ANALYSIS FOR ADAPTIVE COMPLEX PUBLIC ENTERPRISES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in
the Graduate School of The Ohio State University

By
Yushim Kim, M.P.A.

The Ohio State University
2006

Dissertation Committee: Approved by:
Professor Anand Desai, Adviser
Professor Robert Greenbaum
Professor Ningchuan Xiao

Adviser
Graduate Program
in Public Policy & Management
ABSTRACT

The objective of this study is to explore the interactions and interdependence among autonomous and purposeful agents within a complex policy system. From a policy perspective, the focus is on studying systematic patterns that emerge from the interactions among adaptive agents, how these patterns describe actual behavior, and implications of this approach for policy analysis and decision-making. The underlying conceptual question is framed within the context of fraud in the delivery of public services. This conceptualization leads to the following research questions. First, what macro patterns emerge from the interactions among heterogeneous agents in a public service delivery program? Second, what are some of the mechanisms that underlie emergent macro patterns? The conceptual framework is illustrated using data from The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) in Ohio.

Store choice models using certain deterministic factors that characterize the context in which these decisions occur have been used to describe and explain the shopping behavior of individuals. A common criticism of these models is that they do not capture the inherent complexity in the behaviors or the decisions. However, using recent advances in modeling complexity and emergent behaviors it is now possible to develop models that capture nonlinear, random, and self-organizing adaptive behaviors.
This dissertation consists of a set of three interconnected essays. The first essay builds upon the criticisms of the linear Newtonian-Cartesian models of the policy process and decisions and argues that recent advances discussed in the literature on complexity hold promise in providing realistic abstractions of policy issues. The second essay argues that the classical store choice model can be an effective tool for modeling the behavior of program participants as they travel to receive services. A product of such modeling is the spatial patterns that would emerge from the program participants following simple decision rules in selecting the service providers. Deviations from these patterns can be construed as being indicative of potentially fraudulent behavior and therefore deserving of greater scrutiny in efforts to reduce fraud. As mentioned earlier, such models do not capture the interactions among the service providers, program participants and the rules that govern these interactions. The third essay describes an implementation of the general framework developed in the first essay and addresses some of the limitations of the models described in the second essay. An agent-based model of the interactions and interdependence among heterogeneous agents is implemented and used to replicate the results obtained from the store choice model. The flexibility of these models is further illustrated by discussing how such models can be used to simulate not only the deterministic models but can also be used to explore adaptive behaviors in an interdependent system of individuals, service providers, and public agencies.

In sum, the dissertation illustrates that concepts drawn from the literature on complexity and the computational simulation models based on these concepts provide richer abstractions of complex policy problems. Above all, this dissertation contributes to the body of knowledge by investigating the link between micro-behavior of heterogeneous agents and macro-patterns that emerge from such behaviors.

Keywords: Complexity, Agent-Based Models, Decision-Making, Policy Analysis
In Memory Of
My Father,
Yong-Sub Kim
(2. 24. 1942 ~ 9. 8. 1999)
ACKNOWLEDGMENTS

“An introduction is the place for acknowledgements; but my sense of indebtedness leaves me dumb. Socialized and humanized by being claimed from birth onwards as a member of so many communicating human groups; ushered into self-awareness through a language, every word of which resonates with the meanings of ancient usage; heir to several cultural traditions, each far too abundant for my assimilation – how can I name or number or know the living and the dead who have shaped my thoughts and me?” Sir Geoffrey Vickers (1970, p. 14)

One debt I must identify and am glad to acknowledge for this particular piece of work: It is due, which many other less explicit, to my dissertation committee (Dr. Anand Desai, Dr. Robert Greenbaum, and Dr. Ningchuan Xiao) who have unwaveringly supported my thoughts over the past years.
VITA

Dec 30, 1971 .................... Born – Kumsan, South Korea
Feb 25, 1997 .................... MPA, Seoul National University, South Korea
                                   Ohio Department of Health, USA
AU2005, AU2006 ................. Graduate Research Associate
                                   The Ohio State University, USA

FIELDS OF STUDY

Major Field: Public Policy & Management
TABLE OF CONTENTS

Abstract................................................................................................................................. iii
Acknowledgments ................................................................................................................ vi
Vita........................................................................................................................................ vii
List of Tables ...................................................................................................................... xi
List of Figures .................................................................................................................... xii

Chapters:

1. Introduction .................................................................................................................. 1
   1.1. Problem statement ................................................................................................. 2
       1.1.1. Policy context ............................................................................................... 2
       1.1.2. Problem in operation .................................................................................. 6
   1.2. Scope of research .................................................................................................. 9
       1.2.1. Purpose of research ...................................................................................... 9
       1.2.2. Research objectives .................................................................................... 13
       1.2.3. Research questions ..................................................................................... 15
   1.3. Research design ................................................................................................... 16
       1.3.1. Research frameworks .................................................................................. 16
       1.3.2. Methods ...................................................................................................... 18
   1.4. Significance ......................................................................................................... 22
   1.5. Organization of the dissertation .......................................................................... 24

2. A short essay on complexity and policy analysis ....................................................... 26
   2.1. Introduction ......................................................................................................... 27
   2.2. Complexity .......................................................................................................... 28
       2.2.1. Complexity science ..................................................................................... 28
       2.2.2. Complexity of policy problems ................................................................... 30
   2.3. Managing complexity .......................................................................................... 38
       2.3.1. Social complexity and algorithms ................................................................. 39
       2.3.2. Computational social science models ......................................................... 40
       2.3.3. Development of agent-based models .......................................................... 43
       2.3.4. Some perspectives on policy modeling ....................................................... 47
   2.4. Conclusions ......................................................................................................... 50
3. Identifying fraud in a public delivery program ................................................................. 52

3.1. Introduction .................................................................................................................. 53
3.2. Store choice ................................................................................................................. 56
  3.2.1. Store choice models ............................................................................................... 56
  3.2.2. Applications of the store choice models ............................................................... 59
3.3. Method and data ......................................................................................................... 61
  3.3.1. Vendor propensity measures ................................................................................ 61
  3.3.2. Data ....................................................................................................................... 63
3.4. Results .......................................................................................................................... 65
  3.4.1. Descriptive statistics ............................................................................................. 65
  3.4.2. Risk groups .......................................................................................................... 66
  3.4.3. Distribution of vendor propensity measures ......................................................... 67
  3.4.4. Explaining the anomalies ..................................................................................... 69
3.5. Discussion .................................................................................................................... 72
  3.5.1. Summary .............................................................................................................. 72
  3.5.2. Policy implications ............................................................................................... 73
  3.5.3. Limitations ........................................................................................................... 74

4. Simulating fraud using an agent-based model ................................................................. 76

4.1. Introduction .................................................................................................................. 77
4.2. Background ................................................................................................................... 78
  4.2.1. Economics of crime .............................................................................................. 79
  4.2.2. White collar crime ............................................................................................... 80
  4.2.3. Geography of crime ............................................................................................ 81
  4.2.4. Management perspective ..................................................................................... 82
4.3. Method ........................................................................................................................ 83
  4.3.1. Policy problem in WIC ........................................................................................ 84
  4.3.2. Agent-based modeling ......................................................................................... 85
4.4. Simulating patterns of fraud ....................................................................................... 89
  4.4.1. Descriptive statistics from the empirical data ...................................................... 89
  4.4.2. Replicating descriptive statistics ........................................................................ 91
4.5. Testing the effect of a policy intervention ..................................................................... 107
  4.5.1. Deterrence hypothesis ......................................................................................... 107
  4.5.2. Sanctions in the WIC program .......................................................................... 110
  4.5.3. Implementing a policy intervention ..................................................................... 111
  4.5.4. Testing effects of a warning letter ...................................................................... 113
4.6. Discussion .................................................................................................................... 118
  4.6.1. Future trends ....................................................................................................... 118
  4.6.2. Summary ............................................................................................................ 119
5. Discussion .................................................................................................................. 121

   5.1. Contributions ......................................................................................................... 121
       5.1.1. Theoretical contributions .............................................................................. 121
       5.1.2. Practical contributions ................................................................................ 125
   5.2. Policy implications .............................................................................................. 127
   5.3. Limitations & future research ............................................................................... 128
       5.3.1. Vendor propensity measures ...................................................................... 128
       5.3.2. Fraud simulation ...................................................................................... 130

References .................................................................................................................. 132

Appendix:

   A. Vendor propensity measures ................................................................................ 153
   B. Fraud simulation user interface ........................................................................... 156
   C. Class diagram of fraud simulation ....................................................................... 157
   D. Fraud simulation ............................................................................................... 158
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Basic assumptions of complexity were compared with those of traditional positivistic approach. ............................................................................................................. 38</td>
</tr>
<tr>
<td>3</td>
<td>Program activities in an urban county (A) in Ohio WIC ....................................................... 65</td>
</tr>
<tr>
<td>4</td>
<td>Vendor propensity measures of specialty vendors - a. Risk Code (RC) met in the State high-risk vendor identification method (HR: High-risk, AR: At-risk) b. Compliance Investigation (CI) - NV: no violation; S: substitution; OC: overcharge; NF: no food vouchers submitted; NT: no price tags............................................................................................................. 72</td>
</tr>
<tr>
<td>5</td>
<td>Initial parameters of the basic properties used in the simulation ........................................... 89</td>
</tr>
<tr>
<td>6</td>
<td>Program activities in an urban county (B) in Ohio WIC........................................................ 90</td>
</tr>
<tr>
<td>7</td>
<td>Descriptive statistics of the simulated data at Time 299 (the total number of vendors is 200) ............................................................................................................................... 94</td>
</tr>
<tr>
<td>8</td>
<td>Percentage of vendors at different risk levels in the empirical vs. simulated data – gray rows are considered as “high-risk.” ................................................................................................ 97</td>
</tr>
<tr>
<td>9</td>
<td>Type of sanctions imposed by state and the percentage of vendors receiving sanctions – Note that the total percent does not add to 100 because vendors could have received more than one sanction during the survey period. ................................................................................ 111</td>
</tr>
<tr>
<td>10</td>
<td>Percentage of high-risk vendors at the last five steps of the simulation in different scenarios. The policy option influenced the interaction of the vendor agents with new comers as well as those participant agents currently involving in fraud at the vendor at Time 150................. 114</td>
</tr>
<tr>
<td>11</td>
<td>Percentage of high-risk vendors at the last five steps of the simulation in different scenarios. The policy option influenced the interaction of the vendor agents with new comers as well as those participant agents currently involving in fraud between Time 100 and Time 150.... 116</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>WIC program participation and costs – data show total participation and program costs as of September 26, 2006. Participation data are annual averages except FY 1974 (6 months). Program costs consist of food costs for participants and Nutrition Services and Administrative (NSA) costs. NSA costs accounted for approximately 30 percent of the total costs in 2005. By comparison, it was 14 percent in 1975; 19 percent in 1980; 23 percent in 1990; and 28 percent in 2000.</td>
</tr>
<tr>
<td>2</td>
<td>WIC business mechanism in Ohio - depending upon state’s delivery systems, the business mechanism can be slightly different. However, service delivery through retail stores is a major delivery mechanism in the US. Note that numbers show how many participants, local clinics, and vendors are of Ohio WIC. ODH stands for Ohio Department of Health. Figure 2 was drawn from my observation of Ohio WIC.</td>
</tr>
<tr>
<td>3</td>
<td>Defining &quot;high-risk vendors&quot; – this figure describes vendor management mechanism to monitor fraud in Ohio WIC. Identification criteria were excerpted from Federal Register Part 246. Fraud categories mainly occur through the interactions between vendors and participants. Other fraud mechanisms also exist and have been identified as listed.</td>
</tr>
<tr>
<td>4</td>
<td>Multidisciplinary problem – public service delivery programs consist of several heterogeneous actors who are interdependent and who dynamically interact. Public policy and management are interested in aggregated patterns emerging from their dynamics. On the other hand, crime literature in the social sciences has focused on the cause and nature of criminal behavior at an individual level or aggregated levels. Therefore, the literature has limited utility in exploring fraud that occurs from the dynamic interaction among heterogeneous actors. In recent years, there has been rich discussion on proximity, spatio-temporal patterns, and GIS for spatial analysis as a method of understanding complex social phenomena. Some of the tools were utilized for crime mapping and others were used for allocating locations based on the concept of spatial distributions. In this dissertation, I explored the gray area forward through the overlap of these interdependent research boundaries.</td>
</tr>
<tr>
<td>5</td>
<td>Research framework – research objective one is to develop statistical measures for identifying high-risk vendors using a traditional store choice model. This model is based on the assumptions of classical economics. Research objective two is to model the non-linear processes underlying the statistical and spatial patterns of fraud identified from the marketing literature and research objective one. A pattern-oriented agent-based model is built to achieve the second research objective.</td>
</tr>
</tbody>
</table>
Methodological framework – the smaller box inside the larger box presents the relationship between patterns in reality and their representation using a mathematical formula. For example, the Huff spatial interaction model explains consumer store choice behavior with a simple and elegant equation. Research objective one is to use this representation to confirm a reality in the WIC setting. The larger box presents the framework of agent-based models. Often agent-based models aim to replicate the representation of the reality or one’s understanding of reality rather than the reality itself. These multiple representations are achieved by assuming multiple realities and multiple behaviors of agents at micro levels. The model assumes that their dynamic nonlinear interactions self-organize to create certain patterns at macro levels. This also allows to project future patterns based on the nonlinear interactions among heterogeneous agents.

Java mechanism – Java is a language that allows programmers to build a virtual reality. The virtual reality consists of agents who have certain attributes and whose behavior is based on certain rules. Each agent class has its own members as instances of the agent class.

Modeling fractals and social complexity to inform policy decisions.

Schelling’s segregation model was conceptually introduced in his *Micromotives and Macrobehavior* (1978). Simulation above was captured by running a demo in MASON.

Comparison between social science models and policy analysis - the basic mechanism of policy analysis is not different from social science models. However, decision-making and action are crucial components to successful policy analysis. Therefore, policy analysis needs to continuously iterate and revise the whole process to support decision-making and purposeful actions in different contexts.

Simulated theoretical spatial patterns – this spatial pattern shows several clusters or trading areas that each vendor may serve when the Huff model is deterministically used. Numbers inside squares represent the number of lanes in each vendor.

Data synthesis and analysis – data were extracted from three different information systems for a county in Ohio WIC (certification system, vendor management system, and payment data system). Analysis was performed using GIS for identifying the locations, Matlab for calculating the probability matrix, and STATA for analyzing the final results.

Scatterplot of the MEAN & MAX values for 78 vendors.

Comparison of vendor propensity measure by risk groups that were categorized by using two different methods. Left-hand-side figure shows the distribution of MEAN vendor propensity measures when the vendors are grouped by the state high-risk vendor identification index. Right-hand-side figure shows the distribution when the vendors are grouped by the result of field investigation. Note that the total number of vendors included is not that large. Therefore, these figures should be interpreted with caution.

Adaptive complex enterprises – agents, interdependence, and interaction were defined above. While the Huff spatial interaction model addresses the interaction between participant and vendor agent, the agent-based model also includes activities of a public agency agent in the WIC system.
16 Distribution of store sizes – gray bars shows the percentage of vendors by the number of lanes in the empirical data (total vendors = 188) and black bars show the distribution of vendors by the number of lanes in the simulated data (total vendors = 200). 

17 Percentage of participant agents involved in fraud at Time 1 – it was programmed that more participant agents successfully commit fraud at the smaller vendors. 

18 Vendor redemption ratios over time by store size – the thick line in the middle shows vendor redemption ratio for all vendors. 

19 Distributions of sales activities by three criteria used in the empirical and simulated data to identify outliers. 

20 Distribution of vendors by risk level and store size in the (a) empirical (April 2004) and (b) simulated (at Time 299) data. 

21 Patterns of trading area - (a) a normal pattern in which the majority of customers of a store live close to the store; (b) a pattern that may indicate possible fraud where a significant numbers of customers of a store come from far away while loosing those who were supposed to visit. Three circles surrounding vendors represent primary, secondary, and fringe areas in the trading area analysis. 

22 Theoretical vendor choices of participant when the Huff model was deterministically implemented at Time 0. Two small vendors were highlighted: one with relatively high risk propensity (0.86) and the other with relatively low risk propensity (0.12). 

23 Fraud negotiations occurred at Time 1. 

24 Fraud negotiations continue for those participants who have a relatively high risk propensity (> 0.6) and are not involved in fraud. While the small vendor with low risk propensity remained stable (risk propensity: 0.12 to 0.15), the high-risk vendor was actively attracting participants (risk propensity: 0.86 to 1.0) at Time 292. Another one lane vendor was developing as a high-risk vendor (risk propensity: 0.66 to 0.86) at Time 292. Other vendors were maintaining their original trading areas. 

25 Highlighting the interaction between small vendors with high risk propensity and participants at Time 308 (black solid and dotted circles) while the small vendor with low propensity was faded out (gray solid circle). All vendors are small, having one lane. The vendor’s risk propensity was 0.86 at the beginning and quickly reached to 1.0 at the end of the simulation. The vendor was attracting participants from all over the place. 

26 Percentage of high-risk vendors as a response to a warning letter when the letter was sent to randomly selected high-risk vendors at Time 150. This policy option influences not only vendor risk propensity, but also the behavior of vendors in terms of recording actual sales for those participants involving in fraud. 

27 Percentage of high-risk vendors as a response to a warning letter when the letter was sent to randomly selected high-risk vendors at Time 100. The policy option was removed at Time 150. The option influenced not only vendor risk propensity, but also the behavior of vendors in terms of recording actual sales for those participants involving in fraud.
INTRODUCTION

*The origin of thinking is some perplexity, confusion, or doubt.*

*J. Dewey (1997)*

We have long known that public policy problems are “wicked” (Churchman, 1967; Rittel & Webber, 1973) or complex. However, we have not had good practical approaches to modeling this complexity. As a matter of fact, most models assume away the complexity inherent in any human system. Furthermore, current performance monitoring and evaluation approaches focus on constant and consistent goals and objectives in static settings. In an effort to overcome many of the shortcomings of standard performance studies, I draw upon recent developments in the study of complex organizations and explore the use of agent-based models. Such models allow me to study the complex interactions that exist in dynamic organizational settings.

In recent years there has been considerable discussion regarding the use of concepts based on complexity, networks and agent-based models. However, applications of these concepts and models to real problems, particularly at the scale in which they appear in the public sector, are rare. This research attempts to implement these concepts in a real context and provides an example of how a truly multidisciplinary approach to addressing complex problems can yield rich dividends.

---

1 The original version was published in 1910 by D.C. Health & Co., Publishers in Boston.
The traditional model of a hierarchical organization with stable and well-understood relationships has been effective in addressing routine problems. However, in rapidly changing unpredictable environments where individuals and entities react in response to each other and to changes in the environment, there is a need for adaptive models that can emulate some of these interactive and non-routine behaviors. A typical example of non-routine dynamically changing behavior is fraud in the delivery of public services. For instance, the misuse of funds for The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) is a common example of fraud in the public sector.

In this study, I develop a model of the relationships between state agencies, program participants, and service providers to explore their static as well as dynamic interactions in the WIC program. In particular, I am interested in exploring how such models can be used to monitor public delivery systems and to detect and deter fraud.

1.1. Problem Statement

1.1.1. Policy Context

WIC aims to safeguard the health of low-income women, infants, and children up to age 5 who are at nutritional risk. The program provides nutritious supplemental foods, nutritional education, and referrals to health care and other social services. This program is available in all 50 states, the District of Columbia, 34 Indian Tribal Organizations, and the US territories. These 90 WIC State agencies administer the program through 2,200 local agencies and 9,000 clinic sites.²

² Information was retrieved from USDA WIC fact sheet in November 1, 2006.
The origins of WIC were related to the growing concern about malnutrition among low-income mothers and children in the 1960s. As a part of several activities in response to this concern, the WIC program was formally authorized on September 26, 1972 by an amendment to the Child Nutrition Act of 1966. The legislation sponsored by Senator Hubert H. Humphrey established this program as a 2-year pilot program, and was made permanent on October 7, 1975 (Oliveira, Racine, Olmsted, & Ghelfi, 2002). The United States Department of Agriculture (USDA) was given responsibility for administrating the program. WIC is not an entitlement program, but a Federal grant program for which Congress authorizes a specific amount of funds each year.

