness later, when it is needed by the original altruist. But this theory is belied by the psychological evidence that surround acts of heroism. Clearly, a hero's actions are altruistic. But there is no substantial data to suggest that heroes are such because of a superior desire to accumulate reciprocal acts of kindness in the future. Indeed, heroism seems to act just the opposite, often being a spur-of-the-moment, almost spontaneous act. And, while a hero attracts the admiration of the fellows in his/her social group, such admiration does not seem to be a motivating factor in the act itself (Coon, 2007). In other words, heroic acts are a form of pure altruism. The altruist expects nothing in return. And very often, especially in time of war, heroism results in the death of the hero, which is clearly counter to the flow of evolution.

**2.8.8:** I believe the work of Sober and Wilson establishes that complex human behaviors can certainly evolve from biological foundations coupled with appropriate group definition and interaction. The details of their argument and conclusions are offered for the interested reader in Appendix D. Surely if the evolution of altruism has been demonstrated to be possible, then, since the characteristics of humans relevant to marketing are no less counter to survival than altruism, they could also be expected to arise under the appropriate group interaction structures. Sober and Wilson, in fact, make a compelling case for the emergence of social norms in the second half of *Unto Others* (Sober and Wilson, 1998, pp. 199-362), from which Cialdini's (discussed in Section **2.9** below) attributes of social proof, liking and authority directly derive.

**2.8.9:** To my knowledge, no one has traced a likely evolutionary scenario specifically for these marketing traits, but Micheal Shermer (2004), has assembled a convincing analysis of the evolutionary basis of human ethics, vital to any theory of marketing. Indeed, I find his treatment considerably more compelling than that of Hunt (2000) and Hunt and Vitell (1993), but a comparative analysis is not needed for this discussion. This is important because morality is generally accepted as a prerequisite for successful market structures (all deals are ultimately based on trust, which is a moral concept.) The reader is invited to consider the details of Shermer's views, which are described in some depth in Appendix E.

**2.8.10:** Shermer's presentation gives a pathway for the evolution of ethical behavior. This is beyond merely establishing that a complex trait like altruism *could* have evolved, it goes on to demonstrate that such traits did indeed evolve. This type of argument is central to evolutionary science. One must show a pathway by which an observed trait evolves from other observed traits in that species. That Shermer is able to display an evolutionary path strongly supports the evolutionary origins of all universal human behaviors.

## 2.9: Universal Characteristics of Marketing Behavior

**2.9.1:** The marketing community itself has not ignored the impacts of these discoveries. A number of authors have used these findings as a platform from which to elucidate commonly believed attributes of people as they engage in, and are engaged by, marketing activity. Recent contributions can be found by Cialdini (2001), Ariely (2008),[27] Taleb (2007), Mlodinow (2008), Kunda (2001), De Merchi and Hamilton (2009) and a number of others. Arguably the most useful, if somewhat populist, treatment for the purposes of this discussion can be found in Robert Cialdini's *Influence: Science and Practice.* (Cialdini, 2001). He specifically concentrates on seven patterns of behavior that occur in the interpersonal transactions that allow one individual to persuade another to behave in a specific way – in short, how influence works. These characteristics are important because the properties which govern the susceptibility of an individual to influence also condition how one receives and processes information, which, in turn, contributes to the purchase choice. Furthermore, as will be argued in the discussion of the narrative construct (Chapter 3), these behaviors are expressions of underlying thought processes that provide what will be called "good reason," and are therefore important to the modeling of human agents. Cialdini tags the seven behavior patterns with these labels: 1) fixed-response; 2) reciprocation; 3) commitment and consistency; 4) social proof; 5) liking; 6) authority; and 7) scarcity. Indeed, these are the chapter titles of the book. Each will be reviewed in turn below.

**2.9.2:** All sentient creatures make choices in their environments given limits on the time, the resources and the skill they can apply to the decision problem at hand. *Fixed*

[27] Ariely is one of the more egregious culprits in an empirical error that is seen throughout academic social science; the use of college students as subjects in various social and psychological experiments. While certainly convenient, such a population is hardly representative of the general population. So, while many of the conclusions of such research are enlightening for significant aspects of human behavior, they are not as broad as the authors often portray. Hopefully, as greater understanding is sought of human behavior across the lifetime of the individual, and more resources are made available, a less one-dimensional picture of how we *really* behave will emerge.

*response* is one of the most common forms of such bounded rationality. It is an automatic response mechanism typified by the launching of a series of automatic actions and decisions triggered by a particular set of environmental stimuli.

**2.9.3:** *Reciprocation* is the tendency for people to expect and respond to assistance from others. If an individual needs help, most will offer a hand to assist. But Cialdini argues that the offer and acceptance of that assistance instills in each party automatic responses. For the individual who received the assistance, there arises a need to put forward assistance in kind to the individual who put initially offered the help. The person who gave the assistance, on the other hand, expects to be repaid. Further, the transaction can escalate, but in a somewhat unexpected way. The recipient seems to have a desire to pay back the help with *more* than was originally received. This, in turn, creates a need on the part of the original helper to respond back, also with more than is needed to balance the books, and so on. There is no need for the reciprocal transactions to occur immediately after one another. The response can happen at some later time and be equally effective. The evolution argument supporting reciprocation can be made based on the need for cooperation among individual in human social groups, and is modeled with agent models by Axelrod and others. Cialdini (2001, pp. 22-30) suggests that the triggering mechanism is the original offer of help, and after the trigger event, the response is essentially automatic: the expectation of a return favor, with nominal interest on the obligation, and the necessity of returning the favor, also with interest. However, the analysis of altruism by Sober and Wilson (1998) suggests the logic of reciprocation is much deeper than Cialdini's argument portrays.

**2.9.4:** *Commitment* and *consistency* are behavior patterns that increase the likelihood that an individual will be able to predict and deal with the future behavior of another. Commitment is essentially a promise to perform some action or take some position in the future, if the appropriate circumstances should arise, circumstances which have been previously stated and understood. Consistency refers to the need to meet the commitment when the time comes. Of course there is tremendous social pressure to maintain consistency once a commitment is made. Indeed, ethics, and to some extent religion, laws, formal and informal codes of conduct, and other such social norm mechanisms give social and institutional substance to commitment and consistency. When engaged with reciprocity, moreover, there can be a tendency for commitment and consistency to escalate. If an individual commits to do something, he can often be convinced to commit to do additional things because they are consistent with the original obligation and they are part of a reciprocity pattern that has been triggered somewhere in the mutual interaction. Clearly commitment and consistency descend from group adaptation, and are effective means for social survival, both for the individual and for the group.

**2.9.5:** Cialdini's third mode of behavior is *social proof*. He defines social proof as the tendency to "…view a behavior as correct in a given situation to the degree that we see others performing it." (Cialdini, 2001, p 100.) In other words, we tend to believe that something is true, or that particular behavior is proper and appropriate, or that a given choice is the correct (wisest, most clever, highest utility, most moral) to the extent that others support its truth, appropriateness, value or morality. From a societal survival viewpoint, this behavior can be very successful. If others, especially many others, are making a particular decision and suffering no apparent harm, or increased security and good fortune from it, why should we not follow that path? Basically, it's reaffirming that 50 million Frenchmen can't be (or, most likely, aren't) wrong. Another causative factor in social proof is shared narrative. The reason that the observed behavior is applied to a particular circumstance is not only because others seem to benefit (or are not harmed) by it, but also it fits a shared understanding of the anticipated explanatory chain containing the behavior. (See Chapter 3 for further details.)

**2.9.6:** It is easy to see how reciprocity, commitment and consistency, and social proof interact. Reciprocity will create an interaction between two individuals that can be extended and codified by the need for commitment and consistency. Consistency can be validated by means of social proof, so one who commits to a set of beliefs can substantiate those beliefs based on the number of others who also hold them. Examples of the application of social proof to marketing are everywhere. Establishing the credibility and desirability of a product by showing that a large number – the larger the better – of consumers prefer it is a standard marketing tactic. Indeed, the power of mass marketing through the electronic media is enhanced by the theory of social proof, since it is possible to artificially create significant social proof by the very act of advertising on such a massive scale. The viewer sees an ad on TV, realizes that for the ad to even command TV time the product must command the purchase decision of a significant number of other people, so the social proof is supplied.[28]

**2.9.7:** The concept of *liking* is the next social behavior Cialdini explores (2001, pp. 143-177). In addition to subscribing to a decision supported by large numbers, people are apt to adopt that decision themselves if someone they like has made it as well. Cialdini cites substantial research that indicates there are a number of aspects to liking that are important. One of the most prominent, contrary to what the more egalitarian would like to believe, is physical attractiveness. See Eagly *et al.* (1991) for a review. There is ample evidence that physically attractive individuals more easily gain social and economic advantage. The presumed superior status afforded by physical attractiveness also carries over to other behavioral traits such as intelligence, talent, wisdom and kindness.[29] Another is similarity, rather like the role of similarity in social proof. We have a greater tendency to like people who are like us. We also like people who give us praise or compliments, reinforcing our self-esteem and giving us cause to feel good. A fourth aspect of liking is familiarity. We tend to like people and things with which we have had prior (successful) experiences. Individuals with whom we have bad experiences we tend to avoid, thus reducing the likelihood we gain any familiarity. Finally, liking has an aspect of association. We tend to associate ourselves with things we like, such as sports teams or clubs, and we tend to like people who choose to associate with us. Shared narratives play an important part in supporting this attribute.

**2.9.8:** The remaining two socio-psychological features Cialdini discusses are *authority* and *scarcity*. Authority (Cialdini, 2001, pp. 178-202) refers to the tendency to comply with others who are, or appear to be, in positions of authority. Very often people act automatically to symbols of authority, whether the authority is indeed valid or not. For example, a building security guard may be treated as though he were a police officer while in fact having no official authority whatsoever. This automatic, unquestioning

---

[28] A curious 'inverse' social proof application surrounds a product advertised in America called *Head-On*. The product is a headache remedy, and its advertising is considered especially obnoxious by most viewers. It was selected (by some prestigious sounding group) as the least-liked advertisement on television. The makers of Head-On then took advantage of the social proof established by the obnoxiousness of the original ad to modify it just slightly – bowing to 'pressure' from the public – thus using that proof to create demand for the product.

[29] Indeed, many individuals who have strongly attractive physical characteristics often feel it is those characteristics that give them personal worth, rather than what they deem to be the genuine talents they possess or skills they develop.

response to authority is a valuable survival tool, of course, but can be dangerous if invoked in an inappropriate place and time. Another form of authority is the expert. We naturally defer to individuals with greater expertise when we are presented with a choice in a less familiar circumstance. The use of this kind of authority is common in marketing. The expert opinion expressed by the 'doctor' on the TV ad, for example.[30] The recent economic downturn has thrown into question many in the financial expertise field. Their advice proved wrong, and seriously so. Taleb (2007, pp. 145-156) explores how expertise, and belief in it, can lead to bad decisions and dangerous behavior. He, Mladinow (2008) and others have made it clear that for phenomena that are inherently random, like the stock exchange, expertise in forecasting the future cannot logically exist. Moral: fire your financial advisor – you're as good at it as he is!

**2.9.9:** *Scarcity* (Cialdini, 2001 pp. 2-3-232) is interesting because there is significant evidence to support the idea that it derives from basic human evolution, and it helps us understand, at least in part, how we determine the value of a good. The concept is that a thing becomes more valuable the less available it is. This idea is clearly behind the economic principle that rare equals valuable. But more than that, common things that suddenly become less available become more sought after. Further, even if the quality of the good does not change, or if the need for it from an objective, utilitarian perspective is not altered, the fact that it is less available causes it to be perceived as more important. One hypothesis that would explain this increase in value because of scarcity, regardless of intrinsic value or utility, seems to be an expression of the implicit restriction to freedom that results from reduced availability. The item is more desirable because it might no longer be available. Psychological studies have shown that human children around the age of two and in their teenage years are particularly affected by such elimination of choice (Levine, 1983). And these are two key periods of human development. At about age two we move from self as being immersed in the surrounding environmental milieu to recognition that we are independent entities that

[30] A particularly notorious example of the use of expert opinion in advertising was the use of Marcus Welby, MD, in television advertising in the United States. Marcus Welby was a physician character in a long-running, successful television series of the same name, played by an actor named Robert Young. After the series was off the air, Mr. Young was hired by companies to pitch their products based on his "expert" opinion as a medical doctor. Clearly, he had no basis for expressing such opinions, and later in his career he withdrew from participating in such advertising because of the misrepresentation that had been generated, even though they contained clear disclaimers regarding his medical expertise.

can interact with and control our environments[31]. When we move from strong social dependency to independent adulthood as we move through our teenage years is the other development period.[32] Because scarcity implies lack of availability, and lack of availability implies lack of choice, the reaction to decreased choice is the imputation of higher value. This phenomenon is termed by Cialdini as *psychological reactance* (2001, pp. 210-218).

**2.9.10:** Ciadini commits one glaring omission in his catalog of marketing behaviors, however. He does not discuss *ratiocination*, which is the application of reason to the problems of marketing. Occasionally in the literature it is popular to harp on the "irrationality" and ineptness of humans as they make choices. Taleb (2007) and Areily (2007), for example, sometimes enjoy scolding us for being driven by the vagaries of emotion and the errors of careless thought. But in spite of these issues, most people, perhaps more often than is realized, apply careful analysis and rational, utility-grounded decision structures to many of the problems before them. Major purchases, such as stock market securities or real estate for example, are often made only after reasonable and careful financial analysis of the costs and benefits of the available options are studied.

## 2.10: Summary

**2.10.1:** The discussions of this chapter have established several important factors supporting the application of agent-based modeling as a marketing science paradigm. First, agent-based modeling as a technique and tool was defined, and its epistemological foundations verified. It is properly classified as an expression of computational science, which has itself been shown to be a legitimate and sophisticated addition to human cognitive and reasoning power. So, there is no risk in seeking to apply it to marketing science, provided it is suited to the inherent nature of human marketing behavior.

**2.10.2:** But the foundation of human marketing behavior is human evolution. First it was established that human beings share a common set of behavioral traits that have been found in all cultures and eras. This supports the hypothesis that they *could evolve*, at least to a substantial degree, from our biological heritage, and are not strictly the result of cultural factors. I then assert that behaviors that are eminently social and

[31] thus the "terrible twos," familiar to all parents.
[32] from which teenage rebellion emerges.

clearly counter-survival, such as altruism, *can evolve* in the human species from a combination of individual and multilevel adaptation.

**2.10.3:** I have also established that there is strong evidence supporting the hypothesis that human morals and ethics, essential for the interpersonal and social trust structure critical to the operation of markets, *have evolved*, from which economics and markets emerge as complex adaptive systems. If the ethical behavior required of a marketing system has evolved, then I argue that the evidence suggests that an identified set of purely *marketing behaviors have evolved* in the human species.

**2.10.4:** Therefore, those behaviors form at least the foundation of a consistent set of traits which can be, at least in theory, incorporated into models of human actors in agent-based virtual markets. Now the problem is to establish a structural framework and context in which these traits can be modeled, so that similar capability can be reflected in the design of computing objects that represent how people behave when participating in marketing.

**2.10.5:** Thus the first two steps in the research program outlined in Chapter 1 have been accomplished. Agent-based models are valid representations of real things, and in particular things associated with evolution. Furthermore, human marketing behavior is best understood from the perspective of evolutionary science, so agent-based models should be applicable to describing and exploring that behavior. The next step is to delineate a structural framework for that description. That framework is the subject of Chapter 3.

# Chapter 3:

# The Narrative Framework

## 3.1: Introduction

**3.1.1:** I now propose a concept that ties the universal traits and marketing behaviors discussed in Chapter 2 into a coherent structure specifically designed for the implementation of agent-based models. This is the third step in the research program. I call this the *narrative framework*. This chapter will present the background and important concepts from the work of linguists, literary analysts and communications theorists. I will discuss below what I mean by the term 'narrative,' and how narratives act to organize human thinking and understanding of the physical and social world. The discussion then moves to the framework itself and examine some of its most salient properties. Following that, I examine how marketing traits fit the narrative structure. These strands of reasoning will be encapsulated in the conclusion that the essential component of the narrative as it relates to agent-based modeling in a marketing context is the fundamental role of choice. Finally, a sampling of choice mechanisms and how they might be treated in an agent-based model will be offered.

**3.1.2:** A narrative, in the ordinary sense the word is used, is a story that relates a sequence of events, either real or fictional. It stems from the Latin verb *narrare*, "to recount or tell." The term as a number of applications, meaning something slightly different in each. It is often used to describe a particular type of story presentation wherein the story is related by an individual – the narrator – to an audience, such as readers in the case of a written novel or the attendees at the performance of a play. Narratives as literary artifacts form a substantial proportion of the scholarly literature on the subject. The term *narratology* refers to the systematic study of the techniques and methods of narrative as art – books, movies, plays, paintings – and their transmission and reception. Bal (2009) is prominent in the field, and can be consulted as a reference. But many of the concepts and structures she defines have use beyond just the literary context in which they were developed.

**3.1.3:** In psychology, the term can refer to psychological processes including self-identity, the construction and manipulation of memories, and many consider it a basic requirement for existence of a concept of an independent, salient self (e. g. Hevern, 2004 or Dennett, 1991). They are also a fundamental tool of anthropology, where ethnographic studies are often related in a narrative format. Brown (1991), who is an anthropologist, ties many of the human universals (see Chapter 2) directly to the concept of narrative, especially elements of culture (**2.8.6.1**) and language (**2.8.6.2**), the recognition of mind (**2.8.6.4**), some aspects of group association (**2.8.6.6**) and ethics and morality (**2.8.6.8**). In fact, he also asserts that the concept of narrative is fundamental to understanding all aspects of human behavior, including the formation of markets and the behavior of individuals and groups in a marketing context.

**3.1.4:** Because narratives are stories, they play a prominent role in linguistics, especially the semiotic aspects of linguistic studies. This is natural enough, since language is the vehicle most encountered in the portrayal of narratives.[33] There has been considerable attention given to the narrative structure of linguistics. In addition to the structure of language as relates to the presentation and interpretation of narratives, (Gannett, 1972/1980), there is also significant attention paid to the so-called ‘structuralist perspective,’ which implies that the linguistic instantiation of a narrative is homologous with the narrative itself. People thus think in words (Toolin, 2001). There is even considerable activity in the area of mathematical linguistics (see Kornai, 2007), often aimed at computer-assisted language translation (e. g. Hutchins, 1999).

**3.1.5:** Narratives have found application in artificial intelligence work. Bringsjord and Ferrucci (2000) describe a computer program called BRUTUS, which can create stories based on well defined narrative structural elements, such as betrayal. Turner (1994), Ryan (1991), and Schank (1990) relate the development of artificial intelligence directly to narrative theory. Most of these efforts are aimed at creating computer-assisted story writing applications or the development of characters in computer games. They tend to use rule-based methods of artificial intelligence, such as those found in expert systems. There also have been attempts to apply the concept of narrative to the development of knowledge management systems, where narratives are suggested as possible mechanisms to acquire, store and distribute knowledge (Snowden, 2002, Denning 2001). However, more advanced analytical techniques – neural networks, Bayesian method, Markov models – have apparently not yet to move into the field. Some discussion of how these methods may enter into the narrative framework is offered later in this chapter.

[33] Other media include painting and music.

## 3.2 The Central Role of Narrative in Human Understanding

**3.2.1:** The role in communications of the narrative has become a central theme of study in that field. Walter Fisher is a leading proponent of the view that all human thinking, including any communications, exists and is portrayed by a narrative structure. The details of his analysis are offered in Appendix F. Based on his conclusions, two important features of narrative emerge.

**3.2.2:** First is the concept of *good reason*. (Fisher, 1987, p 57). Good reason means that the rationality of a particular situation (or argument, presentation, story, discussion, opinion, etc.) depends on the underlying narrative that is supporting an individual's understanding of that situation. This, in turn, must imply that many, if not all, human mental processes are devoted to the creation and maintenance of internal brain patterns that are of a narrative structure. If a narrative is a recounting of a sequence, Fisher's context refines the definition to mean a mental pattern of time-based (therefore at least sequential in time) changes in perceived states, which he defines as *events*. That human beings build and maintain such patterns is widely supported by substantial psychological, physiological, biological and physical evidence. In other words, the logic inherent in a narrative creates the instantiation of good reason. An argument is thus considered to be rational, not because of rigorous logic, but because it makes sense as part of a narrative.

**3.2.3:** The time dependency of the narrative sequence structure is vital. While sequences in space can be recognized without reliance on memory or other neurological pattern retention mechanisms, that cannot be true for events that unfold in time.[34] The reality of ten minutes ago no longer exists, except in the memory of one who observed and remembered them. Furthermore, in order for the temporal pattern to carry meaning, the events in the sequence must be perceived to create a *cause-and-effect* chain. Thus a narrative must be made up of patterns like "if A occurred, then B will occur," where the meaning of the word 'then' is at the minimum, the passage of time. It is this time-sequenced cause and effect chain that Fisher is talking about when he refers to "good reason." A communication, such as a legal argument, contains good reasons if it asserts

[34] It takes some small amount of time to observe any particular scene with the eyes, for example, but the information recorded in each instant is substantial. The retention mechanism here might well be the retinal cells themselves, not considered part of the brain. The principle is the same, but the physical organ that does the work may not be specifically cerebral.

causal relationships between earlier and later events in the narrative sequence. This is the basic requirement to understanding. To understand something is to fit it into a mental pattern of a cause-and-effect chain.

**3.2.4:** But the description of narrative as a time-framed cause-and-effect chain is so general as to apply to almost any human thought process. How is it distinctive when applied to human thought in a marketing context? The distinction lies in the duality between the narrative and the course of events as they occur in the "real world." Apart from whatever is contained in a narrative, it is obvious that the world outside of the narrative unfolds over time.[35] In agent-based modeling this real world is referred to as the *environment* in which the agent is embedded. The environment can be described in terms of set of time-dependent cause-and-effect chains, (although the description itself can be no more than a narrative), and therefore contains events, and sequences of events. However, there is no history or future associated with the environment, only narrative *recordings* of such.

**3.2.5:** The individual is clearly part of the real world, and constantly interacting with it. Some of these interactions are called, appropriately enough, *actions,* in that they involve the individual doing something to alter the course of the unfolding of the environment's time-sequenced cause-and-effect chain. That is, the individual steps in and influences the outcome of, or perhaps even generates, an event, the outcome of which causes the course of the real world to change. This is done by changing those aspects of the outside world that are represented by an internal mental representation in a fashion that alters the *probability,* according to the individual's narrative, that one of a particular set of outcomes of the event occurs, thus (if it does happen) altering the course of the environment. Changes in the elements of the environment that correspond to representations in a narrative can be accomplished in a number of ways, but arguably the most important method involves the *allocation of resources*. By resources are meant features of the environment over which the individual has discretionary control. For example, resource allocation includes selling a possession and the taking cash from the sale and buying a good, or creating or transforming something from other resources (using soil and seeds to grow food, ore and heat to forge metal).

[35] There is no need to engage in the epistemological argument about the existence of the world outside of the sensory apparatus of the human being. I assume it is there, whether I am or not.

**3.2.6:** How are narratives created? They are learned by experience, trial and error, formal education, by the recording and recalling of circumstances and events that surround everyday life. Kenneth Burke (1957) – using a heavy dose of (literary) narrative – describes this wonderfully when he portrays the role of the narrative in human communication:

> "Imagine that you enter a parlor. You come late. When you arrive, others have long preceded you, and they are engaged in a heated discussion, a discussion too heated for them to pause and tell you exactly what it is about. In fact, the discussion had already begun long before any of them got there, so that no one present is qualified to retrace for you all the steps that had gone before. You listen for awhile, until you decide that you have caught the tenor of the argument; then you put in your oar. Someone answers; you answer him; another comes to your defense; another aligns himself against you, to either the embarrassment or gratification of your opponent, depending on the quality of your ally's assistance. However, the discussion is interminable. The hour grows late, you must depart. And you do depart, with the discussion still vigorously in progress." (Burke, 1957, pp. 94-97)

**3.2.7:** What other evidence is there that narrative has the central role in human thinking that Fisher ascribes to it? Many might find the rhetorical frame of reference he uses difficult to support, and thus his argument less compelling. Acceptance of the concept gains momentum from a number of other areas, however. Theoretical mechanisms that could function to make the capture and manipulation of time-sequenced patterns possible have been demonstrated with neural networks. See Hertz *et al.* (1991) for an early survey, Masters (1995) for a more technical computing perspective, or Haykin (1999) for a rigorous foundation. Strong evidence has emerged indicating that mechanisms similar to neural nets actually are found in biological systems, including primates and humans. A good introductory discussion is given in Sternberg and Ben-Zeev (2001, pp. 74-77). A more fundamental argument is in the classic treatise by Minsky and Papert (1969). Rumelhart and McClelland (1986) include a discussion of parallel processing which is vital capability for understanding cognition. Functional Magnetic Resonance Imaging (fMRI), very popular now with the marketing and advertizing community, reveals distinct patterns of blood flow to different parts of the brain in response to various stimuli.[36] Montague (2006) makes a compelling argument

[36] For no substantial reason that I can detect. Simply showing an association between a laboratory stimulus and a portion of the brain that also is activated when some emotion is triggered stretches the

for the physiological basis of narrative structure through the dopamine reward system (see Montague, 2006, pp. 119-161). Psychology has maintained a central role for narrative-type neural organization for many decades with paradigms such as transactional analysis, founded by Eric Berne. See Berne (1964) or Stuart and Joines (1987) for a review. Marketing has used the narrative concept for many years. Examples abound, but the work of Rapaille (2006) is unusually entertaining, even if it does hark back to the 'infinitely pliable' cultural perspective mentioned in Chapter 2. But it has been dismissed by marketing theorists, perhaps because of its non-technical logic background.

## 3.3: A Formal Narrative Construct - Definition and Properties.

**3.3.1:** The reason for exploring the concept of narrative is that a structure is needed to describe the relationship between what an agent *perceives* the world to be and what the world *actually* is. This is required because the creation of an agent-based model of sufficient complexity to model market behavior involves representations of both *actual* reality, and a given agent's *understanding* of reality. Therefore, for my purposes, the definition of *narrative* that applies to the modeling of human agents is as follows. *The term narrative will be taken to mean a pattern maintained internally by the agent that represents temporal cause-and-effect chains which define events that are perceived by the agent, with which the agent 'understands' the events, and based on which the agent takes action.* In humans, the internal maintenance is carried out by the mental processes supporting memory and pattern recognition. Virtually all animal organisms show evidence of memory-based pattern creation and maintenance. Montague (2006, pp. 69-72) argues that all mobile creatures at any level of evolution must be able to foresee the future, if even in a rudimentary sense. The purpose of narratives is to guide the entity as it lives its life within the environment in which it finds himself. The agent compares the perceived state of its current environment with its internal narrative, and uses the cause-and-effect chain represented by the narrative to determine what it expects the future state of the environment will be. It also uses the narrative to forecast the future of the environment given various courses of action by the agent.

---

science, in my view. That there *is* a relationship is reasonably clear. What it is, however, is another matter.

**3.3.2:** Thus the narrative is the basis for *action* by the agent. The notion of action means that the individual reaches into the "real world" and modifies something in the event structure to affect the outcome. This action always entails two steps: the choice of what to do, and the execution of the actions dictated by that choice. Such a choice point in a narrative only occurs at an *event*, which is defined as an opportunity for change in the relationship between the narrative and the individual's perceived reality which can alter the anticipated course of the narrative in future time. The duality of narrative/reality illustrated by actions and events is an essential characteristic of narratives. Since narratives are temporal cause and effect chains maintained in a mental structure, they are essentially independent of the reality they represent. This distinction is captured in part with the terms *narrative time* and *real time*.

**3.3.3:** The narrative can be given a number of features and properties. All narratives have a distinct beginning and ending, and cover a finite interval of narrative time. These limits are referred to as narrative *horizons*, and each narrative has a *past* and *future* horizon. As narratives are created and discarded, these horizons move. It might be asserted that older individuals have different past and future horizons than younger ones, for example. The perceived state at the past horizon must be different from the perceived state at the future horizon, otherwise there can be no meaningful notion of cause and effect. Therefore there must be at least one event in the course of the narrative. An event is, as defined above, a change in the state as described by a

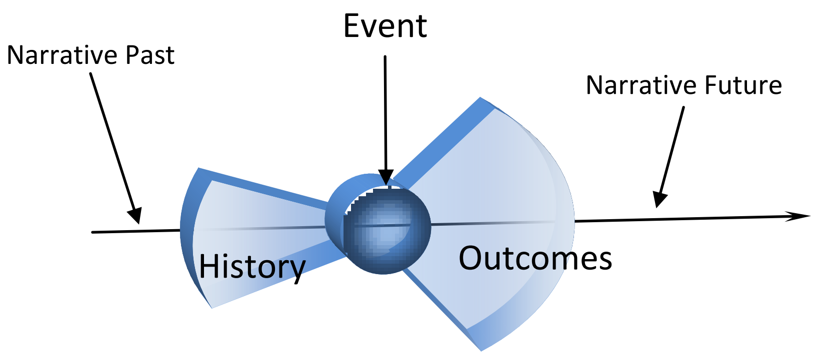



**Figure 3.1: Schematic Conceptualization of an Atomic Narrative**

narrative. The *state* of the narrative is a configuration of perceptions which describe objects, forces or concepts. These perceptions are referred to generically as narrative *variables*. A narrative with a single event is termed by Danto (1985) as an *atomic* narrative. A simple narrative made up of a chain of events (ordered by the time instants when the events occur) is then referred to as a *molecular* narrative. Figure 3.1 shows a schematic conception of an atomic narrative. To the left is an arc segment that represents the *history* of the narrative event. This is the state space that obtained before the event occurs. To the right is an arc representing the set of possible outcomes of the event. The history here, of course, is history in narrative time, and compared with history in real time, is woefully incomplete. That is, there must exist states in real time history that are not in a narrative history, no matter how thorough the recording of that history might be. An atomic narrative can have a set of *discrete* outcomes, or a *continuous* range of outcomes, as shown in Figure 3.2

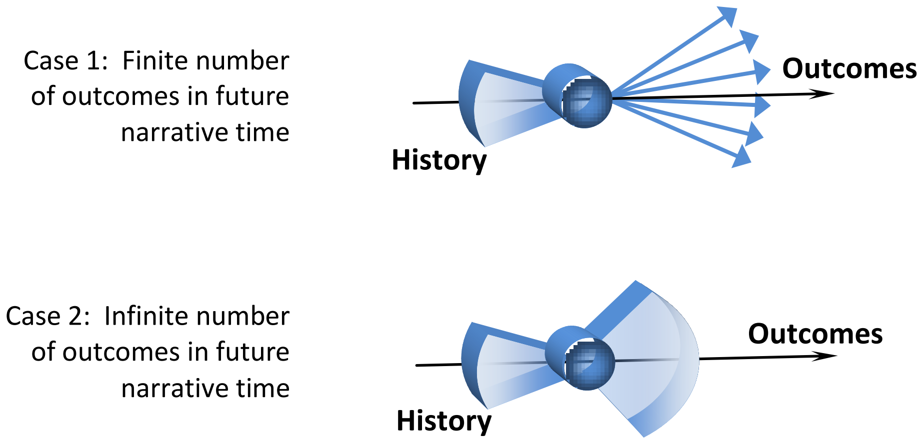



**Figure 3.2: Discrete and Continuous Outcome Sets**

**3.3.4:** That there can be more than one possible outcome is a vital property of the event. An event occurs when something in reality changes because the agent recognizing and responding to the event promulgates the change. The change is actually realized as an event outcome. Within a given narrative, there is a family of probability distributions associated with the outcome set of every event. In general, this family is a stochastic process indexed by a finite set of narrative variables referred to as *resources*, and is thus called the *event stochastic process*. In fact, that a variable is found in the index set of an event stochastic process is a useful, working definition of resource. The probabilistic relationship between the history, resources and outcome associated with an event can be denoted $P_\Lambda(Y|X)$ which is the probability that outcome $Y$ occurs given history $X$ and resource allocation $\Lambda$. A *molecular* narrative is simply of a sequence (in time) of atomic narratives. A narrative can contain multiple molecular narratives, simultaneous in narrative time, and these molecular narratives can coalesce into a single narrative, or diffuse into several. A narrative that contains such sub-narratives is referred to as *compound*. This structure of a molecular narrative is illustrated in Figure 3.3. As real time transpires, the course of the narrative is *realized*, which is illustrated in Figure 3.3 with the solid line connecting the relevant events.

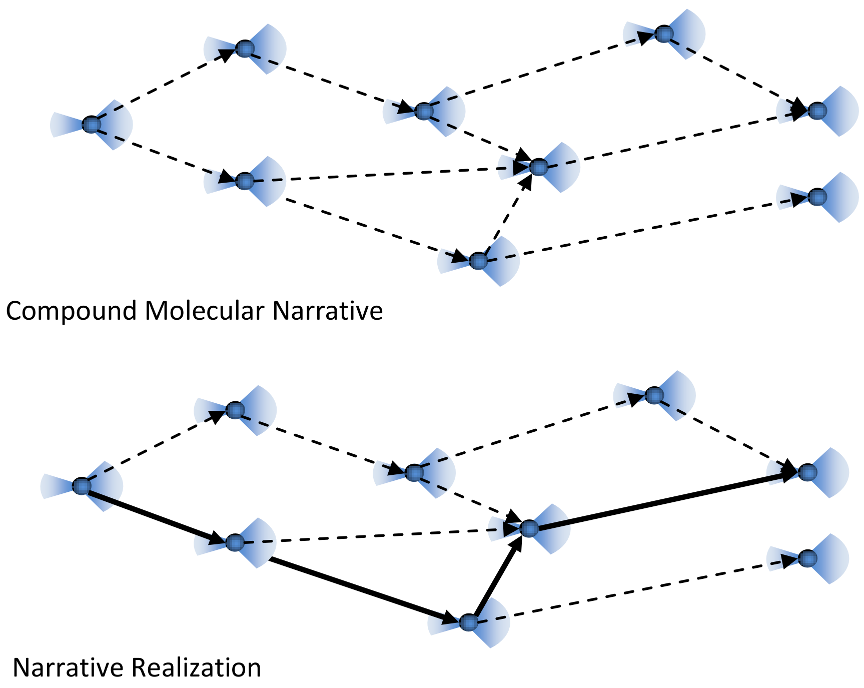



**Figure 3.3: Molecular Narratives**

**3.3.5:** In other words, narratives then can be viewed as memory structures that are retrieved given a current perceived state configuration. That invocation includes not only the present, but also a representation of the past and of the future. The future portion of the narrative is created from the contingency aspect of the expectation property of the narrative. Events, by definition, have multiple outcomes. If one outcome results, the narrative will head down one course, while if a different one obtains, it will follow a different path. The narrative carries with it a description of the expected consequences of these event outcomes. Thus the description of the future in a narrative consists of a set of branches, each opening a new course for the narrative to take. Furthermore, whether explicit or not, there is a probability distribution associated with each potential outcome, yielding a probability that the branch initiated by that outcome will be realized. The stochastic structure of the outcome probabilities cannot be assumed to be simple. The likelihood of outcome A versus outcome B may depend on outcomes that were the result of previous events, or sets of previous events. More succinctly, every atomic narrative has a probability distribution associated with its outcomes, but the distribution associated with the outcome of the last atomic narrative in a molecular narrative is not independent of the other atomic narratives which make up the molecular narrative. Thus it is required that every narrative have a past in order to contain the information required to support the future outcome stochastic processes.

**3.3.6:** But the narrative past is not a completely accurate description of what happened leading up to the present and conditioning the future. It has inherent inaccuracies caused by imperfect recall and incomplete data. Memory deteriorates as events recede into the past, and hence there must be a probability associated with the accuracy of the recall, and hence the past of a narrative also has an associated stochastic process

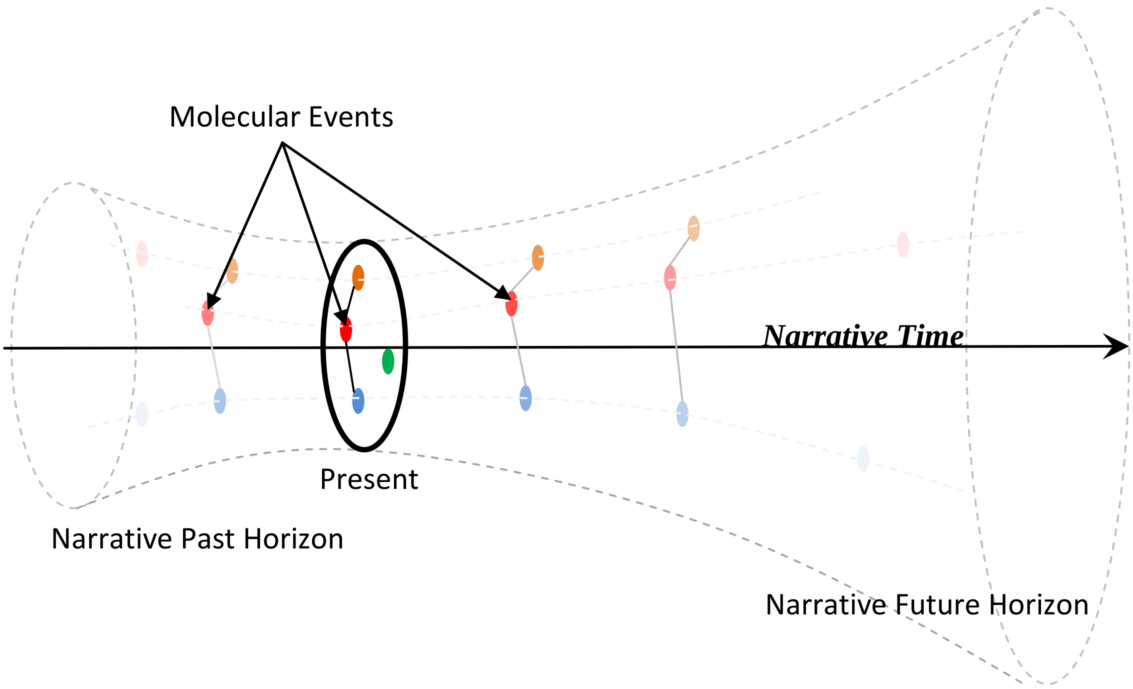



**Figure 3.4: The Time Structure of Narratives**

structure. A convenient way of envisioning the time relationships associated with a narrative is illustrated in Figure 3.4. In this conceptualization, both the past and the future are shown as the widening ends of a hyperbolic cone representing the increasing uncertainty surrounding the past and future of the narrative. The present is the narrowest part of the hyperbolic surface. In Figure 3.4, the small circles connected by the lines represent the events of the narrative and their relationships, as perceived at that point in the narrative. These event conditions can change from past to future, but the degree of uncertainty in the narrative regarding the stochastic structure of the events increases as the past recedes and the looming distant future is contemplated. This is illustrated in Figure 3.4 by the loss of resolution of the event description.

**3.3.7:** It is assumed that a particular individual agent carries around a number of narratives simultaneously, only employing the one that is useful in interpreting the current situation. A narrative that is selected and applied in given situation is said to be

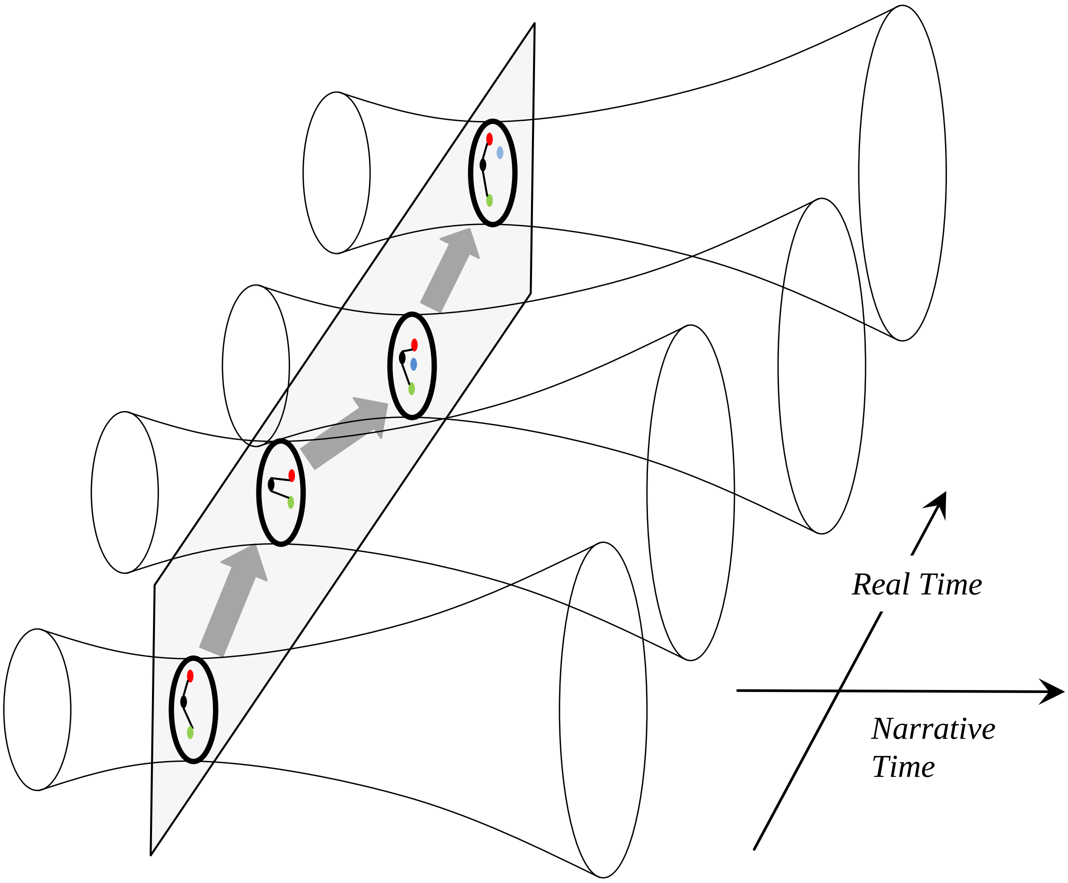



**Figure 3.5: Multiple Narratives Invoked over Real Time**

*invoked*. Not all narratives are active all the time, and as time passes, the environment in which a narrative would apply changes. This is illustrated conceptually in Figure 3.5. Invoking the narrative pattern in a particular circumstance requires that the state be updated with the current state of its variables, and the event stochastic processes be updated.

**3.3.8:** The individual that creates and maintains a narrative is called its *owner*. All narratives have owners, and all narratives are unique, to a greater or lesser extent, to their owners. But narratives are also shared. Language and communications are important adaptive tools for the human species, and a vital purpose served by language is the sharing of narratives. Indeed, if Fisher's hypothesis is accepted, it is the sole purpose of language, since all communications in his view is the description of a narrative. Shared narratives are not, however, perfect copies held by each individual that is part of the sharing group. Each individual owner at least modifies the shared narrative structure to fit their own unique set of compound narratives that include the shared one. Shared narratives become institutionalized into laws, codes of conduct, norms, and other social constructions that serve to assist in group adaptation and evolution. Shared narratives are a very powerful social and evolutionary tool. In fact, it could be argued that the interconnections of the atomic narratives that make up shared compound narratives may set the framework for the multilevel selection adaptation evidenced by group evolution. While substantially beyond the scope of this discussion, future research in this area using agents designed with the narrative construct might yield very useful results.

**3.3.9:** I maintain that narratives are mechanisms that are a result of human evolution. Much like the arguments presented by Sober and Wilson (1998) and Shermer (2004), the narrative hypothesis can be reinforced by evolutionary feasibility and their proven success as an adaptive device in support of both individual and group survival. This leads to the inference that the outcome of a narrative is either *desirable* or *undesirable*. In its most elemental form, a particular atomic narrative may consist of something like this:

- State description: Here is a plant with red berries.
- State description: I am hungry.
- State description: Some berries are food, some are poison.
- Event: I eat the berries.

- Outcome One:  I am no longer hungry.
- Outcome Two:  The berries are poison, and I die.
- Outcome Three:  Do nothing with the berries.  Go search for other food.

Outcome one is desirable, outcome two is not, based purely on its relationship to survival.  Outcome three may depend on the relationship between this atomic narrative and an earlier one, relating to some event where the owner learned something about the safety of red berries.  At the other extreme, the compound narratives that support religious beliefs and institutions can be invoked to portray a future where death is merely a change in physical state from this world to another, which may be Heaven (desirable), Purgatory (not so well favored) or Hell (clearly undesirable).  Given this property of desirability, narratives then can be considered to be the vehicle by which value is expressed by the individual, and it is from narratives that values arise as identifiable attributes of human behavior.  This construction fits well with Fisher's assertion of narratives as value-carrying mechanisms (discussed in Appendix F)**.**

**3.3.10:** The narrative framework proposed here can be considered as a storage device for not only memories, but also expectations.  The role of expectation in narrative structure – the projection of the effect associated with some cause in the future – is quite clear.  The maintenance of a set of expectations associated with the various outcomes in a particular event-based choice situation is required for the mental storage of such expectations.  It is implying too much inherent human analytic ability to assert that each such expectation set has a formal probability distribution also maintained as part of its mental representation, especially since such distributions would be dependent on the pathway by which the event itself arose, and not only on the fact that the event came about at all.  Nonetheless, narratives provide both a context and framework within which expectation can be stored, recalled and manipulated.  It will be seen in Section 3.9, in fact, that humans are prone to a number of significant deficiencies (e. g. base rate error discussed in **3.9.4**) when contemplating formal probabilities.  However, for the purposes of modeling human agents, the expectation-storage property is important. Indeed, one of the most useful applications of an agent model might be that both the "correct (logical)" application of probabilities, or the "deficient (logically false but commonly held)," expectation distributions can be modeled, and the results as expressed in specific choice contexts compared.

**3.3.11:** Recall that the individual is part of a world and is in constant interaction with it. The intervention in the real world by the agent occurs by means of the allocation resources to the event at hand in order to alter its perceived probability distribution of the outcomes, either toward those it favors, or away from those it dislikes. In this way, the values held in the narrative are expressed in action. Since compound narratives are molecular, and thus made up of sequences of atomic narratives, any attempt by an agent to affect what it thinks will be the desired course of a narrative involves choosing a specific outcome from the set of outcomes of the narrative representation of that event. And it must do so one event at a time. Otherwise contradictions, uncontrolled feedback, process deadlock or other dangerous anomalies would emerge. *Thus, in the narrative framework, the relationship between the narrative owner and its environment is always an event, and therefore always a choice structure.*

**3.3.12:** Notice there is absolutely no requirement that any of the events or sequences of events described by a narrative be true, in the sense that there is Karl Popper's intersubjective verifiability of narrative content (Popper, 1959/2002, pp. 79-82). Indeed, everyone has experienced erratic and unpredictable behavior on the part of others. Such behavior is seen because the narrative driving the erratic individual is different in some important way from what the observer expects, given her constellation of narratives. Since narratives exist as neurological entities (fundamentally memories), study of their physical existence and attendant properties resides in the domain of the neuroscientist. But what is represented by the narrative pattern, and whether or not that has any reality outside of the narrative itself, is an entirely separate question. It can be as fanciful or factual as the owner wishes.

**3.3.13:** Moreover, all narratives are models, in the sense discussed in **2.2.1**. Therefore they never correspond in all regards with all characteristics of a real situation, but are always to some extent approximations, and thus they are always in some measure wrong. It is with this conception in mind that Fisher incorporates the rationality of traditional logic into the narrative paradigm (with a somewhat critical slant):

> "Narrative rationality is not simply an account of the 'laws of thought,' nor is it normative in the sense that one *must* reason according to prescribed rules of calculation or inference making. Traditional rationality prescribes the way people should think when they reason truly or toward certainty. … Traditional rationality is, therefore, a normative construct. Narrative rationality is, on the other hand, descriptive; it

> offers an account, an understanding, of any instance of human choice and action, including science." (Fisher, 1987, p 66).

**3.3.14:** A "plan" is a definition of what a future state should be. It can also be a blueprint on which to pattern something. But that something does not now exist. One important element of a plan is a process, or sequential set of steps, to get from the current state to a future state. That is, it is necessary to describe not only what the anticipated future state will be, but also how to get from here to there. Furthermore, effective planning requires one additional component: an evaluation process. Evaluation itself has two questions to answer. How do the results of the course of action compare with the desired future state encompassed in the plan? And, given progress in accordance to the steps in the plan, is the desired future state still as desired as at the outset? Both can be addressed from the perspective of a narrative, and useful answers can be derived therefrom. In fact, those familiar with the construction and utilization of scenario writing as a planning motif will immediately recognize the narrative qualities of the future history so often associated with such approaches. See Schoemaker (1995) or van der Heijden (1996) for further discussion.

**3.3.15:** How does this narrative construct relate in the modeling of human agents in virtual markets? As was stated at the beginning of this section, the narrative provides the mechanism for separating reality from what the agent *thinks* is reality. For a narrative that incorporates in its state description some aspect of the external agent environment, the plan and its attendant sequence of steps determine what choices the agent has available to it for action. That is, it defines the context, values and resources required to change the realization of a narrative to a desired outcome. There is much to be done before there is sufficient data to characterize all the observed marketing behavior, such as that noted by Cialdini (2001) in Section 2.9, into the formal narrative construct put forth here. Many in marketing consider individual narrative discovery as the central problem of marketing research, as exemplified by the strong methodological presence of ethnology in some marketing research quarters, such as ESOMAR (a European society of marketing research). That seems to me to be excessively ambitious given the current state of the data. But it is clear that the choice process is a vital aspect of any such description. In fact, the narrative framework provides an ontological justification for pursuing the study of choice as the critical component of replicating the behavior of humans in agent-based models. It is not necessary to know the full constellation of narratives maintained by an individual to incorporate the concept into agent models. It is often sufficient, at least as a starting point, to model a few essential atomic narratives.

**3.3.16:** Formally speaking, all narratives are models, in the sense of Section 2.2.[37] Moreover, since a narrative is a sequence of events, there exists a finite sequence *{1, 2, 3, ... k}*, each member of which has an associated stochastic process $E_i = P_{\Lambda(i)}(Y|X_i)$. So a narrative can be expressed as the ordered *k*-tuple $(E_1, E_2, E_3, \ldots, E_k)$. Of major importance in describing the relationship between the various branches of a compound narrative it its dynamic, stochastic structure. Bayesian methods are an obvious choice, as are Markov models. Bayesian methods are explored in depth by Rossi *et al* (2005). Markov chains, (Markov processes with a finite number of outcomes) suggest themselves as descriptive mechanisms. Discrete choice models can be employed to create empirically-based Markov state probabilities. Adaptive discrete choice models, such as that alluded to in the discussion of the pags of AirVM (Chapter 6) can also capture the dynamics of compound narrative structure. The precise form of the formal narrative structure, at least at this stage of development, seems to depend heavily on the specific market of interest in the construction of the agent-based model.

**3.3.17:** Narratives are needed for the construction of agents to support marketing science applications because the distinction between the *reality* of the environment that an agent finds itself in and the *perception* and *interpretation* of that environment by the agent must be kept clear. In summary, then, these are the main points of the narrative framework:

> **3.3.17.1:** Human beings create time-framed cause-and-effect chains called narratives to understand, explain and communicate both the real world around them and their perception of it.
>
> **3.3.17.2:** Narratives are made up of a state space with one or more changes in that space, called an event. A narrative with a single event is called an atomic narrative.
>
> **3.3.17.3:** Atomic narratives can be linked in a time-dependent sequence to form a molecular narrative. Molecular narratives can be further tied together to form compound narratives, which may be simultaneous in narrative time.

[37] But not all models are narratives. There is no need for a time dimension in a many models.

**3.3.17.4:** Events have a set of possible outcomes to which the individual who maintains the narrative – the narrative owner – ascribes, either consciously or not, a probability distribution.

**3.3.17.5:** Narrative owners assign value to narrative outcomes in that some are favorable and others less so. The value of a molecular narrative outcome is a function of the favorability of its constituent atomic narratives.

**3.3.17.5:** Two time dimensions are maintained in the narrative framework; the time line of the narrative, and the time line of the real world. The narrative exists only in the narrative time line.

**3.3.17.6:** The narratives maintained by an individual are applied to that individual's perception of the real world, which they intersect. The narrative owner allocates resources to affect the real-world outcome of an event, either favorably or unfavorably, as determined by the associated probability distribution.

**3.3.17.7:** Thus the points of intersection between the narrative structure and the real world always are at points of choice. This may mean simply the use of a simple discrete choice model, or the engagement of a more complex Bayesian, Markovian or other dynamic stochastic model formulation, depending on the skill of the modeler and the nature of the available data.

## 3.4: Comparison of the Narrative Framework with Brown's Human Universals.

**3.4.1:** Having posited the narrative framework as the mental structure which drives human marketing behavior, how well does the construct do in the face of the empirical data of how humans behave in the marketplace? This question is addressed in this section of the paper. The human universals and the observed behavior classified by Cialdini (2001) in Chapter 2 are examined now to see if they have reasonable explanations within the narrative construct. If so, then an empirical argument is created which supports the validity of at least this simple version of the construct. This is the topic of this last section in the chapter.

**3.4.2:** The narrative framework should be consistent with the human universals presented in Chapter 2. While not necessarily directly supported or explained by the construct, the universal trait also should not be contradicted directly or by the implications of the construct. Some may have no relevance. Each will be examined with this test in mind. The center column in each of the tables that follows classifies that particular universal as either supporting or is supported by the narrative construct (coded S), contradicts or is contradicted by the narrative (coded C), or is not relevant (coded N). Recalling Section **2.8.6,** each family of universals will be examined in the same order as presented there.

**Table 3.1: The Culture Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| **1** | *Humans are aware of culture, and behaviors and thought processes are spread among individuals and from generation to generation by learning.* | S | This is an attribute of the shared narrative structure discussed in **3.3.8.** It reflects both trial and error learning within a relatively stable surrounding environment, and the central place narratives hold in communications theory. |
| **2** | *We are aware of our apparent uniqueness among known living things in having culture, although it is not strictly true.* | S | Uniqueness is a state in a narrative. Supports the "us-them" requirement of group identification. |
| **3** | *All human cultures use speech, often ceremonial or otherwise distinguished from ordinary, day-to-day usage, to tell stories that explain how things came to be or what they will become.* | S | This is, of course, an essential feature of narrative in the classic, literary sense of the word. It also supports the role of shared narratives. |
| **4** | *All human cultures share a common external world and internal existence.* | N | Undoubtedly the material from which narratives arise, and required for shared narrative structure. But would support any other kind of communications mechanism and is true of all biological entities. |

**3.4.3:** First consider the family of four human universals that were associated with culture. The narrative context of each is described in Table 3.1. Clearly the universal existence of culture and its equally universal manifestations reflect the impact of a narrative structure. However, number four – shared common world – seems independent of any idea of memory or mental construction.

**3.4.4:** Language, of course, is one of the primary vehicles for narrative expression and communication. So it is not unexpected that language universals are tightly connected with the idea of narrative. As can be seen in Table 3.2, the universal use of language, that it is an independent entity different from the things it is used to denote, its use as a manipulative tool, and its role in group structure all fit the narrative scope. Specific language attributes that are common across cultures, gender specification for instance, seem to neither contribute to nor contradict the narrative construct, but are rather universally the topic of narrative; that is, reside in the state space and have common events associated with them. Perhaps the most relevant universal characteristic of language, however, is the extensive use of metaphor (universal 23). Metaphor, of course, is intimately connected with narrative in that a metaphorical reference imputes the cause-and-effect chain of one narrative from the cause-and-effect chain of another, more familiar narrative, thus assisting in the communications of the narrative. Also, in many regards, the term metaphor is synonymous with narrative.

**Table 3.2: The Language Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| **5** | *All human cultures have language. We use language to think about and discuss among each other both our internal states and the states of the external world around them.* | S | The existence of language is fundamental to the creation of narratives. In fact, language makes the communications of narratives possible. |
| **6** | *Language is not a perfect representation of a human's views or thoughts. There are discrepancies between what we think, and what we say and do.* | S | This universal is also a reflection of the separate existence of narratives *per se*. |
| **7** | *Because the use of language is not a literal description of the world and our place in it, we must, and do, distinguish between the world as it actually is and the world as we conceptualize it.* | S | Narratives exist independently from reality. They can therefore propose cause-and-effect sequences, alternative choice outcomes, and fictional structures. They can be used to communicate imagination as well as relating apparent facts. And they can be used for argument and other rhetorical communications. This universal is the basis for "good reason." |
| **8** | *Humans use language to organize, respond to and manipulate other humans* | S | These characteristics are representative of the key features of a narrative construct, especially in the context of rhetorical argument. |
| **9** | *An important means of verbal manipulation is gossip.* | S | Gossip often relies on (fictional) stories made up to misrepresent the actions or attitudes of another. These are, of course, examples of narrative. |
| **10** | *Language is used to misinform as well as inform.* | S | Another example of the distinction between reality and narrative. |
| **11** | *All humans lie and mislead others, some to a greater extent than most.* | N | Other than the mechanics, there is no apparent relationship. |
| **12** | *All humans understand what a lie is and have methods of detecting it when it occurs.* | S | This is substantial evidence for the existence of narratives, in that a lie detection scheme would have to depend, to some extent, on the differences between what someone says and what the recipient expects to be said. |
| **13** | *Those who are more proficient in the use of language are of higher status, in the view of others, and are more capable of manipulating the behavior of others.* | S | This supports the notion of narratives rhetorical as rhetorical instruments. |
| **14** | *Language is highly symbolic in structure, and even though the exact sounds and words used to denote a given entity or concept may differ, rather arbitrarily, from one group to another, all languages contain a basic set of common features. Thus different cultures can communicate.* | S | This substantiates the existence and value of shared narratives among all people in all cultures, and reinforces the narrative as a construct of broad application. |
| **15** | *All languages have a similar structure in that they contain nouns, verbs, and possessive forms.* | S | Similar to 14, this supports the universal use of narratives. |
| **16** | *A few concepts are found in all languages, such as face, black and white, male and female.* | N | Other than narrative subject matter, no apparent relationship. |
| **17** | *Male and female gender is always distinguished in languages, creating a sexual content in all human language.* | N | Other than subject matter, no apparent relationship. |

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| 18 | *Our sexual terminology is dualistic – male and female. Other terms can portray different sexual contexts, but only in reference to this basic duality.* | N | As above, no apparent relationship. However, this may speak to the role of narratives as essential elements of the intercommunications structure which leads to behavioral evolution. |
| 19 | *All languages describe family and kinship relationships in terms directly related to procreation, such as father, mother, daughter, etc.* | N | As above, including the potential role in behavioral evolution. |
| 20 | *All languages express the concept of time, and units thereof.* | S | Time is, of course, an essential feature of the narrative construct. |
| 21 | *All language contains groups of contrasting terms that could be expressed in three ways.* | S | Goes to the state space and event structure of a narrative. |
| 22 | *Language changes over time, adopting new words and relegating others to disuse.* | S | Addresses the adaptive ability of the narrative construct. |
| 23 | *All languages make extensive use of metaphors* | S | This is essentially a restatement of the definition of narrative. In other words, narratives are a feature of all known languages, and therefore all known human communications. This may be the most powerful evidence supporting the narrative argument. |

**3.4.5:** The Analysis family of universals (shown in Table 3.3) also contains traits that are important elements of the narrative construct, and therefore lend weight to the validity of the idea. Of particular note is the learning by trial and error, which creates and supports the existence of expectation, required for narrative structure. Also important is the universal availability of basic logical inference and deductive ability demonstrated by universals 31 and 32. Traits that establish the ability to measure, create distinctions, and create and manage taxonomies are also essential to the construction and use of narratives. These capabilities are necessary skills if narrative is to be a functional entity. That they can be found among all human cultures argues persuasively that the narrative is also a universal capability.

**Table 3.3: The Analysis Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| 24 | *Humans learn by trial and error, by which is meant if some event or sequence of events does not occur as expected, our expectations change, not the perceived events or sequence of events.* | S | This reinforces the concept that a narrative is a repository of expectations, and that the existence of a narrative structure is a valid construct. It represents an instantiation of the concept of expectation and how it is created. |
| 25 | *Humans can measure and express the size of things.* | S | State spaces require that things be categorized and classified. Continuous state variables, in particular, require measurement. |
| 26 | *All human cultures use language to describe physical properties –speed, motion, dimensionality – and actions, such as giving, lending, and affecting* | S | Those elements from Table 3.2 which discuss language bear equally well here. |

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| | *other things and other people,* | | |
| **27** | *All human languages have terms which identify and discuss parts of the body, internal states such as emotions or thoughts, behavior, the physical world, the weather, tools, and space.* | S | These are necessary material for the construction of narratives. |
| **28** | *An essential feature of our ability to recognize, learn and understand the world is that we can identify distinctions and name them; that is, we can create taxonomies.* | S | An expression of the narrative requirement of state spaces. |
| **29** | *The basic form of distinction is that of binary discriminations, the basis for all categorizations (a thing either is, or is not, in a specified category, or has, or has not, a particular property).* | S | As above, implied by the nature of state spaces. |
| **30** | *All human cultures can also classify things that do not necessarily fall into discrete categories by ordering them along some kind of continuum.* | S | This implies that state spaces must be capable of representing continuous concepts. And such continuity is required for a proper expression of narrative as well as real time. |
| **31** | *All human cultures use basic logic constructions such as "not," "and," "same," "if … then," "equivalent," and "opposite."* | S | These are expressions that lead to the communications of cause-and-effect. Thus a narrative structure provides content for these terms, and use these terms to organize the narrative. |
| **32** | *Humans have the ability to deduce clues the current state and future condition of things, often very inaccurately.* | S | In many respects, this is the purpose of a narrative. The deductions stem from fitting the perceived state space into a narrative structure. |

**3.4.6:** The universals classified as Recognition of Mind deal directly with the ability of the human species to both construct narratives, use them as a vehicle for the capture and recall of causal chains, and communicate them between members of a group. These are discussed in Table 3.34. One of the traits in this family, however, seems to contradict the narrative concept. In fact it is the only trait in the set of 76 that does. It is the pervasive use of body language in communication (universal 38). This behavior suggests that the structure of a narrative may consist of content other than that which can be verbalized into language. There is no doubt this capability exists, but it is unclear if it is integral to narratives as defined here, or merely perceptions, that when converted to verbal form, become part of the narrative. The other universals of this family, however, speak directly to the means and mechanisms by which shared narratives come about, are recognized, and are communicated among a group. As noted earlier, the narrative structure may also imply important features of how group adaptation and evolution works. Finally, included in this family is the use of mind-altering substances such as alcohol and other drugs. That these external agents have a significant effect of on mental processes is clear. They clearly also have substantial impact on what is desired, or considered good. Thus they change the inherent structure of a narrative by altering the basis of value which is represented in the probability processes associated with event outcomes. They undoubtedly affect the perception of the state space, and even the nature of the outcomes, tied to an event as well.

**Table 3.4: The Recognition of Mind Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| 33 | *All individuals have a distinct concept of person and of themselves separate from others.* | N | This is, of course, required for the idea of group to make sense. While the narrative contributes to this conceptualization, the distinction neither supports nor contradicts the narrative construct. |
| 34 | *We can easily and intuitively get into the minds of others and imagine what they are thinking and feeling.* | S | This narrative is a method for executing this capability. |
| 35 | *We know that other people are like us have inner thoughts, make plans, and make decisions and choice.* | S | As noted above, the narrative construct could be a mechanism to allow this empathy to exist. |
| 36 | *We are able to think not only of our own relationships with others, but of the relationships between others and themselves.* | S | This is direct outcome of the narrative structure. |
| 37 | *All of us also mask or modify expressions to mislead and confuse.* | N | Narratives could contribute to this skill, but not necessarily. |
| 38 | *All of us communicate non-verbally, especially with facial expressions.* | C | This contradicts the narrative concept in that it suggest within group communications which does not conform to the narrative framework, at least as it has been offered here. |
| 49 | *We respond to sexual attraction.* | N | This is more the material of narratives, rather than supportive or contradictory to the concept. |
| 40 | *We use reciprocal exchanges in all aspects of our lives. This includes its negative forms, such as retaliation and revenge.* | S | Reciprocation is based on expected future action, which is directly addressed in the narrative construct. |
| 41 | *We are able to distinguish between normal and abnormal mental states.* | S | Abnormal may be defined as sufficiently strong exceptions to shared narratives. |
| 42 | *Humans have ways of making themselves feeling better, including stimulants, narcotics, and intoxicants* | S | The idea of "feeling better" can be defined as the desirable outcome of a narrative. Assistance in doing so with the use of chemicals is a ready example of a specific action to support that definition. |

**3.4.7:** The small Tools Family of universal traits (Table 3.5) includes aspects of resource acquisition and allocation. Tools give humans a much wider availability of a broader range of resources, and thus enable a much wider variety of narrative courses. In particular, the universality of shelter from the weather illustrates the application of a set of shared narratives to (relatively) simple resource allocation problem.

**Table 3.5: The Tools Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| 43 | *We make and use far more tools than any other animal, and use tools to make other tools.* | S | Tools allow us to more flexibly allocate resources to generate outcomes deemed desirable by the associated narrative. |
| 44 | *We know how to use fire.* | S | Clearly a resource. |
| 45 | *All cultures have means of shelter from the elements* | S | Also an example of narrative-driven resource allocation. |

**3.4.8:** Group Association (Table 3.6) traits speak to the relationship between multilevel adaptation and the narrative framework. Some of these behaviors merely confirm that living in groups is a general property of the human species. But others reinforce the role of the narrative in the sociology of the group. These include the education of children and the formation and characteristics of group hierarchy, status and prestige. The ethical dualism mentioned in universal 55 is, however, a direct result of the evolution of ethics as described earlier by Shermer (2004). It is further interesting to note that the traits of government in universals 56 and 57 are capable of being recognized because of the existence of shared narratives. These shared narratives act to describe the beneficial features of these hypothetical outcomes, but actual, complete realization of those ideals are not to be found.

**3.4.9:** The fundamental relationships that create and maintain markets are contained in the shared narratives that are implied by the Trade and Economics family (Table 3.7). The existence of economics, division of labor, materialism and the inequality of the distribution of material wealth are all reflective of the resource allocation requirement of the narrative construct. But universal 59, the definition of trade as the exchange of goods not based on future reciprocation, I believe is incorrect. I would argue that every exchange as an element of reciprocal behavior, even if it is no more than shared expectations of value received for value given. Cheating could not occur or be recognized without expectation. Reciprocity is irrelevant, at least in the sense of reciprocity as an altruistic activity. Envy and generosity, however, represent particular resource allocation characteristics. Envy is an expression on the desire for resources, while generosity to some degree opposite in that it reflects a redistribution of resources based on factors other than event management, such as expressions of affection or respect. Again, a narrative drives both of these traits.

**Table 3.6: The Group Association Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| 46 | *Most of us live part or all of our lives in groups.* | N | The narrative is not necessary for group formation. Many animals have other means of group association. |
| 47 | *We have a sense of territory associated with the groups to which we belong, and are well adapted to the environment associated with that territory.* | N | Territoriality does not require narrative ability. |
| 48 | *There is a distinct sense of "us" and "them" between the groups to which we belong and other groups.* | N | Group membership is does not speak to the existence of properties of narratives. |
| 49 | *We judge members of other groups in terms of those qualities found in our own.* | S | Judgment is an exercise of expectation, and therefore supports narrative skill. |
| 50 | *Marriage, in the sense of a well understood protocol for access to females of child-bearing age, exists in all cultures.* | N | While many narratives support marriage, long-term mating by species is common without any demonstration of narrative ability. |
| 51 | *We have a pattern of socialization in the sense that child-rearing responsibilities are shared among adults, who are obliged to teach the young.* | S | Most teaching is done by narrative, although it is not the only method, so is not strong evidence. |
| 52 | *Children learn from adults by copying and mimicry.* | S | This is the essence of a shared narrative. |
| 53 | *The existence of roles and social structure is implied by kinship, sex and age statuses found in all cultures.* | S | It can be argued that these factors are the subject of narrative development, and hence that status is a derivative of narrative. |
| 54 | *Prestige is not equally distributed among group members.* | S | As above, prestige is a facet of status. |
| 55 | *An ethical dualism exists between the groups to which we belong and others.* | S | The dualism is based on the idea that members of our group share our views – our narratives – and hence the dualism supports the narrative mechanism. |
| 56 | *We have leaders, but they are never completely powerful.* | S | The concept of leader, and the power they hold, can only exist when a notion of future and expectation exists. |
| 57 | *There is never a complete democracy nor an absolute autocracy. Thus there is always an oligarchy* | S | Democracy, autocracy, even oligarchy are concepts that exist because of narrative. That there is no complete democracy or autocracy speaks to the imperfect replication of narratives shared among a group. |

**3.4.10:** The final family in the human universal collection is the Ethics, Morals, Arts and Metaphysics. These traits represent expression of narrative structures that are clearly untrue, thus substantiating that characteristic of the narrative construct. The universals either speak to the existence of such narratives, or discuss specific subject matter of such narratives. The only universal of this set which does not seem to be relevant to this analysis is the aspect of sexual modesty. This would seem to be an artifact of a narrative that is shared among all human cultures. But it could also stem from more basic sources, since other animal species, apes in particular, seem to have expressions of modesty in their social interaction.

**Table 3.7: The Trade and Economics Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| 58 | *All societies have economies – systems for barter or trade.* | S | Trade exists because of shared expectations. In exchange for this good or service now, I am confident *will* receive that good or service. |
| 59 | *We engage in trade, defined as exchange of goods that is not based on future expected reciprocal behavior.* | N | This universal may, in fact, not be true. The definition of trade here – exchange not based on reciprocation – seems itself inconsistent. Exchange always depends on expectation. |
| 60 | *All societies have divisions of labor, and customs of cooperative labor.* | S | This is a reflection of shared narrative. |
| 61 | *We are all materialists to some extent, and distinguish who owns what.* | S | Reflects the resource allocation requirements of actions based on narrative events. |
| 62 | *Members of groups are not economically equal, in that some have more materiel resources than others.* | S | A reflection of the impact of real world events and outcomes on individual narratives. |
| 63 | *We are envious.* | S | A particular result of need for resource allocation. |
| 64 | *We give gifts* | S | Another aspect of resource allocation. |

## 3.5: Comparison of the Narrative Framework and Cialdini's Marketing Behaviors

**3.5.1:** The alignment of the content of the 76 universals just discussed creates a substantial argument in support of the existence and effects of the narrative construct. What of those behaviors are aimed more directly at marketing, as explored by Cialdini (2001) in Chapter 2? To determine this, I will reexamine his seven behavior patterns – fixed-response, reciprocation, commitment and consistency, social proof, liking, authority, and scarcity – in light of the narrative construct. Fixed response can be viewed as the conformance to a narrative construct with an outcome set that has one of its members assigned the probability of one. It does not mean that the triggering event does not create the possibility of other outcomes, merely that the owner of the narrative does not assign any probability to other options. This differs somewhat from Cialdini's assertion that the automatic response is initiated by environmental stimuli. In narrative construct terms it is more precise to say that the environment, as perceived by the narrative owner in his state space, creates an event for which the invoked narrative has only one non-zero probability outcome.

**Table 3.8: The Ethics, Morals, Art and Metaphysics Family of Human Universals**

| No. | Universal | S/C/N | Comment |
|---|---|---|---|
| 65 | *We all have a definition of, and can distinguish, right from wrong.* | S | Shermer (2004) establishes that this is a result of evolution. The narrative capability arises because of our biological heritage, and it can be asserted that this is one of the fundamental narratives. |
| 66 | *All cultures have standards of sexual modesty.* | N | The concept of modesty may or may not be a product of narrative. Some other animals have exhibited characteristics of modesty. |
| 67 | *We all believe things that are clearly and demonstrably false.* | S | Narratives don't have to be falsifiably true. In fact, there is no relationship, in general, between objective truth and narrative truth. |
| 68 | *We all practice magic, especially in trying to prolong our lives or attract others in a sexual context.* | S | Magic exists only in a narrative form, often as mysterious resources or causes that defy 'logical' explanation. |
| 69 | *All humans attempt to predict and to plan the future.* | S | This is a primary function of a narrative. |
| 70 | *All cultures have theories of fortune and misfortune.* | S | The distinction between desired and undesired outcomes, an inherent part of narrative. |
| 71 | *We attempt to control the weather.* | S | A particular form resource allocation |
| 72 | *We all have a coherent world view, in that we understand the world as a unitary whole regardless of the sensory mode by which it is experienced.* | S | This trait can be considered as a direct result of the narrative as a cause-and-effect storage mechanism. |
| 73 | *Our worldview plays into our mythology and concepts of the supernatural.* | S | This is a manifestation of a particular application of (untrue) narratives. |
| 74 | *All cultures have rituals, especially to demarcate changes in states, and all cultures mourn the dead.* | S | Rituals exist because of their role in appropriate narratives. Mourning of the dead, however, is probably not narrative related, except in the basic sense of loved ones being a major part of a narrative. Other species also mourn the dead, and it is unclear if narrative plays a major role in their evolution. |
| 75 | *All cultures have aesthetic standards and preferences.* | S | These are a direct result of narratives. |
| 76 | *All cultures have music, poetry, play and story-telling* | S | These are the standard forms of the literary narrative. |

**3.5.2:** Reciprocation is a trait that lends significant support to the narrative concept, because it employs resource allocation and expectation to achieve the desired outcome. Unlike what Cialdini (2001) suggests, however, it does not have to be initiated by an automatic response mechanism. It can also arise as an intentional allocation decision. That reciprocation calls for repayment of more than originally put forth illustrates the future expectation and planning associated with the narrative that is supporting the reciprocal sequence. The phenomenon also is supported by shared narratives, since it is difficult to see how the transactions could arise without a mutual understanding of the course of the future relationship. Moreover, that altruism is a behavior resulting from evolutionary results argues in favor of narratives also being of such origin, which in turn supports their universality. It could be further hypothesized that it is the narrative basis of reciprocation that is the primary vehicle enabling that evolutionary development.

**3.5.3:** As noted earlier, commitment and consistency are behavior patterns that increase the likelihood that an individual will be able to predict and deal with the future behavior of another. These properties are clearly narrative elements, engaging cause-and-effect, resource and value components of the narrative structure. Indeed, these behaviors can only exist if they are tied to a narrative, in that the expectations and results of both are expressed as future-time favorable resource allocations, an essential aspect of the narrative construct. And the importance of commitment and consistency, as they play a central role in the development of social norms and standards, must also be viewed within the context of the shared narratives they represent.

**3.5.4:** The pattern of social proof can be explained by the concept of shared narrative. That narratives need not be true implies, of course, that shared narratives also need not be true, which could result in pluralistic ignorance. Suppose a set of circumstances creates an event which requires action, and that set of circumstances also provides a context for the application of a shared narrative, such as in a crowd of people. Then if an event occurs which is completely unexpected, there is no shared narrative to guide the choice of event outcomes, and inaction results.

**3.5.5:** Similarity is a clear manifestation of shared narrative. That, in fact, is what the term means in this context. That similarity can range over a number of characteristics, including obvious ones such as race, age, gender, and income, reinforces the shared narrative notion if it is reasonably assumed that such characteristics support appropriately shared narratives.

**3.5.6:** To some extent similarity in a stronger form is the basis for liking. But unlike similarity, liking can arise within the context of narratives in another way. This could be the case with physically attractive people. Attractive people will occur as part of another's individual narratives. Surely all of us are owners of narratives that would be classified as sexual fantasies. Physically attractive people must often represent desirable values of state spaces in such narratives (otherwise there would be no such fantasies), and so it should be no surprise that they would enter other individual narratives. Thus, while liking certainly has a strong shared narrative flavor, it can also exist in private narratives. Indeed, in an extreme case, the stalker could be characterized as one who has incorporated a particular individual into a narrative that requires the person being liked as part of the state space to the point of personal danger.

**3.5.7:** Authority is another trait that arises out of shared narratives. This was mentioned above when discussing one of the group association universal traits, the emergence of leaders. To carry the thought further, leadership implies authority in that the choice of outcome in selected events (in the narrative) is left to another individual, one in authority. Thus the authority role is assumed by an element that exists in the narrative of its owner. The leader in the narrative may be identified with an individual in real time who in fact has no such authority, but the narrative owner will behave as though they did.

**3.5.8:** The last of Cialdini's (2001) influence factors is scarcity. Instead of falling back on the freedom argument to support scarcity, a narrative argument can be invoked that may be more convincing. A resource that has high value in the narrative structure of an individual will assume more value if there is less of it simply because the demands of the narrative that supports its value have not changed. The resource becomes more important because it is now less likely to be available, but the purpose within the narrative has not changed the amount needed. I suggest this may be a more useful explanation of scarcity than psychological reactance.

**3.5.9:** The comparison of the features and traits that are encompassed in Cialdini's (2001) presentation of important marketing behaviors thus seem to stand well in the company of the narrative construct. Most of the essential marketing traits can be reasonably viewed as expressions of characteristics of the narrative. This, I submit, provides additional empirical evidence of the existence and central properties of the construct. It is reasonable, then, to base the general definition of a human agent in a virtual market on the narrative construct.

## 3.6: A Sampling of Rational Choice Protocols

**3.6.1:** The basic unit of a narrative – the atomic narrative – is a change in the state space element of the narrative called an event. An event can have a number of outcomes, some desirable and others not so desirable to the narrative owner, and to which the owner applies a probability distribution. Moreover, individuals or agents have at their disposal resources that can be applied to the intersection of the narrative and the real world to alter the probabilities of desirable event outcomes. These narratives are also the mechanisms by which agents communicate with each other – so-called shared narratives – and understand the world around them. They need not, however, be true descriptions of the world. It is hypothesized that all narratives are compounded from sequences of atomic events (molecular narratives) and, most importantly, from a marketing perspective, the choice process of the atomic narrative is a key focus. Since it is human behavior that is being modeled, there are a number of scholars who have studied human decision-making that can be drawn upon, including Simon (1957, 1996, 1997), Tversky (1972), Gigerenzer (2000), Gigerenzer and Selten (2001), Kunda (1999), and a many others. In this discussion, the set of choice protocols are classified into four broad groups: 1) rational methods, including various concepts of bounded rationality, 2) heuristics, which are quick and easy (but often very inaccurate) rules-of-thumb, 3) social network protocols, which rely on communications between individuals to make choices, and 4) biases, which are significant errors often found in choice-making. However, this survey is by no means complete, but does serve as an introduction and is complete enough for the development of virtual markets as they are described in the remaining discussion of the dissertation.

**3.6.2:** Consider what are usually referred to as protocols for *rational* choice. Perhaps the easiest of these would be rule-invocation methods. This is the situation where the agent making the choice has an available set of rules, and, depending on value of the state space when the event is encountered, one or more of these rules are used to determine the choice. The fixed response choice structures described by Cialdini (2001) are rule-invocation protocols, for example. Rule-based agents are widely used in agent based models. Axelrod (1997), Epstein (2006a), and Wooldridge (2002) insist that rule-based agency is the wisest course of agent construction. This press for simplicity is in response to the need to explore and understand some of the unusual emergent results that are observed with agent-based models. A complicated agent structure makes analysis of such emergent structure much more arduous. And such a protocol is trivial to build into an agent. Merely specify the action to be engaged for each appropriate set of state space variable values. But this is not a *choice* protocol, since the outcome of the choice is predetermined by the rule set. It's a pre-defined action invocation, and since there is no probability associated with the rule-invocation, there can be no associated narrative event. Therefore this kind of choice mechanism is not within the purview of agent-based marketing models as defined here, which require such a stochastic mechanism.

**3.6.3:** At the opposite end of the rational protocol spectrum, in a manner of speaking, is another very simple form of rational choice called simply *random selection*. A set of choices (finite, countable, or uncountable) is available, and the agent tosses a coin, spins a wheel, or calls the `random` function on his personal computer to come up with the selection. Many might argue that this really is not rational choice. However, there are many choice protocols which engage the random selection process at some point, so it is important to remark on its existence. Elimination-by-aspect, described below in Section **3.7.5.3**, is one such protocol.

**3.6.4:** The classic statistical decision problem is perhaps the oldest, and most 'rational' of this class of choice protocols. The fundamental problem of statistical decision theory is to select a possible action from a set of actions that minimizes expected loss. The loss function $L(a,\theta)$ associates a real-valued number (the loss) with an action *a* in some set of possible actions *A* and a state-space value $\theta \in \Theta$ (referred to as the *state of nature* in the statistical literature). The triple $(\Theta, A, L)$ represents the statistical decision problem. Generally, the choice at hand is which value of $\theta$ represents the "true" state of nature. Nominally, there exists empirical data represented by the random variable *X* (which could be a multidimensional entity), the probability distribution of which, $P_\theta(x)$ depends on this true state of nature. A decision rule *d* for a given value of the random variable *X*, say *x*, maps the value of one of the actions in *A* to *x*; that is $d(x) \in A$. The loss is therefore the random quantity $L(\theta, d(x))$, and the expected value of $L(\theta, d(x))$, when $\theta$ is the actual state of nature, is called the risk function

$$R(\theta,d) = \mathsf{E}\left[L(\theta,d(X))\right] = \int_{x \in X} L(\theta,d(X))dP_\theta(x). \tag{3.1}$$

The choice problem is to select the decision rule *d* from the set of all possible decision rules *D* that minimizes *R*. If it is assumed that each $d \in D$ is such that, for each $\theta \in \Theta$, the distribution function $F_X(x|\theta)$ is continuous on a set of probability one, then the above is the Lebesgue integral

$$R(\theta, d) = \int_{x \in X} L(\theta, d(X)) dF_X(x \mid \theta). \tag{3.2}$$

Ferguson (1967) delineates a conceptualization of much of the field of statistics based on this definition, coupling it in with a game-theoretic construction. In fact, the triple $(\Theta, A, L)$ is a formal game in this sense.[38] It is easy to see that this is a quite well-defined problem, but implementing that definition in a specific context can be a considerable endeavor. However, creating a computer routine to implement a statistical decision rule in an agent does not seem to be conceptually prohibitive, although it might require significant time and resources to build and execute.

**3.6.5:** In a general sense, all "rational" choice protocols are described by the statistical decision process defined above. Indeed, some would consider this formulation the axiomatic definition of a rational decision. This general formulation says nothing about the nature of the decision rules $d \in D$. They could be hugely complex or trivially simple. Moreover, the set of actions *A* can be finite or infinite. Much more familiar to the marketing community is the choice process described in the random utility discrete choice case. The general formulation of this family of protocols is as follows. There exists a finite set *J* of possible choices with *#(J)* being the number of elements in the set *J*. Assume that the choices and the choosers are characterized by a vector of variables $\mathbf{z}_{ij}$ for decision-maker *i* and choice *j*. Each decision-making agent has an associated real-valued function $U_i(\mathbf{z}_{ij}): J \to \mathcal{R}$ that assigns a *utility* to each choice. The alternative with the highest value of $U_i$ is defined as the choice made, that is, the value $j^*$ for which

$$U_i(\mathbf{z}_{ij^*}) = \max_{j \in J} U_i(\mathbf{z}_{ij}) \tag{3.3}$$

This utility function is assumed to have an observable part $V_i(\mathbf{z}_{ij})$ and a stochastic component $\varepsilon_i(\mathbf{z}_{ij})$, and is therefore written as:

[38] Ferguson goes on to justify the application of Bayes' theory to statistics with his treatise, arguing that this game-theoretic perspective created a persuasive demonstration of the superiority of Bayesian statistical analysis. The text was written amid the roiling Bayes vs. frequentist debate within the statistics community prominent in the latter half of the 20th century, and which chugs along today with a steady, if tedious, background din.

$$U_i(\mathbf{z}_{ij}) = V_i(\mathbf{z}_{ij}) + \varepsilon_i(\mathbf{z}_{ij}). \tag{3.4}$$

Very often the $\varepsilon_i(\mathbf{z}_{ij})$ terms are assumed to have an Extreme Value Type 1 (EV1: see Johnson *et al.*, (1995) for a complete discussion) distributed with common location parameter $\gamma$ (which, without loss of generality, can be set to zero) and common scale parameter $\mu$. Then it can be shown (e.g. Ben-Akiva and Lerman, 1985, p 106) that

$$P_i(j^*) = \frac{e^{\mu V_i(\mathbf{z}_{ij^*})}}{\sum_{j \in J} e^{\mu V_i(\mathbf{z}_{ij})}}, \tag{3.5}$$

**3.6.6:** The general discrete choice problem is based on the assumption that there exists a set of alternatives $J$ with finite number of element *#(J).* Furthermore the agent making the choice has determined a complete preference order $\prec$ over the elements of $J$. A complete ordering on a set is a relation having the following conditions: [the idea is easier to understand by reading $a \prec b$ as "*a* is less than or equal to *b*."]: 1) antisymmetry: if $a,b \in J$ and $a \prec b$, and $b \prec a$, then $a = b$; 2) reflexivity: for every $a \in A$ it is true that $a \prec a$; 3) transitivity: If $a \prec b$ and $b \prec c$ then $a \prec c$, and 4) completeness: for *every* pair $a,b \in A$ it is true that either $a \prec b$ or $b \prec a$. If the preference order fails to meet these conditions, then the utility function does not necessarily exist, and the discrete choice problem cannot be formulated in a utility maximization context. However, human beings aren't bound by the definitions of preference orders. Non-transitive, circular orderings are common. For example, when Mary is asked to choose between chocolate and vanilla ice cream, she selects chocolate. When asked her preference between vanilla and strawberry, she chooses vanilla, and when asked her choice between strawberry and chocolate, she prefers strawberry. The ordering is not transitive. This situation occurs because humans generally determine orderings pair-wise over some (possibly quite short) period of time, and the circular inconsistency is quite easy to manage if the time of the preference comparison can vary. Moreover, there's no reason why the completeness property needs to be met in real-life situations. (An ordering that does not meet the transitivity and complete conditions is termed a *partial ordering*.) Fortunately, it is possible to derive a set of completely ordered sets from any partially ordered set (by considering each completely ordered set as a separate entity, and ignoring singleton sets), so the utility maximization problem reduces to a bookkeeping issue, (assuming there is sufficient data to estimate the number of models that might arise). In some cases, the collection of completely ordered sets can be represented in a hierarchy. But the point is that agent designers do not need to insist on complete choice sets. In fact, any kind of pairing relationship can be used as a preference ordering, and each can have a unique (empirically derived) utility function.[39]

**3.6.7:** Of course the domain of rational choice models is not exhausted by the utility maximization, discrete choice structure. Indeed, most choice situations are not even discrete, often requiring the selection of a parameter vector from a multiple-dimensioned real-valued vector space. Other methods are called upon here. Bayesian statistics have faded in and out of fashion over the past two centuries. See Zellner, 1971/1996, for the economics perspective, and Rossi *et al.* (2005) for applications in marketing. Of special note are recent simulation approaches to statistical parameter estimation, confidence interval determinations, and hypothesis testing. Among these are some of the more powerful computing methods, such as bootstrap (Efron and Tibshirani, 1993) and Monte Carlo Markov Chain (MCMC) approaches (Barndorff-Nielsen *et al.*, 2001). In this context, the meaning of the word simulation is somewhat different from when it is applied to an agent-based model. What is referred to in this case is actually artificial sampling, where data values are generated with a computer from a known probability distribution, so that complex and otherwise intractable parameter estimates can be determined without expensive, perhaps even impossible, data collection. And last, but certainly not least, there is the thoroughly delightful assortment of linear, dynamic, stochastic and other programming and optimization methods used by the operations research community for a range of commercial and industrial choice optimization problems. However, when they enter the realm of choice based on stochastic characteristics, they lose much of their elegance, and applicability.

[39] A decision support computer application named *Dewaffler* was available corporate decision support market for a while. It allowed the user to determine pair-wise preference orderings on a set of alternatives according to several criteria. Because it had no mechanism of determining the completeness of the resulting overall order, it would crash the computer if the user did not carefully, and manually, manage the order determination process.

**3.6.8:** The significant power for efficient allocation of resources in aid of narrative fulfillment represented by these rational protocols gives them superior positions in the pantheon of choice processes. It is this superior performance, historically incontrovertible, that suggests an ever-widening venue of application. But most humans cannot enlist the aid of these methods without extensive training and the assistance of a variety of other parties. Even as powerful as they are, they are bounded by the time and other resources required for their utilization in the face of the urgency and importance of the particular choice problem at hand. That is, they are examples of *bounded* rationality, in spite of what they may seem. Bounded rationality is almost always portrayed in contrast to the messianic alternative of the demonic methods noted by Gigerenzer. This "Laplacian Daemon" is the all-seeing, all-knowing supreme intelligence that can solve any resource allocation problem and select the globally best option for all individuals for all time.

**3.6.9:** Usually, the phrase "bounded rationality" refers to the inherent limits to computation and analysis required to support rational decision making. But in the background there are the basic limitations created by the cognitive and logical requirements necessary to comprehend and exist in the world. Limitations to reasoning are well known. Russell and Whitehead (1910-13) examine the foundations of mathematics, forming the basis for key work on the bounds of mathematical formulation and reasoning by Gödel (1933) and Freankel (1953). This sets definite limits (albeit quite broad ones) on the extent of application of logic to real phenomena. Philosophies of science (and epistemology in general) lend texture to the conversation (see Chalmers, 1999, for an introduction, and Kuhn, 1996 or van Fraassen, 1989 for more explorations). More pointedly to choice-making, Belnap and Steel (1976) examine the logic of information organization and query formulation. Pollock (1990) and Halpern (2003) have explored the underpinnings of reasoning about uncertainty, a central issue in the study of rationality under the narrative framework. Of course Popper (1959/2002) gives extensive treatment of these questions in his studies of the scientific method.

**3.6.10:** All of the rational choice mechanisms mentioned above are available for the specification of agent choice protocols. All have computer programs that fully specify how they should be executed, what the data should look like, how the results should be presented and the limitations on and conditions of their application. Moreover, it is clear that a number of organizations and institutions make extensive use of these methods. Companies routinely use operations research for a variety of optimal resource allocation tasks. Perhaps one of the more interesting and successful applications of operations research is the revenue management process used in the sale of tickets in the airline industry, now being extended to similar perishable goods such as hotel rooms, rental cars, and theater seats. See Talluri and Van Ryzen (2004) for a thorough discussion of revenue management. In fact, revenue management as an aggressive price setting structure is an important consideration of the airline agent's state vector in AirVM, discussed in Chapter 6.

**3.6.11:** But the individual human being, acting as consumer or seller, does not routinely engage such mechanisms in making choices. In fact, as noted above, they are very likely to be reductive and not structural models, and therefore may describe no actual process found in the real world. There is no evidence that human beings actually make routine choices using any of these tools. Humans tend to employ much less complex approaches to day-to-day choices, and in many instances extend these simple protocols to serious, far reaching and life-changing circumstances where the more sophisticated, rational methods would seem to be called for. Given that an agent model of the human consumer must describe what the modeled human actually does, and not what it could or should do, these less rigorous and more *ad hoc* choice protocols must also be made available to the human agent modeler.

## 3.7: A Sampling of Heuristic Choice Protocols

**3.7.1:** The models of rational human behavior implicit in the realization in economic theory briefly described so far – *homo economicus* – are not the only kind of human choice that is possible. In fact, there is scant evidence that people behave in anything like the optimizing behavior suggested by these protocols. Virtually every scholar who examines the problem derides the idyllic nature of rational, economic humans. Indeed, the economic man model is often more normative than descriptive. While quite useful, as demonstrated by its singular success, applying a normative model ultimately begs the question of how people 'really' make choices, and to that extent the economic theory that is the foundation of discrete choice and utility maximization methods is defective. And it is not necessary with the development of agent-based simulation models. The alternative to economic man is often couched somewhat inappropriately in terms of bounded rationality.

**3.7.2:** Beyond the philosophical context noted in **3.6.9**, bounded rationality refers to the limitations in resources available to undertake and perform data collection and analysis leading up to a decision, in other words, the execution of benefit/cost analyses, formal or otherwise. In this regard, two streams of thought are discernable in the analysis of human rational behavior. The one, stemming from the game theory work of von Neumann and Morgenstern (2004/1944), explores human decision making as a *real-valued trade-off* endeavor. Indeed, classic economic theory assumes that all the entities in a given economy converge to such utility function rationality as equilibrium is reached. Among the many tributaries of this line of thought is the utility structure that leads to discrete choice theory and the modern study of consumer choice behavior discussed previously. The second stream originates from the work of Simon (1957, 1996, 1997), who suggested the notion of *satisficing* as a decision structure. Satisficing is choice-making based on being "good enough" rather than "utility maximizing." This idea fits rather neatly into the narrative framework.

**3.7.3:** Gigerenzer (2001) captures this idea with a simple taxonomy of rational choice. Shown in Figure 3.6, it describes what he refers to as "visions of rationality." He breaks rationality down into two broad classes. One he calls *demons*, referring to the demonic capabilities he views as necessary to carry out rational decision-making in the real world without regard for constraints of time or resources, as mentioned in the previous section. Gigerenzer puts it thus:

> "For example, models that seek to maximize expected utility or perform Bayesian calculations often must assume demonic strength to tackle real world problems. Real decision makers … need, first of all, to search for information. This search cannot go on indefinitely; it is constrained by limited time, money, attention or other finite resources." (Gigerenzer, 2001, p. 38)

Demonic reasoning is dissected into *unbounded rationality* and *optimization under constraints*. The former is literally applying limitless resources to a decision problem, or being presented with a decision problem so simple (such as a statistical estimation problem) that all relevant issues are easily known. The latter refers to concepts such as those frequently seen in operations research, in which the problem at hand has been constrained to become manageable. In this case, however, the choice of the nature and values of the constraints are subject to the same resource limitations as any decision problem, thus only begging the issue of what level of demonic strength is available. It is for that reason that I classify the analytically rational methodologies described in Section 3.6 above as bounded rationality.

**3.7.3:** Gigerenzer suggests that the bounded rationality side consists of two components; the search for options or alternatives, encompassed under the label of "satisficing," and the actual choice among alternatives, referred to with the term "fast and frugal" heuristics. The searching activity includes methods for finding options and methods for stopping the search. Some of the search methods Gigerenzer notes include: *random search*, where the agent explores the decision environment without any apparent organization until time runs out; *ordered search*, using the validity of environmental cues as they apply to the choice problem at hand as the ordering mechanism; search by *imitation* using apparent similarities of this decision problem to those encountered in the past (imitation allows us to know where to look and what to look for, but does limit results if the environment within which we are searching is

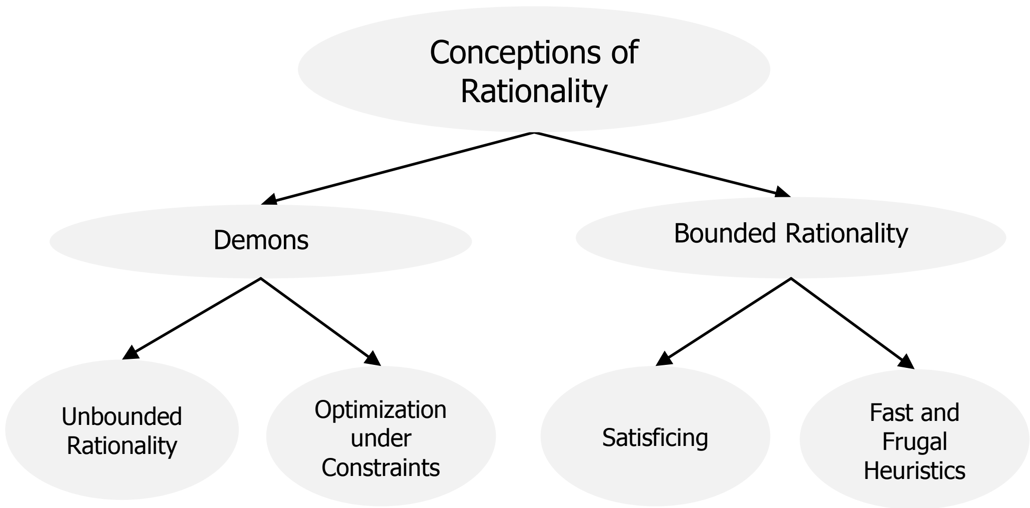



**Figure 3.6: Gigerenzer's Taxonomy of Rationality**

novel or unexpected); and *emotions*, which apparently act to narrow down the search space in effective, but not well understood, ways. Other search methods readily come to mind, but all of them can be interpreted as being enabled by the narrative context in which the choice event is presented. That is, the search process is governed by what the agent, because of the controlling narrative that is creating the decision context, considers important to the resource allocation and outcome probabilities associated with the atomic narrative.