To qualify for WIC, participants must meet categorical, residential, income, and nutritional risk eligibility requirements. Categorical eligibility requirements include

- Pregnant women: up to 6 weeks after birth or after pregnancy ends
- Breastfeeding women: up to infant’s 1st birthday
- Non-breastfeeding postpartum women: up to 6 months after the birth of an infant or after pregnancy ends
- Infants
- Children up to their fifth birthday

WIC applicants must reside within the state where they establish eligibility. The family income of WIC applicants must meet specified guidelines (i.e. 185 percent of the US Poverty Income Guidelines in Ohio). Applicants must be at nutritional risk which will be defined by a health professional. According to the study of *WIC Participant and Program Characteristics 2004* (USDA, 2006), approximately half of total participants are children (49.8 percent). Infants account for 25.7 percent and women are 24.5 percent of those enrolled in WIC.

More than 8 million women, infants, and children received WIC benefits each month in 2005. In 1974, the first year WIC was permanently authorized, 88,000 people participated in the
US. By 1980, participation was at 1.9 million; by 1985 it was 3.1 million; by 1990 it was 4.5 million; and by 2000 it was 7.2 million. Therefore, the program has grown approximately 100 fold during the last 30 years. Congress appropriated $5.2 billion for WIC in FY 2006. By comparison, the WIC Program appropriation was $20.6 million in 1974; $750 million in 1980; $1.5 billion in 1985; and $2.1 billion in 1990. Figure 1 presents WIC program participation and total costs since 1974. The figure shows that the program has been expanded dramatically in terms of participation and costs.

Figure 1: WIC program participation and costs – data show total participation and program costs as of September 26, 2006. Participation data are annual averages except FY 1974 (6 months). Program costs consist of food costs for participants and Nutrition Services and Administrative (NSA) costs. NSA costs accounted for approximately 30 percent of the total costs in 2005. By comparison, it was 14 percent in 1975; 19 percent in 1980; 23 percent in 1990; and 28 percent in 2000.

3 Date was retrieved from USDA website in November 1, 2006.
WIC operates through a federal-state-local partnership. State agencies are responsible for program operations. They contract with local WIC sponsoring agencies, allocate funds to them, and provide assistance to the local agencies. Local WIC agencies provide services to WIC participants either directly, or through local service sites (clinics). The clinics certify applicants, provide nutritional education, make referrals to other social services, and distribute food vouchers to be used at WIC participating retail stores (Oliveira, Racine, Olmsted, & Ghelfi, 2002). State agencies use three different types of food delivery systems. (1) Participants obtain supplemental foods by exchanging food vouchers at authorized retail outlets. (2) Supplemental food is delivered to the participant’s home. (3) Participants pick up supplemental foods from WIC storage facilities operated by the state or local agencies. States can use any combination of the three delivery systems.

This dissertation focuses on Ohio WIC as a policy context. In terms of participation, Ohio WIC is eighth largest program in the US as of July 2006. Ohio WIC serves approximately 277,000 participants each month with over $150 million each year (110 million for food costs and $45 million for NSA costs in 2005). Ohio WIC has contracts with over 200 local clinics and 1,400 vendors. Figure 2 presents the WIC business mechanism in Ohio. Each month, participants receive three or four vouchers with food benefits at local clinics. These participants are expected to redeem their benefits at WIC vendors within a specified period since Ohio WIC uses the retail delivery system. Each voucher specifies what products and quantities the participant can purchase, as well as maximum prices that the state will pay for an allowable food. The state monitors the overall flow of transactions in the WIC system.

---

4 Data were retrieved from USDA website in November 1, 2006.

5 Vendors refer to private entities that contract with public agencies to deliver public services such as Food Stamp benefits and WIC foods.
Figure 2: WIC business mechanism in Ohio - depending upon state’s delivery systems, the business mechanism can be slightly different. However, service delivery through retail stores is a major delivery mechanism in the US. Figure 2 was drawn from my observation of Ohio WIC. Note that numbers show how many participants, local clinics, and vendors are of Ohio WIC. ODH stands for Ohio Department of Health.

1.1.2. Problem in Operation

It is unrealistic to assume flawless implementation of policy design for service delivery. Unexpected challenges frequently occur in the web of service delivery systems. In such complex systems, a small flaw such as failure to monitor illegal behavior can have serious consequences.

---

6 Holland & Miller (1991) defined complex systems as follows: (1) complex systems consist of a network of interacting agents, (2) complex systems exhibit a dynamic, aggregate behavior that emerges from the individual activities of the agents, (3) this aggregate behavior can be described without detailed knowledge of the behavior of the individual agent, and (4) the agents behave adaptively.
for the whole system and ultimately damage the integrity of the public program. Therefore, effective monitoring has been a crucial part of public service delivery programs.

The Food Stamp program and WIC are well-known service delivery programs implemented by USDA. Currently, these programs together spend over $35 billion (Food Stamps approximately $31 billion and WIC approximately $5 billion in Fiscal Year 2005)\(^7\) to provide food and nutritional supplement for the needy. For such large programs, even a small fraction of the money that can be attributed to any type of illegal benefit transaction creates substantial monetary loss at the aggregate level. For example, trafficking – the exchange of public service benefits for cash – diverted about $395 million per year from Food Stamp benefits between 1999 and 2002 (Macaluso, 2003). This accounts for 2.2 percent of total costs of approximately $18 billion during these years.

While in the private sector, theft or “shrinkage” is mainly due to customers or employees acting as individuals, fraud in public programs more often requires the active participation of the program participant as well as the service provider such as vendors. In such a loosely coupled system,\(^8\) participants can commit fraud with or without malicious intentions, but its feasibility is enhanced by complicity on the part of vendors. The multiplicity of different types of interactions complicates the issue thus increasing the likelihood of unanticipated non-routine problems. For instance, a study of national WIC retail vendors showed that 18.1 percent of the retail vendors were involved in overcharging violations one or more times, including 12.4 percent that overcharged once, 4.2 percent that overcharged twice, and 1.5 percent that overcharged in three separate investigations (USDA, 2001). For vendors in the WIC system, overcharging is the main source of fraud. Some other illegal activities include forcing unwanted purchases, substitutions of

\(^7\) Information was retrieved from USDA website in November 7, 2006.

\(^8\) In the loosely coupled system, when one of the variables is disturbed, the disturbance tends to be limited rather than to ramify, or it takes a long time to affect other variables (Weick, 1979, pp.110-112).
WIC foods for unauthorized items (substitution), and trafficking. Each of these is different type of fraud, and they are all considered to be illegal behaviors in the WIC system.

The focus of traditional WIC fraud detection methods has been primarily on identifying high-risk vendors\textsuperscript{9} using criteria such as store characteristics, sales volume, participant complaints, and field investigations\textsuperscript{10} (USDA, 2000).\textsuperscript{11} Recently, the effort has been strengthened with the implementation of Electric Benefit Transfer (EBT)\textsuperscript{12}. However, these methods are static, less effective in uncovering the interaction between corrupt agents, and are easy to evade because of familiarity with detection procedures. The major weaknesses of these methods are as follows:

- Store characteristics: tend to target small vendors.
- Sales volume: identifies only those who show large volumes.
- Participant complaints: are limited to specific situations and are rare.
- Field investigations: are expensive and can be easily avoided.

Figure 3 provides a schematic summary of vendor management mechanism surrounding fraud in the WIC system. As the figure shows, the standard controlling mechanisms do not mirror the complexity of the issue. Further, the simple fraud detection efforts are no match for the dynamic and adaptive processes that they are supposed to monitor and regulate. Hence, there is

---

\textsuperscript{9} High-risk vendors are “those who have high probability of committing vendor violations.” (USDA, 2000) Vendor violations are not only limited to administrative violations, but also to illegal monetary transactions. Administrative violations not necessarily indicate monetary violations. However, monetary violations automatically encompass administrative violations.

\textsuperscript{10} An undercover investigator poses as a participant and attempts to exchange WIC vouchers inappropriately, such as for nonapproved food or nonfood items.


\textsuperscript{12} “Electronic Benefit Transfer (EBT) is an electronic system that allows a recipient to authorize transfer of their government benefits from a Federal account to a retailer account to pay for products received.” http://www.fns.usda.gov/fsp/ebt/
the need for an alternative framework to address complex and non-routine problems in public management.

Figure 3: Defining "high-risk vendors" – this figure describes vendor management mechanism to monitor fraud in Ohio WIC. Identification criteria were excerpted from Federal Register Part 246. Fraud categories mainly occur through the interactions between vendors and participants. Other fraud mechanisms also exist and have been identified as listed.

1.2. Scope of Research

1.2.1. Purpose of Research

Fraud is a crime that violates social norms, uses secretive processes, injures victims, and benefits perpetrators unfairly (Barker & Roebuck, 1973; Vandenburgh, 1999). There have been a variety of forms of fraud throughout history. They include money laundering, credit card fraud,
and telecommunications fraud. In the public sector, fraud in welfare, health care, and child care programs is well-documented by government agencies such as Government Accountability Office (GAO) and USDA.

Academics have also studied crime including fraud. It has been studied under various names across different disciplines. For example, economists approach crime and punishment as a special case of the more general economic theory of choice. According to this theory, criminals maximize their expected utility subject to constraints in an uncertain world. Crime is also a topic in the sociology, criminology, and legal literatures. According to Sutherland (1940), fraud is a white-collar crime. In its original conceptualization, white-collar crime occurs in business contexts, and has been interpreted as crime committed by a person of respectability and high social status in the course of his occupation (Braithwaite, 1985; Baker, 2004). Geography of crime studies the spatial dimension of crime and the effect of place on crime (Chainey & Ratcliffe, 2005). This approach is based on the assumption that crime occurs more frequently in certain places than in other places.

Therefore, the academic literature offers multiple accounts of the nature and causes of criminal behavior. However, the literature has focused on the relationship between criminal behaviors of individuals and causes of the behavior based on discipline-specific assumptions. These studies have not provided sufficient understanding of the mechanism underlying the practice of fraud, especially in public delivery programs where several entities are dynamically and loosely interrelated. Therefore, they do not describe the complex interactions that allow such fraud to occur. They have limited utility for public managers in providing practical and realistic advice for preventing and detecting fraud.

As with many types of crime, the management literature has also studied fraud. The focus of this literature has been a description of criminal activities in specific settings such as organizations or public programs (Ziegenfuss, 1996; Wait, 1997) rather than prediction or
explanation of criminal behavior of individuals. Further, analysis has also had to prescriptions about reducing crime. Literatures on fraud published from government agencies focus on normative guidelines (GAO, 1999) and numerical estimates of fraud in public programs (USDA, 2001). Fraud prevention and detection methods have been developed by focusing on separate players in the programs rather than seeing them as a system. Statistical methods and actual field investigations have been most common approach to dealing with fraud in practice (Bolton & Hand, 2002).

In recent years, scholars have attempted to understand fraud as an outcome of complex interactions. For example, Wilhelm (2004) framed fraud as an issue that requires dynamic, evolving, and adaptive management due to its complexity. Provost (2002) argued that the adaptive nature of fraud in public programs is not addressed well in the typical information system life cycle based on linear processes. These studies imply that fraud needs to be understood within a dynamic and adaptive framework. Spatio-temporal patterns need to be considered carefully into the framework in order to address them effectively. This dissertation is an effort to respond to that demand in public management.

In Figure 4, I identify fraud in the delivery of public services at the intersection among three different research boundaries that I will discuss or use throughout this dissertation. This is the area where individual behaviors and aggregated patterns overlap. The interaction and interdependency among heterogeneous agents is key to the exploration of the issue. I use a comprehensive framework by linking micro behavior and macro patterns to study fraud at the complex intersection of these boundaries.
Figure 4: Multidisciplinary problem – public service delivery programs consist of several heterogeneous actors who are interdependent and who dynamically interact. Public policy and management are interested in aggregated patterns emerging from their dynamics. On the other hand, crime literature in the social sciences has focused on the cause and nature of criminal behavior at an individual level or aggregated levels. Therefore, the literature has limited utility in exploring fraud that occurs from the dynamic interaction among heterogeneous actors. In recent years, there has been rich discussion on proximity, spatio-temporal patterns, and GIS for spatial analysis as a method of understanding complex social phenomena. Some of the tools were utilized for crime mapping and others were used for allocating locations based on the concept of spatial distributions. In this dissertation, I explored the gray area forward through the overlap of these interdependent research boundaries.
In this dissertation, I (1) frame the issue of fraud in the delivery of public services as a complex system, (2) implement a classical store choice model and an agent-based model for capturing the interaction and interdependency within the policy system, and (3) discuss the utility of the frameworks and simulation models for providing practical insights for managers. I use the WIC Program to illustrate the applicability and utility of the proposed modeling framework.

1.2.2. Research Objectives

The main conceptual issue I address is how one makes sense of the interaction and interdependency between autonomous actors within a policy system. From a policy perspective, I am interested in systematic patterns that emerge from the interaction between adaptive agents within a policy system,\(^\text{13}\) how these patterns describe actual behavior, and the implications of this process for policy analysis and decision-making.

To answer the conceptual question, this research has two study objectives. First, I discuss how commonly used store choice models can be successfully implemented in the design of monitoring systems to identify the spatial patterns that discriminate between different types of agent behavior. In spite of the apparent success of these models in predicting consumer choice, they suffer from a variety of theoretical shortcomings, particularly when applied to the delivery of public goods and services. The second objective of this study is to address some of these shortcomings using an alternative method.

\(^{13}\) Adaptive agents are those who influence one another in response to the influence they receive from each other (Marcy & Willer, 2002).
**Objective One:** I utilize a consumer store choice model\(^{14}\) to develop vendors’ propensity measures. By propensity measures, I mean spatial interaction statistics based on store choice probabilities. In other words, these are measures of the likelihood of an individual selecting a vendor among alternatives. The store choice probabilities are measured from the interaction between vendors and participants. My first objective will be to explore how these propensity measures can be used to identify high-risk vendors who are likely to commit fraud in the WIC system. I use Geographical Information Systems (GIS) to provide visual displays of the underlying data, propensity measures, and interaction patterns. I claim that ‘anomalous’ spatial patterns can be effective in identifying high-risk vendors. Anomalous patterns are those which differ from theoretically predicted patterns.

**Objective Two:** Despite their success in prediction, some assumptions underlying these choice models and tools have been questioned as being unrealistic (Bankes, 2002), even undesirable (Starbuck, 2004; Talbot, 2005), and at best, crude approximations (Drezner & Eiselt, 2002). These store choice models are also limited in their ability to address other decision criteria such as random choice and neighborhood influence. My second objective is to conduct computer experiments and simulations to explore the underlying mechanisms of emerging statistical and spatial patterns of fraud using an agent-based model. The models allow me to simulate the complex behavior of adaptive agents and to investigate patterns that emerge from the dynamic interaction (Axelrod, 1997a; Bankes, 2002).

\(^{14}\) The models can address the interactions between vendors and participants. Although the models may not be able to address other interactions such as among vendors or among participants, I assume that final interaction patterns between vendors and participants encompass substantial portions of other interactions.
1.2.3. Research Questions

I have two sets of research questions. Research objective one tests the utility of a classical store choice model that was developed for general consumers in the context of a disadvantaged group such as public assistance program participants. In this context, question set one focuses on how to identify high-risk vendors using the store choice model. Here fraud is seen as an independent choice of rational individuals, and the model captures static interactions between independent vendors and participants. Research objective two simulates the process of emergent statistical and spatial patterns of fraud in the public program. Therefore, question set two frames fraud as a complex system and attempts to understand the underlying mechanisms of the emergent patterns in the public delivery system.

Question Set One:

- What spatial patterns can be expected from the interaction between vendors and participants?
- How can I identify high-risk vendors who have the potential of committing fraud by capturing the interaction between vendors and participants?
- Is this propensity measure effective in identifying high-risk vendors?
- What are the shortcomings of this approach?

Question Set Two:

- Can fraud be understood as a complex system?
- How can I study fraud when it is framed as a complex system?
- Can I emulate the empirical patterns of fraud using an agent-based model?
- How well does a policy option deter fraud in the simulated policy system?
1.3. Research Design

1.3.1. Research Frameworks

Social phenomena have two inseparable aspects: patterns and processes. The literature treats these two as distinct phenomena that require different approaches and understanding. Figure 5 presents a research framework to include both aspects of a social issue. The framework includes assumptions regarding human-beings, spatio-temporal interaction, and the link between micro-behaviors and macro-patterns. For example, by developing statistical propensity measures of vendors using the spatial interaction model, I identify high-risk vendors and their spatial patterns. This propensity measure captures static interactions between vendors and participants at a certain time when a pattern of fraud exists. The efficacy of the measure is also tested using the result of actual field investigation.

This prediction-oriented model does not explain how the equilibrium pattern was formed. One can only speculate based on the assumptions of the models. Therefore, I extend my research objective to process modeling, incorporating assumptions on human interactions using computer simulations. In this stage, I include interaction (behavioral) rules between participants and vendors as well as their basic properties. I replicate the statistical and spatial patterns of fraud in empirical data using an agent-based model. This model also allows me to simulate the effect of policy options in deterring fraud in the simulated policy system.
Figure 5: Research framework – research objective one is to develop statistical measures for identifying high-risk vendors using a traditional store choice model. This model is based on the assumptions of classical economics. Research objective two is to model the non-linear processes underlying the statistical and spatial patterns of fraud identified from the marketing literature and research objective one. A pattern-oriented agent-based model is built to achieve the second research objective.
Figure 6 presents a methodological framework that I implement for the research framework. I revised Drogoul & Ferber’s figure (1994, p.134) that contrasts classical stochastic simulation models and agent-based models, after reviewing James (1996),15 Maturana & Varela (1987), Lash (1990), and Mingers (1995). This is a comprehensive framework that encompasses patterns and processes. It also shows that agent-based models can implement traditional research models as well as enhance the models with realistic assumptions.

1.3.2. Methods

Statistical Measure

I utilize the Huff model (1964) to develop vendor propensity measures. The basic assumption underlying this approach is that if a vendor attracts a substantial number of participants who are predicted by the model to have a low probability of visiting the vendor, then their spatial pattern is anomalous in comparison with these obtained from the trading area analysis pattern. I hypothesize that the anomalous spatial patterns indicates potential fraud through collaboration between vendors and participants. I evaluate the predictive power of these propensity measures in discriminating between high- and low-risk vendors as identified by the state’s monitoring system and compliance investigations. After establishing test-beds to explore the applicability of store choice models to monitor and identify potential fraud, I also enhance the model by incorporating some of the complexity and dynamics inherent in the interaction among players in the policy delivery system using an agent-based model.

15 Original version was published in 1909.
Figure 6: Methodological framework – the smaller box inside the larger box presents the relationship between patterns in reality and their representation using a mathematical formula. For example, the Huff spatial interaction model explains consumer store choice behavior with a simple and elegant equation. Research objective one is to use this representation to confirm a reality in the WIC setting. The larger box presents the framework of agent-based models. Often agent-based models aim to replicate the representation of the reality or one’s understanding of reality rather than the reality itself. These multiple representations are achieved by assuming multiple realities and multiple behaviors of agents at micro levels. The model assumes that their dynamic nonlinear interactions self-organize to create certain patterns at macro levels. This also allows to project future patterns based on the nonlinear interactions among heterogeneous agents.
Dynamic Modeling

The self-referential loop of theory implies that we know “what our theories are and use them to generate information that will continually challenge and upgrade our existing theories” (McMaster, 1996, pp. 28-29). Simulation is an appropriate tool to fulfill the self-referential aspect of theory. Like deduction, simulation starts with a set of explicit assumptions regarding agent behavior and interaction. Unlike the classical deduction, the modeler cannot prove theorems. Instead, simulation generates data in different simulation experiments (Axelrod, 1997b; Parker, et al., 2003). Agent-based models can provide a bridge from deductive analysis of closed systems to interactive analytic support for inductive reasoning about open systems (Bankes, 2002). Since the theoretical foundations of agent-based models are discussed in Chapter 2, here I focus on the technical aspect of agent-based models that I use in Chapter 4.

Technical Aspects of the Agent-Based Model

In agent-based models, each agent possesses both states and behaviors. States are properties, variables, or attributes. Behaviors are codified as interaction rules. Agents’ states and behaviors are most conveniently represented in software as objects (Axtell, 2000). An object is defined as something material that may be perceived by the senses or something mental or physical toward which thought, feeling or action is directed (The Merriam-Webster dictionary). From a programming perspective, an object is a software construct or module that bundles together state (attributes) and behavior (functions or methods) (Barker, 2005). This is a reason why complexity modelers say that agent-based models can model interactions between agents (having attributes and behavior rules), not just variables (attributes). Technically, I develop my
agent-based model based on the architecture of Object-Oriented Programming (OOP). Java is an example of OOP languages. Figure 7 represents a simple mechanism of Java programming.

Figure 7: Java mechanism – Java is a language that allows programmers to build a virtual reality. The virtual reality consists of agents who have certain attributes and whose behavior is based on certain rules. Each agent class has its own members as instances of the agent class.

All programming languages provide abstractions (Eckel, 2003). Any program abstracts complexity of reality that a researcher perceives into a model. The model consists of different types of agents who are modeled in different classes (in case of Java). Each object is an instance (member) of the class. According to Eckel (2003), the most important distinguishing characteristic of a class is ‘what message can you send to it?’ It can also be argued that the question like ‘what goals does the agent pursue?’ can be a criterion for a class or an agent (Simon, 1996).
In this research, an agent-based model is implemented as a method of exploring complex system behaviors. My model captures the interaction and interdependency among individuals and other entities in the delivery of public services. I have based my agent-based model within the programming structure of “Multi-Agent Simulator Of Neighborhoods... or Networks... or something...” (MASON). MASON serves as the basis for a wide range of multi-agent simulation tasks (Balan, et al., 2003; Luke, et al., 2004). The simulation model and toolkit allow me to explore the interdependency and interaction that give rise to the spatial patterns of fraud in the policy system. I named this model “Fraud Simulation.”

1.4. Significance

My dissertation idea has its roots in significant research in the social sciences (Merton, 1936; Sutherland, 1940; Huff, 1964; Newell & Simon, 1972; Simon, 1996). The contribution of this dissertation is primarily to the synthesis and implementation of their ideas for modeling the interactions and interdependency in a complex policy system, thus investigating the underlying process of macro patterns associated with criminal activities. This approach will provide a framework for better exploration of such social phenomena as crime and public health where the process is difficult to study. By successfully completing this study, I aim to contribute to three different bodies of knowledge.

Primarily, this study contributes to the field of policy analysis. While statistical and other modeling techniques have been extensively and successfully used in policy analysis, their shortcomings are well documented (Morçöl, 2002; Fisher, 2003a). These shortcomings are closely related to the complexity of policy problems in the world of politics (Moe, 1989), competing values (Quinn, 1991), interdependency (Pfeffer and Salancik, 2003), and uncertainty (Stacey, 1992). For those problems, the classical approaches of predictive modeling and
optimization are not appropriate (Schön, 1983; Bankes, 2002). By drawing upon recent work on applying complexity concepts in the natural and human sciences and operationalizing these concepts through the use of agent-based models, I explore the potential utility of these concepts and tools for public policy analysis.

Second, the prominence of agent-based models is not limited to the fact that they can implement what traditional policy analysis tools usually address. They can also be useful for exploring the implications of imperfect rationality (bounded rationality), the effects of learning, and social structure. These aspects of human and social systems have been well acknowledged in the literature (Simon, 1955; Argyris and Schön, 1974; Schön, 1983, Senge, 1990; Michael, 199716). However, it is only recently that scholars can model them (Axelrod, 1997a; Gilbert and Troitzsch, 1999; Bankes, 2002). This ability opens a door to reevaluate the concepts that the policy community has implicitly accepted and revisit the approaches that they have practiced primarily because of their technical tractability. As an example of this argument, I focus on the assumptions of rational choice models and the improvement that agent-based models can make. In this process, this study contributes to a developing area of complexity and management.

Third, deviance in a complex system is another area to which my study can contribute. Organizational deviance has a long history in social thought (Merton, 1936, 1940), and has received considerable scholarly attention in recent years.17 The literature indicates that the theoretical progress has been fragmented because there is no single accepted meaning of organizational deviance. Organizational deviance is a very broad term that covers a multitude of sins. It can be positive in the sense that deviance can be simply innovative behavior that deviated

16 The first edition was published in 1973. I cite the second edition here.

17 See Vandenburgh (2004) for a comprehensive theoretical review; Vaughan (1999) for the review of organization deviance from a sociology perspective; see Kidwell & Martin (eds.) (2005) for a recent discussion from a management perspective; see Levitt & Dubner (2005) for white-collar crimes from an economics perspective.
from the norm. However, organizational deviance usually has a negative connotation commonly associated with mistakes, misconduct, and disaster. The focus of the study is on that aspect of organizational deviance that pertains to misconduct in public service delivery systems. My study contributes to the area by modeling dynamic interactions between adaptive agents and investigating emerging patterns between corrupt agents. This effort also has obvious application to the newly developing areas of homeland security and coordinated emergency management (Nunn, 2005; Choi & Brower, 2006) as well as the more traditional areas of law enforcement and public health.

1.5. Organization of the Dissertation

This dissertation consists of five chapters. Chapter 2 reviews the literature on complexity to explore theoretical foundations of agent-based models and their implications for policy analysis. Therefore, this chapter focuses on theoretical bases and methodological insights that complexity provides. In Chapter 3, I discuss monitoring fraud in a public service delivery program using a classical store choice model and empirical data. It shows that rational choice models developed for general consumers have some efficacy even in the context of a disadvantaged group in a public assistance program. However, many shortcomings still remain. These shortcomings provide the motivation for exploring an alternative approach. Chapter 4 is an effort to address some of the limitations. I present a simulation that replicates statistical and spatial patterns in empirical data using an agent-based model. I also show computer experiments to test the effect of policy interventions for deterring fraud. This shows that the simulation

---

18 According to Vaughan (1999), disaster as organizational deviance means a physical, cultural, and emotional event incurring social loss, often possessing a dramatic quality that damages the fabric of social life.
provides a framework for further studies. Chapter 5 discusses theoretical and practical
contributions, policy implications, and research limitations.
CHAPTER 2

A SHORT ESSAY ON COMPLEXITY AND POLICY ANALYSIS

A way of seeing is also a way of not seeing.

R. K. Merton (1940)

M. Weber (1949)

A. Kaplan (1964)

Abstract

We have long known that public policy problems are “wicked” or complex. However, in order to make it manageable we have “reduced” them to the linear Newtonian-Cartesian models. Recent advances in complexity have given us conceptual framework, language, and models that provide us the capacity to implement alternative frameworks. In this chapter, I discuss the role of agent-based models as an enhanced policy analysis tool for dealing with the complexity of policy problems.
2.1. Introduction

“If a frog improves its ability to catch flies by developing a sticky tongue, the fly can respond by developing slippery feet.” (Bak, 1996, pp. 123)

A field study was conducted in day-care centers to see whether the introduction of a fine has some effects on deterring parents’ late-coming which forces a teacher to stay after closing time. The deterrence hypothesis predicts that the introduction of a penalty that leaves everything else unchanged will reduce the occurrence of the behavior subject to the fine. When they implemented this policy in the day-care centers, as a result, the number of late-coming parents increased significantly. Even after the fine was removed, no reduction occurred. (Gneezy & Rustichini, 2000, pp. 1-17)

The Shell group launched the Unified Planning Machinery (UPM) in 1967 – the planning system to end all planning systems. The UPM procedures, set out in a thick manual that all managers were supposed to follow, contained all of the elements of a state-of-the-art financial prediction system. Each year the UPM delivers its estimates of future activity, and the group as a whole bases its investment decision on those estimates. There is only one problem. Whenever times are turbulent, and anticipating the future is the most critical, the UPM is wrong. It failed to foresee the spike in oil prices in the 1970s. It failed to anticipate the collapse of oil prices in the mid-1980s. And the UPM failed to predict the restructuring of the oil business in the 1980s and 1990s. (de Geus, 1987, pp. 41-44)

There is growing recognition of the limitations of traditional linear views and optimization methods in capturing the complex reality that is the subject of public administration, policy analysis and management (Moss, 2002; Lempert, 2002; Henrickson & McKelvey, 2002; Moss & Edmond, 2004). “No model less complex than the system itself can accurately predict in detail how the system will behave at future times” (Bankes, 2002, p. 7263). This methodological reinterpretation of Ashby’s principle of requisite variety (Ashby, 1957, pp. 206-212) guides us in our reflection on current practice of policy analysis.

In recent years, complexity science has made broad claims about its relevance for policy science. However, most papers do not have properly motivated complexity models as a useful tool for policy analysis. In this chapter, I discuss the nature of policy problems using the language of complexity as well as critiques of traditional policy analysis. I also discuss a method that is being proposed to supplement or enhance tools and techniques based on statistical and optimization principles in policy analysis.
2.2. Complexity

2.2.1. Complexity Science

While its roots go much deeper (see François, 1999), complexity science emerged as an academic activity in the 1970s, gathered momentum in the early 1980s, and was enveloped in controversy by the mid-1990s (Wilson, 1998). The multidisciplinary research in modern biology, physics, economics, and computer science initiated this new wave in the U.S. (Waldrop, 1992). From the very beginning, complexity science has relied on a process of gaining knowledge from shared ideas, methods, and experiences from diverse disciplines in order to understand the irreducible complexity of reality as a whole. According to Weaver (1948), such areas as biology, medical science, psychology, economics, and political science focus on ‘the problem of organized complexity.’ They are “problems which involve dealing simultaneously with a sizable number of factors which are interrelated into an organic whole” (Weaver, 1948, p. 539). Due to the nature of problems such as complex systems, these areas may not be successfully understood when only a small number of variables are considered or explored by statistical techniques.

Complex systems are generally defined as dynamic systems that exhibit recognizable patterns of organization across spatial and temporal scales (Holland & Miller, 1991; Parker, et al., 2003). Organizations and policy systems can also be understood as complex adaptive systems where several elements in society are dynamically interrelated for some purposes. A major interest of complexity science is to study emergent phenomena in different complex systems and

---

19 There have been a plethora of theoretical studies, which argues the applicability of complex sciences in the field of policy, management, and organizations. This includes, but is not limited to, the application of chaos and complexity theory for public management (Kiel, 1994), management (Battram, 1999), governance (Amin & Hausner, 1997), knowledge management (Baets, 2005), organization (Stacey, 1996), policy evaluation (Sanderson, 2002), and health service delivery organization (Kernick, 2004).
to recognize common patterns and underlying mechanisms.\textsuperscript{20} Emergent behavior can not be predicted or even envisioned from knowledge of the properties of each component in the system (Casti, 1997). For instance, an atom does not have a temperature, but a collection of atoms has a temperature. Each molecule of sugar does not have a sweetness property, but as a whole it does. Similarly, there are no individuals in a mob, but mobs form and their spatio-temporal properties can be studied systematically. Therefore, it is suspected that the organization of the components plays an important role in the emergent phenomena of complex systems.

Simon (1996) mentioned three important stages in complexity and complex systems. Keywords in each stage are summarized in Table 1.

<table>
<thead>
<tr>
<th>Period</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between World War I and II</td>
<td>Holism, Gestalts, Creative Evolution</td>
</tr>
<tr>
<td>After World War II</td>
<td>Information, Feedback, Cybernetics, General Systems</td>
</tr>
<tr>
<td>Now</td>
<td>Chaos, Complex Adaptive Systems, Genetic Algorithms, Cellular Automata</td>
</tr>
</tbody>
</table>

Table 1: Keywords in three distinct stages of complexity – Simon, H. A. (1996). The Sciences of the Artificial, pp.169-181

As they imply, complexity science shares the holistic view of the world that humans are part of a greater whole. This holistic understanding somewhat reflects frustration with the praxis of

\textsuperscript{20} Several efforts have been documented by scholars in different areas. Examples can be found from Casti (1989), Bak (1996), Wilson (1998), Auyang (1999), Buchanan (2002), and Barabasi, (2003).
of modern scientific knowledge and technical solutions. For instance, a tenet of social and policy science has been social reform and progress. It has attempted to fulfill the tenet through the advance of science and technology with a rosy picture of total predictability. Since then, scientific progress has been a technologically mediated struggle with unending complexity. The more advanced the technology, the more problems we find. Today’s problems come from yesterday’s solutions. The problems are ‘wicked.’ In such complex situations, one is advised to not focus on only one goal. By focusing on one goal, other goals are lost.\textsuperscript{21} Complexity or complex systems view attempts to position us within the holistic context of imminent issues so as to approach them cautiously.

2.2.2. Complexity of Policy Problems

It is known that scientists who subscribe to complexity see through a different lens than the traditional positivist (Morçöl, 2002; Baets, 2005). For example, the positivist paradigm assumes total predictability. Even if we do not know reality in its entirety, it is not because it is inherently unknowable, but because our current knowledge is limited. On the contrary, complexity acknowledges the limited ability of humans to comprehend the complexity of reality and therefore requires humble reflection on the knowledge that we have built (Morçöl, 2002). “Our search is ultimately devoted not to a precise knowledge of the universe, but to a grasp of the role which we play in it – to the meaning of our life” (Jantsch, 1980, p. 310).

This leads us to the debate on whether scientific knowledge should be verifiable and empirically testable in order to establish its objectivity or if scientific objectivity is something that

\textsuperscript{21} Several scholars discussed complexity and wicked nature of problems in policy and management (Churchman, 1967; Rittel & Webber, 1973; Senge, 1990) and the failure of rational approach for such problems (Quinn, 1991; Dörner, 1996).
policy science must pursue.\footnote{There is also a plethora of discussions on the objectivity of scientific knowledge. For example, see Oreskes, Shrader-Frechette, & Belitz (1994) for discussion in natural science and Moore (1983, 2002) and Vickers (1970) in social and policy science. They provide equally good discussions on the limitation of such a concept of objectivity for scientific knowledge. Epistemology based on biology has provided a strong foundation on such an understanding. See Bronowski (1978), Maturana & Varela (1987), Dyke (1988), and Hawkins & Blakeslee (2004). Also, see Casti (1989) and Popper (2002) for the issue approached from the philosophy of science.} What does scientific objectivity mean in human and social affairs? One cannot see the world without the intervention of the physical sense. The world that one sees reflects the image of one’s sensory interpretation. Our nervous system brings forth a world in the process of cognition. The regularity of the world we experience at every moment is created each time by the neural capacity of the brain. Even if the world is out there, the world has no ‘essence’ to be discovered (Czarniawska, 1997). In that sense, interpretation, subjectivity, and social construction become an inevitable domain for policy analysis.

Recent advances in complexity science have given us a framework and language for discussing such complexity. Below, I will discuss some insights that complexity science provides to policy science, along with their critiques of traditional policy analysis.

\textit{Actors and Interactions}

Complexity defines human beings and entities as situated within the environment that consists of other human beings and entities. They are depicted as adaptive, autonomous, and purposeful actors (Nohria, 1992). These actors are adaptive in that they are “guided by information from the environment, must control its essential variables, forcing them to go with the proper limits, by so manipulating the environment that the environment acts on them appropriately” (Ashby, 1952, p. 82). These actors are also autonomous. They are “situated within a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to affect what it senses in the future” (Franklin & Graesser, 1997, p. 18).
Further, these actors are purposive in that they “deal with conscious decisions or adaptations in the pursuit of goals within the limits of their information and their comprehension of how to navigate through their environment toward whatever their objectives are” (Schelling, 1978, p. 18). Therefore, the behavior of these actors is much more complex than physical particles. Human behavior does not usually permit any simple summation or extrapolation to the aggregate (Schelling, 1978).

Traditional policy analysis has implicitly subscribed to the assumption of rational human-beings and mechanistic entities within a de-contextualized situation.23 Periodicity and generalization play an important role in this approach. Thus, a logical consequence of this approach is to seek the best alternatives and best solutions. From the perspective of complexity, the world is not a collection of isolated rational objects, but a representation of a network of autonomous and purposeful actors who are fundamentally interconnected and interdependent. Understanding human beings and entities as adaptive, autonomous, and purposeful actors highlights the limitation of the traditional view and also acknowledges the importance of the distinction between natural laws and social rules.

Living systems are partially reacting to action due to history, context, and randomness. Change is an inevitable component of living systems. The notion of fundamental laws in human and social systems may be neither feasible nor desirable.24 The concept of social laws must be interpreted differently than natural laws. “While behavior in the physical domain is governed by cause and effect (laws of nature), behavior in the social domain is governed by rules generated by the social system and often codified into law” (Capra, 1996, p. 211). Social laws are not

---

23 One of most recent books on this topic is Talbot (2005).

24 Vickers (1970) argued that “the only reason why men are by and large more predictable than the weather is that they are concerned to be predictable; concern to meet each other’s expectations by accepting common self-expectations. Also, the web of mutual expectations creates an order of which the regularities obey neither general nor statistical laws. The order is created rather than discovered, imposed rather than induced” (p.101).
immutable. They need to be seen as general explanatory statements.\textsuperscript{25} This view shifts our conceptual interest from rational objects to the interaction and interdependency among autonomous and purposeful actors in order to refine social rules related to complex human behavior. In other words, human behavior can be perceived as communication. A message is a single unit of communication. A series of messages exchanged among people is called interaction. The higher level of human communication is a pattern of interaction (Watzlawick, Bavelas, & Jackson, 1967). While social rules are designed to deal with certain patterns of human interactions, this theoretical base of social rules has not been properly incorporated in traditional policy analysis. Complexity and human interactions were assumed away in many analyses.

\textit{Knowledge and Learning}

Complexity shows that systems or processes should not be frozen and need to stay somewhere between too much and too little order. At the ‘edge of chaos’ where complexity exists, human creativity and innovation can thrive.\textsuperscript{26} Uncertainty is at the heart of creativity.\textsuperscript{27} This provides a perspective on two key components in human understanding and actions, knowledge\textsuperscript{28} and learning. Is knowledge a transferable commodity or process? Can interpretation be knowledge? Can knowledge exist independent of the human being who uses it, learns it, and transfers it? How should we understand learning in organizations or learning organizations? What does it mean for analysis?

\textsuperscript{25} For discussion on the difference between natural laws and social rules, see Hon (1999), Capra (1996), and Conte, Hegselmann, & Terna (1997). Hon’s discussion (1999) is particularly concise and insightful.


\textsuperscript{27} Biology has provided a foundation on such a view. See Jacob (1982) and Maturana & Varela (1987).

\textsuperscript{28} For discussion on knowledge from the complexity perspective, see Rescher (1996) and Baets (2005).
Knowledge is not something that exists independently from human beings or something that can simply be stored as substance or a framework. Knowledge is seen as an endless process and praxis. There is little use of knowledge in a theoretical framework if it is of no use in the dynamic world. The pragmatics of knowledge emphasizes the relation of signs to users in a context, and behavior can only be studied in the given context. The meaning of the behavior pertains to the context at a certain time. Therefore, the context, dynamic world in which one lives redefines one’s knowledge at every moment.

When environments are stable, standards and procedures are valued. Well-defined hierarchy and neutral bureaucracies are highly regarded. Uncertainty and errors are considered as something to be eliminated. When the environment is rapidly changing and uncertainty prevails, creativity and learning become crucial to effectively deal with complexity. One must be free to reflect on perplexity, confusion, or doubt in order to acquire learning. From this learning process, they become reflective practitioners (Schön, 1983) and reflexive practitioners (Cunliffe & Jun, 2005). There is no room for knowledge and learning if all aspects of social activities are pre-specified or pre-determined. Knowledge and learning are crucial factors of understanding adaptive, autonomous, and purposeful actors.