**3.7.4:** Stopping the search is where satisficing comes in – when have I searched long enough and established enough options? When I am satisfied that further research will not add any important alternatives, or when I have no more time to gain additional knowledge? Gigerenzer (2000, pp. 129-165) proposes what he calls the *probabilistic mental model (PMM)* as a construct to account for the satisficing and fast and frugal heuristic choice protocols. In a PMM, the individual puts the choice event in hand into a mental construct of similar choice situations it has encountered in the past, or has learned by one method or another, and uses that context as the satisficing criterion. In other words, people fit decision problems into models that seem somehow appropriate to the problem, make the decision, then modify the model if expected results are not realized. This approaches argues that the limited cognitive and computational ability of humans mitigate against a purely analytic benefit/cost structure in favor of agile and adaptive, if less than optimal, heuristic decision rules.

**3.7.5:** Heuristic protocols are choice mechanisms that rely on relatively little information and rule-of-thumb thinking. There is strong evidence that much of the choice behavior of humans is of this variety, if for no other reason that bounds on available time for decision-making prohibit any other approach. For example, Malcolm Gladwell's discussion of virtually instantaneous human decision making in his book *Blink* (Gladwell, 2005) addresses this phenomenon. Todd (2001) offers a simple listing of some of the more important fast and frugal heuristics:

> **3.7.5.1:** When choosing between two alternatives, one recognized and the other not, the *recognition heuristic* says choose the recognized one. The basis for this heuristic seems to be that recognized alternatives are apt to be more successful, and therefore more likely to be recognized, and thus choosing them is a better idea. Note that increased option search can reduce choice efficiency if more recognized options are added (Todd, 2000, p 57).
>
> **3.7.5.2:** In the *take the best* heuristic the agent selects the best alternative as measured by one single criterion (e.g. price). Other dimensions which characterize the issue in question are not considered at all. One can see how this heuristic fits neatly into the narrative framework if the criterion reflects a resource deemed supremely important for the realization of the narrative, as it

becomes the dominant factor in the decision. And different agents may have different criteria for what constitutes "best."

**3.7.5.3***:* An extended form of the take-the-best heuristic (which in fact can be shown to actually be rational) is *lexicographic ordering*. This is a multi-dimensional extension of take-the-best. This line of thought has been explored more fully by Tversky (1972) with his *elimination by aspects* approach. Elimination by aspects is a choice method wherein the individual has a set of criteria in mind on which he will evaluate a set of alternative choices. He ranks the criteria from most to least important, and then proceeds to evaluate each alternative against the first criterion. If two more alternatives have equal values according to that criterion, he eliminates all the others from consideration and moves on to evaluate the remaining options using the second most important criterion. He proceeds this way, moving down the prioritized list of criteria, until only one alternative remains. If he gets through the last criterion evaluation and he has more than one choice alternative left, he selects among the remaining alternatives at random; that is, engages a random protocol. The phrase 'lexicographic' is also used to describe this protocol, since alphabetic ordering is done this way. Tversky (1972) also showed the equivalence of elimination by aspects to discrete choice thus moving this seeming heuristic into the domain of the rational.

**3.7.5.4:** Another fast and frugal heuristic approach that lies on the boundary between pure rule of thumb and the rational choice operations is *Dawes's rule* (Dawes, 1979). This is a type of linear result choice method. Evaluate the alternatives against a set of criteria by determining if the alternative is either positive or negative with respect to each criterion, and subtract the number of negatives from the number of positives. The option with the highest score is chosen.

**3.7.5.5:** Other heuristics are described by scholars in several fields, especially cognitive psychology. Kunda (1999) offers an extensive array. She mentions the *representative heuristic*, wherein a choice is based on the similarity of the choice situation a category of choices that have been faced or witnessed before (Kunda, 1999, pp. 57–77). The determination of similarity is based on characteristics of the situation at hand to the one or more of the attributes that define a class of situations, but may differ from members of the class in details. This is conceptually coherent if viewed from a narrative perspective, in the sense that the value of resources and the weight put on the factors which assess that value that play in a narrative may be the among the criteria that define the class similarity. In this sense it is somewhat like the recognition heuristic described above.

**3.7.5.6:** Another family of heuristic methods cited by Kunda (1999, pp. 77 – 89) is the *statistical heuristics*, referring to statistical rules-of-thumb most people seem to have learned and carry around with them. They generally seem to arise as a result of dealing with the pervasive uncertainty life brings. For example, Kunda cites the observation that while having all nine of one's grandchildren be of the same gender would seem quite unusual to most people, having all three of your grandchildren would not seem that odd. But why? The suggestion is that people apply an elementary bit of statistics to the problem, reasoning that the gender of a child is a fifty-fifty proposition, and, equating that to the tosses of a coin, where tossing nine heads in a row happens much less often then than tossing three in a row.

**3.7.5.7:** One final contribution to the heuristic arsenal is put forth by Kunda (1999, pp. 102 – 109), rather more subtle than the others. This is *anchoring and adjustment*. This is the tendency for people to base a decision about a specific issue based on a reference to other (perhaps completely irrelevant) situations. That is, some change in an element of the context in which the choice operation is taking place may cause a choice to vary from one instance to another, even though the context change is not part of the choice event. Tversky and Kahneman (1974) and Strack and Mussweiler (1997) have studied this tendency, and found that setting of an anchor (the changing context element) will cause individuals to tend to adjust their choices to be consistent with the anchor, even though the anchor does not bear on the choice event itself.

**3.7.6:** There are many additional heuristic processes that could be identified, and this would seem to be a fertile area for further research. There is a considerable psychological and sociological literature that could be explored by students of marketing to extract current understanding of the choice mechanisms and formulate computing structures that would be applicable to agent-based models. There is little doubt these mechanisms are used very frequently in many day-to-day choice situations, including those found in the marketplace, and they should be available to the human agent modeler as much as the more flamboyant rational methods are. But in their implementation their sometime severe bias must also be recognized. That is just as much a part of the protocol as the actual choice itself. This issue of choice bias is a topic that will be explored in Section 3.9.

## 3.8: A Sampling of Social Network Choice Protocols

**3.8.1:** Humans rarely make decisions completely alone. Many choices are subject to consideration and examination by not only the chooser directly, but also by other individuals who are connected to her in some way. Friends, relatives, other respected (or not so respected) experts, celebrities, people in authority, co-workers and many others enter into the choice making process in a host of ways, and with a variety of consequences. I consider only a few such mechanisms here. Like the discussions above, there is inadequate space in this treatise to do more than sample the landscape.

**3.8.2:** In a sense, social network choice protocols are somewhere between the rational approaches and the individual heuristics. Gigerenzer poses the dilemma:

> "In many real world situations, there are multiple pieces of information, which are not independent, but redundant. Here Bayes' rule and other 'rational' algorithms quickly become mathematically intractable, at least for ordinary human minds.[40] These situations make neither of these two views [laws of probability, and reasoning error] look promising. If one was to apply the classical view to complex, real-world environments, this would suggest that the mind is a supercalculator – carrying around the collected works of Kolmogoroff, Fisher and Neyman – and simply needs a memory jog …. On the other hand, the heuristics and biases view of human irrationality would lead us to believe that humans are hopelessly lost in the face of real-world complexity, given their supposed inability to reason according to the canon of classical rationality, even in simple laboratory experiments." (Gigerenzer, 2000, p. 167)

That most people survive without falling into Gigerenzer's abyss is due in part to social network choice methods. Clark (1997) makes a compelling argument in support of this vital role, suggesting that the "scaffolding" of the social network in which all humans are embedded is central to our ability to make decisions, survive and advance. Kunda

[40] It is easy to create a simple example where Bayesian analysis generates formulations which are not only mentally intractable, but mathematically intractable as well. Consider virtually any non-trivial situation where there is no conjugate prior.

(1999) provides a broad but insightful survey of the field, and Sternberg and Ben-Zeev (2001) offer an excellent introduction. The rapid rise of social networking sites on the internet – Facebook, MySpace, Twitter – testify to both the importance of the social network and the ease with which people adapt to new forms of it.

**3.8.3:** The development of formal methods of social network analysis has become quite active, as well, partly because of the advances in computing and agent-based modeling. Network analysis as a formal field of academic endeavor dates back at least to Erdos and Renyi (1960/2006) but the emergence of the worldwide web has spurred more recent advances, including the exploration of scale free and stochastic network analysis. An easily accessible introductory survey of modern methods can be found in Barabasi (2002) or Buchanan (2002). A more advanced and formal treatment is offered by Dorogovtsev and Mendes (2003). Newman *et al.* (2006) have compiled a compendium of more recent developments in the field. Recent advances in social network analysis can be reviewed in the collection edited by Carrington *et al.* (2005). Exploration of one formal approach to networks as decision support structures (as argued by Clark, 1997) is Jensen (2001), who proposes an extension to the theory of decision trees with his development of Bayesian networks. Closer to marketing, Rossi *et al.* (2005) discuss the application of hierarchical Bayes methods to marketing research.

**3.8.4:** In an agent-based modeling context, networks are an expression of the topology of the explicit space required by Epstein's definition of agent model (Epstein, 2006). That is, a network defines which agent is "close to" which other agent. That is, it defines what the word "close" means in a particular model. Epstein and Axtell illustrate the network role on the *Sugerscape* grid with respect to economic and social interaction (Epstein and Axtell, 1996, pp. 130-135). As they show, a network in an agent-based model is a communication connection between one agent and another. A single agent can have such connections with a number of other agents. The connection can be one-way or two-way. Different kinds of connections can reflect differences in the nature of the inter-agent communication. Networks change over time, with new connections being made and older ones dying out.

**3.8.5:** A convenient way to consider the network structure of an agent-based model is to stipulate that each agent will maintain a list of the other agents to whom it is connected. Separate lists can be kept for different kinds of communications. If two agents have each other in their individual lists, then the network communication link is mutual, otherwise it is just one way. The *message posting* function of the computer implementation of an agent model (see **4.4.6** for more details) can then be engaged to manage the communications between agents during the simulation.

**3.8.6:** However, network structures are not a requirement of an agent model. Explicit space can be portrayed in other ways. In a cellular automaton, as is used in the word-of-mouth simulation discussed in Appendix I, agents reside on a grid where communications between agents is based on being physically next to each other on the grid. In the AirVM virtual market, simple agent communication networks are used to define the relationships between distribution system agents and airline agents.

**3.8.7:** Perhaps the most common form of social network-dictated choice protocol is *imitation*. There is strong evidence human beings learn by imitation, and thus it is reasonable that the same approach would be called on when faced with a new choice situation. What action did others do in this same situation? Very often the narrative event which creates the choice is encountered in the context of a shared narrative, and thus the choice by the individual is apt to follow the course of the underlying narrative supporting the event definition. In terms of agent-modeling, the agent which uses this protocol must be linked through a social network to the individual or group of individuals whom it wants to imitate. The imitation cannot be certain, however, for that is reserved for consilvocation, discussed below. There must be a randomizing mechanism that allows for the imitation to have a stochastic element, such as being linked to two or more individuals who can be imitated, but with a randomization device that dictates which one is followed in a particular event.

**3.8.8:** Closely related to imitation is *expert advice*. It is natural that someone believed to know more about a particular narrative event – an expert in the field – would make a wiser choice than a novice. And as humans learn as children and adolescents, the courses of action suggested by experts with more experience are important techniques in determining the probability of the outcomes of various choice options. From an agent modeling perspective, clearly a social network link is needed between the agent and the expert. Again, some stochastic mechanism needs to be present if the requirements of the narrative construct are going to be met. In this case, however, in addition to the selection of one of a possible set of experts, whether or not to follow the expert advice can be employed as the randomization method. Gladwell (2000) in his popular book *The Tipping Point* describes the maven-connector network which is often found in expert consultation choice. Taleb examines what he refers to as the "expert problem" in some depth (Taleb, 2007, pp. 145 – 156), classifying experts into those who have expertise in subjects for which expertise exists, such as science and medicine, and reserves the phrase "empty suits" for those who claim expertise in things for which no expertise can possibly exist, such as the forecasting the value of a stock exchange index tomorrow morning.

**3.8.9:** This raises an issue to be addressed by an agent model that uses either imitation or reference to experts: How are those experts going to make a decision which can be imitated or on which expert opinion can be founded? One might suggest that the "imitand" or the expert use a rational choice method, for example. Or there could be hierarchies of imitators or experts, each imitating others while providing expertise to other agents. It would be quite interesting to explore how such networks might work with a simple agent model. In particular, emergent properties of agent-based simulations which contain such social mechanisms could be most curious. Finally, note that the required stochastic property of the choice process of an agent using imitation or experts could be inherited from the stochastic property of the imitated or expert agents.

**3.8.10:** A third kind of social network choice protocol is *voting*. An individual can make a choice by polling a set of other individuals to see what they would choose, and then determine his or her choice by tallying the results. For a simple binary choice, this technique is trivial. (Again, how those who cast votes determined their respective choices is a modeling design issue.) For choice events where three or more options and three or more individuals are polled, however, Kennith Arrow's Impossibility Theorem enters the picture, and some form of bias has to be introduced to guarantee an outcome (Arrow, 1963). Again, implementation of this type of social network decision protocol is straightforward in the design of an agent-based model. The agent in question maintains a list of voters, and polls each one by supplying the voter with the choice problem and accepting that voter's choice as the vote. The agent then tallies the votes, and determines its choice.

**3.8.11:** A particularly strong form of social choice is *consilvocation*.[41] In this situation, an agent turns over to another agent complete control over the choice to be made. A

[41] This term was coined by a colleague and business partner, Nick Lanyon, who graduated from Oxford and has a sharp interest in the English language. He was assisted – although he didn't know it – by a friend of mine, Rebecca West, whose command of English is also admirable, and American. Rebecca refined Nick's original construction, which was virtually unpronounceable.

trivial example is the husband leaving to the wife the choice of restaurant for dinner. Consilvocation happens all the time in a democratic political context. The individual citizen elects a representative to sit in a legislative body and make decisions on his behalf. The citizen has thus turned over the choice function to the representative. Consilvocation is a way of eliminating choice events that are out of an individual's control but will have an impact on the course of a compound narrative. It simplifies matters considerably. Another attribute of the consilvocation choice protocol is that the consilvocated right to make the decision can be revoked. The process by which revocation occurs can be simple (the husband elects to pick the restaurant himself) or complex (the citizen must wait for the next election or invoke a recall process). Adding consilvocation to an agent model is not difficult. In the AirVM virtual market described in later in this dissertation, a single agent actually represent multiple passengers, and one agent buys tickets for everyone in the group it represents. If the model had to represent the individuals in the group for reasons other than ticket purchase (say ideal departure or arrival time determination), then the members of the group would consilvocate the itinerary choice to that agent, but engage other decision protocols for other choice events.

**3.8.12:** As noted above, only a few social network decision structures have been described. This area of virtual market development can benefit from significant additional exploration, research and discussion. Indeed, many social process agent models explore some of these processes in depth. However, an examination of the results and trends in this important area is beyond the scope of this presentation.

## 3.9: A Sampling of Bias in Choice

**3.9.1:** The term *bias* in the context of this analysis refers to the difference between the choice that is actually made and the choice that would be made if the "correct" alternative were selected. Obviously, this calls for a definition of "correct." In a formal statistical decision problem, bias refers to the difference between the expectation of a particular statistic that is used to estimate the parameter and the value of the estimate itself. That is, a statistic $S$ used to estimate a parameter $\theta$ is *unbiased* if $\mathrm{E}[S] = \theta$, and the search for unbiased estimators is a long-standing topic of statistical research. The definition of correct is not so easily determined in the agent choice circumstance, however.

**3.9.2:** Many authors identify and describe choice bias in a manner similar to the statistical definition, holding that the choice that results from the engagement of a rational decision protocol is the correct one, and other heuristic or social network protocols that lead to different choices for the same narrative event are biased. I do not take that view. As has been said before, if human beings operating in the context of marketing activity are to be validly represented by agent-based models, then they must be represented as they are, not how someone thinks they ought to be.[42]

**3.9.3:** Some biases are perceptual in nature, and arise from inaccurate representations of reality, which in terms of the formal definition of agent given in Section **2** can be accommodated with properties of the perceptor component. Festinger *et al.* (1964)[43] and Tumminia (2006) explore some of the implication of cognitive dissonance, which occurs when an individual believes in a reality that is directly contradicted by the sensory evidence before him. The *mistakes-were-made* assertion described by Shermer (2008, pp. 67-71) is an example of *self-justification* bias (Tarvis and Aronson, 2007). *Inattentional blindness* is the failure to recognize some feature of the surrounding environment because attention is focused on some other environmental feature (Simons and Chabris, 1999). *Blind spot bias* is the ability to see biased perception on the part of others, but fail to see it in oneself. It is similar to *better-than-average* bias, which causes a person to think they are more capable at any given skill or talent than the average individual (Pronin *et al.*, 2002, 2004). Humans also tend to see themselves in a more positive light than they see others (Kruger, 1996), and therefore create a *self-serving bias*. People tend to accept credit when they behave in socially acceptable ways, and blame circumstances outside themselves when they don't do so well. This is an example of an *attribution bias* (Nesbitt and Ross, 1980).

**3.9.4:** But not all biases are perceptual. Two, in particular, are based on misunderstanding fundamental concepts in probability theory. Kunda (1999, p. 54) notes that probability theory, as a formal mathematical discipline, dates only back three hundred years,[44] and that its relatively late development speaks to its intuitive difficulty.

[42] However, there are potentially some interesting agent-based models to be built that explore the effects of the biased choice protocol vs. the rational on the outcome of a narrative.
[43] The references in this paragraph are drawn from Shermer (2008), and are intended to direct the reader to the cited work in order to delve deeper into the concept described. I have not consulted these works in the research for this treatise.
[44] She's wrong, in one respect. Formal probability theory, where probability is defined in terms of normed measure spaces, dates back only 80 years or so to Kolmogorov (1931).

One is the *base rate bias*, described nicely by Kunda (1999, pp. 57-62), and the other is the so-called *Let's Make A Deal* fallacy, described by Shermer (2008, pp. 83-84) but first explored by Gillman (1992). Base rate bias is very common. It stems from misunderstanding the incidence of some particular characteristic in the underlying population, and thus from a miss-application of Bayes' Rule to an intuitive inference. For example, consider John, who is a small gentleman with a quiet demeanor, who wears glasses, dresses conservatively, and often is seen carrying a book. Is John a factory worker or a librarian? Many would say a librarian. But the likelihood that he is a factory worker is far higher than he is a librarian. None of the distinguishing criteria disqualifies him from being a factory worker, and there are far more factory workers than librarians. This is the base rate fallacy; failing to take into account the actual rate of incidence of a factor in making a judgment. It is a primary reason for bias in the representative heuristic.

**3.9.5:** The Let's-Make-A-Deal fallacy significantly more subtle. The name refers to a game show popular on American daytime television. (It is also called the Monte Hall fallacy after the individual who hosted the show for a number of years.) This game show situation is this. A contestant stands before three closed doors, behind one of which is a valuable prize, usually a car. Behind the other two are valueless prizes, historically goats (but animal rights advocates objected, so some other worthless offering is now used). Which door hides which prize is unknown to the contestant. The contestant chooses one of the doors and is awarded the prize behind that door, hopefully the car. But before the chosen door is opened, the host, Monty Hall, opens one of the other two doors (always revealing a goat), and asks the contestant if she want to change her door choice, and take the prize behind the remaining door. Should the contestant take Monty's offer, or should she stay with his original selection? Most people will say it doesn't matter. Assuming that the likelihood of a car being behind any of the three doors is the same (one third) for each, then knowing that it is not behind one of them only means that the probability of it being behind either of the remaining two is now changed to one half, and therefore switching does not affect the odds of winning. But that is incorrect. In fact, the probability that the car is behind the door that was chosen originally by the contestant is one third, and that it is behind the unrevealed other door is two thirds.[45] The explanation is clear (but for many, not convincing). The contestant faces three possibilities: the doors can hide 1) car, goat, goat; 2) goat, car, goat; or 3) goat, goat, car. Suppose she starts the game by selecting door number one. Then if she switches and the first possibility is the true situation, she loses. But if either of the other two possibilities are the case and she switches, she wins. Thus the probability of winning by switching is two thirds. In Appendix G is presented a simple application written in C# using Microsoft's Visual Studio 2008 that simulates the game any desired number of times. Execution of that simulation verifies the correctness of the analysis. The source code is also included in the appendix, as an example of a simple simulation program, so the reader can copy the code and test the conclusion directly.

**3.9.6:** There are, of course, many similar examples of incorrect reasoning. That they exist and should be avoided in the making of careful decisions is obvious. But it is equally obvious that these "errors" can be subtle, and difficult to detect. Once again, human agents in virtual markets must be modeled as they are, not as they should be. That means determining when bias is an important part of the agent behavior, and building that bias into the agent. In fact, the sampling of choice protocols sketched out in the last four Sections of this discussion emphasizes how complex and involved modeling human choice can be. But as difficult as the task might seem initially, it is ameliorated by the knowledge that choice protocols span cultures and societies, so what is learned on one context can be applied in others, and, with the connection of the choice protocol to the narrative framework, complex behaviors can be built up out of simpler, atomic elements.

## 3.10: Conclusion

**3.10.1:** In Chapters 2 the first two steps in the research program outlined in Chapter 1 were accomplished. Agent-based models were determined to be valid representations of real things, and human marketing behavior is best understood from the perspective of evolutionary science, so agent-based models should be applicable to describing and exploring that behavior. The next step was to define a structural framework for that description. That framework is the subject of this chapter.

[45] This puzzle was first presented by a columnist named Marilyn vos Savant in a U. S. national Sunday newspaper supplement a few years ago. When she explained how the probability of winning went to 2/3 if the other door were selected, a firestorm of protest erupted from the academic statistics world. She is right, however.

**3.10.2:** I have completed setting the stage for the implementation of agent-based modeling in the context of virtual market simulations as a tool for marketing science. Given the answers supported by the evidence for the first two questions, then I may ask "Is it possible to create a construct which can be used to unify and explain these common traits, and form a coherent context in which the definition of human marketing agent can be offered?" That is, in the definition of the computing objects and properties that are agent-based models, is there a unified formulation that will represent human marketing behavior in a consistent and parsimonious manner?

**3.10.3:** I offer as a candidate for this responsibility the narrative framework. I assert that individuals maintain an internal, mental understanding of the world and their place in it by means of time-framed cause-and-effect sequences called narratives. Simple, atomic narratives represent a single event in time, defined as a change in state, and molecular and compound narratives are made up of strings or other combinations of these atomic narratives. The narrative events are different from events in real time, because the narrative owner can imagine alternative outcomes of the event, and assign probabilities to those outcomes. Moreover, narratives can be fictitious and imaginary and thus need not represent real cause-and-effect relationships. The owner of a narrative has a sense of the favorability or unfavorability of the outcome of any event in the narrative, so the concept of value is inherent in the narrative construct. Finally, individuals have at their disposal resources that they can allocate to change the likelihood of given event outcomes, and they intervene into reality when appropriate to bend event outcomes in favorable directions. Thus the key role of choice in the narrative construct is substantiated.

**3.10.4:** The most important implication of the narrative construct for agent-based modeling is: *In order to represent human behavior in virtual markets, it is necessary to define the narrative event context of the agent and to give the agent appropriate choice behavior with respect to the products or services being offered.* To that end, this chapter closed with a sampling of choice protocols, including rational, heuristic, social and bias. All of these are readily described by relatively simple agent programming objects. With that, I am now ready to define the agent-based computer simulation I call a virtual market.

# Chapter 4:

# Virtual Markets

## 4.1: A Working Definition of Agent for Use in Virtual Markets

**4.1.1:** The objective in this chapter is to set out the design for agents to be used in virtual market simulations that will efficaciously represent the human actors in actual marketing contexts. This general definition of agent has been presented in Parker and Bakken (2007) and in Parker and Perroud (2008). At this point, such a definition must be quite general, but will take on greater depth as specific agents are designed and implemented. I will also assume that the agents to be used in virtual markets always will be implemented in the form of computer programs, so an occasional reference to appropriate computing structures may be needed. The environment for one agent usually contains other agents, either like itself or with different capabilities. Within the agent are mechanisms for perceiving the environment, making choices about what to do given the information from the environment in the context of and the agent's motivating narrative, and taking actions that advance the agent's narrative. This simple agent structure consists of four components, each of which will be briefly examined separately. This structure is illustrated in Figure 4.1.

**4.1.2:** Corresponding to the memory for a human being, the *state vector* is an array of variables which describe the current state of the agent in the simulation. These can be very simple variables, the existence or absence of some condition, or quite complex, such as an array of parameters defining the probability distribution of some choice protocol, or the history of the agent up to this point in the simulation, or its expectations for its future. Referring to this array as a vector does not imply that it is necessarily a precise set of real number laid out in a row or column manner. However, it does imply the memory object is both precisely well-defined and finite. The state vector also encodes the appropriate narrative structures that represent the beliefs and aspirations of the agent within the virtual market being considered. For example, it contains the probability distributions of the outcome sets relevant to the events the agent will encounter during the simulation. The state vector is maintained as appropriately defined data structure within computer memory.

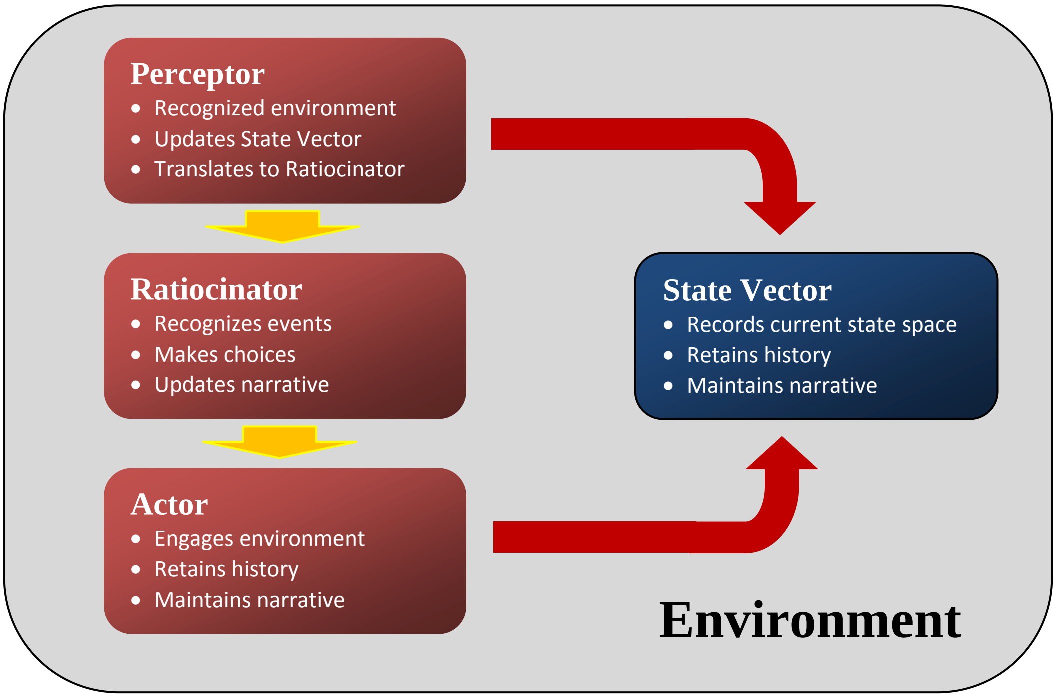



**Figure 4.1: The General Structure of an Agent**

**4.1.3:** Like the ability of humans to receive, filter and understand information, the agent has code that allows it to receive information about the current conditions in its environment through the *perceptor*. Almost without exception, perceptors are message-handling routines within the agent computer code. Messaging in this context will be discussed in greater detail in **4.4.6**. The purpose of the perceptor is to 'observe' the current state of the agent's environment, and filter and translate that information into a form that can be used by the choice-making component of the agent, converting the messages from the environment into an internal form of use to the agent. This internal translation can be unique to each agent, which thus allows for the design of agents that interpret the same external message differently. This would be important, for instance, if different agents had different narrative events that were triggered by the same external environmental conditions.

**4.1.4:** Opposite of receiving information about the environment through the perceptor is taking actions on the environment, such as implementing a choice. Such actions are managed by the *actor* component of the agent. In a virtual market, the activity carried out by actor is very often the notification to the environment (more particularly, specific agents in it) of a choice that has been made by the agent. This is done by means of a messaging structure, quite similar to the messaging operations handled by the perceptor.

**4.1.5:** The actions of the agents to external messages, both those which are put into the environment as output messages by the agent actors and those required for changes in the internal state vector of the agent, are managed by the *ratiocinator* component. The ratiocinator fulfills the role of choice maker and adaptation mechanism. This is the part of the agent which replicates how the human intersects with the simulated world in accordance with the then-active narrative structure. In relatively simple virtual markets, such as those that are developed for this dissertation, the narratives of interest are atomic narratives, dealing with a single narrative event. In the word-of-mouth agent-based simulation presented in Appendix I, the atomic narrative is a choice to buy, wait or ignore, depending (stochastically) on the input about the product the agent receives from its immediate physical neighbors. In the AirVM airline passenger virtual market examined in detail in Chapters 5 through 7, the atomic narrative is the selection of a particular itinerary from those available at the time of purchasing a ticket for a flight. How the stochastic choice mechanisms are implemented are within the purview of the agent's ratiocinator.

**4.1.6:** Among the data stored in the agent's state vector are the relevant details of at least one atomic narrative that governs the acquisition of the product on which the agent model is focused. For even the simplest case of the atomic narrative with a single event, the agent components must contain significant data. The perceptor needs to be designed to recognize the occurrence of the event and the perceived state of the environment at the time of the occurrence. The ratiocinator must be programmed to perform one of the set of choice protocols the agent will apply to exercise the choice required by the event, including the availability and allocation conditions of the resources at the agent's disposal. And the actor must have, in its repertoire of possible actions, those that are suitable given the event perception and protocol requirements. If the narrative construct driving the event recognition and intervention is a molecular narrative, with a number of events connected in a time-ordered and contingency network, then each atomic narrative must be delineated as described above. Furthermore, the connections between the component events must also be precisely represented in order to reliably represent the agent's actions.

**4.1.7:** This definition of agent contains the basic outline of all the pieces of the programmer's art necessary to build and execute a virtual market simulation. At first glance, these requirements may seem onerous. But in the two examples to which the construct has been applied, the problem reduces itself to a tractable, if perhaps complex, computer algorithm programming task, as will be seen as the discussion progresses. As with all computing programs, two elements are present: the data on which the algorithms operate and the computing code which executes the algorithms themselves. Looking at the agent definition from that perspective, the state vector holds the data, and the perceptor, ratiocinator and actor consist of algorithms. From experience to date, and from reflection of how a particular agent behavior might be implemented in a variety of other contexts, the programming of the choice protocol set that resides in the ratiocinator seems the most daunting.

## 4.2: Virtual Markets Defined

**4.2.1:** The phrase virtual market is used throughout this dissertation, and the meaning is quite direct. *A virtual market is an agent-based simulation of the market for a defined set of products being offered for sale by a collection of sellers and available for purchase by a collection of customers.* The phrase is really only meaningful in that it distinguishes agent-based models of use in marketing science from other forms of modeling that are often applied to marketing analysis, such as hierarchical Bayes (Rossi *et al.,* 2005) or diffusion theory (Bass, 1969). It is the method of analysis that is distinctive, not the marketing entities that are being studied or the management objectives that are being addressed.

**4.2.2:** There must be at least two kinds of agents in a virtual market – *customers* and *sellers*. Both need to be agents as defined here, with perceptors, actors, ratiocinators and state vectors. Customer agents must have sufficiently detailed representations of the atomic narrative which is the focus of the simulation. That is, the decision context of the customer when considering alternative products to fulfill their narrative-based needs must be described. And the choice protocol for the customer agent is an essential element of the agent object used in the simulation program. Very often this can be extremely simple: the perceptor recognizes the availability of two or more products; the ratiocinator employs a simple weighted random choice mechanism to make the choice. In such a simple case, the state vector is simply temporary memory for the variables which describe the available product options, and minimal internal agent parameters, such as the weights to be applied. The actor's range of operations is generally restricted to the purchase of the chosen product. On the other hand, it is also possible to build agents which utilize quite complex molecular narratives, where one decision leads to a course of additional choices, while another takes the agent down a different path. In all cases, however, the ratiocinator's choice protocol must be stochastic, but it does not have to be rational in the sense of Chapter 3. It can be a heuristic or a social choice, and it can be biased. Recall that agents are models of how people actually behave, not how one might think they should behave.

**4.2.3:** For an agent-based model to be a virtual market there must also be agents that are sources of products for the customer agents to consider. These are termed, for obvious reasons, seller agents. These agents also have narratives which motivate their behavior, but they are distinguished from customer agents in that generally such narratives are merely ways to make a profit. In the simplest case the agent simply sets a price. But it is quite conceivable that much more complex seller interaction, such as competition using various economic game choice protocols, could be represented in a virtual market.

**4.2.4:** A virtual market can serve a number of roles depending on the frame of reference of the community of individuals or groups that build and use it. In the world of software applications development, these different frames of reference are sometimes referred to as *use cases*. Several use cases can be distinguished for a virtual market. Perhaps the most obvious role, and certainly the basis for the position that agent-based models are a relevant, indeed important, contribution to marketing science, is that of a *laboratory*. Clearly experiments can be done in a virtual market that would be impossible otherwise. Of equal importance, given the inherently stochastic nature of a virtual market (due at the very least to the assumed stochastic structure of the narrative), Monte Carlo methods can be employed to yield probability distributions for variables of interest that validly represent the probability structures found in the actual market place, without recourse to the Central Limit Theorem or other reductive techniques.

**4.2.5:** Closely akin to the agent-based model as laboratory is its role as *instrument*. Humphries (2004), as argued in Appendix A, asserts that computational science is an extension of humanity's ability to observe and understand the world. This function is an attribute of all of computational science, and is thus inherited by virtual markets. In other words, a virtual market can be expected to reveal aspects of the market place it represents that would otherwise not be visible or apparent. This role of a virtual market as an instrument is illustrated in the application of AirVM to the estimation of air travel demand between a selected pair of cities, which is significantly improved because of the instrument property of AirVM. Another manifestation of this instrument use case is *forecasting* the future course of a market. The sharper view of the dynamics of the virtual market creates the opportunity for forecast mechanisms of greater validity, and for specifying realistic confidence bounds on such forecasts. Furthermore, where such mechanisms cannot exist – such as in the stock exchange – the reason they cannot exist may well be more convincingly demonstrated through the virtual market instrument.[46]

**4.2.5:** Considering the broader marketing community beyond just the marketing science world, virtual markets present a third use case. Much of the motivation for the development of a virtual market comes from firms and institution that participate in that market. Airlines want to use AirVM, for example, because they want to understand the behavior of passengers when presented with travel opportunities they are considering offering. In this case, the seller agent in the simulation is a representation of the client for the agent-based model. These virtual markets are characterized by the fact that the narratives of the seller agents are the narratives of the clients they represent. Moreover, the ratiocinators of the client seller agents are incomplete in that the client institution expects to provide choice protocols which reflect its narrative within the simulation, which requires access to the agent ratiocinator. That is, when a decision is called for, the agent doesn't execute the ratiocinator process, the client participating in the simulation does.[47] In other words, the virtual market offers an *avatar* for the simulation client. In addition, other seller agents often represent competitors which react in some way based on choices made by the client agent. The AirVM virtual market has client agents in this sense.

[46] A wonderful experiment would be to use a virtual market representation of a stock market to test the hypothesis that the confidence interval on any market forecast is so wide as to make moot the forecast itself. In other words, the forecast can never be better than a random guess.

[47] Agent-based models are finding some interesting uses in this context, where one or more of the agents in the simulation are actually people. It's obvious that this is the case in many electronic simulation games, where the player manages an avatar which behaves under the direction of the player, and has no independent volition. Agent-based models also are becoming useful in training and medical treatment contexts, as discussed in Shimoda (2002).

**4.2.6:** Customer agents and seller agents are the minimum necessary for a virtual market, but there sometimes can be other agents that are important. Distribution channels, for example, can affect the availability of products at certain times or under certain circumstances that are important for replication of market dynamics. AirVM has agents that represent the institutions and organizations that manage airline ticket sales – reservation systems and travel agents – since they can play a key role in the available options presented to the passenger. Other players important to the dynamics of the market being modeled include suppliers to sellers, regulators, such as government or professional bodies, and other business stakeholders, such as investors or labor groups. All of these agents can be modeled using the general structure presented here.

**4.2.7:** It is also the case that every market represented by a virtual market simulation is embedded in an environment that affects its behavior, and the proper characterization of the dynamics of these environmental influences is also an essential part of the virtual market. As an example, the demand for air travel in the AirVM simulation is external to the simulation, being governed by factors related to general economic activity and geography. The demand data is presented to AirVM as a matrix of numbers with associated probability distributions (representing the weekly number of trips from one city to another) from which the simulation builds essential components, but the derivation of the demand itself is not an integral part of the computer simulation.[48] Any virtual market that is intended to represent an actual market is most likely to have such environmental considerations included. And very often external environmental factors are modeled using rather sophisticated statistical analyses and advanced complex stochastic systems, which themselves are quite challenging. (For a very interesting survey, see Barndorff-Nielsen, 2001.) But the improvement in the overall fidelity of the virtual market is usually worth meeting this challenge.

## 4.3: Synthetic Populations and Incidence Distributions

**4.3.1:** A significant part of the environment of most virtual markets is made up of the population of agents in the simulation. Every agent outside of any individual agent is part of that individual agent's environment. Rarely would there be only a few agents in

[48] It should be noted, however, that AirVM facilitates a method of deriving more accurate estimates of demand, referred to as Network and Channel Adjusted demand. As pointed out earlier, this is a case of the virtual market in the role of instrument being engaged to derive better understanding of the external environment in which it must function, thus improving the fidelity of the virtual market itself.

a virtual market. For many consumer markets, the number of customer agents can rise into the millions. And many markets worthy of simulation with a virtual market have a number of seller agents as well. A collection of agents of the same type (e. g. customer agents) is referred to as a *synthetic population* of those agents. In most virtual markets, there should be at least two synthetic populations – one for the customers and one for the sellers.

**4.3.2:** One of the powerful aspects of agent-based modeling is that the collections of agents can be heterogeneous. Agents of the same type have similar state space structures (which, basically, forms the definition of agent "type"), and heterogeneity is reflected in different values for variables maintained in the state vector. For example, if gender is an important variable in purchase decision, such as might be the case if the product were cosmetics, then the individual customer agents would have to be assigned a gender when they are created in the virtual market simulator. Therefore, since the state space carries data necessary for the representation of the underlying narrative supporting the agent's behavior in the virtual market, the *distribution* of the state space variables in the synthetic population is an important datum for the representativeness and calibration of the simulation. When the agents in a virtual market are created (usually in the initialization phase of the simulation computer program) this distributional data is used to set the relevant state space values of the agents in the synthetic populations. Such distributions are referred to as *incidence distributions*, as they represent the incidence of a certain characteristic or attribute in the synthetic population.

**4.3.3:** If a virtual market is being developed for use as a laboratory or to explore some hypothetical market condition, then incidence distributions can be assumed to be of some simple, convenient form. The purchase probability in the word-of-mouth virtual market presented in Appendix I is the same for all customer agents, and so is set as a constant, but the memory length is some value between one and five (time units), which is assigned using a random number generator applied to uniform [1, 5] integer distribution. For virtual markets which are designed to represent real markets, the incidence distributions must be derived from analysis of the properties of the real customers being modeled. Sometimes this is quite simple, other times complex. In the AirVM virtual market, for example, the length of the total passenger journey is a discrete variable with four values, and so can be assigned using a discrete probability distribution with values found from population survey research. On the other hand, ideal flight departure time is modeled with a mixed normal probability distribution, and ticketing instant (when, prior to flight departure, the passenger buys the ticket) is represented by a non-homogeneous, compound Poisson stochastic counting process.

**4.3.4:** It is important to note that incidence distributions and their relationship to synthetic populations create a coherent structure and rationale for the market research that must be done to build and calibrate a virtual market. If the set of governing narratives requires a corresponding set of variables for their accurate portrayal, then the incidence distributions of those variables are necessary. This gives justification, including quantitative justification, to the field and industry research necessary to generate that data. It also can be used to highlight the costs and risks associated with taking shortcuts and with *not* collecting such information.

**4.3.5:** Finally, synthetic populations are often referred to in this discussion as synpops, trading two syllables for seven. However, a similar contraction of incidence distributions – perhaps to something like *incidists* – is simply too cute!

## 4.4: Computing Concepts for Virtual Markets

**4.4.1:** The identification and design of the agents in a virtual market is, of course, critical to the success of the simulation as a tool for research or as an aid to management decision and policy making. What are the appropriate customer agents? What choice protocols should their ratiocinators call upon? How are the important aspects of the seller agents to be characterized? What other agents are required, and how are important environmental effects to be modeled? These and related questions are aspects of the design of a computer simulation of any type, so now attention turns to the practical consequences and requirements of building a computer simulation.

**4.4.2:** Agent-based models exist because of the capabilities of modern computers. While certainly not necessary for the conceptualization and exploration of the agent concept, as both a theoretical and practical matter they rely for their usefulness on the ability to perform large numbers of computations in a reasonable (human) time. The theoretical rationale is explored fully in Chapter 2 and Appendix A. The conclusion from that discussion is that computational power, specifically simulation, and within that class agent-based simulation, is *sui generis* an extension of man's ability to perceive, contemplate, and manipulate the environment in which he lives. How agent-based models are implemented with today's computing technology needs to be addressed, and is key to understanding that implementation is the concept of object-oriented programming.

**4.4.3:** Object-oriented programming (referred to with the irreverent and slightly sarcastic abbreviation "OOP") was developed as part of a general movement in computer programming away from "folklore" to "engineering." By folklore I mean that programming in the 1960's, 70's and 80's was largely accomplished through the application of skills and techniques that were handed down from one practitioner to another in a context of a specific work environment. To be sure, even in these early days there were a number of schools or courses within schools which taught essential elements of computer programming.[49] But actual programs that did commercially (or even scientifically) useful work were built by individuals who learned not only the formalities of programming, but used undocumented tips and techniques defined by colleagues and unique practices that emerged from the proliferating programming shops. As the field matured, it became increasingly clear that the software modules developed for one purpose might find application in other, unrelated activities. Put another way, software development could learn from engineering, constructing self-contained mechanisms for application in a variety of machines, and the creation of portable modules which accomplish the same function in different contexts would be a worthy objective of the programming art.

**4.4.4:** Simulation had a key role in the development of OOP. *Simula*, a subset of *Algol* intended for programming discrete event simulations, was the computing language that began the move to OOP, first using many of the basic ideas. The first language to call itself object-oriented was *SmallTalk*, developed at Xerox PARC[50] in the 1970's. Many of its features foreshadowed the advent of agent-based modeling computer implementations, such as encapsulation and messaging. But Kernihan and Plauger (1976), with a little book entitled *Software Tools*, established one of the key concepts that moved OOP into the mainstream of programming. In it, they described a "preprocessor" which took code written in easy-to-understand logic, such as simple "if

[49] It was my fate to learn my first computer language – FAP, for Fortran Assembly Program – from a mathematician who knew only slightly more than we, his students, did about computers. He would learn it the night before each lesson.

[50] Xerox PARC (Palo Alto Research Center) also gave the world the windows operating concept, commercialized very successfully by Microsoft.

… then … else” construction, and converted it to classic Fortran 77, which could then be submitted to a Fortran compiler to generate a running program. This idea separated what the human needed to know from what the machine needed to know, and this added layer of translation laid the foundation for the C language (developed by Bell Telephone Labs), which was the predecessor of C+, C++, Objective C, Java and C#, the current top-end Microsoft development language. C# is the language used for the implementation of AirVM. And along the way, the objects picked up other names, like *module* or *class*.

**4.4.5:** In object in an OOP sense is a block of code that stands alone and is treated as a complete entity by other code that uses it. To the reader with some programming experience, a *subroutine* is an adequate metaphor for the moment, although there are a number of technical differences. Objects have *properties* and *methods*. Properties are equivalent (more or less) to data.[51] They include variables, constants, parameters and other information that the object uses to perform its functions. Methods are blocks of codes that actually do something. As a simple example, consider an object that accepts a set of numbers and produces a multivariate linear regression analysis of a set of variables. The properties would be such things as an array containing the independent variables, an array containing the dependent variable, another array containing the estimated coefficients of the regression, and probably a set of variables which hold the regression statistics such as the $R^2$, standard errors, and residuals. The methods would include procedures for error checking, inverting the data matrix, and computing the coefficients, $R^2$ values, standard errors, and so forth. The regression object would be built and tested independently of any particular analysis application, and could be invoked whenever regression was required by some other analysis activity. It is clear that in an agent-based modeling context, agents would be programmed as objects. The state vector is a set of object properties, and the agent object could be expected to have methods representing the activities of the perceptor, actor, and ratiocinator.

[51] Properties can also include code that is to be treated as data, a condition which is referred to as *reflective*. This enables the adaptation of code during the execution of a program.

**4.4.6:** The paradigm of programming with objects is characterized by several distinctive attributes that objects can have. It is these properties which distinguish an object from an ordinary subroutine. Perhaps most distinctive is *messaging* as a vehicle for inter-object communication. Messaging is the paradigm which lies at the core of Microsoft's Windows operating system. When one object wants another to do something, it posts a message to a *queue*. The object to which the request is directed may be busy at the time the message is posted, but it will interrogate the queue at some point (called *pinging* the queue), and proceed to execute the method required by messages addressed to it. The messaging mechanism is also used to move data and results between objects. This process is illustrated in Figure 4.2. On the left, the simple flowchart shows the message reading/action/posting cycle. The queue, shown on the right, keeps track of which object sent the message, the addressee to which it is intended, and, of course, the message contents. Note that many objects can share a queue at the same time, as illustrated in Figure 4.3. The operating system manages the

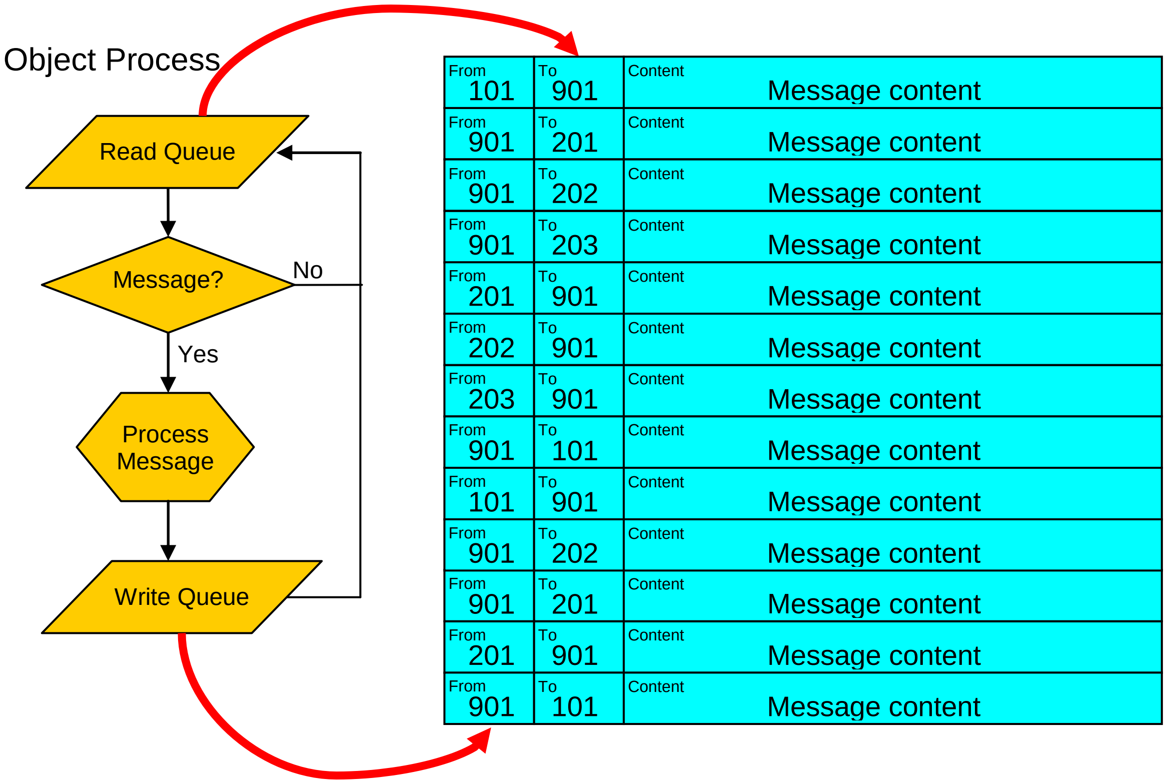



**Figure 4.2: Messaging and Message Queues**

allocation of machine cycles to various objects to optimize the use of the system's resources. Some activities inside the computer, such as moving data to and from disk memory, require substantially longer to complete than others. Messaging allows other resources – computing registers, for example – to be used by other objects while waiting for such long-term activities to finish. In modern operating systems, the programmer sees this multitasking ability as *threads*. Threads are sets of code that operate independently of one another, and essentially simultaneously. A single program can be managing a number of threads, and the programmer can achieve substantial performance gains by the use of threads. Where the hardware allows, several physically separate CPU's can be assigned different threads, which in turn contain threaded portions of the same program. This is essentially how multi-processing works, and this capability also is valuable for programming agent-based models.

**4.4.7:** Other aspects of object structure are also useful to the implementation of agent-

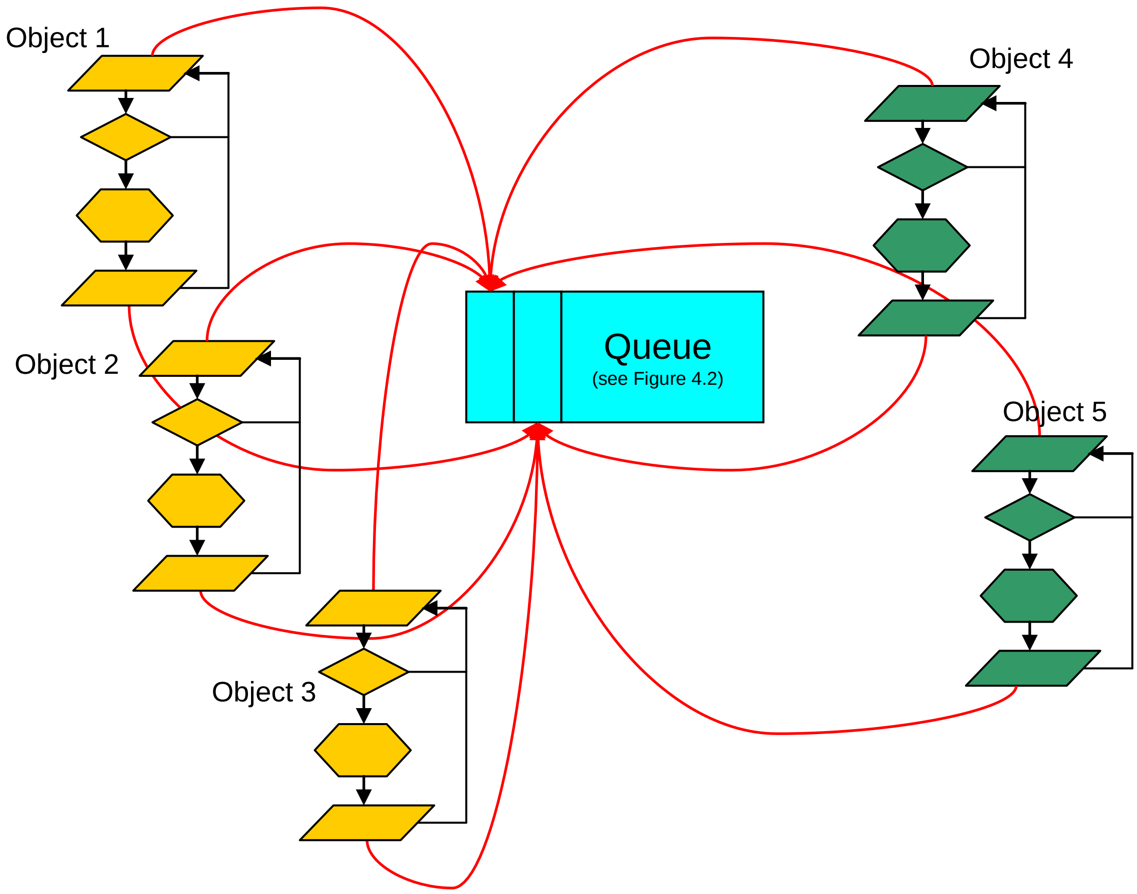



**Figure 4.3: Multiple Objects Using the Same Queue**

based models. *Encapsulation*, which is the hiding of internal operations of an object from view, or molestation, by other objects, keeps the coding cleaner and less prone to error. *Inheritance* allows objects to be created by taking the properties and methods of a base object and adding new properties and methods. Overloading replaces a method in a base object with a similar method in a derived class, but one more specifically suited to the derived class's needs. Thus one can design a generic agent with properties and methods common to all agents, and then specific types of agents that share these aspects, but which have important unique characteristics, can descend from these basic objects. Valenti (1999) carries out this kind of agent construction in his doctoral dissertation. His context is somewhat close to the marketing world, as his work is in the field of computational economics, which makes extensive use of agent modeling.

**4.4.8:** The programming required to build a particular virtual market depends heavily, if not completely, on the computing skills of the team constructing the simulation. In my personal case, I have been programming as part of my work for over four decades, so I have little need to use application systems that are designed for general agent-based modeling use. However, there are a number of available software packages specifically designed for the construction of agent simulations. Appendix H lists some of the more important ones.

**4.4.9:** How the passage of time is managed by the simulation is a vital consideration in design. By definition agent narratives require the passage of time. An agent cannot engage in a choice event until the elements that define that event are in their appropriate condition, an activity that takes some time. For example, in the word-of-mouth simulation described later in Appendix I, an agent cannot become "aware" of a product until one of its immediate neighbors is becomes aware of it, and cannot purchase it until it is aware of it. Time is very often broken into discrete *cycles*, and all the simulation activity designated to occur at a given cycle is completed before the next cycle is initiated. But a computer, in spite of all the power of parallel processors and multiple threads, is essentially a device that operates sequentially over time. First execute this instruction, then that, and so on. True simultaneity is not possible, but it frequently occurs in reality, so the timing mechanism used must assure that simultaneous actions are carried out in a way that reflects the true nature of the relationship. If the simulation designer is not careful in his implementation of the cycle structure, significant unintended results can occur. Consider the following simple example. In the

*Sugarscape* world of Epstein and Axtell's book (Epstein and Axtell, 1996), agents residing on a particular square on the landscape can "feed" off the food in an adjacent square, and if the agent runs out of food, it dies. Suppose two agents are on opposite sides of a square containing food, one of which needs the food to continue. If the other one eats it first, then the hungrier of the two is gone. So the order in time whereby the agents recognize and eat the food affects the path the simulation takes as it runs. If the simulation designer writes the program to operate on each square in the same order at each time cycle, the result will be different than if he randomizes the order, thus bringing the external factor of the programming skill of the designer into the dynamics of the simulation, probably in an unintended way.

**4.4.10:** Usually, the passage of time is marked in the simulation by messages being posted on a queue for all agents to ping, and those which can or must react initiate that action upon receiving the appropriate time posting.[52] One frequent mechanism to handle the simultaneity problem is some type of semaphore, where each agent of a set of agents that need to carry out activity at the same time post a semaphore message that prevents later time activity by other agents until all in the set are finished. If a semaphore mechanism is being used, then the semaphores also are posted as messages on a queue. In AirVM, a semaphore system is used.

**4.4.11:** Another somewhat technical, but operationally and theoretically vital, consideration is the computer generation of random numbers. Random numbers are used to represent the operation of aspects of the simulation that are governed by stochastic behavior. Assuming such behavior is well-defined, then it has an associated probability space $(X, \Omega, P)$ with distribution function *F(x).* If $y \in [0,1]$ is a number with a uniform probability distribution over the domain of the real line between zero and one, then the variable which is the inverse of the distribution *F(x)*, $Y = F^{-1}(y)$, has the probability distribution of *F*.[53] Thus when a random phenomenon is needed in the simulation, a uniform random number is generated by the computer and the appropriate inverse distribution function applied. Of course computer generation of random

[52] This can sometime create a problem when using standard message queue architectures. Many queues purge a message when pinged, to prevent the queue from being any larger than necessary. But when multiple objects must ping the queue, this behavior is not suitable. Rather, the queue structure must allow the message to reside on the queue until no longer needed.

[53] The problem is slightly more complicated than this. If the distribution function has multiple values for a given *y*, a rule must be used to resolve the ambiguity. Often a uniform random number is employed to select one of the equivalent values.

numbers does not produce numbers that are truly random. If one knows the underlying algorithm, and if one also knows the starting place for the algorithm, then one can predict exactly every number to be generated thereafter – hardly random. In fact, it is important that a random number generator be able to produce the exact same sequence of random numbers at will, for this is a necessity for both simulation development and testing and in a number of research applications where the phenomenon of interest needs to be separated from the effects of the randomization.

**4.4.12:** So what is meant by "random?" At the least, the number cannot have any relationship to other factors or variable in the simulation, unless that relationship is part of the formal probability distribution to which the random number is going to be applied. This would guarantee that the number generator is independent of any internal simulation process. The numbers also should not have apparent statistical regularity. Formally, this means that autocorrelations of all orders should be zero. This statistical quality is an ideal, of course, that can never be truly reached, if for no other reason than the computer is a finite state device and eventually must repeat itself.

**4.4.13:** Extensive research has gone on to develop random generators that meet the needs of modern simulation techniques. Most are variations on congruence methods. Congruence methods take a numeric value, multiply it by itself, and take a span of bits out of the middle of the resulting product. That span is the random number, and is used for the value in the multiplication the next time a number is generated. The original value that starts the process is called the random number *seed*, and the identical set of numbers is generated every time the same seed is used. At this time, the Marsenne twister algorithm seems the most preferred. It was developed by Matsumoto (1998) and is based on matrix linear operations on binary fields. It offers fast generation of high-quality pseudorandom numbers. Its name derives from the fact that the period length is chosen to be a *Mersenne prime*,[54] which is a prime number that is one less than some power of two. There are a number of variants of the algorithm, differing only in the size of the Mersenne primes used. The most common is the Mersenne Twister MT19937, with a 32-bit word length and its variant with a 64-bit word length, MT19937-64, which generates a different sequence. The code for this algorithm is freely available. In

[54] Named after Marin Mersenne (1588-1648), a French mathematician who was a contemporary of Descartes and Pascal.

particular, the Troschuetz (2007) suite is quite extensive, with a number of common probability distributions represented. It is free, and the source code is provided.

**4.4.14:** It is evident that the computing requirements to build a virtual market can be considerable, but they are not prohibitive. Many of the deeper technical issues have been addressed and successfully resolved. The capabilities of current software and the wide-spread availability of inexpensive computing resources definitely bring the tool within the reach of most researchers and practitioners. Clearly, it is not the case that a fixed formula for computer simulation design and implementation has yet been adopted. But with the considerable practical experience that has been accumulated, and the sufficient talent and expertise that can be readily found, quite complex problems can be successfully attacked.

## 4.5: Practical Issues in Building Virtual Markets

**4.5.1:** The time aspect of a simulation highlights one of the practical issues that arise in the design and construction of simulation models. Turning more generally to that concern, Grimm and Railsback (2005) provide an excellent reference on how to build agent-based models (or individual-based models, IBMs, as they put it), with an important emphasis on the practical problems that must be addressed and resolved. Although their subject matter is ecology (where IBMs are being used more and more), their observations and advice extend, in my view, to any agent-based development program. They characterize the creation of an agent-based model with a set of sequential steps which are worth review.

**4.5.2:** The first task is c*onceptual design*, where the things being modeled are defined and their characteristics specified. This specification includes: a) the definition of the various agents in the simulation; b) how their essential behaviors are to be modeled, including a precise specification of any behavioral equations that might be needed; c) what kind of stochasticity is to be replicated, specifically the probability spaces and distributions that are to be used to represent those stochastic elements; d) the time factors that affect the behavior of the agents; e) any characteristics of the environment that are important to the purposes of the simulation but not represented by agents in the simulation; f) the important interrelationships among the agents in the simulation and between the agents and the environment; and g) how the results of the simulation are going to be observed and recorded for future analysis (Grimm and Railsback, 2005, pp.71-121).

**4.5.3:** With the conceptual design in hand, the second activity is *programming* the simulation. This encompasses: a) selecting an appropriate computing hardware platform; b) choosing a programming language or application development software package suited to the skills of the simulation development team and the capabilities of the computing platform; c) writing and testing the code that implements the simulation; and d) validating that the code accurately and validly represents the models and interrelationships described in the design (Grimm and Railsback, 2005, pp. 270 - 311). From my own experience, these tasks are the most difficult and time-consuming of the entire simulation development effort.

**4.5.4:** The third activity is the *analysis* of the simulation results. An agent-based model can easily generate a prodigious quantity of output. The AirVM simulator described later in this dissertation generates approximately 280Mb of output data for each run of the simulation. A Monte Carlo AirVM simulation with 500 iterations (a rule-of-thumb nominal size for distributional analyses) can then produce nearly 140 Gb of output data. Analysis of this quantity of data can be a forbidding endeavor if some care is not exercised in determining what to look at and how to summarize the data into meaningful statistical and distributional quantities. Grimm and Railsback identify four issues that they feel must be addressed in this phase of simulation development (Grimm and Railsback, 2005, p. 314-315). First, is the software still representing the models faithfully and accurately? Software development and implementation invariably causes alteration in the original design of a simulation, and these changes are often forgotten as work proceeds. Thus the software has to be checked. Second, assuming the software is faithful in its representation of the theory, how valid are the models themselves? Note that the question is not whether the models are valid or not, but rather *how* valid are they. All models, by definition, are incomplete descriptions and therefore are invalid to some degree. How valid a simulator must be to be acceptable for the purposes for which it is being built is an essential criterion that must be delineated as the simulation is developed. Third, calibration (or as Grimm and Railsback put it, parameterization) needs to be verified. This encompasses evaluation of the data and statistical analyses used to estimate model parameters to establish that they are reasonable representations of the phenomenon being modeled. Finally, the fourth step is determining if the simulation is addressing the set of problems and opportunities the agent-based model is being developed to solve. In the case of virtual markets, this means, among other things, applying the simulation to actual marketing issues and policies to improve management and organizational decisions, and increase profitability for the client firm or organization.

**4.5.5:** Notice, however, that Grimm and Railsback do not include *model accuracy* as one of their criteria. What is meant by accuracy? The first intuition is to compare the model's output with observations to the real world, and see if the two differ. But no model is one hundred percent accurate, for models are always simplifications of what is being studied, and aspects of the phenomenon being modeled that are simplified away can have some effect on the real world observations. Indeed, they had better have some effect, for otherwise they are irrelevant, and would never have come up in the model building anyway. So rather than measure accuracy against real-world observations, because that will always be wrong, perhaps it is wiser to measure the accuracy with respect to other models. This can be done by applying whatever criterion is used to determine the accuracy with the original model to the new model, to see which is better.

**4.5.6:** With agent-based models, however, the situation is exacerbated by the nature of the thing being modeled. When applied to complex adaptive systems, (which is often the case with respect to marketing) each execution of an agent-based model represents one dependent path of the evolution of the system being modeled. Moreover, each independent observation of the complex system being modeled is also the result of a single dependent path. It is highly unlikely that the simulation and the real phenomenon would match, purely because of the path dependency. (In fact, with systems that have an uncountable number of dependent paths, the probability that two paths are identical is zero.) This situation can be detected by measuring the actual phenomenon multiple times. If path dependency is at work, it will be very difficult, if not impossible, to get two measurements that are identical. An example of this is found in the passenger airline virtual market described in the next three chapters. Repeated measurements of even simple variables – say the load on a flight from point A to point B – rarely yield equal values. But, one might suggest, take the average or some other statistical summary of the real observations, and compare that with the a similar statistic derived from the simulation. That only begs the question. It is difficult in most cases to get *independent* observations from a complex system (path dependency is a wicked thing), so the probability distribution of the observed values is not knowable, and thus the probabilities associated with the statistical comparison, whatever it is, is also not knowable. Grimm and Railsback argue, by implication, that if the components of the agent-based model are built carefully, and the parameter estimates are valid, then the model is valid. If the insights that result from the agent-based model are useful (to management, in the case of marketing), then the model is as accurate as it needs to be.

**4.5.7:** Another set of issues that must be addressed in building and using an agent-based model is the *presentation* of the simulation to those who need to use its products in their business or research contexts. (Grimm and Railsback, 2005, pp. 349-358) For agent-based simulations, successful accomplishment of this facet of the development is critical to the success of the effort. I have coined the term *perspectives and portrayals* as a label for this element of a simulation. Essentially, the outcome of the simulator must be relevant to questions that the business user has, and how the simulation represents that relevance must be portrayed to the client in a way s/he can understand and implement. There are no hard and fast rules for determining the best methods of defining perspectives and designing portrayals, but there has been significant attention given to their development in the literature. The wonderful books by Tufte (1990, 1997, 2001, 2006) are indispensable for the successful design of perspective and portrayal interfaces. Harris (1996) offers a less heartwarming, but nonetheless fully satisfying, compendium of visual tools. Card *et al.* (1999) discuss specifically computer-generated methods of information display. And Thalmann and Thallman (1990) lay out basic principles of computer animation, which allows the addition of a fourth dimension – time – to a two or three dimensional display. I am convinced that as the field of agent-modeling matures, animation will take its place as one of the essential features of a successful agent-simulation. Within computing, substantial advances have occurred in giving applications developers and designers access to significant graphics capabilities. The Windows Presentation Foundation (MacDonald, 2008) is one such substantial advance.

**4.5.8:** The final practical rule in simulation model development is to be aware that things will change dramatically from the beginning of the development process to the delivery of the final product. Defining the problem is a cyclic process. Creating a computer program causes repeated revisions, redesigns, and sometime wholesale replacement of pieces of the model based on results to date. The analysis phase always requires revisions in the design and implementation. And the entire effort must be able to be altered to meet the needs for perspective and portrayal if the work is to be of any use. In short, rigid adherence to a set procedure for simulation model development is a recipe for failure. Ultimately, flexibility and imagination, and, frankly, a rather dogged determination, must be the guiding principles of for successful development. On the other hand, therein lies the fun. Many of these practical issues are illustrated in the simple Word-of-Mouth Cellular Automata virtual market described in Appendix I. The reader is invited to look there for more insight into some of these aspects of agent-based modeling.

## 4.6: Summary

**4.6.1:** The fourth step in the research program is now complete. Virtual markets have been defined, which encapsulate the narrative framework coming out of Chapter 3. Some insight has been offered into how to build them on a practical level. The stage is set, then, for the actual construction of a working virtual market simulation. This process begins with Chapter 5.

# Chapter 5:
# The Airline Passenger Virtual Market – AirVM

## 5.1: Introduction

**5.1.1:** In the preceding chapters of this dissertation, the definitional, epistemological, theoretical, empirical, and operational groundwork has been laid for the development of a virtual market. With this chapter I begin the process of constructing a working virtual market used for applications in the airline passenger market. This agent-based simulation is called "AirVM" (for Airline Virtual Market). It is being developed and is owned by Virtual M1nds, SA, of Vevey, Switzerland. The author is one of the principals of that firm, along with partners in Switzerland and England. The airline passenger market was chosen for several reasons. First, the author has thirteen years of experience with the marketing department of a major player in the industry, the Boeing Company, as their senior marketing scientist, so there is ample understanding of how the industry works and how the passengers, airlines, airports and other players in that marketplace interact. Second, as a consumer product, an airline ticket is fairly simple to characterize, and therefore relatively simple to structure into a virtual market description. There is little glamorous or fashionable about a ticket to fly on an airplane (although there are "scales" of tickets in the sense of first class vs. economy). Therefore, many of the characteristics of the product being sold are not as complex as, say, consumer goods like automobiles, appliances or beauty aids. Third, the market is reasonably "closed" in the sense that major drivers and influences within the market are comparatively well understood, and therefore less open to various interpretations and explanations. All these features when taken together offer an opportunity for the construction of a virtual market that is both realistic and useful, but not beyond technical skills of a single person.

**5.1.2:** In this chapter, the nature and workings of the market for airline passenger tickets is presented in sufficient detail for the reader to follow the logic of AirVM's structure and its relationship to that market. Following that background, the three important narrative structures of the simulation, those of the airline passenger, of the air carrier, and the ticket sales and distribution systems used in the industry are considered. Then, after a conceptual overview of the AirVM simulation (to give context), the three agents in the virtual market model are examined in detail. These are the *pags* (*p*assenger *ag*ent*s*), *arasags* (*a*irline *r*evenue *a*nd *s*chedule *ag*ent*s*), and *dsags* (*d*istribution *s*ystem *ag*ent*s*). In Chapter 6, the incidence distributions of these agents are described in technical detail. These include the stochastic structure of the ticketing purchase time (in terms of time before departure), the travel party size, the ideal time of day distribution, and, perhaps most importantly, the decision protocol model of itinerary purchase.

**5.1.3:** Features of arasags and dsags are also discussed in Chapter 6. Among these are representations of the commercial airline network and the generation of potential itineraries and the various revenue management pricing protocols used by the airlines. These models are not as extensive as those relating to the pag, but are nonetheless important to the composition of the virtual market. Also, arasags and dsags are client agents in the structure of AirVM. By that is meant the narratives that drive their choice protocols are intended to come from client users of the virtual market. Therefore they can be considered as avatars for customers using AirVM.

**5.1.4:** In Chapter 7, attention will then be turned to the actual simulation computer program. I will describe its current computing architecture and logic, its input data requirements, its operating characteristics (speed, memory requirements, etc.) and the nature of its user perspectives and portrayals. I then discuss how the tool can be applied by various clients – airlines, airports, regulatory authorities, the financial community, aircraft manufacturers, and industry data suppliers. Model estimation and calibration, and examples of execution and output of the programmed are then described. The discussion ends with a critique of AirVM, including its contributions and its shortfalls.

**5.1.5:** The author of this thesis is Senior Scientist and Partner in the firm of Virtual M1nds, SA, of Vevey, Switzerland, and AirVM is a commercial product owned by Virtual M1nds and currently in use within the airline industry. However, who uses it, how it is applied by specific organizations to various problems, and the technical details of its application in these multiple contexts are generally protected by nondisclosure agreements or other intellectual property vehicles. Similarly, the calibrated values of the empirical parameter of many of the models used in these applications are trade secrets or otherwise not in the public domain. These are legal agreements that restrict access to some details of the workings of AirVM, and they must be respected in this discussion. Nonetheless, the concepts on which AirVM is based, the sundry stochastic processes that are engaged, and other important conceptual structures that are applied to the virtual market simulation come from published works or activities of my own carried out in the preparation of this dissertation. These are openly shared in this presentation. However, the reader will occasionally see notations to the effect that a certain parameter value, for example, cannot be represented as used by a particular customer because of its confidential status. Also, only selected AirVM users will be identified in this discussion, as the disclosure of others would be a violation of the legal arrangements between the customer and Virtual M1nds.

**5.1.6:** Patent protection for several of the unique aspects of AirVM has been applied for under the intellectual property laws of the United States. The United States was chosen for this filing because its intellectual property law extends patent coverage to business processes, including computer software and systems, to a greater extent than that of other countries. This approach, as opposed to trade secret or other methods of intellectual property protection, was chosen to allow full disclosure and peer review of the concepts and methods created and implemented as part of the development process of AirVM.

**5.1.7:** Much of the development of the models, data and data analyses that support AirVM have involved the work of others, both my colleagues at Boeing and members of the research team who were under contract to my group in Boeing Marketing. Where ever this is the case in the ensuing discussion I have noted their contributions, often jointly with my own. Those aspects of the work that are solely mine are also identified. In particular, all the computer development of AirVM has been done by me alone, and so I must take responsibility for errors and omissions in that regard.

## 5.2: A Brief Exploration of the World's Airline Passenger Market

**5.2.1:** The market for airline tickets by passengers wishing to travel is the main focus of AirVM. In order to understand how the simulation must behave, it is necessary to have some level of understanding of how the air industry market is structured. Overall the airline industry is made up of seven major sets of players: 1) passengers – the individuals who want to travel from an origin to a destination on board an airplane; 2) airlines – the companies and organizations that provide the means by which passengers can take an airplane trip; 3) airports – local organizations who make available facilities that allow the transition from ground transport to air transport, as well as support facilities required for the operations of airlines (fuel, maintenance, etc.); 4) local government – the government entities which have jurisdiction over the area in which an airport is located, and which are responsible for providing access, utilities, and other supporting facilities; 5) regulators – offices of government (usually) responsible for the monitoring of the air transport system to ensure public safety, often including the regulation of air space and air traffic; 6) distributors – entities which go between the passengers and the airlines providing brokerage or clearing house functions for the purchase of airline tickets; and 7) supporting companies – companies which manufacture and sell airplanes, firms that provide maintenance and support services, financial groups that supply capital for to the industry, and airport retail services that serve the traveling public.

**5.2.2:** From the passengers' perspective, the main "good" provided by the industry is a seat on an airplane going from the desired origin point to the desired destination, at a price and time acceptable to the passenger and with reasonable comfort and assured safety – in short, access to the airline network. The network is the collection of routes and schedules of airplanes with various capacities, ranges and features. Routing and scheduling is done to maximize revenue while minimizing airline operating and capital costs. Airports and the local political structure within which airports operate have a dramatic impact on route availability and schedule timing. For example, airport capacity coupled with aircraft performance constraints, can determine if two cities can be connected directly or only through an intermediate stop.[55] Also, local flight noise restrictions and air pollutant emission standards, enforced by airport curfews, can prohibit departures or arrivals at night, causing possibly otherwise desirable flight times to be unavailable.

**5.2.3:** Airline seat pricing is among the most sophisticated of any consumer industry in the world, with advanced revenue management systems in use by virtually every commercial airline in the industry (Littlewood 1972; Belobaba, 1987, 1989; Brumelle and McGill, 1993; Robinson, 1995; Li and Oum; 2002; Talluri and van Ryzen, 2004,

[55] Specifically, a particular aircraft may be able to connect city A to B, but the runway length at A will not allow a full airplane to take off in most weather conditions. Thus a smaller, lower take-off-weight plane may be needed, which does not have the range to reach B, and therefore requires an intermediate stop.

Berger and Fuchs, 2005). These systems constantly monitor realized demand and adjust price for every flight in an airline's system, trying to anticipate future demand at the departure time and competitive position to squeeze every drop of potential revenue from the consumer.[56] Revenue management is an important characteristic of the airline industry, and it is being adopted across an increasingly broad range of services and goods outside of air transport. It is now used in the rental car and hotel markets, and is gaining increased attention from a number of other fields.

**5.2.4:** The relationship between competing airlines, their revenue management systems, their schedules and their routing structures as it interacts with the behavior of the passengers and under the business and social models of the airport authorities I term the *fundamental dynamic* of the industry. The revenue available to the industry, and its division among the airline participants, is determined by how these entities behave as they respond to each other's actions and reactions.

**5.2.5:** As industries go, the airline industry can be considered moderately sized. There are approximately eight to ten million passengers traveling world-wide each day. Obviously the rough average of nine million per day does not translate to 9 x 365 million per year, due to multiple flights per year per person. It is not known how many people in the world have never taken a commercial airplane flight. The passengers move in and out of over 3900 commercial airports around the globe (Innovata, 2009). There are about 1000 carriers who provide scheduled airline service between those airports (Innovata, 2009), flying some 23,000 aircraft (Boeing, 2008).

**5.2.6:** The price of the good (the seat on the plane) to the consumer, the passenger, ranges from a few dollars for intra-Europe travel on a low cost carrier such as Ryanair to upwards of $10,000 or more for a transatlantic or transpacific first class fare on a full service carrier like Qantas. At the other end of the scale, the price of an airplane can vary from $10M or so for a regional jet, to over $280M for an Airbus A380. The acquisition of the aircraft is the most expensive capital expenditure required for an airline, rather on the scale of a house purchase for a family. Therefore, it is reasonable to expect the process of airplane sales to be extensive and involved. In a given year, relatively few airplanes are sold (only in recent years has annual sales exceeded 1000 units for either of the two major manufacturers, Airbus and Boeing) to relatively few

[56] Often to the complete bewilderment of the traveling public, a situation that has been capitalized on by a number of emerging low-cost carriers who tout 'transparent pricing.'

customers (rarely do more than a few dozen airlines place orders in any one year). And each unit is a substantial capital investment.

**5.2.7:** Demand for air travel – specifically demand for individual flights between a given set of origins and destinations – is very sensitive to externalities. War, disease and air accidents can all suddenly cause travel demand to plummet. An example in the United States was the terrorist attacks of 11 September 2001. Across most of the US domestic market, and to a significant degree across the international market as well, demand dropped as much as 40% within just a few days of the event. In some markets the recovery time was as long as four years (Cinecizoglu, Carson and Parker, 2007). This triggered (although was not the sole cause of) the deepest decline ever in air travel demand and the greatest losses in the industry in history.[57] Financial losses to US air carriers exceeded US$26bn alone. And while this has certainly been the most devastating down cycle, it is not the only such event in the post World War II era. There were similar downturns during the 1991 Gulf War and in Asia at the SARS epidemic in early 2003.

**5.2.8:** Another feature of the industry is more theoretical in nature, but perhaps more important to understand in the long run. It is hypothesized that the airline industry is what an economist would term "empty-core." As defined by Button, an economy has a core if, (2003, p. 5):

> "A simple economy can be modeled as a set of economic actors (or players) who produce goods and services and transact with one another. An allocation of goods and services is said to be in the core when there is no group within the economy that could be better off trading amongst itself. [Thus] … For an outcome to be in the core of any economy no further gains from trade are possible for any group or subgroup."

Therefore an economy has an empty core if there is no (Nash) equilibrium. Since the deregulation of the airline industry, which began in the US in the late 1970's and has since migrated across the world, there has been mounting evidence that the core of the industry is indeed empty. This was suggested as early as 1995 by Button and has been reiterated by Julius (2002) and Alamdari (2007). Of course, if the industry does have an empty core, then the implications for the long term future are indeed dramatic, for the

[57] It may turn out that the recession of 2008/9/10 outstrips the 9/11 attack in terms of impact on air travel.

only effective remedy to the empty core condition is outside regulation, a complete reversal of the trend of the last thirty years. While neither the cyclic nature of the industry nor the empty core hypothesis is central to the focus of this discussion, the laboratory application of AirVM creates new tools with which new light can be shed on both issues.

## 5.3: Passengers Buying Tickets and Airlines Selling Them

**5.3.1:** Central to the marketing structure of the industry is the purchase behavior of the airline passenger. There can be no doubt that deeper understanding of how and why passengers choose particular airlines to meet their travel needs can only lead to smarter responses by airlines in offering services that passengers will buy. With the exception of cargo movement, every aspect of the airline industry – from the range of fare products offered to the very survival of the carrier – depends on the passenger. If the passenger won't pay for the services that an airline offers, quite simply that airline will fail. Low cost carriers have historically put a premium on passenger relationships (over and above low price, it should be noted) and their (largely negative) impact on the business operations of the legacy, "full service" airlines is evident. It is important to be aware of some of the more salient features of the air passenger market that are essential to the understanding of passengers' and airlines' narratives. A number of terms-of-art the industry are also introduced, and these are summarized for ease of reference in Appendix J of this dissertation.

**5.3.2:** In essence, an airline is offering to passengers opportunities to safely move from an origin to (one or more) destinations by means of a seat on an airplane, in a reasonably well appointed cabin, departing and arriving at reasonable times, hopefully with as few stops and plane changes as possible. That the passenger desires these features (and to what extent) is empirically established through a number of passenger choice studies. Specific applications of the technique to the airline industry can be surveyed by considering Yoo and Ashford (1996), Train (1998), Proussaloglou and Koppelman (1999), Ben-Akiva and Bierlaire (1999), Bates (2000), Algers and Beser (2001), Coldren *et al.* (2003), Hess (2003), Parker (2004), Adler *et al.* (2005), Caldron and Koppelman (2005), Garrow and Parker (2005), Lonsdale *et al.* (2005), Parker and Lonsdale (2005), Garrow *et al.* (2005), Bhat *et al.* (2006), Parker and Garrow (2006) and Parker and Walker (2007).

**5.3.3:** The vehicle by which services are offered to meet these needs is the *regularly scheduled commercial airline network*. The network is a set of airports in different cities that are connected by scheduled, commercial flights. The term *connected* means that a passenger can board a plane at one airport and get off the plane in another, perhaps having to change planes at some intermediate airport on the way. The flights are *commercial* because the passenger purchases the right to sit in a seat on each plane required for the connection. Sometimes the seats are available in cabins which offer different classes of in-flight comfort and service, for which the passenger usually pays differently. The flights are *regularly scheduled* in that the departure and arrival times of

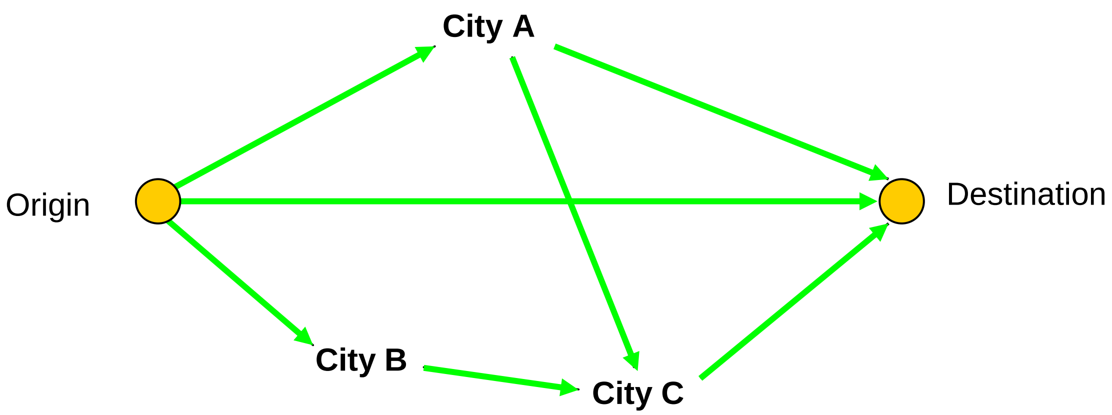



**Figure 5.1: A Simple Airline Network Description**

the connections are known in advance, published as a schedule for anyone to see, and therefore become part of the passenger's purchase decision. In fact, that passengers can purchase tickets prior to the departure of the flight is an important airline revenue dynamic. Also, airplane cabins have *capacities*. They cannot carry more than a fixed number of passengers on any one flight.[58]

[58] In Europe, cabin definition can be somewhat flexible, given that a movable partition is used to separate business from economy cabins, thus allowing adjustment of how many seats are available in each.

**5.3.4:** As an example (and to set some terminology) take a look at the simple network shown in Figure 5.1. Here the origin and destination can be reached with flights that are nonstops, or flights that go through cities A, B or C. A sequence of flights (also called *legs*, *flight legs*, or *segments*) that allows the passenger to travel from the origin to the destination is referred to as an *itinerary*, or as a *path* (the two words are interchangeable). A *nonstop segment* is a flight that doesn't land for passenger embarkation or debarkation between the time it leaves the origin and the time it arrives at the destination.[59] Passengers travelling on a nonstop are often referred to as *local passengers*, as they are in a "local" market. A *direct* segment uses a single airplane that stops *en route*, with the passengers often remaining onboard during the stop, and usually the legs have the same flight number. Passengers in this case are referred to as *through passengers*. *Connecting* segments use more than one airplane and more than one flight number, and the passenger deplanes and enplanes at the connecting airport(s). There are two types of connections made by connecting passengers: *online*, where the passenger is connecting between flights of the same airline, or between flights of airlines within a common alliance, and *interline*, where the passenger is connecting between flights of

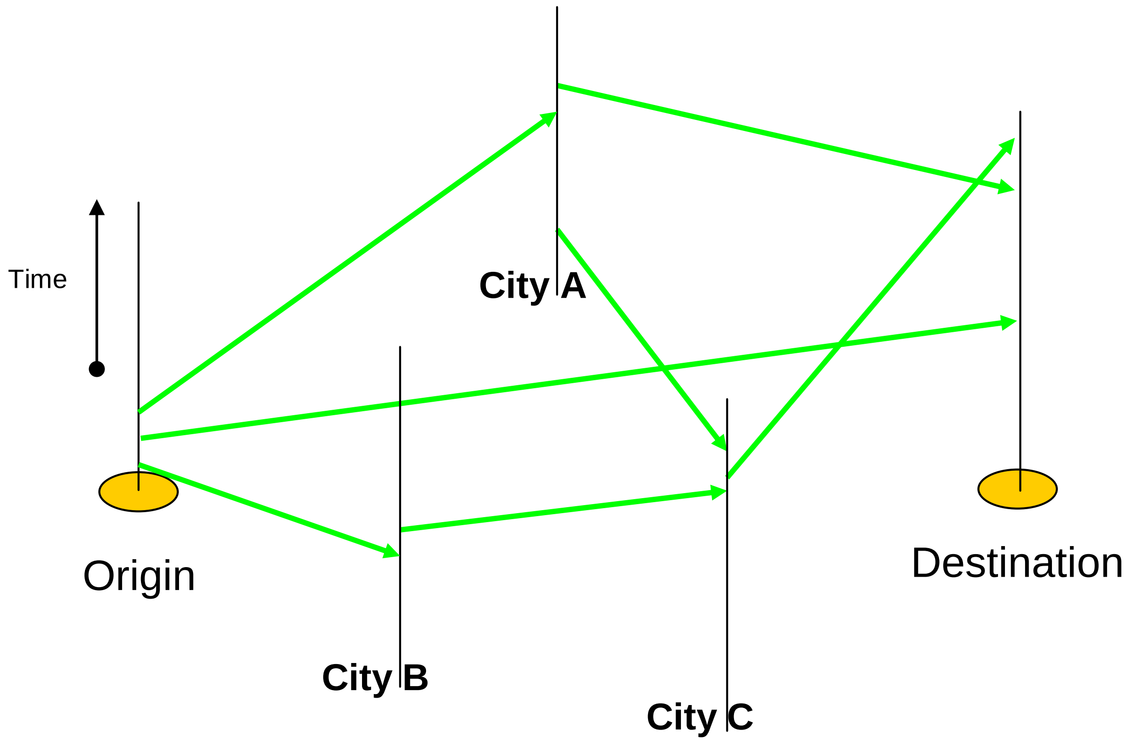



**Figure 5.2: The Airline Network in Three Dimensions**

[59] Thus a plane that lands simply to refuel is still considered a nonstop.

different airlines. Online connections benefit the passenger by providing through check-in, faster baggage transfer, better gate proximity and usually shorter connect times. An itinerary can comprise any combination of these segments. For example, a direct segment can be followed by an online connection to a nonstop segment. Another path might consist of two nonstop segments separated by an interline connection, and so on.

**5.3.5:** In an important sense, the representation in Figure 5.1 is inadequate because there is no indication of the schedule flown by the airplanes that move between the various cities. A more accurate diagram is shown in Figure 5.2. Here, time is shown as a vertical axis, with the departure time at the tail of the arrow and the arrival time at the head. From this perspective, it can be seen, for example, that the connection from City A to City C really doesn't work as an itinerary for this origin-destination pair, because it arrives in City C after the flight leg from City C to the Destination has departed. This three-dimensionality of the airline network has important implications for network analysis.

**5.3.6:** One other characteristic of airline networks needs to be highlighted. Looking at the simple network of Figure 5.1, it is clear the leg that forms the path from the Origin to the Destination that goes through City A is also a (nonstop) itinerary from the Origin to City A. That is, a specific flight leg can be a part of a number of itineraries, not only with respect to the OD pair in question, but with respect to a number of other OD pairs as well. This means that when a passenger buys a ticket on a segment serving one itinerary, he removes that seat from consideration by other passenger that book later, thus altering the choices the later passengers has available to him. This is a vital aspect of the dynamics of the airline network, and has significant ramifications for the design of AirVM.

**5.3.7:** Most people in developed countries today have had some experience with buying an airplane ticket, now some seven decades after the widespread introduction of commercial air travel after World War II. An individual decides, for any of a multitude of reasons, that he or she wants to go from where he lives or works to some other place, and generally return fairly quickly thereafter. She then decides if the mode of travel will be airplane, as opposed to rail, boat or automobile, and proceeds to contact a ticket seller. Airplane tickets are sold through any of a number of *distribution channels*, including directly from the airlines, through online and brick-and-mortar travel agencies, and through intermediary resellers, such as package tour agencies. The fraction of tickets sold through each channel varies across markets and countries. Also, tickets are almost always purchased in advance of the scheduled departure of the desired flight. Many, but not all ticket sellers, allow a reservation to be made prior to when the ticket is actually sold. This practice, however, has diminished in recent years due to the increased use of revenue management systems by air carriers. To maintain the distinction between reservation and sale I use the term "booking" for a reservation and "ticketing" for an actual sale.

**5.3.8:** On average, a little less than half of air travel is because of business requirements, the remainder for personal travel. The industry commonly classifies trip purpose as *business* or *leisure*, reflecting this business and personal split, respectively. It is indicative of the state of customer orientation and focus in the industry that these two very broad categories are not routinely broken down to finer segments. For example, leisure can be further classified into casual (short notice) or planned (vacation). Business can be further classified into internal company, external supplier, external customer, or personal business. Some researchers, notably Forrester Research (2007), do attempt to apply such a more discrete classification.

**5.3.9:** Another aspect of passenger travel that is (relatively) little studied in the industry is what I call *journey structure*. Journey structure is the pattern of itineraries that make up the total air journey. A simple trip to and from a single destination is a *round* trip, while a trip from home to one city, then another, and perhaps more before returning home is called a *multi-stop* trip. Most journeys start and end at the same place, often the traveler's home, but those that do not are termed *migratory* trips, because it usually occurs as part of a residence move to another city, such as a student going to college for a term. There is one particular aspect of journey structure that is important for the implementation of AirVM. This is *departure* vs. *arrival time sensitivity*. This time orientation refers to the element of itinerary schedule that the passenger decides is most important, when the itinerary starts (departure) or when it ends (arrival). For example, for a business trip, the flight to the destination is usually governed by the arrival time, in that the passenger must get to the destination in time for whatever business he has to deal with, while the flight returning home is usually departure time oriented, since he wants to leave for home as soon as he can after his business is done.

**5.3.10:** Booking and ticketing occur over time prior to the departure date the passenger wants. As departure time approaches, it is usually the case that the number of tickets sold per unit time increases. This time dimension is important, for seats on board an aircraft fill up as ticketing proceeds, and a specific fare class on a given flight may no longer be available if the passenger is buying too near the departure time. The time a booking is made is called the *booking instant*, and the time a ticketing event occurs similarly called the *ticketing instant*. The booking instant is when the reservation is made, while the ticketing instant is when the ticket is actually purchased. Clearly the booking instance for a purchase occurs before the ticketing instant, but very often the two time points are simultaneous. The distinction is made because the rate of cancellation of a *booking* is much higher than the rate of cancellation of a *ticketing* event, and cancellation rates directly affect availability. Later in this discussion the stochastic processes which represent the ticketing time-dependent activity will be modeled for the simulation using a non-homogeneous, compound Poisson processes (see Chapter 6). As it happens, the parameters of these processes have different values depending on trip purpose. AirVM does not consider booking instances in its formulation. This is because they are becoming less and less frequently used (airlines are forcing passengers to buy rather than reserve), and research has shown (Illiescu *et al.*, 2007) that the estimation of cancellation behavior parameters of passengers with respect to bookings is substantially less reliable than that for ticketing.

**5.3.11:** Often an individual buying a ticket will purchase more than one, as would be the case if he were buying vacation tickets for himself and his family. The number of tickets purchased at one ticketing instance is called the *group* size, or *party* size. Party size is also modeled (using a simple truncated Poisson distribution) for AirVM. Like the timing of the ticketing events, it also is different for different trip purposes. Business groups are somewhat smaller, on average, than leisure groups. Also, there is a "bump" in the right-hand tail of the probability distribution of the size of the traveling parties. This is created by large group sales, such as tour groups, which occur in both leisure and business categories. At this time, because of lack of data, AirVM ignores these effects.

**5.3.12:** The setting of fares on a commercial airplane is one of the more complex aspects of airline operations. From an operational economics perspective, the airplane seat is considered a perishable good, since once the plane departs unsold seats have zero value – they parish. Almost all carriers in the world now employ a system of *revenue management* to maximize the revenue that can be generated on any given flight. Prices are assigned to not to seats, but to groups of seats referred to as *fare classes*. One cabin, for instance the main cabin, can have a number of fare classes. All the seats in a fare class are at the same price. Fare classes are distinguished by rules called *fences*, such as advance (of day of departure) purchase requirements or refundability restrictions (full refund, exchange only, change fees, non-refundable, and so forth), and there usually are a number of fare classes in the same cabin. Fare classes are rarely apparent to the passenger.[60] Typically, the pricing team in an airline establishes a base price for each fare class, determined by cost factors and corporate revenue considerations. The revenue management team, on the other hand, adjusts the fares in each fare class as a function of the realized demand-to-date or some kind of forecast of the final demand-at-departure. The objective is to maximize revenue by raising prices for seats as they become less available during the ticketing period, or lowering them (or keeping them low) to fill up seats that otherwise would go empty. Revenue management protocols can be quite elaborate, and extensive operations research methods are applied to the problem. And for good reason: revenue management is often given credit for more than a 2% improvement in carrier gross revenue, which sometimes can be the entire profit margin for an airline (see Belobaba, 2008 for a quick summary, or Talluri and van Ryzin, 2004, for an extensive analysis).

**5.3.13:** Of fundamental importance to the passenger is how long it takes to get from where she is to where she wants to go using a particular itinerary – the total travel time. Precisely defined, the travel time from the passenger viewpoint is the total time from when she leaves home (or her office) at the origin to when she arrives at the hotel, meeting site, company facility or friend's home at the destination. This includes ground travel time to the origin airport, time required to check in, clear security, and board the plane, time in flight for each segment in the itinerary, time spent on the ground at intermediate airport which are stops, and the time to pick up any baggage, clear customs and immigration if needed, exit the airport, and utilize ground transportation to get to the final destination. Within this travel time is the itinerary *duration*. This is the clock

[60]Giving rise, no doubt, to the frequently-cited stories of two individuals sitting side-by-side on the plane, one having paid three times as much his ticket as the other, with no apparent difference in the service received.

time between when the aircraft used in the first leg of the itinerary closes its the doors after all the passengers have boarded, to when they open for deplanement at the end of the final leg of the itinerary. Every itinerary has an associated duration, and the shortest duration for a set of itineraries serving a given origin-destination pair is referred to as the *base duration* for that pair. For origin and destination cities that are served by single airports, the time spent before and after the actual itinerary duration that is engaged is the same, and can thus has no bearing on which itinerary is chosen. For cities supported by multiple airports, that may not be the case.

**5.3.14:** Finally, some tickets are sold as blocks in special circumstances. Cruise companies, for example, often bundle airfares into the costs of a cruise. They do this by acquiring the option to buy a large block of tickets on a specific flight designed to bring individual tourists to a central point to start the cruise. This often results in blocks of ten or twenty tickets being purchased as a unit on a single leg. Also, large companies often make arrangements to buy tickets for their employees at lower prices. These are not large groups of passengers, as seen in the cruise example, but the air fares are different for these corporate customers, and thus the choice protocol for the corporate passengers in these arrangements, and the resulting revenue value of those passengers to the airline, are different from the general case that has been described. These two peculiarities, while important to the industry, are not yet recognized by AirVM.

## 5.4: The Narratives of the Agents in AirVM

**5.4.1:** The conceptualization that underlies the development of agent models for application in virtual markets is the narrative framework presented in Chapter 3. That construct asserts that the behavior of an agent represents the behavior of a human being or institution as it plays out an internal 'story' or narrative. A narrative is a time-framed cause-and-effect chain of events that contains the structure by which the narrative owner understands the environment in which it finds itself. Events are situations where the narrative owner must make a choice which potentially changes course the narrative in future time, based on his available resources and desired choice outcomes. All events, and hence all choices, in a narrative are stochastic in nature, which is to say that selection of a choice is subject to uncertainty, and therefore it is the case that all choice events have an associated probability distributions. Moreover, choices require the utilization of resources, and resources available to the narrative owner can be applied to affect the probability distribution associated with the alternatives. The simplest narrative is an atomic narrative, with exactly one event, and more complex, molecular narratives can be built up of sequences of atomic narratives. Thus the initial stage in the design of a virtual market is the description of the narrative structures that agents in the simulation are going to have to represent.