Organizations that embrace complexity are living companies (de Geus, 1987) and learning organizations (Senge, 1990; Michael, 1997). These organizations value openness and flexibility. They embrace uncertainty and error as learning opportunities. For example, it is suggested that organizations focus on recognizing patterns and building networks to amplify positive feedback rather than trying to achieve optimal performance at all times. In analysis, therefore, an optimal policy based on a best estimate model may not be robust across the range of

---

29 A long tradition of such a view is found from Morris (1938), Watzlawick, Bavelas, & Jackson (1967), and Maier, Hadrich, & Peinl (2005).

possible behaviors of complex adaptive systems (Bankes, 2002). These analyses oversimplify issues and limit learning opportunities.

Further, depending upon whether one observes the issues from inside or outside a particular system, the quality of understanding and description are significantly different.\(^{31}\) When one is part of the system, trial and error, adjustment, and change of one’s thoughts and actions become crucial ingredients for better decision-making. Analyses from outside the system may not fully capture the complexity with which these organizations deal.

*Process and Emergence*

It has been implicitly believed that the basic process of nature is deterministic and reversible. Once the particular state of a system is measured, the reversible laws are supposed to determine its future. Therefore, the emphasis of the studies is on time-independent laws. This concept of the simple world governed by time-reversible fundamental laws has been questioned (Pierce, 1961; Prigogine & Stengers, 1984). Reversibility and determinism may apply only to limiting and simple cases. Data, processes, and issues are dependent upon their epoch and upon the forms of process dominant in the time. Irreversibility and randomness are rules in nature and lie at the origin of most processes of self-organization (Prigogine & Stengers, 1984).

Traditional policy science as a search for simplicity contributed to our understanding of policy-making. Nevertheless, there is a need for comprehending the nature of complexity to improve policy processes. Public policy can be seen as an emergent phenomenon (Morçöl, 2003). During implementation, policy is interpreted and enacted based upon the interpretation. Public policy is not reducible to the original intentions of its initiators or the text of the law. Once it

\(^{31}\) This argument is found from critiques on traditional system and management theories (Senge, 1990; Stacey, Griffin, & Shaw, 2000), and also from a discussion on relativity in physics (Casti, 1997).
emerges, public policy does not stay the same, but it constantly evolves. This revisit of *Implementation as evolution*[^32] is rooted in the philosophical tradition of Heraclitus who realized that one cannot step twice into the same river.

Advances in modern science such as chaos theory and quantum physics have provided examples that show that one cannot forecast the future based on the past. Two versions of the impossibility of prediction with certainty are found in the literature. One version starts from Poincaré, who said that small differences in the initial conditions generate very large differences in the final phenomena, to Merton (1936), Lorenz (1964), and Baets (2005). The other version is from Gell-Mann (1994), who stated that even in the classical limit and even when the laws and initial conditions are exactly specified, indeterminacy can still be introduced by any ignorance of previous history… whereby future outcomes are arbitrarily sensitive to tiny changes in present conditions. This implies that there is not always a best solution. Best principles might provide better insights for policy design.

The evolving nature and emergent property of public policy make us reflect on current practices of policy evaluation and evaluation methods. The causal relationship in nature is considered to be a scientific achievement. The relationship between social programs and outcomes may not be the same because programs and outcomes co-evolve. What has passed is gone forever. Therefore, even if the causal relationship between a program and an outcome was present at one time, the relationship might be true only in the given time and under the assumption that the context was fully considered. However, we still do not know how to fully include contextual complexity in our study.

Therefore, some argue that instead of searching for causality, the concept of synchronicity may provide better insights on business dynamics (Peat, 1987, 2002; Baets, 2005).

[^32]: Majone & Wildavsky (1973). This paper was republished in *Public policy: the essential readings* edited by Theodoulou & Cahn in 1995.
Synchronicity refers to ‘meaningful coincidence, significantly related patterns of chance,’ (Peat, 1987) or ‘being together in time’ (Baets, 2005). The notion of synchronicity is indebted to Flatland (Abbott, 1963; Stewart, 2001). Complex objects that fall through flatlands “may look like a correlated series of events that are separated in space” (Peat, 1987, pp. 116-117). Normal events in the three-dimensional world may look like synchronicities in the flat world of two dimensions. One may not be able to draw the whole picture of the complex object at this moment, but one can still catch and follow the flow at the moment of synchronicity. This alludes to the idea that business and organizational performance may be improved by learning how to take opportunities when synchronicity is presented, even if one can not draw the whole picture.

The practice of policy analysis is still under the influence of positivism (Durning, 1999; Fischer, 2003b). However, a massive paradigm shift has conceptually been noticed. This shift implies changes in our world view, from a simple to a probabilistic world, from hierarchy to heterarchy, from mechanistic to holographic universe, from deterministic to indeterministic view, from direct to mutual causality such as symbiosis and nonlinearity, from metaphor of assembly to morphogenesis (creation of new forms), and from pure objectivity to perspectival (Lincoln, 2005). The basic assumptions in complexity are well aligned with the changing world view. The complexity views are compared with the traditional positivistic approach in Table 2.
<table>
<thead>
<tr>
<th></th>
<th>Positivistic Approach</th>
<th>Complexity Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td>Mastering nature</td>
<td>Conversation with nature (Jantsch, 1975; Prigogine &amp; Stengers, 1984)</td>
</tr>
<tr>
<td><strong>Ontology</strong></td>
<td>Substance</td>
<td>Process (Rescher, 1996, 2000)</td>
</tr>
<tr>
<td><strong>Search</strong></td>
<td>General Law</td>
<td>Principles or Rules (Hon, 1999), Synchronicity (Peat, 1987; Baets, 2005)</td>
</tr>
<tr>
<td><strong>Models</strong></td>
<td>To Predict</td>
<td>To Anticipate (Holland, 1998) or understand</td>
</tr>
</tbody>
</table>

Table 2: Basic assumptions of complexity were compared with those of traditional positivistic approach.

2.3. Managing Complexity

What does this understanding of complexity of policy problems mean for policy analysis?

In general, policy questions address two things. What purposeful actions should we take to improve human conditions? What consequences can we anticipate from the actions? If the present becomes self-destructive and the predicted future is barely realized, then where is the role of scientific evidence based on retrospective experience for decision-making or policy analysis?

Over the past few decades, alternatives to the positivist approach have been making some inroads. For example, there has been an effort to understand the argumentative and narrative
nature of policy analysis (Fischer & Forester, 1993). What policy analysts do is to provide policy arguments. This argument is not separated from value judgments and is used by policy makers during public debates. Methods and conditions to improve policy debate and discourse are more important than objective measures (Majone, 1989).

While these perspectives are intellectually stimulating, their applicability in policy analysis remains unclear. How can we use some of these ideas on evaluating public programs and processes? We know how to measure efficiency, effectiveness, and productivity within the framework of traditional policy analysis. However, we do not have many clues regarding how to model complexity and design our future in this uncertain world in order to eventually improve future performance.

2.3.1. Social Complexity and Algorithms

In its most simplistic form, complexity science searches for algorithmic rules that can represent complex patterns in reality. Algorithms are a set of simple rules that repeat over and over again (Peat, 2002). Mandelbrot (1997) has shown that algorithms can produce ‘fractals’33 which are scale-free self-organizing complex patterns in nature. Fractal geometry shows that seemingly complicated shapes and patterns can arise from simple and humble beginnings. Coastlines, clouds, and trees are physical examples of fractals. These natural shapes display self-similarity on many scales. They are made up of many smaller copies of themselves (Stewart, 2001). In other words, each part of a shape is geometrically similar to the whole. Although it may not be true for all patterns, simple mathematical rules can lead to such complicated patterns.

33 Mandelbrot (1977) coined ‘fractals’ from Latin adjective, fractus, and verb, frangere, which means to ‘break’ to create irregular fragments (p.4).
When we do not know whether there are any underlying rules, patterns will help us to look into them because they are not just a random mess (Mandelbrot, 1977; Grimm, et al., 2005).

“Patterns are observations of any kind showing nonrandom structure and therefore containing information on the mechanisms from which they emerge.” (Grimm, et al., 2005, p. 991)

The notion of fractals may not be limited to nature. We can also draw an analogy of fractals for social complexity. For example, collaboration, coordination, and diffusion present enough complexities and irregularities. Nevertheless, certain mechanisms can be found in these behaviors (Axelrod, 1984, 1997b). For example, the traditional notion of organization has been presented as a black box that has complex impersonal relationships and standardized procedures mainly in a hierarchical structure. Recent studies are opening the black box with relatively simple mechanisms such as ‘Sense-and-Respond’ (Haeckel, 1999) and ‘Request-Execution-Delivery’ (Ramanathan, 2005). Organizations are made up of self-similar copies of these mechanisms on many scales. These mechanisms have also been keys to understanding the complexity of global supply chains and health service delivery systems.34

2.3.2. Computational Social Science Models

Computational social science models are developing to simulate such social complexity using computer algorithms. The original ideas of computational models for social complexity can be found in Weiner (1948, 1954), Ashby (1952, 1957), Newell & Simon (1972), and Simon (1996). In cybernetics, it is assumed that human cognition is not different from information

34 Communications of the ACM (the Association of Computing Machinery) published a special issue on adaptive complex enterprises in May 2005. See Jones & Deshmukh (2005) for the application of complexity to supply chain management and Tan, Wen, & Awad (2005) for the application of chaos theory to health service delivery.
processing in computers. Human cognition works through the manipulation of symbols based on a set of rules. Communication is a transmission of information. This simplistic view of humans has been criticized by biologists in the 1970s (Capra, 1996). Whether human cognition can be duplicated by a computing machine is still in question (Dreyfus, 1972; Casti, 1997).

Newell & Simon (1972) further extended such an idea of cybernetics for studying human problem solving. When they explored how integrated activities constitute problem solving in such tasks as chess and puzzles, information processing theory provided a foundation of their understanding of what symbols and symbol manipulation can do for us. Later, Simon (1996) distinguished three important components of ‘the artificial’ and specified their relationships. The three components are: the purpose (goal), the character of the artifact (inner environment), and the environment in which the artifact performs (outer environment). In Simon’s (1996) framework, social complexity is a result of the adaptive interaction between the artifact and its outer environment, rather than from some inner complexity within the artifact.

“Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves.” (Simon, 1996, p. 53)

This idea is known as Simon’s Conjecture on which the epistemology of computational social science is based (Cioffi-Revilla, et al., 2004). This conjecture allows us to anticipate certain behaviors from the knowledge of goals and its outer environment, even when we have very minimal knowledge on the inner environment, such as physical properties and characters of the artifact. For instance, uneven traffic flow as collective behavior emerges, mainly due to the complexity of the environment, such as random positioning, signals, and distance between cars (Resnick, 1994), rather than the capacity of the engines, models, and types of the cars. If the organization of some components, rather than their properties, largely explains social complexity, the consequences of alternative organizational assumptions for human behavior can be explored
using computer agents. The computer agents are organized somewhat in the image of man by having properties and showing certain behaviors (Newell & Simon, 1972; Simon, 1996). The behavioral assumptions can be codified into a set of algorithmic rules.

Complexity theory is a multidisciplinary science. It takes into consideration elements of very different disciplines, such as cybernetics, systems theory, chaos theory, artificial intelligence, artificial life, cognitive sciences, computer science, ecology, economy, evolutionary biology, games theory, linguistics, philosophy, social sciences, and management. I simplified the notion of complexity in many areas with fractals. Certainly, social complexity belongs to, overlaps, and goes beyond complex patterns in nature. As several complexity models have been developed and introduced, there has been an effort to simulate social complexity using computational models. In many cases, they aim to understand and model values, interactions, uncertainty, learning, process, and emergence in social complexity using simulation models (Figure 8).

![Diagram](image.png)

Figure 8: Modeling fractals and social complexity to inform policy decisions
2.3.3. Development of Agent-Based Models

An agent-based model is a recently emerging modeling technique within the tradition of computational social science models. In this chapter, I interchangeably use three different names: complexity models, computational models, and agent-based models. These models are the same in that they codify organizational assumptions as computer programs (algorithms), and the inference is performed by executing the program (Edmond, 2001). Complexity models are a synonym of complexity science, and they are not separable from each other. Computational models are given explicitly in mathematical terms or implicitly by coding the relationships among the variables and rules constituting a computer program (Casti, 1997). An agent-based model is a computational model in that it implicitly codes the interdependency of agents and action rules using symbols of programming language.

However, agent-based models are different from classical simulation models. If macro-simulation in the 1960s used sets of differential equations for macro-level forecasting, micro-simulation in the 1970s used the individual as the unit of analysis for macro-level forecasting. However, individuals do not directly interact or adapt in these simulations (Marcy & Willer, 2002). The basic variables determining the outcome of decisions are aggregated quantities rather than the actions of individuals (Casti, 1997). By making agents behave based on their own properties and interaction rules, users of agent-based models are more interested in theoretical bridges between micro- and macro-levels and in gaining insights than in mathematical solutions.

Characteristics of Agent-Based Models

Agent-based models go beyond deductive analysis of closed systems to provide interactive analytic support for inductive reasoning about open systems (Bankes, 2002). Agent-
based models can be implemented as a method of exploring complex problems. The advantages of agent-based models include the ability to accommodate various differences among individuals, to simulate complex decision-making by an individual, and to address interactions over time and space (Gimblett, 2002). Repetitive competitive interactions among agents are a feature of agent-based models. Even a simple agent-based model can exhibit complex behavior patterns and provide valuable information about the dynamics of the real-world system that it emulates (Epstein & Axtell, 1996). Agents may be capable of evolving and allowing unanticipated behaviors to emerge. Sophisticated agent-based models sometimes incorporate neural networks, evolutionary algorithms, or other learning techniques to allow realistic learning and adaptation (Bonabeau, 2002).

Agent-based models consist of agents and action rules. Agents are the basic unit of action in simulations and are specified by defining the complex system studied and specifying the interdependency of system components. Agents can be humans, institutions, robots, computers, objects, concepts, and even ants. They are heterogeneous and autonomous with behavior that can be rational, adaptive, and random in response to the environment. Learning occurs through the adaptive behavior (Ashby, 1952) and thus influences future decisions. Action rules reflect organizational assumptions in complex systems. The rules determine how the agents interact. Flexibility in designing new action rules in a simulation allows researchers to test alternative assumptions underlying complex social phenomena in the simulated reality (Resnick, 1994; Simon, 1996; Marcy & Willer, 2002).

Above all, agent-based models aim to enrich our understanding of fundamental processes that may appear in a variety of systems and to support our intuition on the target system (Axelrod, 1997a; Edmonds, 2001). Once we model patterns and processes as a dynamic system, we can test some options for purposeful actions. This will allow us to approach complex issues with the awareness of consequences.
Illustration and Examples

Agent-based models have been used for many purposes, such as modeling emergence (Holland, 1998), catastrophic phenomena, far-from equilibrium behaviors (Bak, 1991, 1996), constructivist learning and challenging assumptions (Resnick, 1994), virtual laboratories (Casti, 1997), technological or engineering applications, and planning. Here I present three policy-relevant examples for which agent-based models were used: Schelling’s segregation, Axelrod’s computer games, and an artificial world to address traffic congestion.

Schelling’s segregation model (1978) represents one of the first constructive models of a dynamic system that is capable of self-organization based on simple rules. In Figure 9, the simulation shows that initial random agents that a computer generated (left figure) segregated after a certain time (right figure). Agents were represented using each cell with two different colors. Neighbor cells are environments of agents. The action rule used in the simulation is as follows: “an [blue or red] agent decides whether it wants to move. It scans all of the neighbors and sums the total number of similar agents. If the sum is below the threshold, then it moves.” The simulation did not specify detailed properties of the agents other than their colors and the action rule. Yet, the simulation reaches the equilibrium of segregation.
Figure 9: Schelling's segregation model was conceptually introduced in his *Micromotives and Macrobehavior* (1978). Simulation above was captured by running a demo in MASON.

In their early stages, some modelers presented complexity without altering the rigid assumptions on agents (Axelrod, 1984; Axtell, 1999). Later, their interests were broadened to organizations and social systems. This resulted in relaxed assumptions regarding rationality (Axelrod, 1997b; Axelrod & Cohen, 2000). For example, Axelrod (1984) studied under what conditions cooperation will emerge when egoists compete in a game without central authority. The study showed that the norm of reciprocity made it possible for cooperation to emerge (Axelrod, 1984). In other words, cooperation emerges because of the possibility that the players will meet again even when the assumption of self-interest rational agents is not abandoned. In a later study on the emergence of norms as solutions to dilemma of collective action, Axelrod
(1997b) chose to implement an evolutionary approach. In this simulation, the initial strategies are chosen at random, and strategies also undergo some random mutation. Agents no longer need to be rational. Players are given the opportunity to defect and to punish the defections they observe. The study identified ‘metanorm’ (the treatment of non-punishment as if it were another form of defection) as a mechanism that could sustain a partially established norm.

If the former focuses on understanding the underlying processes of social phenomena, there are also examples of artificial worlds. Many problems that policy deals with require intervention. However, scientific, repeatable, and controllable tests on human subjects are not easy due to ethical, theoretical, and practical issues. Artificial worlds can be a laboratory for testing policy interventions. Chris Barrett at Los Alamos built TRANSIMS in order to tackle Christmas-shopping congestion on Louisiana Boulevard in Albuquerque, New Mexico (Casti, 1997, pp. 131-142). The main questions in the simulation were how a proposed change in the system creates traffic patterns and how these patterns impact the environment. The structure of the artificial world consists of travel demand and transport system data, trip route plan generation, traffic micro-simulation, and environmental simulation. This example shows another usage of agent-based models in the context of policy analysis.

2.3.4. Some Perspectives on Policy Modeling

I discussed the conceptual basis and models of complexity for policy analysis. Based on this discussion, I draw some perspectives on policy modeling. Traditional policy analysis has been effective in identifying measurable factors that are presented with some patterns at a certain time and context. However, many of the underlying processes that give rise to such patterns are hidden. Conventional research methods and tools have limited utility in studying such hidden processes. For example, statistical tools are built on the principle that one can make inferences
about a population based on samples. Statistical analysis presents inductive and historical facts with assumptions specific to statistical techniques and data rather than processes. Economics models and techniques are built upon the basic assumptions of economics. The relationships identified using economic theories or tools do not necessarily mean that the process leading to the relationship followed such assumptions. Linear programming is genuinely solution-oriented in a given context. As the conditions change, the solutions change. What does not change is the assumption behind this tool, which is looking for optimal solutions subject to certain conditions.

As assumptions in social theories easily conceal individuals’ cognitive capabilities and values, assumptions behind analytic tools and techniques also do so. By selecting an analytical tool, one implicitly subscribes to the values on which the tool is based. Understanding the process may reveal much richer stories and make individuals reflect on the underlying assumptions.

Given that policy analysis requires substantial contextual knowledge as well as scientific knowledge for conscious actions, policy analysis tools need to be flexible in incorporating contextual knowledge in many different policy settings. In other words, policy analysts build testable models to represent the patterns they observe. The models are continuously revised until they align well with the observed pattern. Multiple theories and assumptions need to be tested in the process of modeling. This process will improve the analyst’s own understanding of the reality as well as its representation for others. This model-based epistemology (Figure 10) mimics the way people think and allows for investigating analyst’s own assumptions, values, and beliefs of the dynamic worlds that they experience. Therefore, it provides a chance for policy analysts to reflect on what they believe, what they value, and what they do.
Figure 10: Comparison between social science models and policy analysis - the basic mechanism of policy analysis is not different from social science models. However, decision-making and action are crucial components to successful policy analysis. Therefore, policy analysis needs to continuously iterate and revise the whole process to support decision-making and purposeful actions in different contexts.

A crucial aspect of policy analysis is to inform practice even when only partial knowledge exists (Murray, 1983). Policy and decision-making cannot be delayed until social science is able to explain all social processes (Moore, 2002). Policy questions do not necessarily go together with analytic questions in a discipline. Actions need to be taken or not taken, which is fundamentally a policy question. In that sense, policy arguments and debates have prospects for policy analysis. The fact that there are different views and values surrounding policy and decision-making should not discourage people about the role of policy analysis. It only implies
that there are demands for an enhanced analytic and synthetic approach that can incorporate different interests and values into analysis. We must be able to evaluate evidence within a larger spectrum of experience through analysis so as to facilitate dialogue among stakeholders.

2.4. Conclusions

The prominence of agent-based models is not limited to the fact that they can implement what the traditional approach has successfully addressed. These models can also serve to explore the implications of imperfect rationality (bounded rationality), the effects of learning, and social structure. While these aspects of human and social systems have been well acknowledged in the literature,\(^{35}\) it is only recently that scholars are able to develop operational models.\(^{36}\)

Early simulation and modeling had difficulties finding the right agents and describing the interactions among these agents (Casti, 1997), as well in modeling emergence and surprise. Many of the obstacles were overcome with the advance of technology and the progress of knowledge in other disciplines. Today, we can create artificial worlds in our computers. We can conduct repeatable scientific experiments on complex systems. We can even address the consequences of a policy intervention. It seems that we are ever closer to ‘the sciences of the artificial’ (Simon, 1996). I believe that this is the right time to bring up this advance in modeling and advantages for the policy community so as to enrich policy analysis.

In line with this argument, I present two independent studies in the following chapters. These studies were designed within the research framework in Chapter 1, in order to implement

---

\(^{35}\) There is no shortage of this understanding in organization literature (Simon, 1955; Argyris & Schön, 1974; Schön, 1983, Senge, 1990; Michael, 1997).

\(^{36}\) Many of the methodological papers have recently been published (Axelrod, 1984, 1997b; Gilbert & Troitzsch, 1999; Bank, 2002; Lempert, Popper, & Bankes, 2004).
some of the ideas that I discussed in this chapter. Chapter 3 is an empirical study to provide some background about a policy context. Chapter 4 is a simulation to explore an alternative approach for policy analysis. Nevertheless, the objective of both studies is the same in that I aim to investigate the interactions among different players (vendors and participants) in a policy delivery program (Ohio WIC) for different purposes. I investigate static interactions between vendors and participants to use for identifying high-risk vendors in Chapter 3. Their spatial interactions are captured at a certain point in time, to develop vendor propensity measures. Statistical and spatial patterns associated with fraud are identified. This study serves as a background of a simulation in Chapter 4. I replicate the statistical and spatial patterns of fraud in empirical data using the simulation. I also present an example that shows the utility of this model by testing the effect of a policy intervention in deterring fraud. I achieve this objective by capturing dynamic interactions among vendors, participants, and a public agency in the simulation.
CHAPTER 3

IDENTIFYING FRAUD IN A PUBLIC DELIVERY PROGRAM

Advance is partly the gathering of details into assigned patterns.

A.N. Whitehead (1938)

Abstract

The nature of fraud in public delivery programs is different from fraud in the private sector. Fraud in public delivery programs more often requires the active participation of the program participant as well as the service provider. In this chapter, I utilize a store choice model to develop statistical measures that can capture the interaction between participants and service providers, thus providing a practical tool for public managers. I show that this tool can be utilized for monitoring fraud in public delivery programs for the purpose of ensuring public accountability.

---

37 This study was conducted when the author was working as a researcher for the Ohio Women, Infants, and Children (WIC) program between 2003 and 2005. This study was presented at the 27th Annual APPAM Research Conference in 2005 and approved for publication without identifying county and vendor names. It has benefited from discussions with Ohio WIC vendor specialists. Special thanks to Pam Hamilton, Todd Reeves, Robin Ridenour, Nicole Wilson, and Yongwan Chun. The author also thanks to Anand Desai, who read the paper and provided valuable comments.
3.1. Introduction

The basic premise of this chapter is that individuals, in general, obtain their services from providers in close proximity to where they live or work. Using this premise as the starting point, I argue that spatial models which go beyond the deterministic use of the concept of proximity can help detect the patterns that ensue from these habits and deviations from such patterns are anomalies that could be indicative of unusual, perhaps illegal, behavior.

Fraud is defined as a misrepresentation of asset values (Sutherland, 1940). Fraud is a well-known problem in public delivery programs. The misuse of public assistance programs is well documented (GAO, 1998; USDA, 2001, 2003). For example, fraud, waste, and abuse are serious issues in the Unemployment Insurance Program (Heddell, 2002), the workers’ compensation system is vulnerable to fraud (Texas Department of Insurance, 2001), and fraud appears in home health care programs as well (Vandenburgh, 2005). Fraud in public delivery programs creates a double burden for taxpayers by increasing the cost of programs and by damaging the integrity of public programs.

Monitoring fraud is challenging due to the peculiar nature of fraud in public delivery programs. First, these programs are implemented as a system involving several loosely coupled entities, such as state agencies, local agencies, contract agencies, and benefit recipients. When fraud is committed in such a system, the disturbance takes a long time to affect other entities or the whole system (Weick, 1979). Second, the process of fraud, by its very nature, is difficult to detect and to study using traditional research methods that focus on single cause and effect relationships. The focus of much of the literature on fraud is on that which has been detected and the underlying processes revealed and not on detection and prevention. Third, fraud is dynamic in nature. Once a fraud detection method is in place, it immediately begins to lose its effectiveness.
Therefore, monitoring fraud is not just a case of identifying and dealing with a set of simple deviance issues.

**Fraud in Food Assistance Programs**

The Food Stamp program and the WIC program are well-known public delivery programs implemented by USDA. Currently, these programs together spend over $35 billion (Food Stamps: approximately $31 billion and WIC: approximately $5 billion in Fiscal Year 2005) to provide food and nutritional supplement for the needy. For such large programs, even a small fraction adds up to a substantial monetary loss at the aggregate level (Bolton & Hand, 2002). For example, between 1999 and 2002, trafficking (the exchange of public service benefits for cash) diverted about $395 million per year from Food Stamp benefits (Macaluso, 2003), which is a loss of 2.2 percent of total costs of approximately $18 billion during these years.

While in the private sector, theft or “shrinkage” is mainly due to customers or employees acting as individuals, fraud in public programs more often requires the active participation of the program participant as well as the service provider, such as vendors. Vendors are private entities that contract with public agencies to deliver public services. In such a loosely coupled system, participants can commit fraud with or without malicious intentions, but its feasibility is enhanced by complicity on the part of vendors.

For instance, a study of WIC retail vendors (USDA, 2001) showed that 18.1 percent of these retail vendors were involved in overcharging violations one or more times, including 12.4 percent that overcharged once, 4.2 percent that overcharged twice, and 1.5 percent that overcharged in three separate investigations. For vendors in the WIC system, overcharging is the main source of fraud. Other illegal activities include forcing unwanted purchases, substitutions, and trafficking.
There are a variety of fraud detection methods in practice. In general, fraud detection methods can be categorized into statistical analyses and field investigations (Bolton & Hand, 2002). For example, the focus of traditional WIC fraud detection methods has been primarily on identifying high-risk vendors, using criteria such as store characteristics, sales volume, participant complaints, and field investigations. High-risk vendors are those who have high probability of committing vendor violations (USDA, 2000). Vendor violations are not limited to administrative violations, but also include illegal monetary transactions. The violations are identified through comparison of observed data with expected values or by conducting field investigations.

While these methods have widely been used, they are static and less effective in uncovering the interactions between corrupt vendors and participants and easy to evade because of familiarity with detection procedures. For example, these procedures are fairly simplistic, focusing on small vendors, large sales volume (outliers) or vendors against whom complaints have been received. Field investigations are time consuming and expensive and easy to avoid by vendors. Therefore, the standard control mechanisms do not mirror the complexity of the issue. In other words, simple fraud detection efforts do not match the dynamic and adaptive processes that they are supposed to monitor and regulate. Hence, there is an on-going need for alternative frameworks and methods to address complex and non-routine problems in public management.

In this chapter, I explore the use of store choice models to monitor fraud in a public delivery program. In particular, my objective is to develop statistical measures that can capture the interactions between vendors and participants in the program, thus providing a practical tool for public accountability and performance management.
3.2. Store Choice

3.2.1. Store Choice Models

The literature on how consumers choose stores is vast. These models can be categorized into three approaches based on normative assumption, revealed preference, and direct utility (Craig, Ghosh, & McLafferty, 1984).

The early models rely on observations and normative assumptions regarding consumer behavior. For example, the nearest center hypothesis assumes that consumers patronize the nearest store that provides the required goods and services. Gravity models are developed as an analogy to Newton’s theory of gravitational attraction. The degree of attraction between two stores is based on the size of the stores and the distance between them. The main criticism of this approach is that it does not take into consideration consumer behavior and patronage.

While the gravity models use a standard Newtonian attraction to predict consumer store choices, the revealed preference approach uses past behaviors to explain consumer store choice. Huff’s spatial interaction model (1964) occupies a unique position between gravity models and the models based on the revealed preference. The Huff model incorporates two important principles. First, the patronage area of the consumer is probabilistic rather than deterministic. Second, the probability of a consumer visiting a particular store is a relative measure equal to the ratio of the utility of that store to the sum of utilities of all the stores considered by the consumer (Huff, 1964; Huff, 2003). In the model, as in the gravity models, the utility of the store is based on store size (as the amount of floor space) and distance from the consumer.

The Huff Spatial Interaction Model may be expressed as
\[
p_{ij} = \frac{U_{ij}}{\sum_{j=1}^{J} U_{ij}} = \frac{S_j^\alpha \cdot D_i^\beta}{\sum_{j=1}^{J} S_j^\alpha \cdot D_i^\beta}
\]

where, 
- \( p_{ij} \): Probability of consumer \( i \) visiting store \( j \); 
- \( J \) is the set of competing stores in the region 
- \( U_{ij} \): Utility of store \( j \) for consumer \( i \); 
- \( S_j \): Size of store \( j \); 
- \( D_{ij} \): Distance between consumer \( i \) and store \( j \); 
- \( \alpha, \beta \): Parameters

Empirical studies have found the Huff model to be reasonably accurate in predicting the market share of shopping centers (Craig, Ghosh, & McLafferty, 1984; Yrigoyen and Otero, 1998). However, this model has been criticized for its simplicity and the context-dependency of parameter estimation.

Extensions of the Huff model have included other store attractiveness attributes such as service or product quality, image, price, availability of parking (Craig, Ghosh, & McLafferty, 1984; Drezner & Eiselt, 2002), loyalty (Knox & Denison, 2004), and reputation (Ou, Abratt, & Dion, 2006). Two commonly used extensions are the Multiplicative Competitive Interaction (MCI) models and the Multinomial Logit (MNL) models. MCI models were introduced by Nakanishi & Cooper (1974, 1982) and Cooper & Nakanishi (1983), to study retail location in a competitive environment (Ghosh & Craig, 1983). MNL models with discrete choice data have been used to estimate store choice probabilities (McFadden, 1980; Hauser, 1980; Gaver, 1980; Suárez, et al., 2004).

Many of the early applications of the model stayed close to the physical science interpretations, including the Newtonian analogy where the square of distance is the appropriate power function. However, larger exponents were introduced to indicate the friction of distance becomes increasingly important in reducing the expected level of interaction between two
locations (Haynes & Fotheringham, 1984). Extensions of the Huff model that include varying the distance-decay parameter (Eppli & Shilling, 1996; Haynes & Fotheringham, 1984) or allowing for variable size have found that size has a greater influence on store choice than distance (Eppli & Shilling, 1996).

Models that incorporate a revealed preference approach seem to perform better than those based on a normative approach. However, some problems still remain in that consumers reject stores beyond a certain distance, and parameters estimated in the models are context-dependent. Direct utility was introduced as an approach to overcome context-dependency by calibrating parameters from simulated choice data using hypothetical store descriptions (Craig, Ghosh, & McLafferty, 1984).

Applications of these models to a policy context raise a number of other issues (Moore, 1983; 2002) regarding the application of store choice models to contexts where poor people are seeking services. Calibration of such models is context dependent particularly for policy analysts whose goals are very different from wanting to maximize profit or market share. While it may seem that calibration of these models is best done with real data, that also raises a number of issues pertaining to the quality of the data\(^{38}\) and their availability (Huff, 2003, Murray; 1983; Tilly, 2006).

\(^{38}\) The quality of data implies not only the messiness of data, but also the fundamental nature of data in the social science. Tilly (2006) argued that people negotiate reasons when they were asked to provide the reason of certain behavior. Therefore, we may have to ask a fundamental question on the nature of data which social scientists collect and analyze by asking for reasons for their choices.
3.2.2. Applications of the Store Choice Models

The traditional uses of store choice models vary. They include trading area analyses (Berman & Evans, 1995; Houston & Stanton, 1984), spatial accessibility analyses (Guagliardo, 2004; Kwan & Weber, 2003; Fortney, Rost, & Warren, 2000), and location allocation decisions (Drezner & Hamacher, 2002; Serra & Colome, 2001; Benoit & Clarke, 1997). These spatial interaction models were frequently used to locate businesses and other social infrastructure amenities.

The main difference between the public and private sector models is to be found in the objective function of the models (Revelle, Marks, & Liebman, 1970). The objective in public sector applications is to maximize social utility or to minimize social costs (Marianov & Serra, 2002). On the other hand, the objectives of location decisions in the private sector applications are to minimize the total costs of transportation, to maximize profits or market share.

For example, public facility location models have focused on building social infrastructure such as schools, emergency centers, and health clinics to meet the basic needs of the population (Rosenberg, 1984; Hansen, Peeters, & Thisse, 1983; Dear, 1978). These models help identify, within a set of policy constraints, better location choices for providing program benefits and goods to spatially dispersed populations. Spatial interaction models have been used to secure spatial accessibility (Kwan & Weber, 2003; Guagliardo, 2004). Private sector location models have been developed to choose locations in competitive markets for franchises, grocery stores, and convenience stores (Drezner & Eiselt, 2002; Leszczyc, Sinha, & Timmermans, 2000; Houston & Stanton, 1984). The most common application of the store choice models has been to the analysis of trading areas and the location of retail outlets.

Consider for the purposes of illustration the hypothetical spatial patterns of store choice or location decisions emerging from an application of the Huff model. In Figure 11, squares
represent stores along with their size (between 1 and 20), and circles represent customers. There are 10 stores and 2,000 customers shown in the rectangular region. If the customer chooses stores that are allocated strictly based on the Huff model (with size parameter, 1 and distance parameter, -2), spatial patterns will look like those in Figure 1. This is an example where the model was deterministically used. This basically shows the simplified spatial pattern of trading area analysis (Berman & Evans, 1995).

Figure 11: Simulated theoretical spatial patterns – this spatial pattern shows several clusters or trading areas that each store may serve when the Huff model is deterministically used. Numbers inside squares represent the number of lanes in each store.
In spite of the limitations of the store choice models, they provide a simple starting point to explore the differences between the predicted and actual choices of participants selecting vendors provide public services. One possible explanation for anomalous patterns could be fraud and that the vendors that have these anomalous patterns could be at high risk for committing fraud.

To test this idea, I construct vendor propensity measures based on the interactions between vendors and participants. After calculating predicted store choice probabilities of all participants for all vendors, I summarize statistics based on observed store choices of participants at each vendor. This will capture the general propensity of vendors in terms of the kinds of participants they interact with, at least, probabilistically. This exercise serves two purposes. First, it provides a resource-effective method for identifying high-risk vendors in practice, thus helping to monitor fraud. Second, it helps with identifying the spatial patterns between high-risk vendors and participants for further investigation in a given context.

3.3. Method and Data

3.3.1. Vendor Propensity Measures

I begin the implementation of the Huff model by using standard values for the parameters, that is, $\alpha =1$ for size and $\beta =-2$ for distance. I assume that participants’ store choices are mainly decided by whether the store is close enough and whether they can save money. There is a higher chance of saving money on other grocery items in larger stores, such as national chains, than in convenience marts. The first criterion is reflected in distance and the second criterion is mainly influenced by store size, and both are incorporated in the model.
I use the number of checkout lanes as a proxy measure for store size. Locations were identified using GIS. Euclidean distances (not travel distance) between vendors and participants were calculated using Matlab. After I obtained two critical factors, store size and distance, I calculated participants’ vendor selection probabilities and constructed a probability matrix among all participants and vendors. Hence, I know both the model’s estimated probability and actual behavior of the participants in their visits to the vendors.

The steps involved are as follows:

**STEP 1**: Collect data on the location of vendors and participants and other relevant information regarding vendor and participant characteristics to be included in obtaining vendor propensity measures.

**STEP 2**: Compute the store choice probability, which is the theoretical probability \((p_{ij})\) of participant \(i\) shopping at vendor \(j\) based on the characteristics.

**STEP 3**: Create \(p_{ij}\) matrix for all vendors and participants.

**STEP 4**: Group participants’ \(p_{ij}\) s by vendor where the participants actually redeemed the benefits.

**STEP 5**: Calculate statistics (minimum (MIN), maximum (MAX), and average (MEAN) values) of \(p_{ij}\) s in each vendor to use as vendor propensity measures.

**STEP 6**: Examine the spatial patterns of the vendors.

**STEP 7**: Explore possible explanations of vendors showing abnormal spatial patterns.

For any vendor, MEAN is the average propensity that a participant will visit that vendor and MAX is the highest probability of a participant visiting that vendor. Ideally, one would expect participants to visit vendors with high MIN and MAX values. A low MIN, and therefore low MEAN, implies that the vendor attracts participants who have a low probability of visiting
the vendor, according to the store choice model. On the other hand, vendors with low MAX are those who are losing participants who have a theoretically high potential of visiting the vendor. Therefore, vendors with low MEAN and/or low MAX are exhibiting anomalous behavior from a standard store choice model perspective.

3.3.2. Data

The first effort was to demonstrate that by using these models I could do at least as well as other investigation methods in identifying anomalies in the patterns of vendor-participant interactions and potentially identify fraud.

Data were compiled from three different sources in Ohio WIC. The WIC certification system provided detailed information on program participants. The Vendor Management System (VMS) collects demographic information of all WIC vendors that contract with the state to provide foods to participants. Vendor locations and demographic information were retrieved from the VMS. Finally, one month (April 2004) of payment data were retrieved to identify participants who redeemed their benefits along with vendors from whom the participants redeemed these benefits. Figure 12 presents a basic structure of data synthesis and analysis.
Figure 12: Data synthesis and analysis – data were extracted from three different information systems for a county in Ohio WIC (certification system, vendor management system, and payment data system). Analysis was performed using GIS for identifying the locations, Matlab for calculating the probability matrix, and STATA for analyzing the final results.

Data consisted of 78,236 transactions in an urban Ohio county in April 2004. These transactions correspond to the number of food vouchers redeemed by participants that were paid for by the State. Ohio WIC participants receive three to four food vouchers each month. The transactions involved 22,874 participants and 78 vendors. I analyzed the interactions between vendors and participants using ArcGIS, Matlab, and STATA.
3.4. Results

3.4.1. Descriptive Statistics

There appears to be considerable variety among the vendors licensed by the state to provide food for women in the WIC program (Table 3). They vary in size. The average number of lanes is 9, ranging from 1 to 37, with average sales per lane of approximately $3,300. The average number of transactions at each vendor was approximately 1,000. Estimates of the food vouchers from the vendors suggest that only approximately 75 percent of the food vouchers were redeemed. Food package costs per participant at each vendor were $38.45.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N=78</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanes (#)</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Transactions (#)</td>
<td>1,003</td>
<td>1,117</td>
<td>12</td>
<td>4,520</td>
<td></td>
</tr>
<tr>
<td>Vendor redemption ratio (%)</td>
<td>74.5</td>
<td>9.1</td>
<td>53.4</td>
<td>96.58</td>
<td></td>
</tr>
<tr>
<td>WIC sales per lane ($)</td>
<td>3,300.65</td>
<td>3,901.41</td>
<td>23.00</td>
<td>18,418.00</td>
<td></td>
</tr>
<tr>
<td>Food package costs per participant at vendor ($)</td>
<td>38.45</td>
<td>10.02</td>
<td>12.00</td>
<td>69.00</td>
<td></td>
</tr>
<tr>
<td>High-risk vendors (%)</td>
<td>0.09</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Vendor propensity measures

| MEAN      | 0.190 | 0.114 | 0.002 | 0.468 |
| MAX       | 0.765 | 0.286 | 0.010 | 0.998 |

Table 3: Program activities in an urban county (A) in Ohio WIC
3.4.2. Risk Groups

I grouped the vendors by their risk levels utilizing two different methods, (1) the state’s routine high-risk vendor identification index based on monthly payment activities and (2) actual compliance investigations in the county. Neither of the methods is perfect for grouping the vendors by risk levels. Each method has advantages and limitations. The high-risk vendor identification index is based on practical experience and is used as a routine monitoring procedure to identify outliers. The compliance investigations serve as a check and complement the state high-risk vendor identification index.