**5.4.2:** The passenger is the central focus of AirVM. The entire economic structure of the commercial air travel industry rests on the passenger's decision to fly and the passenger's choice of air travel itinerary. Therefore, the agent that represents the passenger in AirVM must be based on an understanding of the narratives in which these two choice events are found. Both are atomic narratives, in that a single, stochastic choice opportunity is present. This agent is called the *pag* (rhymes with 'bag') for *p*assenger *ag*ent. (The other two agent types used in AirVM represent airlines and distribution channels, and are described below.) The decision to fly or not – the mode choice – is somewhat outside the realm of AirVM, since the demand for air travel is considered a given in the AirVM program (as an OD Demand Matrix, see Section 6.2 below). Of much more interest is the itinerary choice atomic narrative. That is, the pags in AirVM must emulate the atomic narrative that describes the choice of itinerary to make a specific trip.

**5.4.3:** It is tempting to delve into the motivations for travel that humans have demonstrated over their history on earth.[61] Travel as a way to expand one's horizons, meet new people, find new markets, discover new resources, explore one's history, see fanciful places, find needed suppliers, learn new concepts, express one's ideas, all are narratives that warrant closer definition. But, as is inevitably the case, boundaries must be drawn to determine what is germane to the construction of the AirVM simulation and what is not. Historically, within the industry, two classifications of trip purpose have been considered – business and leisure – and to the extent that any distinctions at all are made regarding the different characteristics of passengers, this has been the primary one. In spite of there being clear, albeit largely intuitive, distinctions among different types of business and personal travel, such as who pays, internal company business or meeting with customers, vacations vs. visiting distant family, etc., rarely have data been

[61] It would be interesting to explain the evolutionary basis that leads to the seemingly natural tendency for humans to want to travel. Such a dissertation would be a valuable contribution to the travel marketing literature.

collected that captures these important differences. Indeed, one is challenged to find even a consistent and generally accepted definition of business and leisure. Clearly trip purpose is an essential narrative property of travel, but little is available from which to construct the passenger narrative that surrounds the choice of how to travel.[62] So, as weak as the concept and available data may be, describing trip purpose is an essential, and available, narrative for the passenger.

**5.4.4:** The narrative of a passenger's travel is usually imbedded in a simple molecular narrative represents the journey structure. The simplest journey structure is a single flight from home to a specific destination, and then return home. Migratory flights are rare (relative to all commercial flights), and are not specifically considered by AirVM. More complex journey structures are made up of multi-stop flights, where the passenger moves from destination to destination before returning home. Modeling journey structure is an important area of air travel not yet well considered in AirVM. At the present time, all flights are considered in isolation from any other flights that may be occurring during the same journey.[63]

**5.4.5:** Another variable of the narrative is group size. People do not always travel alone. Very often, couples, families or work groups travel together. In this case the narrative that is driving the choice protocol is shared among the group, and very often the itinerary choice is a topic of discussion among members of the group. AirVM explicitly models group size. In fact, a passenger agent does not always represent one person, but can represent several. The other passengers in the group are using a consilvocation decision protocol. The itinerary choice then affect not one ticket from the available options, but one for each individual in the group represented by the pag.

**5.4.6:** It is the choice of itinerary to make a specified trip that is the atomic narrative which the pag must represent. What are the important attributes of the choice event? A consensus in the industry has been substantiated by a number of studies. See Proussaloglou and Koppelman (1999), Coldren *et al.* (2003), Parker (2004), Adler *et al.* (2005), Caldron and Koppelman (2005), Garrow and Parker (2005), Parker and Lonsdale (2005), Hess (2005), Garrow and Koppelman (2006), Bhat *et al.* (2006),

[62] This is not to say it cannot be done. Bunch (2003), in an unpublished work, delves into the economics of vacation travel to a considerable depth, creating an analytic structure of the activity pattern that surrounds a vacation trip. From this, significant improvement in the supporting narrative can be inferred.

[63] One of the difficulties lies in modeling multi-stop journeys. One approach might be using Markov chain processes. There is potentially sufficient data available within the industry to calibrate such a model, but I have not seen one reported in the literature.

Parker and Garrow (2006), and Parker and Walker (2007). That consensus has identified the following attributes as being of value to the prospective airline passenger:

**5.4.6.1:** The *cost* of the itinerary. Cost includes fares, taxes, excess baggage costs, on-board service costs, and whatever surcharges have been added to the price by carriers or other operators in the airline industry. Lower cost is, naturally, preferred.

**5.4.6.2:** The *duration* of the itinerary is important to most people, with a shorter flight being preferred to a longer one. Travel is hardly ever done for the sheer joy of riding on an airplane. In the early years of commercial flight, there was a novelty effect that had positive value to many. But now that most people took their first commercial flight when they were quite young, the novelty has long since worn off, and generally people want a flight to be as short as possible.

**5.4.6.3:** When does the flight *depart* and when does it *arrive*? As noted above, people travel for some other reason than to just ride on the plane, and those other reasons dictate not only the ideal departure and arrival times, but also which of the two time frames governs the itinerary choice. Business travelers generally want to arrive at the destination in time for some other activity, but no sooner than that activity requires, while leisure travelers often want to get to their vacation destinations as soon as they can to maximize the time spent there. In these two cases, the arrival time is dominant. On the other hand, a business traveler may want to get home as quickly as possible after the work is over, and thus is looking at itinerary departure times as most important. Clearly only one of the two times can be applied in a given trip, since one determines the other. Which time dominates in a particular passenger journey circumstance is called the time *sensitivity*. Furthermore, a passenger is considered to have an *ideal* departure and arrival time to which she compares available times in the evaluation of the alternative itineraries. By ideal time is meant the most suitable time for departure or arrival (depending on sensitivity) regardless of the scheduled service that is available.

**5.4.6.4:** The number of times the plane has to *stop* – land and then take off again – between when the passenger emplanes and when he arrives at his destination is also a concern. Stops involve landing, which is uncomfortable to some people, often getting off the plane, moving to another plane, and then boarding and taking off again, another activity which causes many people discomfort. Some people will also be concerned about the airport where the stop occurs, but there is little other than anecdotal data to clarify the role of this factor.

**5.4.6.5:** Airlines go to considerable lengths to encourage customer loyalty. Virtually every carrier in operation today has a *frequent flyer program* which rewards passengers for repeated use of their services. Other amenities, executive traveler lounges for example, are also tied to frequent carrier use.

**5.4.6.6:** The *cabin* on the airplane is also important to many individuals. Passengers seated in the first class or business class of an airplane have more room, more privacy, and better in-flight service, than those in the main cabin. But it universally costs more. Indeed, often much more. Business cabin fares on a transpacific flight can be four times the cost of an economy cabin fare, and a first class fare often ten times as much.

**5.4.6.7**: There are some features of the flying experience that most people care very little about. The particular aircraft – Boeing 737, Airbus A320, etc. – has little impact on the flight choice, although smaller aircraft, such as regional jets, are generally less preferred than larger aircraft, in part because turbulence is less dramatic on a larger plane. But these seem to have little bearing on itinerary choice, or at least research to date has not detected any noticeable effect.

**5.4.6.8**: One characteristic that is missing from these narratives is fear of flying itself. While there are certainly aspects of flying that frighten almost everyone, such as sudden turbulence or a bumpy take-off or landing, individual who are truly afraid to fly rarely, if ever, take a trip on an airplane. Most flight attendants are trained to help the fearful passenger through the ordeal, but such individuals are rarely seen. They craft their narratives which include travel to avoid airplanes as much as possible.

**5.4.7:** The passenger narratives that drive airline use then depend on trip purpose, the cost in money and time of the journey, the journey structure, the group size, and elements of the flight experience itself. These attributes are the basis for the mixed logit discrete choice protocol that is used by the passenger agents in AirVM, described in Chapter 6. The incidence distributions of the values associated with these narrative variables among the population of passenger agents are modeled using several different stochastic processes, and are delineated in Chapter 6, as are the details of the random utility discrete choice model itself. In Figure 5.3 these factors are summarized using the working definition of agent offered in Chapter 4.

**5.4.8:** The airline and the distribution system agents in AirVM are *client* agents. Client agents derive their narratives from the participation of individuals or institutions that are using AirVM in their decision and planning analysis operations. They are avatar agents. Therefore the agent design must offer to the client representations of actions that the client can take at the beginning of or during the course of a simulation. First consider the agent representing the airline. The general definition of this agent is illustrated in Figure 5.4. Only those aspects of the airline that are relevant to the pag fulfilling its primary choice event of selecting an itinerary and purchasing a ticket are relevant to the virtual market. These are ticket pricing and schedule development. Thus the airline agent in AirVM is called the *arasag*, which is shorthand for *a*irline *r*evenue *a*nd *s*cheduling *ag*ent. As noted earlier, prices are set for each fare class on each cabin on a flight in a dynamic fashion using a *revenue management (RM) protocol*. The RM protocol adjusts the fares as a function of the realized demand-to-date and/or some kind

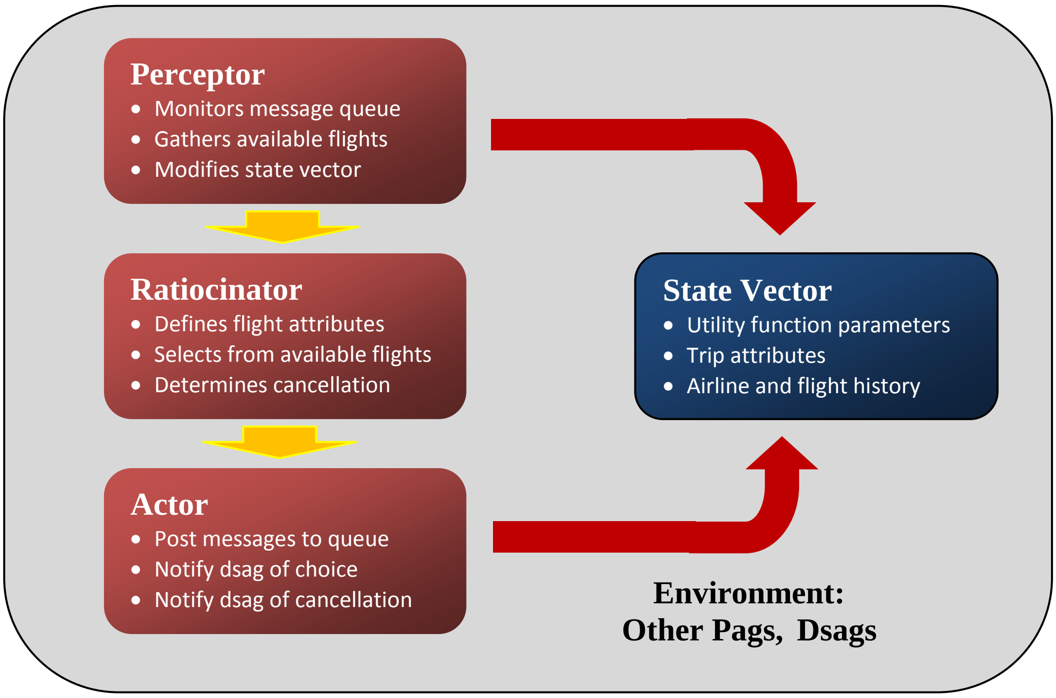



**Figure 5.3: The Abstract Definition of the Passenger Agent (Pag)**

of forecast of the final demand-at-departure, where the objective is to maximize revenue by raising prices for seats as they become less available during the ticketing period. AirVM allows any number of fare classes in each cabin on each flight, and each cabin can have its own protocol. At the present time, AirVM defaults to four protocols the client user can assign to any single flight or group of flights. Each has a set of associated parameters that also can be set by the client. The four available protocols are:

**5.4.8.1:** The *fixed* protocol. The prices in a fare class are fixed throughout the entire ticketing period, and are set by the user or by default using a price probability distribution, as described in **6.9.5**. This is the default protocol for all itineraries in the world when an AirVM scenario is initially defined.

**5.4.8.2:** The *days-before-departure* protocol. Here the price in a fare class increases at pre-defined days before departure as they occur in the simulation. For example, at 21 days before the flight leg is to depart, the price may increase in each fare class by 10%. This is equivalent to the protocol that closes a fare class to further ticketing as departure approaches, with unsold seats in the fare class being allocated to the next higher-priced fare class.

**5.4.8.3:** The *realized-demand* protocol. In this protocol, prices remain the same in a fare class until a pre-set number of tickets have been sold, then the price rises in response to the realized demand. The ticket sales level that triggers the price rise, and the amount of the price increase, is determined by the client.

**5.4.8.4:** The *forecast-demand* protocol. In this method, the number of seats in each fare class that are forecast to be sold is used to set the size of the fare class during the ticketing period, and the forecast is adjusted as ticket sales are realized. There are a number of variations of this approach, using different forecasting methods. Currently in AirVM only Expected Mean Seat Revenue (Belobaba, 1987) and linear forecasting methods are available. The client can set the size of and base fare for each fare class. The price in the fare class is sometimes referred to as the *bid price*, which is defined as the minimum price at which a seat will be sold in a fare class (Talluri and van Ryzin, 2004, pp. 34-35).

**5.4.9:** Revenue management activity – for example updating the protocol forecast parameters – occurs periodically during the ticketing period. Many modern RM systems update as often as every few hours, or even after every ticket sale. During the

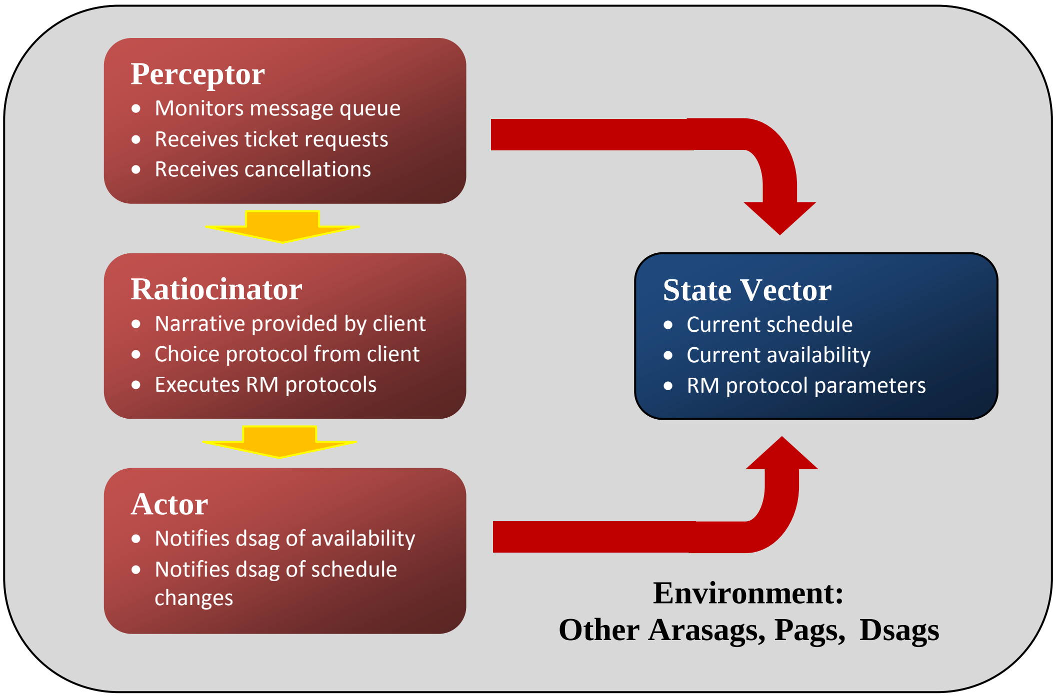



**Figure 5.4: The Abstract Definition of the Airline Revenue and Scheduling Agent (Arasag):** [Client Agent]

execution of an AirVM simulation the arasags monitor the message queues and will initiate RM updates on the schedule required by the protocol (which is under user control). Moreover, other revenue management protocols can be easily designed and implemented in the AirVM simulation. However, thus far in the development of the system this area has received little attention, since development resources have been prioritized to other features.

**5.4.10:** The other part of the arasag is maintenance of the airline schedule. This is managed by one or more teams in a carrier that develops and publishes the flight schedule. In AirVM, the client user has complete flexibility to alter the schedule associated with a particular scenario. An initial schedule is created (from published sources, such as OAG or Innovata), and the client can modify it to suit their needs. They can add flight segments, change the departure and arrival times, change airplane capacity (seats in each cabin), delete flight segments, for their own or any airline in the world. This general process is referred to as *schedule editing*. Schedule editing can cause the regeneration of the itinerary set.

**5.4.11:** The third agent in AirVM represents the ticket distribution systems. This also is a client agent, and so its narrative is supplied by a client/user during the simulation activity. This agent is called the *dsag*, for *d*istribution *s*ystem *ag*ent, and its basic structure is shown in Figure 5.5. Its state vector contains the information required to construct itineraries for the pag to choose from. The dsag handles the requests for tickets from the pags, transmits them on to the arasags, collects the availability as reported by the carriers, and manages the actual ticket purchase transaction. There can be more than one dsag in a simulation scenario. One dsag can represent the travel agent community while another represents direct sales by an airline, for example.

**5. 4.12:** The generation of the components of the itinerary sets for an AirVM scenario is a complex process. Of course all nonstop flights are itineraries between the two cities connected by the nonstop. One-stop flights are made up of connections between non-stops, where 'connection' implies that the arriving flight gets to the intermediate stop in

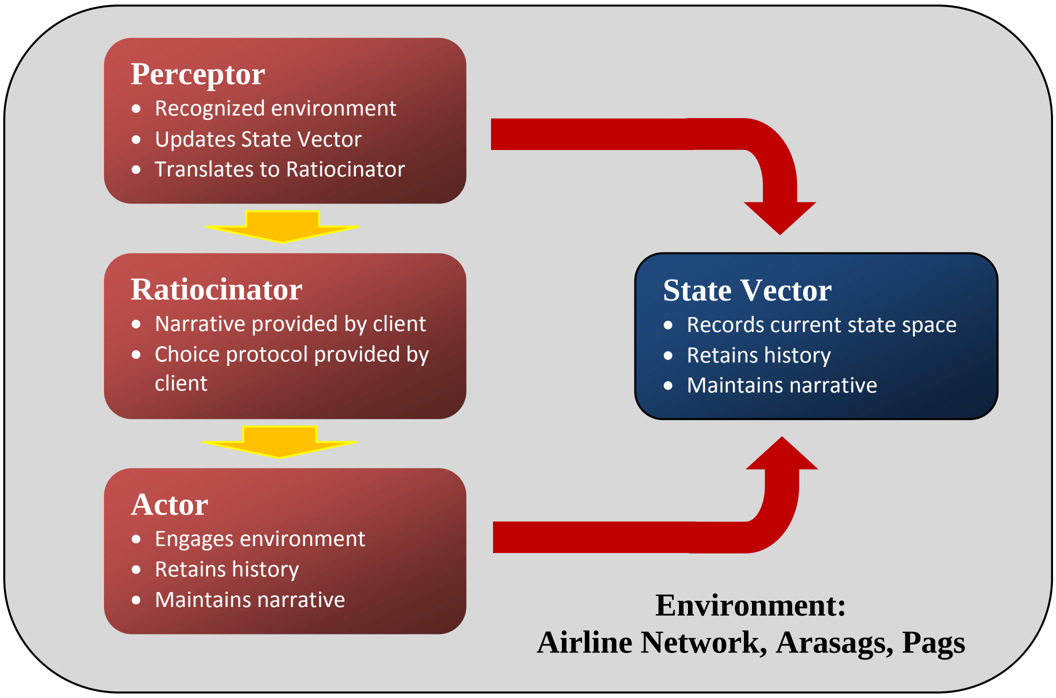



**Figure 5.5: The Abstract Definition of the Distribution System Agent (Dsag).**
[Client Agent]

time to deplane passengers and let those passengers board another flight, (called *connect time*). The number of stops in an itinerary can be called its *order*. Two-stop, three-stop, and so on itineraries are made up of combinations of itineraries of lower order.

One of the operational issues in both the AirVM simulation (and in the industry itself) is efficient ways of generating itineraries. A common method is *depth-first search* (also called Dijkstra's algorithm), where each route between an origin and destination is traced through all possible intermediate cities. [64] In practice, several million itineraries can be generated for each city pair with this method, so generally some form of *stopping rule* is invoked to halt the search, such as a maximum number of stops. Parker, Lonsdale *et al.* (2005) have proposed a passenger-utility based stopping rule. An alternative method is used for itinerary generation in AirVM, building itinerary components as part of the scenario creation process, and is managed by the dsag. It uses a modified breadth-first search process to create data structures of nonstop, one-stop, etc. itineraries, which are then used to build actual itineraries to be offered to the pag. See Section 6.11 for details of this methodology.

**5.4.13:** In summary, AirVM consists of three agents. The pag agent captures the atomic narrative of a passenger or group of passengers as they are faced with the choice of itinerary from the set of possible and available itineraries to meet a desired travel objective, expressed by specific origin destination demand. The other two agents, the arasag and dsag, are avatar agents, meaning their narratives are represented by individuals or institutions that supply the ratiocinators for the agents. These agents must therefore implement resources that the client can use to invoke the appropriate narrative interaction. For the arasag, these resources are ticket pricing (and revenue management) and schedule modification. For the dsag, they are itinerary generation mechanisms. The arasag capabilities are reasonably extensive in the current implementation of AirVM, but the dsag structure is quite elementary.

## 5.5: An Overview of the Simulation Logic

**5.5.1:** At this point in the discussion it is advisable to consider an overview of how the logic of the simulation proceeds. This will make later presentations easier to follow and keep the objectives of the simulation clearer as further details are explored. The general flow of the computing implementation of AirVM follows quite closely the description offered earlier, and is illustrated in Figure 5.6. But before a simulation can be executed,

---

[64] One calculation suggests that the approximately $10^{14}$ itineraries can be found to travel from London to New York. That's about 25,000 for every person on earth. Operationally, it would take more time than the age of the universe to date to calculate all possible itineraries for all possible city pairs in the world.

certain preparatory activities have to be completed. The synthetic populations of pag agents have to be created, and their state vectors populated with incidence distributions of data based on the characteristics of the individuals represented by those agents. Arasag and dsag synthetic populations are also created, but here there is fewer incidence distributions used since there is data on the exact state of existing airlines and distribution systems.

**5.5.2:** During the simulation, pags will purchase tickets in every OD market pair with scheduled service in the world. The pags are created to reflect the demand for travel from each origin city to the destination city in the market pair. If the total demand originating in city *A* is, say, *D(A)*, then the number of pags required to generate that number of passengers (recalling that an individual pag can purchase more than one ticket) is estimated, and that number of pags are generated. As part of the generation process, each pag is assigned a destination city, trip purpose, a group size, a journey structure, a maximum willingness-to-pay, a time sensitivity, an ideal departure or arrival time, and the parameters of the itinerary choice utility model it will use. They also are assigned a ticketing time based on the stochastic process that governs the ticketing process for that OD. (The ticket cancellation model currently in AirVM is the same for all pags, and only depends on the ticketing instant.) The simulator iterates through every city in the world, generating the required number of pags for all the markets for which that city has non-zero demand. With the OD demand data currently available for the execution of AirVM, about 47,000,000 pags are created in this manner. This is a random process, so if the pag generation process is repeated with a different seed for the random number generator, a different synthetic population is created. This establishes one of the bases for the AirVM to execute a Monte Carlo simulation.

**5.5.3:** Also before the actual simulation can run, the synthetic populations of the arasags and dsags must be created. Both of these agent sets are initialized based on the published schedules of airline service. There are two major sources of this data; OAG[65], Inc. and Innovata, Inc. The two companies are competitors, but both produce similar data, with only slight differences. Essentially, the schedule contains a record for each nonstop leg flown in the world for a given week. That record shows 1) the carrier that operates the flight, 2) the origin city, 3) the origin airport, 4) the destination city of the

[65] OAG was formerly called the Official Airline Guide, which is still the name of the printed schedule they publish.

leg, 5) the destination airport, 6) the local departure time, 7) the local day of departure of the flight, 8) the local arrival time, 9) the local day of arrival of the segment, 10) the day offset between the origin and destination, 11) the *code-share* relationship between the carriers that are represented on the flight, and 12) the equipment being flown, from which cabin configuration (the number of seats in each physical cabin on the airplane) is determined. All times and days are local, which means that the values are in reference to the local time at the respective origin or destination city. The day offset is a value that must be added or deducted from the difference in the departure and arrival times when adjusted for the time zone difference between the origin and destination, to account for crossing the international dateline or passing over midnight. Code-share refers to the practice of allowing more than one carrier to sell tickets on a given airplane. The flight owner is called the *operating carrier*, the others *code share partners*.

**5.5.4:** The schedule data used is for a standard time period, which is the week from Monday through Sunday. This time period is chosen because virtually across the world airline schedules repeat every week. Changes tend to occur slowly enough that there are weekly periods in every season where no change is observed. But there are enough changes to argue that several standard weeks should be prepared for each year, to reflect the major seasonal changes that occur in the schedule. This is of no consequence to AirVM, however, since the only impact on its operation is to require different seed scenarios. (Seed scenarios are discussed in more detail below.) The weekly schedule also solves the so-called *wrap* problem. Many flights do not fly every day of the week. For a long time there was daily service from Seattle to London, and, in addition, an extra flight (departing a few hours later) was offered four times a week (Sunday, Monday, Wednesday and Friday). That means that someone who had an ideal departure time on Saturday had some useful itineraries on the next Monday as well as the previous Friday. This situation is handled by letting the itinerary generator "wrap" around the end of the week into the next week. In other words, the system will search for itineraries three days before and three days after a desired departure date. The standard week enables this capability.

**5.5.5:** Before the schedule can be used, however, it must be turned into an itinerary set. When the schedule data is loaded into the system, the individual flights are assigned to the state vectors of the respective arasags. The dsags then use these individual flights to create sets of itineraries for those markets that have non- and one-stop itineraries. These itinerary sets include all possible itineraries connecting the origin city with the destination city. It is no more than a combinatorial problem. If a market is not served by this one-stop or better service, then a single two-, three- and four-stop itinerary set is produced for that market. These higher order itinerary sets only contain zero and first order flights with the highest utility. The itinerary set for a given schedule, along with the OD demand matrix that applies to that set, is called a *scenario*. The client who is represented by an arasag in the simulation can alter the contents of a scenario – change the schedule or pricing structure – and study the impacts of the changes by comparing the results of the altered scenario to the original scenario.

**5.5.6:** The overall conceptual flow of the actual simulation is illustrated in Figure 5.6. Ticketing begins 127 days before the last departure day of the standard week. For each time point (say each minute) prior to midnight on the seventh day of the standard week, a randomly chosen origin city is selected. Then, a random pag is selected from the synthetic population from that city, and the destination of that pag determined. Pag processing then consists of first determining if the pag has ticketed, and if so, the cancellation process is executed. If not, the pag requests ticket availability, receives a list of available options, chooses one, and purchases the ticket. Control then moves on to arasag processing. If an RM update is due, (which is by default once per day, but can be changed), then that is executed. This repeats until all the pags due to purchase tickets for that time period in all of the origin cities have been processed, and then the simulation moves on to the next time period. When all pags have been ticketed or their tickets canceled, the simulation execution is complete and the results are produced. Pags for which tickets have been canceled are not allowed to return to the system to purchase tickets again. The implementation of this logic is discussed in detail in Chapter 7.

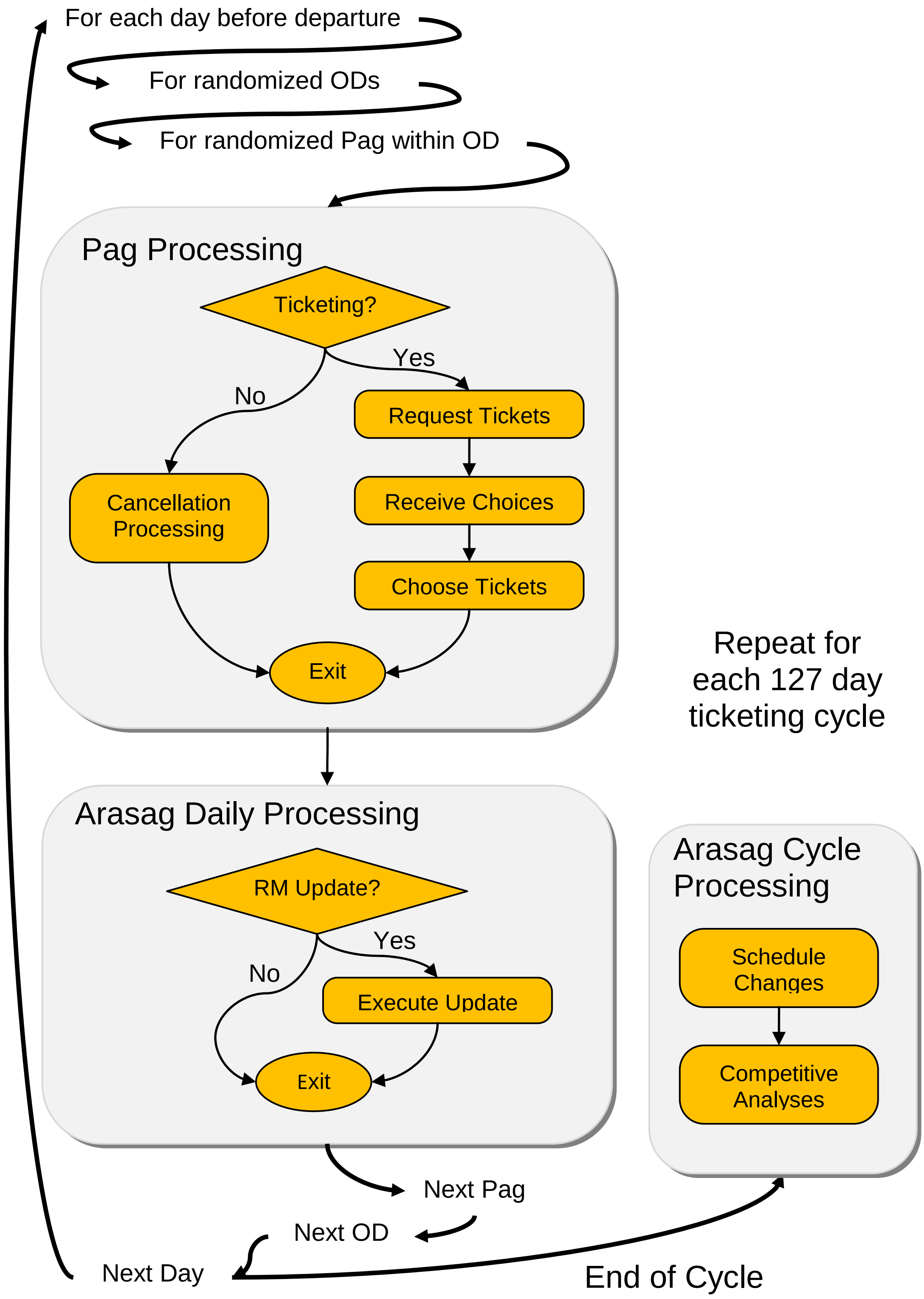



**Figure 5.6: General AirVM Logic Flow**

**5.5.7:** AirVM coordinates and manages the actions of the agents by means of what has been referred to as the messaging architecture, discussed in **4.4.6**. Basically, all the agents in the simulation communicate with one another by means of passing messages through a message queue. Messages addressed to other agents are posted there, and agents read messages from the queues when they have time available. This allows agents to operate independently of one another while maintaining coordinated action. This process of selection and booking is managed in AirVM by employing the messaging features of Microsoft's .NET platform. This service allows independent processing threads (possibly on different computers) to communicate with one another by passing messages through the queue maintained by a designated processor. This message flow is illustrated in Figure 5.7. The main messages are:

**5.5.7.1:** A pag requests one or more seats (depending on group size) on an itinerary from City A to City B from the dsag. The pag supplies how much it is willing to pay per ticket, and its ideal departure or arrival time.

**5.5.7.2:** The dsag takes the pag's request, identifies the arasags that have service in the market, and requests from those agents seats in fare classes that have space and that the pag can afford. There are usually far more itineraries available than a pag can consider, so a limitation on the total number of options returned by the dsag, called the *scope*, can be set by the user.

**5.5.7.3:** The arasags check availability and return messages to the dsag indicating what is available, including departure and arrival times, fare, cabin, and so on. Seats that are offered are held out from further availability until the pag makes its choice.

**5.5.7.4:** The dsag assembles the responses from the arasags, and submits them as one packet to the pag.

**5.5.7.5:** The pag invokes its choice protocol to select one of the available itineraries, and notifies the dsag of which option has been chosen.

**5.5.7.6:** The dsag notifies the arasags of either what the choice was, so seats can be removed from inventory and the revenue credited, or that their alternatives were not chosen, so the seats can be freed up for subsequent pag ticketing events.

There are additional messages that handle pag ticket cancelation, RM protocol execution, and other aspects of the processing. Because of the independent and asynchronous nature of the messaging structure, the computing architecture of AirVM allows implementation as a set of independent processes operating in different threads on one or more separate machines. This makes the simulation scalable while maintaining reasonable operating characteristics.

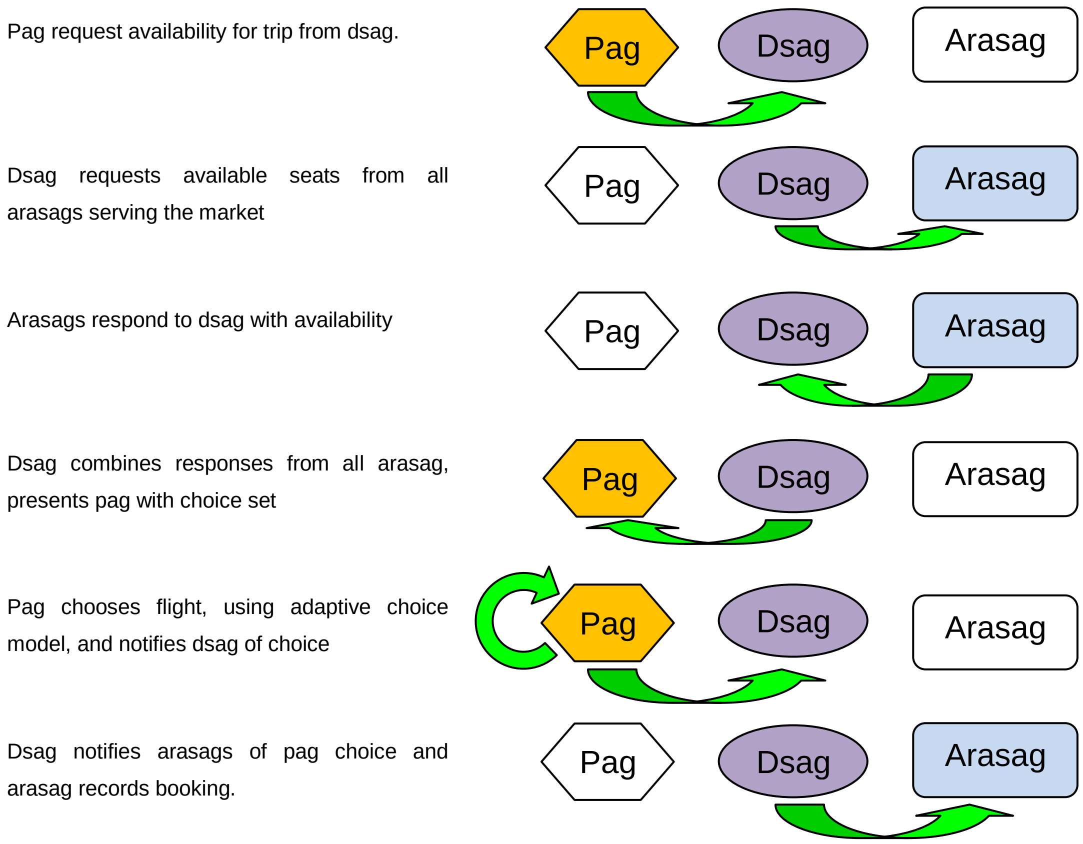



**Figure 5.7: The AirVM Agent Message Flow**

**5.5.8:** How the pag uses the choice protocol to make the selection is typical of how probabilities enter into a virtual market. The pag uses its discrete choice model to compute the probability of selecting each of the offered itineraries. Conceptually, one can think of the construction of a roulette wheel with wedges that correspond in width to the probability of each available itinerary. The wheel is spun, and the winning itinerary is the one that collects the white ball. More prosaically, the options are put in some arbitrary order, and the probability distribution function of the available choices created. Of course, this is a discrete probability distribution, so it consists of discrete jumps equal to the probability of each choice. Then a random number generator produces a uniformly distributed random number between zero and one, and the choice corresponding to that number found by searching the cumulative discrete distribution function. The random numbers for the simulation are created using a Marsenne Twister random number generator. The algorithms used for the number generation are modeled after those developed by Troscheutz (2007), and are shown in Appendix N. The seed of the generator can be set by the user to allow exact duplication of simulation runs.

**5.5.9:** With this summary of the air passenger market in mind, attention will now turn to the details of the incidence distributions that apply to the creation of the synthetic populations of pags, arasags and dsags in the AirVM virtual market. There is considerable effort required to build these populations in a realistic manner, each relying heavily on individual research activities into the occurrence of those features in the narratives that drive the behavior of the agents in the market of interest. In this fashion a virtual market effectively ties together several strands of market research into a unified structure, often highlighting important relationships and missing data. This will become quite evident in the discussion in the next chapter.

# Chapter 6:

# AirVM Synthetic Populations and Incidence Distributions

## 6.1: The Incidence Distributions

**6.1.1:** The reader will recall that the incidence distributions are the stochastic processes that describe the data that are contained in the state vectors of the agents in a virtual market. Generally, as is the case with AirVM, these distributions are intended to be representative of the incidence of the respective variables in the market population of interest, although that is not necessarily always the case. This chapter describes the models, data sources and estimation methods that support the incident distributions for the agents in AirVM.

**6.1.2:** Some incidence distributions are trivial, merely the presences or absence of a property or one of just small set of discrete possibilities. In AirVM, trip purpose (business or leisure) is a two-valued state, and its stochastic structure is a Bernoulli distribution. Also, journey length can take on only four values – less than seven days, seven to ten days, eleven to twenty days, and longer than twenty days – and the associated probability distribution is simply the four probabilities applicable to each (or three, since the fourth is implied by knowledge of the other three). But many incidence variables are governed by significantly more complex stochastic structures, and so require a more complete explication.

**6.1.3:** Most of the incidence distributions are concerned with the pags, since these agents represent the primary consumer in the passenger air travel market. Two are simple binary variables; trip purpose as noted above and arrival/departure time sensitivity. Journey structure has four possible values, while the desired day-of-travel distribution has seven, one for each day of the week. Seven sets of pag state vector data are governed by somewhat more complicated structures. These include 1) the probability that a given pag will want to travel from a given origin to a given destination (the OD demand); 2) the parameters of the random utility itinerary choice model with which the pags selects a ticket; 3) the distribution of passenger ideal departure and arrival times; 4) the distribution of ticketing group size; 5) the stochastic process that governs the ticketing instance as booking approaches flight departure date; 6) the parameters of the stochastic process that determines if a purchased ticket is cancelled before departure, and 7) the fly/no fly choice model that determines willingness-to-pay for a pag. Each of these stochastic processes are discussed in depth in the following sections.

**6.1.4:** The arasags and dsags have many fewer associated incidence distributions, in part because they are relatively simpler agents and partly because they can be client agents (avatars), and so inherit much of the state vector from the client herself. The most important incident distribution is the fare structure for a market, which is examined in detail below. The other state vector structures are not distributions at all, since they are literal descriptions of the current state, such as the set of flight legs in the world, which is fully known. For the dsags, the most important state vector component is that which governs itinerary generation, necessary for the management of the dynamics of the ticketing process over the time frame of the simulation.

## 6.2: The Estimation of Origin-Destination (OD) Demand

**6.2.1:** The number of passengers that want to travel from a given origin city to a specified destination city – *OD demand* – is an important input variable to AirVM. In the airline industry, an origin-destination pair is commonly denoted as an *OD market*. The set of OD markets is presented as a matrix where the row of the matrix is the origin, the column the destination, and with diagonal elements of zero. The demand number used by AirVM is for the standard week. For example, if the OD demand matrix contains an entry of 845 representing the origin Sydney to the destination Townsville, then 845 passengers are expected to want to buy tickets on flights from Sydney to Townsville in the standard week. Currently, the OD demand matrix for AirVM is generated by a third party, for example the PaxIS data product from IATA (International Air Transport Association, 2007), although there are a few other sources available.[66]

[66] The Airline Reporting Corporation is preparing its own version of an OD Demand forecast as of this writing, an effort to which I contributed in a small way. Other vendors in the industry, Sabre Holdings for example, are also bringing similar products to market. There are significant problems with some of the approaches taken in the development of these products.

**6.2.2:** Demand modeling is an important activity in transportation planning, and has received substantial attention in the literature. Brown and Watkins(1968), Verleger (1972), Jung and Fuji (1976), Levine (1987), Cline *et al.* (1998), Wang and Pitfield (1999), Profillides (2000), Abed (2001), Brons *et al.* (2002), Transport Canada (2002), U. S. Transportation Research Board (2002), Bahdra (2003), Gillen *et al.* (2004), and Berry *et al.* (2006) provide various historical perspectives on the efforts to estimate demand in the airline industry.

**6.2.3:** There are a huge number of OD markets in the current world's airline network, over 11,000,000 possible city pairs in August of 2008. (These and the following network statistics are derived from OD data provided by IATA through The Boeing Company, and used by AirVM.) The vast majority of them have no demand at all. Of these markets more than 244,000 that show a demand of at least one over the period of a year, (equal to approximately 0.02 passengers per week), fully 80% have demand less than 50 per year, or one per week. Figure 6.1a and 6.1b show the relationship between the average number of weekly passengers in an OD market and the rank of that market (6.1b is the log base 10 of the values shown in 6.1a, which better illustrates the behavior at the scree point[67] of the sharp power curve). As can be seen, the vast majority of passengers travel in a very small number of markets. The biggest 0.5% of markets carry 43.5% of world's average weekly demand, and the top 1% of about 54.6%. Ninety-five percent of all demand is carried in 20.4%, and the bottom 43.1% carry less than 1.0% of the world's OD demand. The biggest market in the world (in the August 2008 scenario) is Los Angeles to San Francisco, LAX>SFO[68], with 129,147 average passengers per week, or 18,459 per day. As Figure 6.1 shows, market size is represented by a power curve. The equation for that power curve is given by this formula:

$$y = 5,021,263.37x^{-1.68095}. \tag{6.1}$$

Here, $y$ is the average weekly demand, and $x$ is the rank of the OD market in the list of world markets. The $R^2$ value for this equation is 0.97777. This equation form, with

[67] The term "scree point" refers to the area where the graph takes the sharp turn near the bottom left. *Scree* is a common term for the debris found at the bottom of a cliff.

[68] I will frequently use the notation XXX>YYY to indicate an OD market originating in city XXX and terminating in city YYY, where XXX and YYY are industry standard city three-letter codes.

rank position as the *x* variable, is frequently used as measure of network connectedness (see Bonabeau, (2002) for a discussion). Note that *x* is an ordinal variable, so different sets of OD's will produce different parameter values in equation 6.1.

**6.2.4:** But the quality of the observed demand data is a serious issue within the airline industry. There are essentially three sources of observed air travel demand data that are available for use in developing and calibrating econometric models of passenger demand, the carriers, the ticketing channels, and the government. Each has significant shortcomings. Carriers, of course, have records of who booked seats on their flights,

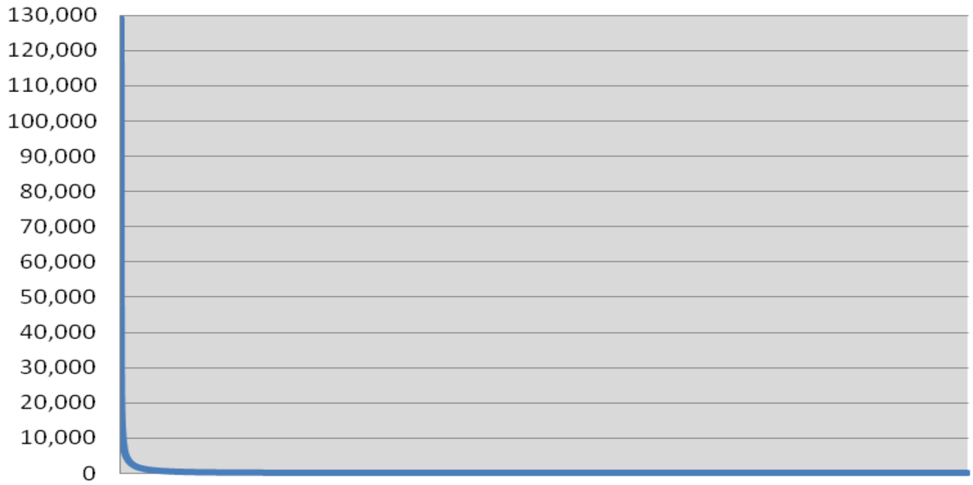



**Figure 6.1a: Market Size by Rank**

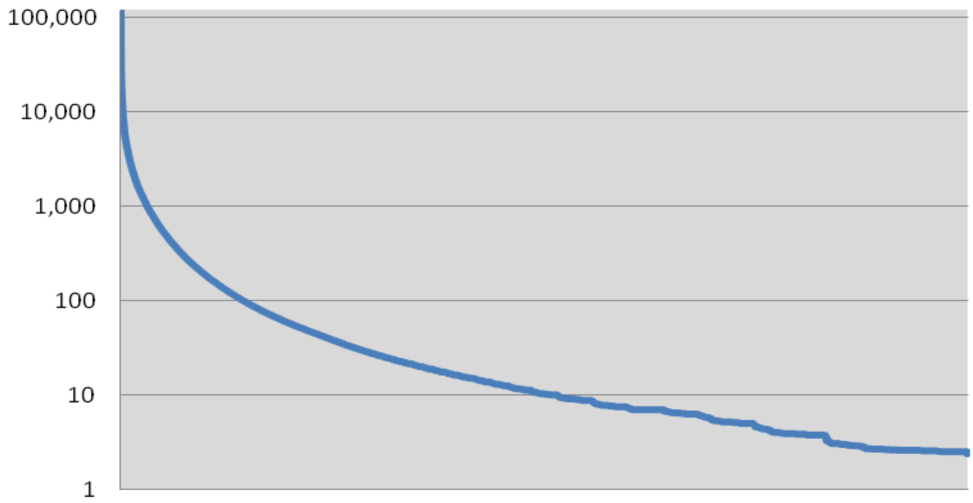



**Figure 6.1b: $\text{Log}_{10}$ Market Size by Rank**

**Figure 6.1: Distribution of OD Market Size** (Source: AirVM Aug 2008 Seed Scenario)

and so can determine accurate numbers of tickets sold on any flight they control, including the markets the passengers are flying in (if they can see the entire passenger journey structure, so they can determine if the flight they operate is part of a larger itinerary containing segments they do not operate.) But they see no accurate data for competitors, and thus have any no direct data on OD demand for any markets in which they have competition. The ticketing channels, in particular reservation services, commonly called *GDS*s (for *G*lobal *D*istribution *S*ystems), and the *BSP*s (which stands for *B*ank *S*ettlement *P*lans) can see a great number of tickets sold on a great number of markets, but only about half of all the tickets sold in a particular market. That is because these organizations traditionally handle sales through travel agencies, and they cannot see the tickets sold through other channels, such as those purchase directly from the airline or over the internet. Government departments occasionally have visibility of all channels and all markets within their jurisdiction, but availability and data definition details vary sharply from country to country. Appendix L provides more depth on these classic methods of demand estimation.

**6.2.5:** In addition, to compound the demand estimation problem, the data that does exist demonstrates the extreme variability of the observed demand values for a given market day after day. Figure 6.2 is a graph of the daily observed demand on all carriers reported by the Airline Recording Corporation (ARC) for the market from Miami, Florida, to Seattle, Washington, USA (MIA>SEA) for the calendar year 2005 (ARC, 2006). The degree of variation is dramatic. Close examination of the figure suggests that there might be a consistent weekly pattern (which is reasonable given the normal business week activity pattern), but as shown in Figure 6.3, significant variability within the same day-of-week is evident. Examination of a histogram of the demand observation shows a decidedly skewed distribution (Figure 6.4). The assertion is often made in the industry (Swan, 1996) that the observed variation in demand can be represented with a normal distribution, an assumption belied by the analysis of a significant number of markets similar to those illustrated here. However, in AirVM, assuming sufficient data is available, the unique variation pattern of a particular OD market can be captured using an empirical distribution function based on actual observations. In addition, AirVM itself can be used to explore the relationship between underlying demand distributions and the resulting demand data observed on the network, as will be seen.

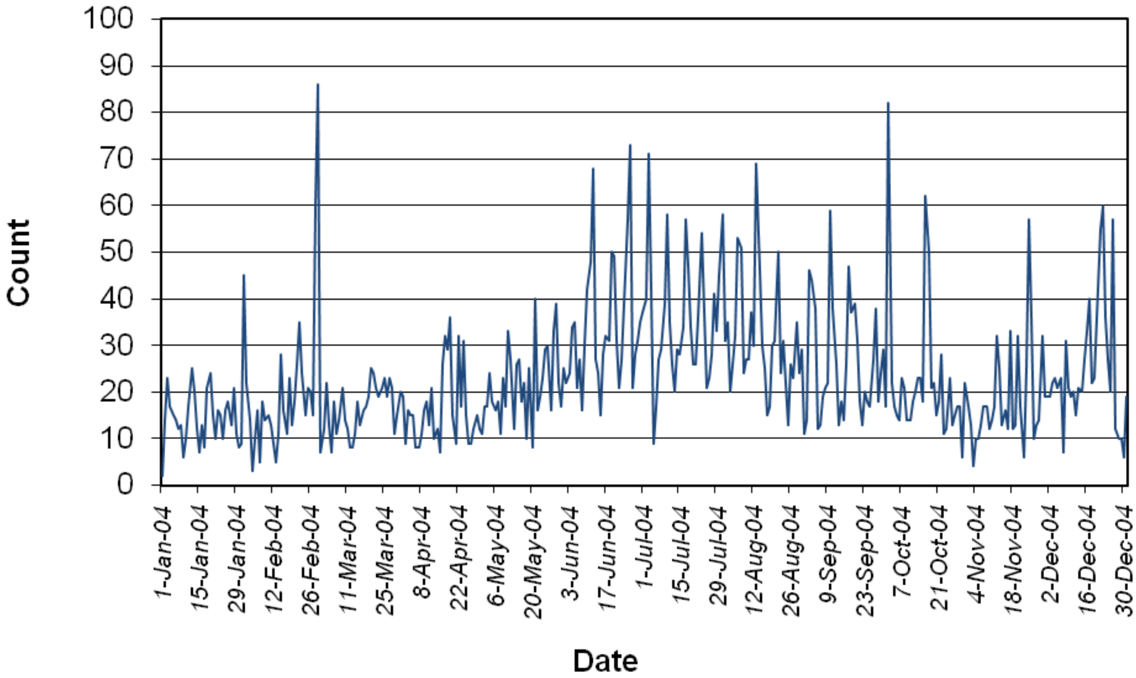



**Figure 6.2: Daily Ticket Sales, MIA>SEA Market.** (Source: ARC, 2006)

**6.2.6:** Because of these deficiencies, vendors of demand data such as IATA, with their Pax-IS product (IATA, 2009), ARC with their data product called COMPASS (ARC, 2009), and Sabre Holdings with their demand data product (Sabre, 2009) use various heuristics and data comparison methods to create demand estimates that are consistent with observed data while accounting for the missing pieces. But these approaches are weak at best, and are impossible to verify or calibrate. Carson *et al.* (2010) have studied the demand problem from a classical transportation planning perspective, developing econometric trip generation models and trip distribution models from the US Department of Transportation data bases, which are probably the best and most consistent available. Appendix M contains a description of an econometric model derived from the Carson work but using AirVM results as an impedance measure. The interested reader can get a good sense of how this classic approach works from that material. Yet, without better quality of observations of demand, extension to other parts of the world, or even refined calibration of these sophisticated econometric models, is suspect.

**6.2.7:** AirVM, viewed from the role that an agent-based model can take as an instrument for the more acute observation, also can provide a method of improving the observed demand data from which econometric models can be estimated. The method is largely untested, and the specific problem is outside the scope of this dissertation, so a description of the approach is relegated to Appendix M. Meanwhile, AirVM uses demand data supplied by others. Currently, that demand data is based on IATA's PaxIS, and Virtual M1nds is working with other data suppliers to refine the OD matrix.

## 6.3: The Pag Itinerary Choice Protocol Model

**6.3.1:** Central to the valid operation of AirVM is the pag itinerary choice protocol used in the simulation. It is a mixed logit with random coefficients model. Each pag in the simulation actually has four itinerary choice utility functions associated with it. The four versions arise due to the combination of trip purpose (business or leisure) and time sensitivity (arrival or departure). They all are of the same form, only differing in the

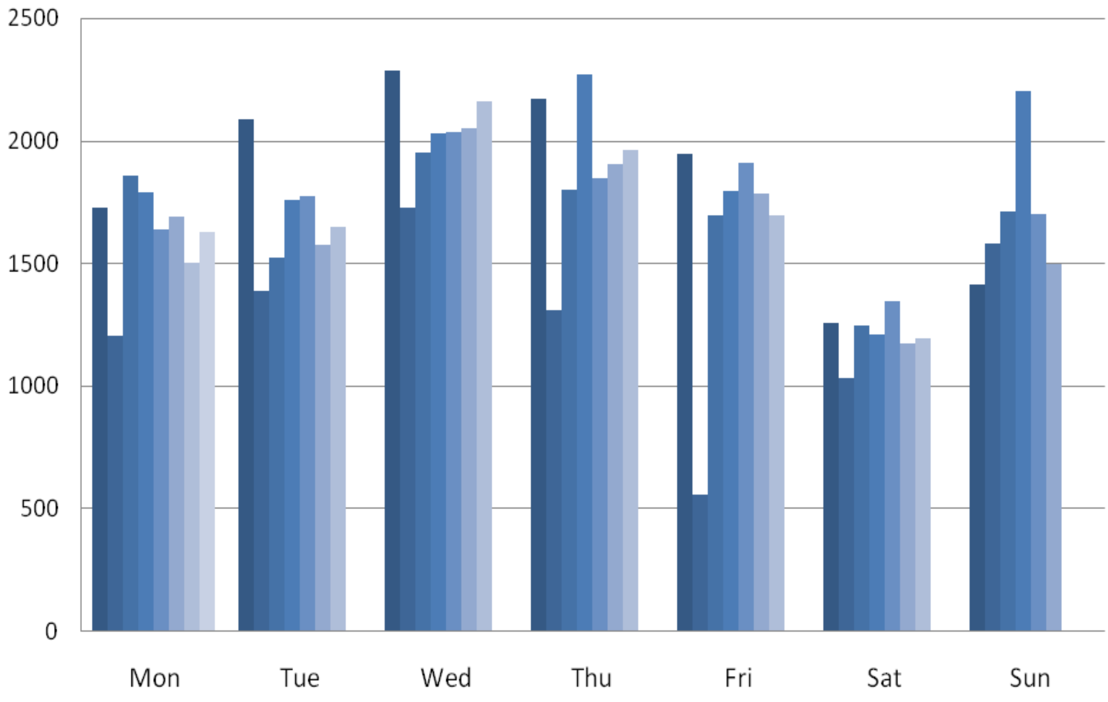



**Figure 6.3: Day-of-Week Variation, MIA>SEA.** (Source: ARC, 2005)

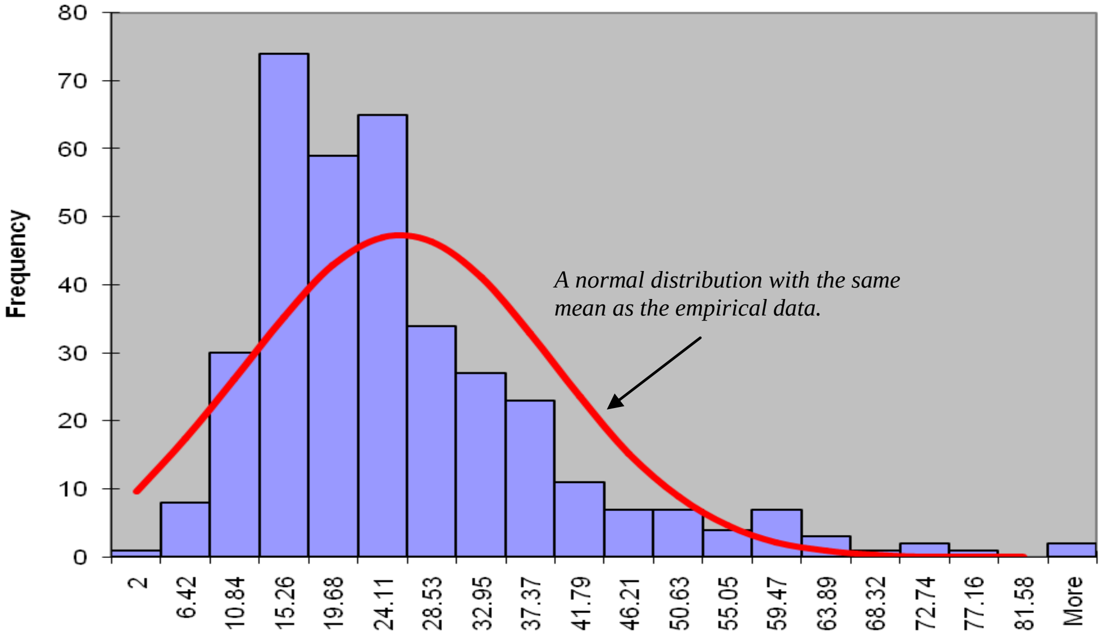



**Figure 6.4: Histogram of MIA>SEA Demand.** (Source: ARC, 2006)

empirical coefficients. For example, the fare coefficient ($\beta_f(i)$ below) is generally more negative for leisure than business travel for the same pag, reflecting the higher sensitivity to fare of an individual compared to a business. In addition, depending on journey structure, a pag can be arrival time sensitive or departure time sensitive. A pag may consider arrival time more important on an outbound business trip, or alternatively the departure time may be more important returning home.

**6.3.2:** The general form of the utility function for the pag choice protocol is as follows. Consider any market *m* for which air travel demand is not zero over the standard time period (one week), and denote the set of available itinerary fare classes in *m* by $\Phi(m)$. In the general random utility formulation, it is asserted that the decision-maker (here a pag), denoted by *i*, has before it a finite set of choices for each of which that decision-maker has determined a real number called a utility. The utility $U(i,j)$ for pag *i* and that alternative *j* is modeled as the sum of an observed component $V(i,j)$ and an unobserved component $\varepsilon(i,j)$. That is,

$$U(i,j) = V(i,j) + \varepsilon(i,j). \tag{6.2}$$

The term *V(i,j)* is termed *observable* because experiments can be conducted which allow the estimation of the utility as a function of attributes of the alternative and characteristics of the decision-maker. The unobserved term is, however, a random variable and represents that aspect of the decision process that cannot be observed by the experimenter. It is assumed here, in keeping with the underlying random utility conceptualization, that $\varepsilon(i,j)$ is independently and identically distributed as an Extreme Value Type 1 (EV1) distribution with mean zero and standard deviation one for all *i* and *j*. See Ben-Akiva and Lerman (1985), Anderson, de Palma and Thisse (1992), Train (2003) and Louvierre, Hensher and Swait (2000) for more thorough discussions of random utility theory. Then, for pag *i*, the utility $V(i,j)$ for that pag and for itinerary fare class $j \in \Phi(m)$ is defined by the following utility equation:

$$\begin{aligned} V(i,j) = {} & \beta_f(i)\ln f(j) + d(j)[\beta_d(i) + \beta_{bd}(i)\ln d_{base}] \\ & +\beta_{dc}(i)N_{dc}(j) + \beta_{ic}(i)N_{ic}(j) \\ & +\beta_{1st}(i)X_{1st}(j) + \beta_{ec}(i)X_{ec}(j) + G(\tau(i) - t(j)) \\ & +\sum_{a\in\Psi}\left\{I(a)[\beta_\Psi(i,a) + \beta_F(i)F(i,a)]\right\} \end{aligned} \tag{6.3}$$

**6.3.3:** The quantities in the first seven terms are defined below. In all cases the coefficients are assumed to be random quantities distributed across the population of pags according to a normal or truncated normal probability distribution.

**6.3.3.1:** $\beta_f(i)$ is an empirically-estimated coefficient giving the relative weight of fare in the utility function for pag *i*. It is assumed to be distributed $N_{T^+0}(\mu_f, \sigma_f)$, where $\mu_f < 0$ and the subscript $T^+0$ indicates that the distribution is truncated above zero. That is, all values are less than or equal to zero. Each pag has a value randomly drawn from this distribution. The truncation ensures that no pag has a positive utility associated with fare, which would imply that paying more is better. *f(j)* is the fare (in 2008 US dollars) of the itinerary fare class *j*, and ln *f(j)* is the natural log of that fare. Natural log is a better fit to the data, reflecting the fact that the impact of a $100 fare increase on a base fare of $1000 is perceived differently than on a base fare of $100.

**6.3.3.2:** $\beta_d\,(i)$ is the empirical coefficient giving the weight of itinerary duration in the utility function for *i,* where *d(j)* is the duration of itinerary fare class *j*. It is also assumed to be distributed $N_{T^+0}(\mu_d, \sigma_d),\ \mu_d < 0.$ Each pag has a unique value drawn from this distribution. Once again the model assumes that a longer itinerary has less utility than a shorter one.

**6.3.3.3:** $\beta_{bd}(i)$ is the empirical coefficient giving the weight of the base (shortest) duration of all the itineraries in the set $\Phi(m)$, denoted $d_{base}$. This coefficient is also considered a random variable assumed to be distributed $N_{T^+0}(\mu_{bd}, \sigma_{bd}),\ \mu_{bd} < 0,$ and from which a value is drawn for each pag. The base duration arises in the formulation because the pag is considered to compare a given itinerary with the best available itinerary, which, *ceteris parabus*, is the alternative with the shortest travel time.

**6.3.3.4:** $N_{dc}(j)$ is the number of direct connections in the itinerary. $\beta_{dc}(i)$ is the empirical coefficient measuring the penalty of a direct (or online, between aircraft of the same airline or airlines in the same alliance) connections in the utility function for *i*. It is distributed $N_{T^+0}(\mu_{dc}, \sigma_{dc}),\ \mu_{dc} < 0.$ Again, there is less utility associated with more stops.

**6.3.3.5:** $N_{ic}(j)$ is the number of interline connections in the itinerary, and $\beta_{ic}(i)$ is the coefficient giving the penalty of indirect (or interline, between aircraft of the different airlines) connections in the in the utility function for *i*. It is distributed $N_{T^+0}(\mu_{ic}, \sigma_{ic}),\ \mu_{ic} < 0,$ with more stops decreasing utility.

**6.3.3.6:** $X_{1st}(j)$ is a dummy variable equal to one if the itinerary fare class uses the first class cabin on the aircraft, and $\beta_{1st}(i)$ is the coefficient giving the value of the first class cabin in the utility function for *i*. It is assumed to be distributed $N_{T^-0}(\mu_{1st}, \sigma_{1st}),\ \mu_{1st} > 0.$ Note that this distribution is truncated below zero, indicating that only positive values are allowed. Thus the first class cabin is better than other cabins.

**6.3.3.7:** $X_{ec}(j)$ is a dummy variable equal to one if the fare class uses the main cabin $X_{ec}(j)$ on the aircraft. If both $X_{1st}(j)$ and $X_{ec}(j)$ are zero, then the business class cabin is assumed (the business cabin is the reference value for the indicator variables $X_{1st}(j)$ and $X_{ec}(j)$). $\beta_{ec}(i)$ is an empirical coefficient expressing the weight of the economy cabin in the utility function for *i*. The distribution individual pag values are drawn form is assumed to be distributed $N_{T^+0}(\mu_{ec},\sigma_{ec}),\ \mu_{ec} < 0.$ The economy cabin is always of less utility than the reference, which is the business class cabin.

**6.3.4:** Recall that a pag can be either departure or arrival time sensitive. As above, each time utility function has the same form, but different parameters. The function $G(\tau(i) - t(j))$ defines the time-of-day utility structure, and is given by a pair of Box-Cox transformations, one for early times and one for late times, which surround an interval of time about the ideal time called the *indifference window*, within which the pag doesn't have any time-related disutility. Specifically, the function $G$ is defined as

$$G(\tau(i)-t(j))=\begin{cases} \beta_E^G(i)\dfrac{(t(j)-\tau(i)-a+1)^{\lambda_E}-1}{\lambda_E} & \tau(i)-t(j)<-a \\ 0 & -a<\tau(i)-t(j)<b. \\ \beta_L^G(i)\dfrac{(\tau(i)-t(j)-b+1)^{\lambda_L}-1}{\lambda_L} & \tau(i)-t(j)>b \end{cases} \tag{6.4}$$

Here, the following quantities are used.

**6.3.4.1:** $\tau(i)$ is the ideal (departure/arrival) time for pag *i*.

**6.3.4.2:** $\beta_E^G(i)$ and $\beta_L^G(i)$ are the empirical coefficients for disutility associated with deviations from ideal departure/arrival times for early and late, respectively, itinerary departure/arrival times. They are each distributed with a normal distribution truncated above zero, as above, e.g., $N_{T^+0}(\mu_L^G,\sigma_L^G),\ \mu_L^G < 0,$ since deviations from ideal is always assumed to have non-positive utility.

**6.3.4.3:** $t(j)$ the departure time of itinerary fare class *j*.

**6.3.4.4:** $\lambda_E$ and $\lambda_L$ are empirically derived parameters characterizing the Box-Cox representation of the disutility curves for early itinerary departure/arrival times, respectively.

**6.3.4.5:** *a* and *b* are the bounds of indifference window within which the pag is indifferent to the itinerary departure/arrival time.

This specification thus stipulates that the disutility of not departing at the desired departure time (or, alternatively, arriving at the desired time) differs depending on if the actual departure/arrival time is before the desired (early) or after the desired (late). The utility curve is therefore shaped like that shown in Figure 6.5. These curves are fitted with a general purpose Box-Cox transformation from empirical data, which seems to describe the available data reasonably well. The indifference window (the flat line between *a* and *b*) reflects the fact that to some degree the passenger doesn't care if a flight is early or late.[69] This expression and its derivation and estimation is described in detail in Parker and Walker (2007). The distribution of the ideal departure and arrival times is described in Section 6.4**.**

**6.3.5:** The final terms in the utility shown in Equation 6.3 depend explicitly on which airline is operating itinerary fare class *j*. Let $\Psi$ be the set of all airlines operating in the market in question, and let the function *I(A,j)* be an indicator variable which is one if $A \in \Psi$ operates itinerary *j*, and zero otherwise. The term *F(i,A)* represents the frequent flyer mileage pag *i* has with airline *a*, with its associated $\beta_F(i)$ coefficient (also assumed to be a truncated normal random variable). The term $\beta_\Psi(i, A)$ is a constant

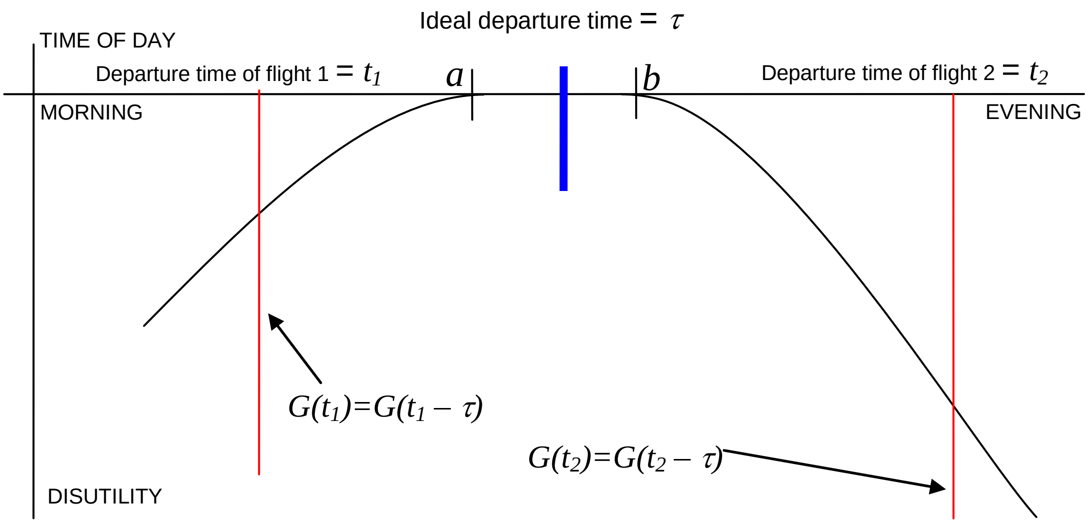



**Figure 6.5: Ideal Time Disutility Curve**

[69] This corresponds roughly to the window conceptualized in the Boeing Decision Window Model of Passenger Choice. See Baseler (2002) for a discussion.

associated with airline *A* by pag *i*, and represents the value the pag puts on airline *A*, while the frequent flyer variable reflects the adaptation of pag *i* with respect to its flying experience on airline *A*, in that increasing flying experiences will improve the likelihood that an itinerary from that airline will be chosen again. This term is not yet implemented in the model, since no data yet exists that allows the estimation of the $\beta_{\Psi}(i,a)$ or $\beta_F(i)$ coefficients. There is a good deal of research yet to be done on the relationship between a passenger's itinerary preference and the airline operating the itinerary. Carriers have mixed opinions on the importance of cabin attendee attitude, on-time performance, cabin cleanliness, and other features of the flight experience directly under the control of the carrier. Incorporating such measures into a discrete choice model is also difficult, although item response theory through Rasch scaling (Fisher and Molenaar, 1995, van der Linden and Hambleton, 1997, Fox and Bond, 2007, and von Davier and Carstensen, 2007) holds great promise in this regard.[70]

**6.3.6:** Under this utility function structure, the probabilities associated with the utility functions are of mixed logit form, with the mixing distribution being the distribution of ideal departure/arrival times in the population of the market. That is, the probability of choosing $j$ from the set of available itinerary fare classes is given by the following:

$$p_i(j) = \int_0^W \frac{e^{V(i,j|\tau(i))}}{\sum_{k \in \Phi(m)} e^{V(i,k|\tau(i))}} d\Theta(\tau), \tag{6.5}$$

where [0,*W*] is the standard time interval and $\Theta(\tau)$ is the distribution of ideal departure/arrival times in the population over the period [0,*W*]. The distribution function $\Theta(\tau)$ is established with empirical data. In AirVM, the mixture probability $p_i(j)$ is easily computed. If *S*(*m*) is the set of passengers traveling in market *m*, (the OD demand for market *m*), and *#(S(m))* the number of elements in *S(m),* then

$$p_i(j) = \int_0^W \frac{e^{V(i,j|\tau(i))}}{\sum_{k \in \Phi(m)} e^{V(i,k|\tau(i))}} d\Theta(\tau) = \frac{1}{\#(S(m))} \sum_{i \in S(m)} \frac{e^{V(i,j|\tau(i))}}{\sum_{k \in \Phi(m)} e^{V(i,k|\tau(i))}} \tag{6.6}$$

[70] Recently, Virtual M1nds has begun developing a survey instrument for use in a panel of air travelers to measure the effect of frequent flyer membership and similar variables on itinerary choice.

**6.3.7:** Recall that each pag actually carries four different utility functions in its state vector. They all are of the form described above, depending only on trip purpose (business or leisure) and time sensitivity (departure or arrival). These distinctions can be considered as discrete probability distribution on the synthetic population. Let $p_i^b$ represent the probability that an individual is taking a business trip in market *m*, while $p_i^l$ is the probability of a leisure trip. Obviously $p_i^b + p_i^l = 1$. Further, let $p_m^a$ be the probability of the pag being arrival sensitive and, as would be expected, $p_i^d$ the probability that it is departure sensitive. Since we also have that $p_i^a + p_i^d = 1$, then, assuming that trip purpose and time sensitivity are independent, we have

$$1 = (p_i^b + p_i^l)(p_i^a + p_i^d) = p_i^b p_i^a + p_i^b p_i^d + p_i^l p_i^a + p_i^l p_i^d, \tag{6.7}$$

which is also a discrete distribution, and therefore

$$\begin{aligned} p_i(j) = p_i^b p_i^a \int_0^W \frac{e^{V_{b,a}(i,j|\tau(i))}}{\sum_{k\in\Phi(m)} e^{V_{b,a}(i,k|\tau(i))}} d\Theta_{b,a}(\tau) + p_i^b p_i^d \int_0^W \frac{e^{V_{b,d}(i,j|\tau(i))}}{\sum_{k\in\Phi(m)} e^{V_{b,d}(i,k|\tau(i))}} d\Theta_{b,d}(\tau) \\ + p_i^l p_i^a \int_0^W \frac{e^{V_{l,a}(i,j|\tau(i))}}{\sum_{k\in\Phi(m)} e^{V_{l,a}(i,k|\tau(i))}} d\Theta_{l,a}(\tau) + p_i^l p_i^d \int_0^W \frac{e^{V_{l,d}(i,j|\tau(i))}}{\sum_{k\in\Phi(m)} e^{V_{l,d}(i,k|\tau(i))}} d\Theta_{l,d}(\tau). \end{aligned} \tag{6.8}$$

In this equation, the subscripts on the observed utilities *V* and the mixing distributions $\Phi$ represent the differing incidence of these properties in the synthetic population. Clearly the integrals here become sums when considering a finite synthetic population, as shown in the original mixing structure described earlier.

**6.3.8:** To this point, the itinerary choice discussion has been couched in terms of alternative flights or sequences of flights that connect a given origin-destination market. But that is not exactly the choice being made by the pag. In fact, the pag is selecting a cabin and fare class on board a flight. For example, a US domestic flight might offer five different fares on the same physical flight. There might be two different offerings in the first class cabin, one which is unrestricted and another which does not allow the fare to be refunded if the ticket is not used, only applied to another flight (with a change fee deducted). In the main cabin, there might be three different options, such as a full fare option which is fully refundable if not used, a slightly less expensive fare class which only allows the cost of an unused ticket to be applied to another flight, and a third, very low fare for which there is no refund if cancelled and must be purchased at least 30 days in advance. It is the practice in the industry to offer a number of fares on-board the same flight, which differ not only in cost to the passenger but also in refund conditions, advance purchase options, or other subtle features. Recent practice in the US, for example, is to charge extra for checked luggage in less expensive fare classes. The itinerary choice options then are different fare classes in addition to different flight itineraries, and nothing else in the analysis changes. One of the current deficiencies of AirVM is itinerary choice impact of these restrictions and conditions, since the choice model does not contain terms which reflect them.

## 6.4: The Population Distribution of Ideal Departure and Arrival Times

**6.4.1:** One of the important characteristics that is explicitly represented in the pag choice model is ideal departure/arrival times. By ideal is meant the best time, in the view of the pag, for an itinerary to depart or arrive given the nature of the journey. (To simplify matters slightly, this part of the discussion will be phrased in terms of only 'ideal' time, since the same arguments hold for ideal departure or ideal arrival times.) The ideal time is considered to be represented by a probability distribution across the standard time interval of one week. That interval is measured in minutes from 00:00 AM on Monday morning. There are thus 10,800 minutes in the standard time interval. It is also assumed that the distribution repeats every 10,800 minutes, so that any translation by 10,800 yields the same result. This accommodates schedule wrap as discussed earlier in **5.5.4**.

**6.4.2:** Two approaches have been explored in the literature to characterize the distribution of ideal times. One suggests that it can be represented by a mixture of normal distributions that represent the morning and afternoon surges of demand observed in most markets. The other offers Fourier series sine-cosine decomposition of these cyclical ideal time curves. Parker and Walker (2007) have reported on the normal decomposition and Koppelman, Coldren and Parker (2007) have studied the Fourier analysis representation, in each case considering the decomposition for a single day rather than the standard week.  A six-term Fourier decomposition is given by

$$\begin{aligned}\phi(\tau) = \varphi_1 \sin\left(\frac{2\pi\tau}{1440}\right) + \varphi_2 \sin\left(\frac{4\pi\tau}{1440}\right) + \varphi_3 \sin\left(\frac{6\pi\tau}{1440}\right) \\ + \varphi_4 \cos\left(\frac{2\pi\tau}{1440}\right) + \varphi_5 \cos\left(\frac{4\pi\tau}{1440}\right) + \varphi_6 \cos\left(\frac{6\pi\tau}{1440}\right)\end{aligned} . \tag{6.9}$$

**Figure 6.6a:  Empirical Densities**

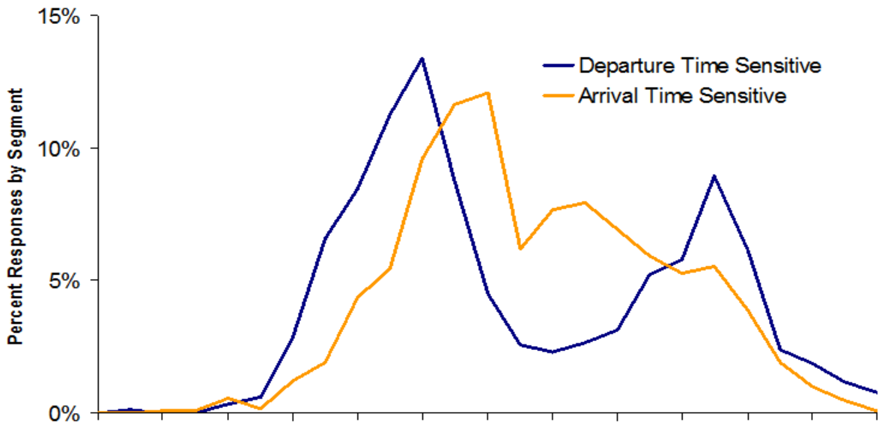



**Figure 6.6a:  Mixed Normal Densities**

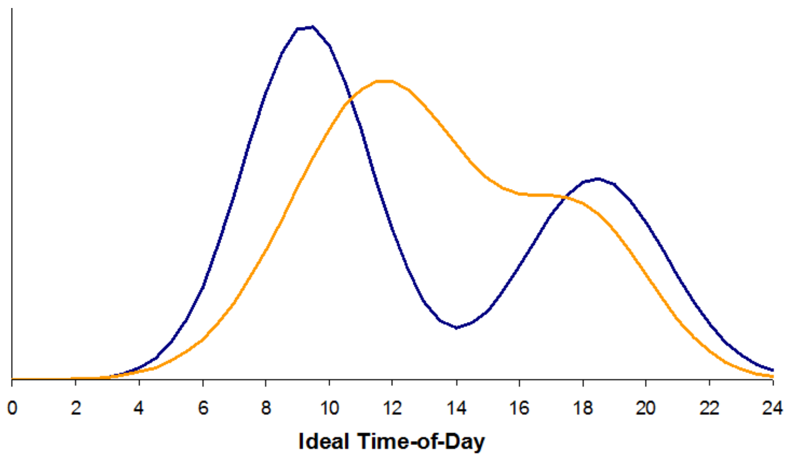



**Figure 6.6:  Mixed Normal Representation of Ideal Time-of-Day**

**6.4.3:** In Figure 6.6, the decomposition into a mixture of two normal distributions, $N_1(\mu_1,\sigma_1)$ and $N_2(\mu_2,\sigma_2)$, with mixing value $\pi$ is shown. Figure 6.6a shows the empirical data (collected from the Sidestep Survey discussed in Appendix K) and 6.6b the fitted normal mixture. The values of the parameters are given in the Table 6.1. In AirVM, four curves are described, one for each of the four possible combinations of arrival/departure sensitivity and business/leisure trip purpose. The coefficients reported are typical, but those used in the commercial version of AirVM are confidential.

**6.4.4:** Neither the mixed normal nor the Fourier decomposition structures have a strong behavioral motivation. They are pretty much curve fitting exercises (nearly pure reductive models, in the sense of Section 2.3). There may be some relationship between the parameters of the mixed normal, however, that could be revealed by further research. Perhaps the means or standard deviations could be regressed against population characteristics. While there is data available to do this research, there has been no opportunity to carry it out. Meanwhile, AirVM uses the normal decomposition, since random number generators with normal distributions are more readily available than distributions based on sine and cosine functions.

**6.4.5:** The analyses presented above consider time-of-day, but AirVM works with a standard week, consisting of a Monday through Sunday time frame. To extend the time-of-day curve to encompass a week, one should consider different time-of-day curves for each day of the week, and concatenate the seven resulting curves to form the weekly curve (since each daily curve goes to zero in the early morning hours. Unfortunately, there was insufficient data in the Sidestep data set from which the curves described in Table 6.1 and Figure 6.6 were derived to determine a curve for each day separately. A close approximation can be determined by using the basic mixed normal curve and combining seven versions of that curve, using a mixing distribution describing the fraction of weekly travel on a given day of the week. This data is illustrated in Figure 6.7 (using data from ARC, 2006). This

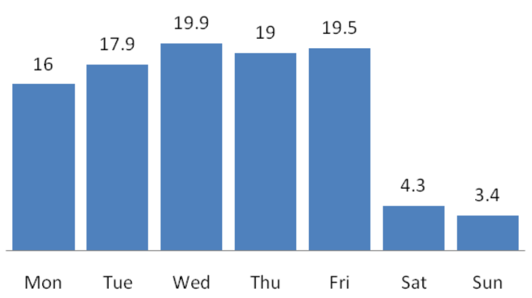



**Figure 6.7: Percent of Weekly Travel by Day of Week.** (Source: ARC, 2006)

is the form of the distribution as it is used in AirVM.

**Table 6.1: Ideal Time of Day Mixed Normal Parameter Estimates** (Source: Parker and Walker, 2007)

| | Estimate (Hours from Midnight) | t-stat |
|---|---|---|
| **Departure Time** | | |
| *Distribution 1: Morning Peak* | | |
| Morning Peak Mean | 9.31 | 139.2 |
| Morning Peak Variance | 4.00 | 19.2 |
| *Distribution 2: Afternoon Peak* | | |
| Afternoon Peak Mean | 16.42 | 186.3 |
| Afternoon Peak Variance | 5.04 | 14.7 |
| *Mixture Value* | 0.46 | 9.0 |
| Observations | 1877 | |
| R-squared | 0.8649 | |
| **Arrival Time** | | |
| *Distribution 1: Morning Peak* | | |
| Morning Peak Mean | 11.66 | 44.4 |
| Morning Peak Variance | 6.07 | 0.4 |
| *Distribution 2: Afternoon Peak* | | |
| Afternoon Peak Mean | 16.09 | 46.2 |
| Afternoon Peak Variance | 4.49 | 5.0 |
| *Mixture Value* | 0.96 | |
| Observations | 1437 | |
| R-squared | 0.7239 | |

## 6.5: The Distribution of Ticketing Group Size

**6.5.1:** Up to this point, the probabilities associated with the choices being made by the pags have been the focus of attention. Among the data in the pag's state vector are several variables that represent the stochastic dynamics of the ticketing process itself. These include the party or group size (the number of individual tickets being sought at the time of booking), the ticketing instant (when, during the ticketing time prior to departure that the pag will request tickets), and the possibility of ticket cancellation. The models that govern the assignment of these data values to specific pags are external to the agents themselves, but reflect the incidence of these variables in the travelling population. A synthetic population must be representative of the population of consumers that are to be represented by the virtual market being built. The relevant parameters of the models are estimated using standard statistical techniques applied to empirical data, and are outlined below.

**6.5.2:** Recall that a pag consists of one or more passengers. This situation becomes important when the objective is to estimate observed demand and passenger market share by multiplying the number of tickets sold on a given itinerary by the reciprocal of that itinerary's probability, since the observed frequency will reflect the effect of the ticketing group size. Let $Q_m$ be a discrete probability distribution defined over the positive integers $Q_m \equiv \{q_m(1), q_m(2), \ldots q_m(k), \ldots\}$ representing the probability of a ticketing group being of size $k$ = 1, 2,… . Then the number of passengers who want to travel on a specific itinerary fare class $j$ in market $m$, denoted $d_m(j)$, is the expected value under $Q_m$ of the number of pags requesting that itinerary. That is, if $d_m^*(j)$ is the number of pags expected to choose itinerary $j$, then,

$$d_m(j) = \sum_{k=1}^{\infty} k q_m(k) d_m^*(j) = d_m^*(j) \sum_{k=1}^{\infty} k q_m(k), \tag{6.10}$$

where $q_m(k)$ is the probability that $k$ tickets are being requested by the pag. Now, suppose $D_O^*(m)$ is the total number of pags moving in market $m$ on all itineraries (the demand). Then $d_m^*(j) = D_O^*(m) p_m(j),$ and so

$$d_m(j) = d_m^*(j) \sum_{k=1}^{\infty} k q_m(k) = D_O^*(m) p_m(j) \sum_{k=1}^{K} k q_m(k) \tag{6.11}$$

and, of course,

$$D_O^*(m) = \frac{d_m(j)}{p_m(j) \sum_{k=1}^{K} k q_m(k)}. \tag{6.12}$$

Finally, if $D_O(m)$ is the total number of *passengers* in market $m$, then it must be the case that

$$D_O(m) = D_O^*(m) \sum_{k=1}^{\infty} k q_m(k) = \frac{d_m(j) \sum_{k=1}^{\infty} k q_m(k)}{p_m(j) \sum_{k=1}^{K} k q_m(k)} = \frac{d_m(j)}{p_m(j)}. \tag{6.13}$$

**6.5.3:** Now consider the probability distribution of group size, $Q$ (ignoring the market $m$ here to simplify the notation slightly). I assert (subject to empirical verification, described below) that $Q$ is a *Poisson distribution truncated at zero*; that is, if $Y$ is the random variable distributed as $Q$ then

$$\Pr\left[Y \le y\right] = \frac{1}{\left(1-e^{-\eta}\right)} \sum_{i=1}^{y} \frac{e^{-\eta}\eta^{i}}{i!} = F_Q(y) \tag{6.14}$$

for some empirical parameter $\eta$, where $F_Q(y)$ is the distribution function corresponding to the discrete probabilities defined by $Q$**.** Then, denoting the expectation operator as $\mathsf{E}[\bullet]$ and the variance operator as $\mathsf{V}[\bullet]$,

$$\mathsf{E}\left[Y\right] = \frac{\eta}{1-e^{-\eta}} \tag{6.15}$$

and

$$\mathsf{V}\left[Y\right] = \frac{\eta\left(1-\eta e^{-\eta}-\eta^{2}e^{-\eta}\right)}{\left(1-e^{-\eta}\right)^{2}}. \tag{6.16}$$

The parameter $\eta$ can be estimated from a sample mean $\bar{Y}$ using the expectation equation above, except that the equation cannot be solved explicitly for $\eta$. However, Haight (1967, p. 87) cites the following Lagrange series solution

$$\hat{\eta} = \bar{Y} - \sum_{j=1}^{\infty} \frac{j^{j-1}}{j!}\left[\bar{Y}e^{-\bar{Y}}\right]^{j}. \tag{6.17}$$

**Table 6.2: Group Size** (Source: Coy, 2003)

| Group Size | Observed Frequency | Expected Frequency |
|---|---|---|
| 1 | 164 | 160.70 |
| 2 | 45 | 47.45 |
| 3 | 5 | 9.34 |
| 4 | 4 | 1.38 |
| 5 or more | 1 | 0.16 |

**6.5.4:** Data supplied by Coy (2003) from Continental Airlines yields the analysis shown in Table 6.3. Using the Lagrange method described in Equation 6.22, $\hat{\eta}$ is estimated to be $\hat{\eta}$ = 0.5904 from the $\bar{Y}$ = 1.324 of the data. A simple chi-square goodness-of-fit test yields $\chi^2$ = 0.259 (p = 0.879). Notice that the deviation from the Poisson is in the tail, and even with a data set of this size, contributes a substantially larger proportion of the deviation from Poisson expected values than the smaller values. These results are not stunning, but they do lend some support to the hypothesis that the truncated Poisson fits these observations, at least to a reasonable first-approximation.

**6.5.5:** The examination of additional data sets suggests, in fact, that there are separate group size parameters for the leisure and business trip purpose. Examination of the group size distributions illustrated in Figure 6.8 illustrates this phenomenon indirectly. Here group sizes are calculated from ARC-reported ticket sales for three different markets – Seattle to Miami (SEA>MIA), Chicago to Honolulu (CHI>HNL), and Atlanta to Houston (ATL>IAH). The CHI>HNL market contains a substantial fraction of travelers heading out for a holiday in Hawaii, that is, leisure trips. ATL>HOU, on the other hand, contains a relatively higher proportion of business travelers, while SEA>MIA contains a more even mixture of each. This implies that the actual, observed party size distribution is a mixture of two truncated Poisson distributions, one for leisure and one for business trips, with different values for $\eta$, say $\eta_L$ and $\eta_B$ respectively. The observed distribution is then

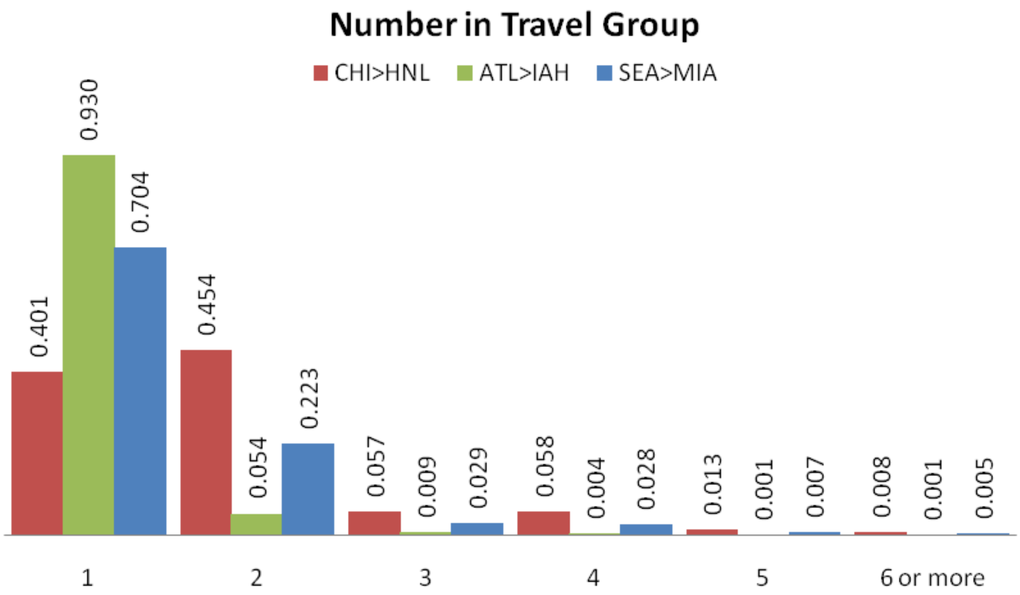



**Figure 6.8: Travel Group Size Distributions.** (Source: ARC 2006)

$$\Pr\left[Y \le y\right] = \pi_g \left[ \frac{1}{\left(1-e^{-\eta_L}\right)} \sum_{i=1}^{y} \frac{e^{-\eta_L}\eta_L^i}{i!} \right] + (1-\pi_g) \left[ \frac{1}{\left(1-e^{-\eta_B}\right)} \sum_{i=1}^{y} \frac{e^{-\eta_B}\eta_B^i}{i!} \right] \tag{6.18}$$

where $\pi_g$ is the mixing fraction. However, since the party size is assigned to each pag individually, there is no need to be concerned with mixing, because the trip purpose is assigned separately and the appropriate values of $\eta$ can be used based on that assignment.

**6.5.6:** It is also suggested by Figure 6.8 that the group size distribution parameters, even given trip purpose, differ from market to market. Given the importance of group size in the simulation of the airline ticketing process, differences between markets could have a noticeable effect on itinerary distribution. It is for this reason that within AirVM it is possible to assign a different group size parameter to any specific market, even though a default value is provided at systems initialization. Fortunately, data for group size is readily available from routine industry sources, so market-specific values are not difficult to acquire.

## 6.6: The Stochastic Process of Ticketing Instances

**6.6.1:** Airline ticket bookings occur randomly over the time from some appropriate starting point up to time of departure of the first segment in the itinerary. In general, as the time of departure of a flight approaches, ticket sales become more frequent. The model of the ticketing-over-time phenomenon used in AirVM is a *non-homogeneous Poisson model*.[71] Let *B(t)* be the total number of tickets sold in the time period [*0*, *t*], $t \in \left[0, T\right]$, where *T* is the departure time of the flight and time 0 is the opening of the ticketing period. Write, for $x$ = 1, 2, …

$$b_x\left(t\right) = \Pr\left[B(t) = x\right] \tag{6.19}$$

Define a *ticketing instant i* as a time $i \in \left[0, T\right]$ when a sale is initially requested from the available itineraries, and denote the positive-integer valued random variable $Y_i$ as the group size associated with that pag, as discussed in Section 6.5. Denote the stochastic

[71] I have been unable to find a reference to the approach in the airline operations literature.

process representing the number of ticketing instances observed in a time period [*0*, *t*] as $\{N(t); t \in [0,T]\}$. Then

$$B(t) = \sum_{i=1}^{N(t)} Y_i \tag{6.20}$$

where it assumed that all $Y_i$ are independent and identically distributed as some random variable *Y*. Further, assume *Y* is independent of *B(t)* for all choices of *t*. The mean and variance of *B(t)* are given by

$$\mathsf{E}\left[B(t)\right] = \mathsf{E}\left[N(t)\right]\mathsf{E}\left[Y\right] \tag{6.21}$$

and

$$\mathsf{V}\left[B(t)\right] = \mathsf{E}\left[N(t)\right]\mathsf{V}\left[Y\right] + \mathsf{V}\left[N(t)\right]\mathsf{E}^2\left[Y\right] \tag{6.22}$$

respectively.

**6.6.2:** I will now focus the discussion on the stochastic process *N(t)*, the number of ticketing instances occurring in the time period [0,*t*]. A *counting process* is a stochastic process that assigns a non-negative integer to some parameter *t*, (usually time). Let $\{N(t); t > 0\}$ be any counting process, and assume that the following properties hold for that process:

**6.6.2.1:** $N(0) = 0$.

**6.6.2.2:** $\{N(t); t > 0\}$ has *independent* increments: That is, for all choices of $t_0 < t_1 < \ldots < t_n$, $N(t_j) - N(t_{j-1})$, $j = 1, 2, \ldots n$ are independent random variables.

**6.6.2.3:** For any $t > 0$, $0 < \Pr[N(t) > 0] < 1$.

**6.6.2.4:** For any $t \geq 0$, events cannot occur exactly simultaneously. That is,

$$\lim_{h \to 0} \frac{\Pr\left[N(t+h) - N(t) \geq 2\right]}{\Pr\left[N(t+h) - N(t) = 1\right]} = 0 \tag{6.23}$$

**6.6.2.5:** *N(t)* has *stationary* increments. That is, for any two points $t > s \geq 0$, and for any $h > 0$, the random variables $N(t) - N(s)$ and $N(t+h)-N(s+h)$ are identically distributed.

Then $\{N(t); t > 0\}$ is a *Poisson stochastic process* for some empirical parameter $\nu > 0$. In other words,

$$\Pr\left[N(t) = x\right] = \frac{e^{-\nu t} \nu t^{x}}{x!} \tag{6.24}$$

**6.6.3:** However, **6.6.2.5** is not true for the ticket sales process. The closer to departure time the pag gets, the more intense the ticketing rate becomes, and thus the process does not have stationary increments. The relaxation of **6.6.2.5** gives rise to the *non-homogeneous Poisson process*. Specifically, replace **6.6.2.5** with **6.6.3.1**,

**6.6.3.1:** Suppose, for some continuous, differentiable, non-negative function $\nu(t)$ it is true that

$$\lim_{h \to 0} \frac{\Pr\left[1 - N(t+h) - N(t) = 0\right]}{h} = \nu(t) \tag{6.25}$$

Then, if the counting process $\{N(t); t > 0\}$ satisfies assumptions **6.6.2.1** through **6.6.2.4**, and **6.6.3.1,** it is true that

$$\Pr\left[N(t) = x\right] = \frac{e^{-m(t)} m(t)^{x}}{x!} \tag{6.26}$$

where

$$m(t) = \int_{0}^{t} \nu(u) du \,. \tag{6.27}$$

$\nu(t)$ is called the *intensity function*, and *m(t)* the *mean value function*. *N(t)* is called, under these circumstances, a *non-homogeneous* Poisson process. If $\{N(t); t > 0\}$ is such a process, then, since it is a Poisson process, it must be that

$$\mathrm{V}\left[N(t)\right] = \mathrm{E}\left[N(t)\right] = m(t). \tag{6.28}$$

An early, but quite readable, treatment of these types of Poisson processes is given by Parzen (1962, pp. 117-159).