First, the 78 vendors in the dataset were grouped by the state’s high-risk vendor identification index. The index consists of three sub-criteria for WIC sales volume at the vendors. The sub-criteria are (1) vendor redemption ratio, (2) WIC sales per lane, and (3) food package costs per participant at the vendor. The vendor redemption ratio is based on two facts. Not all food benefits in food vouchers are redeemed by participants. The state over-issues 10 percent of the total price of products in food vouchers to allow for price variance or fluctuations in different regions or over time. Therefore, those vendors who charge more than 90 percent of the total food voucher values are flagged for illegal transactions. The other two criteria are based on the same logic to identify outliers. Vendors were categorized as high-risk when they were over the 75th percentile for the vendor redemption ratio and 90th percentile for the other two criteria. The state high-risk vendor identification index is the sum of the three sub-criteria. Of the 78 vendors, 8 vendors (10 percent) met more than one sub-criterion. These 10 percent were categorized as high-risk vendors and others as non-high-risk vendors.

39 In the state, the vendor redemption ratio is used to identify vendors as high-risk when the ratio is greater than 90 percent. The other two measures are based on absolute values (sales per lane >$5,000 and food package costs per participants > $50). Here I slightly modify the criteria for the research. While I used 75th percentile for the vendor redemption ratio due to the unique characteristic of the vendor redemption ratio, the 90th percentile was used for the other two criteria.
Second, some vendors in the county were grouped by the results of compliance investigations that were implemented from April to June 2004. Of the 78 authorized vendors in the counties, 51 vendors were selected for compliance investigations. Three selection criteria were used to select these vendors: (1) those who had been previously identified as high risk vendors by the state’s high-risk vendor identification index, (2) vendors identified by specialists in the field, and (3) random sampling from those who were not identified from either method. This procedure overestimates the likelihood of a vendor committing fraud, and therefore the potential number of vendors who might violate their contract.

Among these 51 vendors, 12 vendors (24 percent) were found to have overcharged. The compliance investigations also identified the 6 high-risk vendors that were identified by the state criteria as having overcharged. One vendor was excluded from the investigation due to administrative reasons. In addition, six vendors who were not categorized as high-risk vendors by the state’s high-risk vendor identification index were also identified as high-risk in the compliance investigation.

3.4.3. Distribution of Vendor Propensity Measures

I analyzed the interactions between vendors and participants. For example, a MEAN of 0.40 at a hypothetical Vendor X indicates that on average, Vendor X attracted participants who have approximately a 40 percent chance of visiting the vendor. If the MEAN is higher, it indicates that the vendor attracts participants who have higher probabilities of visiting the vendor among the alternatives. On the other hand, if the MEAN is lower, the vendors attract participants who have a low chance of visiting at the vendor. The other measure (MAX) also provides

---

40 It took approximately 12 months to collect all vouchers that were used for the investigation. Therefore, the state was able to have the result in May 2005.
interesting information. The vendors with low MAX are losing participants who have a theoretically high chance of visiting at the vendor. Therefore, vendors with low MEAN and/or low MAX are potentially attracting those participants who would not visit under normal circumstances. Alternatively, they are loosing participants who would normally visit them.

Figure 13 presents the distribution of vendors by MEAN and MAX. A cluster of vendors with extremely low MEAN and MAX is circled in the lower left corner. For example, the vendor with almost zero MEAN and MAX implies that the vendor was not attracting those who were likely to visit the vendor (low MAX) and attracting those who were not likely to visit the vendor (low MEAN). This provides a starting point for further investigation.

![Figure 13: Scatterplot of the MEAN & MAX values for 78 vendors](image-url)
3.4.4. Explaining the Anomalies

If a vendor attracts participants by providing illegal benefits and certain participants respond to this offer, decision factors such as size and distance in the model may not play a critical role in participants’ vendor selection decision. If a vendor is known to commit fraud, it is highly likely that some participants may be attracted to the vendor or be repelled by the illegal activity. In such situations, the model prediction will be different from actual behavior.

**Figure 14** presents MEAN box-plots by risk group in both identification methods. In the left figure, the vendors were grouped by the state high-risk vendor identification method. The vendor propensity measure (MEAN) was significantly lower in the high-risk group compared to the non high-risk group. The right figure shows MEAN for the high-risk group and non high-risk groups when the vendors were grouped by the actual field investigation. In both figures, the non-high-risk vendors attracted those who have a relatively high probability of visiting the vendors compared to high-risk vendors. The high-risk vendors attracted substantial numbers of participants who were not expected to visit.
Figure 14: Comparison of vendor propensity measure by risk groups that were categorized by using two different methods. Left-hand-side figure shows the distribution of MEAN vendor propensity measures when the vendors are grouped by the state high-risk vendor identification index. Right-hand-side figure shows the distribution when the vendors are grouped by the result of field investigation. Note that the total number of vendors included is not that large. Therefore, these figures should be interpreted with caution.

The marketing literature suggests another explanation, which is that the anomalous pattern may occur due to cultural coherence between stores and consumers (Ogden, Ogden, & Schau, 2004; Wang, 2004; Eckman, Kotsiopulos, & Bickle, 1997). If a vendor with certain cultural characteristics, such as nationality, attracts participants who have the same characteristics, the model based on store size and distance may not be able to explain the anomaly since the model did not include these factors. However, to conclude that these anomalies imply illegal
behavior would be also incorrect. I analyzed the investigations and demographic information to explain whether cultural factors could explain these differences.

The county from which these data were obtained has a substantial population of poor immigrants. I found that the clustered eight stores in Figure 2 were owned by people of the same nationality as these immigrants. I refer to them here as specialty vendors. In terms of interactions, seven of the eight vendors showed low MEANs. Of the seven high-risk vendors, four vendors with low MEAN were reported to have had overcharge violations, and two vendors had other problems such as not having price tags or not returning food vouchers. The remaining vendor had substitution violations. In particular, vendor V8 was reported to have overcharge violations in the compliance investigation, while they seem to have relatively high propensity measures (MEAN=0.234 (MAX: 0.800)). Therefore, cultural coherence could explain the anomaly, but it does not refute fraud in this context.

I wondered why the vendor V8 was different from others in terms of propensity measures. Among the eight specialty vendors, four vendors participated in the program since August 2002, and the other four vendors participated since August 2003. Vendor V8 belonged to the August 2002 cohort. When I compared the historical redemption activities (October 2002 to April 2004), among the four vendors, vendor V8 was the only vendor that was not categorized as a monetary high-risk vendor by the state high-risk vendor identification index until October 2003. After October 2003, vendor V8 started meeting one of the three state high-risk vendor identification criteria. I also noticed that the location of the vendor V8 geographically appeared out of place compared to others.

Table 4 presents a summary of vendor propensity measures of the eight specialty vendors with their risk codes and compliance investigation results. The vendors consist of the top seven vendors who showed low MEANs. The vendors showed a difference not only in the propensity measures, but also in the number of transactions compared to their store size.
<table>
<thead>
<tr>
<th>ID</th>
<th>RC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CI&lt;sup&gt;b&lt;/sup&gt;</th>
<th>MEAN</th>
<th>Std. Dev.</th>
<th>MAX</th>
<th>MIN</th>
<th>Store Size</th>
<th>Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>HR</td>
<td>S</td>
<td>0.002</td>
<td>0.003</td>
<td>0.010</td>
<td>0.000</td>
<td>1</td>
<td>172</td>
</tr>
<tr>
<td>V2</td>
<td>HR</td>
<td>OC</td>
<td>0.009</td>
<td>0.010</td>
<td>0.041</td>
<td>0.000</td>
<td>1</td>
<td>701</td>
</tr>
<tr>
<td>V3</td>
<td>AR</td>
<td>OC</td>
<td>0.013</td>
<td>0.030</td>
<td>0.150</td>
<td>0.000</td>
<td>2</td>
<td>347</td>
</tr>
<tr>
<td>V4</td>
<td>HR</td>
<td>NT</td>
<td>0.016</td>
<td>0.011</td>
<td>0.038</td>
<td>0.000</td>
<td>1</td>
<td>287</td>
</tr>
<tr>
<td>V5</td>
<td>HR</td>
<td>OC</td>
<td>0.024</td>
<td>0.028</td>
<td>0.092</td>
<td>0.000</td>
<td>2</td>
<td>1344</td>
</tr>
<tr>
<td>V6</td>
<td>HR</td>
<td>NF</td>
<td>0.025</td>
<td>0.028</td>
<td>0.107</td>
<td>0.000</td>
<td>2</td>
<td>840</td>
</tr>
<tr>
<td>V7</td>
<td>HR</td>
<td>OC</td>
<td>0.032</td>
<td>0.087</td>
<td>0.632</td>
<td>0.000</td>
<td>1</td>
<td>952</td>
</tr>
<tr>
<td>V8</td>
<td>HR</td>
<td>OC</td>
<td>0.234</td>
<td>0.307</td>
<td>0.800</td>
<td>0.001</td>
<td>2</td>
<td>141</td>
</tr>
</tbody>
</table>

Table 4: Vendor propensity measures of specialty vendors - a. Risk Code (RC) met in the State high-risk vendor identification method (HR: High-risk, AR: At-risk) b. Compliance Investigation (CI) - NV: no violation; S: substitution; OC: overcharge; NF: no food vouchers submitted; NT: no price tags

3.5. Discussion

3.5.1. Summary

I have tested the utility of store choice models by constructing vendor propensity measures for monitoring fraud in the context of public delivery programs. This approach provides an opportunity to test whether the traditional economic models based on rationality can be utilized in such a context. I also document the abnormal spatial patterns associated with vendors.
who were involved in fraud. The study shows that these measures can be resource-effective tools for monitoring vendor performance. However, it is more important to realize that the context should be carefully understood to reduce the risk of alternative explanations for the observed results.

The state of Louisiana built an excellent GIS and Business Intelligence (BI) framework for discovering fraud (Frontenot, 2004). Louisiana’s fraud detection application shows how the interaction between GIS and BI can help detect food stamp fraud. Similarly, Missouri has also followed in Louisiana’s footsteps to develop a fraud detection program (Perlman, 2005; Douglas & Matlack, 2005). In these examples, however, GIS maps simply function as a visualization tool to illustrate the analyses of business intelligence. In other words, once the basic data and information were in place, the state of Louisiana began extending its fraud detection capabilities by incorporating GIS into the BI. By visualizing these data on a map, state employees could quickly peruse anomalies in the data and make informed decisions (ESRI, 2004).

I propose a different way of utilizing GIS and the notion of spatial proximity for monitoring criminal activity in public delivery programs. GIS can not only present complex information in a visual format, but it can also be used as a modeling tool to incorporate spatial interaction in fraud detection methods. I believe that this example with propensity measures addresses the limitation of the Louisiana system and also improves the utility of GIS for policy analysis. The method can be transferred to similar public delivery programs. This work can be naturally extended to the emerging areas of homeland security and emergency management (Nunn, 2005).

3.5.2. Policy Implications
States need to send the right signals to vendors and participants to make them aware that their activities are being monitored periodically. The minor violations identified need to be emphasized during the vendor training process. Continuous training, education, and communication will prevent vendors from becoming high-risk vendors. Using such tools and making people aware that these tools are being used can have deterrent effect on would-be violators. However, one should be careful in interpreting the results of such analysis outside of the context in which they occur. Finally, states need to invest resources to understand participants’ benefit redemption patterns and interactions with vendors since the target group can be very different from the general consumers that most academic literature has been focused on. This effort will pay dividends in the form of highly effective vendor management practices and a program with integrity.

3.5.3. Limitations

The simplicity and elegance of utility maximizing rational choice models can be very attractive. However, these models do not fully capture complexities of the vendor-participant interactions (March, 1978; Simon, 1996; Drezner & Eiselt, 2002). Their static approach does not leave much room for randomness, adaptive behavior, and learning, which are critical aspects of human decision-making (Fawcett & Provost, 1997; Wilhelm, 2004). The use of a county as the geographical unit of analysis may not be appropriate, since differences in geography or community can influence participants’ vendor choices. However, that analysis is beyond the scope of current focus.

My objective in the next chapter is to address some of these limitations of rational choice models by introducing random choices and adaptive behavior using a simulation model. Basic characteristics of the WIC program and statistical and spatial patterns that I identified in
this chapter will serve as a reference for the simulation. Therefore, a major focus of the next chapter is on modeling salient features of the WIC system inside the computer based on their interdependency and programming the dynamic interactions among the players such as vendors, participants, and the public agency.

In particular, the next chapter strengthens the idea that fraud in the delivery of public services can be framed as a complex system. This chapter also highlights that the traditional crime literature has limitations in addressing the dynamics of fraud. By running a store choice rule which is built from the discussion in this chapter and a fraud negotiation rule which is learned from the discussion with Ohio WIC vendor specialists, I simulate the underlying process of fraud in the policy system. This process will be validated by replicating statistical and spatial patterns associated with fraud that has been explored in this chapter and empirical data.
CHAPTER 4

SIMULATING FRAUD USING AN AGENT-BASED MODEL

The future is plural.

de Geus (1987)

Abstract

Fraud in public delivery programs often involves several entities that are loosely interrelated and dynamically interact. The crime literature, however, does not sufficiently address the mechanisms underlying this type of fraud and therefore are of limited use to public management. In this chapter, I consider fraud in a public delivery program as a complex system and develop an agent-based model to help understand the interactions and interdependency among the players in the system. I demonstrate my approach using both empirical and hypothetical data. I conclude that agent-based models can closely replicate the statistical and spatial patterns of fraud and provide a framework for future work in this area.
4.1. Introduction

Fraud is a crime that violates social norms, uses secretive processes, injures victims, and benefits perpetrators unfairly (Barker & Roebuck, 1973; Vandenburgh, 1999). There have been a variety of forms of fraud throughout history, including money laundering, credit card fraud, and telecommunications fraud. In the public sector, fraud in welfare, health care, and child care programs have been well-documented by government agencies such as GAO and USDA.

Public agencies have utilized fraud prevention and detection mechanisms in order to minimize fraud. Fraud prevention focuses on the procedure to avoid fraud using some methods such as watermarks and personal identification documents before fraud occurs. Fraud detection focuses on identifying those who committed fraud after-the-fact using statistical methods or actual investigations. However, many of these traditional mechanisms are static and ineffective in preventing or uncovering interactions among corrupt agents. These mechanisms are also easy to evade because of familiarity with the procedures. In other words, they do not effectively deal with the non-stationary nature of fraud. Once a method is put into place, it begins to lose effectiveness because the pattern of fraudulent behaviors changes as a response to the method (Bolton & Hand, 2002; Fawcett & Provost, 1997; Glover & Aono, 1995). Therefore, there is a need to develop an alternative framework to address such complex and adaptive problems in public management.

In an effort to provide an alternative framework, this study investigates the statistical and spatial patterns of fraud in a public delivery program using an agent-based model. More specifically, I focus on formulating fraud occurring in the delivery of public services as a complex system, implementing an agent-based model to capture the interactions and interdependency among players in the public delivery system, replicating the distribution of fraudulent vendors in empirical data, and testing the effect of policy interventions.

I begin by providing a brief overview of the assumptions in the crime literatures. I
describe the research method including the problem statement and agent-based modeling. I next turn to the results of the chapter, which compare statistical and spatial patterns of fraud in the empirical and simulated data. In the context of fraud in a public delivery program, I explore (1) how the statistical patterns of fraud occur, (2) how the spatial patterns of fraud emerge, and (3) what effect a policy intervention have on deterring fraud. I argue that the patterns of fraud emerge primarily from the negotiation between vendors and participants based on their risk propensity and store size. I conclude that my agent-based model closely replicates the pattern of fraud in empirical data. I also discuss the potential of this approach for studying crime and informing decisions.

4.2. Background

Literature shows that crime, including fraud, is not a subject that has been uniquely addressed within a specific discipline. It has been studied under various names across different disciplines. For example, economics of crime is a well developed topic in the economics literature, while white collar crime is a topic that appears in the legal, criminology, and sociology literatures. Geography of crime has focused on the effect of location on criminal activity. Social scientists have offered multiple explanations for criminal behaviors that range from individual choice to cultural determinism. Here, I review the basic assumptions of criminals, criminal behavior, and crime in the literature. This helps me to understand fraud when it is defined as a crime, and allows me to identify the limitations in the crime literature from a management perspective.
4.2.1. Economics of Crime

Economists approach crime and punishment as a special case of the more general economic theory of choice. Criminals are assumed to be not different from other people as far as rationality, maximizing behavior, and preferences are concerned (Becker, 1968, 1976). Criminals are neither victims nor irrational agents. Criminals behave in the same manner as an average person and act to maximize their expected utility subject to constraints in an uncertain world. Hence, according to this theory, choosing to become a criminal is no different than choosing any other occupation.

Criminal behavior is also price elastic. The price of crime is an indicator of deterrence, which is measured in terms of the severity and certainty of punishment (Grogger, 1991; Ayres & Levitt, 1998). Criminals react to changes in prices in the same way as consumers and workers. Mixed empirical results have been reported on this deterrence hypothesis (Cooter & Ulen, 2000), and some improvements have been made by using exogenous variables to determine the level of policing (Levitt, 1996, 1997).

While scholars interested in the economics of crime were trying to refine the deterrence hypothesis within the framework of rational choice, some scholars pointed out critical limitations of the rational choice approach to understanding risky behavior (Cornish & Clarke, 1986; Vandenburgh, 2004). Deterrence is a communicative process. The deterrent aspects of punishment vary with individual perceptions. There is no a priori reason to assume that every individual would perceive a particular penalty to be equally severe or certain (Grasmick & Green, 1980; Lattimore & Witte, 1986). Still, how to address or model this heterogeneous risk preference of individuals remains an interesting and important research question.41

41 Axelrod (1984, 1997) has written extensively on complexity models of individual preferences.
4.2.2. White Collar Crime

Traditional sociologists recognize crime to be influenced by multiple dimensions of social life. A traditional theory of criminal behavior finds criminality to be caused by poverty or the psychopathic and sociopathic conditions associated with poverty. Crime is a function of socioeconomic factors, or possibly biological factors. Criminals are not different from those who are irrational or have mental illness due to societal oppression or disease. An appropriate way to eliminate crime is to attack the root cause of crime through job creation and income maintenance.

Sutherland (1940) argued that the traditional conception of crime and explanations were misleading and incorrect. He argued that the traditional theory was derived from data provided by criminal justice agencies, which focused on lower class crime. A general theory of crime should include white collar crime, and white collar criminality is not different from the criminality of the lower class. Both types of crime are learned rather than simply influenced by the psychopathic and sociopathic conditions.

Fraud is also a white collar crime in that it is “concerned in relation to business” (Sutherland, 1940, p. 1). White collar crime has been interpreted as crime committed by a person of respectability and high social status in the course of his occupation (Braithwaite, 1985; Baker, 2004). White collar criminals are mostly recidivists, and their illegal behavior is much more extensive. They often do not lose status among associates, even after violating the laws designed to regulate business (Sutherland, 1982). However, the usefulness of the distinction of crime by social status has been a source of debate in crime theory (Hirschi & Gottfredson, 1987; Griffin, 2002; Lynch, McGurrin, & Fenwick, 2004). Empirical studies on social class and punishment have also yielded mixed results (Wheeler, Weisburd, & Bode, 1982; Benson & Walker, 1988).

The current definition of white collar offenses by the Department of Justice shows that it constitutes those classes of non-violent illegal activities, which principally involve traditional
notions of deceit, deception, concealment, manipulation, breach of trust, subterfuge or illegal circumvention (Baker, 2004). Levitt & Dubner (2005) argued that “despite all the attention paid to rogue companies like Enron, academics know very little about the practicalities of white collar crime” (p. 46). A more recent study (Friedrichs, 2004) suggests an integrated theoretical approach to explain cases such as Enron as “an outcome of a complex interaction of many different factors and variables, operating on various levels” (p.116).

4.2.3. Geography of Crime

Geographers have played a unique role in explaining crime. Geography of crime studies the spatial dimension of crime, focusing on the effects of place on crime (Chainey & Ratcliffe, 2005). This approach is based on the assumption that certain characteristics of places or neighborhoods influence crime (Cohen, 1941). Therefore, place is a factor that can be used to explain crime. Depending upon the nature of this relationship, crime control and interventions targeting specific locations can be fruitful or wasteful. Advanced spatial analysis techniques have been developed to investigate the relationship (Anselin et al., 2000).

Spatial analysis of crime has focused on two distinct approaches that are slightly different in nature. One approach is to describe and map the spatial distribution of crime since crime is not distributed evenly or randomly over space. For example, point pattern analyses, such as hot spot mapping (Lockwood, 2005), have been used to see whether the concentrations of crime within an identifiable boundary are clustered, randomly distributed or uniformly distributed. Polygon data type analyses are common applications. Several statistical techniques have been suggested to assess whether spatial patterns in the polygon type data analyses are systematic. In recent years,

---

42 The Moran’s I statistic measures the deviance from spatial randomness. The Moran scatterplot visualizes the degree of spatial autocorrelation in a dataset. The Getis-Ord G* statistic was designed to find clusters of high or low values within a given distance band. For more details, see Anselin et al. (2000).
personal computers, combined with GIS software, have helped to infuse this type of crime analysis on organizational and management problems in public agencies (Baum, 2005).

Another approach is to study the spatial patterns of crime with the environmental, social, historical, psychological, and economic variables that may explain these patterns (Georges, 1978). Socio-economic causes, including neighborhood characteristics, have been identified as important factors in predicting the incidence of crime (Herbert, 1976; Brown, 1982; Krivo & Peterson, 1996; Sampson, Raudenbush, & Earls, 1997; Roncek, 1981). While regression analysis plays a crucial role in modeling the causes of crime, it is also acknowledged that spatial concentration tends to result in spatial autocorrelation, which violates the assumptions of classical regression analysis (Anselin, et al., 2000). Attention has been given to addressing spatial heterogeneity in regression models, especially based on arbitrary boundaries.

4.2.4. Management Perspective

Two findings can be noted from the crime literature. First, the literature has focused on the relationships between the criminal behaviors of individuals and causes of that behavior based on discipline-specific assumptions. By focusing on limited factors, they fail to adequately describe the complex interactions and dynamics that allow crime to occur in the real world. Second, empirical findings are mixed. Methods utilized in empirical studies are primitive and static. They are limited to uncovering the process of any type of deviance. In many instances, these studies simplify complex issues that emerge from the adaptive behavior of people or entities, with linear and static models primarily using cross-sectional data, resulting in errors due to over-simplification.

Therefore, while the academic literature offers substantial accounts of the nature and causes of criminal behavior, these studies have not provided sufficient understanding of the
mechanisms underlying the practice of fraud, particularly in public delivery programs, where several entities are dynamically and loosely interrelated (Weick, 1979). These studies have limited utility for public managers in providing practical and realistic advice for the prevention and detection of fraud.

A number of scholars have recently attempted to understand fraud as an outcome of complex interactions (Provost, 2002; Wilhelm, 2004). They recognize that fraud is an issue that requires dynamic, evolving, and adaptive approach due to its complexity. Those who commit fraud have substantial knowledge on the program and modify their behavior in response to the change in management strategies. Therefore, the adaptive nature of fraud in public programs is not well-addressed in the traditional framework. It requires a dynamic and equally adaptive approach, and spatio-temporal patterns need to be considered carefully in order to address the issue effectively.

4.3. Method

I draw an integrated framework from recent advances in complexity science and models. The approach sees natural, human, or social phenomena as complex adaptive systems. Complex systems are generally defined as dynamic systems that exhibit recognizable patterns of organization across spatial and temporal scales (Holland & Miller, 1991; Parker et al., 2003). In other words, complex systems consist of a network of interacting adaptive agents who exhibit a dynamic aggregate behavior that emerges from individual activities of the agents (Holland & Miller, 1991).

To implement such an integrated framework, I consider fraud in public delivery programs as a complex system and explore the applicability of a simulation model. Agent-based models allow me to study fraud in public programs as an outcome of a complex interaction of
heterogeneous actors (Bankes, 2002; Epstein & Axtell, 1996). I explore the underlying mechanism of fraud by replicating statistical and spatial patterns in empirical data. In Figure 4, I identify fraud in public delivery programs at the intersections (the shaded area) of context, issue, and interest boundaries. I use a comprehensive framework that can link micro behavior and macro patterns in order to study the issue of fraud at this complex junction.

4.3.1. Policy Problem in WIC

The WIC program is also subject to managerial and operational breakdowns and their undesirable consequences. Fraud is an example of these consequences. A study of national WIC retail vendors (USDA, 2001) showed that 18.1 percent of retail vendors were involved in overcharging violations\(^{43}\) one or more times, including 12.4 percent that overcharged once, 4.2 percent that overcharged twice, and 1.5 percent that overcharged in three separate investigations. For vendors in the WIC system, overcharging is the main source of fraud and is the focus of this chapter.

Within the WIC system, four possible fraud mechanisms can be noted. First, fraud can occur through the illegal exchange of benefits between vendors and participants. Well-known fraud activities such as overcharging, substitution, and trafficking belong to this category.\(^{44}\) The transaction occurs through negotiations between vendors and participants. Second, fraud can occur through the improper exchange of benefits among participants. By contract, participants agree to follow the WIC policies and guidelines, which require that they satisfy program

\(^{43}\) Overcharging means a charge by a WIC authorized store to the program through redemption of a WIC check for an allowable food in excess of the store’s shelf price for that food or in excess of the price charged a non-WIC participant for that food.

\(^{44}\) Some other illegal activities include forcing unwanted purchases, substitutions of WIC foods for unauthorized items (substitution), and the exchange of public service benefits for cash (trafficking). Each of these is a different type of fraud, but all are considered to be illegal behaviors in the WIC system.
eligibility and do not hand over the benefits to non-WIC participants or other WIC participants. In reality, compliance on this contract is uncertain in some cases. Third, fraud can also occur through a network of corrupt vendors. Fourth, in addition to improper conduct by vendors or participants, there is a chance that a third party can be involved. For example, an external group can get into the business of fraud by committing organized crime for expensive items. The last three mechanisms, while relevant, are beyond the scope of this research. Here, I focus on fraud committed by the interactions between vendors and participants and the macro patterns that emerge from these interactions, because these are more common than others.

4.3.2. Agent-Based Modeling

My agent-based model captures the interactions and interdependency among individuals and other entities in a public service delivery system. I have based this agent-based model within the programming structure of “Multi-Agent Simulation Of Neighborhoods... or Networks... or something...” (MASON). MASON serves as the basis for a wide range of multi-agent simulation tasks (Balan et al., 2003; Luke et al., 2004).

I also used players in Figure 2 as basic agents. In Figure 15, agent “public agency” represents the functions of local clinics and state administration. For example, local clinics issue food vouchers for participants and update information given to the state. The state contracts with vendors for the delivery of food benefits and monitors the WIC program. These functions have been integrated for the public agency agent. A “vendor” agent is responsible for delivering food

---

45 To be eligible for WIC, participants should meet all of the following categories: under 185 percent of the Federal poverty income guidelines, residential guideline, and nutritionally at risk.

46 In July 26, 2005, the U.S. Department of Justice announced that the United States Attorney Northern District of Ohio returned a seven-count indictment for conspiracy to commit money laundering and conspiracy to commit food stamp fraud, WIC fraud, mail fraud, wire fraud, and witness tampering. The indictment alleges that fraud has been committed by a fraud network for a long time (1995 - 2001).
and nutritional supplements to “participant” agents, who redeem their vouchers. The public agency agent is not spatially explicit, while the other types of agents are spatially referenced.

Figure 15: Adaptive complex enterprises – agents, interdependence, and interaction were defined above. While the Huff spatial interaction model addresses the interaction between participant and vendor agent, the agent-based model also includes activities of a public agency agent in the WIC system.

During the initialization of the simulation model, each participant is assigned a vendor. This assignment is based on a store choice rule that is adopted from the spatial interaction model developed by Huff (1964) in order to have some sense of how participants will choose vendors upon joining the program, as well as in general. As discussed in the previous chapter, the Huff spatial interaction model consists of two decision factors for store choice (size and distance). In
this model, the probability of a consumer visiting a particular store is calculated as a relative measure equal to the ratio of the utility of that store to the sum of utilities of all stores considered by the consumer. More formally, 

\[ p_{ij} = \frac{U_{ij}}{\sum U_{ij}} \]

where \( i \) and \( j \) indicate the consumer and store respectively.

Once the store choice rule is executed, a fraud negotiation rule is used in each iteration. Risk propensity and vendor size are crucial properties on which I focus. Risk propensity is a hypothetical property used to model the change of an agent’s propensity toward risk, and is based on decisions and behaviors. I assume that at the initial stage, agents’ risk propensity follows a truncated Gaussian distribution with a mean of 0.40 ranging from min 0 to max 1 for both participant and vendor agents regardless of other properties. The assumption is that the higher the assigned risk propensity, the higher the probability of committing fraud will be.

Store size is a proxy of business type in the WIC program. Larger vendors are most likely national chains while small vendors with one or two lanes are owned by single owners. It is reported that fraud occurs more frequently among small vendors than among large vendors (USDA, 2001). Hence, the smaller the vendor, the larger the probability of committing fraud. I programmed that vendors’ involvement in fraud is influenced by store size and dynamically changing risk propensities.

For example, among small vendors (< 3 lanes), if risk propensity is greater than 0.9, there is a 95% chance of committing fraud, while if risk propensity is less than 0.3, the vendor has only a 10% chance of being involved in fraud. For larger vendors (> 5 lanes), those with a low risk propensity (< 0.3) have a 0.01% chance of participating in fraud, while vendors with a high risk propensity (> 0.9) have a 10% chance of committing fraud. This process will lead to a skewed distribution of vendor risk status by store size when they reach equilibrium. In other words, smaller vendors will be more likely to be categorized as high risk than larger vendors.
because they have a higher chance of being involved in fraud, even when risk propensity was equally distributed among the small and large vendors at the initialization.

I modeled four possible outcomes of fraud negotiation between participant and vendor agents. Fraud negotiation fails when one party refuses to be involved. In this case, participant agents move to the next choice based on the store choice rule. If both agents agree to commit or not commit fraud, participant agents continue to visit the vendor and use their benefits throughout the simulation.

Along with this process, I set up four other mechanisms to simulate vendors’ sales activities. First, a public agency agent issues vouchers, with a mean of $45.00, minimum of $2.00, and maximum of $100.00, to participant agents. A participant agent uses approximately 75% of the voucher value, with a range of 4% to 100%, when exchanging the vouchers for WIC foods. Second, actual dollar amounts used by participant agents will be recorded differently by vendor agents, depending on the result of the negotiation. The vendor agents record that a participant agent used 100% of the benefits if that participant agrees to commit fraud, while the vendor records the actual amount used by participants who are not involved in fraud. The third mechanism is that participant agents who were not involved in fraud, but have a relatively high risk propensity (> 0.6), will have a 1% of chance of moving to random vendors during each step. The participant agents use the fraud negotiation rule with new vendors until they find a comparable vendor. Finally, the risk propensities of participant and vendor agents will change, depending upon their decisions and behaviors. I assume that risk propensity increases or decreases faster for smaller vendors (<3 lanes) than larger vendors when they are involved in fraud. I continued to test and revise the parameters to identify most appropriate ones that replicate the statistics from the empirical data. In Table 5, I summarize the initial parameters for the basic properties in the simulation.
Table 5: Initial parameters of the basic properties used in the simulation

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor risk propensity (VRP)</td>
<td>0.40</td>
<td>0.40</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Participant risk propensity (PRP)</td>
<td>0.40</td>
<td>0.30</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Voucher values ($)</td>
<td>45.0</td>
<td>40.0</td>
<td>2.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Voucher usage by a participant (%)</td>
<td>0.75</td>
<td>0.10</td>
<td>0.04</td>
<td>1.00</td>
</tr>
</tbody>
</table>

4.4. Simulating Patterns of Fraud

4.4.1. Descriptive Statistics from the Empirical Data

In the state vendor monitoring system, vendors meeting more than one risk criterion are considered to be high risk. These criteria are based on a data-driven risk assessment in the state system that serves to identify outliers (i.e. >90th percentile for all criteria in this research). For example, here we focus on the pattern of WIC sales volume by vendors. Sales volume is investigated for (1) the ratio between the estimated total voucher costs from collected vouchers and actual charges from vendors, (2) WIC sales per lane, and (3) food package costs per participant at the vendor. Categorizing the sales volume and comparing vendors with their peers provides an opportunity to monitor vendors from different angles. The number of risk codes shows how many criteria (out of three) the vendor met each month.

In this study, one month (April 2004) of payment data in an urban county in Ohio were used to obtain an estimate of the distribution of the risk status of vendors in the county. In the
sample county, a total of 188 vendors were included, along with a total of 28,887 participants who redeemed their benefits in April 2004. On average, a vendor has 6 lanes, ranging from 1 to 30 lanes. The state paid, on average, $8,976 to each vendor (state payment to vendors) while the total cost from collected vouchers was an average of $11,589 at each vendor (voucher costs). The vendor redemption ratio was approximately 82 percent of the state’s voucher costs. Average sales per lane were $2,520 per month. Average food costs per participant at the vendor were approximately $40, ranging from $7 to $76. Descriptive statistics are provided in Table 6 as a reference for my simulation.

<table>
<thead>
<tr>
<th>(N=188)</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Note a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanes (#)</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>State payment to vendors ($)</td>
<td>8,976</td>
<td>10,008</td>
<td>28</td>
<td>56,884</td>
<td></td>
</tr>
<tr>
<td>Voucher costs ($)</td>
<td>11,589</td>
<td>13,988</td>
<td>38</td>
<td>82,263</td>
<td></td>
</tr>
<tr>
<td>Participants per vendor (#)</td>
<td>223</td>
<td>248</td>
<td>4</td>
<td>1,402</td>
<td></td>
</tr>
</tbody>
</table>

**Risk criteria**

<table>
<thead>
<tr>
<th>Risk criteria</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Note a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor redemption ratio (%)</td>
<td>82.0</td>
<td>10.7</td>
<td>60.0</td>
<td>100.0</td>
<td>6.4%</td>
</tr>
<tr>
<td>WIC sales per lane ($)</td>
<td>2,520</td>
<td>2,706</td>
<td>3</td>
<td>16,673</td>
<td>9.6%</td>
</tr>
<tr>
<td>Food costs per participant at vendors ($)</td>
<td>40</td>
<td>12</td>
<td>7</td>
<td>76</td>
<td>8.5%</td>
</tr>
</tbody>
</table>

a Percentage of vendors who met each risk criterion

Table 6: Program activities in an urban county (B) in Ohio WIC

47 While the total number of participants who redeemed their benefits was 28,887 in the month, 41,854 unique participants showed up at vendors. Therefore, each participant visited approximately 1.5 vendors, which shows that participants visit more than one WIC vendor to redeem their benefits each month.
4.4.2. Replicating Descriptive Statistics

To make this simulation tractable, I reduced the size of the WIC system to 20 vendors and 1,000 participants, while the distribution of store sizes (numbers of lanes) replicated the statistical characteristics in the empirical data. The simulation ran 10 times to generate random vendors and participants. Figure 16 shows a comparison of the store size distributions in the empirical and simulated data.

Figure 16: Distribution of store sizes – gray bars shows the percentage of vendors by the number of lanes in the empirical data (total vendors = 188) and black bars show the distribution of vendors by the number of lanes in the simulated data (total vendors = 200).
I modeled the idea that small vendors with a high-risk propensity are more actively involved in fraudulent activity. This results in a higher sales volume for small vendors and, ultimately, more vendors with high-risk status among these small vendors. **Figure 17** shows the percentage of participants who were involved in fraud through initial negotiations between vendors and participants. Most frauds occur among small vendors. Approximately 20 percent of participants visiting vendors with 1 to 5 lanes were involved in fraud from the time of initial fraud negotiation. This will aggravate the risk status of the small vendors over time. The majority of participants and vendors will not be involved in committing fraud, especially at larger vendors.

![Figure 17](image)

**Figure 17**: Percentage of participant agents involved in fraud at Time 1 – it was programmed that more participant agents successfully commit fraud at the smaller vendors.
Figure 18 presents the ratio between total voucher costs and the actual dollar amount the state paid to vendors. The ratio increased over time, because the benefits for participants who are involved in fraud continued to be recorded at 100 percent. To show the source of the increase, I present the vendor redemption ratio by selected store size in Figure 18. The vendor redemption ratio of small stores (i.e. 1 and 3 lanes) continued to increase, while the ratio of large stores (i.e. 11 and 18 lanes) was stable over time.

Figure 18: Vendor redemption ratios over time by store size – the thick line in the middle shows vendor redemption ratio for all vendors.
What I expected from this process was to replicate the descriptive statistics found in the empirical data in April 2004. I started from the assumption that the risk propensity of agents is randomly assigned at the initialization of the simulation, regardless of an agent’s characteristics. However, I also programmed that small vendor agents with a high risk propensity will have a higher chance of committing fraud from negotiations with participant agents. I expected that this process would lead to the aggregated statistical patterns in the empirical data after certain time periods.

I examined descriptive statistics in my simulation at Time 299. The results are presented in Table 7. In the simulated data, vendors redeemed approximately 82 percent of the total voucher costs. On average, food costs per participant at vendors were $40, ranging from $11 to $78. The absolute sales amounts are approximately 1/10 of the original size of the WIC business in the empirical data with similar distributions.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Note^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanes (#)</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>State payment to vendors ($)</td>
<td>1,070</td>
<td>964</td>
<td>0</td>
<td>4,594</td>
<td></td>
</tr>
<tr>
<td>Voucher costs ($)</td>
<td>1,362</td>
<td>1,298</td>
<td>0</td>
<td>6,043</td>
<td></td>
</tr>
<tr>
<td>Participants per vendor (#)</td>
<td>28</td>
<td>26</td>
<td>0</td>
<td>122</td>
<td></td>
</tr>
</tbody>
</table>

Risk criteria

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Note^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor redemption ratio (%)</td>
<td>82.2</td>
<td>8.8</td>
<td>65.2</td>
<td>100.0</td>
<td>11.0%</td>
</tr>
<tr>
<td>WIC sales per lane ($)</td>
<td>313</td>
<td>359</td>
<td>0</td>
<td>2,158</td>
<td>10.0%</td>
</tr>
<tr>
<td>Food costs per participant at vendors ($)</td>
<td>40</td>
<td>8</td>
<td>11</td>
<td>78</td>
<td>10.5%</td>
</tr>
</tbody>
</table>

^a Percentage of vendors who met each risk criterion

Table 7: Descriptive statistics of the simulated data at Time 299 (the total number of vendors is 200)
Figure 19 presents the distributions of the three criteria that were used to evaluate the risk status of vendors. The distribution in the empirical data is presented on the left, and the distribution of the simulated data at Time 299 is presented on the right for comparison. This figure also shows that the simulation closely replicates the distributions of the three criteria in the empirical data with current parameter settings in the simulation.