**6.6.4:** The problem now is to estimate $\nu(t)$. Empirical data suggests that an exponential function on the domain [0, *T*] might be a good fit. That is, $\nu(t)$ might be expressed as a function of the general form

$$\nu(t) = \alpha e^{-\xi(T-t)} \tag{6.29}$$

for some empirically estimated parameters $\xi$ and $\alpha$. The graphs in Figure 6.10 show the number of tickets sold per day (the left side) and the accumulative number sold for each day up to departure (the right side) for four markets. The four markets illustrated in the figure all show similar patterns, with a fair amount of variation in the per-day ticket counts caused by inherent daily variations when tickets are sold (e. g. when travel agencies are open for business), but a reasonably uniform accumulative tickets sold pattern. Indeed, the $R^2$ measure of fit for the per day data is around 0.82, while total accumulated tickets sold reaches $R^2$ values of over 0.96.

**6.6.5:** The results shown in Figure 6.9 also illustrate other interesting aspects of the ticketing stochastic process. Notice, for example, the Chicago to Honolulu market in Figure 6.10d contains a burst of ticket sales during the time span from about 45 days to 20 days before departure. This is most likely due to a higher-than-usual number of groups and vacation package ticket sales, which would be offering discounted fare tickets for purchase far enough in advance. This burst of activity also, of course, affects the accumulative distribution. The caution here is that, even though this model represents the process reasonably well in most cases, markets will differ substantially from one to the next. Within the operations of AirVM, the parameters of the accumulative distribution for a given market are estimated and then adjusted for group size, which is also measured using the same data set, with the application of Equations 6.21 and 6.28.

**Tickets Sold/Day** | **Accumulative Tickets Sold**

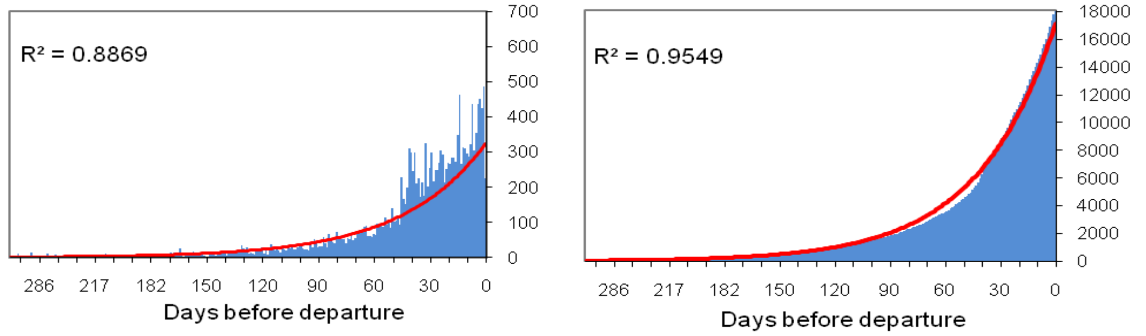



**Figure 6.9a: SEA>MIA**

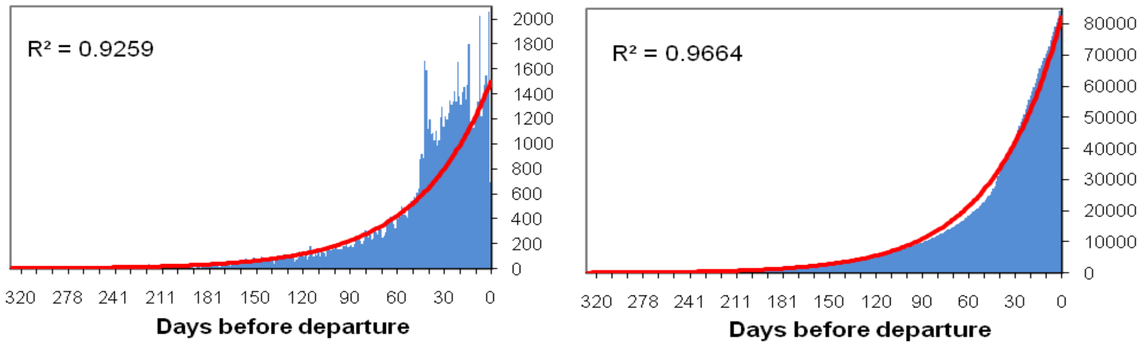



**Figure 6.9b: BOS>MIA**

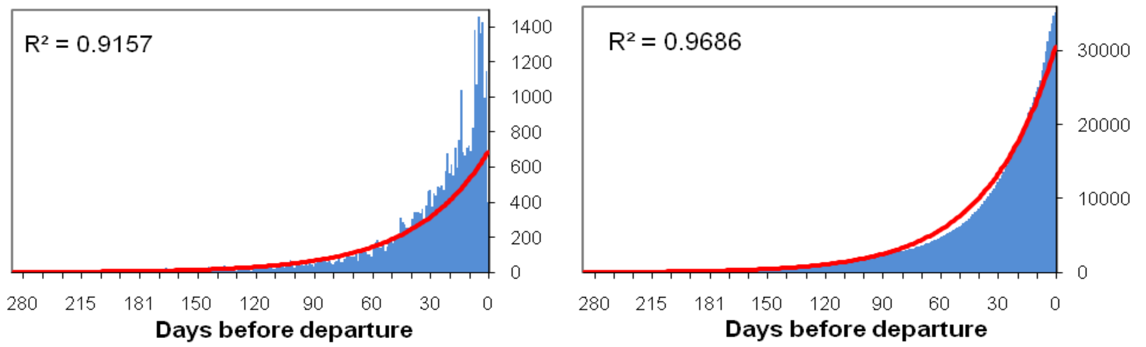



**Figure 6.9c: BOS>SEA**

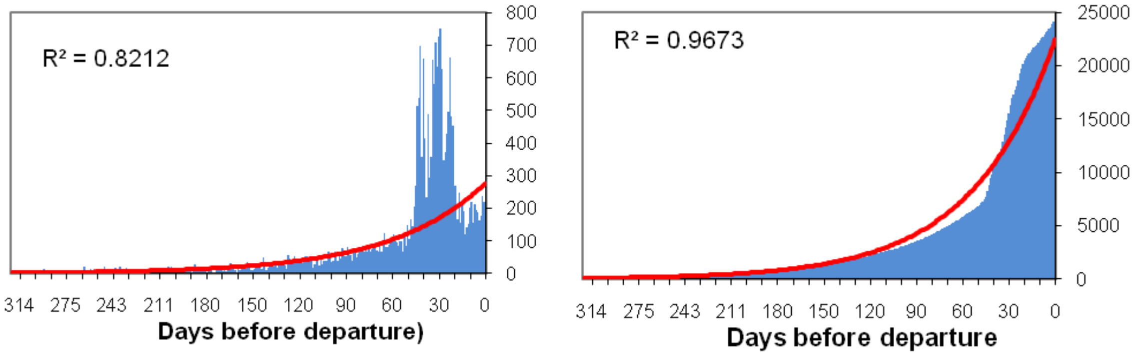



**Figure 6.9d: CHI>HNL**

**Figure 6.9: The Ticketing Stochastic Processes** (Source, ARC, 2006)

## 6.7: The Ticket Cancellation Stochastic Process

**6.7.1:** Other models relevant to the ticketing context are the booking cancellation models and day-of-departure no-show and go-show models. These last are not currently implemented in AirVM, but are described in extensive detail in Garrow (2004). The cancellation model uses hazard analysis to characterize purchased ticket cancellations as functions of several associated variables. Illiescu, Garrow and Parker (2007) discuss this approach in detail. In general, hazard models are used to represent time-to-event phenomenon, such as time to onset of illness after exposure to hazard. The terms and phrasing in the field reflect its heritage from the medical research world. In the AirVM case, the "event" is either the use or cancellation of a ticket.

**6.7.2:** The method used for the cancellation model is referred to as the Discrete Time Proportional Odds (DTPO) model. Consider for ticket *i*, say, the time to either use or cancellation of a ticket, denoted $T_i$. DTPO partitions the time to event into discrete time units (e. g. days) of represented by the half-open intervals $(t_0, t_1], (t_1, t_2], \ldots, (t_{k-1}, t_k]$, where $t_0$ is the date of issue of the ticket and $t_k$ represents either the date of departure if the ticket is used or the date of refund or exchange in the case cancelation. The *discrete hazard* of ticket *i* being cancelled in period *k* is the probability that *i* is cancelled in period *k* given that it has not been cancelled in any period from 0 to *k*-1, that is:

$$h_{i,k} \equiv \Pr[T_i = k \mid T_i > k-1] \tag{6.30}$$

Thus, from conditional probability, the probability that the cancellation event will occur in time period *k* is

$$\begin{aligned} \Pr[T_i = k] &= \Pr[T_i = k \mid T_i \geq k-1]\Pr[T_i \neq k-1 \mid T_i \geq k-1]\ldots\Pr[T_i \neq 1 \mid T_i \geq 1] \\ &= h_{i,k}(1-h_{i,k-1})\ldots(1-h_{i,1}), \end{aligned} \tag{6.31}$$

and the probability that a ticket will *not* be cancelled prior to departure at time *k*, the *survival probability*, is given by

$$
\begin{aligned}
S_{i,k} &= \Pr[T_i > k] \\
&= \Pr[T_i \neq k \mid T_i \geq k] \Pr[T_i \neq k-1 \mid T_i \geq k-1] \ldots \Pr[T_i \neq 1 \mid T_i \geq 1] \\
&= (1-h_{i,k})(1-h_{i,k-1}) \ldots (1-h_{i,1}).
\end{aligned} \tag{6.32}
$$

Now apply the log odds transformation to the hazard probability to create a model of that probability as a function of other variables of interest, such as

$$
\log\left(\frac{h_{i,k}}{1-h_{i,k}}\right) = \Psi_{i,k} + \xi_1 X_{i,1} + \xi_2 X_{i,2} + \ldots \xi_j X_{i,j}. \tag{6.33}
$$

From this, compute $h_{i,k}$ as

$$
h_{i,k} = \frac{1}{1 + e^{\Psi_{i,k} + \xi_1 X_{i,1} + \xi_2 X_{i,2} + \ldots \xi_j X_{i,j}}}. \tag{6.34}
$$

The constant term $\Psi_{i,k}$ is called the *baseline hazard function*, and represents that part of the hazard not accounted for by the variables $X_{i,1}$, $X_{i,2}$, ... $X_{i,j}$. Note that, if the magnitude of the conditional probabilities in $h_{i,k}$ are small, then the DTPO model approximates the proportional hazards model, described by the classic work of Cox and Oakes (1984).

**6.7.3:** Using maximum likelihood methods, data from ARC (2006) were studied to explore the relevance of several factors to the passenger cancellation probabilities. The result of this analysis is illustrated in Table 6.3. The data shown in this table is reported in Illiescu *et al.* (2007). A model of the same structure is used in AirVM, but the coefficients are proprietary. Among the covariates showing significant effect on the cancellation probability are the number of days since the ticket was issued (days from issue), remaining days before departure, number in traveling party (group size), departure day-of-week, whether a Saturday night stay was included in the trip, OD market, carrier, and fare. Not all of these variables are available to AirVM, however. In particular, the journey structure component of the state vector does not accommodate

Saturday night stay behavior.[72] Further, available data does not allow estimation of the carrier effect around the world. Thus the cancellation hazard model used in AirVM is of the form

$$\log\left(\frac{h_{i,k}}{1-h_{i,k}}\right) = \Psi + \xi_1(i)t_p^d + \xi_2(i)t_d^d + \xi_3(i)g(i) + \xi_3(i)\ln f^*(i), \tag{6.35}$$

where $\xi_k(i)$ are empirical coefficients that are drawn from normal distributions, as in the itinerary choice model (Section 6.3) and $t_p^d$ is the number of days from the ticket purchase date, $t_d^d$ the days remaining before departure, *g(i)* the group size for pag *i*, and *f*(i)* the fare paid by pag *i*. The baseline hazard, $\Psi$, is assumed constant for all pags.

**6.7.4:** Detailed examination of actual passenger ticketing behavior reveals some anomalies that are worthy of note, although they are not reflected in the current AirVM model. It is not uncommon in a business context, for example, to buy tickets for the same individuals on several multiple flights in the same market on the same day, not knowing in advance exactly when the flight will be needed. If fully flexible and fully refundable tickets are purchased, the unused ones are simply refunded. Of course the carrier can't sell them to other passengers, and hence lose revenue. Also, occasionally a block of 20 or more tickets are sold at one time on a flight. This is a group booking, often associated with tour companies. Further, many corporations have special arrangements with carriers to purchase tickets for their business travelers at special rates. These passengers often have little say in the choice of itinerary or even desired time of travel. As AirVM matures, hopefully there will be opportunities to introduce models of these situations. Theoretically there is nothing in the AirVM structure to prevent it.

[72] A long standing practice in airline revenue management was to charge less for a round trip fare if the passenger stayed at his destination over a Saturday night. This moves business traffic off of Friday and on to Sunday, where loads are usually lighter. This practice is seen less today, however.

**Table 6.3: Estimated Cancellation Model** (Source, Illiescu *et al.*, 2007)

| Variable | Estimated Value | Standard Error | 95% Confidence Interval | |
|---|---|---|---|---|
| Days from Issue (reference category 46–90 days) | | | | |
| 0 to 3 | 2.135 | 6.19 | 1.780 | 2.559 |
| 4 to 7 | 2.966 | 12.32 | 2.495 | 3.526 |
| 8 to 14 | 2.461 | 10.36 | 2.075 | 2.918 |
| 15 to 21 | 2.048 | 7.88 | 1.714 | 2.448 |
| 22 to 45 | 1.681 | 5.98 | 1.417 | 1.993 |
| Days before Depart | 0.963 | -24.22 | 0.960 | 0.966 |
| Group size (reference = one person) | | | | |
| 2 people | 0.440 | -13.93 | 0.392 | 0.494 |
| 3 or more people | 0.304 | -10.88 | 0.245 | 0.377 |
| Saturday night stay | 0.779 | -5.66 | 0.714 | 0.849 |
| Outbound day-of-week (reference = Sunday) | | | | |
| Monday | 1.297 | 4.33 | 1.153 | 1.460 |
| Tuesday | 1.275 | 3.86 | 1.127 | 1.442 |
| Wednesday | 1.135 | 1.99 | 1.002 | 1.287 |
| Thursday | 0.862 | -2.14 | 0.753 | 0.987 |
| Friday | 0.823 | -2.74 | 0.716 | 0.946 |
| Saturday | 0.945 | -0.71* | 0.809 | 1.104 |
| Market (reference = BOS>MIA) | | | | |
| BOS>SEA | 0.653 | -7.69 | 0.586 | 0.728 |
| HNL>ORD | 0.441 | -5.69 | 0.333 | 0.585 |
| MIA>BOS | 0.618 | -7.28 | 0.543 | 0.703 |
| MIA>SEA | 1.347 | 3.86 | 1.158 | 1.567 |
| ORD>HNL | 0.669 | -4.85 | 0.569 | 0.788 |
| SEA>BOS | 0.629 | -7.61 | 0.558 | 0.709 |
| SEA>MIA | 0.625 | -5.86 | 0.535 | 0.732 |
| Carriers (masked information) | | | | |
| Carrier 2 | 1.133 | 2.20 | 1.014 | 1.266 |
| Carrier 3 | 0.392 | -10.79 | 0.331 | 0.465 |
| Carrier 4 | 0.804 | -2.20 | 0.662 | 0.977 |
| Carrier 5 | 1.089 | 1.20* | 0.948 | 1.250 |
| Pro-rated Fare | 1.001 | 20.35 | 1.001 | 1.002 |

Number of observations = 151,401

Pseudo $R^2$ = 0.0577

*Not statistically significant at the 0.05 level

## 6.8: The Pag Willingness-to-Pay Modeling

**6.8.1:** The term *willingness-to-pay* (WTP) is a synonym for the demand vs. price curve, or simply the demand curve. It is usually presented as a graph with cost on the vertical axis and the number of individuals in the marketplace willing to pay that cost on the horizontal axis. WTP curves equivalent in concept to elasticity, and are important because, if known, they provide a mechanism for the computation of demand increase (or decrease) caused by a price decrease (increase). This is also referred to as *stimulated demand* in the airline industry. This is, of course, critical to the understanding and estimation of the effects of network enhancements on the demand supported by that network.

**6.8.2:** While the simplest discussion of WTP only concerns changes due to price (and that case will be the basis for the intuitive discussion and example below), it is easy to extend the concept to include inducement effects caused by other network feature changes, such as frequency, distance, and so on. This type of generalized price viewpoint is closely akin to the definition of consumer surplus in the context of the random utility passenger choice models. (See McFadden (1999), Parker and Lonsdale (2004), or Parker, Lonsdale and Erwin (2007) for a more detailed discussion). From a choice model perspective, and in most transportation planning fields, WTP can be determined by modeling two choice situations: the choice of making the trip or not, and the choice of making the trip by air, or not. These are the *trip generation* and *mode choice* models, respectively (see McNally, 2000). For this discussion, only the mode choice component in this process is of interest, since it is all that is required to meet the data requirement of AirVM. The issue of WTP can be considered from a much more sophisticated viewpoint, however, such as by using a nested logit model. See Garrow, Jones and Parker (2007) for a specific example.

**6.8.3:** Consider a binary choice random utility model where the two choices are: 1) to *fly* between the origin and destination in a particular market, or 2) *not to fly* (either not go or go by some other means). Let $A_F$ indicate the fly choice and $\sim A_F$ the not to fly choice. Recall the basic random utility formulation for a binary choice case: $U_{F,i}$, the utility to passenger $i$ of flying, and $U_{\sim F,i}$ the utility of not flying are given by the stochastic equations

$$U_{F,i}(\mathbf{x}_F) = V_{F,i}(\mathbf{x}_F) + \varepsilon_{F,i},$$
$$U_{\sim F,i}(\mathbf{x}_{\sim F}) = V_{\sim F,i}(\mathbf{x}_{\sim F}) + \varepsilon_{\sim F,i} \tag{6.36}$$

where the $V$'s are the observable portions of the utility functions and the $\varepsilon$'s the stochastic terms. As before, it is assumed that the $V_F$ $(x_F)$ is a function of a vector of attributes of flying and characteristics of the passenger denoted $x_F$, and that $V_{\sim F}(x_{\sim F})$ is a function of another vector $x_{\sim F}$ of the attributes of not flying (along with the passenger characteristics), not necessarily the same as $x_F$. Assume that each $\varepsilon$ is independent and identically distributed EV1 across the two alternatives and across all passengers, and so the assertion of a logit probability is valid. To wit:

$$\begin{aligned}\Pr[i \text{ chooses } A_F] \equiv P_i[A_F] &= \frac{e^{V_{F,i}(\mathbf{x}_F)}}{e^{V_{F,i}(\mathbf{x}_F)} + e^{V_{\sim F,i}(\mathbf{x}_{\sim F})}} \\ &= \frac{1}{1 + e^{V_{\sim F,i}(\mathbf{x}_{\sim F}) - V_{F,i}(\mathbf{x}_F)}}.\end{aligned} \tag{6.37}$$

The probability of not flying is, of course, $1 - P_i[A_F]$. Note that there is no requirement in random utility theory, or its logit incarnation, that $V_F$ and $V_{\sim F}$ in any way be the same function, or even the same functional form.

**6.8.4:** For use as the default structure of the fly/no-fly choice model, consider the following. Define three relevant attributes of the market: a constant that is interpreted to mean the (negative) utility of not making the trip by air, the minimum available fare to make the journey, and market distance (determined to be the great circle distance between the city pair that defines the market). Passenger characteristics are ignored. Now, let $\mathbf{x}_F$ be a function just a minimum fare term and $\mathbf{x}_{\sim F}$ be a function of a constant disutility (of not flying) and the distance between the cities. Then, assuming a linear-in-the-parameters expression, define

$$V_F(\mathbf{x}_F) = \eta_f f_{\min}$$
$$V_{\sim F}(\mathbf{x}_{\sim F}) = \eta_c + \eta_d d \tag{6.38}$$

where $f_{\min}$ is minimum fare and $d$ is market distance. Notice that $\eta_c$ is an alternative-specific constant for the no-fly alternative. This and the Equation 6.38 yields

$$P(A_F) = \frac{1}{1+e^{V_{\sim F,i}(\mathbf{x}_{\sim F}) - V_{F,i}(\mathbf{x}_F)}} = \frac{1}{1+e^{\eta_c + \eta_d d - \eta_f f}}. \tag{6.39}$$

Note that the subscript $i$ has disappeared from P, since there is no distinction in this model between one passenger and another. Figure 6.10 is a graph of this probability as a function of fare for representative values of $\eta_c$ = *-2.62304*, $\eta_d$ = *-0.00158*, $\eta_f$ = *-0.00235* and *d=240* minutes. It behaves exactly as would be expected, dropping off rather steeply then leveling out, and having a small positive probability that some won't fly at any price.

**6.8.5:** Given data which indicates current values of the attributes in a fly/no-fly model, and the current observed demand that exists with those attribute values, then computing the stimulated demand created by changing one of those attributes is easy. Simply calculate the new probability that exists with the new attribute values and apply that to

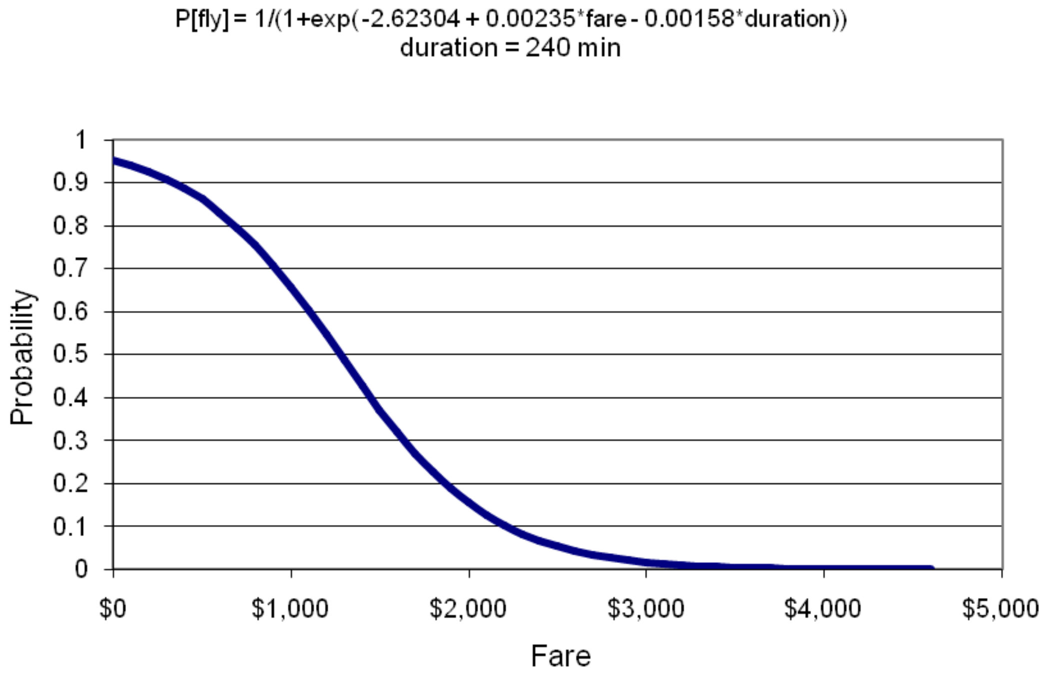



**Figure 6.10: Sample Willingness-to-Pay Curve**

the known demand. Specifically, let $P(A_F)$ be the probability of flying under the initial attribute conditions $\mathbf{x}_F$, and $P^*(A_F)$ the revised probability under the modified values of the attributes, $\mathbf{x}_F^*$, and suppose $D$ is the known demand at attribute level $\mathbf{x}_F$. Then the demand under the new values, $D^*$ is given by

$$D^* = D\frac{P^*(A_F)}{P(A_F)}. \tag{6.40}$$

For example, in the WTP curve shown in Figure 6.10 suppose that the demand at a fare of \$1500 is 550. The probability of flying at a \$1500 fare is about 0.3724. Now lower the fare to \$1000, and the probability climbs to 0.6577. From this equation, then, the demand will increase to 550 x (0.6577/0.3724) = 971. The stimulated demand is then 971 – 550 = 421.

**6.8.6:** The data for this estimated WTP function comes from the Sidestep survey (discussed in Appendix K). In that survey, after the respondent had chosen among the itinerary alternatives, he was then asked that, if the offered choices were the only options available, would he still take the trip. The responses to this choice question – a perfectly legitimate stated preference experiment – formed the data for the estimation of this simple WTP model.

## 6.9: The Structure of the Arasag State Vector

**6.9.1:** So far in this discussion the concentration has been focused on the passenger agent. But there are two other agents in the AirM virtual market. The arasag represents that portion of the air carrier relevant to the passenger ticket purchase activity, that is, the scheduling team and the pricing and revenue management teams. Also, the arasag is a client agent, and its choice protocol is the responsibility of the user client as that client interacts with the simulation, and therefore an arasag can be an avatar for a simulation client. The contents of the state vector for an arasag consist primarily of data describing scheduled flights and their availability. As the simulation proceeds, seats on these flights are offered for sale at stipulated prices, are purchased, and the revenue credited to the arasag. The arasag also invokes the appropriate revenue management protocols to modify prices and availability on the segments operated by the arasag carrier. Thus the state vector for the arasag consists of descriptions of all nonstop flight segments along with each fare class and associated revenue management protocol for all flight segments. There is also data associated with alliance relationships that is used to calculate connection times for direct vs. interline connections. Virtually all aspects of the arasag can be set a client user. However, there are over a thousand arasags in a typical simulation run, and any particular client can be expected to be interested in only a few of them. Therefore, as elsewhere in AirVM, the arasags need default values provided for the state vectors of those arasags that operate in a simulation, but are not engaged by a client.

**6.9.2:** The fare class structure for a flight leg depends on the cabin configuration of the airplane assigned to the leg. This information is available from the suppliers of schedule data, OAG and Innovata, and is included when this input is incorporated into AirVM. The main cabin always exists. By default, for an airplane with fewer than 100 seats, that is the only cabin that is considered to be available. Economy fare classes are always in the main cabin. If there is one other cabin on the airplane, then that is assumed to be a first class cabin (even though the carrier may call it something else). If there are two additional cabins, then the third one is considered the business class cabin. Intercontinental segments usually have three cabins. If the data indicate that a fourth cabin is available, then it is assumed to be an economy plus configuration, but these are fairly rare. By default, every main cabin has two fare classes, termed unrestricted and restricted economy, respectively. The fraction of seats in each defaults to 40% unrestricted and 60% restricted. If other cabins exist on the aircraft, a single fare class is assumed. All revenue management protocols are set by default to fixed price. See **5.4.8** for a description of the RM protocols used in AirVM.

**6.9.3:** The default fare class RM protocol is the fixed price protocol, which simply means that the same fare is charged to the pag for all seats in the cabin throughout the 127 days of the ticketing time frame. What should that fare be? The default fare is determined by applying a standard fare curve to the cabin arrangement of the aircraft, and using that standard curve to determine the default fare to be assigned to each cabin fare class. The standard curve is found by computing the empirical distribution function (EDF) of the fares from available data. The EDF of a data sample of size *n*,

$\{X_1, X_2, \ldots X_n\}$ is found by first ordering the data points in a sample from smallest to largest, and then computing the following function:

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x) \tag{6.41}$$

where $I$ is the indicator function defined by

$$I(X_i \le x) = \begin{cases} 0 \text{ if } x < X_i \\ 1 \text{ if } \mathrm{X}_i \le x \end{cases}. \tag{6.42}$$

It is true that an EDF is a sufficient statistic for any data sample. In fact, it contains (obviously) all the information that is available in the data sample. When graphed, an EDF looks like the example in Figure 6.11. For fare data, there are considerably more data points available, so the EDF is a smooth (if not terribly common) curve as

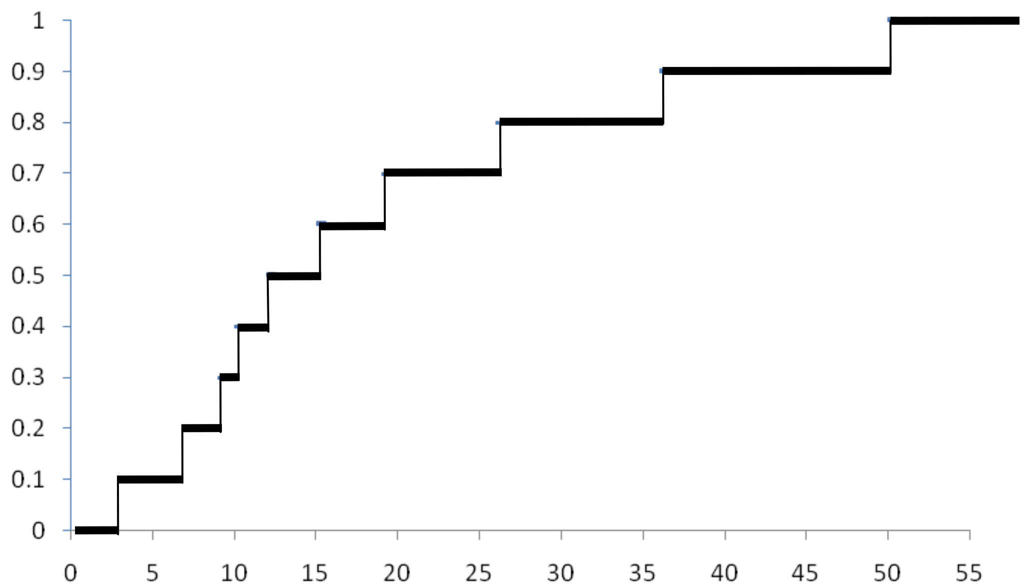



**Figure 6.11: Example of an Empirical Distribution Function (EDF)**

illustrated in Figure 6.12.

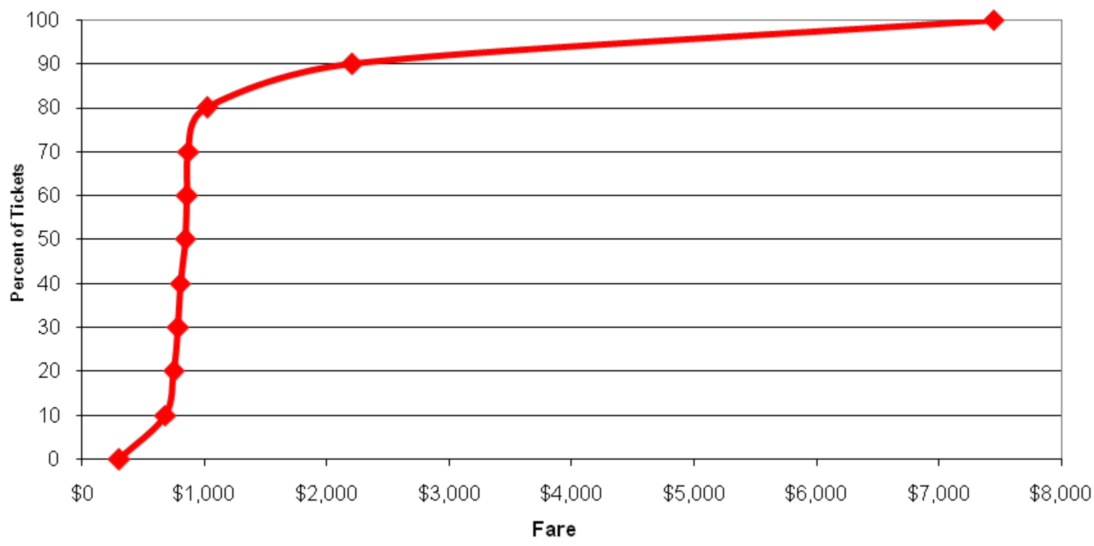



**Figure 6.12: Fare EDF for LAX>NRT Market. Nov 2007.** (Source: ARC, 2007)

**6.9.4:** This curve is characteristic of almost all markets that have been explored. The steep rise at the left end of the curve is typical of the closeness in price for the cheaper fares on a segment. The relatively wide curve at the upper part of the steep rise, and the long, almost horizontal line across the top following that, are characteristic of the consistent price increases and very high prices associated with the business and first class cabins. With appropriate data, one can associate the fare with the fare class, and hence the cabin in which the fare is charged. This translates into points on a graph like Figure 6.13 which demarcate the fares charged in each cabin and fare class. In addition, the horizontal axis of the curve can be associated with the distance between the two cities in the market. First compute the mean fare as a function of the great circle distance between two cities. Then use a multiplier to compute the consequent fares in each fare class. These multipliers are shown in Table 6.4. With this adjustment, a fare EDF can be created for any OD pair. Of course the fit is not perfect, but as a default construct, it is adequate.

**6.9.5:** The EDF representing fares on a flight segment is an example of an incidence distribution for the arasag synthetic population. And it is somewhat unusual in AirVM because there exists an actual list of the world's airlines with data for each from which the population of arasags can be created. The details of the segments flown by each arasag is a clear example. Thus there is little stochastic about the arasag population –

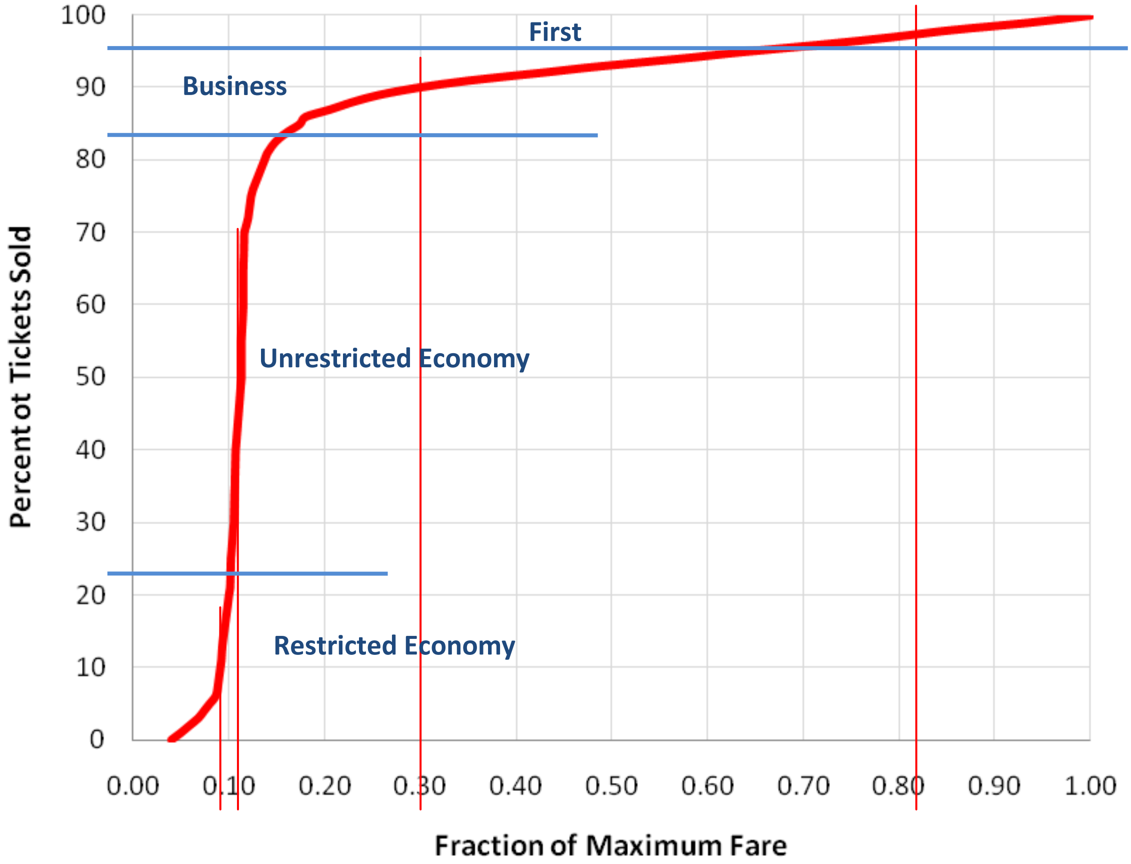



**Figure 6.13: Fare EDF with Fare Class Bounds and Nominal Fare Levels**

except the fare structure for each segment. In the absence of such fare data, the generalized EDF discussed above is used. At this time, this is the only stochastic component of arasag agents in AirVM.

**6.9.6:** One of the peculiar marketing elements of the airline industry is the concept of *code sharing*. Code sharing occurs when one airline sells tickets under its own name on another carrier's flight. For example, a Qantas flights from Los Angeles to Sydney would be operated by Qantas, but an American Airlines customer might purchase a ticket through American as though it were American's flights, not knowing it was Qantas until boarding the plane. Frequently, when looking at departure or arrival timing board, one will see the flight number and carrier of the same physical listing change. This is a manifestation of code sharing. Code sharing is done by contractual agreement between carriers. The airline that actually provides the airplane and crew is called the *operating carrier*, while the other code share partners are referred to as *marketing* *carriers*. Code sharing agreements are confidential between the carrier partners, involving the number of seats made available to the partners, revenue sharing arrangements, and so forth. At this point in the development of AirVM there has been no call for explication or representation of code share arrangements, although it would not be difficult to incorporate that aspect of the ticketing process at some future time.

**Table 6.4: Fare Class Nominal Fare Multipliers from EDF**

| | |
|---|---|
| Midpoint, First Class Fare Range (0.95-1.00) | 0.8200 |
| Midpoint, Business Class Fare Range (0.82-0.95) | 0.2960 |
| Midpoint, Unrestricted Economy Fare Range(0.22-0.82) | 0.1130 |
| Midpoint, Restricted Economy Fare Range (0.00 – 0.22) | 0.0920 |
| | |
| Mean Fare as Fraction of Maximum Fare | 0.1722 |
| | |
| Mean Fare Multipliers by Fare Class | |
| First Class Fare = Mean Fare x (.82/.1722) | 4.762043 |
| Business Class Fare = Mean Fare x (.296/.1722) | 1.718981 |
| Unrestricted Economy Fare = Mean Fare x (.113/.1722): | 0.656233 |
| Restricted Economy Fare = Mean Fare x (.092/.1722): | 0.534278 |

**6.9.7:** Another feature of the arasag set is airline *alliances*. Airlines join in alliances with each other in an attempt to reach a broader market by developing schedules that are mutually complimentary and therefore appealing to passenger whose travel requirements extend beyond the reach of any one carrier. For example, Alaska Airlines in the United States is a member of the One World Alliance, which also includes British Airways. Thus, for a passenger traveling from Anchorage, Alaska to London, England, the options offered by Alaska Airlines might be more attractive to a customer than an itinerary using carriers not within such an alliance. Moreover, alliance partnerships often include code sharing arrangements. The material effect of alliances on travel itineraries, however, is that the connection times at the intermediate airports are generally better than between carriers that not allied. This is represented in AirVM by using a different minimum connect time between segments when constructing itineraries (see Section 6.11). The minimum connect time is the minimum amount of time between the arrival of one flight and the departure of the next leg in an itinerary.

Nominally this is set at 90 minutes, but for alliance partners it is reduced to 45 minutes. (These values can be replaced if more accurate data is available).

**6.9.8:** Another aspect of multiple leg itineraries is the sharing of the fare revenue between the owners of the respective legs. The default methodology that is applied is a distance-proportional allocation. Specifically, the fare structure for a market is established either from external input data or from the default EDF approach described in **6.9.4**. Then the total distance travelled by the itinerary is calculated simply by summing the distances flown by the separate legs, and the fraction of that total distance flown by a specific carrier multiplied by the fare of a sold ticket to determine the share of the revenue earned by the airline operating that leg. If the user or client sets different fares as part of its analysis, then those fares are used.

**6.9.9:** As the ticketing process proceeds throughout a simulation, individual tickets are sold on individual flights. Tickets can be sold either on legs acting as nonstop flights between an OD pair, or as part of any of a number of itineraries which contain that particular leg. The respective arasags keep track of both the total tickets sold and the total revenue earned as part of their state vectors. Moreover, if designated by the simulation user, the *manifest* (the list of passengers holding tickets on a segment) of any segment or set of segments can be maintained during the simulation as part of the state vector for an arasag. This enables a client user to closely study the composition of specific flights as in terms of the traffic they carry and the markets they serve.

**6.9.10:** The basic data required to build the arasag set is the current airline operating schedule. This can be artificial, but actual schedules are easy to acquire and give a full picture of the world's airline network as it is available to the passenger customer. In addition, as indicated earlier, default values for fare, cabin capacity, and fare class structure are established for each arasag. While these default values are adequate for many AirVM studies (e. g. comparative analyses of alternative schedule modifications), increased accuracy can be gained by using data that describe real price and RM structures. This is generally available for its own flight segments for an airline that is a client of AirVM, but is almost never available for that carrier's competition (antitrust and fair competition statutes prohibit such detailed knowledge of pricing), so the default settings allow reasonable analysis without risk of legal consequences. There are other data that can improve the default settings, however, and still maintain required confidentiality. Fare EDFs can be extracted from the settled ticket data processed by bank settlement plans. These are organizations that act as intermediaries between travel agents and passengers, on the one hand, and airlines, on the other. They take the money, issue the ticket to the passenger, and periodically (e. g. daily) transfer the funds to the airlines. They also keep records of these transactions, and so have detailed information regarding observed demand in specific OD markets – prices paid, itineraries used, when the tickets were sold, group size, and so forth. Data from both the International Air Transport Association (IATA) and the Airline Reporting Corporation (ARC) have been cited in earlier discussions. Virtual M1nds, the company developing AirVM, is constantly working with these sources to improve the default estimates of the model parameters used in AirVM.

## 6.10: The Dsag State Vector

**6.10.1:** The dsag is the agent in AirVM that represents the ticket distribution system. There are several sources from which a passenger may shop and buy tickets. Of course, tickets can be bought directly from the airline, either over the web, by telephone, or at a ticket counter in the airport or at a store-front location. (This last has all but disappeared in the US and Western Europe in recent years.) But most tickets are not sold through this channel. Although it is diminishing as time goes on, the travel agent still holds the largest share of airline ticket sales, but no longer accounts for the majority (Forrester Research, 2007). Travel agents deal with global distribution systems (GDSs), of which there are only about nine worldwide. These organizations consolidate the offerings of the various carriers into a central database so the passenger can see the available options more easily. Also in the travel agent sales channel are the bank settlement plans (BSPs), which act as clearing houses for the funds and airlines. A third major channel is internet sales. Some of this is directly from the airline sites, where intermediary software applications marshal the availability across a range of airline web sites. Also, internet discount sales involve a reseller that will acquire surplus seat availability directly from the carriers and essentially auction their inventory off to the highest bidder at the last minute. And there are travel agents which operate almost exclusively with a web sales mechanism, such as Travelocity.

**6.10.2:** The costs of distribution are significant to the airline. A ten dollar booking fee charged by a travel agent can be up to 15% of the total cost of the ticket. Airlines, for example, prefer internet direct sales over telephone sales since it is substantially cheaper to maintain. Low cost carriers keep their costs down by avoiding expensive distribution channels, such as travel agents. And of course carriers prefer direct sales to agency sales because the options given to the customer by the carrier only range over its availability. This control over the choices offered the passenger led, some three decades ago, to antitrust problems that have a significant effect on the nature of the market data available to the industry even today. Within the United States, prior to deregulation of the airline industry in the late 1970's, reservation services (which evolved into the GDSs) were owned by airlines. Sabre Holdings, Inc., for example, was initially created as a division of American Airlines. Upon deregulation, these carrier-owned channels began offering tickets for flights offered by other airlines, as well the original carriers (and hence became 'global'). Naturally, these reservation service were partial to the flights offered by their owners, and so would offer only those flights to potential customers. To remedy this unequal access to the market, the US Government adopted regulations which required that the reservation services treat all carriers that elected to use them as a distribution channel equally. This was enforced by the periodic detailed reporting of the booking through reservation systems using a system called the *Marketing Information Data Transfer*, or MIDT. It is available for purchase by those in the industry,[73] and eventually became the industry standard for marketing data. The MIDT essentially prevented a reservation service from restricting what it offered the passenger, and enforced that equality by reporting for anyone to see the market share each carrier had of a specific market. By the first decade of this century, however, the reservation services had evolved into proper GDSs, were no longer owned by carriers, and the enforced equality of opportunity was no longer necessary. In 2003, the US Government rescinded the regulations that created the MIDT. While still available, and still quite costly, it can no longer be relied on to accurately portray the distribution channel market share.

[73] And the data is costly. The MIDT for one year for the entire world is estimated to cost over US$7,000,000 (Shepherd Systems, 2006).

**6.10.3:** The market share of each of the various channels varies widely around the world. In the United States, Western and Central Europe, and Australasia travel agent sales accounts for somewhat less than 50% if all sales, and is thought to be declining (Forrester Research, 2007). In India, agencies operate out of storefronts, and account for the majority of sales, with the remainder being internet. In Japan, airlines sell virtually all their seats to intermediate dealers, who actually apply revenue management techniques as they resell to the public. And channel share varies widely from market to market. If an OD market is served by a low cost carrier, for example, the travel agent share is often sharply reduced, since many low costs carriers do not offer their product through travel agents. Actual market share figures are not readily available in much of the world, and when such data is available in can be of questionable accuracy, since it is in the competitive interests of the carriers to keep such information closely guarded. The result of this secrecy is that no one channel can see what other channels are doing, except when they undertake research of their own to specifically determine share in specific markets. Data gathered in this fashion is of course proprietary. Carriers can see its own sales through all the channels it uses, but can only guess at what other carriers in the same market are doing. One channel cannot see other channels at all. The MIDT, once a reasonably accurate portrayal of the channel share, has deteriorated substantially since its *raison d'être* dissolved. This makes the estimation of OD demand extraordinarily difficult, since it is very hard to get accurate measures of observed OD sales.

**6.10.4:** Within AirVM, only one essential function of the distribution system is represented, because the resources necessary to build an extensive structural model of the intricacies of the distribution channels – carrier pricing, geographic coverage, etc. – are not yet available. For AirVM, therefore, the critical role of the distribution agent is the creation and management of itineraries. Itineraries, it will be recalled, are single or sequential groups of flight legs that accomplish the travel objective of the traveler in moving from an origin or destination. The segments in an itinerary can be all operated by the same carrier, or by carriers in the same alliance, or by different carriers not otherwise related. The provision of itineraries is the primary function of the dsag, and the data needed to construct the itineraries form most of the content of the dsag's state vector. Itinerary management within the simulation is carried out via the messaging architecture, so message management is the other important function of the dsag. The current version of AirVM also allows the creation of multiple dsags, differentiated by geographic and carrier coverage, but this feature has not been fully tested, and is not available to the general user. Finally, the dsag is also designed to be a client agent, in that user clients would provide the ratiocinator for a specific dsag. This role has yet to be fully implemented.

## 6.11: The Generation of Itineraries

**6.11.1:** As has been described, the dsag creates and maintains the itinerary set for use by the pags and arasags as they engage in the ticketing process. This section discusses the process and its products in more detail, establishing terminology required to fully understand the simulation architecture laid out in Chapter 7. As noted earlier, the source data from which the itineraries are built consists of a commercially-available global schedule description from a vendor such as Innovata or OAG. The data supplied is a simple flat file, usually with one record for each nonstop segment. That record contains all the data necessary to construct the itineraries, with the exception of minimum connect time, as noted below.

**6.11.2:** The itineraries that are maintained by the dsag actually do not contain leg details. Rather, these are kept in the individual arasag state vectors, and only references (memory addresses) are stored in the itinerary structure. This saves considerable storage space, since a given leg can occur in a number of itineraries. Obviously, all multi-stop itineraries are made up of two or more nonstop legs, so these elementary segments are the building blocks of all other itineraries, and preparing them is the first task of the itinerary generator. Given the format of the incoming data, this simply consists of loading the schedule in its vendor format, confirming that all airports and cities coded in the segment descriptions are valid, and creating a single leg record for each individual leg flown in a week. (Usually one record in the incoming data set contains a set of fields indicating which days of the week the leg is flown. This is expanded to create one segment record for each day.) The internal objects that represent the legs are then supplemented with default fare class data for the nonstop OD market served by that leg, including fare class capacities, RM protocols, and default pricing. They are then associated with the arasag representing the operating carrier, thus creating the arasag state vectors.

**6.11.3:** Itinerary generation using the depth-first search is the method used in the Global Market Allocation System (GMAS) used by Boeing Commercial Airplanes, which was developed by a team led by the author and is described Parker *et al.* (2005). However, this approach is not used in AirVM, for it is extremely time consuming.[74] By itself this is not a crippling issue – computing resources are inexpensive. But the same order of magnitude of performance is required whenever the schedule changes, such as when a client avatar alters his schedule to analyze the changes in share or revenue of such an alteration. This kind of performance hit, for even a small change, is severe when viewed in this operating context.[75]

**6.11.4:** AirVM does not generate itineraries with the depth-first method. Instead, it prepares the components of itineraries in a manner that allows the virtually instantaneous creation of pag choices as needed, rather than in advance. For the nonstop itineraries, this is of course trivial, but since these single-leg itineraries are the building blocks for multi-stop paths, some attention is paid to organizing them in a way that makes building more complex structures more efficient. First all the OD markets that are served by nonstop legs are identified. Then the set of nonstop legs that serve each of those markets is collected into the market *nonstop set* for that market. Two copies of each nonstop set are kept, one ordered by departure time and the other by arrival time. (Recall that the standard time period is a Monday through Sunday week. Times variables for a leg, and hence for an itinerary, are kept in units of minutes from 00:00 AM on Monday, so a flight departing at 23:50 on Saturday night, local time, has a departure time of 10,790.) When a pag then requests a selection of itineraries, the arrival/departure time sensitivity determines which nonstop set is used.

**6.11.5:** One-stop itinerary sets are composed of pairs of nonstops that connect. By connect is meant that the two nonstop legs share an intermediate airport. The first leg

[74] Nearly 2000 processor hours are required to produce the itinerary set used by GMAS. This is not an issue if enough processors are available, since the algorithm is linearly scalable. But even with 200 processors, ten hour are required.

[75] There are some ways to speeding up the process. One such method is called the Worthy Airport method (patent pending). In this approach, when the original schedule is created, the depth-first search process tracks all the airports visited before a branch of the tree leads to the destination or is abandoned as being infeasible. Then, when a change is made, only the schedules in those so-called *worthy* airports need be examined. This can reduce the schedule recalculation time by from 10 to 60%.

arrives at the airport in time for the passenger to move to the departure gate of the second leg, which departs sometime thereafter to arrive at the final destination. The minimum time between the arrival of the first and the departure of the second is the minimum connect time, which varies with the relationship of the carriers operating the two legs. However, notice that since the schedule repeats across the standard time interval, every initial leg will connect with every departing leg. For example if leg 1 from A to B arrives too late for a connection to leg 2, the copy of leg 2 departing 10,800 minutes from then – in the next week – does connect (albeit with an atrocious travel time). Therefore all that is needed to provide every one-stop itinerary in a market is ordered pairs of nonstop itinerary sets, the first being all nonstops from the origin to the common intermediate city, and the second from the intermediate city to the destination. The preparation for this is simply the creation of the one-stop pairs. The set of these pairs for a given market is called the *one-stop itinerary set* for that market. The itineraries that are offered to a pag, then, are the nonstop and one-stop itineraries that are closest to the ideal time for that pag. (Not all itineraries are returned. Recall that the scope determines how many are considered.) If the pag is departure sensitive, the departure time is used to define "closest" to the ideal, and the creation proceeds from the earliest to the latest leg. If the pag is arrival time sensitive, then the arrival-time-ordered nonstop set is used, and the creation proceeds in reverse order. A mixture of nonstops and one-stops can be returned.

**6.11.6:** As a practical matter, the vast majority of traffic between two cities is carried on non- or one-stop itineraries. If there are none such available, then it is because there is little demand for travel in the OD pair. However, there is sufficient demand across the world that two- and higher-stop itineraries are often used. Therefore, *n*-stop itineraries are necessary. They, of course, are created using combination of non-, one-, up to $n - 1$ stop itineraries, exactly like the creation of one-stops from nonstops. However, the number of possible itineraries grows beyond exponentially as the itinerary stop structure expands, quickly becoming impossible to manage with even the largest computer. So AirVM modifies the procedure outlined above in recognition of this relative rarity. The first simplification is to retain only a single set of two-stop itineraries, rather than all possible ones. A two-stop path can be made from matching a nonstop up with a path from a one-stop (in two ways, with the nonstop being either the first leg in the path, or the third), which is logically equivalent to a triple consisting of three nonstop sets. Thus all possible two-stops can be readily computed. However, during the computation loop only the one-stop/nonstop pairing that has the highest estimated pag utility is retained, since the demand for two-stops is relatively low. Three-, four- and *n*-stop itineraries are created in exactly the same way, with one added feature: If a market has non-, one- or two-stop itineraries, no additional itineraries are generated. There are no itineraries available in these sets (because they are all full), the rare itinerary process (described immediately below) is invoked to find higher-stop opportunities. If there are no lower order itinerary sets for a given OD, then itinerary set generation continues until with higher order combinations until exactly one set is defined. In any case, itinerary sets of order higher than four are not routinely created. If there are no itineraries available in the collection of itinerary sets, then the process for rare itineraries is invoked.

**6.11.7:** When a pag requests itineraries as part of the AirVM simulation execution, the dsag examines the itinerary set structure defined of the OD market in which the pag is traveling and requests availability from the arasags on appropriate legs. This is done during the simulation because the identification of available paths from the itinerary sets costs little in terms of performance. If there is no availability, either because all paths are full or none exist, then the dsag executes the basic procedure of generating *n*-stop itineraries from *(n – 1)*-stop itineraries as outlined above, but just for that pag on that journey. Since the pag ticketing process occupies only one thread in the computer's processor, it can go about this tedious task while other pags continue to be processed. This is the rare itinerary process.

**6.11.8:** The effect of generating itineraries using the approach set out above addresses the performance issues mentioned in **6.11.3**. The initial process to create the itinerary sets from the schedule data requires approximately four hours of single processer time, as opposed to the 2000 hours needed for the depth-first search approach. More importantly, changes in the schedule – which imply direct changes in the relevant itinerary sets – are reflected almost immediately in the dsag and arasag state vectors, thus effectively eliminating the performance barrier for client input that is associated with the addition, deletion or timing change of a flight.

**6.11.9:** The collection of arasags and their respective flight legs, and dsags with their itinerary sets is called a *scenario*. That term is used because, from a carrier client point of view, the flights and their connections, pricing, RM, and so forth represents the global airline network, and it is that resource that the client agent will want to manipulate to engage its appropriate narrative, and scenario seems the appropriate label for this aspect of the simulation. When a client is actually operating the AirVM application (discussed in Chapter 7 below), it is given a so-called *seed scenario*, which is the scenario that is derived from a single schedule without any changes. The client never modifies a seed scenario directly, but rather creates a *child scenario* from the seed, and makes changes in the child scenario. Children can also be derived from existing child scenarios, forming a hierarchy of related scenarios. The scenario from which a child is derived is naturally enough called its parent. This arrangement guarantees that scenarios maintain comparability across a wide range of analyses.

**6.11.10:** Scenarios are comparatively large data objects, typically requiring roughly 300 Mb of disk storage in the current AirVM software configuration. The data in Table 6.5 gives some detail of the number of itineraries that can be examined in a seed scenario based on a schedule from August, 2008.

**Table 6.5: Summary Statistics for Typical Seed Scenario**

| | |
|---|---:|
| Seed Scenario Name | World Aug08 |
| Standard Week | 11 Aug 08 – 17 Aug 08 |
| Cities served with regularly scheduled flights | 3,956 |
| Unique segments (per week) | 1,087,131 |
| OD markets | 11,012,422 |
| Nonstop markets (best service) | 38,424 |
| One-stop markets (best service) | 723,132 |
| Two-stop markets (best service) | 3,948,632 |
| Three-stop markets (best service) | 4,213,598 |
| Approximately passengers (per week) | 47,000,000 |

## 6.12: Summary

**6.12.1:** This chapter delineates the stochastic models used to create the synthetic populations of pags, arasags and dsags in AirVM. Arasag and dsag synpops are generated by actual field data, and so the stochastic elements are concerned with the values in the state spaces for these agents. For example, the EDF method is used to assign default fares to flights maintained by an arasag.

**6.12.2:** The pag agent synthetic population is a true synthetic population in the sense that the agents are created by the virtual market and not defined by actual people. And a number of stochastic processes are needed for that creation. Specifically: 1) pag origin and desired destination; 2) trip purpose; 3) journey length; 4) the pag itinerary choice mixed logit model parameters; 5) ideal departure and arrival times; 6) travelling group size; 7) ticketing instant; 8) cancellation model parameters; and 9) willingness-to-pay.

**6.12.3:** With the synthetic population structures defined, attention can now turn to the computer simulation program itself. That is the topic of Chapter 7.

# Chapter 7:

# The AirVM Virtual Market Computer Program

## 7.1: AirVM Computing Architecture

**7.1.1:** This chapter describes the computing logic flow of AirVM. It covers in broad brush the entire system, but drills down in detail when necessary to delineate important methods that are applied in the execution of the simulator. More routine tasks, such as result reporting or user input, are only discussed briefly since implementation of such capability can be accomplished a number of ways, and no unique constructions are required. I have written all the code currently in AirVM. Figure 7.1 in this chapter is the program flow chart. Each process identified in the flowcharts is described. When a process in a flowchart has a grey background, then there is another flowchart that takes that process down one logic level to a more specific description. Also, processes are discussed in numerical order, not necessarily in the logical order specified by the flowchart. Processes in a flowchart are numbered from upper left to lower right, for ease in locating the entry within the flowchart.

**7.1.2:** The numbering in the following sections that are shown in italic refers to the flowchart tasks shown in Figures 7.1 which is broken into ten sub-figures, 7.1.1 to 7.1.10. In addition, in the discussion below the description of each logic level is indented, where appropriate, to help distinguish process description from the main course of the dissertation. The four major process groups are in AirVM are shown in Figure 7.1.1. Each of these is delineated at level two logic, as well. Those discussions proceed in numerical order. When one of these processes warrants an additional level of logic, it is discussed under that general topic. One process requires a level four logic diagram (Process *3.11.7*, the Pag Choice process) and it is presented in numerical order under the discussion of process *3.11*).

> *1. Preprocessing, Calibration and Estimation (PCE):* This logic covers the preparation of the raw data, the calibration of mathematical models (estimating the parameters of such models or EDF's, as appropriate), the estimation of OD demand, and the creation of synthetic populations of pag, arasag and agents.

These activities are carried out by the system administrator, and are not available to system users. The system can exit from this process, or continue on to carry out further work. Figure 7.1.2 contains the level two logic for process *1*.

*1.1 Initialize:* Establish initial value for several internal parameters, such as location of key files. Most are read from files where default values are stored. Standard computing application techniques are used.

*1.2 Seed Scenario Creation.* Nothing can be done in AirVM without a scenario. It is built from the published airline schedule vendors. The seed scenario built here is one from which all scenarios generated by users descend. AirVM uses the itinerary generation process described in Section 6 to produce the components needed to create itineraries during a simulation. The level three logic of Figure 7.1.6 describes this process in detail.

*1.2.1 Load schedule data:* The data from a commercial schedule source which describes the world's flight offerings for a given, standard time period (nominally a specified week) are loaded into AirVM. Standard data base techniques are used. In this task, carriers operating for the time period are also recorded, and arasags are created.

*1.2.2 Validate cities against schedule:* The flights in the airline scheduled are compared with a list of known airports in the world to determine which OD are served by that schedule. This results in an (empty) OD matrix.

*1.2.3 Create nonstop itinerary sets; sort by depart/arrival times:* For every OD that is served by one or more nonstop flights, a list is prepared of every flight that serves that OD. This is called a *nonstop itinerary* set. Two lists are prepared: one is sorted in ascending order by departure time (defined in terms of minutes from the start of the standard time period), and the other is sorted in ascending order by arrival time. Two lists are prepared because pags can be departure or arrival time sensitive.

*1.2.4 Create all one-stop pairs; Sort by average disutility:* Nonstop itinerary sets can be matched up (if the destination of one matches the origin of another), and the pair is a set of itineraries from the first's origin to the latter's destination. Every pair of flights, one from the first element of the ordered set, and one from the second, is a one-stop itinerary. This is true because the schedule is assumed to repeat once per standard time period, so multi-stop itineraries "wrap" around the end of the period. Many such pairs exist for most OD's, so the system sorts them by average passenger disutility, which speeds the search for nonstop itineraries.

*1.2.5 Itinerary n-stop computation loop control:* As one-stops can be created from nonstops, two-stop itineraries, three-stop itineraries, etc. can be created from lower order itinerary sets. However, the number of possible two-stop and higher-order itineraries far exceed the feasible demand for travel in that particular OD, so not all possible combinations are kept. Only the single combination with the lowest average passenger disutility is retained.

*1.2.6 n-stop needed branch:* If an OD has non- or one-stop itineraries, only two-stop itineraries are calculated. No further multi-stop itineraries are added. This is because the economics of the industry virtually guarantees, as a practical matter, that three-stop and higher itineraries will not be needed if lower order itineraries exist since demand will be met by those lower order itineraries. So higher-order multi-stop itineraries are only defined if there are no lower order itineraries available. This branch determines if such additional itinerary sets are required. If they are, task *1.2.7* is executed. If not, control moves to *1.2.8*. This branch is implemented with standard techniques.

*1.2.7 Find lowest disutility n-stop.* This task examines the possible *n*-stop itinerary sets and retains the one with the lowest average passenger disutility. This approach, as far as I have been able to discover, is unique in the industry.

*1.2.8 Set default price distribution and RM protocols, dsag alignment:* After the itinerary sets are determined, each market is give a default fare class price distribution. This is an empirical distribution function description of the fares charged for markets of this type, as described in **6.9.5**. No such approach has been found elsewhere in the literature.

Also, default revenue management (RM) protocols are assigned to each flight in the scenario. These can be changed by the user. Finally, the default alignment between the number of dsags and the carriers and cities they serve is set. At the present time, only one dsag is used in AirVM.

*1.2.9 Store the seed scenario.* The seed scenario is stored in an internal, proprietary data structure. Several seeds can exist, generally one for a different standard time period. The parent-child structure of the scenarios guarantees comparability during comparative data analysis.

*1.3 OD Demand, Synpop branch:* Two options are available once a seed scenario has been created. The system administrator can create a new OD demand matrix, or define and create a synthetic population (synpop).

*1.4 Iterative Demand Estimation:* Passenger demand by origin-destination city pair is required for the operation of AirVM. The demand matrix can be supplied externally. If, however, adequate data of the right type is available, AirVM can produce an estimate of that demand. The procedure used for developing this estimate is unique in the airline industry. This is an implementation of the analysis of Appendix M, and the logic of the process is outlined at level three logic in Figure 7.1.7.

*1.4.1 Load and clean ticketing/booking data:* The data needed for the iterative OD estimation process is either ticketing or booking data from a source such as an airline, a bank settlement plan, a global distribution system or other such source. It must contain records of individual ticket purchases, indicating the size of the group traveling, when the ticket was purchased, what itinerary was purchased, and how much was paid. This task controls the loading and cleaning of this data. The ODs from the OD matrix that have data in the data set is used to build the so-called “OD with Data” list.

*1.4.2 Estimate Ticketing Curve Parameters:* Using a compound, non-homogeneous Poisson counting process as a model, as described in Section 6.6, the required parameters of that model for each OD in the OD with Data list is estimated, using standard statistical techniques.

*1.4.3 Estimate Group Size Distribution Parameters:* A mixed, truncated Poisson distribution is used to represent the size of the group addressed in a particular ticketing instance, as discussed in the previous chapter in Section 6.5. Parameters for this model are estimated directly from the data using standard statistical techniques.

*1.4.4 For each OD with Data control loop:* Loop control which manages tasks *1.4.5* through *1.4.12*.

*1.4.5 Determine itinerary impute set:* Knowledge of the probability that a passenger will choose one of a set of itineraries coupled with knowledge of the total number of passengers choosing a single itinerary of that set allows the imputation of the total demand in that OD, using the simple relation that demand for a specific itinerary is equal to the total demand times the probability of that itinerary. Thus, if data on the ODs of passengers on a give flight is known, then data is known on the demand that must be being served by other itineraries that serve those same ODs. The set of ODs the meet this criterion is called the *impute set* for a given flight.

*1.4.6 Find First Itinerary Closure Time:* The number of passengers on a particular flight traveling in a specific OD market is not only a function of the probability of itinerary choice, but also of the availability of itineraries. If an itinerary is closed during the ticketing period, it is no longer available to later-ticketing passengers, and this distorts the allocation of passengers to the itineraries, invalidating the imputation discussed above. However, prior to the closure of the first itinerary in a market, the numbers of passengers represent the choice probabilities, and calculation of demand at that point is valid. This is point in time that the first itinerary becomes no longer available is called the *First Itinerary Closure Time***,** or *FICT*. The FICT is discernable by calculating parameters of the stochastic process describing the expected ticketing rates (steps *1.4.2* and *1.4.3*) sequentially through the ticketing process and noting when the parameter changes beyond a specific threshold. To my knowledge, this methodology has not been reported in the literature, and is believed to be unique to AirVM.

*1.4.7 OD imputation control loop:* This loop controls the calculation of the imputed demand for each OD market for which data is available.

*1.4.8 Impute demand:* The demand at departure is estimated using the FICT total and the ticketing curve/group size models for the specified market.

*1.4.9 Exit control of OD imputation loop:* Exits when all ODs in the impute set are handled.

*1.4.10 Exit control of the OD with Data loop:* Exits when all ODs with Data are processed.

*1.4.11 Expand ODs with Data List:* The results of the estimation process above includes the determination of additional ODs for which demand data is now known. The flights that serve those ODs also serve other ODs not originally in the ODs-with-Data List. This step adds those ODs to that List.

*1.4.12 Exit control of OD with Data loop:* If there are no more ODs added to the list, the process moves to step *1.4.13*.

*1.4.13 Compute OD demand distribution parameters:* OD demand is not a constant, and the imputation process can produce a set of demand estimates for a specified market. This step characterizes the distribution of the demand estimates by the computation of the first four moments of the empirical distribution, so that this natural variability can be accommodated in AirVM.

*1.5 Synthetic Population Generation:* Once a seed scenario has been established, and an OD demand matrix determined, the pags that will be used in the simulation can be defined. A pag represents one or more tickets being sold to passengers on to travel in a specific market. The number of tickets associated with a pag is the group size. The level three logic of process *1.5* is illustrated in Figure 7.1.8.

*1.5.1 Load OD Matrix:* The OD demand matrix for this synthetic population is loaded from disk storage at this step, using standard techniques.

*1.5.2 Load Pag Parameters:* The file containing the parameters for the models of pag state space variables are loaded from storage. These include the parameters associated with ticketing curves, group size distributions, trip purpose and journey structure probability distributions, willingness-to-pay discrete choice parameters, ideal departure and arrival time distributions, and the mean and standard deviation of the coefficient variables for the four itinerary choice models maintained by each pag. See Chapter 9 for a complete discussion.

*1.5.3 Compute number of pags required.* The OD demand matrix determines how many pags are to be created, being equal to the sum of the entries in that matrix.

*1.5.4 Pag pre-generation control loop:* The creation of pags is a two-step process. First a "pre-pag" data structure is created. This is then used to create the full pag. This is the control loop for the pre-generation.

*1.5.5 Randomly set OD:* The origin and destination of the trip to be requested by that pag is randomly set by applying a uniform random number to the discrete probability distribution of OD travel implicit in the OD matrix.

*1.5.6 Randomly set Trip Purpose:* Using standard randomization methods, a trip purpose is assigned to the pag for this trip.

*1.5.7 Randomly set ticketing instant:* The time before departure that the pag will request ticketing – referred to as the *ticketing instant* – is now set using standard randomization methods. However, the time span from which the ticketing instant is drawn is unique to AirVM, and, it is believed, enables a method of simulation timing unique to the agent-based modeling field in general, and specifically unique to airline passenger market simulations. The ticketing instant is measures in $10^{th}$s of a second following the commencement of the ticketing period, which is nominally 90 days before the beginning of the standard week. It continues to the $97^{th}$ day following commencement, allowing for ticketing during the week-long standard time period. The ticketing instant is drawn from a probability distribution representing the ticketing stochastic process for the trip purpose and the OD being considered.

*1.5.8 Exit control for pag pre-generation:* Exits from loop using standard procedures.

*1.5.9 Sort pags by ticketing instant:* This step sorts the pre-generation data structure by ticketing instant, from earliest to last. When sorted this way, the pag ticketing instant becomes the simulation clock, as discussed in process *3*, simulation execution.

*1.5.10 Full pag creation loop:* This is the control loop for the completion of the pag generation process.

*1.5.11 Randomly set group size:* The number of people in the group that will be ticketed by the pag is set using standard techniques applied to the group size distribution for the specific trip purpose or OD for that pag.[76]

*1.5.12 Randomly set choice model parameters:* Using the incidence distributions for the parameter supplied in step *1.5.2* for the pag state space, the coefficients of the itinerary probability distribution needed for the discrete choice model are assigned to the pag. Using the incidence distributions of ideal departure and arrival times, the ideal times for this pag are also randomly selected.

*1.5.13 Randomly set willingness-to-pay:* Using the WTP model, a maximum fare willing to be paid by the pag for a ticket is chosen.

*1.5.14 Store pag in random access file:* The pag is stored in a random access file for use by the simulation execution processes. It is important to note that the pags are stored in ticketing instant order, with the earliest ticketing agent being first.

*1.5.15 Exit control for full pag creation loop:* When all pags have been created, the synthetic population creation process is complete, and the routine exits.

[76] Note that this implies that group size and ticketing instant are independent of one another. That hypothesis has yet to be tested empirically.

*1.6 Continue control for PCE processing:* The PCE can be entered in two ways: as part of the initial set up of a scenario for future processing, or as a result of an edit to the OD demand matrix. If from the OD edit, control exits to the scenario definition and editing functions of AirVM (Process *2*). If not, AirVM closes down.

*2. Scenario Definition and Editing:* This activity includes the specification of user-modifiable parameters of the system, including child scenario definition, (schedule flight addition, deletion and change, revenue management protocols) and editing of the OD demand matrix. It is here that the client user represents his narrative to the arasag or dsag as it will be engaged by the virtual market simulation – the arasag or dsag becomes as avatar. Here the user also sets the so-called *focus*, which determines what kinds of output are initially displayed after the completion of a simulation run. If the editing included changing OD demand values, then control returns to the PCE process to allow creation of an appropriate synthetic population for the altered demand matrix. The level two logic for these processes is illustrated in Figure 7.1.3.

*2.1 Select or create scenario branch:* At the outset of this process, the user chooses to operate on an existing scenario, or create a new one. This branch handles that choice.

*2.2 Create scenario from parent scenario:* All scenarios are derived as children from a seed scenario, or children of such children. This insures that all scenarios that are descended from the same parent that are eligible for comparison are, in fact, comparable, since they are based on the same state of reality contained in the seed scenario. Users may create child scenarios, and children of child scenarios, etc. to any extent they wish to organize and capture key changes are under study.

*2.3 Select scenario family and scenario:* If an existing scenario is to be used in the analysis, it is identified here. Selection of a scenario also identifies the family from which it derives, and siblings of the selected scenario are available for comparative analyses.

*2.4 Edit select branch:* This logic branch is where the user chooses which edit function to engage. The user can engage editing operations repeatedly, so control returns here after each function is complete.

*2.5 View/edit dsags:* Here the user can change the allocation of carriers to dsags, and change the cities and pags (customer base, from the dsag perspective) associated with a specific dsag.

*2.6 View/edit arasags:* In this operation the user can change the schedules (add, delete, change timing or equipment) for the flights associated with an arasag associated with a carrier, or alter the revenue management protocols a carrier applies to a specific flight or group of flights. Individual flights, groups of flights, flights in a specified market of collection of markets, or flights operated by a specified carrier or group of carriers (such as an alliance) or utilizing a specific airport or group of airports can be modified at the same time.

*2.7 View/edit market demands:* The user can change the mean and standard deviation of the market demand for any entry in the OD matrix. However, if this is chosen, the available synthetic populations are no longer valid for this OD matrix, and control returns to the synthetic population process in PCE (Process *1.5*) for creation of a new synpop. In this case, the PCE process exits and returns to the editing control branch. Notice that the creation of a new synpop has no effect on the scenarios.

*2.8 Define focus:* In order to organize the massive amount of output that is produced by an AirVM execution, the user is offered the opportunity to define a focus for an analysis. The focus is the set of output the user is most interested in seeing when the simulation is completed, and those are the results that are initially displayed when a simulation run is finished. This is part of the user specification of the perspectives and portrayals element of the simulation. The focus can be a flight or set of flights, a carrier or set of carriers, an airport, a city, or a market or set of markets. From the focus results display, the user can move to any of the other portrayal opportunities to explore the results of a simulation run.

*2.9 Edit loop control branch:* If OD demand edits were done, control shifts to the synpop processing step (Process *1.5*). If not, it returns to the Edit Select branch.

*3. Simulation Execution:* This process contains the actual tasks associated with the execution of a simulation itself. A simulation run takes a scenario, and applies to it a synthetic population, resulting in the production of an *outcome*. An outcome is the loading of every passenger that is expected to fly during the simulated time period on every flight operated by every scheduled airline in the world. The outcome is a database much like the scenario, except that all flights have manifests of passengers and totals of revenue, from which flight, carrier, market, distribution agent, and various comparative analyses can be performed. The simulation execution is the heart of AirVM, and, as far as is known, the entire process is unique among market analysis products available in the airline industry. Figure 7.1.4 sets out the level two logic of the execution process.

*3.1 Define outcome:* To begin the simulation process the user defines the outcome that will hold the results. There is exactly one outcome per combination of synpop and scenario, except for a Monte Carlo simulation, where there is one outcome for each combination of scenario, synpop, and Monte Carlo simulation cycle. Currently, the capability of saving a partial outcome is being studied.

*3.2 Select pag synthetic population:* The user chooses a synthetic population from those available from the OD matrix associated with the chosen scenario. The user can also choose to oversample (randomly cause some pags to attempt ticketing more than once) or undersample (randomly cause some pags to be skipped in the ticketing process) in order to easily allow exploration of demand jumps or sharp declines. Oversampling or undersampling can be applied system-wide or to specific ODs or groups of ODs.

*3.3 Set monitors control branch:* Monitors are screens that are visible during the simulation to show the status of some part of the scenario during the simulation execution. There are five available monitors, described below, and this branch controls the flow of logic to set up each. The user can invoke as many monitors of each type as he wishes. Each consists of a window which displays appropriate data using standard methods. The results shown in any monitor can be saved to a .csv file if desired. Monitors are also set up automatically when a user defines a focus, as they are the mechanics by which the focus feature is implemented.

*3.4 Monte Carlo control branch:* A Monte Carlo simulation is a repeated execution of the same simulation on the same scenario except that some variable in the simulation is varying at random from one cycle of the Monte Carlo to the next. The result is a probability distribution of the values of some of the key variables resulting from the simulation, such as the total revenue generated by the flight. This is a standard form of Monte Carlo simulation. Two forms of Monte Carlo are currently supported. One, called *inherent variation*, is the case where each simulation cycle is exactly like every other one, except that the choice of itinerary is varying according to the probability associated with the discrete choice model. The second is called *demand variation*, and randomly changes the OD demand for a market or set of markets between simulation runs. The probability distribution found using the Iterative Demand Estimation (Process *1.4*) is the basis for the randomly varying demands. The user can also edit the OD demand matrix to set other values of the demand distribution. Monte Carlo simulation results are not analyzed within AirVM. The results (usually as saved monitor results) are analyzed externally. Support for other forms of Monte Carlo – such as random variation in fare class pricing – is under development as of this writing.

*3.5 Set flight monitors:* The specific flights that the user wants to monitor are selected here. As tickets are booked on a monitored flight, the monitor is updated to show the number of tickets, the revenue value, the market being traveled in, if the ticket is cancelled, and so forth.

*3.6 Set market monitors:* All the ticketing activity in a specified market can be monitored, and the display is updated like the flight monitor as tickets are sold or cancelled.

*3.7 Set airline monitors:* This form of monitor displays changes in the tickets sold by a specified airline.

*3.8 Set dsag monitors:* This form of monitor displays changes in the tickets sold through a given dsag. Dsag monitors have not yet been implemented.

*3.9 Set passenger monitors:* Using the sequence number of the pag, a screen will display when a specified pag purchases a ticket (if it can), or when it cancels. The available itineraries from which it can choose are shown, along with the respective probabilities, as well as the details of the discrete choice model used and the itinerary chosen.

*3.10 Set Monte Carlo parameters:* If a Monte Carlo simulation is desired, this step is where users set the parameters of that simulation. The number of iterations of the simulation is established, and the type of Monte Carlo is fixed. If a demand variation simulation is to be done, the ODs in which demands are to vary are also designated. Monte Carlo simulations allow the storage of only monitors, to save storage space, since otherwise a complete outcome would be saved at each iteration, and that could amount to several terabytes of data. As noted above, analysis of Monte Carlo results is done outside of AirVM.

*3.11 Simulation execution:* This is the process that contains the actual execution of the simulation. It is described in detail below. The simulation takes some time to execute, depending on the hardware configuration supporting the system. As a benchmark, a single Intel quad processor can produce a single simulation with roughly 42,000,000 pags – representing about 57,000,000 passengers – in approximately 45 minutes. The simulation architecture in which the ticketing instant is used as the simulation clock is a very significant improvement over other methods of timing the cycles of an agent-based model. The use of this unique method reduces simulation execution time by over 95%. During the execution, the user can pause and restart the simulation (to study a monitor, perhaps) or abort the simulation (in case something was miss-specified). Figure 7.1.9 shows the level three logic of step *3.11*. [Note: In Figure 7.1.9 and Figure 7.1.10, which describes the pag itinerary choice methodology, hexagonal boxes are used to designate processes under separate control of other agents. Communications between autonomous objects are shown with a dashed line.]

*3.11.1 Pag ticketing process control loop:* Each pag is loaded sequentially from disk storage and processed semi-independently of the others. The number of pags that can be processed simultaneously depends on the number of processor threads available. One pag occupies a single thread, and each thread runs independently of the others, except for semaphore-based interaction with the scenario during availability and ticketing. The processing is not quite simultaneous, however, since the assignment of pags to threads occurs in ticketing instant order.

*3.11.2 Post clock message:* The ticketing instances of the pags acts as the simulation clock. The pags, arasags and dsags communicate with each other using the message/event protocols provided by the computer operating system. They are standalone objects that function in separate threads, so the messaging process is well defined. Both the arasag and dsag execute activities depending on the time. Arasags, for example, usually execute revenue management protocols every few hours, or daily (in simulated time). Dsags must clean ticketing request queues periodically (daily) to prevent lockout. Normally these processes would be governed by an external system clock. In AirVM, this clocking function is provided by the ticketing instant. This is substantially more efficient than an external clock, and has no effect whatsoever on the simulation logic, since everything is driven by passenger action. As far as is known, this is a unique characteristic of this agent model, but one that could be extended to other agent models without difficulty.

*3.11.3 Arasag RM processing:* The arasag agents in AirVM execute their RM protocols periodically, and this process occurs here. The exact procedure depends on the specific protocol, but they are standard practices in the industry, and nothing unique is contained in the processing.

*3.11.4 Dsag queue cleaning:* Dsags maintain the communications between pag and arasag, and during the normal operations of the simulation the message queues can get corrupted by the asynchrony of the messaging process. Periodically, the dsags empty the message queues as a housekeeping function. This is handled in this operation, using standard procedures.

*3.11.5 Scan ticketed pags for cancellation:* All pags that have purchased tickets are processed through the cancellation algorithm to see if they are canceled. If the cancellation occurs, the pag is posted to an available thread so that cancellation notification can occur.

*3.11.6 Post pag to available thread:* This process is a scheduler that monitors the available threads and sends the current pag to the first one available. Once the pag has been posted to a thread, control passes to the pag ticketing loop exit control (*3.11.8*) and, if more pags are waiting for ticketing, the next pag is prepared for processing. If all pags have been processed, the post function waits until the last thread is complete, and then passes control to the exit control, which exits the loop and passes control to *3.11.10*.

*3.11.7 Pag itinerary choice process:* In this procedure, the pag requesting tickets interacts with the dsag, which in turn interacts with the arasags. The logic of this process is central to the operation of AirVM, and is described at level four logic in Figure 7.1.10.

*3.11.7.1 Post itinerary request message:* The pag posts a message to the dsag indicating it wants a fixed number of tickets in the specified market at or below a set willingness-to-pay threshold. It also indicates if it is a business or leisure trip, if the pag is arrival or departure time sensitive, and the ideal times for the pag.

*3.11.7.2 Dsags request availability:* The dsag takes the pag request and identifies the itineraries in the market which meet the pag's requirements. A fixed number of possible itineraries – the scope – are identified. That number can be set by the user. The dsag then posts messages to the affected arasags requesting availability of tickets. The dsag accumulates the arasag response messages and, when all have responded, packages them up and posts them to the pag message queue.

*3.11.7.3 Arasags availability response.* The arasags examine the flights requested and respond to the dsag with an availability message.

*3.11.7.4 Receive itinerary list:* The pag retrieves the itinerary list posted by the dsag, and examines it for availability.

*3.11.7.5 Availability branch:* If there is no availability for the desired travel, the pag records that fact and exits the thread. Availability may not exist if there is no room on a flight or if no flight meets the pags willingness-to-pay threshold.

*3.11.7.6 Random itinerary choice:* Using the discrete choice random utility function of the pag, the probability of each available itinerary is calculated, and then a random number generator is used to select one of the itineraries for ticket purchase. The chosen itinerary is then posted to the dsag message queue.

*3.11.7.7 Dsag requests ticketing:* The dsag posts a message to the carriers operating flights in the chosen itinerary requesting tickets. Carriers operating flights in itineraries that are not chosen are also notified, so that reserved seats can be released if that is the practice of the arasag. The dsag then awaits response by the carriers. If purchase confirmation is received, the appropriate message is posted to the pag message queue. If, on the other hand, availability is no longer available, a message to that effect is posted.

*3.11.7.8 Arasags respond to purchase request:* The carriers determine if seats are still available for the requested flights (they could have been taken by in intervening pag operating in another thread), and returns an appropriate message to the dsag.

*3.11.7.9 Ticket purchased branch:* If the ticket purchase is not confirmed, control returns to the pag itinerary request process (*3.11.7.1*), and the request process is repeated. If not, the ticketing process moves to completion. Request processing continues until all possible itineraries are exhausted, at which point the pag is marked "no availability," and is not ticketed.

*3.11.7.10: Record choice:* The chosen itinerary is recorded for this pag, and the thread exits.

*3.11.8: Pag loop exit control:* If all pags have been processed, the simulation is complete, and the wrap-up processing begins. If not, control is passed to the monitor update task.

*3.11.9: Update status, display monitors:* The status bar on the simulation control screen is updated, and the monitors, if any, are updated with the relevant new data.

*3.11.10: Store outcome:* The result of the simulation, contained as the outcome data set, is written to permanent storage.

*3.11.11: Display focus results:* The results from the outcome that are defined in the focus of the scenario are now displayed. This signals the completion of the simulation, and the user can move from the focus displays to drill deeper into the simulations results.

*4. Analysis of simulation results:* This set of processes include a number of standard screen formats to display the results contained in the outcome of a simulation run, or to compare the results of two or more runs. The user can examine the results, starting from the focus report, as he wishes, drilling down to such detail as necessary. The user can also use the analysis results to design new scenarios. The user can also open AirVM, select a scenario, and if one or more outcomes exist for that scenario, move directly to the focus results displays and proceed to execute any of these analyses. All displays can save the results as text or .csv files for analysis with other programs. The architecture of this process is straightforward and uses standard methods. Figure 7.1.5 shows its level two logic.

*4.1 Result display branch:* This represents the display choice of the user. Control returns here for additional selection until the user exits.

*4.2 Flight-specific results:* These displays show the loads and revenues for selected flights, all flights operated by a selected carrier, or all flights out or into a selected airport. Complete details of the flight – loads by cabin and fare class, revenue by cabin and fare class, and load factors and yield data are displayed. In addition, upline, midline, downline and local loads are identified.

*4.3 Carrier-specific results:* These displays show the loads and revenue results for itineraries which contain flights operated by one or a specified collection of carriers. In addition to data as discussed in *4.2*, load and revenue share information is also provided.

*4.4 Market-specific results:* These displays show the itineraries used in a specified market. Data like that mentioned previously is portrayed, but load and revenue share for each itinerary are also shown.

*4.5 Airport-specific results:* For a specified airport, the flights in and out of that airport are displayed as a function of the operating week. Dwell numbers – the number of individuals in the airport waiting to board, in the process of changing flights, or deplaning and departing the airport are also shown as a function of time of day. This capability has yet to be completely implemented.

*4.6 Comparative results:* A series of displays stand flight, carrier, or market results of two sibling (comparable) scenario outcomes side by side so the user can compare the effects of one or more changes to a scenario.

*4.7 Benefit/cost analysis:* This result display accepts cost data (in a general form) from the user and then shows the benefit/cost ratio of individual flights, itineraries, markets or carriers.

*4.8 Export to database:* This operation writes the outcome of a simulation to an external database in MS-SQL format. Other database structures can then be supported from this common structure.

**17.3:** Other capabilities, especially analysis capabilities, are constantly under development. None of these analyses, however, are in any way unique. Rather, they are more than likely tailored to the needs of specific clients, such as the financial industry. This flowchart discussion does not describe these emerging additions to AirVM.

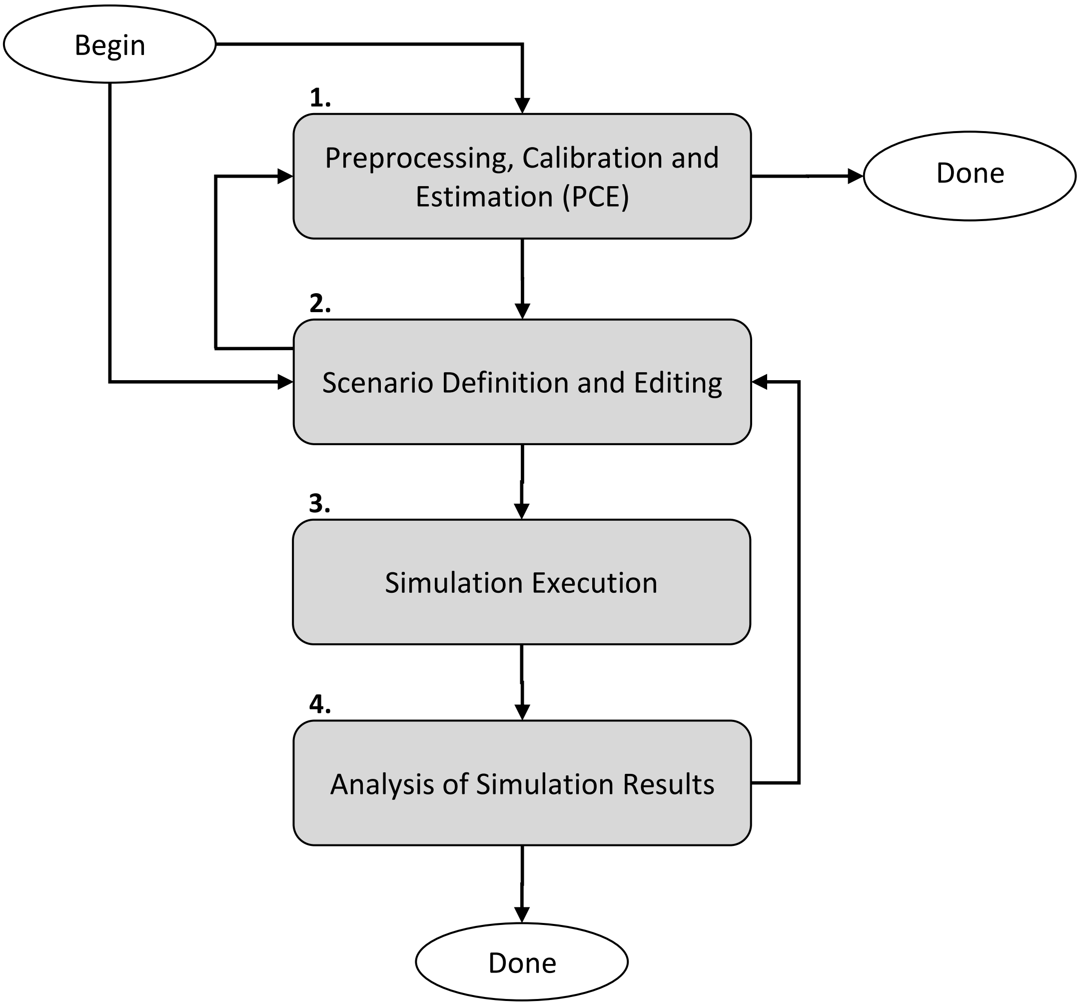



**Figure 7.1.1: AirVM Program Logic Flowchart. Level 1 AirVM Logic**

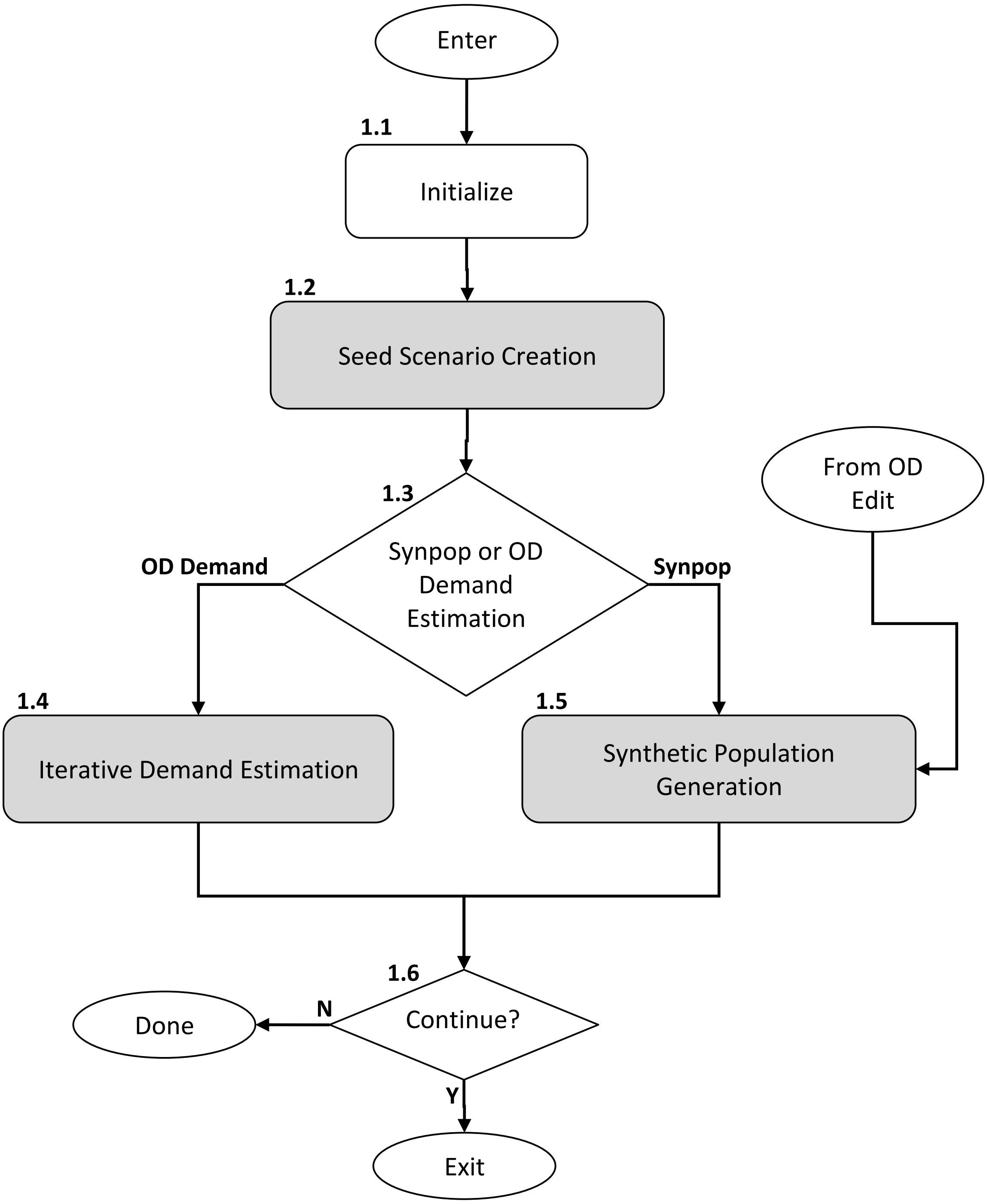



**Figure 7.1.2: AirVM Program Logic Flowchart. Procedure 1. Preprocessing, Calibration and Estimation (PCE). Level 2 Logic**

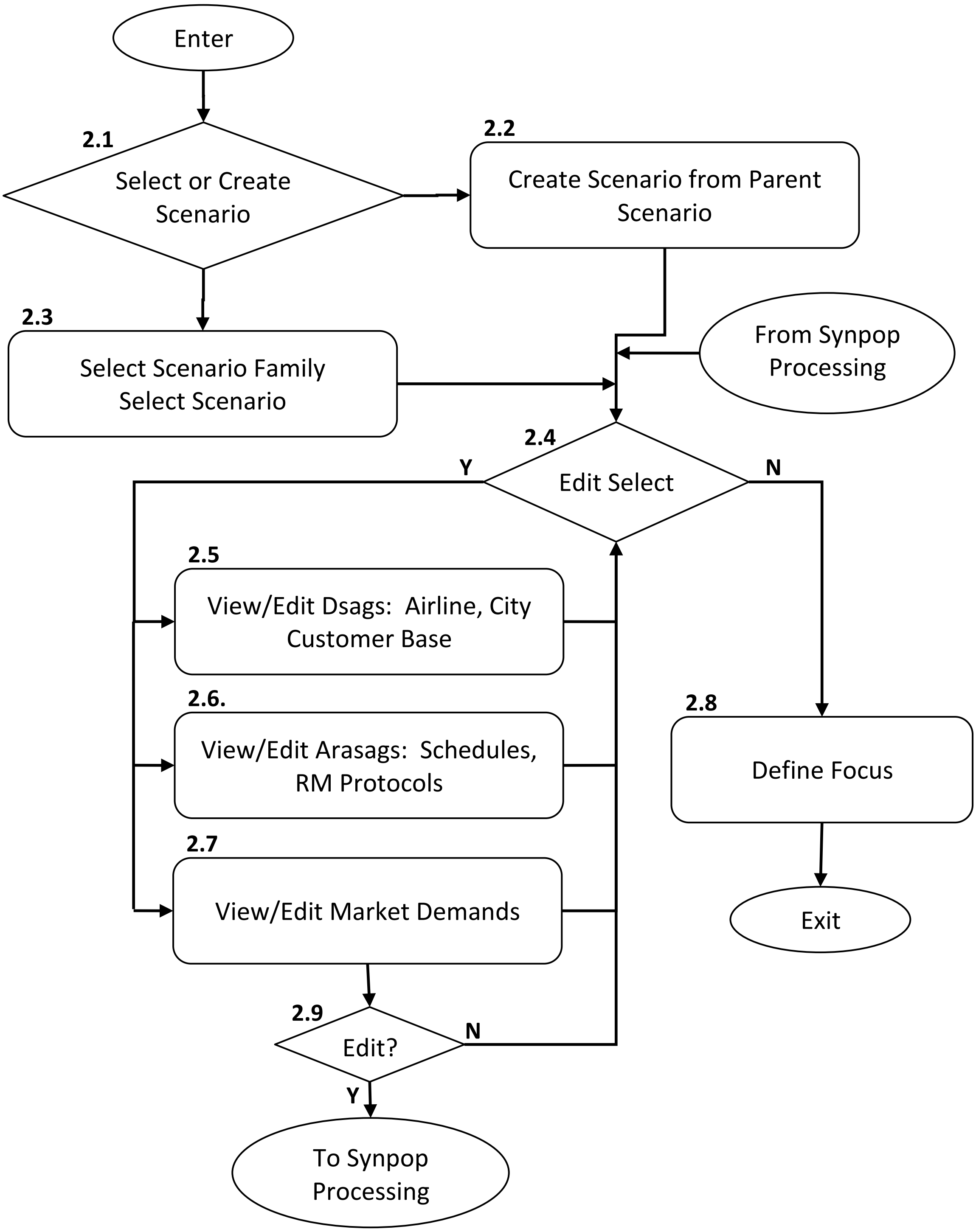



**Figure 7.1.3: AirVM Program Logic Flowchart. Procedure 2. Scenario Definition and Edit. Level 2 Logic.**

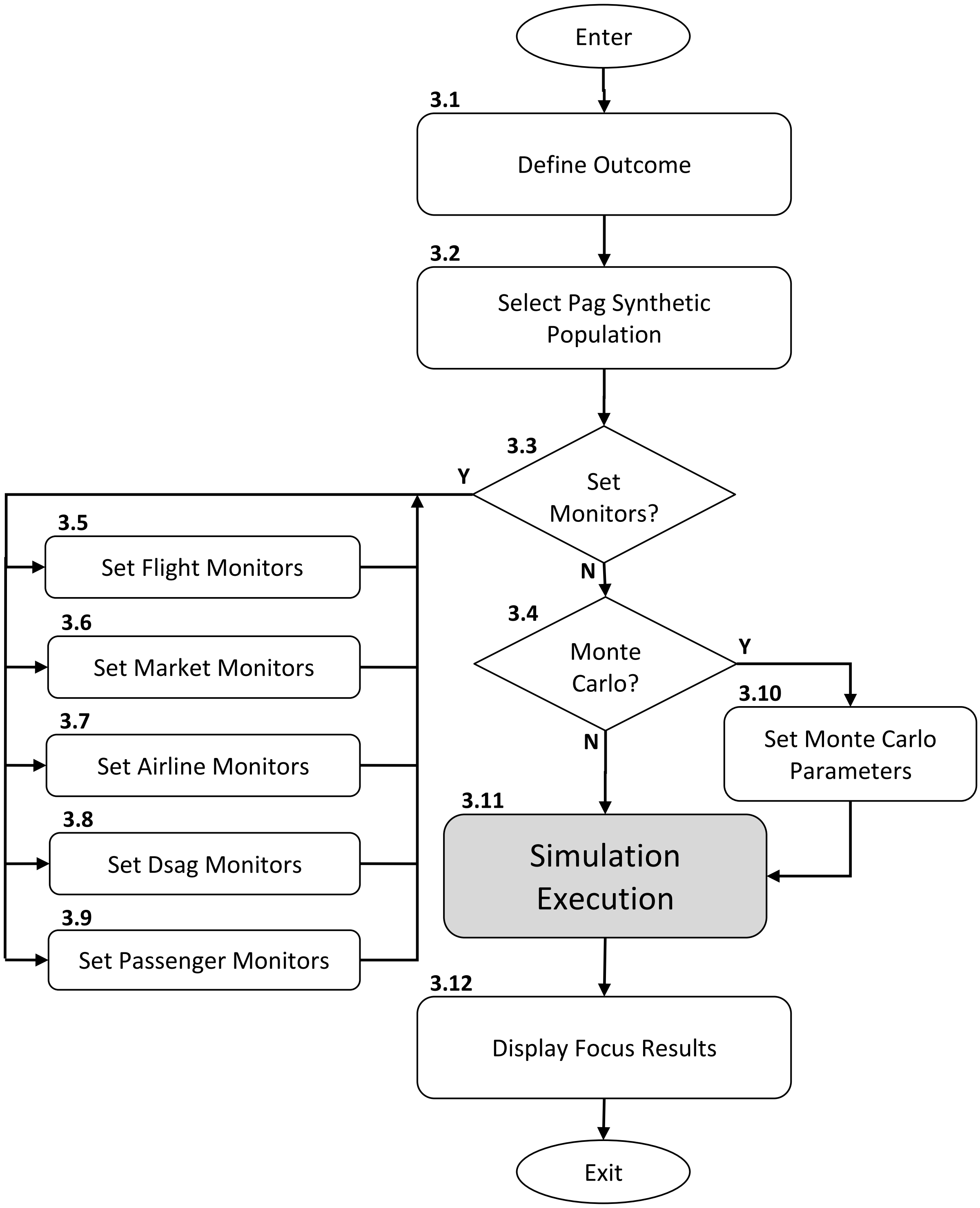



**Figure 7.1.4: AirVM Program Logic Flowchart. Procedure 3. Simulation Execution. Level 2 Logic**

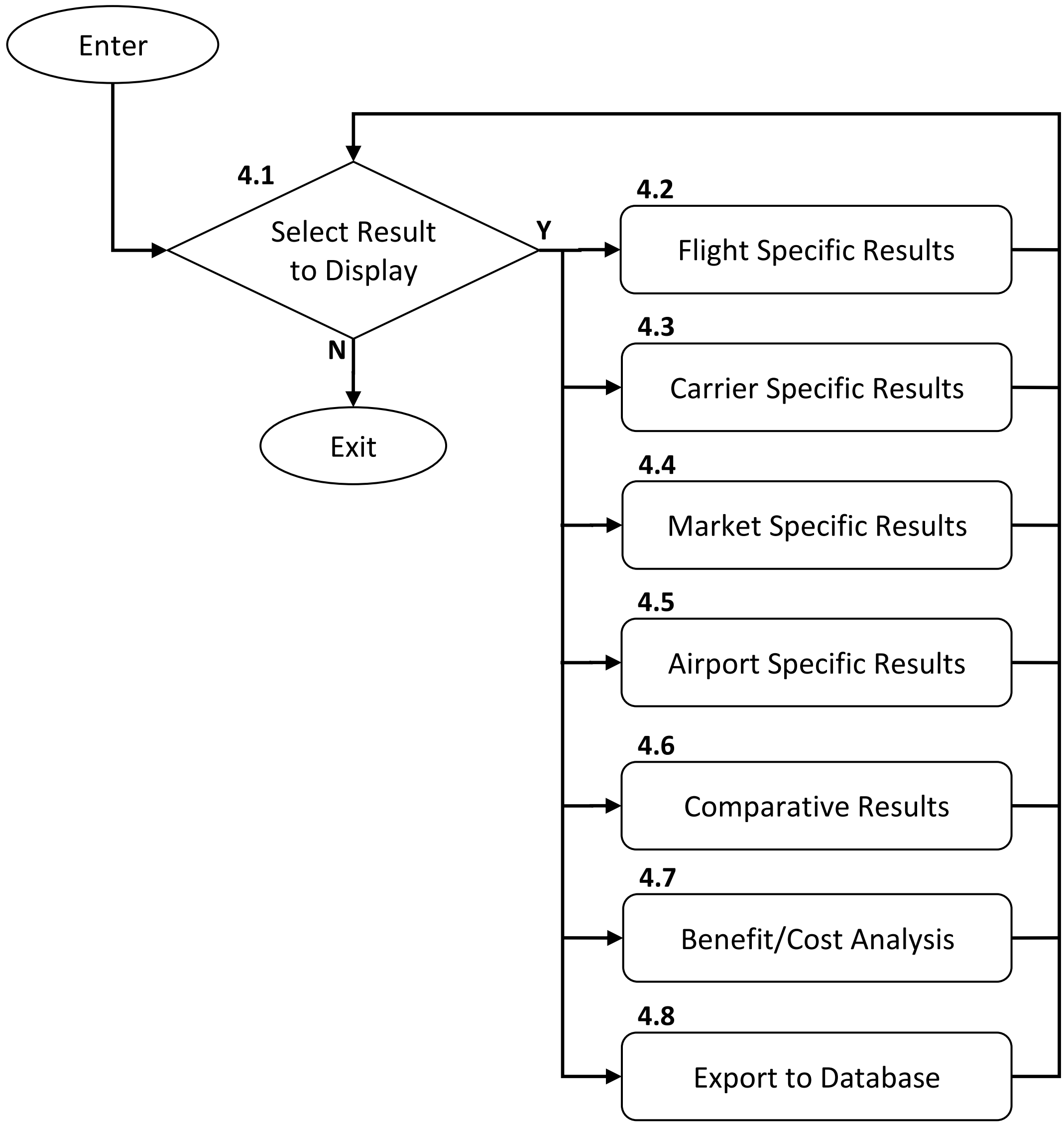



**Figure 7.1.5: AirVM Program Logic Flowchart. Procedure 4: Analysis of Simulation Results, Level 2 Logic**

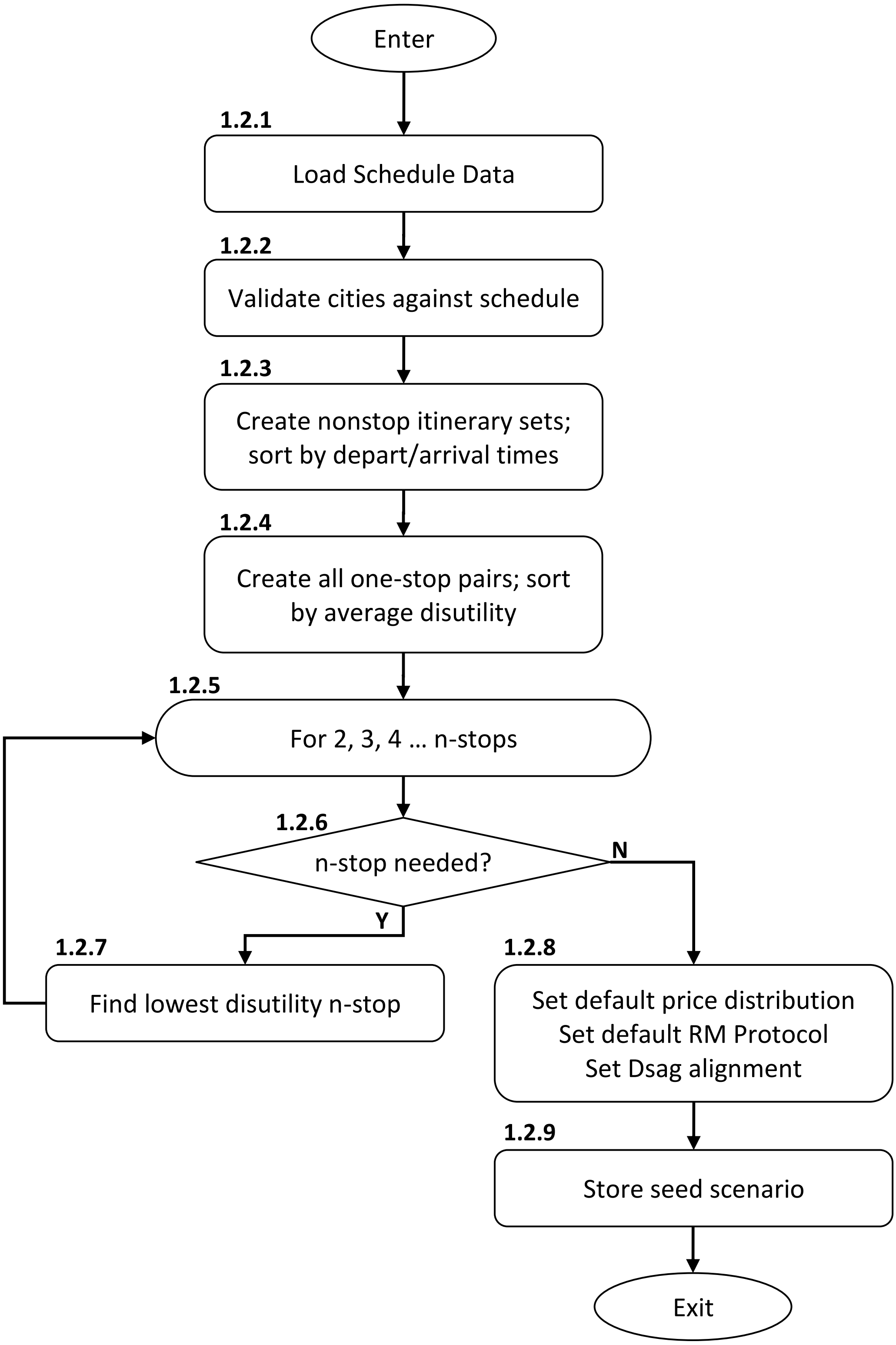



**Figure 7.1.6: AirVM Program Logic Flowchart. Procedure 1.2: Seed Scenario Creation. Level 3 Logic**

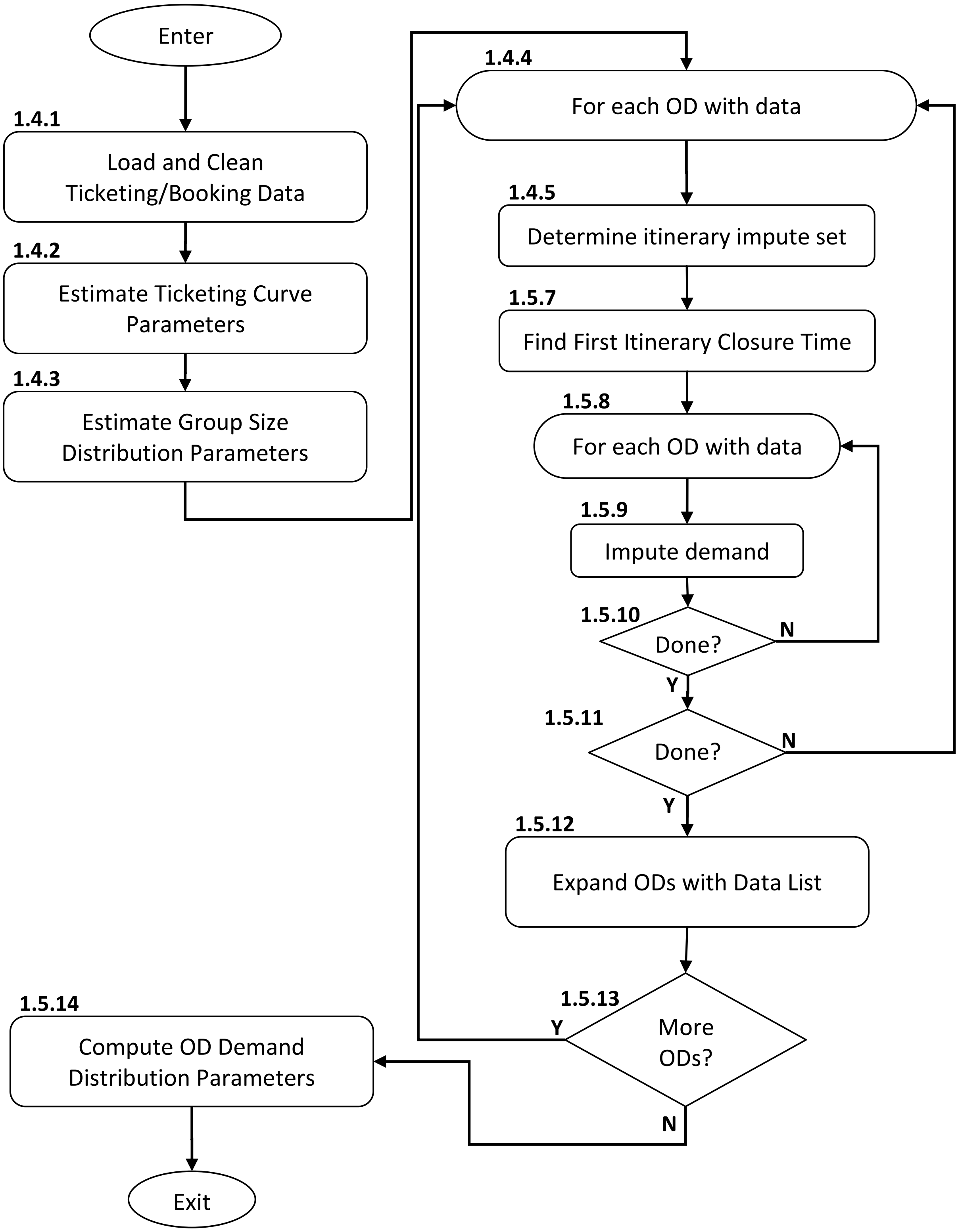



**Figure 7.1.7: AirVM Program Logic Flowchart. Procedure 1.4: Iterative Demand Processing. Level 3 Logic**

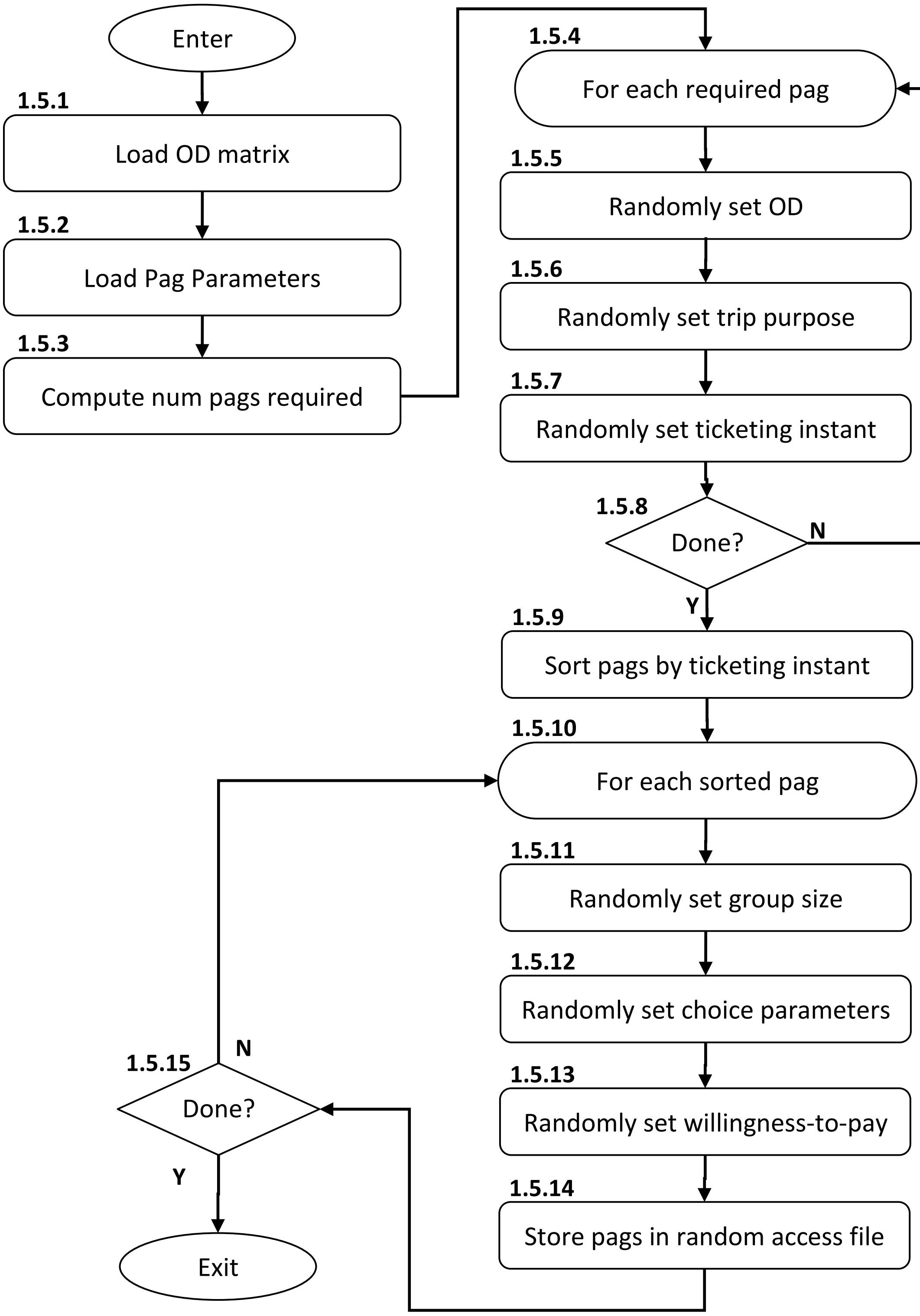



**Figure 7.1.8: AirVM Program Logic Flowchart. Procedure 1.5. Synthetic Population Generation. Level 3 Logic**

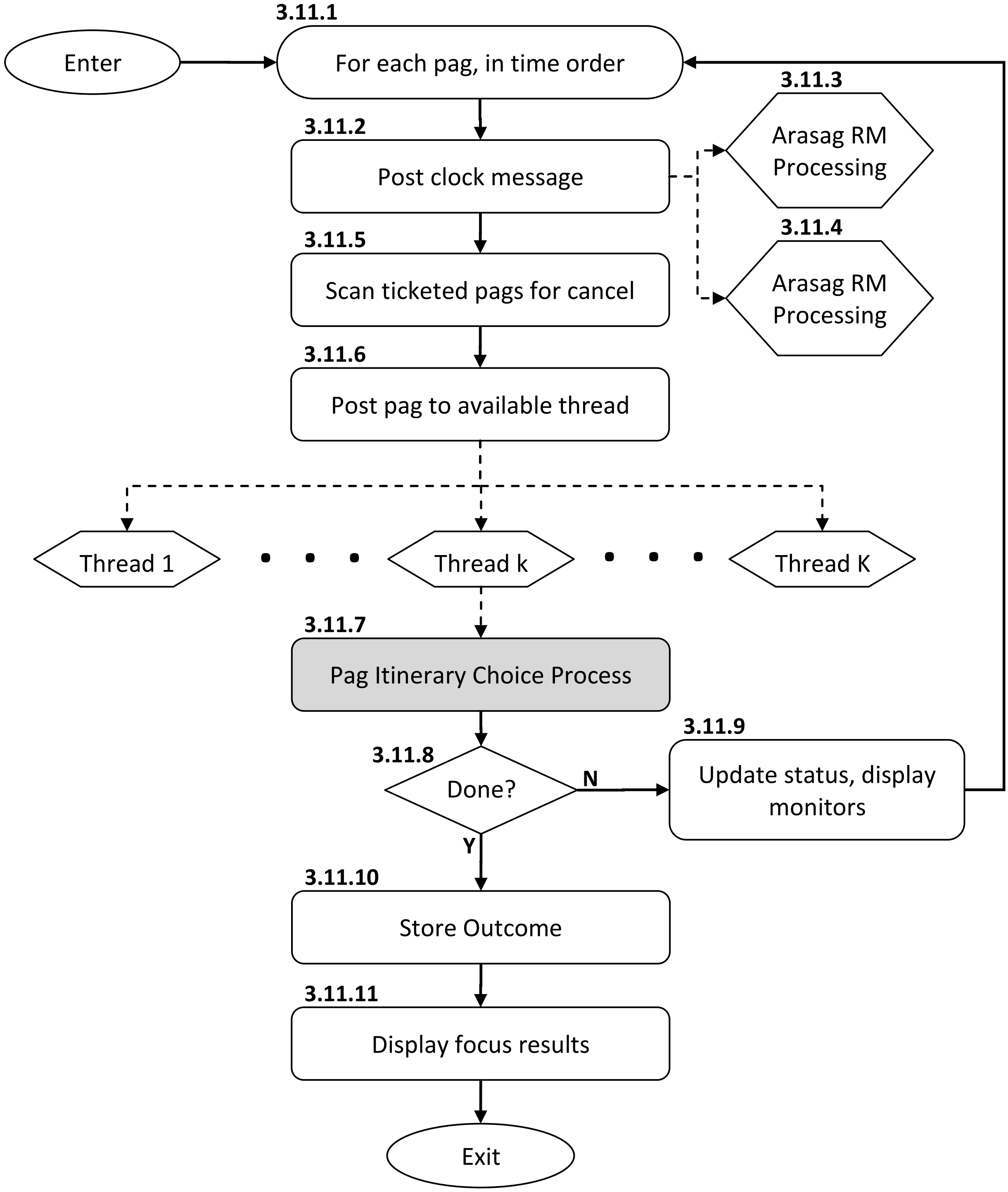



**Figure 7.1.9: AirVM Program Logic Flowchart. Procedure 3.11: Simulation Execution. Level 3 Logic**

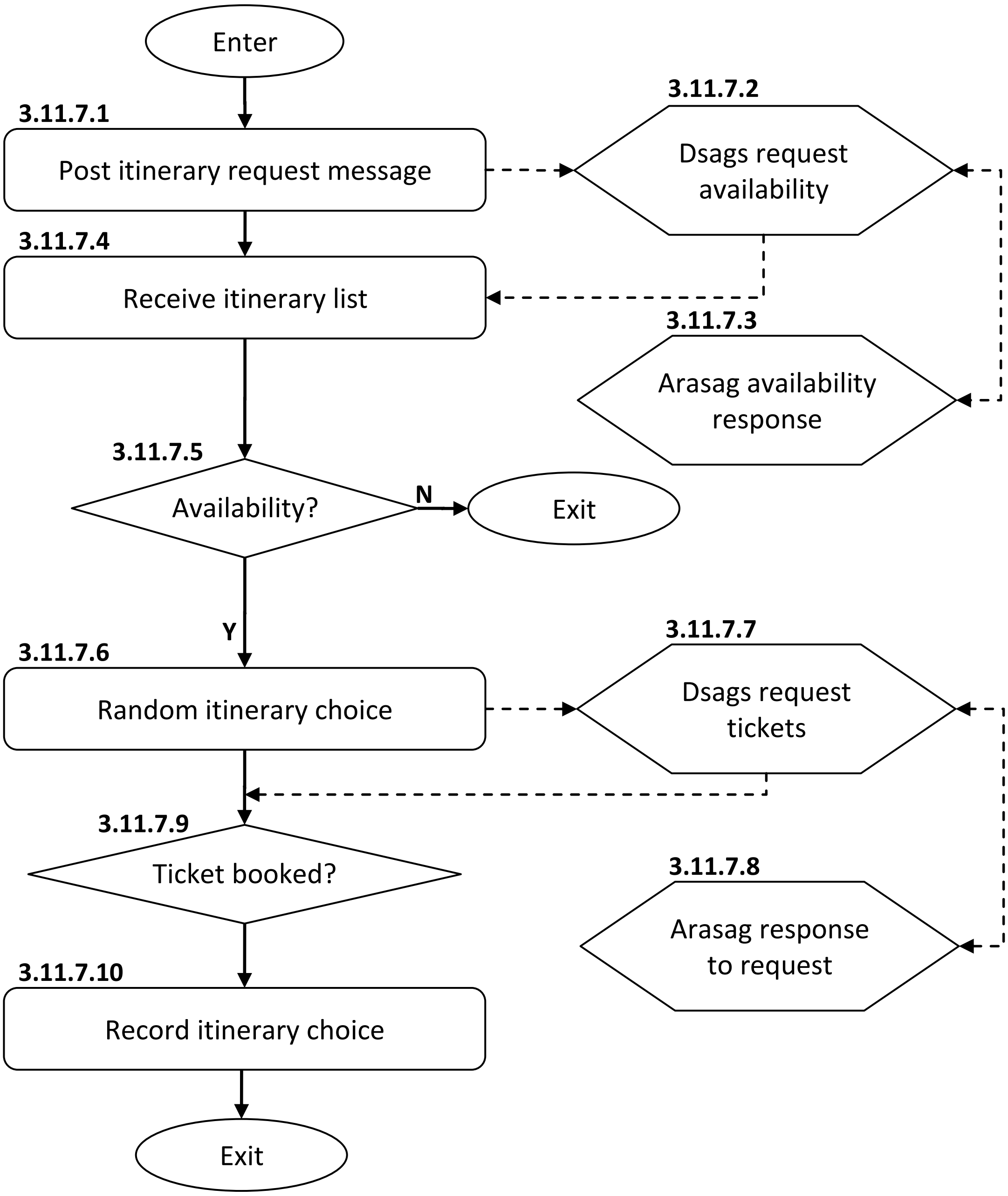



**Figure 7.1.10: AirVM Program Logic Flowchart. Procedure 3.11.7: Pag Itinerary Choice. Level 4 Logic**

## 7.2: Perspective and Portrayal: The User Operation of AirVM

**7.2.1:** The actual physical operation of AirVM is sketched out in this section of the discussion. It is infeasible to illustrate every interaction screen and fully explain all the options and conditions. That is the domain of a user's manual, which, for AirVM, is an online document. However, the general course of how a user would operate the program, what the essential step are, and a general flavor for the interface and displays helps in understanding how a virtual market is realized in practice, and what kinds of considerations are of concern. Within the overall definition of a virtual market simulation, this is the "Perspective and Portrayal" part of the simulation design. See the discussion of 4.5.6.

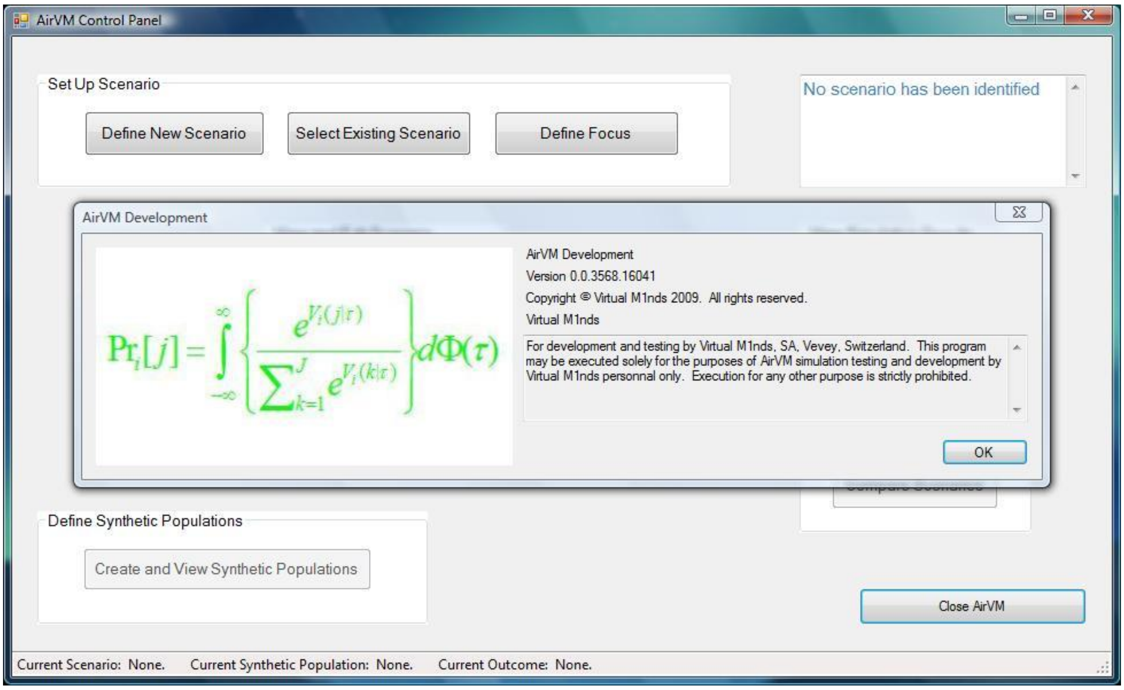



**Figure 7.2: AirVM V.0.0.3568.16041: Splash Screen at Launch**

**7.2.2:** The current implementation of AirVM is designed for Intel-based processors running the Windows 7, Vista or XP operating systems. The minimum hardware configuration is a quad processor (with four CPU's), 8 Gb of memory, and 1.3 terabytes of disk storage. No special video or printing capabilities are needed. The version used for this discussion is V.0.0.3568.16041.[77] The simulation is written in Microsoft C# 2008, using the Visual Studio 2008 Developer Suite. At the present time, it consists of

[77] This form of versioning reflects the *release.major-version.minor-version.revision.build* convention often seen in software development.

approximately 380,000 lines of code. I am completely responsible for all the code in AirVM. The opening splash screen is shown in Figure 7.2.

**7.2.3:** The main control screen is illustrated in Figure 7.3. AirVM follows a design paradigm of a process flow control program, as opposed to the more common document processing paradigm usually seen in Windows applications. That means that there are no menus or document references, rather there are control buttons which open dialog boxes to allow the user to modify aspects of the scenario, execute the simulation under different conditions, and examine the results from different viewpoints. The main control screen illustrates this paradigm. The three buttons shown in the box in the upper left let the user define a new scenario, load an existing one, or define the focus of an analysis. When clicked, a dialog box opens to allow the user to take the steps necessary to fulfill the requested task.

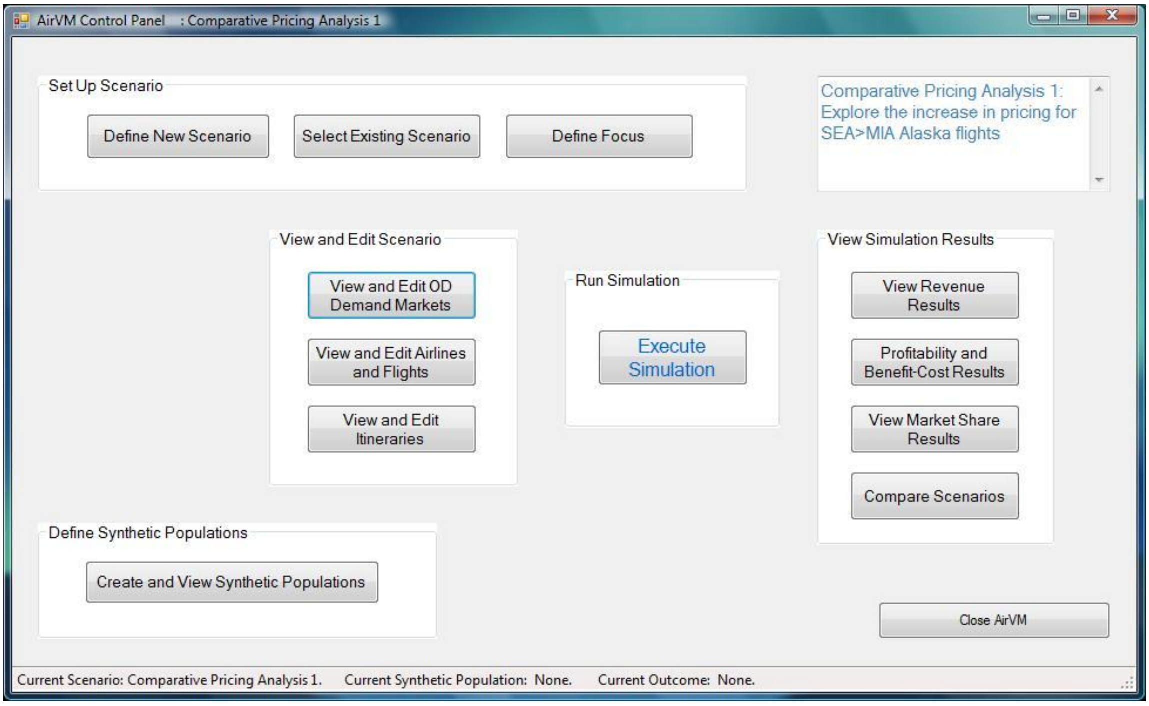



**Figure 7.3: AirVM V.0.0.3568.16041: AirVM Control Panel after Scenario Selection**

**7.2.4:** In the lower left of the control screen is a button that opens the synthetic population creation dialog box. Recall that a simulation combines a scenario with a synthetic population of pags. This control gives the user access to the parameters that define the synthetic population, such as the fraction of business versus leisure travel, the parameters of the willingness-to-pay models, and the ability to change the parameters of the itinerary choice model. Only users who have been granted permission by Virtual M1nds have access to the full functionality of synpop generation, both to protect proprietary information and to prevent users from creating analyses that prove a preconceived concept by adjusting passenger behavior to suit their own view of the world. Figure 7.4, for example, shows the scenario selection dialog box.

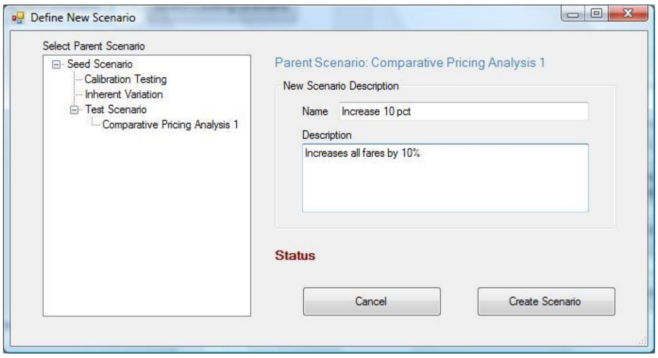



**Figure 7.4: AirVM V.0.0.3568.16041: Scenario Selection**

**7.2.5:** When a scenario and focus have been defined, the user can view or edit a range of properties of the scenario by using the buttons in the center left of the control screen. Notice that the terms "pag," "dsag" and "arasag" are not used in the button descriptions or the text on the forms. These terms mean little to the user, and just create unnecessary confusion at this level of presentation. The view and editing dialogs vary, of course, with the nature of the features being considered. Figure 7.5 shows the fare editing screen. The drop-down box on the left allows the user to select the desired airline from those available, and the

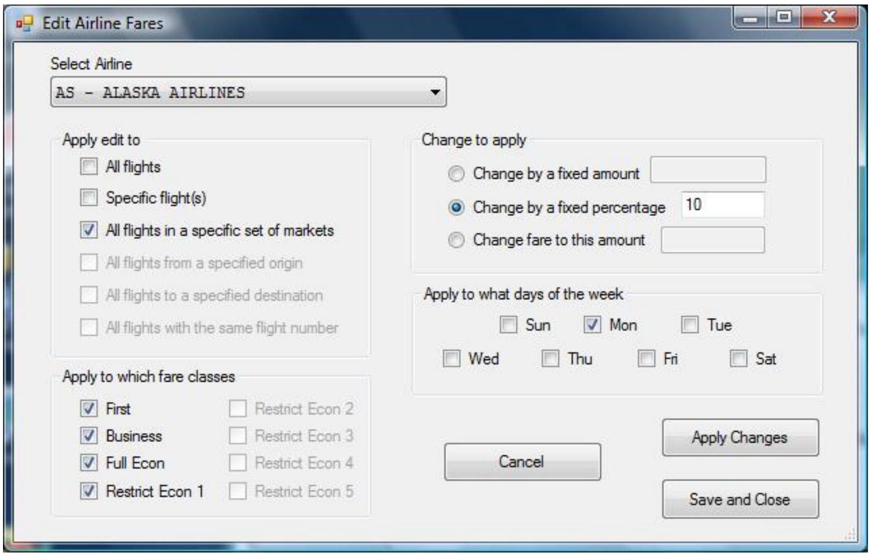



**Figure 7.5: AirVM V.0.0.3568.16041: Edit Fare Dialog Box**

remaining controls allow him to select one or more flights and change the fare in a variety of ways. When an edit occurs to a scenario, that edit is permanent. This is the reasoning behind the scenario family structure. A user can create and compare sibling and offspring scenarios while keeping a maximum level of comparability. It does, however, require significant disk storage

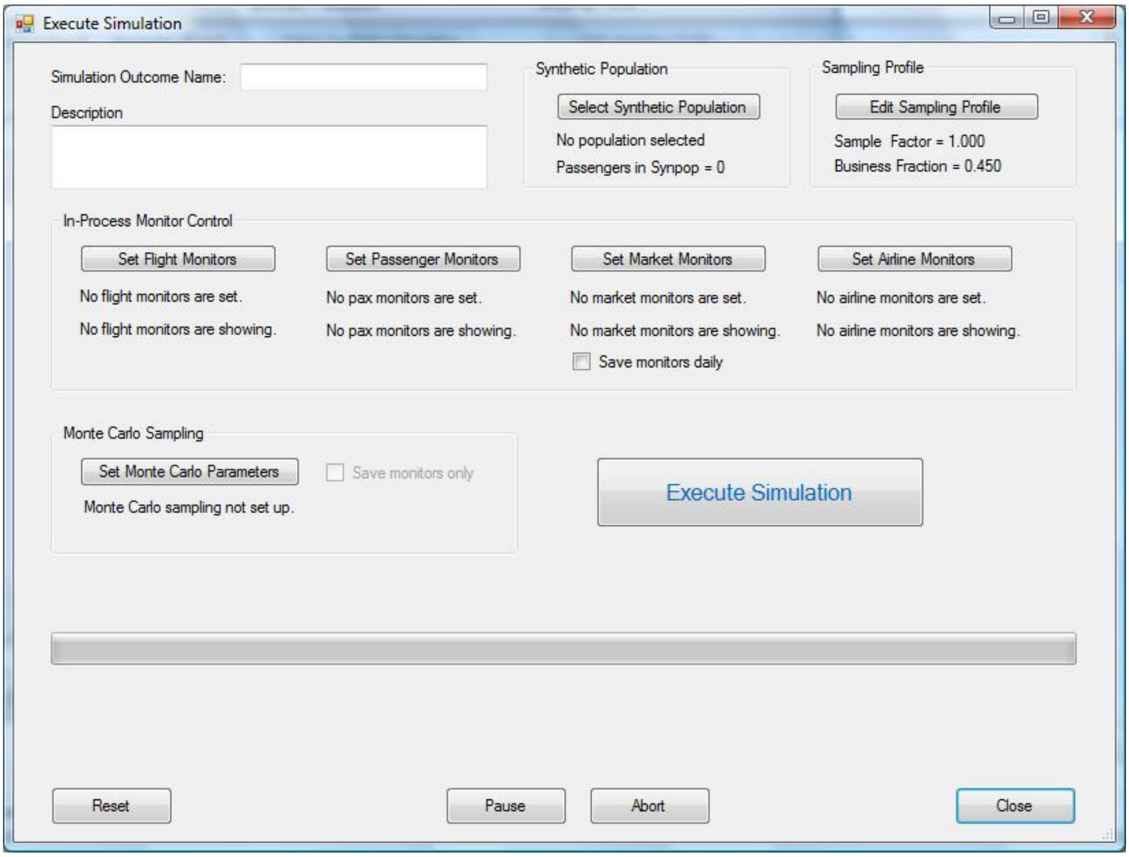



**Figure 7.6: AirVM V.0.0.3568.16041: Simulation Execution Control: Initial State**

**7.2.6:** With a scenario defined or selected, and desired modifications made to that scenario, the user can then invoke the simulation execution form by clicking on the "Execute Simulation" button in the middle of the control screen. This opens the dialog box shown in Figure 7.6. The results of a scenario is an outcome, and the area in the upper left of the execute screen is where the outcome is named and described. Every run must have a unique name, and the description can be anything appropriate the user wishes. Experience has shown that users need the ability to describe how scenarios and outcomes differ in detail to coherently move through a complex analysis. The other component of the simulation is the synthetic population. This is selected using the dialog box invoked by the button in the upper right center. The synthetic population can be adjusted to allow some changes to the way in which the synthetic population is used in that particular simulation execution by opening the sampling profile dialog box invoked by the upper far right button. The random number seed can be fixed (it is normally chosen by the system time), the fraction of business vs. leisure travel for the entire network adjusted, and oversampling or undersampling specified. Oversampling and undersampling are quick ways of stress testing a scenario; more or fewer pags respectively are ticketed by duplicating or dropping, at random, pags during the simulation execution.

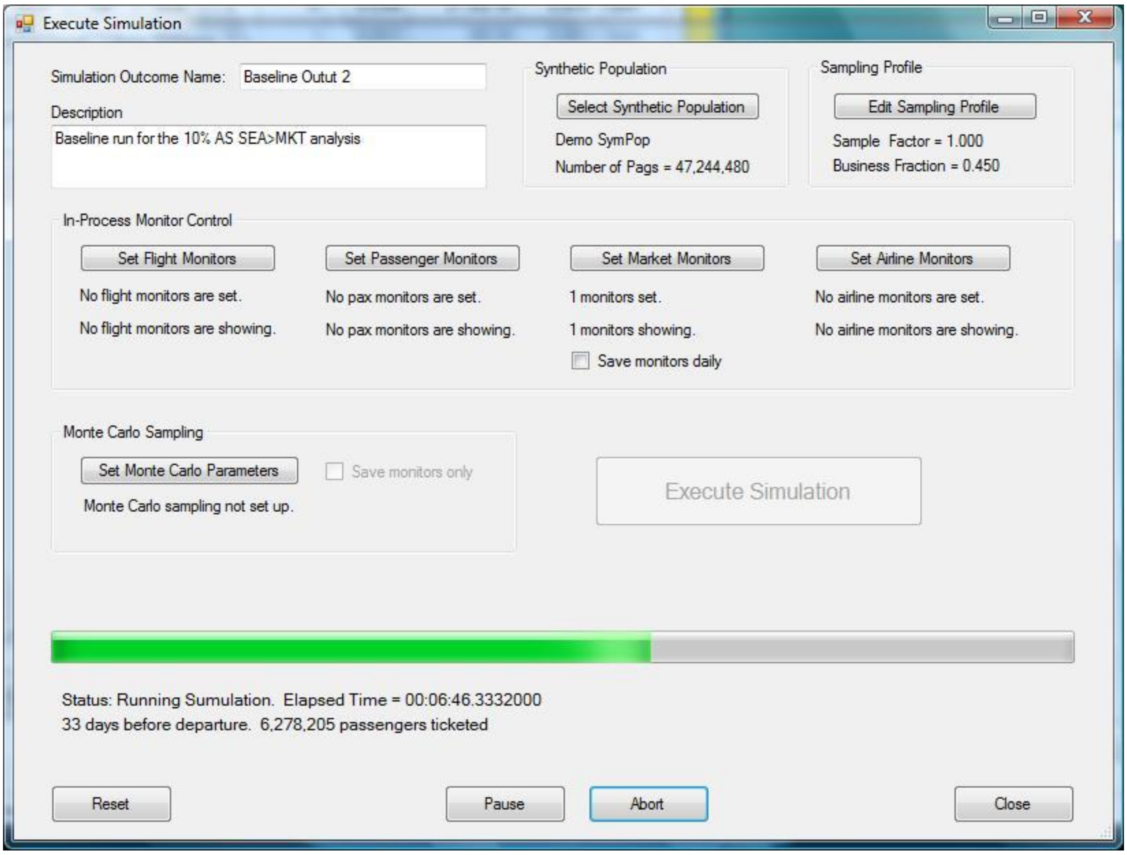



**Figure 7.7: AirVM V.0.0.3568.16041: Simulation Execution Control, In-Process.**

**7.2.7:** Figure 7.7 shows the appearance of the execution dialog box part way through a simulation. The simulation is executed with the specified synthetic population, and with any selected monitors (see below), by pressing the "Execute Simulation" button prominent in the center right of the execute dialog box. Clicking that button begins the

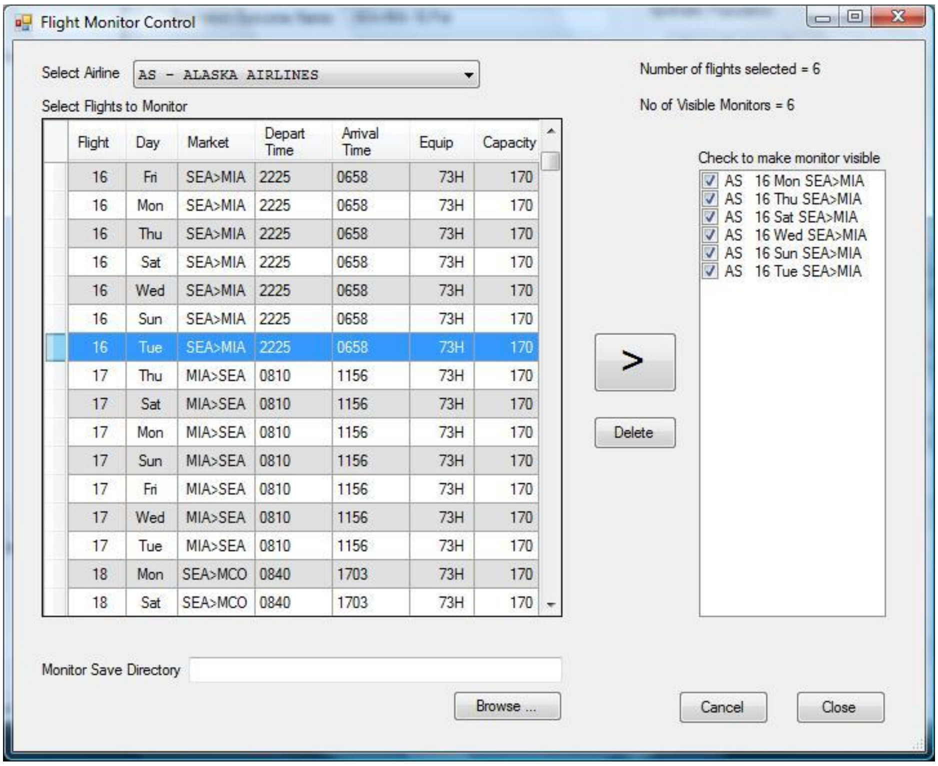
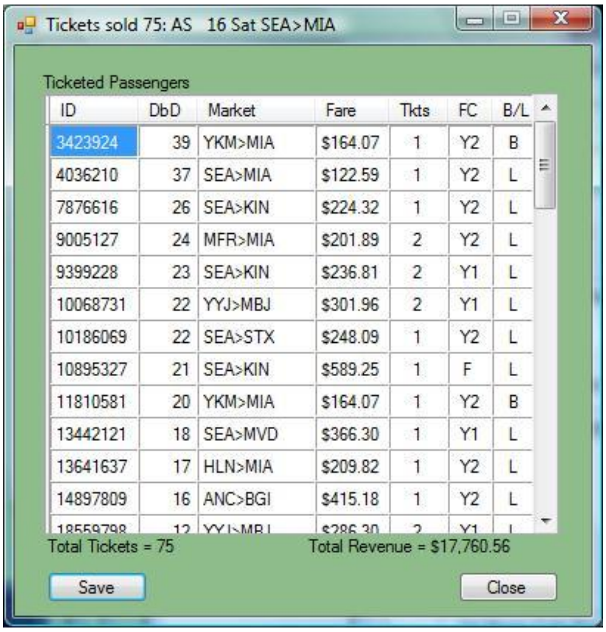



**Figure 7.8: AirVM V.0.0.3568.16041: Flight Monitor Selection Dialog Box**

simulation execution. As the execution proceeds, the progress bar and the two legends below it show the current status of the run. Each simulation in this version takes approximately 45 minutes to execute, depending on the size of the synthetic population and the number of monitors that are operating.

**7.2.8:** The upper center portion of the Execute Simulation form contains the access buttons for the specification of simulation monitors. Monitors are objects that capture specific activity during the course of a simulation, to allow greater detailed study and analysis. The user can monitor a flight, a market, a passenger, or an airline. Monitors can run in the background during a simulation, and are then displayed at the end of the run. They can also be set to be visible during a simulation run, and in this way set to portray some of the dynamics of the simulation process as it works. Figure 7.8 shows the flight monitor selection dialog box. The flights are selected from the list of flights operated by the chosen arasag and listed on the right. The check-box is checked if the user wants the monitor visible during simulation execution. The monitor results can be saved as a .csv file after the run is complete, to a location selected by the user. Figure

7.9 illustrates a flight monitor. As a pag purchases a ticket on an itinerary that uses the monitored flight, its ID number, days-before-departure (DbD), market, fare, number of tickets purchased, fare class code, and trip purpose are displayed. The user can select a pag and click to see the details of that pag's full itinerary and also see what other options it was given to select from. As execution proceeds, this monitor is updated with additional lines being added.


Tickets sold 75: AS 16 Sat SEA>MIA

Ticketed Passengers

| ID | DbD | Market | Fare | Tkts | FC | B/L |
|---|---|---|---|---|---|---|
| 3423924 | 39 | YKM>MIA | $164.07 | 1 | Y2 | B |
| 4036210 | 37 | SEA>MIA | $122.59 | 1 | Y2 | L |
| 7876616 | 26 | SEA>KIN | $224.32 | 1 | Y2 | L |
| 9005127 | 24 | MFR>MIA | $201.89 | 2 | Y2 | L |
| 9399228 | 23 | SEA>KIN | $236.81 | 2 | Y1 | L |
| 10068731 | 22 | YYJ>MBJ | $301.96 | 2 | Y1 | L |
| 10186069 | 22 | SEA>STX | $248.09 | 1 | Y2 | L |
| 10895327 | 21 | SEA>KIN | $589.25 | 1 | F | L |
| 11810581 | 20 | YKM>MIA | $164.07 | 1 | Y2 | B |
| 13442121 | 18 | SEA>MVD | $366.30 | 1 | Y1 | L |
| 13641637 | 17 | HLN>MIA | $209.82 | 1 | Y2 | L |
| 14897809 | 16 | ANC>BGI | $415.18 | 1 | Y2 | L |

Total Tickets = 75 Total Revenue = $17,760.56

Save Close

**Figure 7.9: AirVM V.0.0.3568.16041: Flight Monitor**

**7.2.9:** Figure 7.10 shows the market monitor selection dialog box. The user first chooses an origin, then all destinations with non-zero demand are listed, from which he chooses the destination part of the OD. Again, checking the OD market makes the monitor visible as the simulation executes. Figure 7.11 shows the market monitor. It shows all itineraries that are used by any pag travelling in the selected market. Among the data shown for each itinerary is the carrier, departure time and day, arrival time and day, number of stops, number and share of tickets sold for this itinerary in the market, total revenue and revenue share, and whether the itinerary is open or closed (has available seats in any fare class). As the simulation progresses, itineraries are updated with new bookings. As with the flight monitors, the results can be saved to a .csv file for further study.

**7.2.10**: Referring to Figure 7.6, in the lower left center is the button that brings up the dialog box that manages a Monte Carlo run. A Monte Carlo simulation is the repeated execution of the AirVM simulator in which each cycle consists of a complete simulation using a different value of a key parameter. That parameter is itself governed by a probability distribution, from which the specific value for each cycle is pulled. At the present time, two kinds of Monte Carlo simulations can be run. The first changes nothing at all from cycle to cycle. This means that the only variation possible is the

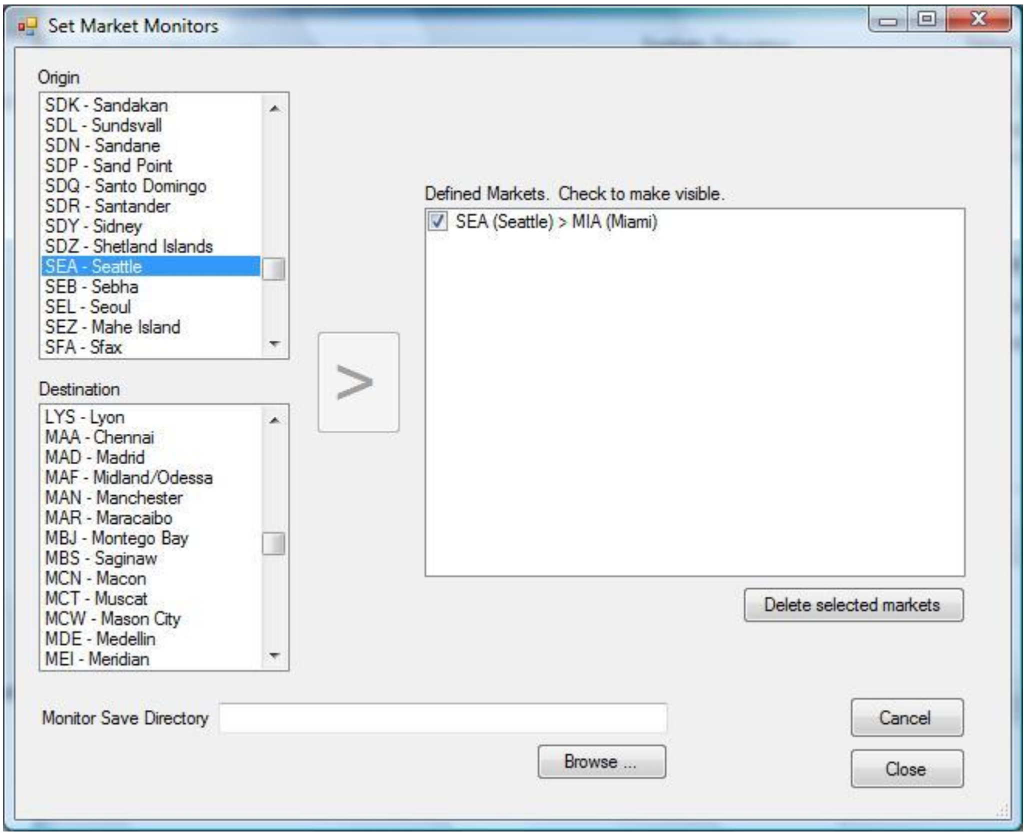
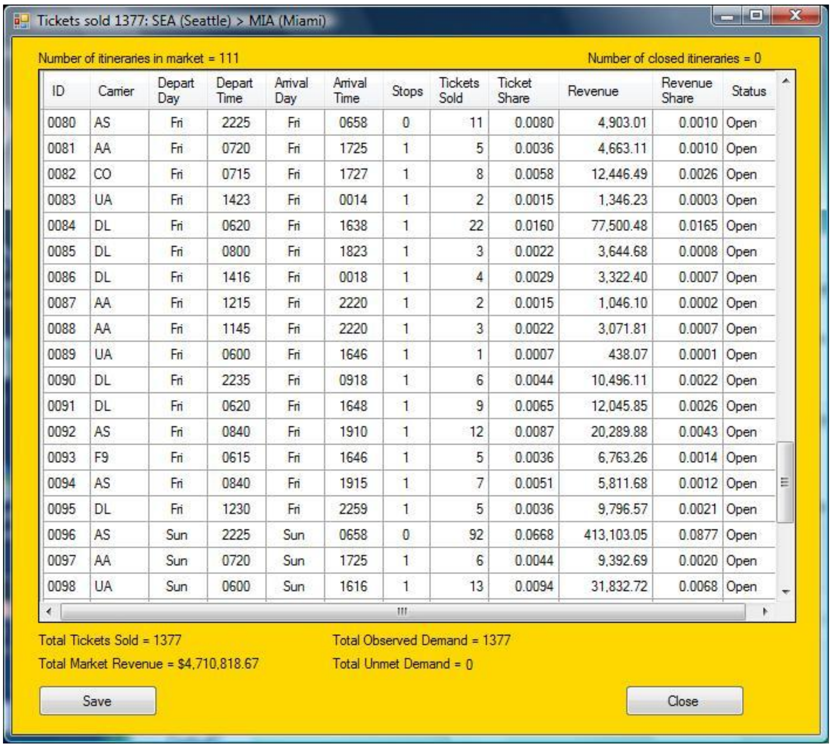



**Figure 7.10: AirVM V.0.0.3568.16041: Market Monitor Selection Dialog Box**

inherent variation that exists because of the random property of the itinerary discrete choice model. This is used to determine the inherent variability in leg demand which is due solely to the choice behavior of the pags. The second, and much more useful in many ways, variable available for Monte Carlo runs is market demand. Recall that market demand in the OD demand matrix is a random variable. A demand Monte Carlo generates a different demand for all or selected markets according to the respective demand distributions. This has direct use in a number of estimation contexts, including the estimation of risk associated with schedule or fare changes and finding confidence bounds on revenue or load estimates. A Monte Carlo simulation uses monitors exclusively to record the results of the execution. These results monitors are saved automatically so they can be used for analysis outside of AirVM.



Tickets sold 1377: SEA (Seattle) > MIA (Miami)

Number of itineraries in market = 111

Number of closed itineraries = 0

| ID | Carrier | Depart Day | Depart Time | Arrival Day | Arrival Time | Stops | Tickets Sold | Ticket Share | Revenue | Revenue Share | Status |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0080 | AS | Fri | 2225 | Fri | 0658 | 0 | 11 | 0.0080 | 4,903.01 | 0.0010 | Open |
| 0081 | AA | Fri | 0720 | Fri | 1725 | 1 | 5 | 0.0036 | 4,663.11 | 0.0010 | Open |
| 0082 | CO | Fri | 0715 | Fri | 1727 | 1 | 8 | 0.0058 | 12,446.49 | 0.0026 | Open |
| 0083 | UA | Fri | 1423 | Fri | 0014 | 1 | 2 | 0.0015 | 1,346.23 | 0.0003 | Open |
| 0084 | DL | Fri | 0620 | Fri | 1638 | 1 | 22 | 0.0160 | 77,500.48 | 0.0165 | Open |
| 0085 | DL | Fri | 0800 | Fri | 1823 | 1 | 3 | 0.0022 | 3,644.68 | 0.0008 | Open |
| 0086 | DL | Fri | 1416 | Fri | 0018 | 1 | 4 | 0.0029 | 3,322.40 | 0.0007 | Open |
| 0087 | AA | Fri | 1215 | Fri | 2220 | 1 | 2 | 0.0015 | 1,046.10 | 0.0002 | Open |
| 0088 | AA | Fri | 1145 | Fri | 2220 | 1 | 3 | 0.0022 | 3,071.81 | 0.0007 | Open |
| 0089 | UA | Fri | 0600 | Fri | 1646 | 1 | 1 | 0.0007 | 438.07 | 0.0001 | Open |
| 0090 | DL | Fri | 2235 | Fri | 0918 | 1 | 6 | 0.0044 | 10,496.11 | 0.0022 | Open |
| 0091 | DL | Fri | 0620 | Fri | 1648 | 1 | 9 | 0.0065 | 12,045.85 | 0.0026 | Open |
| 0092 | AS | Fri | 0840 | Fri | 1910 | 1 | 12 | 0.0087 | 20,289.88 | 0.0043 | Open |
| 0093 | F9 | Fri | 0615 | Fri | 1646 | 1 | 5 | 0.0036 | 6,763.26 | 0.0014 | Open |
| 0094 | AS | Fri | 0840 | Fri | 1915 | 1 | 7 | 0.0051 | 5,811.68 | 0.0012 | Open |
| 0095 | DL | Fri | 1230 | Fri | 2259 | 1 | 5 | 0.0036 | 9,796.57 | 0.0021 | Open |
| 0096 | AS | Sun | 2225 | Sun | 0658 | 0 | 92 | 0.0668 | 413,103.05 | 0.0877 | Open |
| 0097 | AA | Sun | 0720 | Sun | 1725 | 1 | 6 | 0.0044 | 9,392.69 | 0.0020 | Open |
| 0098 | UA | Sun | 0600 | Sun | 1616 | 1 | 13 | 0.0094 | 31,832.72 | 0.0068 | Open |

Total Tickets Sold = 1377

Total Observed Demand = 1377

Total Market Revenue = $4,710,818.67

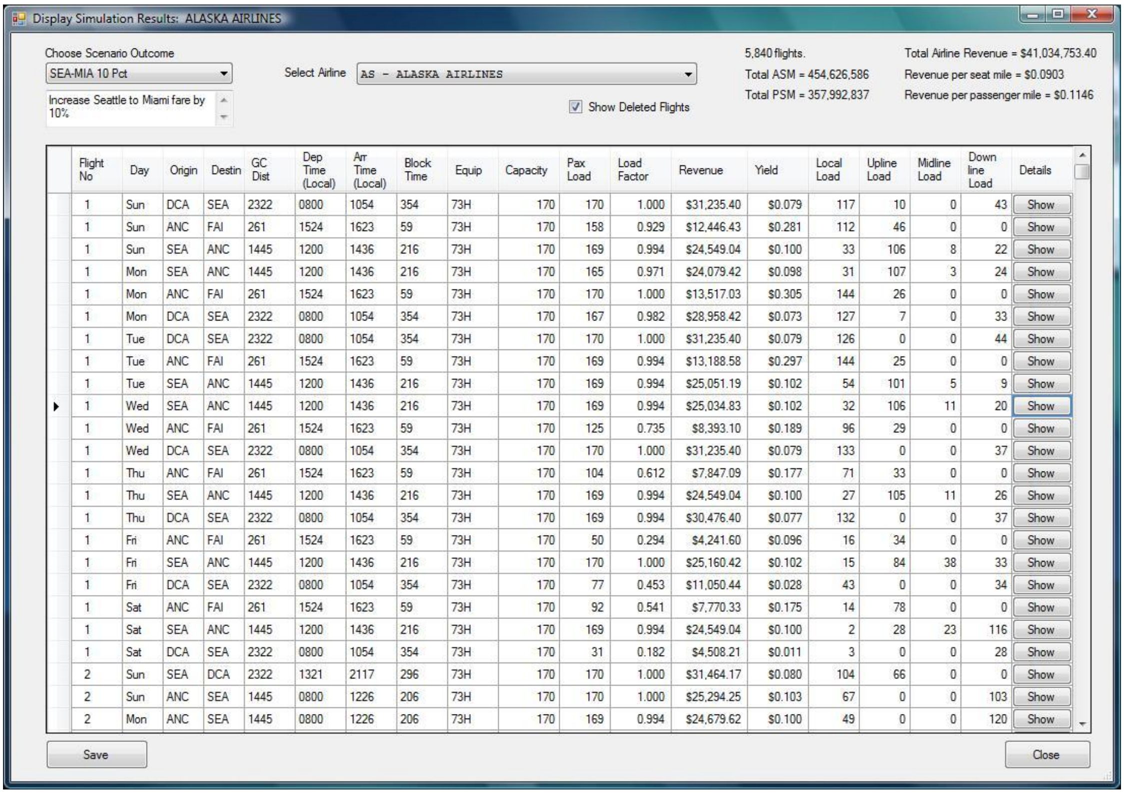
Total Unmet Demand = 0

Save

Close

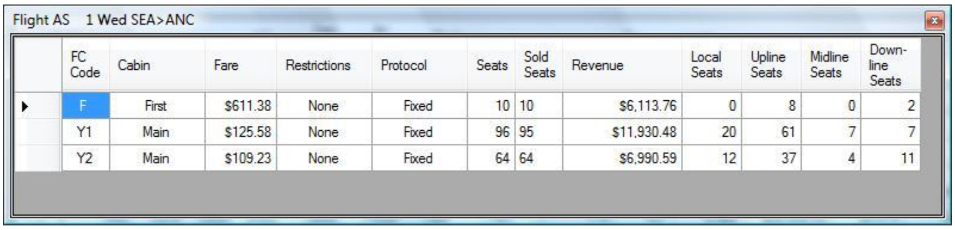
**Figure 7.11: AirVM V.0.0.3568.16041: Market Monitor**

**7.2.11:** At the end of the simulation run, the outcome is saved to disk, and the focus report is produced. From there, the user can open and explore a number of views of the outcome data set. For example, he can display load and revenue data for all the flights for a selected airline, as shown in Figure 7.12. The complete description of each flight is given, including loads, load factor (fraction of seats occupied), revenue, yield (revenue per passenger per mile flown), and the distribution of the load on the flight with respect to the set of itineraries supported by this flight. Summary statistics for the entire airline are also displayed in the upper right corner of the screen. If the user clicks the "Show" button in the far right column, the details of each individual fare class for that flight is displayed, as shown in Figure 7.13. On the other hand, if the user is interested in a particular market, he can bring up the screen shown in Figure 7.14, which depicts the market shares – both with respect to tickets sold and revenue – for all itineraries serving the market, and summarized by all carriers. If the user provides rudimentary cost information, then a benefit/cost analysis can be prepared, as shown in Figure 7.15. This also shows flight and carrier profitability. Finally, two scenarios can be compared using the comparison screen illustrated in Figure 7.16. This is of course useful for showing the effects of changes in a scenario, such as the fare increases as shown in the illustration.




| Flight No | Day | Origin | Destin | GC Dist | Dep Time (Local) | Arr Time (Local) | Block Time | Equip | Capacity | Pax Load | Load Factor | Revenue | Yield | Local Load | Upline Load | Midline Load | Down line Load | Details |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Sun | DCA | SEA | 2322 | 0800 | 1054 | 354 | 73H | 170 | 170 | 1.000 | $31,235.40 | $0.079 | 117 | 10 | 0 | 43 | Show |
| 1 | Sun | ANC | FAI | 261 | 1524 | 1623 | 59 | 73H | 170 | 158 | 0.929 | $12,446.43 | $0.281 | 112 | 46 | 0 | 0 | Show |
| 1 | Sun | SEA | ANC | 1445 | 1200 | 1436 | 216 | 73H | 170 | 169 | 0.994 | $24,549.04 | $0.100 | 33 | 106 | 8 | 22 | Show |
| 1 | Mon | SEA | ANC | 1445 | 1200 | 1436 | 216 | 73H | 170 | 165 | 0.971 | $24,079.42 | $0.098 | 31 | 107 | 3 | 24 | Show |
| 1 | Mon | ANC | FAI | 261 | 1524 | 1623 | 59 | 73H | 170 | 170 | 1.000 | $13,517.03 | $0.305 | 144 | 26 | 0 | 0 | Show |
| 1 | Mon | DCA | SEA | 2322 | 0800 | 1054 | 354 | 73H | 170 | 167 | 0.982 | $28,958.42 | $0.073 | 127 | 7 | 0 | 33 | Show |
| 1 | Tue | DCA | SEA | 2322 | 0800 | 1054 | 354 | 73H | 170 | 170 | 1.000 | $31,235.40 | $0.079 | 126 | 0 | 0 | 44 | Show |
| 1 | Tue | ANC | FAI | 261 | 1524 | 1623 | 59 | 73H | 170 | 169 | 0.994 | $13,188.58 | $0.297 | 144 | 25 | 0 | 0 | Show |
| 1 | Tue | SEA | ANC | 1445 | 1200 | 1436 | 216 | 73H | 170 | 169 | 0.994 | $25,051.19 | $0.102 | 54 | 101 | 5 | 9 | Show |
| 1 | Wed | SEA | ANC | 1445 | 1200 | 1436 | 216 | 73H | 170 | 169 | 0.994 | $25,034.83 | $0.102 | 32 | 106 | 11 | 20 | Show |
| 1 | Wed | ANC | FAI | 261 | 1524 | 1623 | 59 | 73H | 170 | 125 | 0.735 | $8,393.10 | $0.189 | 96 | 29 | 0 | 0 | Show |
| 1 | Wed | DCA | SEA | 2322 | 0800 | 1054 | 354 | 73H | 170 | 170 | 1.000 | $31,235.40 | $0.079 | 133 | 0 | 0 | 37 | Show |
| 1 | Thu | ANC | FAI | 261 | 1524 | 1623 | 59 | 73H | 170 | 104 | 0.612 | $7,847.09 | $0.177 | 71 | 33 | 0 | 0 | Show |
| 1 | Thu | SEA | ANC | 1445 | 1200 | 1436 | 216 | 73H | 170 | 169 | 0.994 | $24,549.04 | $0.100 | 27 | 105 | 11 | 26 | Show |
| 1 | Thu | DCA | SEA | 2322 | 0800 | 1054 | 354 | 73H | 170 | 169 | 0.994 | $30,476.40 | $0.077 | 132 | 0 | 0 | 37 | Show |
| 1 | Fri | ANC | FAI | 261 | 1524 | 1623 | 59 | 73H | 170 | 50 | 0.294 | $4,241.60 | $0.096 | 16 | 34 | 0 | 0 | Show |
| 1 | Fri | SEA | ANC | 1445 | 1200 | 1436 | 216 | 73H | 170 | 170 | 1.000 | $25,160.42 | $0.102 | 15 | 84 | 38 | 33 | Show |
| 1 | Fri | DCA | SEA | 2322 | 0800 | 1054 | 354 | 73H | 170 | 77 | 0.453 | $11,050.44 | $0.028 | 43 | 0 | 0 | 34 | Show |
| 1 | Sat | ANC | FAI | 261 | 1524 | 1623 | 59 | 73H | 170 | 92 | 0.541 | $7,770.33 | $0.175 | 14 | 78 | 0 | 0 | Show |
| 1 | Sat | SEA | ANC | 1445 | 1200 | 1436 | 216 | 73H | 170 | 169 | 0.994 | $24,549.04 | $0.100 | 2 | 28 | 23 | 116 | Show |
| 1 | Sat | DCA | SEA | 2322 | 0800 | 1054 | 354 | 73H | 170 | 31 | 0.182 | $4,508.21 | $0.011 | 3 | 0 | 0 | 28 | Show |
| 2 | Sun | SEA | DCA | 2322 | 1321 | 2117 | 296 | 73H | 170 | 170 | 1.000 | $31,464.17 | $0.080 | 104 | 66 | 0 | 0 | Show |
| 2 | Sun | ANC | SEA | 1445 | 0800 | 1226 | 206 | 73H | 170 | 170 | 1.000 | $25,294.25 | $0.103 | 67 | 0 | 0 | 103 | Show |
| 2 | Mon | ANC | SEA | 1445 | 0800 | 1226 | 206 | 73H | 170 | 169 | 0.994 | $24,679.62 | $0.100 | 49 | 0 | 0 | 120 | Show |

**Figure 7.12: AirVM V.0.0.3568.16041: Revenue Analysis Display**




| FC Code | Cabin | Fare | Restrictions | Protocol | Seats | Sold Seats | Revenue | Local Seats | Upline Seats | Midline Seats | Down-line Seats |
|---|---|---|---|---|---|---|---|---|---|---|---|
| F | First | $611.38 | None | Fixed | 10 | 10 | $6,113.76 | 0 | 8 | 0 | 2 |
| Y1 | Main | $125.58 | None | Fixed | 96 | 95 | $11,930.48 | 20 | 61 | 7 | 7 |
| Y2 | Main | $109.23 | None | Fixed | 64 | 64 | $6,990.59 | 12 | 37 | 4 | 11 |

**Figure 7.13: AirVM V.0.0.3568.16041: Flight Fare Class Ticketing Detail**

**7.2.12:** The user can repeat any of these operations as he deems appropriate, and can exit AirVM and come back later to pick up with further analysis. Since AirVM is not a mature computer product as of this time, it will also continue to change as it responds to the needs of client users who find the simulation of value. The next Section of this discussion suggests who some of those customers might be.

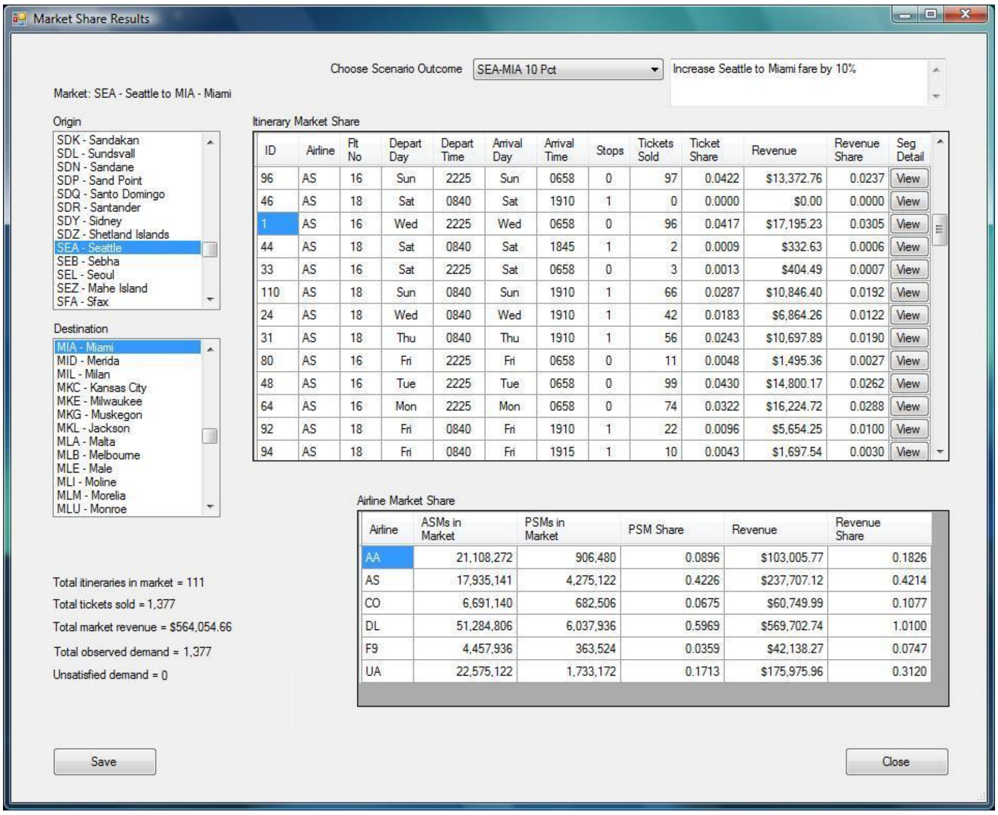



**Figure 7.14: AirVM V.0.0.3568.16041: Market Share Analysis Display**

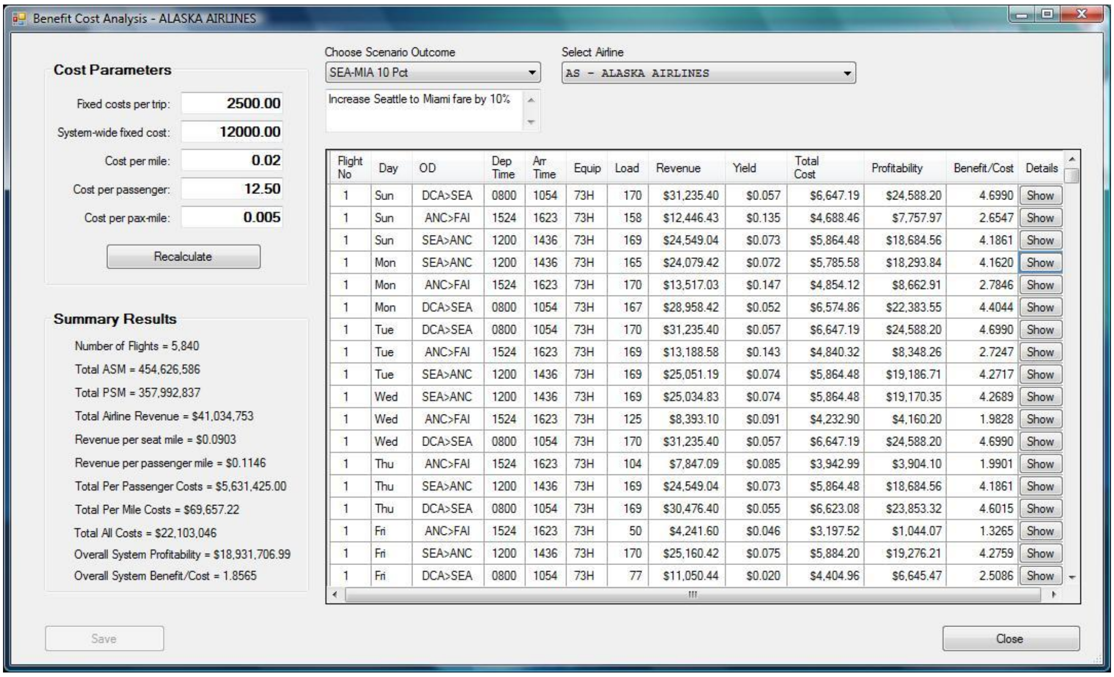



**Figure 7.15: AirVM V.0.0.3568.16041: Benefit/Cost Display**

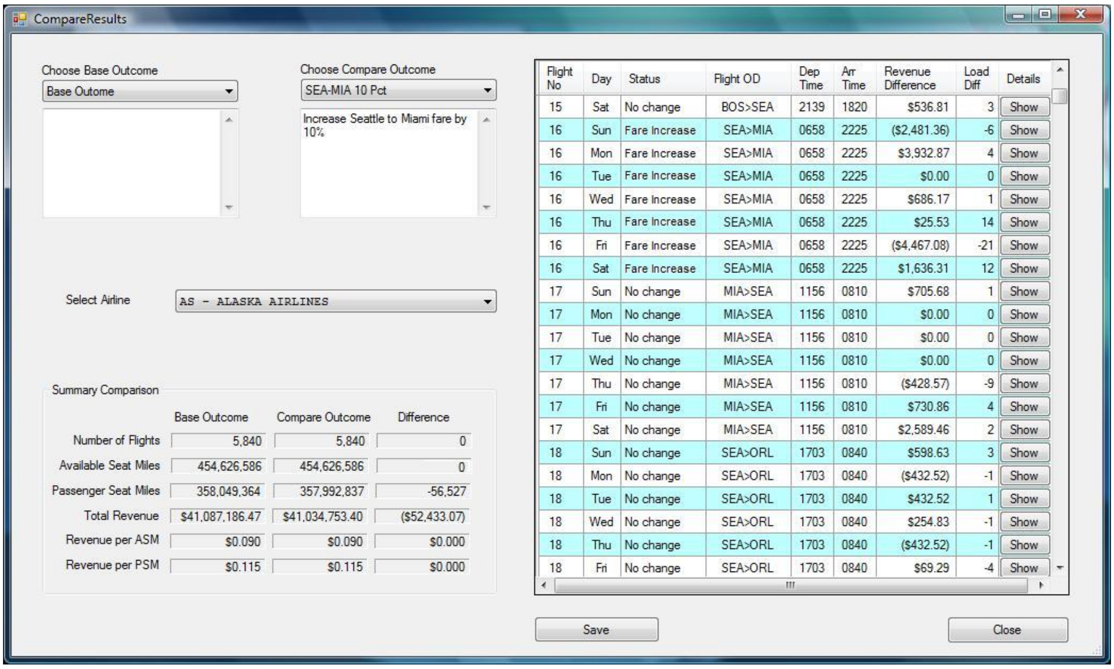



**Figure 7.16: AirVM V.0.0.3568.16041: Scenario Comparison Display**

## 7.3: The Customers of AirVM

**7.3.1:** AirVM can be of direct use to a number of segments of the airline passenger marketplace. How various businesses, government agencies and other institutions could take advantage of the opportunities offered by the virtual market is explored here. Since the client set is an important aspect of a virtual market design, the current spectrum of potential clients as they are now understood is delineated.

**7.3.2:** Airline users can change the airline network – adding or deleting flights, changing pricing, changing departure or arrival times, or changing cabin capacity – for any airline including their own or a competitor. Recall that arasags are avatar-capable agents. They can then execute the simulation and estimate revenue results of the change, comparing them against the current situation. Or they can increase or decrease demand, by general area or specific origin-destination market, and get estimate of the effects on loads and revenues. Each of these clients has its own perspective and needs its own portrayal of results to make the most of the AirVM opportunity. Current development work on AirVM is aimed at improving these perspective and portrayal aspects of the simulation.[78] Some flavor of the roles AirVM can play is suggested by

[78] Discussion of these efforts are beyond the scope of his thesis, since they engage interface design components that are secondary to the operation of the agent model itself.

the utility of AirVM to several important teams within a client airline, AirVM can provide analysis and decision support particularly suited to their needs.

**7.3.2.1:** For airline scheduling teams, who are responsible for the design and evaluation of the airline's regular schedule, these capabilities are important: 1) estimating potential on-board loads on contemplated new services; 2) separating new traffic from traffic shifted from other routes on new service; 3) changes in market share resulting from service alterations; 4) estimates of likely revenue resulting from changes in scheduled service; and 5) the potential value of specific gate slots at airports where service is contemplated, as part of the analysis supporting slot auctions.[79]

**7.3.2.2:** Somewhat surprisingly, pricing and marketing entities for airlines are usually separate from the scheduling teams. The pricing and marketing teams are responsible for setting fares and developing and marketing programs. These features of AirVM are directly useful to these efforts: 1) estimating the likely changes in demand because of changes in ticket pricing; 2) estimating the changes in demand because of changes in the ticket pricing of the carrier's competitors; and 3) determining the revenue value of customer service.

**7.3.2.3:** Revenue management teams, separately from the pricing team, determine the revenue management protocol to be applied by the carrier. This can be uniform across all flights, or can be set for each flight or class of flights separately. Among the AirVM results that are useful here are: 1) determining the changes in demand resulting from revisions to the RM system; 2) the estimated revenue differences that can be expected from the deployment of different RM systems, such as leg-based RM and OD-based RM; and 3) determining the effects of pricing on RM performance and profitability.

**7.3.2.4:** For competitive strategy teams, which are responsible for the preparation of medium and long term strategic plans for the carrier, AirVM offers these supporting services: 1) estimating the effects on carrier load and revenue of pricing and RM protocol changes instituted by competitors; 2) anticipating the effects on the carrier's load and revenue position of competitive

[79] Slots at airports – time periods at fixed gates for parking aircraft while loading and unloading – are acquired by a carrier at a so-called *slot auction*. This is a worldwide meeting of airport authorities and airlines where all carriers bid on available slots at all airports. How much these slots are worth, to the airline or the airport, is an open question in the industry that AirVM can address.

airline routing, schedule or capacity changes; and 3) anticipating changes in market share due to anticipated changes in competitor strategy.

**7.3.2.5:** Finally, for the component of airline management responsible for fleet planning and aircraft acquisition, these results are valuable: 1) calculating the expected effects of capacity changes on the loads, revenues, and market share of other routes; 2) calculating the effects of equipment on competitor operations – route share and revenue – and potential competitor response; and 3) supporting the return-on-investment analyses required to justify capital acquisition proposals in support of airplane purchase or lease.

**7.3.3:** Airport planners and municipal transportation planners benefit from AirVM by estimating passenger traffic into, through and out of airports based on anticipated changes in the service provided by carriers. AirVM provides estimates of the following key planning data for groups involved with many aspects of airport operations.

**7.3.3.1:** For airport facility planners, who are responsible for the development of both groundside and airside installations and support facilities, these data are produced: 1) the number of passengers passing through the terminals and gates, especially the number requiring internal transport; 2) the effects on passenger internal movement of gate slot and carrier service changes; 3) changes in airport security requirements resulting from changes passenger ground traffic patterns; and, 4) determining the value of gate slots as a function of potential airline revenue.

**7.3.3.2:** Local and municipal transportation planners must accommodate the needs for the population to get into and get out of airports. AirVM provides estimates of the number of local passengers departing an airport. (How many parking spaces are needed? What level of ground transport access is required?) From the other side, it can estimate the number of visiting passengers arriving, which directly affects how much local transport is required. And these data are reported by time of day or day of the week.

**7.3.3.3:** For companies that provide airport retail services, AirVM can be used to estimate 1) the number of passengers in the airport by time of day and day of week, 2) the in-terminal dwell time for transiting passengers, and, when coupled with gate assignment and airport layout information, 3) the number of waiting passengers by time of day and internal passenger location. Data of this type is very helpful for market analysis and facility planning.

**7.3.4:** For agencies within government and civil transport authorities responsible for the regulation of air travel and the mitigating the environmental impacts of travel, AirVM provides reliable information on air traffic volumes – both aircraft and passenger – which are necessary for the support of many important functions.

**7.3.4.1:** For air traffic management planning, the distribution of aircraft landings and takeoffs across the week and the passenger numbers moving in, out and through airports across the week are very important. AirVM could also be adapted to estimate the passenger cost of air traffic delays, including missed connections and time loss.

**7.3.4.2:** Those agencies responsible for environmental assessment and management would find valuable the per passenger-trip carbon cost estimates that could be supported by AirVM. In addition, comparison of the environmental costs of alternative network architectures – hub-and-spoke vs. point-to-point, for example – can be generated.

**7.3.4.3:** Finally, public health officials can use AirVM for monitoring likely vectors for communicable disease transmission, the costs of travel restrictions as a disease control strategy, and the expected effectiveness of counter-measures to combat the spread of an air-travel vectored disease.

**7.3.5:** AirVM, when applied as an instrument of observation, becomes of significant benefit to those companies within the industry that supply industry data to carriers and others. As discussed in Appendix M, AirVM is capable of generating network- and channel-adjusted demand data from the incomplete data about the marketplace that is available from several competing sources. It would clearly be difficult to do this based on, say, data from travel agencies in a small country – even though those agencies sell tickets involving travel in many parts of the world. But by exercising certain characteristics of the network model it is possible to do this with surprisingly small samples of data. AirVM thus enables suppliers of industry data to augment what comes from the sources they control.

**7.3.6:** The investment community has expressed some interest in applying AirVM as a tool to evaluate opportunities presented to it by the airline industry. Historically, no such tools have been available. But given the level of investment required to support the industry – large passenger jets cost in the hundreds of millions of dollars – and the extraordinary historic volatility of the industry since deregulation, the analytic value inherent in AirVM becomes abundantly obvious. As with all customer sets, this would require a different set of perspective and portrayals, but no real change to the underlying structure of the virtual market is necessary.

**7.3.7:** Last, but certainly by no means least, the academic community might find AirVM a valuable addition to their research laboratories. It could be used both as a teaching and research vehicle in institutions that specialize in airline and airport development and management – MIT, Georgia Tech, Northwestern, Emory Riddle in the US, and Cranfield in Great Britain come to mind. Indeed, Georgia Tech is currently in serious discussions with Virtual M1nds regarding the acquisition of access to AirVM. Its value as a research tool in the marketing academic world is unknown. Perhaps the UTS Marketing research faculty could share some views on this, and it is certainly possible to arrange for UTS to have an academic license for AirVM.

## 7.4: The Validation and Calibration of AirVM

**7.4.1:** The discussion of **4.5.4** regarding the practicalities of building agent-based models identified four criteria by which the models should be evaluated: 1) Does the software represent the underlying models faithfully and accurately? 2) How valid are the models themselves in representing the phenomena being explored? 3) Are the models calibrated with valid data and appropriate statistical analysis? 4) Is the agent-based model addressing the issues and concepts that it was designed to address? As a way of validating an agent-based model, it is tempting to see if applying the model to actual data yields the same results as the real world shows. But in all but the simplest cases, this approach is doomed. Recall that measuring the value of an agent-based model by measuring its accuracy with respect to observed values in the real world is a fruitless exercise. The model will always be wrong. The reasons for this were explained in **4.5.5** and **4.5.6**. But it is valuable and appropriate to examines AirVM with respect to the four criteria of Grimm and Railsback, as applied to the twelve incidence distribution models described in Chapter 6 which govern the execution of the simulation, all of which are stochastic.

**7.4.2:** The first criterion is the representation of the models in the software. This is generally easy to determine. Since all the models are stochastic, all require the use of random numbers to within the software. The random number generator procedure used in AirVM is the Mersenne Twister method described in **4.1.13**. Appendix N contains the code that execute these generations. All the routines were developed and tested outside the simulation itself with known data values to validate the routine logic. The testing protocols and results are not reproduced here. The reader can verify the validity of the code himself from the listings in Appendix N.

**7.4.3:** Two of the models are simple Bernoulli random variables – trip purpose and departure/arrival sensitivity, and how they are handled is typical of the procedure for discrete random variate selection. Suppose the proportion of business trip purpose in the general population is $p_b$. Then draw a uniform random value between zero and one, say $x$, and if $x < p_b$, set the trip purpose to business, otherwise to leisure. An identical process is used for departure/arrival sensitivity. Three variables – journey length, day-of-week of departure/arrival, and origin-destination of the trip – are finite random variables with more than two outcomes. For an arbitrary finite discrete random variable, a value is chosen as follows: 1) arbitrarily order the possible outcomes; 2) create a distribution function for the ordered outcomes; 3) select a uniform 0-1 random value; 4) determine the location of the random value in the empirical distribution function of the variable, and 5) select the corresponding variable value from that location. From a programming perspective, these algorithms are quite straightforward, and have been thoroughly tested in isolation prior to incorporation into AirVM. Details are described in Appendix N. It should be noted that the origin-destination distribution is extracted from a matrix of OD demand estimates using the ratio of destination demand to total origin generation. See Appendix M for details.

**7.4.3:** The pag itinerary choice model and the ticket cancellation model are used in two ways in the simulation. During the generation of the pag synthetic population, the parameters themselves are generated using the (truncated) normal routines described in Appendix N, since the coefficients of each are considered to be normally distributed with means and standard deviation estimated from data. See Sections 6.3 and 6.7 for more detail. The functional forms themselves are (finite) probability distributions, so during simulation execution discrete methods are applied to determine, respectively, which of the available itineraries are to be used or whether or not a ticket purchase has been cancelled. The normal distribution values are generated using a well-known technique described in Press, *et al.* (1988), the code for which is in Appendix N.

**7.4.4:** The ticketing instant is created during pag synpop generation time using a Poisson distribution, as discussed in Section 6.6. The intensity function is estimated from available ticketing data as shown in **6.6.4**. From that, a mean value function is readily calculated using Equation 6.27. This function is then normalized to one to create a probability distribution. Then a uniform random number is generated and applied to the inverse of the normalized mean value function to yield a ticketing instance. The probability distribution of willingness-to-pay (Equation 6.39) is used to determine the maximum fare a pag will pay (per ticket) using a similar distribution inversion procedure. One minor alteration in the standard procedure is used, however, in the ticketing instance case. If the resulting ticketing instance is more than 120 days from the departure/arrival time, the instant is set to 0. All booking instants equal to 0 are processed before other, later booking instants, thus accounting for those pags that book prior to the 120 days that are covered by AirVM. The other aspect of ticketing, the group size, is generated directly using a truncated Poisson distribution as coded and shown in Appendix N. Finally, there are two empirical distribution functions used to generate fares for arasags and ideal departure/arrival times for pags. These employ the EDF routines in Appendix N.

**7.4.5:** The second criterion suggested by Grimm and Railsback is the estimation of how valid the models themselves are. The third criterion is the appropriateness of the data and estimation techniques. The type of data and statistical techniques are treated in depth for each of the models used in Chapter 6, where the data sources and estimation procedures are described. It should be noted that AirVM maintains set of default values for all parameters in all models, which are the values discussed in Chapter 6. However, there is significant evidence that suggests that many of the models parameters are market dependent. For example, the Chicago>Orlando market is apt to have group sizes bigger than the average OD because of the increased number of families traveling on vacation to Orlando, which is the site of both Disney and Universal theme parks, and hence a big destination of family holidays. This highlights one of the many industry data deficiencies that exist in the air passenger industry. These deficiencies are sketched out more thoroughly in Section 7.5. But fortunately these data problems do not affect the model structure as much as they do the parameter values for the models. So if better data is available, better model estimates are to be had. AirVM currently has the ability to carry individual coefficient values for any origin-destination pair in the world, thus accommodating individual model estimates if they are available.

**7.4.6:** The fourth criterion is the question of whether AirVM is addressing the things it was designed to address. AirVM is intended as a model of the purchase of tickets by airline passengers – the market for passenger air travel – given the dynamics of ticketing and airline network design. I submit it meets that standard. I can see no important element of the market missing from the agent-based model. Some elements of the market are not well developed, however. The differing sales channels – direct, travel agent, internet – are not yet treated in detail, for example. But for studying the effects of increased service, lower fares, or redesigning schedules on revenue or share, the program addresses this challenge more completely than any other available tool.

## 7.5: Overall Critique of AirVM

**7.5.1:** It is of vital importance that the reader understand that, at this point in time, the validation and calibration of AirVM is still underway. The inclusion of AirVM in this dissertation by no means implies that AirVM is suitable or ready for any specific commercial application, and the acceptance of this thesis in fulfillment of the requirements for admission to the PhD for which it is written does not imply such suitability in any way, and the reader is explicitly cautioned that such an assumption is *prima facie* invalid without additional verification and calibration work. As this is written, Virtual M1nds SA, the owner of AirVM, is undertaking primary research to generate the data required to validate and calibrate AirVM. This effort has been initiated because there is convincing evidence that no single or combination of data sources currently in the industry is adequate for validation studies. Thus original data must be collected. While certainly of interest in this discussion, such calibration data is not central to the primary focus of the work, and waiting for it to become available would unduly delay the completion of the work.

**7.5.2:** AirVM is the first example of a virtual market to be implemented using the application of agent-based modeling to marketing science as proposed in this dissertation. It is therefore the first case of the implementation of agent-based modeling which is founded specifically on the notion of virtual markets, using the narrative construct to frame the structure of the agents used in the model. How well does AirVM represent of the nature and capabilities of agent-based modeling when considered in the contexts of both marketing science and operational marketing research? As noted in Section 7.4 it seems to satisfy the Grimm and Railsback validity criteria, and Virtual M1nds has engaged in extensive calibration work using data from a variety of airline industry sources. As might be expected, there is substantial industry secrecy surrounding such efforts, given the potential for a significant competitive advantage to be gained by the organization that operates AirVM with sufficiently accurate empirically calibrated parameters. However, the structure and logic of AirVM is open for all to see. Virtual M1nds made the decision at the outset of the development process to invoke intellectual property protection through the patent process, as opposed to trade secret or other methods precisely so that the inner working of the simulator can be examined and judged – and certainly improved – by both academic scholars and industry experts. Part of the rationale for this position stems from my experience and that of my colleagues in the airline industry. The natural business secrecy that is to be expected in any industry can be most destructive in the airline world. Airlines, like other forms of transportation, cannot avoid assuming a public welfare function, which argues for behavior that is more in line with the general public good than just that needed in a simple market place. Fifty years ago, passenger air travel was a luxury. Now it is tantamount to a necessity for most modern cultures. Patent protection allows for – indeed requires – the public examination of how something is done, and the quality and efficacy of the subject invention is thereby validated both scientifically and operationally. In other words, the true measure of the value of the agent-based virtual market concept is its adoption by the industry. Again as noted earlier, it has determined that no combination of sources is sufficient for a reasonable validation, so it has set about a course to create the necessary data.

**7.5.3:** Does AirVM accurately portray the dynamics of the air passenger market place? This question can be examined from at least two viewpoints. First, is the structure of the real market as understood by those most familiar with it accurately and validly represented by the virtual market? The world's passenger market cannot be known with complete certainty. There is inherent variation in the itinerary choices individuals make. A bad experience with one carrier may lead the passenger to select another the next time. There is also natural variation in the demand, caused, for example, by local disturbances such as conventions or holidays. Local response to overall economic trends varies from market to market. Moreover, even if the total demand were identical day over day, passengers will book at different times from one trip to the next, which causes variation in the resulting loads because of the network connectedness. As noted earlier, these realities mean that it is impossible for AirVM, or any other airline market model for that matter, to be one hundred percent accurate.

**7.5.4:** However, repeated simulations (in Monte Carlo mode, as discussed above) can determine the *probability* characteristics of the loads and revenues, and these can be compared to empirical data to assess the similarity (or not) to actual probability structures in the field. Figure 7.17 illustrates a "typical" flight revenue Monte Carlo result. This is the load distribution on flight AA 1506 (American Airlines flight 1506, in the MIA>BOS market) where the OD demand is being varied according to a positive truncated normal distribution on all markets using the subject flight in an itinerary. One hundred iterations are used in this simulation. Note that the axis values are not given, since to do so might reveal revenue data that is closely held by American Airlines. This shows the distinctly bimodal nature of this particular market's revenue structure, which is both unexpected and difficult to explain. Literally thousands of graphs like this one

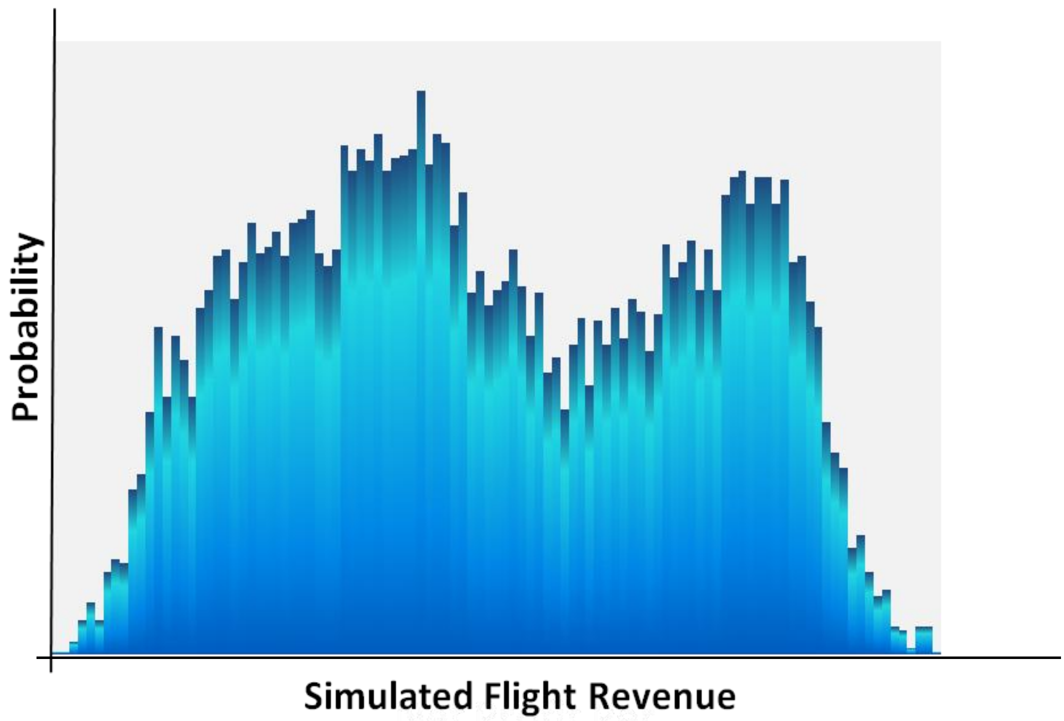



**Figure 7.17: Monte Carlo Simulation of Flight Revenue Distribution for AA 1506 (MIA>BOS)** (AirVM V.0.0.3568.16041. 100 Iterations)

can be prepared, few of them with even remotely similar patterns. If it were available, and because of a number of policies of American Airlines it is not, this distribution could be compared with actual ticketing detail from American Airlines for this flight, and the two compared. But gaining access to revenue data from an airline involves significant obligation to that airline, and Virtual M1nds is not in a position to undertake those obligations at this time.

**7.5.5:** From another viewpoint, the model's representation of some of the underlying structures of the ticketing process can be checked to see if the simulation is revealing structures similar to those found in the field. Consider a flight ticketing curve, for example. Because of the connectedness of the network, the ticketing curve for a flight is a composite of the ticketing curves for the markets being served by that flight. In AirVM, simulated market curves are based on passenger behavior characteristics, and are thus input to the simulated flight ticketing curve, which is an output. Comparison of observed flight curves and simulated flight curves thus provides a measure of the accuracy of the simulation. Figure 7.18 shows such a comparison. The ticketing

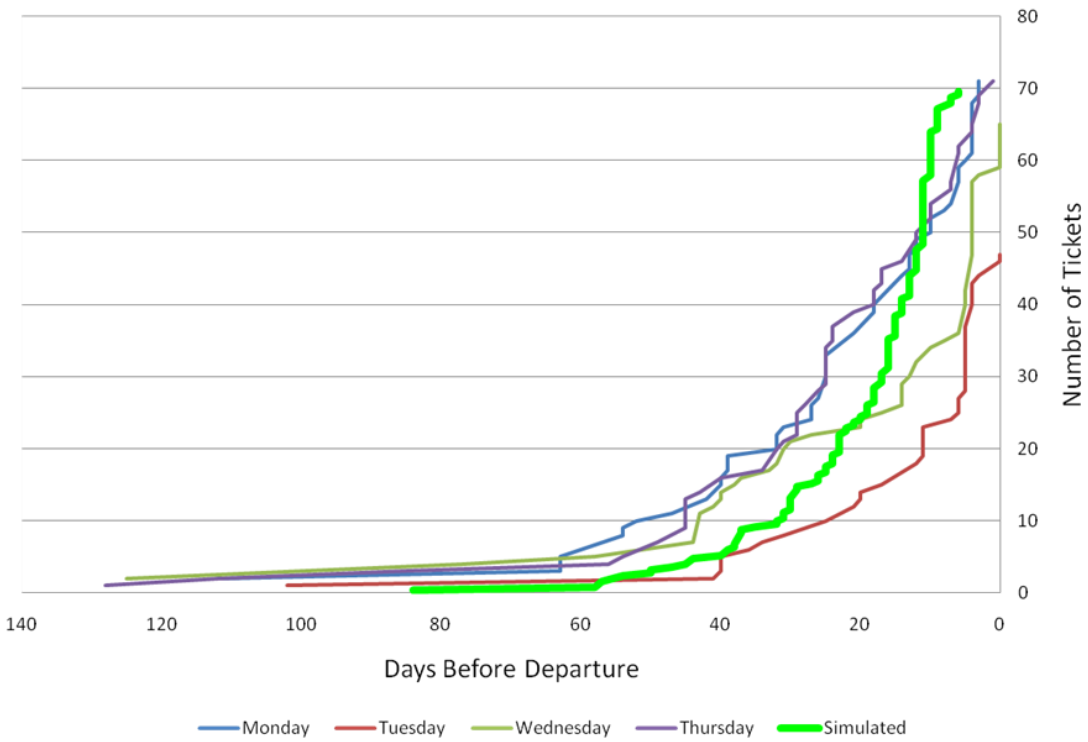



**Figure 7.18: Actual vs. Simulated Ticketing Curves. All Nonstop Segments in the SFO>NYC Market.** (Source: AirVM and ARC (2008).

curves revealed by ARC data (from August, 2008) for all nonstop segments serving the SFO>NYC for four different days of the week market are stood in comparison to the simulated ticketing curve for nonstops in that market. The correspondence seems reasonable. However, the 'bulge' in the 40 to 20 day before departure, also noticeable in Figure 6.9, is not represented by the simulated ticketing curve. The empirical variation across the days of operation of the flight shown is quite pronounced, however, perhaps indicating an area where AirVM's modeling should be improved. Still, the simulated flight ticketing curve replicates the observed curves within this inherent variation.

**7.5.6:** Consider the market share for a given flight, itinerary or carrier produced by AirVM. How does that compare when stood against an observed share value? The probability characteristics revealed by the simulation enable computation of the fit with reality, and that fit is good. The graph in Figure 7.19 is an example of a flight share comparison. This compares the load share for four nonstop itineraries in the SFO>NYC market (AA 16, AA 18, AA 20, AA 24). The share data is normalized across these four flights for ease of comparison and to mask the actual share carried by American Airlines. As can be seen, the matchup is not perfect, but quite close. Calibration of the underlying formulation with real data from actual operations continues to improve the quality of its results. In addition, the models, and the data used to calibrate AirVM in its present version, originate in the United States. Accuracy outside the US is yet to be

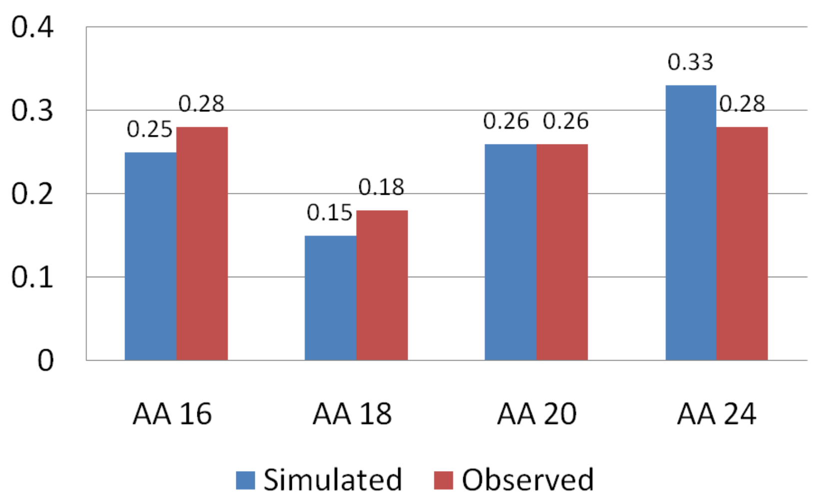



**Figure 7.19: Actual vs. Simulated Market Share. Four Nonstop Flights in SFO>NYC Market.** (Source: AirVM and ARC (2008))

tested. But for that area of the world on which the model parameter estimates are based, the results are good.

**7.5.7:** Only a superficial consideration is needed to challenge the applicability of the itinerary choice model used in the ratiocinator of the pag. It is both intuitively likely as well as empirically substantiated that other factors have a significant impact on the choices an individual makes of his mode of flying. To apply a single discrete choice model, even one with random coefficients, to all passenger choice in all markets is naïve. For example, research by Bunch, Louviere and Carson (2002) suggests that the value of air travel time varies across cultures, with individuals raised in Japan or China having a more acute value of time, whereas those originating in Europe somewhat less. However, a much more thorough data collection is required to establish the particulars of how that choice protocol ought to change for individual passenger throughout the world. Does this deficiency invalidate AirVM? I think not. Any model of air passenger behavior will have the same problem, because it appears to be an empirical fact that itinerary choice varies around the world, and any model must recognize it. Moreover, while a rational choice protocol was chosen in the AirVM case, there is no reason to restrict the passenger choice protocols available to pags to only that form. In fact, one of the useful aspects of a virtual market is to be able to test the sensitivity of the market to changes in the structure of the ratiocinators of important agents in it.

**7.5.8:** A similar set of considerations regarding accuracy can be posed with respect to the various support models used in AirVM. To what extent are the ticketing curve and group size models representative of what is actually seen in practice? What other model forms might be more appropriate? What data would they require to substantiate their validity? How sensitive is the behavior of the virtual market to changes in model parameters? The models suggested in this thesis are simple, and fit the data available. However, as validation and calibration efforts continue, the need for better data, and the rationale for the generation of that data becomes more pronounced. And, in fact, that drives the creation of such data. Already, for example, data sources such as the BSPs have begun to consider capturing journey data in a way that would fit AirVM, which will enable a much deeper understanding of the ticket purchase decision in the context of the kind of journey being undertaken, which in turn implies a dramatic improvement over current practices with respect to pricing flights, and hence revenue management.

**7.5.9:** AirVM is a true agent-based model, so it can be used to explore many aspects of the air travel market hitherto out of reach. The long term evolution of the market under various regulatory schemes, especially unexpected, emergent behavior, can be developed and studied with the current version of AirVM. One specific example is the exploration of the environmental impact of the world's airline network. This can be estimated by AirVM with the addition of carbon-production performance data which is available from the airline manufacturers. The two major network topologies touted by the industry – point-to-point vs. hub-and-spoke – can be easily contrasted and compared. It is even conceptually feasible to test the empty core hypothesis put forth by Button and others mentioned in **5.2.8**.

**7.5.10:** In my view, the value of AirVM has already been established. And here the power of the agent model paradigm becomes apparent. That questions can be formulated about the appropriate choice protocol, or the parameter estimates of the supporting models, or the importance of changes in how the dynamics are represented is, in and of itself, justification for AirVM. As with all science, the agent model is itself a hypothesis, and demands to be rejected by better science. But it is, at the root, a falsifiable statement, and that is its ultimate scientific purpose.

# Chapter 8:
# Conclusions and Contributions

## 8.1: Conclusions from the Research

**8.1.1:** The overall subject of this dissertation is really scientific instrumentation. To paraphrase eminent physicist the late Richard Feynman, science is the skill of observing and forecasting. Observation identifies what has happened. Forecasting says what will happen, and observation indicates to what extent the forecast was true. All science is based on these two simple ideas, including, most certainly, marketing science. The instrument of interest in this discourse is the form of computational science known as agent-based modeling. Agent-based modeling is a structural modeling tool which describes a complex phenomenon from the ground up, starting with the behavior and attributes of elementary objects in the system, delineating their mutual interaction, and creating a computer program which can portray the dynamics of the system as a whole as well as any of its component parts. The specific goal of the research, then, is to define and describe agent-based modeling in a way suitable for marketing science, study the properties of the necessary agents, formulate a general concept of an agent-based market model – a virtual market – and develop a case study example of such a virtual market in sufficient detail to test and refine the abstract concepts.

**8.1.2:** The research program to meet this overall goal consists of five objectives, or steps, as outlined in Chapter 1. The first of these was to determine the epistemological and ontological basis for agent modeling. What can be "seen" using this method that is invisible otherwise, and are those things important to marketing? Modern computers extend the human ability to observe and understand the world, but to what kind of real-world phenomena do they legitimately apply, and does marketing science qualify as a subject for their application? Chapter 2 takes up these questions. In that critical literature review, it was established that a) agent-based modeling is a valid form of mathematical, structural model building, and as such has as much standing in science as any other mathematical model that might be encountered, and b) because of its generative, heterogeneous and dynamic representation capabilities it is especially appropriate for application in evolutionary science. Moreover, evolutionary science as a paradigm for marketing science has received increasing affirmation in the marketing literature, and it thus seems a likely candidate for the successful application of the agent-based approach.

**8.1.3:** Even though significant scholarly study puts marketing into the domain of evolutionary science, those opinions seem based on rather elementary comparisons with biological evolution. The second step in the research program, then, to be explored contingent on the satisfactory response to the first step discussed in Chapter 2, is to test this hypothesis directly. That is, can human marketing behavior be derived from the observed evolutionary attributes of human beings? If so, two key implications emerge: first, models of human behavior that are necessary for agent model design, when validated in practice, can apply to many cultures and circumstances with only nominal adjustments since they are describing characteristics that can be found of all humans, regardless of culture; and second, the structure of the evolutionary heritage itself may suggest a corresponding structure that supports agent design and construction. The evidence, discussion and analysis of Chapter 2 strongly supports this hypothesis, and thus the two important consequences can result.

**8.1.4:** The third step in the research program is to define a framework for the portrayal and observation of human marketing behavior in sufficient depth to use it as a template for models of human agents in real markets. While it is quite reasonable to study marketing as it occurs in a variety of contexts – competitive dynamics, organizational theory, theory of the firm – eventually the discussion must get down to buyer and seller. It cannot avoid examination of human behavior. An instrument that can capture and display this behavior at the individual human level will provide, *a fortiori*, better forecasts of marketing results. Chapter 3 proposes the outlines of just such a general structure, called the narrative framework. The key conclusion from the analysis is that human behavior in marketing contexts is driven by narratives – time-based cause-and-effect chains – the central property of which is events. Events are instances in time where the narrative owner must apply resources to make a choice of future action. They are marketing actions because the resources must be applied. Choices are the focus since otherwise there is no change in the environment, no causes to yield effects, and narratives will proceed on course without intervention. Based on this conclusion, Chapter 3 goes on to offer a sampling of various choice protocols, all of which have legitimate representations that can be built into computer agent objects. The most important conclusion of the discussion is that agent based models for marketing science must be specifically concerned with modeling the choice structure relevant to the product or services being marketed as it relates to the agent's narrative.

**8.1.5:** Chapter 4 moves to the fourth research objective, which is the practical side of the application of agent-based models to marketing science that I refer to as *virtual markets*. The models must be realized. The instrument that is implied by agent-based modeling must actually be built. The instrument is a computer program written in real computing languages to run on actual machines. The important computing concepts on which the implementation of agent-based models rest are object-oriented programming (OOP), parallel processing, and messaging architecture. Agents are independent entities, and as such can be programmed as objects in an OOP structure. There is technique, skill, and a degree of art required to carry out that construction. And because these models extend the human ability to understand, how skillfully they are built directly affects how useful they will be. Chapter 4 therefore offers a general technical definition of agent. An agent has four required components. It must have a memory allocation called a state vector, used for the storage of the current state of the agent and its environment, the relevant history that led to the current state, and the narratives that describe the agent's concept of the future. It also must have three operating capabilities. Two more components govern the relationship between the agent and its environment – the perceptor for receiving and decoding incoming messages, and the actor for transmitting messaging indicating the choices made. The fourth component is called the ratiocinator, and is responsible for the choice functioning of the agent, which is the critical feature of the narrative as it relates to marketing activity. Any of the choice protocols that are described in Chapter 3 can engaged in this role, including irrational behavior, bias, or social choice protocols.

**8.1.6:** The agents in a virtual market are often members of synthetic populations, which represent either real or hypothetical groups of consumers, suppliers, or competitors. Synthetic populations are created with the use in incidence distributions, which describe the heterogeneity of the population observed in the market place or hypothesized in a research context. Virtual markets can be built for academic research purposes, obviously, but also are intended for use by management in the design, analysis and evaluation of marketing programs and policies. In such a role, the ratiocinator of an agent can be supplied by an actual person, such as a marketing manager or someone representing the marketing strategy of a competitor. Such a client agent is an avatar for management, and ideas and programs can be tested directly on the synthetic population of customers or competitors in the virtual market world.

**8.1.7:** The final research step examined in this dissertation is the design and construction of an actual working example of a virtual market. Chapters 5, 6 and 7 provide a detailed description and analysis of AirVM, an agent-based model of the world's airline passenger market. In Chapter 5 key features of that market are described for those readers who are unfamiliar with the airline industry. Chapter 6 goes into the details of the passenger synthetic populations in AirVM, including the mathematical descriptions and the data supporting the dozen incidence distributions used in the simulation. Chapter 7 describes the computer simulation program itself, including the logic flow charts, user interface, and results portrayals.

**8.1.8:** AirVM simulates the booking of every passenger in every city-pair market on every flight in the world for a one-week travel period. This scope includes about 42,000,000 passengers travelling in some 11,000,000 possible markets on about 1,000,000 individual flights. The critical dynamics of flights closing and fares changing during the ticket sale time before departure, due to flights belonging to multiple itineraries and airline revenue management policies, are all captured in the simulation. In AirVM marketing agents and choice protocols are defined and calibrated with operational data. Several aspects of the virtual market are explicitly specified, including synthetic populations with attendant incidence probability distributions, avatar agents, and the application of the virtual market as an instrument for more accurate demand estimation or data collection requirements. Because of the nature of the airline passenger industry, a number of specific managerial objectives are also met by AirVM. Finally, a discussion of issues of the efficacy and accuracy of the AirVM simulation, and a critique of its deficiencies complete the discussion. The AirVM simulation tool is now being deployed within the airline industry.

**8.1.9:** The five research steps have thus been achieved and their attendant objectives substantively met. Agent-based modeling is a worthy extension to the toolbox of the marketing scientist, since it portrays with finer acuity the evolutionary nature of human marketing activity. Because that activity is to a large extent universal among all human groups, a basic, yet widely applicable, framework can be posited which describes it.

Such a structure is suggested in the narrative framework, which in turn determines the definition and properties of vital elements of an agent-model in marketing science – a virtual market. And virtual markets can and have been built. AirVM is a specific example.

## 8.2: Contributions to the Literature

**8.2.1:** The most important contribution, in my view, of this research effort is the demonstration that agent-based models are a feasible and useful instrument in the array of tools available to both the academic and practicing market researcher. Real markets can be modeled using agent-based models, and new discoveries and insights gained by those in the academic world, as well as increased forecast accuracy and reliability for practitioners and the management they support. AirVM is an agent-based model of an actual market, with more than 40,000,000 customers weekly, over a thousand suppliers, and gross revenue in the hundreds of billions of dollars. Yet essential parts of it can be realistically simulated in about an hour, with reasonable fidelity and detail down to the individual flight and market. Many of the things asked of a marketing research tool are provided by AirVM. It provides a format and structure for the integration of a mixture of separate market research activities and results, from customer interview data for the estimation of a discrete choice model to industry-generated data for passenger on-board loads. This integration facility can be used to organize and streamline an otherwise incoherent, wasteful and inefficient company market research program, thus saving money and improving impact for the firm.

**8.2.2:** At the same time, virtual markets like AirVM point to data that is clearly required to model the market with any increased fidelity, but is not available. For example, there is no consistent source of passenger demand and preference data within the global airline industry. AirVM makes this deficiency quite clear, and describes in depth what is needed and how it should be presented. In addition, AirVM provides a mechanism for designing a data collection process that can exactly meet these unfulfilled requirements, efficiently and effectively. It is reasonable to expect that the development of a virtual market for any product or product class would yield similar results and opportunities.

**8.2.3:** For the academic researcher, virtual markets would seem to be a valuable tool. For example, AirVM could be used to investigate the "empty core" hypothesis on the deregulated structure of the industry, as suggested in **5.2.8**. Studies of the impact on and response of the industry to short and long-term external stress could be designed and implemented, such as another 11 September 2001 attack, another SARS-like epidemic, or the discovery of a rich natural resource in an otherwise barren part of the world (as occurred when oil was discovered in Alaska). Exploration of the impacts of other decision protocols – for instance the effects of various forms of decision protocol on the part of the pags – could be done with AirVM. The clarity of the potential applications of agent-based models to the scientific understanding of marketing science is abundant.

**8.2.4:** A second contribution to the literature of this work is the deepening of the empirical connection between marketing science and evolutionary science, thus setting the foundation for a paradigm that can be applied to address problems of heterogeneity and dynamics. These have been largely intractable with earlier approaches. System dynamics, the modeling method using differential equations developed about 40 years ago by Forrester (1975, 1968) and his colleagues and made famous by the "Limits to Growth" results that became popular in the 1970's, attempts to model the dynamics of a complex system directly. But the complex systems that describe evolution are inherently stochastic, and the best one can hope for is a probabilistic description of the evolutionary pathways that may unfold in the future. This is deeply and significantly different from the systems dynamic portrayal, but focuses the attention of the researcher on areas that will, I believe, prove far more productive as the understanding of marketing and markets grows. And the ability of agent-based models to specifically encompass elaborate heterogeneity within synthetic populations is in sharp contrast to, older, more aggregate methodologies. As the saying goes, with respect to marketing science, the devil is in the details, and the evolutionary nature of human marketing behavior is only well represented if those details are retained.

**8.2.5:** But the details that are so important to understanding marketing are not completely incoherent. Another contribution to the literature made by this dissertation is that human marketing behavior is universal across all cultures, and at least the beginnings of a better representation of that universality is suggested in the formulation of the narrative framework. An argument is developed and set forth that both empirically and deductively supports the role of *choice* as a basic, indivisible "atom" of marketing science, which gives the field another focus of its research attention. But the concept of the framework is only sketched out, offering enough detail only to support the definition of virtual market and its major properties. And it may be well pushed aside by more elaborate or general concepts. However, whatever form the concept may eventually take, the central role of choice in marketing will remain in place.

**8.2.6:** Furthermore, the narrative framework can be used as basis for quantitative advances in marketing research that include what have been considered hitherto as qualitative methodologies, such as ethnology. Anthropology and biology have established human universals, from which it has been argued that marketing behavior descends, and recent neuroscience (e. g. Montague, 2006) is beginning to lay a biochemical foundation of the pattern conceptualization that lies at the heart of the narrative structure. I would expect exciting progress in the next few years as the structure of complex behavior and its underlying support becomes better understood in the social sciences in general, and in marketing science in particular.

**8.2.7:** Finally, the narrative framework offers a structure on which behavior outside of what is normally termed “rational” can be described and formally included in analysis. This includes the characteristics and attributes of fads, dogmatic but clearly untrue beliefs, social momentum based on fear or prejudice, persuasive communications and propaganda, voting anomalies, early adoption and new product resistance. Behaviors based on these motivations can be included in the analysis of a market, and the results compared across studies. Marketing science is thus freed from utility maximization as the only basis for validly modeling consumer behavior.

**8.2.8:** The final contribution to the marketing research literature is an example of the process of how an agent-based model for application in marketing is actually designed and built in practice. The kinds of sub-models needed to build a representative synthetic population are delineated. What data is needed, how can be acquired, what’s missing from available data, and the impact of not having such data becomes clear in the construction process of a virtual market. Virtual markets are complex, but complexity in and of itself is no issue. Clearly complex marketing research demands more resources for data collection and analysis than does a marketing research program based on customer satisfaction surveys, but the example of AirVM, I believe, highlights the extraordinary added value of the results. The validity and practicality of a virtual market is demonstrated by the construction of an actual, working virtual market simulation.

**8.2.9:** In addition to the contributions to the literature of marketing research and marketing science summarized above, there are also several contributions to the literature pertaining to passenger air transport planning and analysis that result from this research, which to the best of my knowledge are unique and worthy of note. These results are of interest, of course, to the airline industry as they struggle to create a viable and stable business model for future development. Among these discoveries are the following.

> **8.2.9.1:** The mixed-logit model of passenger itinerary choice is the first choice model that explicitly includes desired vs. actual departure and arrival times. Frequency of service is a long standing problem in airline schedule design that has hitherto been poorly represented with the available arrays of analysis tools. It's incorporation into AirVM allows AirVM to shed light on the effects of flight frequency changes for airlines clients, competitive response analysts, and airport administrators.
>
> **8.2.9.2:** I suspect that the traveling group size distributions treated as a truncated Poisson probability distribution has been applied within the industry before, but there I have been unable to find any reference to it in the literature. It is, of course, critical in assessing the rate of ticketing (failure to adjust for the group size severely understates the realized demand at departure). However, certain aspects of the model, especially the differences in group size among origin-destination city pairs and with respect to trip purpose, are still subjects for further research.
>
> **8.2.9.3:** I am confident that the use of a Poisson stochastic process to represent the time-before-departure ticketing curve, also has been used in the industry, but have found no documentation to support that position. However, I am equally certain that expressing it as a compound, non-homogeneous process is an important advance in the state of the art. The verification of this model to portray the realization of demand over the time span of booking can provide a distinctive improvement over current methods used forecast demand-at-departure by airline revenue management systems.

**8.2.9.4:** AirVM is the first global market simulation at the individual flight level which encompasses all the relevant dynamics of passenger ticket purchase. While a number of simulations of various kinds have long been used in the airline industry, none have encompassed the entire world's system simultaneously. And this is a critical advance. One of the discoveries made with AirVM is that the world's airline network is everywhere connected. That is, changes in any part of the network – flight times, capacities, or routings – can have unexpected effects on any other part of the network. Proscription of a sub-network for analysis purposes is a common practice in the industry, but AirVM obviates the need for such requirements. From another perspective, the ability to model the interrelationships between various components of the network suggests methods by which the probability of an effect of specified impact on a given network component can be calculated, using Monte Carlo techniques. This would be quite valuable, for example, in studying disruption scenarios or constructing disease vector and remedial action strategies.

**8.2.9.5:** The construction of AirVM has revealed in sharp relief the quality and deficiencies in the industry data it uses for business and market planning. It has also suggested a remedy for at least some of these data issues. AirVM can be used as a very efficient basis for a sequential sampling method to be used for the collection of origin-destination data from passengers. Such data would yield direct measurements of a number of the parameters of the pag agent model, and thus improve the knowledge base of both AirVM and the industry. One method of approaching this would be to interview passengers during and after their in-flight experience. But to interview enough passengers to get a world-wide picture seems prohibitively expensive. In fact, however, only passengers at selected airports need to be interviewed. And which airports, which flights in those airports, and the estimated numbers of each can be extracted from AirVM. This can then be used for an optimum survey resource allocation plan.

**8.2.9.6:** In a similar way, the imputation method of origin-destination demand estimation and calibration discussed in this dissertation (Appendix M) is also a contribution to the airline industry that I have not seen previously reported. It is made possible by AirVM, and, when coupled with improved data from other sources, such as discussed above, might lead to much better econometric models of the generation and distribution of air travel.

**8.2.9.7:** Finally, the advanced form of dynamic flight itinerary generation used by the dsag agent in AirVM has not, to the best of my knowledge, been reported in the literature. This might find useful application in several areas of the industry, such as sharper and more relevant on-line choice generation, more accurate capacity planning, and related activities.

**8.2.10:** One other result ranks, in my opinion, as a valuable contribution to the literature, but neither in marketing or the airline industry, but rather in the development and application of agent-based models in general. AirVM demonstrates the value of an agent-based model as an instrument which reveals otherwise hidden characteristics of the phenomenon being studied. That agent-based models can do this has been long known. But to my knowledge this is one of the first cases where agent technology has been applied to study an existing dynamic human social structure such as a market. Such applications are being published with more and more regularity, and AirVM is a substantial contribution to that growing literature. As more and more results of this kind are developed, the value of the technique and its future application opportunities are significantly enhanced.

## 8.3: A Closing Word

**8.3.1:** Scientific research activities often raise more questions than they answer. And this is no less the case here. But the fundamental issue posed by this research – is it possible and feasible to apply computational science in the form of agent-based modeling to marketing research and science of marketing – has been answered in the solid affirmative. I have addressed and successfully explored the fundamental issues of the scientific pedigree of agent-based modeling and found it worthy of the attention of the marketers, the marketing researchers, and the marketing academics. I have constructed a framework for the modeling human behavior as an agent object in a computer program. Even though that framework is elementary, indeed even shallow, it is a viable starting point for the realistic portrayal of the universal properties of humans engaged in what can be reasonably called marketing behavior. And I have successfully constructed an agent-based virtual market simulation of an actual, in-operation consumer market of sufficient size and complexity to be of direct use to those who are undertaking the job of marketing and selling airplane tickets in that marketplace. While many deficiencies can be found, and many improvements readily made, I am proud of the effort.

# Appendices

# Appendix A: Agent-Based Models and Computational Science: The Perspective of Paul Humphreys

**A.1:** Consider the computational science aspects of agent-based modeling. These are important not only for practical reasons (regarding computer hardware requirements, execution speed issues, programming hardware and software architectures, and so on) but also from both an ontological and epistemological perspective. Parker and Perroud (2008) discusses an agent-based model of airline passenger behavior which routinely simulates the actions of tens of millions of passengers, effectively simultaneously. This is so large that under no conceivable human-based computational structure could an equivalent analysis be undertaken. Thus, in a very substantial sense, the agent-based model is an extension of human observational capability because it is able to dramatically increase the amount of empirically valid data. Each run of the agent-based passenger model would be impossible without the assistance of substantial computing capability which didn't even exist two decades ago! That is, each *observation* would be impossible without the computers. This aspect of computational science, of which agent-based modeling is a subset, is explored in depth by Paul Humphreys in his work *Extending Ourselves: Computational Science, Empiricism, and Scientific Method* (Humphreys, 2004, hereinafter PH04). I will now explore his exposition in some detail.

**A.2:** Humphreys views computational science as an extension of human observational ability. Such enhancements, he proposes, can be accomplished by extrapolation, conversion, and/or augmentation. (PH04, pp. 4-5). *Extrapolation* is when our human sensory sensitivity is enhanced, such as magnification with a telescope or microscope. *Conversion* occurs when one mode of sensing is converted to another, as in converting a sound wave into an oscilloscope graph. *Augmentation* is when features of the outside world not normally sensed by us are made available for our observation, such as infrared or ultraviolet light, or detections of subatomic particles. The same forms of enhancement also can be applied to our computational abilities. We are all very familiar with the extension of the speed of computations with computers, which is an illustration of the extrapolation concept. Conversion also occurs often in this realm, as with Magnetic Resonance Imaging (MRI) and Computer Axial Tomography (CAT) scans. Augmentation is encountered less frequently, but agent-based models themselves exemplify this concept, extending our deductive abilities into substantially more complex than hitherto possible.

**A.3:** In passing, let me note that Humphreys takes pains to point out that science neither has to be done for humans or by humans in order for it to be science. While it is perhaps heart-warming to think that we, the (presumably) cleverest species on the planet, are the sole creator and consumer of scientific fact, in reality yet another anthropomorphic center of the universe has been found to be an illusion. Science requires that data (observations) be collected and that data compared to predictions in support of theories. In many cases, for example the analysis of turbulent flow over an airplane wing, as many as $10^3$ observations per second must be recorded and compared. No human, or even a remotely feasible collection of humans, can do this, much less absorb and comprehend it. But a computer can manage this quite readily. As another, perhaps deeper example, also from the air travel industry, when next the reader is sitting in the departure lounge at an airport, take a close look at the housing on the engines of one of the newer Boeing 737 passenger planes. When looked at from the front, it has an asymmetric ovoid shape, not round like the older models. This shape has been found to aerodynamically superior to the older housing. But the specification of the shape exists only in a computer. It was devised by computer, and without the computer it would not be known or used. In a substantive sense, it was discovered and exists by and because computers exist. Thus, "scientific epistemology is no longer human epistemology." (PH04, p. 8) (This, by the way, raises interesting questions regarding empiricism, especially with respect to van Frassen's constructive empiricism (van Frassen, 2002, 1989).)

**A.4:** If computational science is an extension of our observational abilities – an instrument, if you will – then what are the salient properties of instruments that need to be understood to be convinced of their value? Humphreys proposes three: accuracy, precision and resolution (PH04, p 16-22). Accuracy refers to an instrument's ability to measure quantities that reflect the real value of those quantities. Precision means that the "empirical variance of the readings is small." Resolution is a relative term, wherein one instrument has greater resolving power than another if the instrument of higher resolution can distinguish at least one or more pairs of data points than the other, lower resolution alternative cannot. Humphreys offers a convincing case that these three properties are met by computational instruments. His argument is somewhat dense, but the essentials are reasonably accessible.

**A.5:** The *overlap* argument states that what we perceive through the use of instruments is valid to us if the instruments, when they are applied to phenomena with which we *are* familiar, reproduce those phenomena faithfully, that is, with known and acceptable precision, accuracy and resolution. This fidelity arises because we assert that if a device measures what we know well, then we can safely assume it will measure what we don't know equally well, and therefore will generate reliable observations in domains where we have no prior familiarity. Why do we believe this? Because we construct mental configurations of experience and believed fact under which this instrumental overlap is needed. To understand how this works, Humphreys employ the notion of *property cluster definition*, wherein objects are identified by a collection, or cluster, of attributes or properties. This viewpoint, for example, is at the philosophical heart of discrete choice theory, in that the distinction between one choice object and another rests in differing values of the defining properties. Instrumentation, then, allows for the discovery and recording of a broader range of attribute values, or, perhaps more interestingly, a wider range of salient properties. Further, he engages the *property localization principle* (PH04, p 41), which states that each instance of a causal property possesses the entire causal sense of that property. Thus it is properties, not the objects that possess them, which have causal effects. This justifies the use of instrumentation to study causal chains, since instruments measure properties.

**A.6:** Finally, Humphreys observes:

> "There is no reason for scientific empiricism to be so constrained by the highly contingent and limited resources of the human sensory apparatus, and it can be extended without overly weakening its traditional orientation. Epistemology need not be confined to the study of human knowledge obtainable by us as we are presently constituted. None of the

> positions traditionally falling under the category of empiricism can properly capture the knowledge by many new methods in the sciences. It is within the vast territories shunned by these positions that most scientific knowledge is located, and I take the view that instrumentally and computationally guided understanding is *sui generis*, not to be reduced to something more elemental." (PH04, p 48)

Continuing he summarizes the important contributions of advanced computational ability to modern science. To wit:

> "Computational methods now play a central role in the development of many physical and life sciences. In astronomy, in physics, in quantum chemistry, in meteorology and climate studies, in geophysics, is oceanography, in population biology and ecology, is the analysis of automobile crashes, in the designs of computer chips and the next generations of supercomputers as well as of synthetic pharmaceutical drugs, of aircraft, of cars and racing yachts, and in many other areas of applied science, computer simulations and other computations methods have become a standard part of scientific and engineering practice. They are also becoming a common feature is areas such as econometrics, many areas of the social sciences, cognitive sciences, and computer science itself." (PH04, p 50)

**A.7:** But these advances are more than just faster the faster calculation of simple additions and multiplications. These capabilities create a substantive extension of the *language* of science. Things that could not be even addressed before can now be spoken of with logic, clarity and precision. What would be impossible to compute – if for no other reason than the time it would take to make the calculation would engage every human living now, has ever lived, or would ever live (given the current understanding of the biological world) full time, without rest – can now be calculated in a few days or hours. And not calculated only once, but hundreds, thousands or millions of times. Thus it is feasible to explore a consumer market with alternative marketing approaches, appropriately modeled, under thousands of different conditions, considering each of the millions of potential consumers individually in each model iteration.

**A.8:** Several issues can be raised, however, to challenge the point of view that this computational skill is a material extension to human scientific endeavor, that this is truly a uniquely new instrument. One is as mentioned above – that this is just a simple extension of the speed of calculations. But clearly to be able to execute computations at a rate far in excess of anything else heretofore available to humans is a quantum leap in analytic ability. Supercomputers can process multiplications at the rate of $10^{14}$ times faster than humans. As Humphreys (PH04, p 53) succinctly puts it: "Speed matters." Many phenomena are now within reach that would otherwise be beyond any definition of feasibility.

**A.9:** Another issue that might be raised is that computation is not science, since it has no empirical content. But each computation in an agent-based model, for example, is a wff that has been deduced from agent properties, which themselves are perfectly capable of being empirically substantiated or falsified. Indeed, I would argue that empirically validating the structure of an agent usually is an easier task than validating the more complex realization of an agent-based simulation. This issue is also addressed, in part, when we distinguish what is mathematically representable by analytic means to what can be portrayed by computational means. (By analytic representability we mean expressible with equations and theorems that are of closed form.[80]) However, what is conceptually possible and what is within the range of the practical is an important consideration. Computational advances can enhance and extend the ability to represent a problem or theory that otherwise would be intractable. Humphreys comments:

> "To fully appreciate the scope of these [computational] methods, one must be willing to set aside the logical analyses of theory that have been so influential in the philosophy of science, to switch attention from problems of representation to problems of computation, and to relinquish the idea that human epistemic abilities are the ultimate arbiter of scientific knowledge. Our slogans will be mathematics, not logic; computation, not representation; machines, not mentation. " (PH04, p. 53)

**A.10:** A third source of criticism is that since the correctness of the processes underlying theoretical computational science cannot be verified, it cannot meet the basic requirements of the scientific method. But the verification of the underlying process will always be prone to errors that need to be corrected, errors that can and are corrected. The existence of mistakes in understanding a process does not *prima facie* invalidate that process. Such error correction happens in all science, all the time.

**A.11:** As an aid in understanding the value of computational science, Humphreys (PH04, p. 55) suggests locating the methods and tools of science in a three dimensional epistemological space, the three relevant dimensions being observational detectability, mathematical accessibility, and openness to experimental manipulation. He further asserts that "it is the invention of tractable mathematics that drives much progress in the physical sciences," and suggests the same holds true for other sciences as well – economics, biology, sociology and marketing science. The converse is also true. Scientific progress is often constrained by mathematical intractability. For example, the use of multi-dimensional normal random variables to express choice probabilities, as in the probit model (see Train, 2003), becomes computationally infeasible with manual methods after just a few choices enter the choice set. Computational mathematics opens a new avenue of tractability. The computation of areas under multidimensional normal model probability distributions is within reach, if sufficient computational power is applied to the issue. With computer-based computational models, we can do many more things, express many more ideas, and describe whole new vistas of theoretical constructs.

**A.12:** It is possible, indeed even likely, that the era of simple mathematics (that which can be expressed in a relatively few equations) effectively modeling parts of the world

---

[80] Roughly, the term "closed form" refers to the property of an equation or system of equations wherein any variable can be expressed as a function of the other variables. In the equation $y = ax + b$, *y* is solved in terms of *x*, and *x* can be solved in terms of *y*. This is not true if you try to solve for *t* in the integral equation $y = \int_{-\infty}^{t} e^{-x^2} dx.$ This equation has no closed form solution in *t*.

is coming to an end. All the simple stuff has been done. Only more complex, realistic descriptions will serve to advance science. This thesis alone, given the demonstrable lack of progress in marketing science since its inception, suggests approaching that science with computational methods in general, and agent-based modeling in particular. ”The era of anthropocentric mathematics is over.” (PH04, p. 57)

**A.13:** Most computational science implementations, in particular agent-based models, take the form of a computer simulation. This is because, to a significant degree, the dynamics (over time) of the model must be represented. Contrast this with the generally atemporal nature of classical mathematics. Humphreys calls the essential temporal aspect of a simulation the *core simulation*. Indeed, he goes so far as to exclude any simulation which does not have such a temporal core from his working definition of simulation. From this he goes on to conclude that

> “When both a core simulation of some behavior and a correct representation of the output from that core simulation are present, we have a full computer simulation of that behavior. And similarly, *mutatis mutandis*, a full simulation of a system.” (PH04, p. 111)

**A.14:** We noted above that computer simulations need no human interrelationships to be fully accepted as good science. However, the ability for a human to comprehend the essence and results of a simulation is a critical part of the simulation, and effectively links the simulation to human comprehension. “The form of the representation can profoundly affect the understanding of a problem, and because understanding is an epistemic concept, this is not at root a practical matter but an epistemological one.” (PH04, p. 114) Moreover, simulations can be replicated exactly, including their stochastic elements, as many times as are necessary for the desired level of precision, something not generally possibly with normal experimental conditions. In sum, the realm of the tractable, like the realm of the observable, is an expanding universe.

**A.15:** Humphreys goes on to describe how agent-based models overcome the “template” problem of the relationship between agent and environment, so the dynamics of a system become represented directly and precisely. Richard Feynman (Feynman, *et al.*, 1965) is famous for having said (among a number of much more profound things) that “the same equations have the same solutions.” From this comes the notion of computational templates, which are mathematical structures which can be applied across a wide spectrum of specific contexts. This notion of templates corresponds to Kuhn’s idea of “symbolic generalizations.” See Kuhn (1996, page 182-83) for details. Thus agents, when appropriately designed, can act as templates in the sense that they can be applied across a variety of contexts without substantial change. The creation of templates follows five “construction assumptions:” 1) ontology; 2) idealization; 3) abstraction; 4) constraints (including laws); and 5) approximations. The ontology describes the template objects and their relationships. The other features are obvious. (PH04, p. 78). All of these features apply to agents, and therefore the construct can fulfill a basic responsibility in terms of scientific description.

**A.16:** Humphreys also highlights some of the dangers of agent-based modeling. One in particular needs comment. A frequently cited requirement for successful agent models is that the behavior of the agents be as simple as possible (e. g. Epstein and Axtell, 1996) while producing the desired macro properties. However, like any simulation where the input settings can be adjusted to fit the results to empirical data, one can be led to believe that simple representations are the actual causal factors of the emergent behavior. In fact, there may be many patterns of behavior, some not so simple, that yield the observed emergent results. In the absence of any information that contravenes the presumed simple agent behavior, Occam‘s Razor argues for the inference of the truth of the simple explanation. But rarely is that simplicity supportable by empirical fact. With respect to human agents, rare indeed is "simple" behavior ever seen. Being able to duplicate macro behavior with simple agents does not support the contention that the agents are, in fact, behaving simply. In sum,

> "Because it is often possible to recapture observed structural patterns by using simple models that have nothing to do with the underlying realty, any inference from a successful representation of the observed structure to the underlying mechanisms is fraught with danger and can potentially lock us into a model that is, below the level of the data, quite false." (PH04, p. 134).

This critique applies to all simulations, not just agent-based models. What is left out from simplifying assumptions may turn out to be central to the causal structure under study. That's why they are called experiments. And, of course, the very nature of structural modeling requires faithful representation of the underlying process of interest.

**A.17:** Agent-based modeling could be also accused of suffering from a crippling case of what Humphreys refers to as *epistemic opacity*. This opacity results from the inability to observe, and hence explain, all the steps from initial state to the final output of a simulation. If the analytic requirements permit working with averages and other statistical summaries (as is the case, for example, of trying to find equilibrium conditions) then this opacity is not a substantial problem. But if that is not the case, as is certainly true of most agent-based models, then the issue must be addressed. Two kinds of opacity are apparent; computational speed, where the sheer number of calculations is beyond human review, and so-called "irreducible processes," where a process can most simply (and perhaps only) be described by the process itself (PH04, p 149). Agent-based models are very likely to be irreducible processes, where they "occur in systems in which the most efficient procedure for calculating the future states of the system is to let the system itself evolve." Such computational irreducibility does not *a priori* disqualify a procedure with that property from consideration as a modeling tool. The inner workings can be examined on a slow, step-by-step basis verifying that the components of the simulation are working as expected, thus increasing confidence in the outcome of the whole. Further, that the simulation can be exercised repeatedly to test input parameter sensitivity and to attempt to duplicate empirical findings also lends confidence to a represented theory.

**A.18:** Humphreys' analysis supports significant aspects of the key features of an agent-based model as set out by Epstein (2006). Homogeneity of individual population members need not be assumed, and so there is no need to use the concept of "representative individual." Heterogeneity can be an inherent part of the scientific explanation being formulated. Indeed, in marketing the ability to capture that heterogeneity is a significant, if not absolute, necessary advance in the state of the theory. Because of this ability to manage heterogeneity, agent-based models may well lead to a fundamental reorientation with respect to the scientific approach to be made to marketing science, such as the relaxation of the requirement for fixed preference orders within a consumer population.

**A.19:** Moreover, the agent-based modeling construct may resurrect a discussion of methodological individualism, the philosophical position that all complex phenomena can be explained by sufficiently complete understanding of its individual, component parts and how they fit together. From an agent-based modeling perspective, it argues that with a sufficiently sophisticated and detailed description of the nature of the agents, and their interaction with each other and their environment, the properties of the whole system can be derived and therefore explained. It is thus contrary to the holism concept, and runs counter to the view of emergence as a "mystical" entity that arises out of the properties of the inter-agent relationships, and cannot be foreseen by understanding the properties of the agents alone. Epstein clearly favors this view, when he refutes, at length, the mystical characterization of emergence. In fact, he goes on to assert that – indeed it is the backbone of his philosophy – such holistic emergence does not exist. Since agent-based models are wff's, the emergent properties that are observed when agent models execute are theorems, and therefore, in the most strict sense, deducible from the assumptions of the model, themselves wff's. Recall: "If you can't grow it, you haven't shown it."

**A.20:** As a final comment on the value of agent-based modeling, I note that showing that something cannot be done in principle dictates that it cannot be done in practice. Many of the most important epistemological results of the twentieth century are negative facts: the Heisenberg Uncertainty Principle, Gödel's Proof (Gödel, 1938), and the Arrow Impossibility Theorem[81] (Arrow, 1963). But establishing that it can be done is principle is no way implies that a practical realization is readily available. Science must be useful, and therefore must offer practical results, not merely abstract principals. Moving from principle to practicality would seem to be another compelling advantage of agent-based models. For example, if I assume that all human reasoning is, in fact, bounded (by resources and time, if nothing else) then marketing science models which represent such variation from *homo economicus* will be highly valued. Indeed, I welcome the expansion of my limited rational abilities implied by agent modeling, much like I appreciate the telescope as an enhancement of my powers of sight.

[81] Referred to as the PossibilityTheorem by Arrow himself.

# Appendix B: A Simple Agent-Based Model: The Genetic Algorithm

**B.1:** In a genetic algorithm (GA), the 'species' of interest is the set of possible solutions to some properly specified problem. For example, if the problem at hand is the routing of a delivery truck among several destinations (the so-called 'Traveling Salesman' problem[82]), then each possible ordering of the destinations represents a possible truck routing. An individual in the subject population is represented by a string of symbols called, appropriately enough, a chromosome. Each symbol (or selected combination of symbols) itself represents the value of some characteristic needed for the problem solution (and is thus called a gene). A particular member of the population represents one solution to the problem at hand, that solution being described by a string of attribute values of salient features of the solution. In the Traveling Salesman problem, the order of destinations is one solution to the problem. The worth, or fitness, of the particular solution described by an individual is determined by a so-called *fitness function*. A fitness function takes the attribute levels of an individual solution and determines how good that solution is for the problem at hand. It determines the fitness of each solution in a measure suitable for the problem.

**B.2:** It is how GA's mimic reproduction that clarifies how variation is created and maintained in the population. In the typical operation of a GA, a number of individuals are initially defined with random values of the characteristics encoded on the genes in the chromosome. These characteristics are usually arranged linearly, as though on a string, with the order of the variables represented by the string the same for all individuals, analogous to the structure of a biological chromosome. Thus the gene strings for all individuals are the same length and represent the same features in the same places on the chromosome. These individuals then "mate" with each other to produce additional individuals. This is where a GA closely follows the (current) understanding of genetic processes. Breeding within a GA is done by matching two individuals, then creating two new individuals by randomly changing their genetic structure by means of a *crossover*. A crossover occurs as follows: Each parent contributes one chromosome to each offspring. The offspring's chromosomes are created by breaking each parent's string at the same random, but identical, point, and creating two new chromosomes by exchanging the broken strings and joining them at the point of the break. Crossovers are known to occur with actual chromosomes in living organisms, usually during egg fertilization. The offspring of such a mating shares some characteristics of each parent individual, through the exchanged values of the variables on each side of the crossover break, but the resulting chromosome is a viable representation of the species. Thus the degree of variation within a population is maintained through subsequent generations of individuals in those populations. Notice that a population with little variation will have much more difficulty creating individuals with different abilities to respond to changes in the environment, while a population with a high level of variability is more likely to have at least a few individuals that can survive environmental change when such should arise. In biology, crossover is far and away the most important mechanism for the propagation of

[82] The Traveling Salesman problem is one of the more famous early success stories of the application of genetic algorithms. See Haupt and Haupt (2004), pp 151-153, for details.

variation within a population. The same is true of GA's. But if the degree of variability within the population is small, it will not get substantially larger from generation to generation. This is where *mutation* comes in. A mutation is the random alteration of one value of one gene in a chromosome of a randomly chosen individual, basically creating a new individual with a new genetic makeup.

**B.3:** GA's work by breeding successive generations of solution individuals. Each member of the new generation is evaluated using the fitness function, the most fit kept for further evolution, and the remaining individuals removed from the population. Over time, the fitness of the fittest solution continues to increase, and the process continues until some threshold of sufficient fitness is reached, such as an insignificant improvement from one generation to the next. There is one area, however, where a GA does not reflect genetic reproduction in nature. The assumption is made in a GA problem that the environment, as defined by the fitness function, does not change over the course of the solution time. Thus eliminating the weakest (least fit) members of the population poses no risk to the process, and survival of the fittest does indeed occur. However, in nature, the "fitness function" is not used to keep the fittest, but remove only the number of least fit necessary. Individuals only slightly more fit than the bottom of the fitness scale may remain, and may prove significantly more able to survive if the environment changes.

**B.4:** An example of a genetic algorithm can be found in Parker and Zhang (2007) and Parker (2009). In this case, the "species" consist of a set of eight airplanes which are flying between ten cities. The objective is to find a routing of each airplane among two or more cities which results in maximum profit for the entire operation. Each routing has an associated cost, usually computed by summing fixed costs plus per passenger costs plus per mile costs plus per passenger-mile costs. Revenue is estimated by applying the AirVM agent-based air passenger simulator described in Chapters 5 through 7. Research to date has established that this GA seems to have no stable optimum species member. That is, while evolution certainly occurs and a "best" routing for the eight aircraft is found in each case, repeated runs of the GA with different initial random arrangements of the flight patterns yield drastically different final routings. Profitability (the fitness measure) can vary as much as 100% from one GA execution to another. This shows the extent of the path dependency in this particular GA structure, demonstrating that even relatively simple evolutionary structures can create quite complex and challenging features. For more on GA's, see Goldberg (1989) or Haupt and Haupt (2004).

# Appendix C: Detailed Discussion of Marketing-Related Human Universals.

**C.1:** From Brown's (1991) synopsis, I can identify at least 76 human universals that seem related to why humans engage in and respond to marketing. The following have extracted from Brown's narrative (1991, pp. 130-141).

### C.1.1: Culture:

77. Humans are aware of culture, and behaviors and thought processes are spread among individuals and from generation to generation by learning.
78. We are aware of our apparent uniqueness among known living things in having culture, although it is not strictly true.
79. All human cultures use speech, often ceremonial or otherwise distinguished from ordinary, day-to-day usage, to tell stories that explain how things came to be or what they will become.
80. All human cultures share a common external world and internal existence.

### C.1.2: Language:

81. All human cultures have language. We use language to think about and discuss among each other both our internal states and the states of the external world around them.
82. Language is not a perfect representation of a human's views or thoughts. There are discrepancies between what we think, what we say and what we do.
83. Because the use of language is not a literal description of the world and our place in it, we must, and do, distinguish between the world as it actually is and the world as we conceptualize it.
84. Humans use language to organize, respond to and manipulate other humans.
85. An important means of verbal manipulation is gossip.
86. Language is used to misinform as well as inform.
87. All humans lie and mislead others, some to a greater extent than most.
88. All humans understand what a lie is and have methods of detecting it when it occurs.
89. Those who are more proficient in the use of language are of higher status, in the view of others, and are more capable of manipulating the behavior of others.
90. Language is highly symbolic in structure, and even though the exact sounds and words used to denote a given entity or concept may differ, rather arbitrarily, from one group to another, all languages contain a basic set of common features. Thus different cultures can communicate.
91. All languages have a similar structure in that they contain nouns, verbs, and possessive forms.
92. A few concepts are found in all languages, such as *face*, *black* and *white*, *male* and *female*.

93. Male and female gender is always distinguished in languages, creating a sexual content in all human language.
94. Our sexual terminology is dualistic – male and female. Other terms can portray different sexual contexts, but only in reference to this basic duality.
95. All languages describe family and kinship relationships in terms directly related to procreation, such as father, mother, daughter, etc.
96. All languages express the concept of time, and units thereof.
97. All language contains groups of contrasting terms that could be expressed in three ways. For example, we could contrast *good* with *bad* by saying *good or bad*, *good or not good*, or *not bad or bad*. But, interestingly, the third case never occurs in ordinary language.
98. Language changes over time, adopting new words and letting others fall into disuse. Language structure changes more slowly, but does change nonetheless.
99. All languages make extensive use of metaphors.

**C.1.3: Analysis**:

100. Humans learn by trial and error, by which is meant if some event or sequence of events does not occur as expected, our *expectations* change, not the perceived events or sequence of events.
101. Humans can measure and express the size of things.
102. All human cultures use language to describe physical properties – speed, motion, dimensionality – and actions, such as giving, lending, and affecting other things and other people.
103. All human languages have terms which identify and discuss parts of the body, internal states such as emotions or thoughts, behavior, the physical world, the weather, tools that are available, and space.
104. An essential feature of our ability to recognize, learn and understand the world is that we can identify distinctions and name them; that is, we can create taxonomies.
105. The basic form of distinction is that of binary discriminations, the basis for all categorizations (a thing either is, or is not, in a specified category, or has, or has not, a particular property).
106. All human cultures can also classify things that do not necessarily fall into discrete categories by ordering them along some kind of continuum.
107. All human cultures use basic logic constructions such as “not,” “and,” “same,” “if … then,” “equivalent,” and “opposite.”
108. Humans have the ability to deduce, from a variety of even subtle clues, the current state and future condition of things, often very inaccurately.

**C.1.4: Recognition of mind:**

109. All individuals have a distinct concept of person and of themselves separate from others.
110. We can easily and intuitively get into the minds of others and imagine what they are thinking and feeling.

111. We know that other people are like us have inner thoughts, make plans, and make decisions and choice.
112. We are able to think not only of our own relationships with others, but of the relationships between others and themselves.
113. All of us also mask or modify expressions to mislead and confuse.
114. All of us communicate non-verbally, especially with facial expressions.
115. We respond to sexual attraction.
116. We use reciprocal exchanges in all aspects of our lives. This includes its negative forms, such as retaliation and revenge.
117. We are able to distinguish between normal and abnormal mental states.
118. Humans have ways of making themselves feel better, including the use of stimulants, narcotics, and intoxicants.

**C.1.5: Tools:**

119. We make and use far more tools than any other animal, and use tools to make other tools.
120. We know how to use fire.
121. All cultures have means of shelter from the elements.

**C.1.6: Group association:**

122. Most of us live part or all of our lives in groups.
123. We have a sense of territory associated with the groups to which we belong, and are well adapted to the environment associated with that territory.
124. There is a distinct sense of "us" and "them" between the groups to which we belong and other groups.
125. We judge members of other groups in terms of those qualities found in our own.
126. Marriage, in the sense of a well understood protocol for access to females of child-bearing age, exists in all cultures.
127. We have a pattern of socialization in the sense that child-rearing responsibilities are shared among adults, who are obliged to teach the young.
128. Children learn from adults by copying and mimicry.
129. The existence of roles and social structure is implied by kinship, sex and age statuses found in all cultures.
130. Prestige is not equally distributed among group members.
131. An ethical dualism exists between the groups to which we belong and other groups.
132. We have leaders, but they are never completely powerful.
133. There is never a complete democracy nor an absolute autocracy. Thus there is always an oligarchy.

**C.1.7: Economic behavior:**

134. All societies have economies – systems for barter or trade.
135. We engage in trade, defined as exchange of goods that is not based on future expected reciprocal behavior.

136. All societies have divisions of labor, and customs of cooperative labor.
137. We are all materialists to some extent, and distinguish who owns what.
138. Members of groups are not economically equal, in that some have more materiel resources than others.
139. We are envious.
140. We give gifts.

**C.1.8: Ethics, morals, art and metaphysics:**

141. We all have a definition of, and can distinguish, right from wrong
142. All cultures have standards of sexual modesty
143. We all believe things that are clearly and demonstrably false.
144. We all practice magic, especially in trying to prolong our lives or attract others in a sexual context.
145. All humans attempt to predict and to plan the future.
146. All cultures have theories of fortune and misfortune.
147. We attempt to control the weather.
148. We all have a coherent world view, in that we understand the world as a unitary whole regardless of the sensory mode by which it is experienced.
149. Our worldview plays into our mythology and concepts of the supernatural.
150. All cultures have rituals, especially to demarcate changes in states, and all cultures mourn the dead.
151. All cultures have aesthetic standards and preferences.
152. All cultures have music, poetry, play and story-telling.

**C.2:** Language is one feature that is common to all collections of humans referred to as a *culture*. Four other features are categorized as cultural artifacts as well and have bearing on human interaction in marketing. Learning (Universal item 1) is essential to culture, and thus the economic structure which supports trade, hence markets, and hence marketing, is learned. That (we think) human culture is unique in the natural world (2) is a basic manifestation of the "without there being a 'them,' there can be no 'us' " phenomenon that lies at the root of altruism and group evolution. The identification of culture with *specialized* speech (3) creates the possibility of debate, polemic, politics, persuasion, and marketing. And the fact that we all share, within fairly broad parameters, the same external and internal world (4) creates the common frames of reference needed for the effective functioning of language in any context, and in particular in a socially shared narrative.

**C.3:** Language is commonly believed to be one of the characteristics that differentiates the human species from all others. While not literally true (the communications of dolphins and whales must be considered language by any useful definition of the word, and recently it has been suggested that elephants also have an effective language, transmitted at subsonic (to humans) levels), it is certainly the most prevalent vehicle for weaving together our social environment. The language group of human universals highlights various features that are created by this verbal capability. The universality and basic uses of language (5), and its relationship to the outside world (6 and 7) are self-evident. Universal items 8, 9, 10, 11 and 12 describe ways in which language facilitates interpersonal relations, that we lie and can detect lies, and the role of gossip. These points stress the characteristics of language use, which is trivially fundamental to marketing. Language as status (13) is also a valuable tool to the marketer. As might be expected, language structures have similarities across all cultures (14), and all languages share descriptions of phenomena and circumstances that are common to all of us (items 15 through 20). The three contrast modes that all languages share are described in item 21.[83] The last two aspects of language, however, to some extent stand alone. That language changes over time (22) requires that the use of language be characterized not only by the subject and form of the linguistic statement, but also by the time frame in which it is used. The things being described by language are changing, thus the language changes. The last item, the use of metaphor (23) will become important in the discussion of narrative in the next chapter.

**C.4:** The common behaviors listed in the Analysis Family collect together those features that form the foundation of the capability to consciously reason about ourselves and the world around us. Trial and error (24) implies that humans can take actions (make trials) and observe and remember the consequences of those actions (take note the errors). From this simple fact *expectations* are known to be universal (there can be no 'errors' without an expected outcome if error is defined as variation from that expected). Closely aligned with the notion of error and expected outcome are concepts of measurement and the linguistic devices to describe them (universals 25, 26 and 27). The ability to categorize and order, in both discrete and continuous contexts, is essential for any analysis, and especially for scientific analysis (28, 29 and 30). This taxonomic ability, coupled with universal concepts of (first order predicate) logic constructs (31), imply our demonstrated talent for *deducing* states of the world, albeit with often questionable validity. These universal skills together result from one of humanity's most important talents: the ability to recognize, construct and manage *patterns* with substantial degrees of complexity. Language is used as the vehicle for the instantiation of such mental patterns. This observation will lead me in to explore the narrative construct as an elementary building block of agent-based models of human market behavior.

**C.5:** The term *recognition of mind* connotes the ability of one human to put him or herself in the position of another, thereby being capable of some degree of understanding the situation that other individual finds himself in. That is, the ability to empathize and sympathize. This capability allows us to predict, with some (albeit perhaps small) likely success, the behavior of others by asserting that they would behave the same as we in similar circumstances. This predictive improvement makes other humans less dangerous to us, and leads to a variety of market-related aspects of behavior. Distinct individuality (33) is, of course, necessary for any of the other universals to function properly. Without awareness of *self* there is no concept of *other*, and virtually every other trait listed becomes impossible. The ability to put ourselves in the shoes of others (34) in a marketing context leads us to think of what others might want or do, which gives us a meaningful starting point for almost any transaction. The same notion applies to the appreciation of the behavior of others (35 and 36). Because of this inherent understanding of how others behave, we can use that knowledge to lie

[83] It is unclear why the third method – contrasting bad and good by using bad and not bad – is never used.

or confuse (37), and an important mechanism for such capability is non-verbal communications (38). Sexual attraction (39), naturally important to species survival, is also enhanced and managed through recognition of mind, and is perhaps the most elemental form of such, second only to mother-child empathy. Recognition of mind also seems to be a requirement of reciprocity (40), which, of course, is fundamental to any kind of trade or other economic behavior. It also the how we define abnormal behavior (41) by comparison with the behavior of ourselves or others. And I classify mood-changing agents (42) with this group (very much loosely) because of the effect on mental processes which lead to substantial distortions of the views of the mind of others.

**C.6:** Only three behaviors are listed under the heading of tools. This is not an indication of a lack of importance of tools to human kind, but the next level of specificity in such a discussion would get into types of tools, which would extend the description far wider than necessary for my purpose. Suffice it to say, then, that we are the most prolific of tool makers (43), but not the only species that makes and uses tools. We use tools to create and manage fire (44), a ready source of energy for a wide variety of uses, and construct shelter (45) as a method of creating and controlling, to some degree, our surroundings and environment.

**C.7:** Humans live in groups. Later in the presentation that group association will become vital to understanding how behavior evolves, so it is not surprising that a number of attributes of group structure show up as human universals. Group cohabitation (46) seems to imply a sense of territory (47) – home, if you will. And this leads to trait 48, the distinction between those who are members of our group (us) and those who are not (them). In fact, without a "them," there is no "us." This distinction also causes us to consider others outside of our group as different as defined by comparison to us (49). I include marriage (50), child-rearing (51) and education (52) as characteristics of the most elemental and universal human group, the family. Families create the opportunity for learning by mimicry, which is crucial to the development of narratives. Further group definition is maintained by roles (53), which also imply distinctions in the distribution of prestige (54). The us-them disparity is further punctuated by an ethical dualism (55). Finally, all groups have leaders (56), but ideal forms of government/social structures – dictatorship, autocracy, democracy – do not actually exist. Only oligarchies are actually ever found (57).[84]

**C.8:** Those universals classified as economic activity are fairly obvious. The existence of economies (58) and the pursuit of trade (59) seem synonymous, but can be distinguished by noting the difference between *trade* – exchange of goods – and *systems* for trade – methods set up to manage the exchange of goods. The divisions of labor (60), materialism (61) and wealth inequality (62) give rise to the need for trade. In fact, there may be a feedback mechanism in play here, where labor division leads to material possession inequality which leads to trade advantages and disadvantages, which in turn causes systems of trade to be instituted to the advantage of those otherwise without it. Indeed, this cycle has been the subject to a number of agent-based model studies. See, for example, Epstein and Axtell (1996), Anderson and Anderson (2009), Banerjee and

[84] This raises the interesting question of what kind of behavioral patterns would have to exist among individual in a group for any of these ideals to emerge as a likely group structure. This issue could readily be addressed by agent-based models.

Murphy (2006), and van den Kuilen (2009). Envy (63) and material generosity (64) are more complex behaviors, but clearly arise out of the material possession universal behaviors. Envy and generosity also stem from some of the behaviors classified as the recognition of mind family.

**C.9:** The last collection of twelve traits concerns ethics, morals, art, and more generally, metaphysics. These universals address how we form patterns from our experiences and learning that we use to safely operate in the social world in which all humans live. We all have a sense of right and wrong (65), and therefore some basis on which to define morality, and we apply that morality to the sexual aspects of our lives (66), perhaps because of the central role sexuality plays in survival. We all believe in magic, and practice it as we try to manage the future (67 and 68). Controlling the future is a universal trait (69 and 70), especially the weather (71). Clearly such ability is an adaptive advantage, for creatures that can successfully predict the outcome of some action are more likely to avoid survival-threatening circumstances. Such ability seems to imply the necessity for a coherent understanding of the world (71) and how it works (even if that understanding is not true or accurate), which relates directly to the pattern-forming behavior mentioned earlier. That worldview organizes and structures our universal cultural propensity for mythology and the supernatural (73) from which we derive rituals (74), aesthetics (75) and, in turn, art (76). All of these traits will come to the fore when the narrative construct is presented.

# Appendix D: The Evolution of Behavior: The Example of Altruism

**D.1:** Evolution, as it is often simplistically defined, is the survival of the more fit over the less fit. Traits which improve the survival of a species persist, those that do not fade away. But it is very hard to see the evolutionary value of altruism. The biological definition of altruism is behavior that *reduces* the chance of survival (Sober and Wilson, 1998, p. 17). The common social meaning of the word – the granting of a kindness to another for no immediate return – alludes to this inherently anti-survival characteristic. But altruism persists in spite of its evolutionary unsuitability. Mayr (2001), Brown (1991), Shermer (2004, 2008) and even Darwin (1859) offer substantial evidence that altruism is a basic human universal. Further, it is difficult to see how altruism is just an unused artifact of evolution left over from a prior era, or is just there because of evolutionary background. That is, is an *exaptation* (see the discussion in Gould, 2002). It would seem to be downright dangerous to the future of the species. But how did it come about, and how does it continue, if it confers no adaptive advantage?

**D.2:** It is not surprising that the continued existence of altruism has caused substantial consternation among biologists when it comes to understanding human evolution. One line of thought, that fostered by Dawkins (1976), the author of *The Selfish Gene*, requires that altruism cannot really exist, and that it only arises in the context of "reciprocal altruism," where a kind deed done at one point in time serves the purpose of helping to insure a reciprocal kindness later, when it is needed by the original altruist. But this theory is belied by the psychological evidence that surround acts of heroism. Clearly, a hero's actions are altruistic. But there is no substantial data to suggest that heroes are such because of a superior desire to accumulate reciprocal acts of kindness in the future. Indeed, heroism seems to act just the opposite, often being a spur-of-the-moment, almost spontaneous act. And, while a hero attracts the admiration of the fellows in his/her social group, such admiration does not seem to be a motivating factor in the act itself (Coon, 2007). In other words, heroic acts are a form of pure altruism. The altruist expects nothing in return. And very often, especially in time of war, heroism results in the death of the hero, which is clearly counter to the flow of evolution.

**D.3:** How the evolutionary biologist might consider altruism is suggested by Mayr (2001, pp. 256-260). He points out that defining altruism as an act that is beneficial to the recipient but at a cost to the altruist is incorrect. Instead, he proposes three kinds of altruism, and suggests a selection advantage to each. His three classes are distinguished by who the kindness is directed toward, and the importance of that individual to the species survival in the sense that stable groups and social cohesion have been selected for in the evolution of humanity. The three foci are 1) altruism for the benefit of one's own offspring, 2) favored treatment of other close relations, and 3) altruism among members of a social group to which one belongs. He discusses reciprocal altruism (but not the progressive variety explored by Caldini, (2001) and to that shown to distant outsiders. He argues that any altruistic behavior observed or expected of outsiders is only explained by learning behavior, especially on the part of philosophy and religious leaders. He also contends that assuming an altruistic act comes at a cost to the altruist is wrong. Indeed, he notes that many acts of kindness or generosity come at very little or no cost to the individual performing the act. But Mayr seems to be straddling the issue of whether group evolution exists. This issue, it turns out, becomes crucial to the general understanding of how evolution works.

**D.4:** A much deeper analysis of the evolutionary structure of altruism is offered by Elliot Sober and David Sloan Wilson in their book *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Sober and Wilson 1998, hereafter referred to as SW98). These authors lay out a convincing case for how complex behavioral traits like altruism can arise as a natural consequence of our evolutionary heritage. (Also, while they don't do so in their presentation, the application of agent-based modeling to the issue could be an effective demonstration of the validity of the argument.) By demonstrating that altruism can arise as a natural evolutionary phenomenon, they then assert that similar logic can be applied to show the feasibility of other complex traits, such as economic behavior and market convention.

**D.5:** Sober and Wilson's definition of altruism is:

> "… evolutionary biologists define altruism entirely in terms of survival and reproduction. A behavior is altruistic when it increases the fitness of others and decreases the fitness of the actor. The challenge for the evolutionary biologist is to show how such self-sacrificial behaviors can evolve, regardless of how or even whether the individual thinks or feels as it performs the behavior." (SW98, p. 17.)

Given the more "classical" definition of evolution, it becomes immediately evident that altruism cannot exist as an evolutionary trait, for it acts against the survival of the altruist by decreasing its fitness. That altruism does exists is an empirical fact, however. Sober and Wilson resolve the apparent contradiction with the introduction of *multilevel selection*, or *group evolution*. This explanation contends that not only is adaptation occurring at the individual organism level, but there is adaptation at the level of a collection of individuals that is not apparent, indeed would not occur, if the individuals were not acting in some way peculiar to the fact that they are part of a group.

**D.6:** A simple example will serve to illustrate how a counter-fitness trait like altruism could evolve if there are sufficiently strong conditions on the general population of individuals. First, it needs to be established that altruism in the absence of any other forces will be selected out of a population. Following an example first constructed by Wilson (1989) and reproduced in SW98, consider a population of $N$ individuals, who reproduce asexually, in which $p$ is the probability of being an altruist, and $1 - p$ the probability of being selfish. There are thus $Np$ altruists and $N(1-p)$ selfish individuals in the population. Fitness is measured in purely biological terms, that is, in terms of the number of offspring produced, and so let the cost of altruism be $c$ fewer offspring to the altruist, and the benefit is $b$ additional offspring to one other member of the population. Let $X$ be the average number of offspring per individual in the absence of altruism, so that fitness where no altruism is present is simply $X$. The expected fitness of altruists, $F_A$ and of the selfish $F_S$ if altruism is present in the population is, however,

$$F_A = X - c + \frac{b(Np-1)}{N-1},$$
$$F_S = X + \frac{bNp}{N-1}. \quad \text{(D.1)}$$

This is true because each altruist has a cost of *c*, but can expect to receive a benefit of *b* from the one among the $Np - 1$ remaining altruists. After reproduction, the number of offspring $N^*$ and the fraction of altruists among them $p^*$ have now been changed to

$$\begin{aligned}
N^* &= N[pF_A + (1-p)F_S] \\
&= N\left[p\left\{X - c + \frac{b(Np-1)}{N-1}\right\} + (1-p)\left\{X + \frac{bNp}{N-1}\right\}\right] \\
&= N\left[\left\{pX - pc + \frac{pb(Np-1)}{N-1}\right\} + \left\{(1-p)X + \frac{(1-p)bNp}{N-1}\right\}\right] \\
&= N\left[\left\{pX - pc + \frac{pbNp - pb)}{N-1}\right\} + \left\{X - pX + \frac{bNp - pbNp}{N-1}\right\}\right] \\
&= N\left[X - pc + \frac{pb(N-1)}{N-1}\right] \\
&= N[X + p(b-c)],
\end{aligned} \quad \text{(D.2)}$$

and

$$p^* = \frac{NpF_A}{N^*}. \quad \text{(D.3)}$$

If mortality is constant regardless of the presence or absence of altruism, then the population will return to size *N* through normal attrition. However, clearly $F_A < F_S$ at each successive generation, and eventually, with finite *N*, the altruists will disappear.

**D.7:** To see how altruism might evolve, however, consider the case where the population consists of two sub-groups. Let $N_1$ be the size of the first group, which has proportion $q_1$ of altruists, and $N_2$ the size of the second group, with $q_2$ fraction of altruists. Then Table D.1 shows the calculations described in Equation D.2 for each sub-group and the total population before mortality returns the groups to their original size. If it is the case that $q_1 \neq q_2$ then it is quite possible that, when looked at as a whole population, $p^*$ can be greater than 0.5, implying that altruism will survive in that population. All that is necessary is that

$$N_2^* > \frac{N_1^*(q_1^* - 0.5)}{0.5 - q_2^*}, \tag{D.4}$$

which results directly from the expressions in Table D.1. Simpson's paradox (or the *averaging fallacy*) is in play here (Simpson, 1951). Thus, in each subgroup, altruism will die out over time, but if the groups reproduce in isolation and then recombine in some meaningful biological sense, altruism can flourish. This example illustrates the impact on the analysis of allowing evolution and adaptation at a *group* level, not just at the species individual level. In the view of Sober and Wilson, it is this group evolution that explains the emergence of an anti-survival trait such as altruism.

**Table D.1: Total vs Subgroup Population Altruism Survival Parameters.**

| | Subgroup 1 | Subgroup 2 | Total Population |
|---|---|---|---|
| Group size | $N_1 = 100$ | $N_2 = 100$ | $N = N_1 + N_2 = 200$ |
| Proportion of altruists | $q_1 = 0.2$ | $q_2 = 0.8$ | $p = \frac{q_1 N_1 + q_2 N_2}{N} = 0.5$ |
| Cost of altruism | $c = 1$ | $c = 1$ | $c = 1$ |
| Benefit of altruism | $b = 5$ | $b = 5$ | $b = 5$ |
| Average number of offspring | $X = 10$ | $X = 10$ | $X = 10$ |
| Population size after reproduction | $N_1^* = N_1[X + q_1(b-c)]$ $= 1080$ | $N_2^* = N_2[X + q_2(b-c)]$ $= 1320$ | $N^* = N_1^* + N_2^*$ $= 2400$ |
| Proportion altruists after reproduction | $q_1^* = \frac{N_1 q_1 F_A}{N_1^* + N_2^*} = 0.184$ | $q_2^* = \frac{N_2 q_2 F_A}{N_1^* + N_2^*} = 0.787$ | $p^* = \frac{N_1^* q_1^* + N_2^* q_2^*}{N_1^* + N_2^*} = 0.516$ |

**D.8:** The existence of group evolution has been a very controversial issue in biology for the past four decades, and the debate continues to this day. Certainly Darwin was aware of and supported the notion of group selection and evolution, and little heated debate arose over the idea until, in the 1960's, a serious challenge was mounted to the concept. Trivers (1971), Axelrod and Hamilton (1981), Maynard Smith (1982), and Dawkins (1982, 1976) all attacked the idea of a group having traits separate from the traits of the individuals in the group, promoting the alternative concept that all group behaviors were no more than emergent properties created out of the interaction between the members of the group. It seems the discussion could become quite acrimonious at times. A quote from Dawkins (1982, p. 115), cited by SW98, p. 50, gives a flavor:

> "As for group selection itself, my prejudice is that it has soaked up more theoretical ingenuity that its biological interest warrants. I am informed by an editor of a leading mathematics journal that he is continually plagued by ingenious papers purporting to have squared the circle. Something about the fact that this has been proved to be impossible is seen as an irresistible challenge by a certain type of intellectual dilettante. Perpetual motion machines have a similar fascination for some amateur inventors. The case for group selection is hardly analogous: it has never been proved impossible, and never could be. Nevertheless, I hope I may be forgiven for wondering whether part of group selection's romantic appeal stems from the authoritative hammering the theory has received ever since Wynne-Edwards (1962) did us the valuable service of bringing it out into the open."

A thread of viciousness can be readily noted in the quote. Sober and Wilson take exception to the characterization, of course, pointing out that "This account is about as accurate as the school version of American history in which our government can do no wrong." (SW98, p. 51)

**D.9:** At the risk of oversimplification, Dawkins based his theory of how evolution worked on a strict enforcement of the replication ability of the gene. The mechanism of adaptation had to be such that it could replicate exactly what came before, and the adaptive properties of the genetic structure could then be elucidated by observing how the exact replication changed through the processes of mutation and crossover. This further implied that all evolution could be imputed from knowledge of how the DNA chemistry worked in the context of the environment in which the organism found itself, which directly supports the Epstein (2006) notion of "if you can grow it, you can show it," his central precept of agent theory. This view also was adopted by Blackmore (1999)[85] and Brodie (1996) in their theory of the meme. But there is no requirement in nature that the replicative processes demonstrated in the DNA genetic structure be the only kind of mechanism that supports evolutionary adaptation. Other such processes can be identified in the structure which explains the interactions of individual species members in a group, for example. This is the position that the group evolution advocates came to demonstrate in their work.

**D.10:** To counter the argument that no group evolution can exist, Sober and Wilson extensively examine the positions held by the group evolution opponents, and one-by-one dismantle their models, at least in the context of altruism evolution. The oldest of these conceptualizations is the kin selection concepts of Hamilton (1963, 1964a, 1964b), alluded to by Mayr above, which suggest that individuals treat close relations more kindly than strangers, thus preventing the demise of altruism due to the frequency of interaction with those close relatives. The haystack theory of Maynard Smith (1964) carries forward this kinship idea by adding to the model the concept that after evolving for several generations in isolation, the kinship groups break up and the individuals intermix and interbreed, creating new kinship groups. Each clan is originated by a

[85] In fact, Dawkins wrote the preface to Blackmore's work.

single pair of individuals, and if it so happens that both of these individuals carry the "altruism gene" the interbreeding during the isolation period of that clan will prevent altruism from disappearing. Since such conditions (several generations of isolation followed by intermixing followed by another period of several generations of isolation) were very unlikely to be found in nature, true altruism was therefore also not to be found in nature, except in extremely rare and specific circumstances. A third approach – dubbed Evolutionary Stable Strategy by Maynard Smith (1982) – applies mathematical game theory as founded by von Neumann and Morgenstern (1944/2004) to evolution. This is the theoretical conceptualization behind Trivers' (1971) notion of reciprocal altruism, and behind the more recent approaches of Maynard Smith (1982) and Axelrod and Hamilton (1981). This construct simply asserts that groups compete as described by game theoretic concepts, and the evolutionary issue is which group devises the winning strategy. One of the primary purposes of invoking this approach was "to explain the evolution of cooperation among nonrelatives without invoking group selection" (SW98, p. 79). But the trait that is evolving is cooperation, not altruism. Cooperation implies benefit to the cooperating individual as well as to the other individuals involved, whereas altruism need not confer any benefit at all to the altruist. Cialdini (2001) uses reciprocal altruism as a key part of his explanation of human traits important to the conduct of marketing.

**D.11:** The selfish gene notion carries considerable weight in current evolutionary biology. Dawkins popularized the notion of selfish gene, originated (according to Sober and Wilson) by Williams (1966). Williams argued that the unit of Darwinian evolution was the gene, since it is able to replicate itself accurately, which is required for selection to occur. Dawkins amplified this idea to the point where it became the "decisive" (SW98, p 87) element in the manifesto of the anti-group selection advocates. Because groups have no replication mechanism structure similar to the gene, there could be no evolution of groups. This notion, by the way, is maintained by Hodgson (1993) and indirectly by Hunt (2000) in their constructs of economics and marketing as evolutionary systems. That it has been challenged significantly argues for a reappraisal of Hodgson's position, a task that would be worthy to undertake, but is unfortunately beyond the scope of this discussion.

**D.12:** The conclusion of Sober and Wilson's exhaustive analysis is that altruism will evolve because it increases the fitness of a group over other groups, even though within the group the altruists will die out because – by definition – altruists are less able to survive and will become extinct over time. They call such group evolution *multilevel selection and adaptation*, suggesting that evolution and adaptation can be seen at various levels of a group structure, and needs to be recognized and observed in that context. This means that the mechanisms supporting multilevel selection in a particular instance depend on the structure of the interactions among the members of a population as well as the reproductive properties of the specific individuals. In turn, this adds considerable complexity to the understanding of trait survival, but also offers conceptual substantiation for the notion that complex, counter-survival behaviors can arise as a result of evolutionary forces. This idea seems to run counter to Epstein's (2006) basic philosophy of "if you can grow it, you can show it." But the contradiction is only apparent. The resolution can be found in the deeper consideration of Epstein's requirement for Explicit Space, which is the topological relationship within communities of agents. Those topologies can be quite complex, and not at all related to the usual concepts of space in one, two or three dimensions.

**D.13:** To address the increased analytic responsibility implied by multilevel selection, Sober and Wilson offer constructs and criteria for defining groups and determining if multilevel selection and adaptation is occurring. First they define a *group* to be

> "… a set of individuals that influence each other's fitness with respect to a certain trait but not the fitness of those outside the group. Mathematically, the groups are represented by a frequency of a certain trait, and fitnesses are a function of this frequency. Any group that satisfies this criterion qualifies as a group in multilevel selection theory, regardless of how long it lasts or the specific manner in which groups compete with other groups." (SW98, pp. 92-93.)

To determine if multilevel selection is actually involved in a particular situation, they go on to offer a three-step procedure to apply in a given circumstance (SW98, pp. 103-116): 1) determine what would be the evolutionary result if only group selection and adaptation was in play; 2) determine what would be the outcome if only individual selection and adaptation was working; and 3) examine the fundamental properties of the inheritance structure implied by 1) and 2). This last step is described by Sober and Wilson in purely biological terms, relating multilevel selection directly back to Mayr's fundamental evolutionary notions of population, variation, and heritability.

**D.14:** Because of the work of Sober and Wilson and the arguably more pertinent writings of Steven Jay Gould, among others, the debate within biology of the existence and efficacy of multilevel group evolution is essentially over, and the group theorists have won. Gould published in 2002 what can only be described as a true *magnum opus*. His 1400-page tome entitled *The Structure of Evolutionary Theory* (Gould, 2002)[86] puts an end to the debate, effectively burying the concept of the selfish gene as presented by Dawkins and his disciples. Any lingering concepts that human groups did not contribute to the evolutionary flow are dismissed. However, concepts of social inheritance based on the selfish gene continue to flourish. Shermer (2004, 2008) implicitly accepts the idea in his important studies of the evolution of morals and ethics, even though it is not needed. Seemingly, the development of scientific ideas are no less subject to the laws of evolution than biology!

**D.15:** Gould's opus provides a detailed map of how this generalized theory of evolution is structured. His hierarchical theory of evolution depends on three structural components: agency, efficacy, and scope. By agency he is referring to the level (of the hierarchy) where evolutionary change is occurring. He identifies six such levels in the biological context: gene, cell lineage, organism, deme, species and clade. Most of the terms are readily understood, but a couple are less often encountered outside of biology. A deme is a local population of a species that interbreeds but is not a subspecies because of (possibly infrequent) cross-breeding with other demes of the same species. In other words, a deme is a group in the Sober and Wilson sense. A clade is another kind of biological group, consisting of a single member of a species and all of its descendents. The first member and many of its descendent can be extinct. Gould

[86] Gould died suddenly shortly after the work was published.

discusses how these various levels interact, and their specific characteristics affect the evolutionary pathways they trace as such interaction occurs (Gould, 2002, pp. 614-637).

**D.16:** Gould's second fundamental structure is efficacy (Gould, 2002, pp. 146-147 and elsewhere). By this he means the mechanism of evolutionary change. At the core, of course, is Darwin's natural selection – survival of the least unfit. But that's not the only mechanism at work. There are sudden large-scale shifts in environment, such as the Tunguska event in 1907 in Siberia, believed to be caused by a meteor, or the global warming phenomenon now established to be occurring on earth due to the human production of green-house gasses and related pollution. There is also exaptation (a lovely word), which means the use of some evolved feature for purposes different then the selection pressure which favored the feature in the first place. He offers a number of examples of this phenomenon.

**D17:** The third element in Gould's thesis is scope, referring to the range of the effect of natural selection (Gould, 2002, pp. 1276-1277). To the eons-long steady progression of life from "simple" to "complex," Gould adds the catastrophic events and sharp breaks so often seen in evolutionary history by the naturalists, what he calls *punctuated equilibrium*. These unpredictable circumstances and happenings are characteristic of history, not science (at least in its strict sense) and thus (Gould, 2002, p 972):

> "In summarizing the impact of recent theories upon human concepts of nature's order, we cannot yet know whether we have witnessed a mighty gain in insight about the natural world (against anthropomorphic hopes and biases that always hold us down), or just another transient blip in the history of correspondence between misperceptions of nature and prevailing social realities of war and uncertainty. Nonetheless, contemporary science has massively substituted notions of indeterminacy, historical contingency, chaos, and punctuation for previous convictions about gradual, progressive, predictable determinism. These transitions have occurred in field after field. Punctuated equilibrium, in this light, is only paleontology's contribution to a Zeitgeist, and Zeitgeists, as (literally) transient ghosts of time, should never be trusted. Thus, in developing punctuated equilibrium, we have either been toadies or panderers to fashion, and therefore destined for history's ashheap, or we had a spark of insight about nature's constitution. Only the punctuational and unpredictable future can tell."

Would that marketing science were so well defined and understood that a conjecture such as this can emerge as a central issue of its development!

**D.18:** Sober and Wilson go on in *Unto Others* to explore both the biological and social structures needed for altruism to have arisen in nature and in human society. These interesting details are beyond the scope of this investigation, but the closing paragraph of the book has an important message for modeling human agents:

> "Our book has been about altruism, but it has also opened the door to a wide range of other subjects. Altruism can be removed from the

> endangered species list in both biology and the social sciences. Groups can qualify as organismic units. Culture can play a vital role in the evolutionary process. And a study of psychological mechanisms can be as evolutionary as the study of behavior. It is heartening to contemplate the emergence of a legitimate pluralism – for evolutionary theories of social behavior, for theories of psychological motivation, and for the larger intellectual traditions that influence how we think about ourselves and the world around us." (SW98, p. 337).

Sober and Wilson thus demonstrate that complex, even counter-survival, traits can arise from the evolutionary porridge which is the background of all species, including humans. This demonstration, in my view, lends considerable weight to the hypothesis that the universal human behaviors summarized by Brown are based on our evolution, because they are indeed universal and because feasible mechanisms for their adaptive emergence do exist. With the additional emerging evidence cited by Goleman (2006) and Montague (2006), it is safe, in my view, to assert that human evolution is thus the source of the market behavior that must be captured in an agent-based model. Thus I am confident that there exists a class of human behaviors which will apply in any agent-based modeling context, and that identification and characterization of members of that class is a worthwhile, scientific effort.

# Appendix E: Shermer's Discussion on the Evolution of Ethics

**E.1:** Shermer (2004) approaches the issue of the emergence of morals and a moral sense in humans from the perspective of what he refers to as *methodological naturalism* (Shermer, 2004, p. 19). That is, it is assumed that all effects have natural causes, and therefore are within the realm of scientific inquiry. Thus he dismisses concepts and theories on the existence of morals, morality and ethics that rely on the concept of a God or other divine intervention. He does not debate whether or not God exists.[87] Rather, he simply views the issue as irrelevant. From this perspective, he maintains that human morality and the cultural norms that support a moral code in human society are the result of natural forces and unique paths of historical development. This viewpoint reinforces the universality of the human traits that support marketing and market behavior, and thus the search for agent-based conceptualizations of those traits is a far more tractable effort with much higher likelihood of success.

**E.2:** Shermer invokes Sober and Wilson, Gould and others to actually take on the issue of how a particular set of behavioral traits – morals and ethics – came to be through the evolutionary pathways. While not completely abandoning the selfish gene position of Dawkins, he does admit that multilevel selection and adaptation is at work in

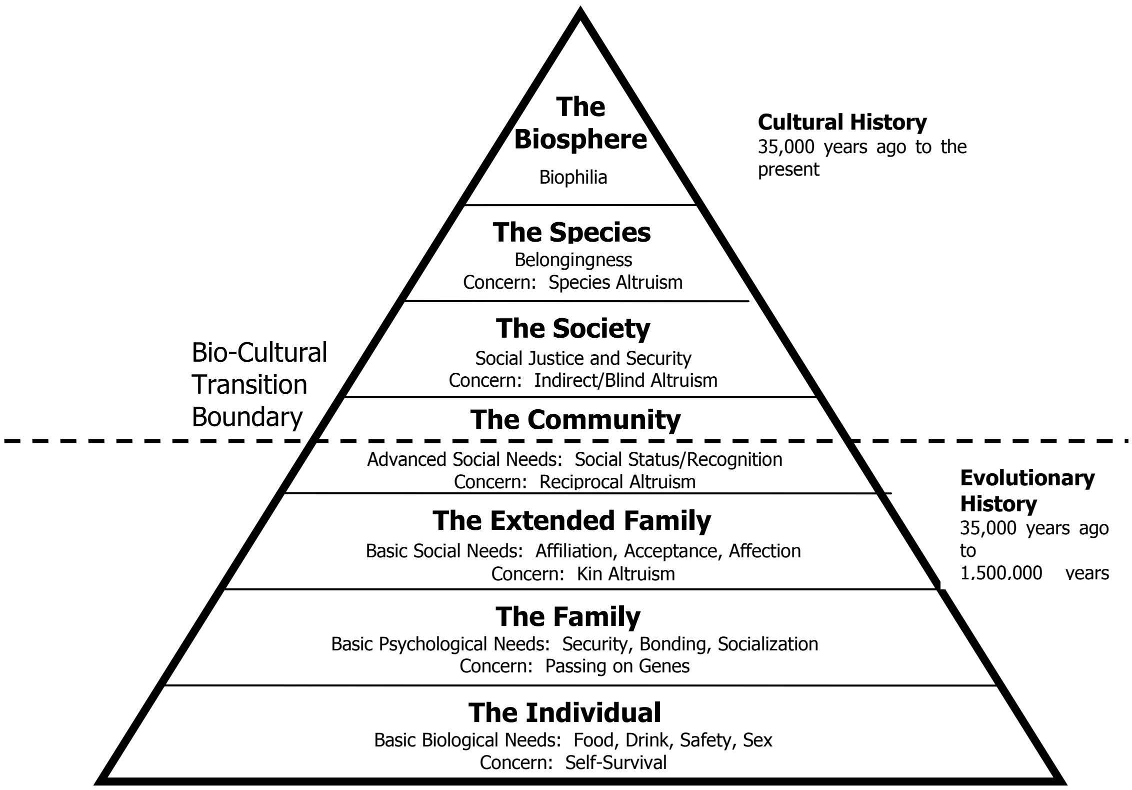



**Figure E.1: Shermer's Bio-Cultural Pyramid**

[87] Shermer refers to himself as a *nonthiest*, meaning that to him belief in a god is beside the point.

evolutionary effort, although he puts it in a secondary role (Shermer, 2004, pp. 50-56). That this is incorrect given the evidence and the logic can be taken up at another time, for it is immaterial to the argument at hand. Addressing the question of the time frame of behavioral evolution (the scope, as Gould would put it), he offers the Bio-Cultural Evolutionary Pyramid (BCEP), as illustrated in Figure 3.1. This illustration concisely describes how Shermer believes ethics evolved over the course of the time of modern *homo sapiens* on the planet. Going from bottom to top of the pyramid represents the increasing complexity of humanity's ethical perspective, from early man whose concerns were basic biological needs such as shelter, food and reproduction, to modern society's concerns over the condition and future of the environment as a whole (the biosphere).

**E.3:** The pyramid also is designed to represent other aspects of our social-biological evolution. Its height represents the extent to which the feelings of altruism and related ethical concepts such as fairness and respect extend beyond the individual, to kin, members of the tribe, the larger community, and eventually the entire biosphere. He uses the kinship motif of group evolution here, while other, more subtle, effects may be at work. The width of the pyramid at any vertical point represents the strength of the ethical sentiment and the extent to which it is under the control of biological evolution, as opposed to social (group) evolution. Of particular importance is the dashed line midway up the pyramid. This is referred to as the Bio-Cultural Boundary. Occurring approximately 35,000 years ago[88], this is the point where the majority of social evolution came under the control of culture and social influences, passing from the more pure survival world of biology that preceded it for a million and a half years. In other words, this is the point where the group traits and communications skills of the human species began to have more influence on our evolution than our biological heritage and inherent genetic adaptation. It is also here where second-order adaptation became feasible, and about the time when narratives began to emerge.

**E.4:** Shermer's work is an example of the application of modern concepts of biological and social evolution to human social and behavioral traits, thus establishing that not only can such traits emerge as consequences of that evolution, but also how it might have worked. It does not matter whether Shermer's model is accurate in every detail; it is feasible and fits the data at hand, so it has empirical credence and is therefore an acceptable, if provisional, theory. It would be remiss if the conclusions of his derivation were not highlighted in this discussion. While they are not directly concerned with the issue of modeling human market agents, at least not given the current level of development of that agency, it does shed light and give background to some of the points that follow. Of course, there will undoubtedly be other interpretations and theories put forth as new data becomes available or implications are worked out, so its provisionality is a key perspective to keep in mind.

**E.5:** In fact, Shermer uses "provisional" as an adjective in all three of the ethical areas he discusses -- provisional morality, provisional right and wrong, and provisional justice. He views all aspects of morality as provisional in the sense that they do not apply to every individual in every culture for every situation for all time (thus meeting

[88] It is unclear how this line came to be drawn at 35,000 years. While largely immaterial to the course of the argument, I suspect that some archeologists and anthropologists, not to mention a few biologists, might quibble with the date.

Mayr's criterion of population thinking). But they apply to most people from most cultures in most situations for most of the time. This is because he maintain that all "humans are, by nature, moral and immoral, good and evil, altruistic and selfish, cooperative and competitive, peaceful and bellicose, virtuous and nonvirtuous" (Shermer, 2004, p. 20). He defines right and wrong with three principles: 1) happiness, in that it is a higher moral standard to seek happiness that does not cause others unhappiness; 2) liberty, in that it is a higher moral position to seek liberty only when it does not reduce the liberty of others; and 3) ask-first, being the concept that if the morality of an act is in question, those affected need to be consulted in making the determination. What he refers to by provisional justice is that all humans are responsible for their behavior and the consequences thereof, and social norms and practices seek to mete out rewards and punishments in the here and now consistent with this idea, as opposed to some far-off, even post-death adjudication. A comparison of this moral code with that enunciated by Hunt, for example, would seem a worthy contribution to the foundations of marketing science.

**E.6:** In a later work, *The Mind of the Market*, Shermer (2008) applies the same bio-social evolution analysis to the emergence of economies, and hence of markets. He makes the case that economies are complex adaptive systems, complete with emergent properties and path dependence. This is not new, of course, the computational and evolutionary economists having made this case some time ago. See, for example, works by Daguchi (2004) who presents several formal results in treating economies as a complex adaptive system, Lude and Tesfatsion (2006) who collect together a number of important discussions of computational economics (which assumes economics is a complex adaptive system), and Luna (2000) and Kohler and Gummerman (2000) who discuss both economic and social analysis using complex system concepts and the "swarm" technology developed by the Santa Fe Institute. In fact, as these references show, a considerable amount of agent-based modeling is applied to the testing and substantiating hypotheses related to this perspective. This reinforces the view that the parameters of modeling human market behavior, and thus human marketing agents, are to be calibrated against evolutionary structures that are common across all of the human species.

**E.7:** To summarize, if Sober and Wilson (1998), and indirectly later Gould, make a persuasive case that human behavior *could* arise out of biological evolution, with the help of the social enhancements and group evolution that implies, then Shermer gives a good example of the kind of analysis that demonstrations that it has happened. And he applied the analysis to a subject inherently important to marketing – ethics and morals. Thus it has been demonstrated that, in fact, human behavior very likely originated in our biology. Therefore the human universals itemized by Brown are, indeed, universal, and can be used safely used as a basis for agent-based modeling efforts. But what of the specific behavior of humans that is seen in a strictly marketing context? To that the discussion now turns.

# Appendix F: The Narrative Conceptualization of Walter Fisher

**F.1:** The term *narrative* in linguistics refers to one of the four forms of classical rhetorical discourse (the other three being exposition, argumentation, and description). Walter Fisher, a sociologist with major influence in modern communications theory, in his treatise titled *Human Communications as Narrative: Toward a Philosophy of Reason, Value and Action* (Fisher 1987, hereafter WF87), makes the case that narratives can capture the essential features of not only rhetorical discourse, but of *all human communications whatsoever*. His approach offers several interesting concepts which bear on the role of the narrative as a construct in which marketing behavior can be considered, and so a deeper exploration of his reasoning is valuable.

**F.2:** Fisher begins his discussion with the ancient Greeks, and from there traces the threads of the development of logic and rhetoric as two aspects of human rational thought. Plato and Aristotle began the tradition by making a distinction between logic, that aspect of human thinking which relies on axiom and deduction as the basis for decision, and rhetoric, which is the form of discourse that seeks to persuade, convince, argue or inform individuals in a group or social context. Through the works and legacies of Bacon, Descartes, Galileo, Kepler, Newton, Locke, and many others, "technical logic", as Fisher puts it, became "the handmaiden of learned discourse" (WF87, pp 30-37). He asserts that the success of this form of communications was responsible for the rise to prominence of logical positivism, formal mathematics, science, and many empirical endeavors such as engineering.[89] This perspective reached the point, in the mid-20th century, where any thinking that was not based on formal logic and empirical verifiability, the essence of Popper's concept of falsifiability (Popper, 1959/2002) was at best entertaining, and at worst an imminent danger to man's tenure on earth. And the human mind became disassociated from the constructions that were borne of it. Citing Bertrand Russell,

> "Throughout logic and mathematics, the existence of the human mind or any mind is totally irrelevant; mental processes are studied by means of logic, but the subject-matter of logic does not presuppose mental processes and would be equally true if there were no mental processes. It is true that in that case we should not know logic; but our knowledge must not be confounded with the truths that we know."[90] (WF87, p 35.)

In Fisher's view, the result of this thread of thought is that

> "Logic is now the province of formalized systems. This means that logic now stands apart from issues such as those addressed by the narrative

[89] An interesting but little explored motive behind this empiricism is described in Russell Shorto's work *Descartes' Bones* (Shorto, 2008). He argues that the quest for human immortality – the ultimate prediction of the future, if you will – has driven the development of abstract logic and empiricism as much as anything. To preview slightly what is to come, such a concept is an important outcome to many, many human narratives.

[90] Curiously, this point is made in a different way by Humphreys (2004), in that a human mind is not needed for there to be epistemology.

> paradigm: How do people come to believe and act on the basis of communicative experiences? What is the nature of reason and rationality in these experiences? What is the role of values in human decision making and action? How can reason and values be assessed?" (WF87, p 35.)

**F.3:** But there are limits to "technical logic." Gödel (1938) established the limits of the application of formal logic. More fundamentally, Heisenberg (1927/1983) established that the universe is unpredictable at a very basic level, violating the hoped-for certainty of formal logic. These discoveries create insurmountable difficulties for the extension of pure logic into all areas of human endeavor. Not everything could be explained by formal axiom and proof (including the validity of the logic itself). Predictions cannot be made for all things, in fact not even for some very fundamental things, like the simultaneous location and velocity of a subatomic particle. Nevertheless, there appears to be a very strong normative character to the writings of the philosophers, logicians and mathematicians, suggesting that reasoning without the benefit of formal logic is somehow deficient, and society should take firm steps to reduce the reliance on other modes of thought in favor of logical reasoning. For very many, logic is synonymous with rationality. Fisher counters by observing that a very substantial body of human activity, including jurisprudence, politics, art, literature, poetry, and theater have rationality, but not necessarily technical logic. And attempts to insinuate such formality into these fields has been, he asserts, singularly unsuccessful.[91]

**F.4:** If "technical logic is the handmaiden of learned discourse," then Fisher asserts that "rhetorical logic is the handmaiden of public discourse." (WF87, p 36). In the last five centuries, as technical logic has come to dominate philosophical thinking, rhetorical logic has been pushed further into the background. A quote from Giambattista Vico,[92] cited by Fisher, illustrates the distinction nicely:

> "… whoever endeavors to devote his efforts, not to physics or mechanics, but to a political career, whether as a civil servant or as a member of the legal profession or of the judiciary, a political speaker or a pulpit orator, should not waste too much of his time, in his adolescence, on those subjects, which are taught by abstract geometry. Let him, instead, cultivate his mind with an ingenious method; let him study topics, and defend both sides of a controversy, be it on nature, man or politics, in a freer and brighter style of expression. Let him not spurn reasons that wear a semblance of probability and verisimilitude." (Vico, as translated by Gianturco, 1965).

---

[91] During my first tenure at University, in the 1960's as an undergraduate student of mathematics and sculpture at the University of Washington, I was privileged to give a talk to students of from both disciplines on the relationship between mathematics and art. Much of the ensuing discussion was if and how the practice of art could be improved with the application of mathematical reasoning. Of course nothing was resolved, but this, in turn, led to many years of friendship with individuals in both communities.

[92] Giambattista Vico (1668 - 1744) was an Italian philosopher, historian, jurist, and clearly a sharp critic of the then-emerging Cartesian rationalism.

The sentiment was expressed over 300 years ago, and the degree to which it is still true measures the rate of shift from the rhetorical to the technical.

**F.5:** It is this rhetorical logic and its integral role in human understanding in such fields as literature and jurisprudence on which Fisher basis his thesis. From the concepts surrounding rhetorical logic he extracts what he calls the *narrative paradigm*. Central to this idea is the concept of *good reason*. (WF87, p 57). Good reason means that the rationality of a particular situation (or argument, presentation, story, discussion, opinion, etc.) depends on the underlying narrative that is supporting an individual's understanding of that situation. This, in turn, must imply that many, if not all, human mental processes are devoted to the creation and maintenance of internal brain patterns that are of a narrative structure. Recalling that a narrative is a recounting of a sequence of events, Fisher's context refines the definition to mean a mental pattern of time-based (therefore at least sequential in time) changes in perceived states, which he defines as *events*. That human beings build and maintain such patterns is widely supported by substantial psychological, physiological, biological and physical evidence. In other words, the logic inherent in a narrative creates the realization of good reason. *An argument is thus considered to be rational, not because of rigorous logic, but because it makes sense as part of a narrative.*

**F.6:** The time dependency of the narrative sequence structure is vital. While sequences in space can be recognized without reliance on memory or other neurological pattern retention mechanisms, that cannot be true for events that unfold in time.[93] The reality of ten minutes ago no longer exists, except in the memory of one who observed and remembered them. Furthermore, in order for the temporal pattern to carry meaning, the events in the sequence must be perceived to create a *cause-and-effect* chain. Thus a narrative must be made up of patterns like "if A occurred, then B will occur," where the meaning of the word 'then' is at the minimum, the passage of time. It is this time-sequenced cause and effect chain that Fisher is talking about when he refers to "good reason." A communication, such as a legal argument, contains good reasons if it asserts causal relationships between earlier and later events in the narrative sequence. This is the basic requirement to understanding. To understand something is to fit it into a mental pattern of a cause-and-effect chain.

**F.7:** When an internal narrative is the object of communication with another – that is, being shared – two other aspects of Fisher's narrative paradigm emerge. One is the requirement of coherence, which Fisher terms *narrative probability*.[94] Coherence is simply whether or not the receiver of the communications understands what is being communicated. Are the symbols used known and understood, and does the structure being portrayed correspond to the time sequence pattern of a narrative that the receiver owns? The second feature is *narrative fidelity*. Does the receiver agree that, given her own internal array of narratives, the appropriate cause and effect chains are or could be

[93] It takes some small amount of time to observe any particular scene with the eyes, for example, but the information recorded in each instant is substantial. The retention mechanism here might well be the retinal cells themselves, not considered part of the brain. The principle is the same, but the physical organ that does the work may not be specifically cerebral.

[94] In reference to the concept of dramatic probability in literature, meaning does the story make sense given the experience and frames of reference of the anticipated audience.

*true*, at least within the context established the narrative? Do they correspond with her prior experience both in her life and as shared by others?

**F.8:** Fisher maintains (WF87 pp. 60-61 and elsewhere) that the narrative is the vehicle by which *value* is developed and maintained. Something has value if its application increases the likelihood that a desired future course of the narrative is realized. As in many aspects of human psychology, value is established by vehicles and behaviors that have served well as the species as it has evolved. Something is valuable if its possession enhances the ability to survive individually or as a group. Fisher makes the point that not only does the narrative paradigm address the modes and methods of how humans understand and communicate about their world, but also the paradigm captures an important concept of value. Value in economic terms is generally defined in terms of preference. Something has value only when compared to something else. That is, “A is valued over B” is a meaningful statement but “A has value” is not. But value in the context of a narrative does have meaning outside the comparison against other things. One example is suggested by the theory of psychological reactance that accounts for some of the scarcity principle (see Cialdini, 2001) – that as an item previously readily available becomes less so, its value rises because of the implied loss of ‘freedom’ inherent in it becoming scarce. This inherent freedom is viewed as such because it is necessary for the realization of a narrative, and loss of it inhibits that realization. Therefore the scarce commodity/service has value because it serves to help realize a desired narrative, and its value increases when it becomes less available because it becomes more unlikely to be available to support the realization of the narrative.

# Appendix G:  The Let's-Make-A-Deal Bias

**G.1:** The Let's-Make-A-Deal bias is an example of the subtle errors that can occur when contemplating probability and stochastic events. That the probability of winning is 0.666666 … if one switches doors can be empirically demonstrated with the use of the simple computer simulation offered here. Figure G.1 shows the interface screen for a windows application that simulates the game. The user sets the number of game iterations, and also selects which door to initiate the game. During each play of the game, the program randomly assigns the car to one of the three doors, and selects one of the remaining doors to open for the play of that iteration. Then the game player is assumed to choose the remaining door. The simulation then counts the number of times the player wins the game, reporting the result as shown on the interface.

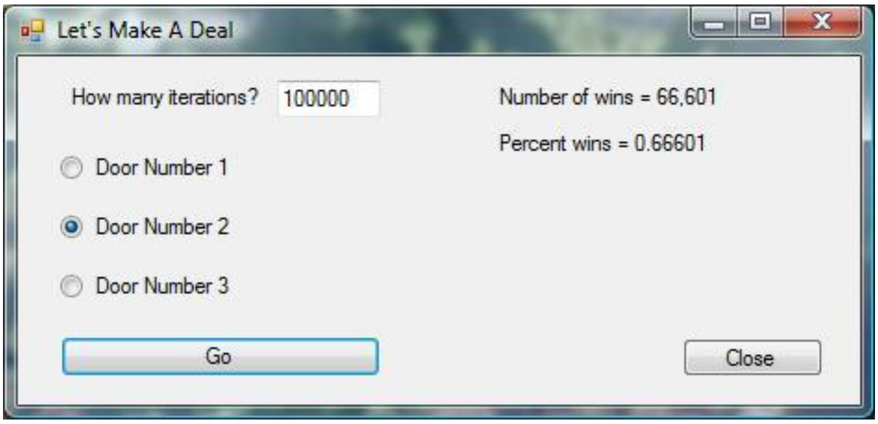



**Figure G.1:  Let's Make A Deal Game Simulation**

**G.2:** The code for this program is presented in below. This simulation is written in C# 2008 using the Microsoft Visual Studio 2008 development environment. The application can be duplicated in that environment by creating a new Windows Application project entitled LetsMakeADeal, and populating the main form (named LMAD) with:

- A text control named textBoxIterations
- Three radio button controls for the three doors named radioButtonDoor1, radioButtonDoor2 and radioButtonDoor3, respectively
- A button control named buttonGo
- Two label controls named labelWins and labelPctWins
- A button control named buttonClose.

The code can then be copied and used to replace the class LMAD definition in the form code

.

**G.3:** The following is the C# code for the LMAD application.

```
/*
 * Application: Lets Make A Deal
 * 20 September 2009
 * R. A. Parker
 *
 * Simulates the TV game of Let's Make A Deal
 *      The contestant is presented with three closed doors.  Behind
 *      one is a car, behind the other two a useless prize
 *      (historically, goats).After the contestant selects a door, the
 *      moderator of the game shows the contestant what's behind
 *      one of the two doors not originally chosen. The contestant
 *      then ALWAYS switches to the third door -- the one he did not
 *      originally choose and the one the moderator did not open. The
 *      probability of winning is consistently above 60%.
 *
 *      The user can decide how many times to simulate the game
 *      to estimate the probability of winning.
 *
 */

using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;

namespace LetsMakeADeal
{
    public partial class LMAD : Form
    {
        int DoorNumber = 1;

        public LMAD()
        // initialize the application window class
        {
            InitializeComponent();
        }

        // see what door the contestant originally chooses
        private void radioButtonDoor1_CheckedChanged(object sender,
            EventArgs e)
        {
            if (radioButtonDoor1.Checked) DoorNumber = 1;
        }

        private void radioButtonDoor2_CheckedChanged(object sender,
            EventArgs e)
        {
            if (radioButtonDoor2.Checked) DoorNumber = 2;
        }

        private void radioButtonDoor3_CheckedChanged(object sender,
            EventArgs e)
        {
```

```
        if (radioButtonDoor3.Checked) DoorNumber = 3;
    }

    private void buttonGo_Click(object sender, EventArgs e)
    {
        int cardoor = 0;            // the door number where the car is
        int showdoor = 0;           // the door number shown to
                                    // contestant
        int chosendoor = 0;         // the door the contestant chooses
                                    // after seeing the opened door
        int winner = 0;             // the number of times the
                                    // contestant wins

        Random rnd = new Random();  // a simple random number generator


        // get the number of iterations.
        int iter = int.Parse(textBoxIterations.Text);

        for (int i = 0; i < iter; i++)
        // repeat the game iter number of times
        {
            // put the car behind a door
            double x = rnd.NextDouble();
            double y = rnd.NextDouble();
            if (x < 0.33333333333333) cardoor = 1;
            else if (x <= 0.6666666666666) cardoor = 2;
            else cardoor = 3;

            // determine the door opened for the contestant
            switch (DoorNumber)
            {
                case 1:
                    // contestent chose door 1
                    if (cardoor == 1)
                    // if the car is behind door 1
                    {
                        // randomly choose to show door 2 or 3
                        if (y < 0.5) showdoor = 2;
                        else showdoor = 3;
                        break;
                    }
                    else if (cardoor == 2)
                    // but if the car is behind door 2
                    {
                        // show the third door
                        showdoor = 3;
                        break;
                    }
                    // if the car is not behind door 1 or door 2,
                    // then its behind door 3, and need to
                    // show door 2
                    else showdoor = 2;
                    break;
                case 2:
                    // contestant chose door 2
                    if (cardoor == 1)
                    // if the car is behind door 1
                    {
                        // show the third door
```

```
                showdoor = 3;
                break;
            }
            // but if the car is behind door 2
            else if (cardoor == 2)
            {
                // randomly choose to show door 1 or 3
                if (y < 0.5) showdoor = 1;
                else showdoor = 3;
                break;
            }
            // if the car is not behind door 1 or door 2,
            // then its behind door 3, and need to show door 1
            else showdoor = 1;
            break;
        case 3:
            // contestant chose door 3
            if (cardoor == 1)
            // if the car is behind door 1
            {
                // show door 2
                showdoor = 2;
                break;
            }
            else if (cardoor == 2)
            // but if the car is behind door 2
            {
                // show door 1
                showdoor = 1;
                break;
            }
            // if the car is not behind door 1 or door 2,
            // then its behind door 3, and need to randomly
            // show door 1 or 2
            else
            {
                // randomly choose to show door 1 or 2
                if (y < 0.5) showdoor = 1;
                else showdoor = 2;
                break;
            }
    }

    // contestant chooses the door not opened from the
    // two he did not initially select
    if (DoorNumber == 1)
    {
        if (showdoor == 2) chosendoor = 3;
        if (showdoor == 3) chosendoor = 2;
    }
        if (DoorNumber == 2)
    {
        if (showdoor == 1) chosendoor = 3;
        if (showdoor == 3) chosendoor = 1;
    }
    if (DoorNumber == 3)
    {
        if (showdoor == 1) chosendoor = 2;
        if (showdoor == 2) chosendoor = 1;
    }
```

```
                // does the contestant win?
                if (chosendoor == cardoor) winner++;
            }

            // display the results
            labelWins.Text = "Number of wins = " + winner.ToString("n0");
            labelPctWins.Text = "Percent wins = " +
                  ((double)winner/(double)iter).ToString("n5");

        }

        private void buttonClose_Click(object sender, EventArgs e)
        // close the program
        {
            Application.Exit();
        }
    }
}
```

**G.4:** A typical result is shown in Figure G.1. Here the user specified 100,000 iterations of the game, selected door number two to start the simulation, and produced 66,601 wins with the switching strategy.

# Appendix H: Available Agent-Based Modeling Packages

**H.1:** The programming required to build a particular virtual market depends heavily, if not completely, on the computing skills of the team constructing the simulation. In my personal case, I have been programming as part of my work for over four decades, so I have little need to use application systems that are designed for general agent-based modeling use. However, there are a number of available software packages specifically designed for the construction of agent simulations. Table H.1 lists some of the more important ones. Most are readily available for no or nominal cost. Each is characterized in terms of its platform and language compatibility, and the contact information available at the time of this writing listed. I make no claim to the completeness of this inventory, however, given that the field is evolving rapidly and that I have no specific use for these applications and therefore do not keep up with current developments. The interested reader is invited to consult Nikolai and Madey (2008), Railsback and Lytinen (2006), or Tobias and Hoffman (2004), which are the sources of the information in Table H.1.

**Table H.1: Available Software Packages for Agent-Based Modeling**

| Package | Purpose | Licensing | Platform | Contact |
|---|---|---|---|---|
| ABLE | Building agents using machine learning and reasoning | Open source. Free for non-commercial use | OS/2, Windows, Java, Unix | www.alphaworks.ibm .com/tech/able |
| ADK | Large scale distributed applications | LGPL* | Windows, Unix, Linux, Java | www.tryllian.com |
| AgentBuilder | General purpose multi-agent systems | Proprietary | Windows, Linux, Java | www.agentbuilder.com |
| AnyLogic | General purpose multi-agent systems | Proprietary | Windows, Java, Linux, Mac OS | www.coensys. com/anylogic.htm |
| AOR Simulation | Agent-based modeling of beliefs, speech | LGPL | Windows, Mac, Unix, Linux | www.AOR-Simulation.org |
| Ascape | General purpose agent-based systems | Proprietary | Windows, Mac, Unix, Linux | ascape.sourceforge.net |
| Brahms | Organizational agent-based simulations | Open source, Free for non-commercial use | Windows, Linux, Sparc | www.agentsolutions.com |
| Cormas | Ecological IBM | Free to modify, but not distribute | Linux, Unix, Mac, Windows | cormas.cirad.fr |

| Package | Purpose | Licensing | Platform | Contact |
|---|---|---|---|---|
| Cougaar | Highly distributed multi-agent systems | Open source | Windows, Linux, Mac OS, Java | cougar.org |
| ECHO | Ecology IBM | Open source, free | Unix, Sparc | www.santafe.edu/~pth/echo/ |
| ECJ | Environmental systems analysis and evolutionary modeling | Free academic license | Java | www.cs.gmu.edu/~eclab/projects/ecj/ |
| JAMEL | Agent-based macroeconomic models | Free | Java | p.seppecher.free.fr/jamel |
| JAS | General purpose agent-based models | LGPL | Java | jaslibrary.sourceforge.net |
| JASA | Computational Economics | LGPL | Java | sourceforge.net/projects/jasa |
| JCA-Sim | Cellular automata | Free | Java | www.jweimar.de/jcasim |
| jEcho | Ecological modeling | Free, Open source | Java | www.brianmcindoe.com |
| MAML | Social science agent-based models | Open source | Windows, Linux | www.maml.hu |
| MASON | Social complexity, AI/Machine learning | Academic free, open source | Java | cs.gmu.edu/~eclab/projects/mason |
| MASS (Multi-Agent Simulation Suite) | General purpose simulations. Participatory simulations | Proprietary | Java | mass.aitia.ai |
| NetLogo | Social and natural sciences. Beginning users | Free | Java | ccl.northwestern.edu/netlogo |
| MIMOSA (Micro- and Multilevel Modeling Software) | Social science. Education | Free | Java, Solaris, Linux | www.uni-koblenz.de/~moeh/projekte/mimose.html |
| Repast (Recursive Porous Agent Simulation Toolkit | Social sciences, computational economics | BSD license | Java, Mac OS, Windows | repast.sourceforge.net |
| Jade's sim++ | Communications and real time network planning distributed systems | GPL** | Sun, Meiko, BBN, HP 9000 | Jade Simulations International 1422 Kensington Road NW. Calgary, Alberta T2N 3P9, Canada |

| Package | Purpose | Licensing | Platform | Contact |
|---|---|---|---|---|
| SimAgent | Research and teaching for highly complex agent systems | Free | Prolog 15, Windows, Mac OS, Linux, Unix | www.cs.bham.ac.uk/research/projects/poplog/packages/ simagent/html |
| | | | | |
| Soar | Human performance modeling | BSD license | Windows, Linux, Mac OS | www.soartech.com |
| StarLogo | Social and natural science modeling. Suitable for K-12 educational use | Free | Mac OS, Windows, Unix, Linux | education.mit.edu/starlogo |
| Sugarscape | Social sciences | GPL | Java | sugerscape.sourceforge.net |
| Swarm | General purpose agent-based models | GPL | Java, Windows, Linux, Mac OS | www.swarm.org |
| VisualBots | Agent simulator in Excel | Free | Windows | www.visualbots.com |
| ZEUS | Rules engine based agent models | Open source | Windows | labs.bt.com/projects/agents/ zeus/ |

*Lesser  Gnu Public License
**Gnu Public License

# Appendix I:  A Simple Word-of-Mouth Virtual Market

**I.1:** To bring some tangible context to the ideas discussed in this Chapter, I will now present a simple word-of-mouth virtual market. This is a rather natural application for agent-based modeling, as it has many characteristics shared by some of the early social science agent models. In order to simulate the spread of a concept, product or service by word-of-mouth, it is necessary define a population of consumer agents on some a spatial landscape. In addition, the consumer agents need some way of communicating with or influencing other agents. That is, there have to be objects that are spatial related to each other so that verbal communications is reasonably represented. A very simple structure for this can is implemented using a grid where the state of each space or cell on the grid is a function of two conditions, the current state of the cell, and the state of one or more of the immediate neighbors of the cell. The agents are represented by the grid's cells. For this word-of-mouth simulation, I assume that each cell starts off in a state of ignorance. Once one agent obtains information (changing state from "unaware" to "aware"), the neighboring cells become potential recipients of the information, and can then can become aware. Once an agent is aware, it can purchase the product, decide it doesn't want the product, or wait and make a decision later. If it waits too long before choosing to buy or not, it 'forgets' about the product and returns to an unaware state.

**I.2:** This is as an example a simple *cellular automata* simulation, somewhat akin to the famous game of life invented by Conway (1970). A large two-dimensional lattice is constructed on which little squares called *cells* reside. Each cell sits in between the vertices in the lattice completely filling the space between vertex points. Each cell can be in one of a known number of internal conditions, called naturally enough *states*. A cell can change from one state to another depending on something internal to it and unrelated to any other cell (such as the passage of time) or based on some interaction with other cells, or a combination of the two. (To keep things simple in this case, the

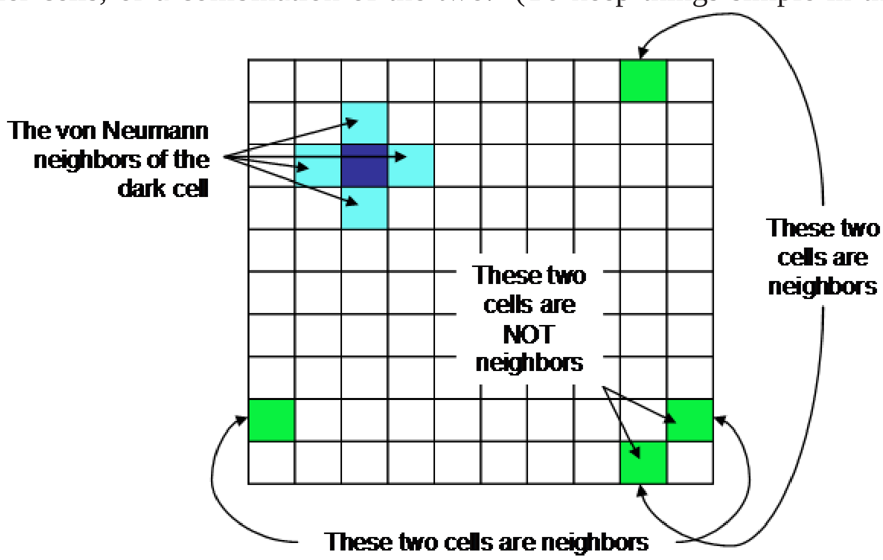



**Figure I.1: The von Neumann Neighborhood**

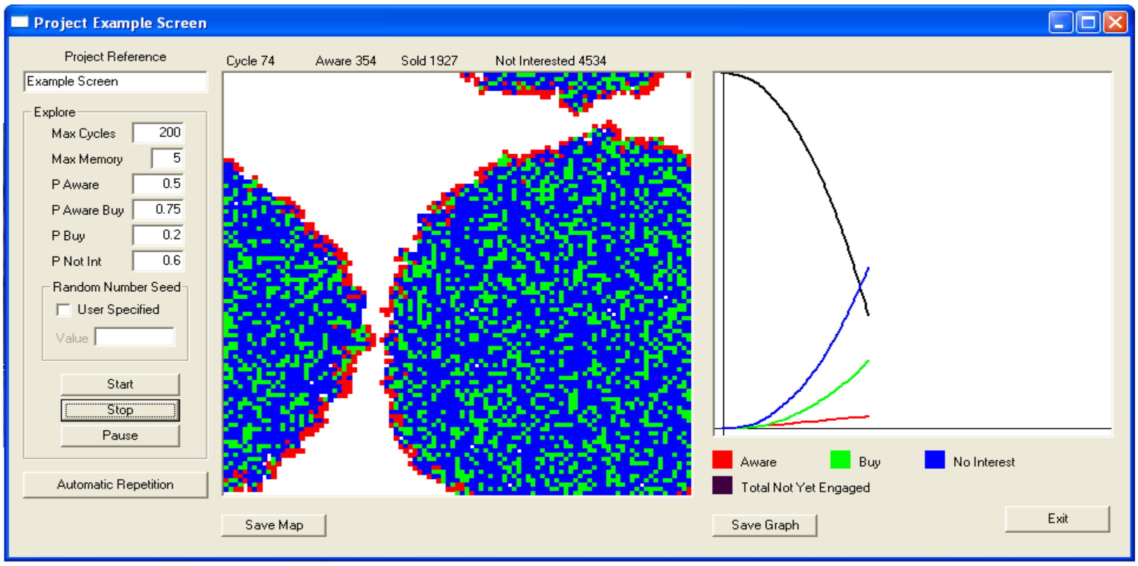



**Figure I.2: Main Control Screen for WOM Simulation**

cells don't move on the lattice.) Cellular automata are the simulation formats used by Epstein and Axtell in their now-classic work *Growing Artificial Societies* (1996). A cellular approach was taken by Shelling, who developed one of first agent models, the manually-executed neighborhood segregation simulation in 1973 (Shelling, 1973), and significant depth added by Wolfram (1994).

**I.3:** Suppose the cells represent individual consumers who have states associated with the purchase of a product – unaware of the product, aware but uninterested in it, aware but undecided, aware and have purchased. A cell will change from unaware to aware only if it is next to a cell that is aware, and then only with probability $p_{a|a}$, or next to one which has bought the good, with probability $p_{a|b}$ (either higher than $p_{a|a}$, due to a favorable recommendation, or lower, due to a disappointing experience). Further, once aware, the cell will choose to purchase with some probability $p_b$, or decide it is not interested with some probability $p_n$, (where $p_n + p_b \leq 1$) during each time period up until it either buys, loses interest, or 'forgets.' Between the time it is aware and chooses to buy or not, or forgets, it is undecided. All the probabilities are constant for all cells. How long it takes for a cell to forget varies from cell to cell, and is randomly set from between one and five time cycles. Once it has become disinterested, it will not return to the aware state again. Once in a purchase state, it will stay there. If it forgets, it can become aware again if the conditions permit.

**I.4:** I will employ a common cellular automata convention in defining what "neighbor" or "next to" means. One cell is considered the neighbor of another if it is right next to it in any of four directions – up, down, left or right. The cells diagonal to it are not considered neighbors. These four cells create the so-called *von Neumann* neighborhood. Further, the lattice on which the cells reside is considered a torus, in that a cell in the top row is a neighbor of a cell in the same column position on the bottom row, and vice versa, and a cell on the right edge is a neighbor of a cell in the same row on the left edge, and vice versa. This allows awareness to flow across the entire surface of the cell space without worry about behavior at edges. If the space is visualized in three dimensions, it is in the shape of a donut, or torus. Figure I.1 shows a simple 10 x 10 grid illustrating these conventions. The torus topology also means that it does not matter which cell is the first one in the awareness-purchase cycle. The basic grid is completely uniform.

**I.5:** This is how the simulation operates. If an agent already possesses the information and its probability of forgetting is below some threshold, the agent retains the

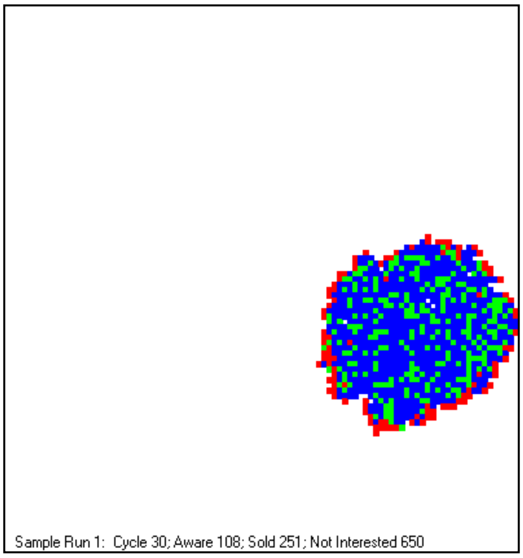



**Figure I.3a: Time Cycle 30**

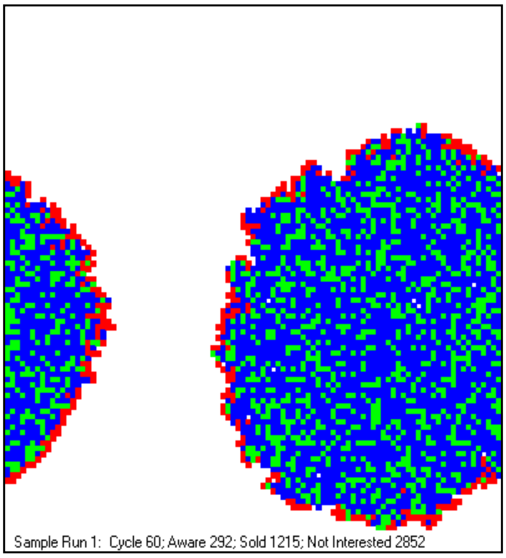



**Figure I.3b: Time Cycle 60**

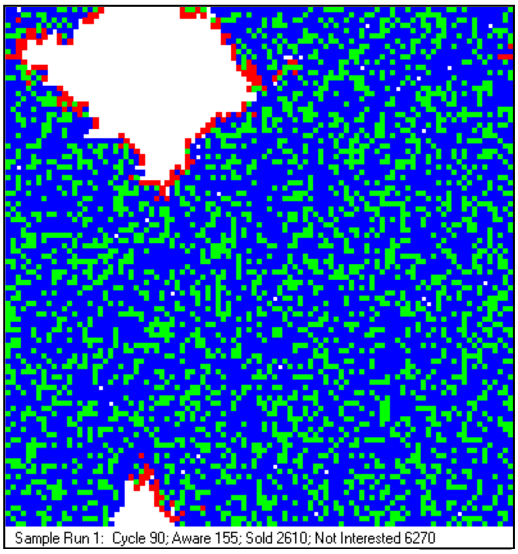



**Figure I.3c: Time Cycle 90**

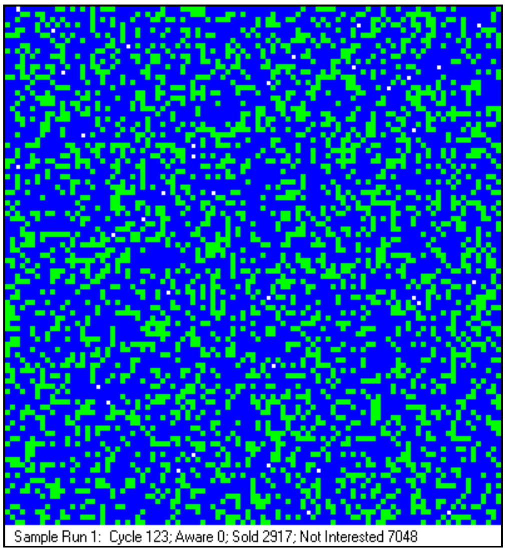



**Figure I.3d: Time Cycle 123 – Stable**

☐ Not aware   Aware   Purchased   Not

**Figure I.3: Sample WOM Simulation**

information through the next simulation time cycle. If an agent does not have the information, it communicates with each of its immediate von Neumann neighbors, and if one has the information and a high enough probability of passing it on, the agent becomes knowledgeable. Time, as would be expected, will be broken up into discrete cycles. During each cycle, the cellular agents will communicate with its neighbors, and change state according to the agent design and the various probabilities. Of interest is the rate of spread of awareness and purchase as a function of $p_{a|a}$, $p_{a|b}$, $p_b$ and $p_n$ over time, given that only one individual cell starts out aware. This is a very elementary model of word-of-mouth (WOM) marketing. While simple, it illustrates some of the key concepts in a virtual market.

**I.6:** The simulation is implemented with simple computer program. The program is written in Microsoft Visual Basic V5.0 for execution on Intel PCs with the Microsoft Windows operating system. It is thus a standalone program with a graphical interface. Figure I.2 is a screen image of the program itself. Values of the various probabilities, the maximum number of cycles before an agent forgets, the number of time cycles to be executed, along with the random number seed (if desired) are input in the text boxes on the left-hand side of the simulation window. Then the 'Start' button is clicked. The program then steps through each cycle, displaying the current status of the cells on the grid with the grid display in the center, and showing the accumulative number of total affected cells, currently aware cells, cells which have purchased, and cells which have decided not to purchase in the graph in the right portion of the window. (The grid is 100 x 100, so there are 10,000 cells in this simulation.) The number of cycles can be set by the user, since the stable state (when no cells change state) is unknown at the outset of a particular simulation run. The timing logic for this simulation is very straightforward: every cell determines whatever operation to engage first – look to its neighbors, forget, buy, etc. – using the then-current state of neighboring cells if appropriate, and then all those operations are executed before the next is initiated. That way, there is no issue with the sequential order in which the cell status is determined.

**I.7:** Figure I.3 shows the simulation in four stages (with nominal values for the input variables). Figure I.3a shows the simulation at cycle 30. The initially aware cell is roughly in the center of the emerging mass of aware/sold/not interested cells. Note that all the aware cells are on the edges of the central mass. Figure I.3b shows results after 60 cycles, where the torus structure of the grid is revealed by the awareness cells coming in from the left side. Figure I.3c is the simulation at a stable state after 90 cycles. Most of the cellular world is occupied, with only a small area in the upper left not yet affected. Finally, Figure I.3d shows the world after all cells have been made aware. Notice that not all cells are buyers or not interested. A few are no longer aware because they 'forgot' before making a decision and there were no neighboring cells in an aware or purchased state to remind them. In this example, the probability parameters have been set to $p_{a|a}$ = 0.5, $p_{a|b}$ = 0.75, $p_b$ = 0.2 and $p_n$ = 0.4. The memory time is randomly for each cell somewhere between one and five. The evolution of a particular simulation can be tracked by plotting the number of cells that are aware, have bought, or are not interested against the time cycles. An example of this graph is shown in Figure I.4 for the same simulation illustrated above in Figure I.3. The S-shaped nature of the curves (except, of course, for the number of aware, which eventually goes to zero, and the number not aware, which is an inverse 'S') is expected, since there is an absolute upper limit on the number of cells in this example.

**I.8:** These simple agents possess several of the properties described in the agent definition. The perceptor and actor are both nothing more than state awareness. In the computer code, each cell, if unaware, "looks" at its neighbor to see if it is aware or has bought, and becomes aware according to a toss of a weighted coin, reflecting the appropriate probability. The ratiocinator is simply the toss of another weighted coin (loaded to reflect the probability of buying, or deciding not to). The state vector is both the current state and time units before it "forgets" its awareness if it doesn't purchase or lose interest.

**I.9:** This example illustrates additional properties of agents in agent simulations, such as autonomy, interaction, and asynchrony. Each agent acts to some degree independently of the others. While there are obvious relationships between neighboring cells, once an agent is aware, the agent decision to purchase or not is independent of its surroundings. Further, how long it will stay aware (assuming it doesn't purchase or lose interest) is internal to the agent. This is an example of agent autonomy. While autonomous, the agents obviously interact with each other. There are two types of interactions that can lead to awareness – either contact with a cell that is currently aware

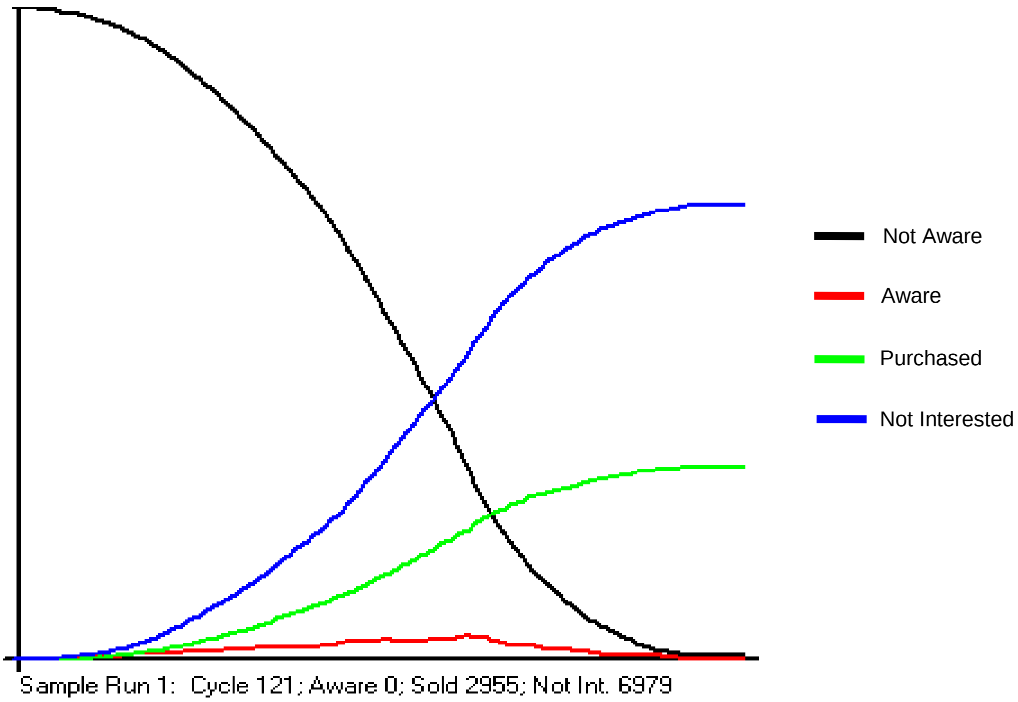



**Figure I.4: WOM Simulation Status Graph**

or contact with one that has purchased. The awareness interaction occurs at each time cycle when the neighborhood structure makes it possible. Even though the simulation moves in an orderly way through time because of the cycle structure, this really is a signal to each agent that a unit of time has passed, and if there is anything the cell must do, proceed to do it (such as decide to buy or not). The time it actually makes a decision is not synchronized with any other agent's behavior. This is asynchrony.

| Property | Mean | Std Dev |
|---|---|---|
| Number of Buyers | 2753.09 | 700.40 |
| Number of Not Interested | 6598.26 | 1675.51 |
| Number Forgot or Never Aware | 648.65 | 2374.07 |
| Number of Cycles to Stable State | 115.59 | 27.82 |

**Table I.1: Macro Statistics for WOM Simulation** (100 runs)

**I.10:** This simple agent model also shares another subtle but important property of agent simulations – emergence. Emergence is an aspect of agent models (and complex adaptive systems in general, see Johnson, 2001, and Holland, 1998) that leads to interesting questions about the ability to predict the future state of even a simple agent world. If the simulation were executed a number of times with the same parameter values used above, the expectation would be to observe consistent behavior from run to run, and the results could statistically analyzed to characterize the simulated world in a macro analysis sense. Effectively, this is a study of the final status of the simulation, leaving out the details of the cell mapping or the evolution curves shown in Figures I.3 and I.4. Table I.1 shows the statistical analysis of 100 runs of the simulation with the nominal parameters used earlier. Here the mean and standard deviation have been estimated for of the number of buyers, the number not interested, the number who forgot or never were aware, and the number of cycles required before no further change occurred (the time to a stable state).

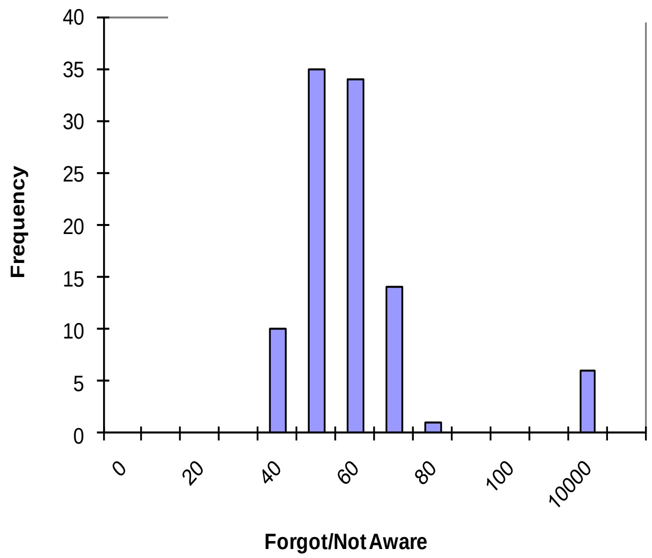



**Figure I.5: Histogram of Forgot/Not Aware.** (100 runs)

**I.11:** The values in Table I.1 seem reasonable, at least at first glance, although perhaps the standard deviations are a little large, especially that associated with the number of cells who forgot or who were never aware. One would expect, based on inspection of a few simulation runs and on the inherent logic of the simulation, to see the mean of this value near zero, with a correspondingly small deviation. But the standard deviation is nearly four times the size of the mean, which is itself far away from zero. If a histogram is drawn of the number of forgot/not aware, a problem is immediately noted. In Figure I.5 it can be seen that far out to the right are what appear to be half a dozen extreme outliers. Closer examination of the outliers reveals that these simulations failed to ever really get started, so the number not aware was very large. Recall that there is no difference from one run of the simulator to another except for the numbers produced by the random number generator. But in about five percent of the cases, the WOM dies out before it gets started. This raises the question of why five percent, as opposed to ten percent or one percent, or why any at all. Further, if only one instance of the WOM evolution were to be observed, this suggests there is a small, but certainly not negligible, probability that nothing of any particular interest would happen. The system would quickly die out, and the marketing manager, true to his spirit of incisive action and bold decision, would cancel the viral marketing program that supported the WOM effort. And he would be wrong.

**I.12:** Exploring this curiosity further, it is found that the effect of changing the probability parameters can be dramatic indeed. Set the probabilities of the simulator to some different values, say $p_{a|a}$ = 0.4, $p_{a|b}$ = 0.45, $p_b$ = 0.05 and $p_n$ = 0.8 (the memory time is left between one and five). A typical result is shown in Figure 4.9. There are far fewer buyers, and the number of forgot/not aware is substantially higher, since larger areas of the cell world are isolated from awareness contact (neighboring cells who are aware or who have purchased), giving rise to the sparse pattern observed. It is clearly much different than the pattern observed in the previous experiment. But more interestingly, the distribution of the outcomes of a number of simulation runs using these low probability settings reveals the bimodality of the result noted in our first example much more dramatically. A histogram plot of the number of buyers in each of 500 runs of this simulation reveals the pattern shown in Figure I.6. There are two distinct classes of results; a fairly large group with none or few buyers, and the remainder clustered about a mean in the range of 480 or so. In fact, approximately 36% of the runs fall in the left cluster. Figure I.7 shows the graphic results for this situation.

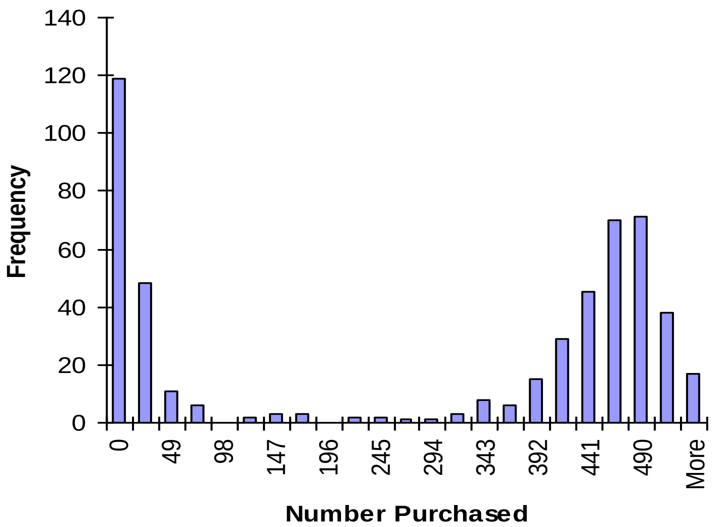



**Figure I.6: Number Purchased, Low Probability Simulation.** (500 Runs).

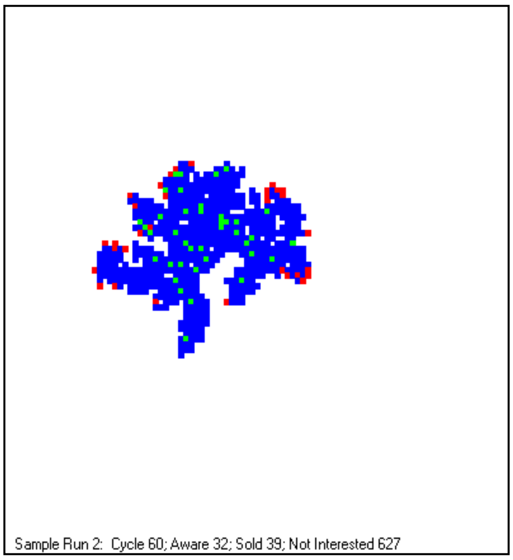



**Figure I.6a: Time Cycle**

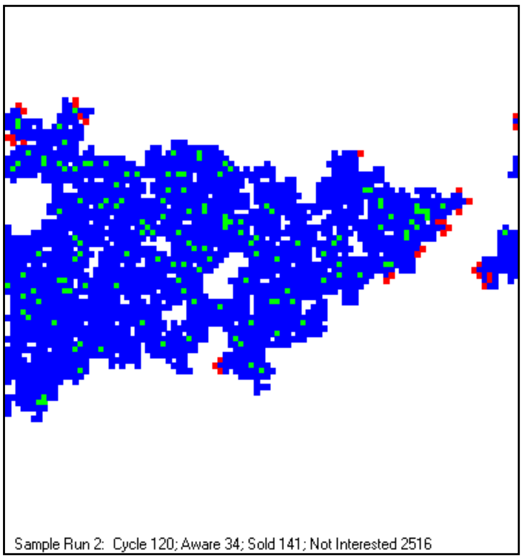



**Figure I.6b: Time Cycle**

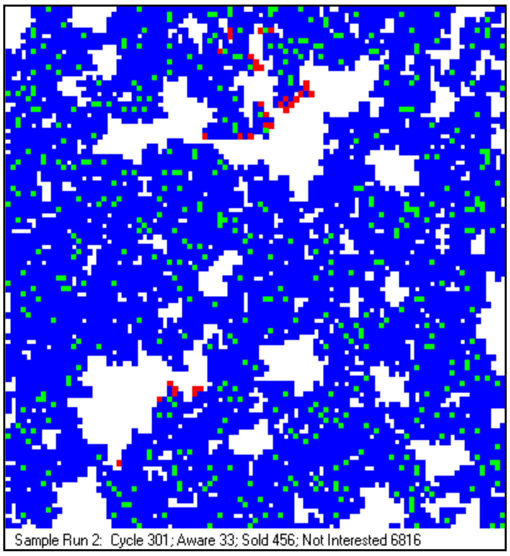



**Figure I.6c: Time Cycle**

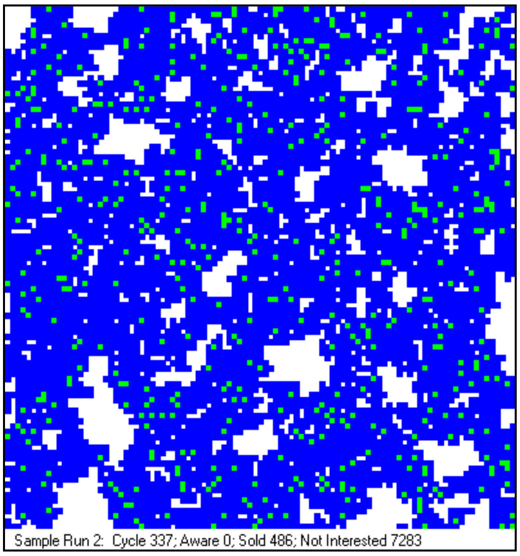



**Figure I.6d: Time Cycle 337 – Stable**

☐ Not aware ■ Aware ■ Purchased ■ Not

**Figure I.7: WOM Simulation with Lower Purchase Probability**

**I.13:** This bimodality of the simulation runs is an example of emergent behavior in an agent model. If all that could be observed was the final outcome of the simulation, with no idea of how came to be, it would be difficult to infer that the observed schizophrenia was due to the properties and interactions of the simple cellular agents that created it. This bimodal result is an example of *bifurcation*, a phenomenon that has been frequently observed in the analysis of complex systems. When bifurcation emerges, the areas of each mode are sometimes referred to as *strange attractors*, perhaps reflecting their unexpected nature. Further insight into this bifurcation behavior is gained by repeating the above experiment for different values of the input parameters. This leads to the graph shown in Figure I.8. This graph is produced by letting the value of the awareness probability vary from 0.05 to 0.45 by increments of 0.05, holding the other probabilities constant, and plotting the resultant bifurcated histogram for each value. For some set of values of the awareness probability, *trifurcation* – three peaks – emerges. This is essentially a graph of the probability distribution of the possible end states of the WOM agent model, and, I submit, could not have been anticipated by the most acute of students. [The reader is cautioned to note the scale changes on the x-axis of Figure I.8. This is done to highlight the modality of the result.]

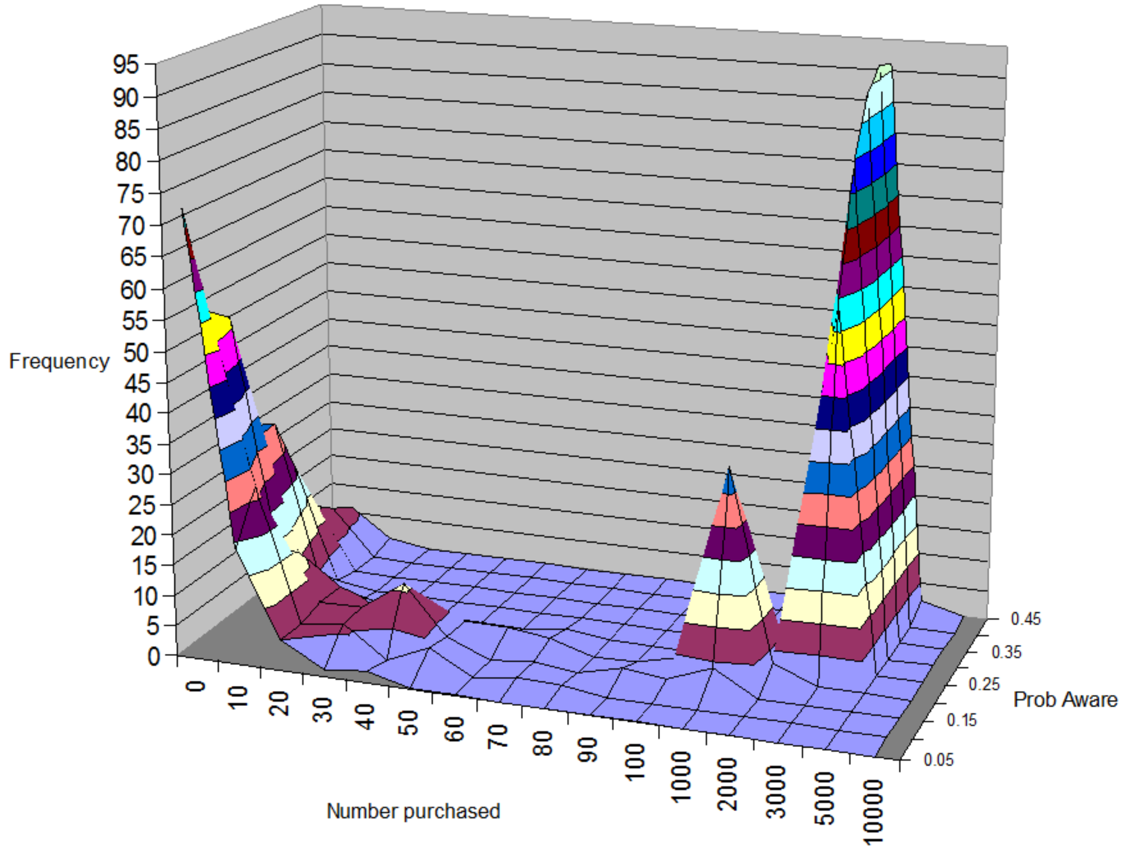



**Figure I.8: Trifurcation of Purchased Frequency**

**I.14:** This simulation example reveals how involved even a simple agent model can quickly become. But the cost in complexity yields substantial benefits in understanding. The trifurcation seen in the WOM simulation can readily be converted to a probability distribution (merely by norming to one the data plotted in the graph by the area under the three-dimensional curve), which completely characterizes the stable-state structure of this complex system. Agreed the distribution does not have an elegant equation that can be written down, but that is of no concern. Recalling Humphries' (2004) admonition that things understood do not necessarily have to be readily understood (if at all) by humans, this is a case where the description, and to some extent the understanding, is best managed on a computer. At this point in their development, computers seem not interested in elegance. A table of the data is adequate. And the possibilities for further analysis are by no means exhausted. One could study the relationship between awareness probability and forgetting, or the WOM could be enhanced by, say, introducing repeat purchases or giving a qualitative flavor to the nature of the communications between a cell and its neighbor, perhaps indicating disapproval or approval of the product. This emphasizes the perspectives and portrayals aspect of an actual agent-based model.

# Appendix J:  Glossary of AirVM Terms

**J.1:** A number of technical terms have been identified over the course of the development of AirVM. The table below collects them into one place for ease of reference.

**Table J.1:  Glossary of AirVM Terms**

| **Term** | **Definition** |
|---|---|
| Abort | To terminate a simulation before it reaches completion, usually because of a miss-specification of one or more conditions for the simulation. |
| Agent | A computer program routine that emulates the salient behavior of an entity important to a complex system. In AirVM, there are three agents, the *Pag*, *Arasag*, and *Dsag*. |
| Agent-based model | A computer model wherein the behavior and interaction of the individual entities – so called agents – in a system are replicated, so that the structure and behavior of an otherwise-intractable complex system can be described, understood and forecast. A technique of complex systems analysis made possible by extensive computing resources. |
| AirVM | Virtual M1nds' Airline Virtual Market. An agent-based model which represents the purchase of air travel tickets in all *markets* in the *global airline network*. |
| Arasag | *A*irline *R*evenue *A*nd *S*cheduling *AG*ent. The class of agents in AirVM that represent those components of airlines concerned with schedules and flight pricing. |
| ASM | Available Seat Mile, which is one seat available for sale for one flight mile. A measure commonly used in the industry for total system capacity. |
| Availability | At any point in time prior to departure, the number of seats on a specific airplane at a given price available for purchase by passengers. |
| Booking | The reservation by a passenger for one or more tickets on a specific itinerary in a specific market. The ticket(s) is not purchased, merely reserved. See *ticketing*. |
| Booking curve | The rate of bookings as a function of time during the *booking period*, usually described as a curve with increasing slope as departure approaches. Also see *ticketing curve*. |
| Booking period | The period of time, measured in days before departure, over which simulated ticketing activity for a given flight occurs. In AirVM, the booking period is usually 90 days. Synonymous with *ticketing period*. |
| Capacity | The number of seats available on an aircraft for sale for a specific flight. Also, with appropriate modifiers, refers to the number of seats in a specific cabin on the airplane. Cannot change during the course of a simulation. |
| Carrier market share | The fraction of travel in a specified market being provided by a specific carrier. |
| Channel | Refers to the sales channel source of an airplane ticket. Common channels are: **agency**, where the ticket is sold by a travel agent; **direct**, where the passenger buys the ticket directly from the airline; and **resellers**, where tickets are purchased from an airline by an intermediary and resold to the passengers. In AirVM, the *dsag* represents the channel. |

| Term | Definition |
|---|---|
| Child scenario | A *scenario* derived from another *scenario*. All scenarios except the *seed scenario* are derived from other scenarios by making changes in the scenario's *network topology* or its *demand matrix*. The derived scenario is referred to as the child scenario, and the one derived from is the *parent scenario*. |
| City | A geographic location served by at least one airport which has commercial, regularly scheduled air service. In AirVM a city is identified by its three-letter IATA code, eg. GVA. Some cities have several associated airports – eg. LON has LHR and LGW, among others. |
| City-pair | The directional market defined by an origin city and a destination city. |
| Closed journey | A *journey* wherein the final *destination* is also the *origin*. |
| Demand | The number of individuals who want to make a journey from an origin to a destination in a specified period of time, and for a specified price on a specified commercial airline network. |
| Demand matrix | An internal AirVM database which specifies the *demand* for each *city-pair* in the *global airline market*. The demand matrix also specifies how demand will change over time, often hypothetically, to enable alternative demand realizations to be tested against a fixed *network topology*. |
| Destination | A city which a passenger travels to satisfy some purpose other than simply the debarkation point of the airplane on which the trip is made. |
| Downline load | The number of passengers who boarded at the local origin that are going to a destination beyond the local one |
| Dsag | Distribution system agent. Represents the behavior of intermediaries, such as travel agents, which act between pags and arasags. |
| Equipment | Synonym for airplane, generally when referring to different types or configurations of aircraft. Of importance to AirVM is the operating speed and capacity of the airplane. |
| Equipment assignment | The assignment of a specific type of *equipment* to the *flights* in a *network*. |
| Fare | The amount paid by the pag for access to a given itinerary in a specified OD market. The fare includes all taxes, surcharges and other fees levied by the airlines or governmental entities. It is the total amount paid for the itinerary. |
| Fare class | A collection of seats in a cabin that are priced the same. The pricing for a fare class is usually set based on restrictions such as advanced purchase requirements or refundability restrictions. The fare class is the unit of pricing for the revenue management protocol. |
| Flight | The movement of one airplane consisting of taking off from one airport and landing at another, wherein passengers may embark or debark at the relevant airports. |
| Flight Leg | A flight that takes off from exactly one airport and lands at exactly one other, both of which allow passenger embarkation and debarkation. Also called simply a leg. |
| GC distance | The great circle distance between two cities, computed as the length of the arc of the spherical triangle with the three points being the center of the earth and the two cities. |
| Group size | The number of passengers traveling together and booking or ticketing as a single transaction. For example, a family. Also referred to as *party size*. |
| Group size distribution | The probability distribution of the group size. That is, the probability assigned to a group of size one, to size two, etc. |
| Global airline market | The collection of all OD markets in the world, considered as a whole. |
| Home | The city in or nearest the permanent residence of the passenger. |
| Impute | Attribute or assign (value) to a product or process by inference from the value of the products or processes to which it contributes (SOD): the term specifically used in AirVM to describe the use of AirVM to determine demand in markets for which data is not available by calculations applied to related markets for which data is available. |
| Inherent variation | Random variation in aspects of the Global airline market that are not controllable or predictable. |

| Term | Definition |
|---|---|
| Itinerary | A sequence of flight legs which connect an origin with a destination such that there is sufficient time at each intermediate stop, if any, for a passenger to safely move from the one leg in the sequence to its immediate next leg. |
| Itinerary market share | The fraction of travel in a market using a specified itinerary. |
| Itinerary choice | The process of selecting one from a set of itineraries serving a given OD market. See Random Utility Model. |
| Journey | A sequence of flights from a home to at least two or more destinations. See *closed journey*, *open journey*, *multi-stop trip* and *round trip*. |
| Journey structure | The specific sequence cities which are visited in a *journey*, specified in the temporal order in which they are visited. |
| Latent demand | The number of passengers that want to move from a specified origin city to a specified destination city, regardless of price or airline network configuration. |
| Load | The number of passengers on a flight leg, often qualified by the cabin in which the passengers are seated, such as First Class load. |
| Load factor | The percentage of seats occupied on a flight leg. |
| Local destination | In an itinerary, the destination of any non-stop leg. |
| Local load | The number of passengers on board a flight leg traveling in that leg's local market. |
| Local market | The market served by any nonstop leg. |
| Local origin | Of a flight, the last airport visited |
| Market | An ordered city pair with the first city referred to as the *origin* and the second as the *destination*. A passenger is in a given market if s/he wants to purchase a ticket to travel from the origin to the destination. It is further necessary that going to the destination must be for some reason other than to complete an itinerary to a third city. That is, there must be some purpose for going to the destination not related to the routing of an airplane. |
| Market share | The fraction of the total passengers taking a specific itinerary (*itinerary market share*), or using a specific carrier (*carrier market share*), or the fraction of revenue (*revenue market share*) being earned by a specific itinerary, flight or carrier. |
| Midline load | The number of passengers on a flight leg which are travelling in markets with origins before the local origin and destinations after the local destination. |
| Monitor | A display that appears during the execution of an AirVM simulation and supplies details of the simulation as it proceeds. Flights, passengers and markets can be monitored. Monitors can be visible or non-visible. |
| Monte Carlo | A Monte Carlo simulation is the situation wherein a number of simulations are run against exactly the same scenario, with selected characteristics changing at random. This allows the study of the stochastic properties of the network under various uncertainty constellations. Only Monte Carlo simulations of Inherent Variation is currently implemented. |
| NCAD | Network- and Channel-Adjusted Demand |
| Network | An abstract description of the regularly scheduled *flights* serving a defined set of *cities*. A network is assumed to be fixed during a simulation, and includes a fixed set of *cities* and *flights* which depart and arrive at those *cities*. |
| Network topology | The collection of *network*, *equipment assignment*, and RM *fare protocols* that describe the air service on which simulation is being run. The network topology is changed if the network, equipment assignment or fare protocols change. |
| OD | Abbreviation for *origin-destination* |
| Open journey | A journey wherein the final destination is not the home. |
| Origin | The city in which a passenger begins a journey. |
| Origin-destination | The *city-pair* connected by an *itinerary*. Used most often to designate *demand* between the *origin* city and the *destination* city, as opposed to the demand associated with a specific *flight*, which is composed of portions of many origin-destination demands. |

| Term | Definition |
|---|---|
| Outcome | In AirVM, the database that contains the output of a simulation. Specifically, the loads and revenues for all flights in all markets in the *global airline market* under the *network topology* and with the *demand matrix* on which the simulation is run. |
| Oversampling | In an AirVM simulation, the operation of randomly selecting *pags* more than once for *ticketing* in that particular simulation run. Used to study the impact of hypothetical sharp increases in *demand* without making modifications to the *demand matrix*. |
| Pag | Passenger agent |
| Parent scenario | The *scenario* from which another is derived by making changes in its network topology or demand matrix. |
| Party size | See *group size*. |
| Passenger | An individual wanting to travel on board an aircraft. |
| Passenger Itinerary Choice | The process by which a passenger selects from the available *itineraries* supporting a given *OD market*. In AirVM, that process is represented by an empirically derived discrete choice random utility model. |
| Passenger market share | The fraction of the count of passengers traveling in a market using a specified service. |
| Pause | In AirVM, an action to stop execution of a *simulation* momentarily, then continuing. Often used to examine the *results* displayed by a simulation *monitor* which is expected to change during the course of a simulation. |
| Prototype week | A weekly time period containing the departure or arrival times for all *flights* in a *network*. In AirVM it is assumed that a prototype week repeats week after week. That is, the airline schedule is exactly the same week after week. Seed scenarios are based on airline schedules as reported by industry sources (e. g. OAG or Innovata) and supplied for a specified seven day period. |
| PSM | Passenger seat mile, which is defined as a ticketed passenger actually flying for one flight mile |
| RM | Abbreviation for *revenue management* |
| Result | In AirVM, a result is the assignment of one or more *pags* to a set of *itineraries* in the *markets* defined by the *network topology,* together with the associated *fare* and required *capacity* changes. See *outcome*. |
| Revenue market share | The fraction of revenue produced by all passengers in a market using a specified service. |
| Revenue Management | The theory and practice of setting fares in response to observed demand in an attempt to maximize total revenue by charging the highest price that demand will meet. Also referred to as Yield Management. Frequently abbreviated to *RM*. |
| Revenue Management (RM) Protocol | The set of fares and rules by which fares are set during the *ticketing period*. Some RM protocols include: **fixed**, where the fares don't change; **fixed time**, where they change by a preset amount at fixed times during the ticketing period; and **forecast**, where various methods are used to estimate total demand at departure as booking/ticketing occur, adjusting fares dynamically as expected final demand fluctuates. |
| Save | In AirVM, operations that produce a *result* have save buttons, which allows the user to store the result – usually as a *csv* file – on permanent storage for future use. |
| Scenario | The definition of the air travel environment that the simulation will run against. This is made up of the *network topology* and the *demand matrix*. |
| Seed scenario | The *scenario* which represents the airline as it is at a particular instant that is used as the baseline for a set of simulations. The network topology is usually based on the actual, observed state of the world's airline network in a defined prototype week and for a defined demand matrix. |
| Session | In AirVM, the act of a user interacting with the simulator to define a *scenario*, define a *demand matrix*, and execute one or more simulations using the defined scenario and demand matrix. |
| Sibling scenario | A scenario created from the same parent scenario as another scenario |
| Stop | The landing and taking off of an airplane at an airport for the primary purpose of letting passengers disembark or embark. See *tech stop*. |

| Term | Definition |
|---|---|
| STOX | Sales Through Other *Channels*. Note that it is generally unknown exactly what these channels are |
| Tech Stop | The landing and taking off of an airplane at an airport where passengers can neither leave nor board the plane. Usually used primarily for aircraft maintenance or refueling. See *stop*. |
| Ticket | The right to travel on board a specific *flight*, in a specific seat or cabin, usually purchased through a *channel*. |
| Ticketing | The process of purchasing a *ticket*. |
| Ticketing curve | The rate at which *tickets* are purchased in an *OD market* as a function of the time before departure across the *booking period.* |
| Ticket cancellation | The act of cancellation of a purchased ticket. The result of a cancellation can be the ticketing on another itinerary, or refund of the fare paid. AirVM removes the fair paid from the revenue for the cancelled flight. It does not handle re-ticketing. |
| Ticket purchase | The transaction where a fare is paid by a passenger for one or more tickets on the flights of a specified itinerary in a specified OD market. |
| Time before departure | The amount of time from a specified date/time to the scheduled departure of the first flight of a specified itinerary. |
| Undersampling | In AirVM, the process of randomly choosing not to ticket some *pags* from the synthetic population. Used to study the effects of declining demand on a particular scenario without modifying the demand matrix. |
| Upline load | The number of passengers who boarded a flight from a market that originated before the local origin, but which has the same local destination |
| Yield | The revenue per passenger seat mile realized by a flight. |

# Appendix K: Estimating the Parameters of the Pag Choice Protocol Model

**K.1:** Discrete choice models like those above are created, like all such empirical constructs, from a combination of wisdom, wit, and experience coupled with data collection, data cleaning and the statistical processes of estimation and validation. By definition, the utility function is made up of an observable part and an unobservable part, with the latter assumed to have a known probability distribution. Choice experiments provide the material to build and validate the observable part of these models. Suppose $\mathbf{z}_{ij} = h(\mathbf{w}_i,\mathbf{y}_{ij})$ is a vector-valued function of a vector of characteristics for passenger $i$, $\mathbf{w}_i$, and vector of attributes of choice $j$, $\mathbf{y}_j$. The basic random utility model can be restated like this: For pag $i$ the utility of choice $j$ is expressed as in the following form:

$$U_i(\mathbf{z}_{ij}) = V_i(\mathbf{z}_{ij}) + \varepsilon_i(\mathbf{z}_{ij})\,. \tag{K.1}$$

Thus choice data consists of a description of the choice (the attributes described by the vector $\mathbf{y}_j$), the characteristics of the chooser, (the vector $\mathbf{w}_i$), and whether or not the alternative in question was selected by the chooser, an indicator variable that is valued at 1 if the alternative is chosen and 0 if not. This is usually laid out as a single data vector denoted by ($\mathbf{y}_j$, $\mathbf{w}_i$, $c_{ij}$) where $c_{ij}$ is the binary choice variable. The total data set on which the analysis is based can then be represented by a matrix of size $n\times(K_y + K_w + 1)$, for $n$ choice observations of passengers (one in each row), each with $K_y$ characteristics and selecting among choices with $K_w$ attributes.

**K.2:** Almost universally, choice models are estimated with the general statistical procedure of maximum likelihood. Maximum likelihood is easy to understand (but often not so easy to carry out). Simply find the value of the set of parameters which has the highest probability of being the true set, given the probability distribution of the data available for the estimation. For example, suppose the MNL model being estimated is linear-in-the-parameters with parameter vector $\boldsymbol{\beta}$. Let the multivariate probability distribution of the data, the vectors $\mathbf{x}$ = ($\mathbf{y}_j$, $\mathbf{w}_i$, $c_{ij}$), and the parameters vector $\boldsymbol{\beta}$ is represented by $F(\mathbf{x}, \boldsymbol{\beta})$. The maximum likelihood method says to determine the values of the elements of $\boldsymbol{\beta}^*$ such that

$$\boldsymbol{\beta}^* = \max_{\boldsymbol{\beta}} F(\mathbf{x},\boldsymbol{\beta}) = \max_{\boldsymbol{\beta}} \prod_{i=1}^{n} f(x_i,\boldsymbol{\beta}) \tag{K.2}$$

On the right side of this equation, $f(x_i, \boldsymbol{\beta})$ is the distribution function of the $i$-th independent observation, and the fact that the joint distribution function of $n$ independent random variables is the product of the individual distribution functions of the random variables yields the result. Write the probability that pag $i$ will choose itinerary $j$ as $P_i(j)$, and let $I(i,j)$ be the indicator function which is 1 if $i$ chooses j and 0 otherwise. Then

$$P_i(j) = \prod_{k=1}^{\#(J)} [P_i(k)]^{I(i,k)}, \tag{K.3}$$

so, assuming that each pag's choice is independent of the others (thus the equation for the maximum noted above applies), it is the case that the likelihood function $L(\boldsymbol{\beta})$ is

$$L(\boldsymbol{\beta}) \equiv \prod_{i=1}^{n} \prod_{k=1}^{\#(J)} [P_i(k)]^{I(i,k)} \tag{K.3}$$

Since the natural log is a one-to-one transformation, it is more convenient to use the natural log likelihood function, which is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}) &\equiv \ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{\#(J)} I(i,k) \ln P_i(k) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{\#(J)} I(i,k) \ln \left( \frac{e^{\boldsymbol{\beta}^T \mathbf{x}_{ik}}}{\sum_{j=1}^{\#(J)} e^{\boldsymbol{\beta}^T \mathbf{x}_{ij}}} \right) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{\#(J)} I(i,k)(\boldsymbol{\beta}^T \mathbf{x}_{ik}) - \sum_{i=1}^{n} \sum_{k=1}^{\#(J)} I(i,k) \left( \ln \left[ \sum_{j=1}^{\#(J)} e^{\boldsymbol{\beta}^T \mathbf{x}_{ij}} \right] \right). \end{aligned} \tag{K.4}$$

Ordinary calculus implies that the maximum of any (continuous) function is a point where the derivative is zero, so take the derivative of $\mathcal{L}(\boldsymbol{\beta})$, set it to zero, and solve for the required values of $\boldsymbol{\beta}^*$. Most discrete choice texts describe the process in operational detail (e. g. Ben Akiva and Lerman, 1985 or Train, 2003), and most statistical software packages (e. g. SAS (2004) or LIMDEP by Economics Software (2002)) handle the calculation directly, starting with a matrix of choice data and producing the maximum likelihood estimate of the parameters in the specified model.

**K.3:** The discontinuity of the likelihood function for the pag itinerary choice utility means that the standard techniques cannot be applied directly, because one has to accommodate the fact that the derivative is zero at the points of discontinuity at *a* and *b* (the necessary derivatives do not exist). Manual estimation must be used to resolve this. As has been mentioned previously, the coefficient values for the pag itinerary choice model are proprietary, but to give some idea of the magnitude of the coefficients and the estimation procedure, consider Table K.1. The values here are similar to those currently used in AirVM (the model is not exactly the same). This table comes from work reported previously by Parker (2004).

**K.4:** Within the airline industry, revealed preference is far and away the most used data source to estimate passenger choice behavior models. Every airline has within its records a substantial basis for the conduct of such studies – the booking and ticketing histories of thousands of flights – which makes such data so inexpensive and exhaustive that to ignore it would be foolish. Indeed, Sabre Holdings, one of the largest vendors of airline schedule planning software, engages its customers in an elaborate calibration exercise using historical ticketing data when it builds the choice model which underlies its schedule analysis software (Venod, 2006). A *revealed preference experiment* is simply the recording of actual choices made by passengers when selecting flights, as evidenced by the booking and ticket purchase behavior. These observations include, of course, the choice made, attributes of the flight (fare, departure time, arrival time, intermediate stops, booking class, cabin, airplane equipment, days in advance of departure date) and, if tied somehow to other databases such as frequent flyer records, some socio-demographic information on the traveler. Clearly, the booking and ticketing data within a typical airline provides a substantial source of revealed preference information.

**K.5:** There are some difficulties with revealed preference data in this case, however. For one, often the choice set from which the revealed preference is selected is not

**Table K.1: Estimation Results for a Pag Itinerary Choice Model** (Source: Parker, 2004)

| Model Term | | Business Trip Purpose | | Leisure Trip Purpose | |
|---|---|---|---|---|---|
| | | Estimated Value | Std Error | Estimated Value | Std Error |
| $\beta_f$ | Log Fare | -4.216 | 0.198 | -4.549 | 0.245 |
| $\beta_d$ | Duration | -1.227 | 0.091 | -0.540 | 0.101 |
| $\beta_{db}$ | Duration x Log Base Duration | -0.301 | 0.029 | 0.117 | 0.041 |
| $\beta_{dc}$ | Inline Stops | -0.168 | 0.057 | -0.097 | 0.039 |
| $\beta_{ic}$ | Online Stops | -0.126 | 0.074 | 0.167 | 0.023 |
| $\beta_{1st}$ | 1st Class Cabin Dummy | 0.359 | 0.056 | 0.302 | 0.066 |
| $\beta_{ec}$ | Economy Cabin Dummy | -0.605 | 0.061 | -0.367 | 0.068 |
| $\beta^G_{AE}$ | Arrive Sensitive Early | -0.161 | -2.518 | -0.092 | -2.622 |
| $\lambda_{AE}$ | Arrive Sensitive Early Exponent | 1.000 | | 1.000 | |
| $\beta^G_{AL}$ | Arrive Sensitive Late | -0.471 | -3.401 | -0.449 | -4.743 |
| $\lambda_{AL}$ | Arrive Sensitive Late Exponent | 0.000 | | 0.000 | |
| $\beta^G_{DE}$ | Depart Sensitive Early | -0.803 | -3.901 | -0.364 | -2.620 |
| $\lambda_{DE}$ | Depart Sensitive Early Exponent | 0.365 | | 0.000 | |
| $\beta^G_{DL}$ | Depart Sensitive Late | -0.522 | -1.781 | -0.094 | -3.681 |
| $\lambda_{DL}$ | Depart Sensitive Late Exponent | 0.470 | | 1.000 | |
| $a$ | Lower Indifference Window Bound | 0.75 | | 1.50 | |
| $b$ | Upper Indifference Window Bound | 0.50 | | 1.25 | |
| Sample Size<br>Pseudo $R^2$ | | 1546<br>0.335 | | 3914<br>0.310 | |
| All coefficients are significant at the 0.05 level | | | | | |

known.  Assumptions can be made about the level of knowledge the decision-maker has about the available options and this issue goes away.  If the IIA assumption is justified, then there is no problem since knowledge of options not considered is of no consequence.  But to *validate* the IIA assumption requires knowledge of all the options available, and that can often be quite difficult in practice.  Another deficiency of revealed preference experiments is that alternatives that don't exist cannot be tested. For example, if one is interested in the value an airline passenger might put on a faster airplane, revealed preference cannot be used to estimate the models because there are no faster passenger aircraft available (at least for context of commercial air travel).

**K.6:** This problem can be remedied by using *stated preference* experiments, another method of choice data collection.  It is referred to as stated preference because the respondent makes choices in a hypothetical choice environment, rather being observed making actual choices.  In other words, he is asked which choice he would make from a list of options, as if those were the only options available.  Often these hypothetical choices are offered in the form of a survey, and such surveys can be conducted on the internet at very low comparative cost.  In 2004, while I was with the Boeing Company, I conducted a stated preference experiment selecting respondents from individuals using a particular flight display engine called SideStep.  SideStep is an internet application which allows the user to see flight options from a number of airlines as provided by the airline's own web sites, as opposed to choices offered through online travel agencies such as Expedia.  While a person is using SideStep to shop for a flight, it is possible to interview her via the internet, offering another choice that might not exist – for example a non-stop where none is now available – thus providing an excellent stated preference environment.  Jordan Louviere and his colleagues at the University of Technology Sydney assisted in important design elements of the SideStep survey.

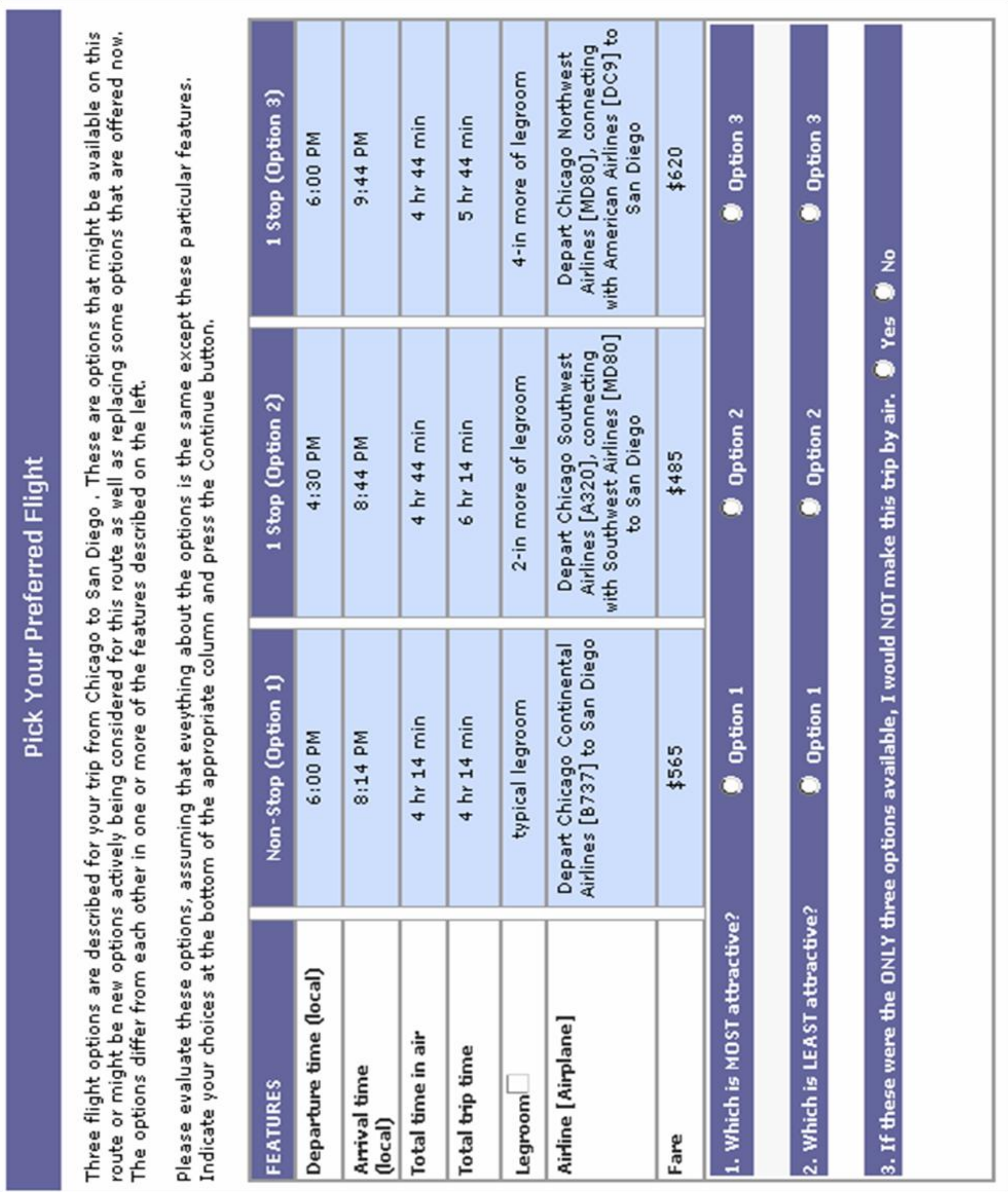
**K.7:** The survey was conducted on the internet by intercepting a customer after they had indicated the origin and destination of the trip they were interested in.  This origin and destination were then incorporated into the survey form, giving a more realistic context in which the respondent could frame his answers.  The screen image shown in Figure 6.6 gives one of the choice pages of that survey.  There were a number of choice combinations available, and respondents only saw three out of the 256 that were being studied.  But because there were more than 3500 respondents, all choice combinations were available to a significant number of surveyed individuals, providing presumably sufficient sample size for statistical validity.[95]  The choice experiment being carried out in this survey is the selection of itinerary based on seven attributes, called "features" on the survey screen, as illustrated in Figure K.1.  These include departure time, arrival time, total time in the air, total trip time, leg room (defined further in a popup window the user could select), airline and airplane routing, and fare.  These seven attributes each were defined with two or more "levels" or values.  For example, there were four different fares that could be assigned to each alternative.  The set of all combinations of levels of all attributes creates the choice set, and each respondent was given three of the possible combinations from which to select.  Under the assumption of the independence of the decision-makers, the results from many interviews are combined to develop a maximum likelihood estimate of the parameters of the models in a manner following the discussion above.  The SideStep survey also asked a variety of socio-demographic

[95] This begs the question of the convergence rate of the coefficient estimates in discrete choice experiments.  Alas, this is a topic for another time.

questions, such a as age, marital status, income and education, and so forth, thus allowing exploration of the relationship of these characteristics of the decision-maker to the attributes of the choices made.

**K.8:** Like revealed preference experiments, stated preference surveys have their drawbacks. The most obvious is the reliability of a choice made in an artificial environment versus the same choice made in the real world. There is a long and involved debate with associated literature on this issue, which is unfortunately beyond the scope of our discussion. Suffice it to say that both methods are widely used, and, if done properly, there is no sound reason not to utilize stated preference methods. Louviere *et al.* (2000) give a thorough discussion.

**K.9:** Those familiar with the application of discrete choice models may note with concern the relative absence of characteristics of the passenger from the model's formulation. There are no interaction terms representing the effects of age or gender, for example, or any explicit consideration of the income or occupation of the traveler. An early form of a similar model did contain a term for the ethnic origin of the traveler, but it has not been reproduced in other work. Perhaps because of its historical orientation toward revealed preference data, and the difficulty associating such data with socio-demographic variates, these features of the demand analysis have not been effectively explored. That they are important there can be no doubt, but, when stood against fare, travel time, number of stops and other itinerary attributes they may be of such subtle impact as to be invisible to the typical researcher.



## Pick Your Preferred Flight

Three flight options are described for your trip from Chicago to San Diego . These are options that might be available on this route or might be new options actively being considered for this route as well as replacing some options that are offered now. The options differ from each other in one or more of the features described on the left.

Please evaluate these options, assuming that eveything about the options is the same except these particular features. Indicate your choices at the bottom of the appropriate column and press the Continue button.

| FEATURES | Non-Stop (Option 1) | 1 Stop (Option 2) | 1 Stop (Option 3) |
|---|---|---|---|
| Departure time (local) | 6:00 PM | 4:30 PM | 6:00 PM |
| Arrival time (local) | 8:14 PM | 8:44 PM | 9:44 PM |
| Total time in air | 4 hr 14 min | 4 hr 44 min | 4 hr 44 min |
| Total trip time | 4 hr 14 min | 6 hr 14 min | 5 hr 44 min |
| Legroom | typical legroom | 2-in more of legroom | 4-in more of legroom |
| Airline [Airplane] | Depart Chicago Continental Airlines [B737] to San Diego | Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego | Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego |
| Fare | $565 | $485 | $620 |

1. Which is MOST attractive? Option 1 Option 2 Option 3

2. Which is LEAST attractive? Option 1 Option 2 Option 3

3. If these were the ONLY three options available, I would NOT make this trip by air. Yes No

**Figure K.1: SideStep Web Survey Page**

# Appendix L: The Estimation of OD Demand

## L.1: Introduction

**L.1.1:** This appendix briefly describes the Trip Generation Model (TGM) and Trip Distribution Model (TDM) of the econometric OD Demand Forecast Model. The TGM produces estimates of the total monthly passenger airline trips to all destinations generated from a selected city, whereas the TDM produces estimates of the proportion of the generated trips that are expected to be observed between the given origin and a specific destination. The econometric theory implemented in this work is based on material from Carson *et al.* (2010). This treatise describes extensive research into OD demand modeling done for the United States. It examines alternative forms, different independent variable combinations, and various aggregate and disaggregate properties of several model structures.

### L.2: The Trip Generation Model (TDM)

**L.2.1:** A separate trip generation model has been derived for each of the largest 184 US cities with sufficient data. The following model is estimated for each city. Let $P_{it}$ be the population of city $i$ in month $t$, and let $T_{it}$ represent the total trips taken by passengers originating in city $i$ in month $t$. Thus $C_{it} = T_{it} / P_{it}$ is the monthly per capita trips from city $i$. The model form is

$$\begin{aligned} y_{it} &\equiv \ln\left[\frac{C_{it}}{1-C_{it}}\right] \\ &= \alpha + \beta g_{it} + \lambda u_{it} + \phi_1 c_{it} + \phi_2 c_{it}^2 + \varphi J_t + \delta \Delta f_t + \sum_{k=1}^{11} \theta_k I_k + \eta y_{i,t-1} + \varepsilon_{it} \end{aligned} \tag{L.1}$$

In this formulation, the Greek letters represent empirical coefficients estimated from available data (discussed below). The independent variables are as follows (in all cases $i$ denotes the city and $t$ the month):

$g$ = the population growth rate for the city,

$u$ = the unemployment rate for the State in which the city is located,

$c$ = the coincident index of economic activity for the State in which the city is located,

$J$ = the average national jet fuel price (cents per gallon) for the month

$\Delta f$ = the change from the prior month of the 3rd nearby NYMEX oil futures price

$I_k$ = 1 if the month equals k (k = 1 being January), 0 otherwise, referenced to December,

$y_{i,t-1}$ = the value of the dependent variable lagged one month.

The dependent variable is the logit transformation of per capita trips. The logit transformation is used to unbound the variable and stabilize the variance. The population, unemployment, coincident index and jet fuel price are straightforward. The oil futures price change accounts for the future price of oil, which is highly correlated with jet fuel price. The summation term accommodates monthly adjustment parameters for seasonal effects, where the empirical coefficients estimate the monthly effect and the indicator functions (the *I*'s) designate the appropriate coefficient to use for the month that is represented by *t*. The lag term reflects the autocorrelation between trip production in one month and trip production in the previous month.

The term $\varepsilon_{it}$ represents the usual error term associated with linear regression analysis.

**L.2.2:** With the logit of the per capita trips known, total number of generated trips is readily computed using the inverse of the logit transformation. That is,

$$\begin{aligned} y_{it} &= \ln\left[\frac{C_{it}}{1-C_{it}}\right] \\ \Rightarrow \left[\frac{C_{it}}{1-C_{it}}\right] &= e^{y_{it}} \\ \Rightarrow C_{it} &= \frac{e^{y_{it}}}{1+e^{y_{it}}} \\ \Rightarrow T_{it} &= \frac{P_{it}e^{y_{it}}}{1+e^{y_{it}}}. \end{aligned} \tag{L.2}$$

**L.2.3:** The air passenger tickets processed by the all data sources – bank settlement plans, global distribution systems, or carriers – are not all the tickets sold in a specific market. Data from an airline industrial source, whose identity must be masked for non-disclosure reasons, is used to estimate the models here. Most ticket data must be adjusted to reflect the share of total travel represented by the channels visible to the source. At this time, there is no data of known reliability from which this share can be estimated. The best that can be done until with existing data is to use the USDOT's T-100 data to estimate channel share. USDOT's DB1B database gives a 10% sampling which does cover all channels, although the sales channel is not identified. The T-100 reports monthly tickets by carrier. However, it does not report markets as defined by this model (which is true passenger origin-destination), but rather the use of itineraries that have a common flight number, but which may serve several markets. For a given data supplier, isolating the tickets that meet the same criteria as the T-100, however, allows the fraction of total tickets that are seen by the supplier to be determined. The rate of supplier tickets per T100 ticket is called the *ratio*, and is denoted $r_{it}$. With $r_{it}$ known for a given city *i*, then total tickets generated, $D_{it}$, is simply $T_{it}/r_{it}$. Thus

$$D_{it} = \frac{T_{it}}{r_{it}} = \frac{P_{it}e^{y_{ij}}}{r_{it}\left(1+e^{y_{ij}}\right)}. \tag{L.3}$$

In **L.2.7** below, the deficiencies in the use of the T-100 data to determine the ratio are examined from an empirical perspective based on the actual ratios calculated for the cities for which trip generation estimates have been prepared. In particular, the computation of an empirical 80% confidence limit for the ratio in each market quantitatively portrays the weakness of this approach.

**L.2.4:** Except for the ticket totals supplied by industry for the dependent variable, all other data required is publicly available. Also, for the model to be used for estimating future trip generation, the independent variables must be forecast to those future time periods.

> **L.2.4.1:** Monthly population estimates are produced by the US Bureau of the Census. In this analysis, the geographic area unit used for population determination is the CBSA which contains the market city. The US Census is considered the most reliable available population data. Estimates are based on projections from the 2000 US Census, and are updated annually. The census data can be found at http://www.census.gov/main/www/access.html.
>
> **L.2.4.2:** The unemployment rate for each state is produced monthly by the US Department of Commerce, Bureau of Labor Statistics. They also produce some rate figures for individual cities, but not all market cities are included, and reporting time lags differ, so the State figures are used. Data for a particular month become available within the first two weeks of the following month. Unemployment data is available at http://www.bls.gov/lau/
>
> **L.2.4.3:** The coincident index of economic activity is also produced monthly for each State by the Federal Reserve Bank of Pennsylvania. There is about a month lag in data availability. Forecasts of this variable by the Federal Reserve are the best available, from http://www.philadelphiafed.org/research-and-data/regional-economy/indexes/coincident.
>
> **L.2.4.4:** Spot jet fuel prices (in cents per gallon) are reported daily by the US Department of Energy. The value used here is the average for the month of prices at all five of the reported jet fuel source ports. Jet fuel prices are very closely correlated with per-barrel oil prices, and forecasting that out more than a few months is fraught with error. It is best to use forecasts produced by the Department of Energy, but even those are very unstable. These can be found at http://tonto.eia.doe.gov/dnav/pet/hist/ rjetnyhm.htm.
>
> **L.2.4.5:** The NYMEX prices (in dollars per barrel) of 3rd Nearby Oil Futures are also reported daily by the US Department of Energy (DOE). The values used here are the monthly averages of that price. Forecasts of this variable are also best taken from the DOE, but futures price is inherently very unpredictable. Data is available at http://www.nymex.com/lsco_fut_cso.aspx.

**L.2.6:** One hundred and eighty-four individual Trip Generation Models are estimated, one for each city of sufficient size in the United States. The parameters of the TGM are estimated for each of the 184 cities using weighted least squares estimation on the data for the 43 time periods from October, 2005 until April of 2009. Only 42 periods are used, since the 1st must be used for the lag term for the 2nd, thus eliminating it from consideration. (The first few months of the estimation are thus prone to higher autocorrelation error.) Thus the coefficients for all terms in the model, except for the monthly seasonal parameters, are based on 42 sample points. The monthly seasonal values are the same for all markets, and are estimated by combining the monthly reported ticket sales for the 184 market cities into one large sample. Estimating 11 parameters with only 42 data points is very unstable, but combining the markets gives a sample size of 8,148, yielding substantially more accurate estimates. The same seasonal parameter values are thus used for all markets. Weighted least squares regression is applied because of the variance effects of the logit transformation. The weighting that is applied is the reciprocal of the variance of the probabilities in the logit, that is, the observed per capita tickets. Table L.1 shows the estimation for a specific city, as an example of the model output. Please note that the city has been masked because of the nondisclosure policy with regard to the dependent variable data source.

**L.2.7:** The ratio data, as noted above, comes from the industry source and the US DOT. There is significant variation in the ratios from one month to another in most market cities. An extensive statistical analysis was therefore done of the ratio data to determine if any regularity, such as seasonality effects, could be found to account for the at least some of the variability. This analysis consisted of 39 data points for each city, rather than 43, since more recent T100 data is not available. Unfortunately, no such regular patterns were detected. Indeed, the results of the analysis showed that: 1) there was no trend, not even a linear one, detected over the 39 monthly data points, and, 2) the dispersion about the mean – measured by standard deviation, skewness and kurtosis revealed widely differing values from city to city. This means that there seems to be no general-purpose probability distribution that could be applied to account for the observed ratio variability. There would be natural tendency to check the accuracy of the estimates by comparing the total generated tickets estimate with USDOT T100 reports for the same period. However, this is not correct. The definition of market for T100 differs from that used in the TGM, so direct comparisons are not valid. Such a comparison could be undertaken for at least for some cities if a matching set of data were generated by AirVM, for example, which might duplicate the DOT definitions.

**L.2.8:** With such unstable variability, however, simply using the mean ratio without some measure of the level of variability in that ratio seems unwise. Therefore, an "empirical confidence interval" is supplied with each estimate of the mean ratio. Given that the probability distributions of the individual city ratios vary widely, a non-parametric interval definition was chosen. Specifically, since there were 39 data points for each city ratio, the 4$^{th}$ and 35$^{th}$ points of the ratio data ranked in ascending order gives an approximately 80% confidence interval for the ratio. The mean ratio is thus used as the value of $r_{it}$ for the generated trip calculation, but also reported are the lower and upper bounds of this empirical confidence interval. Calculation of the trips estimated at these bounds (the lower ratio bound yielding the upper generation bound, since one is dividing by a fraction, and *vice versa*) thus gives some reasonable idea of how much variation in the generation estimate is introduced by the variation in the T100 data.

## L.3: The Trip Distribution Model (TDM)

**L.3.1:** A common approach to Transportation Distribution Modeling (TDM) is called the *gravity model*. The name comes from the mathematical form of the model bearing a (vague) resemblance to Newton's law of gravity, where gravitational attraction between two objects is proportional to the product of the masses of the two bodies divided by the distance between them. In a generalized sense, the TDM described here is a gravity model, in that the proportion of trips from a given origin to a specified destination is a function of some measure of attractiveness between the origin and destination (desire to go from one to the other) and the "distance" between the cities, being not only physical distance but also price (fare) and other factors that hinder travel resulting from the structure of the airline network, which is called impedance. Gravity models are discussed from various economic and transport perspectives in Stewart (1948), Isard (1960), Anderson (1979), Anas (1983), Bergstrand (1985), Mátyás (1998), and Cameron and Trivedi (2005). While a separate trip generation model has been derived for each city, only one trip distribution model is required for the estimation of any OD pair where the origin city is one the original 184.

**L.3.2:** The general model for the distribution of trips is as follows. Let $p_{ijt}$ be the fraction of the passengers who originate in city $i$ and fly to city $j$ in month $t$. The model form is

$$z_{ijt} \equiv \ln\left[\frac{p_{ijt}}{1-p_{ijt}}\right] = \mu f_{ijt}^{*} + \lambda_1 d_{ij} + \lambda_2 d_{ij}^2 + \lambda_3 d_{ij}^3 + \lambda_4 d_{ij}^4 + \pi v_{ijt-1} + \tau \sum_{k\in K} w_k A_k + \xi_{ijt}. \qquad \text{(L.4)}$$

In this formulation, the Greek letters represent empirical coefficients estimated from available data, while the Roman letters signify independent variables. The independent variables in this equation are quite complex when compared to many regression-type equations, and warrant separate discussions to make their derivations clear. The term $\xi_{ijt}$ represents the usual error term associated with linear regression analysis. Also, as is the case with the TGM, the dependent variable is the log odds of the probability (the likelihood) that a passenger will fly to $j$ given he started in $i$. This stabilizes the variance of the bounded variable $p_{ijt}$, and a simple transformation returns $p_{ijt}$ from $z_{ijt}$. The estimation of this model is shown in Table L.2. The $R^2$ value of 0.9146 is reasonable for this type of analysis.

**L.3.3:** There are four independent variables in this estimator – Fare Proxy, Distance, network Impedance and Destination Fixed Effect. Each of these will be described in turn, starting with the most obvious, distance.

**L.3.3.1:** *Distance***.** This is the great circle distance, $d$, in 100's of miles, between the city pair. Clearly, if A and B are two cities, the distance from A to B is the same as from B to A. Notice that there are four terms in the estimation equation involving $d$, being $d$ to the first to fourth power. There are four distance terms to capture anomalies associated with distance (if any), and because they cost next to nothing to add to the model, and improve the fit to a small extent.

**L.3.3.2:** *Fare Proxy.* In a model such as this fare is endogenous. That is, demand depends on fare so the fares are not independent of demand, and cannot thus be a proper independent variable in the model. To address this problem, a so-called *instrumental* variable is used. The effect of an instrumental variable is to create a proxy for fare that is independent of the demand. Several authors discuss instrumental variables and their appropriate use in econometric modeling, including Jung and Fuji (1976), Evans *et al.* (1993), Morrison and Winston (1990), Angrist and Krueger (2001), Katz (2001), Espasa *et al.* (2002), Greene (2003), Hahn and Hausman (2003), Gillen *et al.* (2004), and Morrison *et al.* (2005). The fare proxy instrumental variable used in this model is given by

$$f_{ijt}^{*} = \alpha + \beta t + \sum_{j=1}^{11} \gamma_j I_j + \delta f_{ijt-1}. \tag{L.5}$$

In this equation, as usual, the Greek letters are empirical parameters. $f_{ijt}^{*}$ is the fare proxy value for fares from *i* to *j* at *t*, the trend (time period) variable, where $t = 0$ is February, 2008. The $f_{ijt-1}$ is the actual observed average fare in the *ij* market lagged one month, to time period $t-1$. The summation expression represents a constant which reflects the month effect on fare. The $\gamma$'s are empirical coefficients that result from the time series regression of fare described below, and the $I_j$'s are indicator variables for the respective months; that is, equal to 1 if *t* is that month and 0 otherwise. An analysis of 10,000 fares supplied by the industry source is used to estimate the parameters of the fare proxy. The results of the regression are shown in Table L.3. The adjusted $R^2$ value for the model is 0.937, quite remarkable for econometric analysis. The coefficient values in the estimation equation are all of reasonable sign and magnitude.

**L.3.3.3:** *Network Impedance:* The airline network itself, of course, has an effect on demand beyond just the fare of an itinerary connecting origin *i* with destination *j*. For example, nonstop routes are preferred to one-stop itineraries, as they are more convenient and generally faster. Thus the allocation of travel is to some extent affected by the network configuration. Measuring the network configuration has been an issue in all air travel demand modeling (see, for some perspectives, Jung and Fuji (1976), Saab and Zouein (2001), Brons *et al.* (2002), Brueckner (2004), and Berry *et al.* (2006). It is also true that the travel network always *impedes*, rather than enhances, the ability of a passenger to move from an origin or destination (unless one is flying purely for the joy of riding on a commercial airplane) to a greater or lesser extent. In Carson *et al.* (2010) the ratio of the number of nonstops to the total number of itineraries connecting a market is used to measure this impedance, based on the analysis performed by a proprietary network analysis program used in that work. They also comment on the inadequacy of this approach, which leads to the consideration of more efficacious alternatives. It is suggested in the Carson analysis that the concept of *inclusive value* from discrete choice methodology be considered. Inclusive value is a term that refers to the expected maximum utility for a finite set of choices. That is, if one has a discrete choice model of a choice problem, the concept of the probability of each available choice is well-defined, and thus one can compute the mathematical expectation of functions of that choice structure, in particular the expected maximum utility. See Ben-Akiva and Lerman (1985) and McFadden (1999) for more detailed discussions. As the network connecting an origin city with a destination city changes, with improvements in the service (such as the addition of a nonstop) or reductions in service (such as the elimination of some routes), the air passenger's expected maximum utility as defined by the discrete choice model representing the passenger's itinerary selection decision will also change. Thus the expected maximum utility measures how "good," (or more accurately, how "less bad") a particular network configuration is – a direct measure of impedance. Let $V_{ij}(m)$ represent the utility of alternative itinerary $m$ connecting the origin $i$ to the destination $j$. Let $M$ be the set of all such itineraries. Then, under reasonable conditions, (Ben-Akiva and Lehrman, 1985, pp. 100-129), the probability of a passenger choosing alternative $m$ is

$$\Pr[k] = \frac{e^{V_{ij}(m)}}{\sum_{n \in M} e^{V_{ij}(n)}}. \tag{L.6}$$

This is the classic multinomial logit model of choice. The expected maximum utility for the set of alternatives $M$ is given by (McFadden, 1999, p 259)

$$\mathrm{E}\left(\max_{m \in M}(V_{ij}(m))\right) = \ln\left[\sum_{n \in M} e^{V_{ij}(n)}\right]. \tag{L.7}$$

The quantity for a given OD market can be readily computed from AirVM. AirVM is used to define the itinerary set for a specified market, and then the maximum utility is calculated for a random sample of passenger agents. The impedance measure is the average of these computed maximum utilities. The averaging is required because each passenger agent uses a somewhat different utility function, and expected (average) maximum utility is the desired measure. The results of the calculation are then stored in a matrix for use in the distribution model. When the model is applied, the appropriate value required for the calculation of the distribution proportion is extracted from the table as variable $v_{ijt}$. Note that the impedance measure contained in the equation is lagged one time period, to avoid the endogeneity that would arise from using a contemporaneous value.

**L.3.3.4:** *Destination Fixed Effects (DFE).* To a certain degree, every city has a unique "attractiveness." Orlando attracts many people because of the theme parks located there: Cleveland, not so much. To measure the attractiveness, Carson *et al.* (2009) used USDOT DB1A and DB1B data to estimate a constant for each city in the set of the 181 urban areas that was the subject of that study. The value is denoted $A_k$ in the estimation equation. In some cases, the USDOT-based estimates covered two or more of the OD model cities, so the same DFE was applied to all. The DFE is a constant for each origin city. In the expression $\sum_{k \in K} w_k A_k$, $K$ represents the set of cities in the model, $w_k$ is an indicator function that has the value 0 if $j \neq k$, and 1 otherwise, and $A_k$ is the DFE for city $k$. Thus this expression has the effect of providing the correct DFE for the destination city $j$ being considered. The coefficient $\tau$ in the distribution model equation has the effect of modifying the DFE values for this estimation data set.

## Table L.1: Trip Generation Model for a Sample City. (City name is masked.)

| T | PopGrow | Unemp | CI | CI^2 | JetFuel | Future | Month | Lag | Logit | Est | Resid |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.001454 | 5.4 | 155.72 | 24,249 | 166.835 | -2.72 | -0.02670 | -1.82955 | -1.80870 | -1.80184 | -0.00686 |
| 2 | 0.001452 | 5.1 | 157.10 | 24,680 | 171.506 | 1.19 | 0.00000 | -1.80870 | -1.77559 | -1.77054 | -0.00504 |
| 3 | 0.001133 | 5.5 | 158.08 | 24,989 | 184.680 | 5.88 | -0.10730 | -1.77559 | -1.89652 | -1.86728 | -0.02924 |
| 4 | 0.001132 | 5.7 | 158.86 | 25,236 | 177.920 | -2.52 | -0.11130 | -1.89652 | -1.93939 | -1.91036 | -0.02904 |
| 5 | 0.001131 | 5.2 | 159.35 | 25,392 | 186.080 | -2.25 | 0.06990 | -1.93939 | -1.69787 | -1.68320 | -0.01467 |
| 6 | 0.001129 | 4.7 | 159.50 | 25,440 | 206.738 | 10.89 | 0.00830 | -1.69787 | -1.72378 | -1.69973 | -0.02405 |
| 7 | 0.001128 | 4.8 | 160.03 | 25,610 | 209.749 | 0.08 | 0.04360 | -1.72378 | -1.59291 | -1.64789 | 0.05497 |
| 8 | 0.001127 | 4.9 | 160.59 | 25,789 | 209.776 | -0.65 | 0.10810 | -1.59291 | -1.40169 | -1.51528 | 0.11359 |
| 9 | 0.001125 | 4.9 | 161.27 | 26,008 | 215.813 | 4.29 | 0.15320 | -1.40169 | -1.41124 | -1.39282 | -0.01842 |
| 10 | 0.001124 | 4.8 | 161.76 | 26,166 | 217.317 | -1.13 | 0.15300 | -1.41124 | -1.40442 | -1.39288 | -0.01154 |
| 11 | 0.001123 | 4.5 | 162.90 | 26,536 | 188.529 | -9.75 | -0.02980 | -1.40442 | -1.61326 | -1.64542 | 0.03216 |
| 12 | 0.001122 | 4.2 | 163.26 | 26,654 | 177.973 | -3.56 | 0.04390 | -1.61326 | -1.75510 | -1.63820 | -0.11690 |
| 13 | 0.001121 | 4.8 | 163.57 | 26,755 | 177.889 | 0.51 | -0.02670 | -1.75510 | -1.82584 | -1.78887 | -0.03697 |
| 14 | 0.001119 | 4.7 | 164.31 | 26,998 | 189.118 | 1.29 | 0.00000 | -1.82584 | -1.78402 | -1.77456 | -0.00946 |
| 15 | 0.001177 | 5.4 | 165.35 | 27,341 | 169.287 | -7.60 | -0.10730 | -1.78402 | -1.90344 | -1.90299 | -0.00045 |
| 16 | 0.001176 | 5.3 | 166.73 | 27,799 | 177.485 | 4.95 | -0.11130 | -1.90344 | -1.97062 | -1.95827 | -0.01234 |
| 17 | 0.001174 | 4.7 | 167.34 | 28,003 | 185.523 | 2.09 | 0.06990 | -1.97062 | -1.71046 | -1.73134 | 0.02088 |
| 18 | 0.001173 | 4.3 | 168.18 | 28,285 | 201.248 | 3.75 | 0.00830 | -1.71046 | -1.73708 | -1.71864 | -0.01844 |
| 19 | 0.001172 | 4.2 | 168.59 | 28,423 | 203.948 | -1.04 | 0.04360 | -1.73708 | -1.60900 | -1.67391 | 0.06490 |
| 20 | 0.001170 | 4.4 | 169.10 | 28,595 | 208.905 | 2.82 | 0.10810 | -1.60900 | -1.45653 | -1.54097 | 0.08443 |
| 21 | 0.001169 | 4.4 | 169.03 | 28,571 | 214.719 | 5.15 | 0.15320 | -1.45653 | -1.43699 | -1.43204 | -0.00495 |
| 22 | 0.001168 | 4.3 | 169.37 | 28,686 | 208.781 | -2.39 | 0.15300 | -1.43699 | -1.41944 | -1.42048 | 0.00104 |
| 23 | 0.001166 | 4.2 | 169.41 | 28,700 | 224.946 | 5.93 | -0.02980 | -1.41944 | -1.63834 | -1.67390 | 0.03557 |
| 24 | 0.001165 | 4.1 | 169.90 | 28,866 | 238.159 | 6.23 | 0.04390 | -1.63834 | -1.72089 | -1.64494 | -0.07594 |
| 25 | 0.001164 | 4.5 | 170.37 | 29,026 | 271.894 | 9.00 | -0.02670 | -1.72089 | -1.79617 | -1.75418 | -0.04199 |
| 26 | 0.001162 | 4.7 | 170.48 | 29,063 | 262.922 | -1.48 | 0.00000 | -1.79617 | -1.77386 | -1.73664 | -0.03722 |
| 27 | 0.001161 | 5.2 | 170.69 | 29,135 | 260.864 | 0.90 | -0.10730 | -1.77386 | -1.86882 | -1.87805 | 0.00923 |
| 28 | 0.001160 | 5.3 | 170.39 | 29,033 | 275.713 | 3.06 | -0.11130 | -1.86882 | -1.88384 | -1.92158 | 0.03774 |
| 29 | 0.001158 | 5.1 | 169.81 | 28,835 | 313.900 | 8.16 | 0.06990 | -1.88384 | -1.70159 | -1.68219 | -0.01940 |
| 30 | 0.001157 | 4.5 | 169.16 | 28,615 | 340.570 | 7.64 | 0.00830 | -1.70159 | -1.77377 | -1.70739 | -0.06638 |
| 31 | 0.001156 | 4.9 | 168.14 | 28,271 | 382.988 | 14.03 | 0.04360 | -1.77377 | -1.66568 | -1.68639 | 0.02071 |
| 32 | 0.001154 | 5.1 | 167.24 | 27,969 | 392.862 | 9.70 | 0.10810 | -1.66568 | -1.50272 | -1.56229 | 0.05957 |
| 33 | 0.001153 | 5.1 | 166.46 | 27,709 | 394.651 | -0.26 | 0.15320 | -1.50272 | -1.49312 | -1.44154 | -0.05158 |
| 34 | 0.001152 | 5.3 | 165.23 | 27,301 | 330.597 | -17.21 | 0.15300 | -1.49312 | -1.48770 | -1.46062 | -0.02709 |
| 35 | 0.001150 | 5.0 | 163.55 | 26,749 | 311.719 | -14.10 | -0.02980 | -1.48770 | -1.66337 | -1.74236 | 0.07899 |
| 36 | 0.001149 | 5.4 | 160.80 | 25,857 | 227.166 | -25.22 | 0.04390 | -1.66337 | -1.76526 | -1.74461 | -0.02065 |
| 37 | 0.001148 | 6.2 | 158.24 | 25,040 | 188.135 | -19.61 | -0.02670 | -1.76526 | -1.94156 | -1.90761 | -0.03395 |
| 38 | 0.001146 | 7.1 | 154.67 | 23,923 | 141.308 | -11.34 | 0.00000 | -1.94156 | -1.84809 | -1.96129 | 0.11319 |
| 39 | 0.001145 | 8.6 | 151.23 | 22,871 | 146.322 | 1.93 | -0.02980 | -1.84809 | -2.03236 | -1.96112 | -0.07123 |
| 40 | 0.001143 | 9.1 | 148.24 | 21,975 | 128.044 | -3.61 | 0.04390 | -2.03236 | -2.09464 | -1.90470 | -0.18994 |
| 41 | 0.001142 | 9.6 | 145.19 | 21,080 | 129.648 | 5.29 | -0.02670 | -2.09464 | -1.82584 | -2.00284 | 0.17700 |
| 42 | 0.001141 | 8.9 | 142.14 | 20,204 | 140.633 | 3.17 | 0.00000 | -1.82584 | -1.87821 | -1.83267 | -0.04554 |

MODEL

| | Const | T | PopGrow | Unemp | CI | CI^2 | JetFuel | Future | Month | Lag |
|---|---|---|---|---|---|---|---|---|---|---|
| Coeff | 11.4422 | -0.0057 | -126.5035 | 0.0091 | -0.1589 | 0.0005 | 0.0005 | -0.0014 | 1.3686 | 0.3539 |
| SE Coef | 8.5659 | 0.0025 | 181.1697 | 0.0349 | 0.1058 | 0.0003 | 0.0003 | 0.0021 | 0.1650 | 0.0917 |

R-squared = 0.88473574

F = 27.29144 DOF = 32

Std Dev of Residuals = 0.02524

GENERATED TRIPS ESTIMATE

Mean DepData/T100 Ratio = 0.54001. 80% Confidence Interval = 0.47770 to 0.66710

Generated trips last period = 869,146. 80% Confidence Interval = 703,568 to 982,521

### Table L.2: Trip Distribution Model Regression Results

| Regression Statistics | |
|---|---|
| Multiple R | 0.914550076 |
| R Square | 0.836401841 |
| Adjusted R Square | 0.836236832 |
| Standard Error | 2.002196412 |
| Observations | 12016 |

**ANOVA**

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 7 | 246125.5895 | 35160.7985 | 8770.924486 |
| Residual | 12009 | 48141.56477 | 4.008790471 | |
| Total | 12016 | 294267.1543 | | |

**Coefficient Estimates**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Proxy Fare | -0.530850759 | 0.023585953 | -22.50707274 | 6.5253E-110 |
| Distance | -0.19030678 | 0.029948489 | -6.354470076 | 2.16711E-10 |
| Distance^2 | 0.003884049 | 0.003103099 | 1.251667709 | 0.210715351 |
| Distance^3 | 9.2341E-05 | 0.000117936 | 0.782976803 | 0.433656174 |
| Distance^4 | -2.71648E-06 | 1.40687E-06 | -1.930864761 | 0.053523281 |
| Impedance | -0.064149537 | 0.014517797 | -4.418682726 | 1.00175E-05 |
| Dest Fixed Effect | 0.782500614 | 0.015189118 | 51.51718468 | 0 |

## Table L.3: Regression of Fare Against Trend, Month Seasonality, and Lag

| Regression Statistics | |
|---|---|
| Multiple R | 0.968358234 |
| R Square | 0.937717668 |
| Adjusted R Square | 0.937636588 |
| Standard Error | 46.37114835 |
| Observations | 10000 |

**ANOVA**

| | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Regression | 13 | 323291659.8 | 24868589.21 | 11565.26122 |
| Residual | 9986 | 21472730.03 | 2150.2834 | |
| Total | 9999 | 344764389.8 | | |

**Coefficient Estimates**

| *Term* | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 20.68619328 | 1.963539284 | 10.53515631 | 8.12352E-26 |
| Trend | -0.596158497 | 0.042324217 | -14.08551729 | 1.24526E-44 |
| Jan | -7.447291333 | 2.365202238 | -3.14869114 | 0.001644866 |
| Feb | -16.42411171 | 2.220325215 | -7.397164886 | 1.50318E-13 |
| Mar | 32.09316291 | 2.219131651 | 14.46203649 | 6.26277E-47 |
| Apr | 6.597023694 | 2.213713975 | 2.980070492 | 0.00288876 |
| May | -5.107670749 | 2.382306189 | -2.144002636 | 0.032056745 |
| Jun | 0.294580458 | 2.37825246 | 0.12386425 | 0.901425252 |
| Jul | -10.4810219 | 2.374263379 | -4.414431015 | 1.0234E-05 |
| Aug | 1.922414143 | 2.370800445 | 0.810871344 | 0.417458883 |
| Sep | -3.058891229 | 2.368270323 | -1.291614053 | 0.196520729 |
| Oct | -3.550290531 | 2.366393931 | -1.500295654 | 0.133569449 |
| Nov | -6.779735696 | 2.365244324 | -2.866399731 | 0.004160375 |
| Lag | 0.977285403 | 0.00254198 | 384.4584092 | 0 |

# Appendix M: Using AirVM to Impute Observed Origin-Destination Demand

**M.1:** An exploration of the structure of demand allocation among flights and itineraries reveals some interesting methods of estimating observed origin-destination (OD) demand given several alternative configurations of observed data. This Appendix discusses with some formality the structure of demand allocation and indicates how observed demand can be deduced from only partial knowledge of the complete demand picture.

**M.2:** Use the notation $D_O(m)$ to indicate observed demand in market *m* over a given time period, nominally one week in AirVM. Since the application of the methodology developed in this discussion relates to the estimation of observed demand when only part of that demand can be directly observed, it is necessary to decompose $D_O(m)$ into two parts, that is, $D_O(m) = D_a(m) + D_d(m)$. In this equation, $D_a$ represents the demand met by sales through agency channels (and therefore visible to, say, a BSP) and $D_d$ stands for the demand met by sales through other channels, such as directly from the airlines via the internet. Travel on the itineraries that support a given market *m* is done, of course, on one or more flight legs. Define $\Gamma$ to be the set of all legs $\lambda$ in the airline network, and let *N* be the number of elements in $\Gamma$. An *itinerary* $\pi$ of legs from $\Gamma$ is an ordered set of legs $\pi = (\lambda_1, \lambda_2, \ldots \lambda_{N(\pi)})$ which has the following properties:

> **M.2.1:** The first leg in the ordered set, $\lambda_1$, starts at the market origin.
>
> **M.2.2:** The last leg of the set, $\lambda_{N(\pi)}$ ends its travel at the destination.
>
> **M.2.3:** There are no other legs in the itinerary which start at the origin or end at the destination.
>
> **M.2.4:** The arrival airport of each leg, except the last, is the departure airport of the next leg in the sequence.
>
> **M.2.5:** The arrival time of each leg, except the last, is before the departure time of the next leg in the sequence. (There must be a delay required between arrival times and departure times to allow for passenger transfer, but that is of no consequence to the argument.)

**M.3:** Now let $\Pi(m)$ be the collection of all itineraries that serve market *m* in a reference time period, and write $n(m)$ for the number of elements in $\Pi(m)$. Consider the following *leg allocation matrix* of size $n(m) \times N$:

$$\mathbf{A}(\Pi(m)) \equiv \mathbf{A}_m = \begin{pmatrix} I_{1,1}n_{1,1} & I_{1,2}n_{1,2} & \cdots & I_{1,N}n_{1,N} \\ I_{2,1}n_{2,1} & I_{2,2}n_{2,2} & \cdots & I_{2,N}n_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n(m),1}n_{n(m),1} & I_{n(m),2}n_{n(m),2} & \cdots & I_{n(m),N}n_{n(m),N} \end{pmatrix}. \quad \text{(M.1)}$$

Here $n_{i,j}$ is the number of passengers on the $i$-th itinerary in market $m$ who have seats on leg $j$. In other words, this matrix displays the allocation of passengers on each leg of each itinerary supporting the market. $I_{i,j}$ is an indicator function, defined as

$$I_{i,j} \equiv \begin{cases} 1 \text{ if itinerary } i \text{ uses leg } j \\ 0 \text{ otherwise} \end{cases}. \quad \text{(M.2)}$$

Now, observe that the value of $n_{k,l}$ is either a constant $n_k$ or 0 for all elements in row $k$. This is because a leg is either in itinerary $k$ or not. If not, $n_{k,l}$ must be zero, since the indicator function for that cell would have the value zero. If so, it must be the same value for all legs in the itinerary, since the each passenger selecting a given itinerary will appear on each leg of that itinerary. Obviously, almost all of the entries in every row $k$ must be zero. The description of each allocation can be easily extended to describe that portion which is sold through the agency channel and that portion sold through other channels, such as the Internet or CTOs (City Travel Offices), thus suggesting that $n_k = n_{d,k} + n_{a,k}$ for all $k$, where the suffixes $a$ and $d$ denote the agency and direct channels, respectively. Thus

$$\mathbf{A}_m = \begin{pmatrix} I_{1,1}(n_{a,1}+n_{d,1}) & I_{1,2}(n_{a,1}+n_{d,1}) & \cdots & I_{1,N}(n_{a,1}+n_{d,1}) \\ I_{2,1}(n_{a,2}+n_{d,2}) & I_{2,2}(n_{a,2}+n_{d,2}) & \cdots & I_{2,N}(n_{a,2}+n_{d,2}) \\ \vdots & \vdots & \ddots & \vdots \\ I_{n(m),1}(n_{a,n(m)}+n_{d,n(m)}) & I_{n(m),2}(n_{a,n(m)}+n_{d,n(m)}) & \cdots & I_{n(m),N}(n_{a,n(m)}+n_{d,n(m)}) \end{pmatrix}, \quad \text{(M.3)}$$

keeping in mind that both terms of each element are zero in the vast majority of columns.

**M.4:** Now, recall that the passenger choice model at the core of AirVM describes the probability of each itinerary in a market as a function of attributes of that itinerary by this mixed logit distribution:

$$p_m(k) = \int_{\tau \in T} \frac{e^{V(k|\tau)}}{\sum_{j=1}^{n(m)} e^{V(j|\tau)}} d\Theta(\tau) \quad \text{(M.4)}$$

where $p_m(k)$ is the probability of a pag choosing itinerary *k* in market *m* and $\Theta(\tau)$ is the distribution of departure/arrival times in the synthetic population. This probability is known from the underlying passenger choice model. Now assume that for some itinerary *k* in a market *m*, the value of $n_k$ is known and is equal to *n**. This would be the case, for example, if data from a specific carrier data were available. Then, since

$$p_m(k)D_O(m) = n_k, \ \forall\ k \tag{M.5}$$

it must be that

$$D_O(m) = \frac{n^*}{p_m(k)}, n^* = n_k. \tag{M.6}$$

Furthermore, since the agency-channel ticket count $n_a$ is known by BSP data sources,

$$n_i = n_{a,i} + n_{d,i} = p_m(i)D_O(m) \tag{M.7}$$

and

$$n_{i,d} = p_m(i)D_O(m) - n_{i,a} \tag{M.8}$$

for all values of *i* in the itinerary set for market *m*. Thus *the complete leg allocation matrix is known for a given market if the allocation to a single itinerary in the market is known, and the direct sales on each itinerary (and therefore for each carrier supporting the itinerary) is also known*. Of course, this then also provides an estimate of the carrier share in the market. That is,

$$\mathbf{A}_m = \begin{pmatrix} I_{1,1}p_m(1)D_O(m) & I_{1,2}p_m(1)D_O(m) & \cdots & I_{1,N}p_m(1)D_O(m) \\ I_{2,1}p_m(2)D_O(m) & I_{2,2}p_m(2)D_O(m) & \cdots & I_{2,N}p_m(2)D_O(m) \\ \vdots & \vdots & \ddots & \vdots \\ I_{n(m),1}p_m(n(m))D_O(m) & I_{n(m),2}p_m(n(m))D_O(m) & \cdots & I_{n(m),N}p_m(n(m))D_O(m) \end{pmatrix}. \tag{M.9}$$

**M.5:** The above formulation shows how to estimate the observed market demand from knowledge of the ticket sales total on a single itinerary in the market. It also shows how to estimate the fraction of the demand accounted for by direct channel distribution given that the agency and total demand are known. But there is a relationship between the leg allocation matrix for one market and the leg allocation matrices for other markets, and this relationship, can yield estimates of the observed demand in these other, related markets. Recall that a single leg can be a part of more than one itinerary, and thus will be listed in the itinerary set for more than one market. Consider any leg $\lambda \in \Gamma$ and let $\mu(\lambda)$ be the set of markets $m \in M$ such that $\lambda$ is part of at least one itinerary in *m*. Since there are a finite number of markets in $\mu(\lambda)$, they can be identified with arbitrary integer indices. Now a matrix can be created by "stacking" the respective leg allocation matrices one below the other for each $m \in \mu(m)$, so that the top *n(1)* rows of the matrix are identical to $A_1$, the next *n(2)* rows identical to $A_2$ and so on down to $N(\lambda)$, the number of markets in $\mu(\lambda)$. Thus

$$\mathcal{A}_{\mu(\lambda)} = \begin{pmatrix} I_{1,1,1}p_1(1)D_O(1) & I_{1,2,1}p_1(1)D_O(1) & \cdots & I_{1,N,1}p_1(1)D_O(1) \\ I_{2,1,1}p_1(2)D_O(1) & I_{2,2,1}p_1(2)D_O(1) & \cdots & I_{2,N,1}p_1(2)D_O(1) \\ \vdots & \vdots & \ddots & \vdots \\ I_{n(1),1,1}p_1(n(1))D_O(1) & I_{n(1),2,1}p_1p_1(n(1))D_O(1) & \cdots & I_{n(1),N,1}p_1(n(1))D_O(1) \\ I_{1,1,2}p_2(1)D_O(2) & I_{1,2,2}p_2(1)D_O(2) & \cdots & I_{1,N,2}p_2(1)D_O(2) \\ I_{2,1,2}p_2(2)D_O(2) & I_{2,2,2}p_2(2)D_O(2) & \ddots & I_{2,N,2}p_2(2)D_O(2) \\ \vdots & \vdots & \cdots & \vdots \\ I_{n(2),1,2}p_2(n(2))D_O(2) & I_{n(2),2,2}p_2(n(2))D_O(2) & \cdots & I_{n(2),N,N(\lambda)}p_1(n(2))D_O(2) \\ \vdots & \vdots & \vdots & \vdots \\ I_{1,1,N(\lambda)}p_{N(\lambda)}(1)D_O(N(\lambda)) & I_{1,2,N(\lambda)}p_{N(\lambda)}(1)D_O(N(\lambda)) & \cdots & I_{1,N,N(\lambda)}p_{N(\lambda)}(1)D_O(N(\lambda)) \\ I_{2,1,N(\lambda)}p_{N(\lambda)}(2)D_O(N(\lambda)) & I_{2,2,N(\lambda)}p_{N(\lambda)}(2)D_O(N(\lambda)) & \cdots & I_{2,N,N(\lambda)}p_{N(\lambda)}(2)D_O(N(\lambda)) \\ \vdots & \vdots & \ddots & \vdots \\ I_{n(N(\lambda)),1,N(\lambda)}p_{N(\lambda)}(n(N(\lambda)))D_O(N(\lambda)) & I_{n(N(\lambda)),2,N(\lambda)}p_{N(\lambda)}(n(N(\lambda)))D_O(N(\lambda)) & \cdots & I_{n(N(\lambda)),N,N(\lambda)}p_{N(\lambda)}(n(N(\lambda)))D_O(N(\lambda)) \end{pmatrix}. \tag{M.10}$$

Each column of this matrix represents the distribution of *loads* on a given leg in the network. That is, if $L(\lambda_i)$ is the load on leg $\lambda_i$, then

$$L(\lambda_i) = \sum_{j=1}^{N(\lambda_i)} \sum_{k=1}^{n(j)} (I_{i,j,k}\, p_j(k) D_O(j)). \tag{M.11}$$

Now consider the case where there a given leg carries traffic on only two markets, and further assume that by the argument given above, sufficient data has been collected so that the demand for travel on that market has been estimated. That is, the above equation is simply

$$\begin{aligned} L(\lambda_i) &= \sum_{k=1}^{n(1)} (I_{1,a,k}\, p_1(k) D_O(1)) + \sum_{k=1}^{n(2)} (I_{2,a,k}\, p_2(k) D_O(2)) \\ &= R + D_O(2) \sum_{k=1}^{n(2)} I_{2,a,k}\, p_2(k) \\ &\Rightarrow D_O(2) = \frac{L(\lambda_i) - R}{\sum_{k=1}^{n(2)} p_2(k)} \end{aligned} \tag{M.12}$$

Thus the demand in the second market, $D_0(2)$, can be computed if the load on the leg is known. From that, of course, all the respective agent and direct market quantities can be calculated (provided the agent channel proportion is known for the respective markets). A similar calculation can be performed for a leg serving an arbitrary number, say *s*, of itineraries if the demand of $s-1$ of the markets is known. And that is required is the addition of total load data for the leg.

**M.6:** Load data can be supplied by carriers or from other sources, such as airport departure/arrival traffic reports However, there is another, implied source. It stems from the structure of the demand realization over time. Each of the variables and expressions described in the analysis above assumed that time was not a factor in the realization of demand. But that is clearly not the case. Indeed, effective revenue management systems utilize the interval-before-departure as a important determinant in setting ticket prices, which are the most important attribute of an itinerary, reflected in the fact that the coefficient on fare in the passenger choice model exceeds all others by a considerable amount. In addition, over the booking period prior to departure, flight legs often reach capacity. From that point on in the ticketing time frame the itineraries that contain full flights are no longer available to support their associated markets, which in turn causes an adjustment in the probabilities of the remaining itineraries in those markets. At the present time it is assumed that the individual itineraries in a market comply with the independence of irrelevant alternatives (IIA) assumption. Under that assumption, the change in probabilities due to the elimination of a choice from the choice set is given by the following formulation. Let $p_m(i),\ i \in \Pi(m)$ be the probability of choosing itinerary *i* from the set of itineraries supporting market *m*. Now suppose itinerary *j* is no longer available because one of its constituent legs is full, and let $\Pi(m)-\{j\}$ denote the set of itineraries with *j* eliminated. The probability $p_{\Pi(m)-\{j\}}(i)$ of $i \in \Pi(m)-\{j\}$ is then given by

$$p_{\Pi(m)-\{j\}}(i) = \frac{p_m(i)}{\sum_{k \neq j} p_m(k)}. \qquad \text{(M.13)}$$

That is, the probability of the each remaining itinerary is the probability of the itinerary before the elimination divided by the sum of the probabilities of the remaining options. The time dependency in AirVM is measured in terms of days-before-departure. So a flight open at day *t* may be closed at $t-1$, where *t* is the number of days prior to the departure day of the leg. The equation above can be restated by writing $p_m(i) = p_m(i,t)$ and $p_{\Pi(m)-\{j\}} = p_m(i,t-1)$, so that the expression above becomes

$$p_m(i,t-1) = \frac{p_m(i,t)}{\sum_{k \neq j} p_m(k,t)}, \qquad \text{(M.14)}$$

thus describing a clear relationship between the probabilities before flight closure and after.

**M.7:** Now, let $n(i,t)$ be the number of tickets sold on itinerary *i* in the market *m* up to time period *t*. Then, if $D_0(m)$ is the observed demand in *m*, the following is true:

$$
\begin{aligned}
n(i,t-1)-n(i,t) &= D_O(m)\,p_m(i,t-1)-D_O\,p_m(i,t) \\
&= D_O(m)\left[\frac{p_m(i,t)}{\sum_{j\neq k} p_m(j,t)}-p_m(i,t)\right] \\
&= D_O(m)\,p_m(i,t)\left[\frac{1-\sum_{j\neq k} p_m(j,t)}{\sum_{j\neq k} p_m(j,t)}\right] \\
&= D_O(m)\,p_m(i,t)Q
\end{aligned}
\tag{M.15}
$$

and thus

$$
D_O(m)=\frac{n(i,t-1)-n(i,t)}{p_m(i,t)Q}. \tag{M.16}
$$

So, if the number of tickets sold on any flight in an itinerary set is observed, then this equation can be used to determine when flights are closed in the course of the ticketing sequence. If sales on the leg in question stop entirely, it is the one that is no longer available. If the relative number of tickets on flights changes, it is due to some other flight closing, and direct observation of the other flights, or inference using the probability relationship above, can determine which one is the culprit.

# Appendix N: The Random Number Generators Used in AirVM

N.1: The C# code below is that which produces the random numbers for the incidence distributions of AirVM. The code is produced by Microsoft Visual Studion 2008. This library is the property of Virtual M1nds, S. A. and may not be used without permission. The Mersenne Twister uniform random number generator is that developed by Troescheutz (2007).

```
/* VM1RandomNumberGenerators
* Copyright © Virtual M1nds 2008 - 2010.  All rights reserved.
*     This document may not be reproduced in whole or in part, in any
*     form whatsoever,without express written permission of Virtual
*     M1nds, SA, of Vevey, Switzerland.
* Author:  RAParker
* Module Name:  VM1RandomNumberGenerators
* Module Purpose:  Random number generators based on the Mersenne
* Twister algorithm for use in virtual markets.
*
*/

using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.IO;

namespace VM1RandomNumberGenerator
{
    /// <summary>
    /// The enum Truncated holds the values that are used describing
    /// the truncation needed for a normal distribution
    /// </summary>
    public enum Truncation
    {
        /// <summary>
        /// Enum values
        /// </summary>
        /// <summary>None</summary>
        None = 0,           // no truncation
        /// <summary>Negative (truncation greater than zero; negative values only.
        </summary>
        Negative = 1,
        /// <summary>Positive (truncation less than or equal to zero; positive values
        only.</summary>
        Positive = 2
    }

    /// <summary>
    /// class DistributionParameters holds necessary parameters for a
    /// probability distribution
    /// </summary>
    /// <remarks>
    /// mean = mean.
    /// stdev = standard deviation.
    /// thirdmoment and fourthmoment are reserved for later use.
    /// truncation holds the truncation parameters for the distribution.
    /// </remarks>
    public class DistributionParameters
    {
        /// <summary>mean:  the expectation</summary>
        public double mean;             // mean
        /// <summary>stdev:  standard deviation</summary>
        public double stdev;            // standard deviation
        /// <summary>thirdmoment:  third moment about the mean.
        Reserved for future use</summary>
        public double thirdmoment;
```

```
/// <summary>fourthmoment:  fourth moment about the mean.
Reserved for future use</summary>
public double fourthmoment;
/// <summary>trunc:  truncation enum</summary>
public Truncation trunc;        // truncation paramaters

/// <summary>
/// Null constructor
/// </summary>
public DistributionParameters()
{
    mean = 0;
    stdev = 0;
    thirdmoment = 0;
    fourthmoment = 0;
    trunc = Truncation.None;
}

/// <summary>
/// Constructor with external values supplied
/// </summary>
/// <param name="m">Mean</param>
/// <param name="s">Standard Deviation</param>
/// <param name="t">Third moment value</param>
/// <param name="f">Fourth moment value</param>
/// <param name="tr">Truncation parameters</param>
public DistributionParameters(double m, double s, double t,
    double f, Truncation tr)
{
    mean = m;
    stdev = s;
    thirdmoment = t;
    fourthmoment = f;
    trunc = tr;
}

/// <summary>
/// Constructor with only mean and standard deviation supplied
/// </summary>
/// <param name="m">Mean</param>
/// <param name="s">Standard Deviation</param>
public DistributionParameters(double m, double s)
{
    mean = m;
    stdev = s;
    thirdmoment = 0;
    fourthmoment = 0;
    trunc = Truncation.None;
}

/// <summary>
/// Constructor with only mean, standard deviation and
/// truncation supplied
/// </summary>
/// <param name="m">Mean</param>
/// <param name="s">Standard Deviation</param>
/// <param name="tr">Truncation parameters</param>
public DistributionParameters(double m, double s,
    Truncation tr)
{
    mean = m;
    stdev = s;
    thirdmoment = 0;
    fourthmoment = 0;
    trunc = tr;
}

/// <summary>
/// Loads a distribution parameter object from a binary stream
/// </summary>
/// <param name="br">The stream to read from</param>
public void Load(BinaryReader br)
{
    mean = br.ReadDouble();
    stdev = br.ReadDouble();
```

```
            thirdmoment = br.ReadDouble();
            fourthmoment = br.ReadDouble();
            trunc = (Truncation)br.ReadInt32();
        }

        /// <summary>
        /// Saves a distribution parameter object to a binary stream
        /// </summary>
        /// <param name="bw">The stream to save to</param>
        public void Save(BinaryWriter bw)
        {
            bw.Write(mean);
            bw.Write(stdev);
            bw.Write(thirdmoment);
            bw.Write(fourthmoment);
            bw.Write((int)trunc);
        }

        /// <summary>
        /// Writes the distribution parameter object to a text stream
        /// </summary>
        /// <param name="sw">The text stream to write to.</param>
        /// <param name="indent">The amount of indentation.</param>
        public void Write(StreamWriter sw, string indent)
        {
            sw.WriteLine(indent + "({0}, {1}, {2}, {3}, {4})", mean,
                stdev, thirdmoment, fourthmoment, (int)trunc);
        }
    }

    /// <summary>
    /// Uniformly distributed random number generator.
    /// </summary>
    /// <remarks>
    /// Just a shell around the Mersenne Generator.
    /// </remarks>
    public class UniformDistribution
    {
        private MersenneTwister RanGen;
        private double min;
        private double max;

        /// <summary>
        /// Uniformly distributed [0, 1) random number generator.
        /// </summary>
        public UniformDistribution()
        {
            RanGen = new MersenneTwister();
            min = 0;
            max = 1;
        }

        /// <summary>
        /// Uniformly distributed [minium, maximum) random number
        /// generator.
        /// </summary>
        public UniformDistribution(double minimum, double maximum)
        {
            RanGen = new MersenneTwister();
            min = minimum;
            max = maximum;
        }

        /// <summary>
        /// Returns the next double in the series.  Parameters max and
        /// min set by the constructors
        /// </summary>
        public double NextDouble()
        {
            if (min == 0 && max == 1) return RanGen.NextDouble();
            else return RanGen.NextDouble(min, max);
        }
    }

    /// <summary>
```

```
/// Finite Distribution: A class which manages the selection of
/// one category from a finite set of categories
/// each of which has an associated probability.
/// </summary>
public class FiniteDistribution
{
    private UniformDistribution uni;
    private double[] p;

    /// <summary>
    /// A default constructor
    /// </summary>
    public FiniteDistribution()
    {
        uni = new UniformDistribution();
        p = new double[0];
    }

    /// <summary>
    /// Constructor with uniform generator and category probabilities provided.
    /// Probabities are for each category, and not accumulated.
    /// The routine accumulates the probabilities
    /// </summary>
    /// <param name="un">A uniform [0,1) random number generator</param>
    /// <param name="probs">An array of doubles containing the
    /// probabilities of each of the finite set of categories</param>
    /// <remarks>
    /// If the supplied probabilities do not sum to one, then the
    /// probabiities in each category are adjusted so that they do
    /// sum to one.  This means that counts or conditional
    /// probability results can be supplied, and the algorithm will provide
    /// a valid result.
    /// </remarks>
    public FiniteDistribution(UniformDistribution un,
        double[] probs)
    {
        uni = un;

        double[] q = new double[probs.Count()];
        p = new double[probs.Count()];
        double t = 0;

        // create the pdf
        q[0]=probs[0];
        for (int i = 1; i < probs.Count(); i++)
        {
            q[i] = q[i - 1] + probs[i];  // create the
            // accumulative distribution function
        }
        t = q[probs.Count() - 1];

        for (int i = 0; i < probs.Count(); i++) p[i] = q[i] / t;
    }

    /// <summary>
    /// Produces the next integer value of the distribution
    /// </summary>
    /// <returns>A Poisson distributed integer</returns>
    public int NextInt()
    {
        double x = uni.NextDouble();
        for (int i = 0; i < p.Count(); i++) if (x < p[i])
            return (i + 1);
        return p.Count();       // to avoid the compiler warning
    }
}

/// <summary>
/// Normally distributed random number generator.  Uses Numerical Methods in C
/// algorithm (Press, et al., 1988).
/// </summary>
/// <remarks>
/// Two new normally distributed random numbers are generated every other call to
/// NextDouble.  Stores a precomputed normally distributed random number that will
/// be returned the next time NextDouble gets called.
```

```
/// </remarks>
public class NormalDistribution
{
    private double helper1;
    private bool helper2;
    private MersenneTwister RanGen;

    /// <summary>
    /// Initializes a new instance of the NormalDistribution class, using the
    /// Mersenne generator with no seed.
    /// </summary>
    public NormalDistribution()
    {
        RanGen = new MersenneTwister();
        UpdateHelpers();
    }

    /// <summary>
    /// Initializes a new instance of the NormalDistribution class, using the
    /// specified seed.
    /// </summary>
    public NormalDistribution(int seed)
    {
        RanGen = new MersenneTwister(seed);
        UpdateHelpers();
    }

    private void UpdateHelpers()
    // resets the helper variables
    {
        this.helper1 = 0.0;
        this.helper2 = false;
    }

    // Resets the normal distribution, so that it produces the same random number
    // sequence again.
    public bool Reset()
    {
        bool result = RanGen.Reset();
        if (result)
        {
            this.UpdateHelpers();
        }
        return result;
    }

    /// <summary>
    /// Returns a normally distributed random number with mean 0 and st dev 1
    /// </summary>
    public double NextDouble()
    {
        if (this.helper2)        // return the spare one, of have it
        {
            this.helper2 = false;
            return this.helper1;
        }
        else                     // compute new ones
        {
            this.helper2 = true;
            while (true)
            {
                double v1 = 2.0 * RanGen.NextDouble() - 1.0;
                double v2 = 2.0 * RanGen.NextDouble() - 1.0;
                double r = v1 * v1 + v2 * v2;
                if (r <= 1)
                {
                    double y = Math.Sqrt(-2.0 * Math.Log(r) / r);
                    this.helper1 = v2 * y;
                    return v1 * y;
                }
            }
        }
    }

    /// <summary>
```

```
    /// Returns a normally distributed random number with mean mu and st dev sigma.
    /// </summary>
    /// <param name="mu">Mean value</param>
    /// <param name="sigma">Standard deviation</param>
    public double NextDouble(double mu, double sigma)
    {
        double x = this.NextDouble();
        return x * sigma + mu;
    }

    /// <summary>
    /// Returns a truncated normal random number
    /// </summary>
    /// <returns>Returns a normal random number from the truncated distribution.
        </returns>
    /// <param name="dparams">The parameters of the distribution</param>
    public double NextDouble(DistributionParameters dparams)
    {
        double x = 0;
        while (true)
        {
            x = NextDouble(dparams.mean, dparams.stdev);
            if (dparams.trunc == Truncation.Positive)
            {
                if (x >= 0) break;
            }
            else if (dparams.trunc == Truncation.Negative)
            {
                if (x <= 0) break;
            }
            else if (dparams.trunc == Truncation.None) break;
        }
        return x;
    }

}

/// <summary>
/// Poisson distributed random number generator.
/// </summary>
/// <remarks>
/// Generates random integers n = 0, 1, 2, ... with a Poisson Distribution with
/// parameter lambda.
/// </remarks>
public class PoissonDistribution
{
    private MersenneTwister RanGen;     // Mersenne random number generator
    private double lm;                  // lambda value

    /// <summary>
    /// Poisson (lambda) integer random number generator.
    /// </summary>
    /// <remarks>
    /// Truncation, if necessary, occurs after returned random number using a
    /// rejection method.
    /// </remarks>
    /// <param name="lambda">lambda:  The distribution parameter value</param>
    public PoissonDistribution(double lambda)
    {
        RanGen = new MersenneTwister();
        lm = lambda;
    }

    /// <summary>
    /// Poisson (lambda) integer random number generator with an integer seed.
    /// </summary>
    /// <remarks>
    /// Truncation, if necessary, occurs after returned random number using a
    /// rejection method.
    /// </remarks>
    /// <param name="lambda">lambda:  The distribution parameter value</param>
    /// <param name="seedvalue">seedvalue:  Integer-valued random number seed.
    /// </param>
    public PoissonDistribution(double lambda, int seedvalue)
    {
```

```
            RanGen = new MersenneTwister(seedvalue);
            lm = lambda;
        }

        /// <summary>
        /// Poisson (lambda) integer random number generator using dparams
        /// </summary>
        /// <remarks>
        /// Truncation, if necessary, occurs after returned random number using a
        /// rejection method.
        /// </remarks>
        /// <param name="dparams">The distribution parameter object</param>
        public PoissonDistribution(DistributionParameters dparams)
        {
            RanGen = new MersenneTwister();
            lm = dparams.mean;
        }

        /// <summary>
        /// Poisson (lambda) integer random number generator using dparams with a fixed
        /// seed
        /// </summary>
        /// <remarks>
        /// Truncation occurs after returned random number using a rejection method.
        /// </remarks>
        /// <param name="dparams">The distribution parameter object</param>
        /// <param name="seedvalue">The value of the random number seed.</param>
        public PoissonDistribution(DistributionParameters dparams, int seedvalue)
        {
            RanGen = new MersenneTwister(seedvalue);
            lm = dparams.mean;
        }

        /// <summary>
        /// Returns Poisson (lambda) integer random numbers.
        /// </summary>
        public int NextInt()
        {
            double r = RanGen.NextDouble();
            double p = Math.Exp(-lm);
            double tvalue;
            tvalue = p;
            if (r <= tvalue) return 0;
            for (int i = 1; i <= int.MaxValue; i++)
            {
                p = p * (lm / (double)i);
                tvalue += p;
                if (r <= tvalue) return i;
            }
            return int.MaxValue;
        }

    }

    /// <summary>
    /// ExternalPoisson class:  Generates Poisson random numbers using external uniform
    /// generator and user supplied lambda value.
    /// </summary>
    public class ExternalPoisson
    {
        private UniformDistribution uni;
        /// <summary>
        /// Poisson distribution using and externally supplied random number generator
        /// </summary>
        /// <param name="u">A pre-defined uniform [0,1) random number generator</param>
        public ExternalPoisson(UniformDistribution u)
        {
            uni = u;
        }

        /// <summary>
        /// Return Poisson(lambda) with an externally supplied uniform random number
        /// generator </summary>
        /// <param name="lambda">Externally supplied value for lambda</param>
        /// <returns>Interger value</returns>
```

```
        public int NextInt(double lambda)
        {
            double r = uni.NextDouble();
            double p = Math.Exp(-lambda);
            double tvalue;
            tvalue = p;
            if (r <= tvalue) return 0;
            for (int i = 1; i <= int.MaxValue; i++)
            {
                p = p * (lambda / (double)i);
                tvalue += p;
                if (r <= tvalue) return i;
            }
            return int.MaxValue;
        }

        /// <summary>
        /// Return Poisson(lambda, truncvalue) with an externally supplied uniform
        /// random number generator </summary>
        /// <param name="lambda">Externally supplied value for lambda</param>
        /// <param name="truncvalue">Truncation value:  values less than or equal this
        /// are truncated.</param>
        /// <returns>Interger value</returns>
        public int NextInt(double lambda, int truncvalue)
        {
            double r = uni.NextDouble();
            double p = Math.Exp(-lambda);
            double F = 0;
            for (int i = 0; i <= truncvalue; i++)
            {
                if (i > 0) p *= lambda / i;
                F += p;
            }
            double C = 1 - F;
            double FT = p / C;
            for (int i = truncvalue + 1; i < int.MaxValue; i++)
            {
                p *= lambda / i;
                FT += p / C;
                if (r < FT) return i;
            }
            return int.MaxValue;
        }
    }

    /// <summary>
    /// EDF:  The empirical distribution function class.
    /// An edf is supplied as an ordered array of points.
    /// To generate a random number, supply a random number generator, the value of x
    /// corresponding to the inverse of the edf at the random number y is returned.
    /// The routine does a linear interpolation between the user supplied minimum, each
    /// point in the edf, and a user supplied maximum, based on the assumptions that x =
    /// 0 when y = 0 and x = max when y = 1.</summary>
    public class EDF
    {
        private UniformDistribution uni;
        private double[] pdf;
        private double n;

        /// <summary>
        /// Default constructor.
        /// </summary>
        public EDF()
        {
            uni = new UniformDistribution();
            pdf = new double[0];
        }

        /// <summary>
        /// Constructor which builds the random number generator based on the supplied
        /// empirical distribution function values.
        /// </summary>
        /// <param name="un">A user supplied uniform [0,1) random number generator
        /// </param>
        /// <param name="x">The array of n values of the array x where the edf jumps of
```

```
        /// 1/n occur.</param>
        /// <param name="min">The minimum value of the edf.  Must be less than x(1)
        /// </param>
        /// <param name="max">The maximum value of the edf.  Must be greater than x(n)
        /// </param>
        /// <remarks>
        /// The generator produces a random number y between 0 and 1, and then
        /// interpolates between the two values of the x array in which y occurs to
        /// determine the value of x to return.  The interpolation is linear.
        /// </remarks>
        public EDF(UniformDistribution un, double[] x, double min, double max)
        {
            uni = un;
            List<double> spdf = new List<double>();
            pdf = new double[x.Count() + 2]; // interval boundaries.  Number of points
                                             // plus min and max
            n = pdf.Count() - 1;            // number of intervals

            // sort the points
            spdf.Add(min);
            for (int i = 0; i < x.Count(); i++) spdf.Add(x[i]);
            spdf.Add(max);
            spdf.Sort();

            // create the pdf array.
            for (int i = 0; i < spdf.Count(); i++) pdf[i] = spdf.ElementAt(i);
        }

        /// <summary>
        /// Generates a random number according to this edf
        /// </summary>
        /// <returns>A double value between min and max</returns>
        public double NextDouble()
        {
            double y = uni.NextDouble();
            double xreturn = 0;

            int k = (int)(y * n);       // which interval

            xreturn = pdf[k] + (y - ((double)k / n)) * (n * (pdf[k + 1] - pdf[k]));
            return xreturn;
        }

        /// <summary>
        /// Inverse:  Returns the value of the EDF for the supplied value of y
        /// <param name="y">A double in the interval [0,1).</param>
        /// </summary>
        /// <returns>A double value between min and max</returns>
        public double Inverse(double y)
        {
            double xreturn = 0;

            int k = (int)(y * n);       // which interval

            xreturn = pdf[k] + (y - ((double)k / n)) * (n * (pdf[k + 1] - pdf[k]));
            return xreturn;
        }
    }
}
```

# Bibliography

Abed, S., A, Ba-Fail and M. Jasimuddin (2001), "An Econometric Analysis of International Air Travel Demand in Saudi Arabia," *J. of Air Transport Management,* 7:143-148.

Adler, T., C. Falzarano and G. Spitz (2005), *Modeling Trade-offs in Air Itinerary Choices*, 84th Annual Meeting of the Transportation Research Board, Washington D. C.

Alamdari, F. (2007), Vice President, Future Market Analysis, Boeing Commercial Airplanes, former Chair, Dept. of Aeronautics, Cranfield University, Private communication.

Algers, S. and M. Beser (2001), "Modeling Choice of Flight and Booking Class: A Study Using Stated Preference and Revealed Preference Data," *International J. of Service Technology and Mgmt,* 2:1-2.

Anas, A. (1983), Discrete Choice Theory, Information Theory and the Multinomial Logit and Gravity Models, *Transportation Research B*, 17, 13-23.

Anderson, J. (1979), "A Theoretical Foundation for the Gravity Model," *American Economic Review,* 69: 106-116.

Anderson, S., A. de Palma and J. Thisse (1992), *Discrete Choice Theory of Product Differentiation,* MIT Press.

Andersson, T. and C. Andersson (2009), "Solving House Allocation Problems with Risk-Averse Agents," *Computational Economics,* 33:389-401.

Angrist, J. and A. Krueger (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *J. of Economic Perspectives*, 15: 69-85.

ARC (2004, 2005, 2009), Private data communications, Airline Reporting Corporation, Arlington, VA.

Ariely, D. (2008), *Predictably Irrational: The Hidden Forces That Shape Our Decisions*, HarperCollins.

Arrow, K (1963), *Social Choice and Individual Values*, 2nd Ed., Yale University Press.

Arthur, W. (2006), "Out-of-equilibrium Economics and Agent Based Modeling," in Judd, K and L. Tesfastion, eds. *Handbook of Computational Economics, Vol II: Agent-Based Computational Economics,* North Holland.

Auffhammer, M. and R. Carson (2008), "Forecasting the Path of China's $CO_2$ Emissions Using Province-Level Information," *J. of Environmental Economics and Management*, doi: 10.1016 /j.jeem.2007.10.002.

Axelrod, R. (1997), *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press.

Axelrod, R. and W. Hamilton (1981), "The Evolution of Cooperation," *Science* 211: 1390-1396.

Bahdra, D. (2003), "Demand for Air Travel in the United States: Bottom Up Econometric Estimation and Implications for Forecasts by Origin and Destination Pairs," *J. of Air Transportation*, 8: 19-56.

Bal, M. (2009), *Narratology: Introduction to the Theory of Narrative*, 3rd Ed, University of Toronto.

Banerjee, S. and J. Murphy (2006), "A Simplified Test for Rationality of Two Commodity Choice," *Experimental Economics,* 9:67-75.

Barabasi, A. (2002), *Linked: The New Science of Networks*, Perseus.

Barndorff-Nielson, O., D. Cox and C. Kluppelberg (2001), *Complex Stochastic Systems*, Chapman & Hall/CRC.Barzun, J. (2001), *From Dawn to Decadence: 500 Years of Western Cultural Life, 1500 to the Present*, San Val.

Baseler, R. (2002), "Airline Fleet Revenue Management—Design and Implementation," in D. Jenkens, ed, *Handbook of Airline Economics*, Aviation Week.

Bass, F. (1969), "A New-product Growth Model for Consumer Durables," *Management Science*, 15:215-227.

Bates, J. (2000), "History of Demand Modeling," in D.A. Hensher and K.J. Button eds., *Handbook of Transport Modeling*, Pergamon.

Belnap, N. and T. Steel (1976), *The Logic of Questions and Answers*, Yale University.

Belobaba, P. (2008), *An Update on the State of Revenue Management in the Airline Industry*, AGIFORS Symposium, Montreal.

Belobaba, P. (1989), "Application of a Probabilistic Decision Model to Airline Seat Inventory Control," *Operations Research*, 37, pp 183-197.

Belobaba, P. (1987), *Air Travel Demand and Airline Seat Inventory Management*, PhD Thesis, Flight Transportation Laboratory, MIT.

Ben-Akiva, M. (1973), *The Structure of Travel Demand Models*, PhD Thesis, MIT.

Ben-Akiva, M. and M. Bierlaire (1999), "Discrete Choice Methods and Their Applications to Short Term Travel Decisions," in *Handbook of Transportation Science*, R. Hall, ed., Kluwer.

Ben-Akiva, M and Lerman, S. (1985), *Discrete Choice Theory*, MIT Press.

Berger, B. and M. Fuchs (2005), "Dynamic Pricing – a Future Airline Business Model," *J. of Revenue and Pricing Management*, 4:1, pp 39-53.

Bergstrand, J.H. (1985), "The Gravity Model in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics*, 474-481.

Bernard, R. (1999), *Using Adaptive Agent-Based Simulation Models to Assist Planners in Policy Development: The Case of Rent Control*, Santa Fe Institute.

Berne, E. (1964), *Games People Play*, Grove Press.

Berry, S., M. Carnall, and S. Spiller (2006), "Airline Hubbing, Cost and Demand," in *Advances in Airline Economics: Competition Policy and Antitrust*, D. Lee, ed. Elsevier.

Bhat, C. (1987), "Covariance Heterogeneity in Nested Logit Models: Economic Structure and Application to Intercity Travel," *Transportation Research B*, 31:11-21.

Bhat, C, V. Warburg and T. Adler (2006), *Modeling Demographic and Unobserved Heterogeneity in Air Passengers' Sensitivity to Service Attributes in Itinerary Choice*, 85th Annual Meeting of the Transportation Research Board, Washington, D. C.

Blackmore, S. (1999), *The Meme Machine*, Oxford.

Boeing Commercial Airplanes (2008), *Current Market Outlook*, The Boeing Company.

Bonabeau, E. (2002), “Agent-based modeling:  Methods and Techniques for Simulating Human Systems,”  *Proceedings of the National Academy of Sciences of the United States of America*, 99 (suppl 3), pp 7280-7287.

Bone, J, J. Hey and J. Suckling (2009), “Do People Plan?” *J. of Experimental Economics*, 12: 12–25.

Bringsjord, S and D. Ferrucci (2000), *Artificial Intelligence and Literary Creativity:  Inside the Mind of BRUTUS, a Storytelling Machine*, Lawrence Erlbaum.

Brodie, R. (1996) *Virus of the Mind:  The New Science of the Meme*, Integral Press

Brons, M., E. Pels, P. Nijkamp, and P. Rietveld (2002), “Price Elasticities of Demand for Passenger Air Travel: A Meta-Analysis,” *J.l of Air Transport Management*, 8, 165-175.

Brown, D. (1991), *Human Universals*, McGraw-Hill.

Brown, S.L. and S. Watkins (1968), “The Demand for Air Travel: A Regression Study of Time Series and Cross-Sectional Data in U.S. Domestic Markets,” *Highway Research Record*, 213: 21-34.

Brueckner, J., N. Dyer and P. Spiller (1992), “Fare Determination in Airline Hub and Spoke Networks,” *Rand J. of Economics*, 23: 309-333.

Brumelle, S. and J. McGill (1993), “Airline Seat Allocation with Multiple Nested Fare Classes,” *Operations Research*, 41, pp 127-137.

Buchanan, M. (2002), *Small Worlds and the Groundbreaking Theory of Networks*, W. W. Norton and Co.

Bunch, D. (2003), *Behavioral Models and Estimates for Leisure-Passenger Value of Travel-Time Savings in Long-Haul Air Travel Markets Using Stated Choice Experiments*, Unpublished manuscript.

Burke, K. (1957), *Philosophy of Literary Form*, (rev. ed.), Vantage Books.

Button, K. (2003), “Does the Theory of the ‘Core’ Explain Why Airlines Fail to Cover their Long-run Costs of Capital?” *J. of Air Transportation Management*, 9, pp 5-14.

Cameron, A.C. and P.K. Trivedi (2005), *Microeconometrics: Methods and Applications*, Cambridge Economics Press.

Card, S., J. Mackinlay and B. Shneiderman (1999), *Information Visualization:  Using Vision to Think*, Morgan-Kaufmann.

Caroll, J. (1956), *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf.* MIT Press.

Carrington, P., J. Scott and S. Wasserman, eds, (2005), *Models and Methods in Social Network Analysis*, Cambridge University Press.

Carson, R. Parker, R., B. Conaway, C. Mohn and T. Cenesizoglu, (2010), *Modeling the Demand for United States Domestic Commercial Air Travel*,  in preparation.

Chalmers, A. (1999), *What is This Thing Called Science?* Hackett Publishing.

Chang, C., and H. Keisler (1973), *Model Theory*, Elsevier North Holland.

Chu, C (1989), “A Paired Combinatorial Logit Model for Travel Demand Analysis,”  *Proceedings of the 5th World Congress on Transportation Research*, 4:295-309.

Cialdini, R. (2001), *Influence:, Science and Practice*, 4th Ed., Allyn & Bacon.

Clark, A. (1997), *Being There: Putting Brain, Body and World Together Again*, MIT Press.

Cline, R., T. Ruhl, G. Gosling and D. Gillen, (1998), "Air Transportation Demand Forecasts in Emerging Market Economies: A Case Study of the Kyrgyz Republic in the Former Soviet Union," *J. of Airport Management* 4:11-23.

Coldren, G. and F. Koppelman (2005), *Modeling the Underlying Competitive Dynamic Among Air-travel Itineraries with Ordered Generalized Extreme Value Models*, 84th Annual Meeting of the Transportation Research Board, Washington D. C.

Coldren, G., F. Koppelman, K. Kasturirangan, and A. Mukherjee (2003), "Modeling Aggregate Air-travel Itinerary Shares: Logit Model Development at a Major U. S. Airline," *J. of Air Transport Management*, 9, pp 361-369.

Conway, J (1970), "Mathematical Games: The Game of Life," Martin Gardner, ed. *Scientific American*, October.

Coon, C. (2007), "*On Heroism*," The Humanist, February/March 2007.

Conte, R. and Paolucci, M. (2001), "Intelligent Social Learning," *J. of Artificial Societies and Social Simulation*, 4:1. http://jasss.soc.surrey.ac.uk/JASSS/4/1/3.html.

Cosmides, L and J. Tooby (1987), "From Evolution to Behavior? Evolutionary Psychology as the Missing Link", in the *Latest on the Best: Essays on Evolution and Optimality*, J. Dupre ed. MIT Press, pp 277-306.

Coy, R. (2003), Continental Airlines, Private data communication.

Cox, D and D. Oakes (1984), *Analysis of Survival Data*, Chapman and Hall.

Economics Software, Inc, (2002), *LIMDEP Version 8.0 Reference Guide*, Economic Software, Inc.

Danto, A. (1985), *Narrative and Knowledge*, Columbia University Press.

Darwin, C. (1859), *On the Origin of Species*, John Murray, London.

von Davier, M. and C. Carstensen, (2007), *Multivariate and Mixture Distribution Rasch Models*, Springer.

Dawkins, R. (1976), *The Selfish Gene*, Oxford University Press.

Dawkins, R. (1982), *The Extended Prototype: The Long Reach of the Gene*, Oxford University Press.

Dawes, R. (1979), "The Robust Beauty of Improper Linear Methods in Decision Making," *American Psychology*, 34: 571-582.

Deguchi, H. (2004), *Economics as an Agent-Based Complex System*, Springer-Verlag.

Denning, S. (2001), "The Springboard: How Storytelling Ignites Action in Knowledge-era Organizations," *J. of Organizational Change Management*, 14:6 pp 609-614.

*Dennett, D. (1991), Consciousness Explained, Little, Brown & Co.*

Dorogovtsev, S and Mendes, J. (2003), *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press.

Duckworth, R. (2009), "The Role of Behavior in Evolution: A Search for Mechanism," Evolutionary Ecology, 23: 4: 513-531.

Eagly, A., R. Ashmore, R. Makhijani and L. Longo (1991), "What is Beautiful is Good, But …: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype," *Psychological Bulletin,* 110: 109-128.

Economics Software, Inc. (2002), *LIMDEP Version 8.0 Reference Guide,* Economic Software, Inc.

Efron, B. and R. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall.

Engels, F. (1964), *Dialectics of Nature*, Lawrence and Wishart.

Epstein, J. (2006), *Generative Social Science: Studies in Agent-Based Computational Modeling*, Princeton University Press.

Epstein, J. and Axtell, R. (1996), *Growing Artificial Societies: Social Sciences from the Bottom Up*, MIT Press.

Epstein, J., and R. Hammond, (2001), *Non-Explanatory Equilibria: An Extremely Simple Game With (Mostly) Unattainable Fixed Points*. Santa Fe Institute.

Erdos, P. and A. Renyi (1960), "On the Evolution of Random Graphs," reproduced in Newman, M. A. Barabasi and D. Watts, ed (2006), *The Structure and Dynamics of Networks*, Princeton University Press.

Espasa, A. E. Senra, and R. Albacete (2002), "Forecasting inflation in the European Monetary Union: A disaggregated approach by countries and by sectors," *European Journal of Finance* 8:402-421.

Evans, W.N., L.M. Froeb and G.J. Werden (1993), Endogeneity in the Concentration—Price Relationship: Causes, Consequences and Cures," *Journal of Industrial Economics*, 41: 431-438.

Ferguson, T. (1967), *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press.

Festinger, L., H. Rieckin, and S. Schachter (1964), *Where Prophecy Fails: A Social and Psychological Study*, HarperCollins.

Feynman, R., R Leighton and M. Sands (1965), *The Feynman Lectures on Physics*, Addison-Wesley.

Fisher, G. and I. Molenaar, (1995), *Rasch Models: Foundations, Recent Developments, and Applications*, Springer-Verlag.

Fisher, W. (1987), *Human Communication as Narrative: Toward a Philosophy of Reason, Value and Action,* University of South Carolina Press.

Forrester, J. (1975), *Collected Papers of Jay W. Forrester*, Pegasus.

Forrester, J. (1968), *Principals of Systems*, 2nd ed, Pegasus.

Forrester Research, Inc. (2007), Consumer Technographics, http://www.forrester.com.

Fox, C. and T. Bond (2007), *Applying the Rasch Model : Fundamental Measurement in the Human Sciences*, 2nd ed, Lawrence Erlbaum Associates.

Freankel, A (1953), *Abstract Set Theory: Studies in Logic and the Foundations of Mathematics*, North Holland.

Gannett, G. (1972/1980), *Narrative Discourse: An Essay on Method*, trans. J. Lewin, Cornell University Press.

Garcia, R. (2005), "Uses of Agent-based Modeling in Innovation/New Product Development Research,", *J. Product Innovation Management*, 22, pp 380-398.

Garrow, L. (2010), *Discrete Choice Modelling and Air Travel Demand: Theory and Applications*, Ashgate.

Garrow, L. (2004), *Predicting Air Travelers' No-show and Standby Behavior Using Passenger and Directional Itinerary Information*, PhD Dissertation, Northwestern University.

Garrow, L, S. Jones and R. Parker (2007), "How Much Airline Customers Are Willing to Pay: An Analysis of Price Sensitivity in On-line Distribution Channels," *J. of Revenue Management* 5:4 271-290.

Garrow, L. and R. Parker (2005), *Market Elasticity: How Much Are Airline Customers' Willing to Pay to Fly?* Air Transport Research Society World Congress, Rio de Janeiro.

Gianturco, E., trans. (1965), *On the Study Methods of our Times, Giambattista Vico*, Bobbs-Merrill, p. 22.

Gigerenzer, G. (2000), *Adaptive Thinking: Rationality in the Real World*, Oxford University Press.

Gigerenzer. G. and R. Selten (2001), Eds., *Bounded Rationality: The Adaptive Toolbox*, The MIT Press.

Gilbert, N. and K. Troitzsch (1999), *Simulation for the Social Scientist*, Open University Press.

Gillen, D., W. Morrison, and C. Stewart (2004), *Air Travel Demand Elasticities: Concepts, Issues and Measurement*, Final Report to the Department of Finance, Government of Canada.

Gillman, L. (1992), "The Car and the Goats," *American Mathematical Monthly*, 99:1 3-7.

Gladwell, M. (2005), *Blink*, Little Brown and Company.

Gladwell, M. (2000). *The Tipping Point*, Abacus.

Gödel, K. (1938), "The Consistency of the Axiom of Choice and the Generalized Continuum Hypothesis," *Proceedings of the National Academy of Sciences*, 24.

Goldberg, D. (1989), *Genetic Algorithms*, Addison-Wesley.

Goldenberg, J., B. Libai and E. Muller (2004), "Complex, Yet Simple: Cellular Automata as an Enabling Technology in Marketing Strategy Research," in C. Moorman and Lehmann Eds, *Assessing Marketing Strategy Performance*, Marketing Science Institute, pp 229-246.

Gould, S. (2002), *The Structure of Evolutionary Theory*, Harvard University Press.

Greene, W.H. (2003), *Econometric Analysis*, 5$^{th}$ ed., Prentice Hall.

Greiner, A and P. Flaschel (2009), "Economic Policy in a Growth Model with Human Capital, Heterogenous Agents and Unemployment," *Computational Economics*, 33:175–192.

Grenander, U. and M. Rosenblatt (1952), *Statistical Analysis of Stationary Time Series*, John Wiley and Sons.

Grimm, V. and S. Railsback (2005), *Individual-based Modeling and Ecology*, Princeton University Press.

Haight, F. (1967), *The Handbook of the Poisson Distribution*, John Wiley and Sons.

Hahn, J. and J. Hausman (2003), "Weak Instruments: Diagnosis and Cures in Empirical Economics," *American Economic Review*, 93, 118-125.

Hales, D., B. Edmonds, E. Norling and J. Rouchier, eds (2003), *Multi-Agent-Based Simulation III: 4th International Workshop*, Springer-Verlag.

Halpern, J. (2003), *Reasoning about Uncertainty*, MIT Press.

Hamilton, W. (1964a), "The Genetical Evolution of Social Behavior: I", *J. of Theoretical Biology,* 7:1-16.

Hamilton, W. (1964b), "The Genetical Evolution of Social Behavior: II," *J. of Theoretical Biology*, 7:17-52.

Hamilton, W. (1963), "The Evolution of Altruistic Behavior," *American Naturalist*, 97: 354-356.

Harris, R. (1996), *Information Graphics: A Comprehensive Illustrated Reference*, Management Graphics, Atlanta.

Hayek, F. (1988), *The Fatal Conceit: The Errors of Socialism*, Routledge.

Hayek, F. (1978), *New Studies in Philosophy, Politics, Economics and the History of Ideas*, Routledge and Kegan Paul.

Hayek, F. (1967), *Studies in Philosophy, Politics and Economics*, Routledge and Kegan Paul.

Haykin, S. (1999), *Neural Networks: A Comprehensive Foundation*, Prentice-Hall.

Haupt, R. and S. Haupt (2004), *Practical Genetic Algorithms*, Wiley-Interscience.

van der Heijden, K. (1996), *Scenarios: The Art of Strategic Conversation*, John Wiley and Sons.

*Heisenberg, W. (1927), "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik", Zeitschrift für Physik 43:172–198*, trans: J. Wheeler and H. Zurek, *Quantum Theory and Measurement*, Princeton University Press, 1983, pp. 62–84.

Hensher, D., ed (2001), *Travel Behavior Research: The Leading Edge*, Pergamom.

Hertz, J., K. Andrews and R. Palmer (1991), *Introduction to the Theory of Neural Computing*, Addison-Wesley.

Hess, S. (2005), *Advanced Discrete Choice Models with Applications to Transport Demand*, PhD Dissertation, University of London.

Hevern, V. (2004), "Ong, Interiority, and the Narrative Turn in Psychology," *Explorations in Media Ecology,* 3(1), 53-60.

Hewitt, E. and K. Stromberg (1965), *Real and Abstract Analysis*, Springer-Verlag.

Hocking, J. and G. Young (1961), *Topology*, Addison-Wesley.

Hodel, R. (1995), *An Introduction to Mathematical Logic*, PWS Publishing.

Hodgson, G. (1993), *Economics and Evolution: Bringing Life Back into Economics*, University of Michigan Press.

Holland, J. (1992), *Adaptation in Natural and Artificial Systems*, MIT Press.

Holland, J. (1995), *Hidden Order: How Adaptation Builds Complexity*, Basic Books.

Holland, J. (1998), *Emergence: From Chaos to Order*, Basic Books.

Hughes, G. and M. Cresswell (1968), *An Introduction to Modal Logic*, Harper and Row.

Humphreys, P. (2004), *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*, Oxford University Press.

Hunt, S. (2000), *A General Theory of Competition: Resources, Competences, Productivity and Economic Growth,* Sage.

Hunt, S., and S. Vitell, (1993), "The General Theory of Marketing Ethics: A Retrospective and Revision" in N. Smith, and A. Quelch, eds, *Ethics is Marketing*, Richard D. Erwin.

Hutchins, J. (1999), "Retrospect and Prospects in Computer-based Translation," *Proceedings of the Machine Translation Summit VII,* Singapore, pp. 30-44.

IATA PaxIS (2007), http://www.iata.org/ps/intelligence_statistics/paxis/

Illiescu, D., L. Garrow and R. Parker, (2008), "A Hazard Model of U. S. Airline Passenger Refund and Exchange Behavior," *Transportation Research Board, Part B*., 42B:3, 229-242.

Innovata, LLC. (2009), http://www.innovatallc.com/

Isard, W. (1960), *Methods of Regional Analysis,* MIT Press.

Jensen, F. (2001), *Bayesian Networks and Decision Graphs*, Springer.

Jackson, D. (1941), *Fourier Analysis and Orthogonal Polynomials*, Casus Mathematical Monographs, #6, Mathematical Association of America.

Johnson, S. (2001), *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*, Scribner.

Johnson, N., S. Kotz and N. Balakrishnan, (1995), *Continuous Univariate Distributions, Vol 2,* 2nd ed, John Wiley and Sons.

Judd, K and L. Tesfastion, eds. (2006), *Handbook of Computational Economics, Vol II: Agent-Based Computational Economics,* North Holland.

Julius, D. (2002), Member, Monetary Policy Board, Bank of England, and Chair,The Royal Institute of Foreign Affairs, Chatham House, London, Private communication.

Jung, J. and E. Fuji (1976), "Price Elasticity of Demand for Air Travel: Some New Evidence," *J. of Transport Economics and Policy*, 10: 257-262.

Katz, E. (2001), "Bias in Conditional and Unconditional Fixed Effects Logit Estimation," *Political Methodology*, 9: 379-384.

Keane, P (2006), University of Technology Sydney, Private communication.

Kennedy, J. and R. Eberhart (2001), *Swarm Intelligence*, Academic Press.

Kernighan, B. and P. Plauger (1976), *Software Tools*, Addison-Wesley.

Kleene, S (1971), *Introduction to Metamathematics*, American Elsevier Publishing.

Kohler, T. and G. Gumerman (2000), *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes,* Oxford University Press.

Kolmogorov, A. (1931), "Uber die analytischen Methoden in der Wahrscheinlichkeitsrechnung," *Mathematiche Annalen*, 104:415-458.

Kornai, A. (2007), *Mathematical Linguistics*, Springer.

Koppelman F., G. Coldren and R. Parker, (2008), "Schedule Delay Impacts on Air-travel Itinerary Demand*," Transportation Research Board, Part B*, 42B:3, 263-273.

Koppelman, F. and C. Wen (1998), "Alternative Nested Logit Models: Structure, Properties and Estimation," *Transportation Research, Part B*, 32:358-372.

Kruger, J. (1996), "Personal Beliefs and Cultural Stereotypes About Racial Characteristics, *J. of Personality and Social Psychology*, 71: 536-48.

Kuhn, T. (1996), *The Structure of Scientific Revolutions*, University of Chicago

van den Kuilen, G. (2009), "Subjective Probability Weighting and the Discovered Preference Hypothesis," *Theory and Decision*, 67:1-22.

Kunda, Z. (1999), *Social Cognition*, MIT Press.

Levine, M. (1987), "Airline Competition in Deregulated Markets: Theory, Firm Strategy, and Public Policy," *Yale J. of Regulation*, 4: 393-494.

Levine, L. (1983), "Mine: Self-Definition in Two-year-old Boys," *Developmental Psychology*, 19: 544-549.

Li, M. and T. Oum (2002), "A Note on the Single Leg, Multifare Seat Allocation Problem," *Transportation Science*, 36, pp 252-265.

van der Linden, W. and R. Hambleton, eds. (1997), *Handbook of Modern Item Response Theory,* Springer.

Littlewood, K. (1972), *Forecasting and Control of Passengers Bookings*, 12th AGIFORS Symposium Proceedings, Nathanya, Israel.

Louviere, J. (2001), "What if Consumer Experiments Impact Variances as well as Means? Response Variability as a Behavioral Phenomenon", *J. of Consumer Research*, 18: 506-511.

Louviere, J., R. Carson and D. Bunch (2002), *The Value of Time Analysis for High Speed Airplanes*, Boeing White Paper, unpublished.

Louviere, J., P. Hensher and J. Swait (2000), *Stated Choice Methods,* Cambridge University Press.

Lude, M and L. Tesfatsion, (2006), "Agent-based computational economics: A constructive approach to economic theory," in Judd, K. and Tesfatsion, L, eds., *Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics*, North Holland.

Luna, F., ed. (2000), *Economic Simulations in Swarm: Agent-based Modeling and Object Oriented Programming*, Kluwer Academic.

Lupia, A., M. McCubbins and S. Popkin, eds (2000), *Elements of Reason: Cognition, Choice and the Bounds of Rationality*, Cambridge.

Lusch, R. and Tay, N. (2004), "Agent-Based Modeling: Gaining Insight into Firm and Industry Performance," in C. Moorman, and Lehmann, eds, *Assessing Marketing Strategy Performance*, Marketing Science Institute, pp 213-228.

MacDonald, M. (2008), *Pro WPF in C# 2008: Windows Presentation Foundation with .NET 3.5*, 2nd ed. Apress.

Matsumoto, M. (1998), "Mersenne Twister: a 623-dimensionally Equidistributed Uniform Pseudo-Random Number Generator," *ACM Transactions on Modeling and Computer Simulation* 8: 3.

Marshall, A. (1949/1890) , *The Principles of Economics*, 8th ed. Macmillan.

Marx, K. (1973), *Grundrisse: Foundations of the Critique of Political Economy*, M. Nicolaus, trans, Pelican.

Masters. T. (1995), *Advanced Algorithms for Neural Networks: A C++ Sourcebook*, John Wiley and Sons.

Mátyás, L. (1998), "Proper Specification of the Gravity Model," *World Economy*, 20: 363-368.

Maynard Smith, J. (1964), Group Selection and Kin Selection," *Nature*, 201:1145-1146.

Maynard Smith, J. (1982), *Evolution and the Theory of Games,* Cambridge University Press.

Mayr, E. (2001), *What Evolution Is,* Basic Books.

Mayr, E. (2004), *What Makes Biology Unique?* Cambridge University Press.

McFadden, D. (1999), "Computing Willingness-to-Pay in Random Utility Models," in J. Moore, R. Reizman and J. Melvin, eds, *Trade, Theory and Economics*, Routledge.

McFadden, D. (1980) "Econometric Models of Probabilistic Choice Among Products", *J. of Business* 53: 513-529.

McFadden, D. (1978) "Modeling the Choice of Residential Location", in Karlquist, A., L. Lundquist, F. Snickers and J. Weibull, eds, *Spatial Interaction Theory and Planning Models*, North Holland.

McFadden, D. and K. Train (2000), "Mixed MNL Models of Discrete Response," *J. of Applied Econometrics*, 15: 447-470.

McNally, M. (2000), "The Four-step Model," in D. Hensher and K. Button, eds., *Handbook of Transport Modeling*, Pergamon.

Menger, C. (1990/1892) "On the Origins of Money," *Economic Journal*, (2)2: 239-55.

Menger, C. (1963/1883), *Problems of Economics and Sociology*, F. Nock, trans from the German edition of 1883, University of Illinois Press.

Menger, C. (1981/1871), *Principles of Economics*, J. Dingwall, ed, B. Hoselitz, trans from the German edition of 1871, New York University Press.

de Merchi, S. and J. Hamilton (2009), You Are What You Choose, Portfolio.

Minsky, M. (1985), *The Society of Mind*, Simon and Schuster.

Minsky, M., ed. (1968), *Semantic Information Processing*, MIT Press.

Minsky, M. and S. Papert (1969), *Perceptrons*, MIT Press.

Mladinow, L. (2008), *The Drunkards Walk: How Randomness Rules our Lives*, Pantheon.

Montague, R. (2006), *Your Brain is (Almost) Perfect*, Plume.

Morrison, S.A. and C. Winston (1990), "The Dynamics of Airline Pricing and Competition," *American Economic Review*, 80: 389-393.

Morrison, S.A., C. Winston and V. Maheshri (2005), "Competition and Welfare in the U.S. Airline Industry," Working paper, Brookings Institution.

von Neumann, J. and O. Morgenstern (1944/2004), *Theory of Games and Economic Behavior,* Princeton.

Nesbitt R. and L. Ross, (1980), *Human Inference, Strategies and Shortcomings of Social Judgment*, Prentice-Hall.

Newman, M. A-L Barabasi and D. Watts (2008), *The Structure and Dynamics of Networks,* Princeton University Press.

Nikolai, C. and G. Madey (2008), "Tools of the Trade: A Survey of Various Agent Based Modeling Platforms," *J. of Artificial Societies and Social Simulation*, 12:2, http://jasss.soc.survey.ac.uk/12/2/2.html.

Oluyomi, A, S. Karunasekera and L. Sterling (2007), "A Comprehensive View of Agent-oriented Patterns," *Autonomous Agent and Multi-Agent Systems,* 15:337–377.

Parker, R. (2008), *The Connectedness of the World's Airline Network*, AGIFORS Symposium, Montreal.

Parker, R. (2004), *Evaluating Schedules using the Network Value Index,* AGIFORS Symposium, Singapore.

Parker, R. and D. Bakken (2007), "Predicting the Unpredictable: Agent-based Models for Market Research," in P. Mouncey, and F. Wimmer, eds, *Best Practice in Market Research*, John Wiley and Sons.

Parker, R. and D. Bakken (2005), *Predicting the Unpredictable: Agent-based Models for Market Research,* ESOMAR Annual Congress, Cannes, France.

Parker, R., T. Cenesizoglu and R. Carson, (2005), *Aggregation Issues in Forecasting Aggregate Demand: An Application to U. S. Commercial Air Travel*, Advanced Research Techniques Forum, American Marketing Association, Couer d'Alene.

Parker R., and L. Garrow, (2006), *Modeling Customer Demand: Past, Present and Future Applications,* AGIFORS Symposium, Rome.

Parker, R., T. Glans, R. Lonsdale and Z. Zhang (2004), *The Naïve Universal Market Simulator,* American Marketing Association Advanced Research Techniques Forum, Whistler, B. C.

Parker, R. and R. Lonsdale (2007), *The Application of Agent-Based Simulation Modeling to Airline Passenger Booking Behavior and Revenue Management*, AGIFORS Revenue Management Study Group, Korea.

Parker R. and R. Lonsdale (2004), *Using Expected Maximum Utility to Calculate the Network Value Index for Airline Network Enhancements.* AGIFORS Revenue Management Study Group, Auckland.

Parker, R., R. Lonsdale and F. Ervin (2007), *Methods of Measuring Airline Network Value*, INFORMS Congress, Seattle.

Parker, R., R. Lonsdale, F. Ervin, and Z. Zhang (2005). *GMAS: The Boeing Global Market Allocation System.* AGIFORS Scheduling and Strategic Planning Study Group, Seattle.

Parker, R. and D. Perroud (2008), *Exploring Markets with Agent-Based Computer Simulations*, 60th Annual ESOMAR Congress, Montreal.

Parker, R. and J. Walker (2007), *Determining the Passenger Value of Departure and Arrival Times Using a Mixed Logit Model*, AGIFORS Scheduling and Strategic Planning Study Group, Tokyo, Japan.

Parzen, E. (1962), *Stochastic Processes*, Holden-Day.

Press, W, B.Flannery, S. Teukolsky, and W. Verrerling, (1988), *Numerical Recipes in C*, Cambridge University Press.

Pollock, J. (1990), *Nomic Probability and the Foundations of Induction*, Oxford University Press.

Popper, K. (1959/2002), *The Logic of Scientific Discovery*, Routledge.

Profillidis, V. (2000), "Econometric and Fuzzy Models for the Forecast of Demand in the Airport of Rhodes," *J. of Air Transport Management*, 6:95-100.

Pronin, E. D. Lin and L. Ross (2002), "The Bias Blind Spot: Perceptions of Bias in Self versus Others," *Personality and Social Psychology Bulletin*, 28: 369-381.

Pronin, E., T. Gilovich and L. Ross (2004), "Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others," *Psychology Review*, 111: 781-99.

Proussaloglou, K. and F. Koppelman (1999), "The Choice of Air Carrier, Flight and Fare Class," *J. of Air Transportation Management*, 5: pp 193-201.

Quammem, D. (2006), *The Reluctant Mr. Darwin*, Atlas Books. W. W. Norton.

Railsback, S. and S. Lytinen (2006), "Agent-Based Simulation Platforms: Review and Development Recommendations," *Simulations*, 82: 609-623.

Rapaille, C. (2006), *The Culture Code*, Broadway Books.

Robinson, L. (1995), "Optimal and Approximate Policies for Airline Booking with Sequential Nonmonotonic Fare Classes," *Operations Research*, 43, 252-263.

Rossi, P., G. Allenby and R. McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons.

Rubenstein, A. (1998), *Modeling Bounded Rationality*, MIT Press.

Rumelhart, D. and J. McClelland (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press.

Russell, B. and A. Whitehead (1910-1913), *Principia Mathematica*, Oxford University Press.

Ryan, M. (1991), *Possible Worlds: Artificial Intelligence and Narrative Theory*, Indiana University Press.

Saab, S. and P. Zouein (2001), "Forecasting passenger load for a fixed planning horizon," *Journal of Air Transport Management*, 7:361-372.

Sabre, Inc, (2009), *Sabre Global Demand Data*, PowerPoint presentation, Sabre, Inc.

SAS, (2009), http://www.sas.com/

Schank, R. (1990), *Tell Me a Story: A New Look at Real and Artificial Memory*, Scribner.

Schoemaker, P. (1995), "Scenario Planning: A Tool for Strategic Thinking," *Sloan Management Review*, Winter 1995, pp. 25-40.

Schumpeter, J. (1976), *Capitalism, Socialism, and Democracy*, 5th ed, George Allen and Unwin.

Schumpeter, J. (1954), *History of Economic Analysis*, Oxford University Press.

Shelling, T (1973), "Dynamic Models of Segregation," *J. of Mathematical Sociology*, 1: 143-186.

Shepherd Systems, Inc. (2006), Private communication.

Shermer, M.  (2004), *The Science of Good and Evil*, Henry Holt & Co.

Shermer, M. (2008), *The Mind of the Market*, Henry Holt & Co.

Shimoda, T. (2002), “An Interactive Software-Agent Smoking Cessation Program,” *Proceedings of the 36th Hawaii International Conference on System Sciences*, IEEE.

Shorto, R. (2008), *Descartes' Bones:  A Skeletal History of the Conflict Between Faith and Reason*, Doubleday.

Simon, H. (1957), *Models of Man*, John Wiley & Sons.

Simon, H. (1996), *The Sciences of the Artificial*, MIT Press.

Simon, H. (1997), *Models of Bounded Rationality:  Empirically Grounded Economic Reason*, MIT Press.

Simons, D. and C. Chabris, (1999), “Gorillas in Our Midst:  Sustained Inattentional Blindness for Dynamic Events,” *Perception*, 28: 1059-1074.

Simpson, E. (1951), “The Interpretation of Interaction in Contingency Tables,” *J. of the Royal Statistical Society*, B 13:238-241.

Small, K. (1987), “A Discrete Choice Model for Ordinal Alternatives,” *Econometrics*, 55: 409-424.

Snowden, D. (2002), “Narrative Patterns:  Uses of Story in the Third Age of Knowledge Management,” *J. of Information and Knowledge Management*, 1:1 pp 1-6.

Spencer, H. (1890), *First Principles*, 5th ed., Williams and Norgate.

Spencer, H. (1892), *Essays Scientific, Political and Speculative*, Appleton.

Sober, E. and D. Wilson, (1998), *Unto Others:  The Evolution and Psychology of Unselfish Behavior*, Harvard University Press.

Sternberg, R. and T. Ben-Zeev (2001), *Complex Cognition:  The Psychology of Human Thought*, Oxford University Press.

Stewart, J.Q. (1948), “Demographic Gravitation: Evidence and Application,” *Sociometry*, 1: 31-58.

Strack, F. and T. Mussweiler (1997),, “Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility,”  *J. of Personality and Social Psychology*, 73, 437-446.

Stuart, I. and V. Joines (1987), *TA Today: A New Introduction to Transactional Analysis*, Lifespace Publishing.

Sun, R. and Naveh, I. (2004), “Simulating Organizational Decision-making Using a Cognitively Realistic Agent Model,” *J. of Artificial Societies and Social Simulation*, 7:3. http://jasss.soc.surrey.ac.uk/7/3/5.html.

Swan, W. (1996), *The Spill Model*, Boeing White Paper, unpublished.

Taleb, N. (2007), *The Black Swan*, Random House.

Talluri K. and G. van Ryzin (2004), *The Theory and Practice of Revenue Management*, Springer.

Tarvis, C. and E. Aronson, (2007), *Mistakes Were Made (but not by me)*, Harcourt.

Tesfatsion, L. (2009), http://www.econ.iastate.edu/tesfatsi/ace.htm

Thalmann, N. and D. Thalmann (1990), *Computer Animation: Theory and Practice*, 2nd Ed. Springer-Verlag.

Thompson, J (2000), *Simulation; A Modeler's Approach*, Wiley.

Tobias, R. and C. Hoffmann (2004), "Evaluation of Free Java-libraries for Social-scientific Agent Based Simulation, *J. of Artificial Societies and Social Simulation*, 7:1, http://jasss.soc.survey.ac.uk/7/1/6.html.

Todd, P. (2001), "Fast and Frugal Heuristics for Environmentally Bounded Minds," in Gigerenzer, G. and R. Selten, eds, *Bounded Rationality: The Adaptive Toolbox*, MIT Press.

Toolin, M. (2001), *Narratives: A Critical Linguistic Introduction*, Routledge.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press.

Train, K. (1998), "Recreation Demand Models with Taste Variation over People," *Land Economics*, 74: 230-239.

Transport Canada (2002), *Passenger Origin-Destination Forecasting Model*, Aviation Forecasts Economic Analysis Policy Group, Transport Canada, February.

Trivers, R. (1971), "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology*, 46: 35-57.

Troscheutz, S. (2007), *.NET Random Number Generators and Distributions*, http://www.codeproject.com/KB/recipes/Random.aspx.

Tumminia, D. (2006), *When Prophecy Never Fails: Myth and Reality in a Flying-Saucer Group*, Oxford University Press.

Tufte, E. (2006), *Beautiful Evidence*, Graphics Press LLC.

Tufte, E. (2001), *The Visual Display of Quantitative Information*, 2nd ed. Graphics Press LLC.

Tufte, E. (1997), *Visual Explanations*, Graphics Press LLC.

Tufte, E. (1990), *Envisioning Information*, Graphics Press LLC.

Turner, S. (1994), *The Creative Process: A Computer Model of Storytelling and Creativity*, Lawrence Erlbaum.

Tversky, A. (1972), "Elimination by Aspects: A Theory of Choice", *Psychology Review*, 79: 281-299.

Tversky, A. and D. Kahneman (1974), "Judgment under Uncertainty: Heuristics and Biases," *Science*, 185: 1124-1131.

UIUC (2009), *Guide to the Human Relations Area Files*, University of Illinois at Urbana-Champaign.

Veblen, T. (1919/1990), *The Place of Science in Modern Civilization and other Essays*, Augustus Kelly, 1919 ed, reprinted from the 1990 edition, Brunswick.

Veblen, T. (1914/1990), *The Instinct of Workmanship, and the State of the Industrial Arts*, reprinted from 1914 edition, Brunswick.

U.S. Transportation Research Board (2002), *Aviation Demand Forecasting: A Survey of Methodologies [Number E-C040]*, Washington, DC: Transportation Research Council/National Research Council.

Valente, M. (1999), *A Model for the Evolution of Markets: Consumer Behavior and Technological Complexity in the Evolution of Markets*, PhD Dissertation, University of Aalborg.

Van Frassen, B. (2002), *The Empirical Stance*, Yale University Press.

Van Frassen, B. (1989), *Laws and Symmetry*, Clarendon.

Venod, B. (2006), Executive Vice President, Marketing, Sabre Inc., Private communication.

Verleger, P.K. (1972), Models of the Demand for Air Transportation," *Bell J. of Economics*, 3: 437-457.

Waldorp, M. (1992), *Complexity: The Emerging Science at the Edge of Order and Chaos*, Simon and Schuster.

Wang, P. and D. Pitfield (1999), "The Derivation and Analysis of the Passenger Peak Hour: An Empirical Application to Brazil," *J. of Air Transport Management*, 5:135-141.

Williams, G. (1966), *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*, Princeton University Press.

Wolfram, S. (1994), *Cellular Automata and Complexity: Collected Papers*, Westview.

Wooldridge, M. (2002), *An Introduction to MultiAgent Systems*, Wiley.

Wynne-Edwards, V. (1962), *Animal Dispersion in Relation to Social Behavior*, Oliver and Boyd.

Yoo, K. and N. Ashford (1996), "Carrier Choices of Air Passengers in the Pacific Rim: Using Comparative Analysis and Complementary Interpretation of Revealed Preference and Stated Preference Data," *Transportation Research Record*, #1562, pp 1-7.

Zellner, A (1971/1996), *An Introduction to Bayesian Inference in Economics*, John Wiley and Sons.