While Figure 19 shows the distribution of WIC sales based on different indicators, two questions remain with regards to the original objective of this chapter. One is whether the simulation is replicating the distribution of high-risk vendors in the empirical data when the individual risk criteria were combined. It is expected to find a similar percent of vendors who were identified as meeting greater than and equal to two risk criteria in both data. The other is who the outliers are. It is also expected that small vendors are more likely identified with higher level of risk in both data.
Figure 19: Distributions of sales activities by three criteria used in the empirical and simulated data to identify outliers.
In Table 8, I present the percentage of vendors at different risk levels in both the empirical and simulated data. Approximately 5.3 percent of vendors in the empirical data were categorized as high-risk vendors as a result of meeting at least two risk criteria. In the simulated data, approximately 5.5 percent of vendors were categorized as high-risk. Therefore, this simulation reasonably replicates the distribution of vendors by risk status in the empirical data.

<table>
<thead>
<tr>
<th>Risk level of vendors</th>
<th>Percent (empirical data)</th>
<th>Percent (simulated data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>81.4</td>
<td>75.0</td>
</tr>
<tr>
<td>1</td>
<td>13.3</td>
<td>19.5</td>
</tr>
<tr>
<td>2</td>
<td>4.8</td>
<td>4.5</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 8: Percentage of vendors at different risk levels in the empirical vs. simulated data – gray rows are considered as “high-risk.”

In Figure 20, the distribution of vendors by risk level and store size was presented in the empirical and simulated data at Time 299. Vendor’s risk propensity was randomly distributed at the beginning regardless of vendor’s characteristics. However, the agent-based model was designed to make small vendors more actively involved in fraud compared to larger vendors. Therefore, I expected that small vendors were identified as high-risk. Figure 20 shows the result of the simulation based on such an assumption.
Figure 20: Distribution of vendors by risk level and store size in the (a) empirical (April 2004) and (b) simulated (at Time 299) data.
4.4.3. Investigating Spatial Patterns

Along with the descriptive statistics, I investigate spatial patterns emerging from the interactions among vendors and participants. In the retailing literature, for example, trading area patterns have been identified. A trading area is a geographical area containing the consumers of a particular firm or groups of firms for specific goods and services (Berman & Evans, 1995). The trading area is broken down into three parts: primary, secondary, and fringe. According to this analysis, the majority of consumers (50 to 80 percent) come from the primary area, with an additional 15 to 25 percent of a store’s customers in the secondary area. The remaining, most widely dispersed customers come from the fringe area. I have excerpted the trading area analysis pattern from Berman & Evans (1995) and modified it in Figure 21. These patterns are equilibrium patterns at a certain point in time. I focus on simulating such theoretical spatial patterns (Figure 21 – (a)) by using a classical store choice model as a base, as well as emergent spatial patterns due to fraud (Figure 21–(b)), using an agent-based model.
Figure 21: Patterns of trading area - (a) a normal pattern in which the majority of customers of a store live close to the store; (b) a pattern that may indicate possible fraud where a significant numbers of customers of a store come from far away while loosing those who were supposed to visit. Three circles surrounding vendors represent primary, secondary, and fringe areas in the trading area analysis.
I developed a hypothetical scenario with 10 vendors and 2,000 participants in order to explore spatial patterns. Locations of the vendors and participants were randomly generated within a rectangular landscape. Dots represent participants, while squares represent vendors. Each vendor was randomly assigned to have one to twenty store checkout lanes. The larger the square, the larger the number of lanes. The second number inside each vendor box shows the number of lanes. Vendors and participants were also randomly assigned risk propensities at the initialization. The numbers in decimals shows the assigned risk propensity.

Figure 22 shows the theoretical spatial pattern that can be expected when the store choice model is based on describing individual choice at the micro level. It is assumed that customers will choose stores based on economic rationality, that is, they will maximize their utility. It is possible to incorporate other decision factors, such as store quality and availability of parking, in order to generate different trading area patterns. However, the fundamental logic does not change.

In terms of a high-risk pattern, I assumed that a vendor who participates in fraud will lose participants with low risk propensity who are likely to visit the vendor, because negotiations to engage in fraud will most likely fail. Instead, these vendors will attract participants with a high risk propensity. In other words, high-risk vendors lose those who are likely to visit the vendor and attract those who are not likely to. This is ultimately reflected in the spatial patterns of vendors participating in fraud compared to those who are not involved in fraud. It should be noted that these “low propensity” participants will now go to vendors who are not necessarily the closest, hence, we have a situation in which “low propensity” vendors attract “low propensity” participants from secondary and fringe areas.

48 For the presentation purpose, I draw the same size squares for vendors who have more than 3 lanes. Numbers inside the vendor show the risk propensity assigned at the initialization of the simulation.
In Figure 22, I identified two small vendors with a relatively high (0.86) and low (0.12) risk propensity. Black circles show their locations. I continued watching their interactions with participant agents they were expected to serve. Figure 23 shows the result of fraud negotiations at the initial stage. The vendor with high risk propensity failed to attract participants in the vendor’s trading area and started attracting participants who failed in negotiations with their theoretical vendors and were successful in negotiations with the high-risk vendor.

In Figure 24, the vendor agent with high-risk propensity attracted more participants from secondary and fringe areas, while other vendors kept their original patterns at Time 292. It was found that a small vendor next to the high-risk vendor was also developing as high-risk at Time 292. However, the small vendor with a low risk propensity still maintained its original participants at Time 292. This simulation contrasts the different interactions between vendors and participants based on their risk propensities when the store size is the same.

Figure 25 presents the distribution of participants visiting the vendors at Time 308. The small high-risk vendor lost participants in the primary area and attracted participants from other areas who were not expected to visit from the current theoretical framework. Another developing high-risk vendor (inside dotted circle) also shows similar spatial interaction patterns. The high-risk vendor (inside solid circle) reached to the level of 1.0 (maximum risk propensity) from 0.86, while the low risk vendor remained at the level of 0.15 which is not different from the initial conditions (0.12).
Figure 22: Theoretical vendor choices of participant when the Huff model was deterministically implemented at Time 0. Two small vendors were highlighted: one with relatively high risk propensity (0.86) and the other with relatively low risk propensity (0.12).
Figure 23: Fraud negotiations occurred at Time 1.
Figure 24: Fraud negotiations continue for those participants who have a relatively high risk propensity (> 0.6) and are not involved in fraud. While the small vendor with low risk propensity remained stable (risk propensity: 0.12 to 0.15), the high-risk vendor was actively attracting participants (risk propensity: 0.86 to 1.0) at Time 292. Another one lane vendor was developing as a high-risk vendor (risk propensity: 0.66 to 0.86) at Time 292. Other vendors were maintaining their original trading areas.
Figure 25: Highlighting the interaction between small vendors with high risk propensity and participants at Time 308 (black solid and dotted circles) while the small vendor with low propensity was faded out (gray solid circle). All vendors are small, having one lane. The vendor's risk propensity inside black solid circle was 0.86 at the beginning and quickly reached to 1.0 at the end of the simulation. The vendor was attracting participants from all over the place.
In summary, I modeled the interactions and interdependency among players in a public delivery system and explored the possible mechanisms underlying the patterns of fraud using an agent-based model. This study shows that the agent-based model reasonably replicated the statistical and spatial patterns of fraud in the system. Therefore, the model can serve as a framework for implementing different theoretical assumptions, modeling underlying mechanisms, and testing policy options. Below, I focus on simulating the effects of a policy intervention on deterring fraud in the simulated policy system.

4.5. Testing the Effect of a Policy Intervention

4.5.1. Deterrence Hypothesis

While scholars have seen deterrence as a vague idea developed from the legacies of two moral philosophers, Cesare Beccaria and Jeremy Bentham (Meier & Johnson, 1977; Piliavin, Gartner, Thornton, & Matsueda, 1986; Gneezy & Rustichini, 2000), in the late 1960’s, this hypothesis was revitalized by an influential economist, G. B. Becker (1968), and was later extended by Ehrlich (1975a, 1975b, 1996) and others (Piliavin, Gartner, Thornton, & Matsueda, 1986; Paternoster, 1987; Grogger, 1991; Levitt, 1996, 1997; Levitt & Dubner, 2005).

Becker (1976) emphasized that what most distinguishes economics as a discipline from other disciplines in the social sciences is not its subject matter but its approach. The economic approach assumes stable preferences underlying the objects of choice, market equilibrium, and maximizing behavior more explicitly and extensively than other approaches. Therefore, the economic approach can be applied to many areas of human behavior, such as family, fertility, marriage, and other topics that were not a primary focus of economics. Crime and punishment is one area that was substantially influenced by Becker’s ideas.
In Becker’s model (1968), for example, if individuals are risk-neutral, criminals will respond to the benefits (B) of crime (monetary and psychic), the probability of being caught (P), and the costs of punishment (C) for illegal activity. If the gap between the benefits of the illegal activity and punishment (B-PC) is greater than the wage for legal activities (W), the individual will commit crime, because criminals make rational choices. Therefore, the total number of criminals is a function of the four components in Becker’s model: number of criminals 

\[ n = f(B, P, C, W). \]

The number of criminals increases as the benefit increases, and decreases when the probability of being caught, the cost of crime, and wages in the legal employment sector rise (Glaeser, 1999).

Since *Crime and Punishment* (1968), the deterrence hypothesis has been rigorously studied and developed as an interesting topic in the economics, sociology, and legal literatures. Early economists focused on testing the relationship between the certainty and severity of sanctions and crime rates, using aggregate data. Soon, this approach was criticized when it was realized that reality does not work in the way that the hypothesis predicts. For example, Grasmick & Green (1986) argued that deterrence is a communicative process. Punishment works through the perception of individuals. They argued that the studies should measure the perceived penalty and threat of punishment, because there is no *a priori* reason to assume that all individuals would consider a particular penalty equally severe or certain. How people perceive the severity and certainty depends upon their own risk preference.

This approach has led some scholars to focus on the individual level and explore deterrence from the perspective of perception theory. This shift in interest found a consistent and moderate effect of perceived certainty of punishment on criminal behavior, while severity failed to be confirmed as a crime deterrent (Piliavin, Gartner, Thornton, & Matsueda, 1986; Paternoster, 1987). Along with these findings, however, methodological and theoretical issues were also identified. Major issues were the problem of temporal order, the nature of certainty and severity,
and model specification in empirical studies. For example, many empirical studies estimated the relationship between expected utility of punishment and criminal behavior by asking questions about the current perception of punishment and self-reported past behavior, using cross-sectional data. Obviously, the temporal order was reversed. To accept the result, one must assume that people’s risk preferences are stable, which is doubtful. Therefore, it was realized that cross-sectional data have an inherent limitation for testing the deterrence hypothesis (Paternoster, Saltzman, Waldo, & Chiricos, 1983; Piliavin, Gartner, Thornton, & Matsueda, 1986, Paternoster, 1987). Grasmick & Green (1986) also questioned the additive combination of perceived severity and perceived certainty. To be effective, both the perceived certainty and severity should be greater than zero. If the perceived severity is zero, the perceived certainty will not work. If the perceived certainty is zero, then the perceived severity will not work. More importantly, the perceived certainty and severity can interact. As with empirical studies of any subject in the social sciences, the issue of model specifications and omitted variables was also criticized.

These limitations led to scholars using longitudinal data for testing the hypothesis (see Paternoster, 1987; Bodman & Maultby, 1997), understanding deterrence within the larger framework of social control and social influence (Meier & Johnson, 1977; Hirschi, 1986), or using exogenous variables (Levitt, 1996, 1997). Nevertheless, the issue of circular and simultaneous causality exists at many levels. For example, the crime rate depends on police deterrence efforts, policy deterrence efforts depend on resources, and the level of resources depends on the crime rate (Benson, Kim, & Rasmussen, 1994; Glaeser, 1999). If the volume of offense arrests increases, then the extra load of law enforcement increases, which may decrease their effectiveness, therefore causing a reduction in arrests and related sanctions (Ehrlich & Brower, 1987).

There also exists literature that disagrees with the predictions of the hypothesis. An example of these disagreements is recidivists, who make a career of committing crime. In their
study on recidivists, Wilson & Abrahamse (1992) claimed that career criminals may temporarily overvalue the benefits of crime and discount punishments compared to other people. A more recent experiment (Gneezy & Rustichini, 2000) found that in some conditions, such as imperfect contracts, the deterrence hypothesis is violated. In their experiment of penalizing late-coming parents at day-care centers, the introduction of a fine increased rule-breakers. Even when the fine was removed, the increased numbers of rule-breakers continued rather than return to the earlier equilibrium without the fine.

Therefore, it is not easy to draw certain conclusions. From the literature, one can only articulate general statements regarding the deterrence hypothesis. The hypothesis predicts that the higher expected utility of punishment deters crime. Punishment is more effective when it is certain and immediate. If adaptation occurs, the effectiveness of punishment decreases. Below, I examine the deterrence effect of a sanction in the WIC program by modeling the perceived certainty of a policy option in the dynamic simulation framework, with the assumption that the perceived severity of the policy option is the same, but not zero for the vendor agents.

4.5.2. Sanctions in the WIC Program

Misuse, fraud, and abuse have been issues from the inception of the WIC program, and several regulations and policies have been developed as a result. For example, according to one study (GAO, 1999), state and local agencies detected relatively more fraud among vendors (approximately 9% of all vendors) compared to participants (0.14%) and employees (4%). Therefore, states have imposed various sanctions for vendors, ranging from warning letters for less serious offenses to disqualification for the most serious offenses.

Among these sanctions, a warning letter was most frequently utilized by stage agencies to deter fraud. The GAO report (1999) showed that more than half of the vendors committing fraud
received at least warning letters in their study period. Table 9 lists the sanctions used by states, including temporary disqualification or suspension, a monetary fine or penalty, sanction points, and other penalties.

<table>
<thead>
<tr>
<th>Type of sanction</th>
<th>Percent of vendors receiving sanction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warning letter only</td>
<td>58</td>
</tr>
<tr>
<td>Temporary disqualification or suspension from the WIC program</td>
<td>25</td>
</tr>
<tr>
<td>Monetary fine or penalty</td>
<td>21</td>
</tr>
<tr>
<td>Sanction points</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
</tr>
<tr>
<td>No warning letter or other sanction</td>
<td>2</td>
</tr>
</tbody>
</table>


Table 9: Type of sanctions imposed by state and the percentage of vendors receiving sanctions – Note that the total percent does not add to 100 because vendors could have received more than one sanction during the survey period.

4.5.3. Implementing a Policy Intervention

I use the same simulation framework that was developed earlier in this chapter. This simulation includes 200 vendors with 10,000 participants. One step in the simulation is considered to be one month. Each participant agent receives vouchers for three months at every third step. Vendor agents record their actual and total sales depending upon the result of the fraud negotiation. The public agency agent evaluates vendors’ sales at each step and categorizes certain vendors as high-risk.

Earlier, I used three criteria based on WIC sales volume to identify outliers. Here, I simplify the mechanism by focusing on only one criterion, the vendor redemption ratio. The
vendor redemption ratio shows the difference between the total voucher costs and the actual charge paid to the vendors. Due to the way the state issues maximum dollar amounts for each voucher, it is very unlikely to see those vendors who charge greater than 90 percent of the maximum costs for every voucher. Based on this fact, I programmed the public agency agent to identify the vendors charging greater than 90 percent of the total voucher costs collected. The public agency agent uses this evaluation method at each step and updates a vendor’s risk status as high-risk whenever the vendor charges greater than 90 percent of the total voucher costs.

Below, I consider the unusual charge of total voucher costs as a vendor violation, define those vendors committing the violation as high-risk, and test the effect of a warning letter in deterring vendor violations in the WIC system. I investigate different policy options, focusing on the percentage of vendors getting a warning letter due to the violation. “No warning letter” is used as a baseline. I increase the percentage of vendors who receive a warning letter to less than 30%, 90%, and 100%. Those receiving a warning letter are randomly decided at each step.

While the public agency agent sends out the warning letter, the vendor agents respond to the policy intervention. In terms of the vendor agent’s response, I utilize the vendor risk propensity property. An assumption is made on the vendor agent’s response to policy options. The response to the warning letter will be more immediate and direct for those who have a less than average risk propensity (< 0.4) and also for those who have a relatively higher risk propensity (> 0.8), compared to those who have a risk propensity in the middle range (0.40 ~ 0.8). Therefore, vendors will respond differently to the warning letter, depending upon their own risk propensity.

While there could be various scenarios regarding policy interventions, here I focus on two simulation scenarios. First, the warning letter is introduced in the middle of the simulation (at Time 150). When the vendor receives a warning letter, the vendor agents committing fraud reduce approximately 10% of total voucher value for those participants who are involved in fraud.
It was programmed to record as the 100% of voucher values was redeemed in the case of fraud. Therefore, there will still be a gap between actual sale values and what the vendor recorded. However, the gap will not be large compared to the situation based on the voucher usage (75%) in Table 5. Because this option influences actual sales of vendors, it is expected that the percentage of high-risk vendors based on the redemption ratio will be significantly reduced.

Second, the first scenario is slightly modified. The public agency agent issues warning letters at Time 100, but removes the intervention at Time 150. The same assumption on the vendor agent’s response is used for this scenario. The percentage of high-risk vendors will be monitored before and after the policy intervention.

4.5.4. Testing Effects of a Warning Letter

Scenario 1

The first step is that the public agency agent sends a warning letter to a certain portion of the high-risk vendors. Once the vendor agents receive the warning letter, their risk propensities will be adjusted, as discussed above. This change in risk propensity influences the outcome of fraud negotiations with participant agents visiting the vendor agents, which ultimately influences the probability of becoming a high-risk vendor.

Table 10 shows the percentage of high-risk vendors in the different policy options at the last five steps of the simulation, when the first scenario is implemented. This table shows that the certainty of being caught had a significant influence on the percentage of vendors violating the redemption ratio criterion, once the policy option was implemented at Time 150. There was approximately 16%–19% of difference between “no warning letter” and warning letters to 100% in terms of the percentage of high-risk vendors.
Table 10: Percentage of high-risk vendors at the last five steps of the simulation in different scenarios. The policy option influenced the interaction of the vendor agents with new comers as well as those participant agents currently involving in fraud at the vendor at Time 150.

<table>
<thead>
<tr>
<th>Time</th>
<th>No warning letter</th>
<th>&lt;30%</th>
<th>&lt;90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>295</td>
<td>19.5</td>
<td>2.5</td>
<td>1.0</td>
<td>1.5</td>
</tr>
<tr>
<td>296</td>
<td>18.5</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>297</td>
<td>18.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>298</td>
<td>20.0</td>
<td>2.5</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>299</td>
<td>17.0</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Figure 26 presents the effect of four policy options. The percentage of high-risk vendors ranged from 6% to 12% in the scenario. The option of ‘no warning letter’ showed the continuous increase of high-risk vendors, when there is no policy intervention. However, in other three scenarios, the percentage significantly decreased after Time 150. Nevertheless, the gap among the three scenarios based on warning letters to <30%, <90%, and 100% was not large, while it seems that warning letters to 100% performed slightly better than warning letters to <30% at the end of the simulation. Approximately 0%~4% of the vendors was identified as high-risk at the end under the three scenarios. Figure 26 shows discontinuity at Time 150, due to the policy option.
Figure 26: Percentage of high-risk vendors as a response to a warning letter when the letter was sent to randomly selected high-risk vendors at Time 150. This policy option influences not only vendor risk propensity, but also the behavior of vendors in terms of recording actual sales for those participants involving in fraud.
Scenario 2

Table 11 shows the percentage of high-risk vendors in the different policy option at the last five steps of the simulation, when the scenario 2 was implemented. This table shows that the certainty of being caught also had a significant influence on the percentage of vendors violating the redemption ratio, once the policy option was implemented at Time 100. However, the percentage increased again after the policy option was removed at Time 150. There was approximately 4%~11% of difference between “no warning letter” and warning letters to 100% in terms of the percentage of high-risk vendors in the WIC system.

<table>
<thead>
<tr>
<th>Time</th>
<th>No warning letter</th>
<th>&lt;30%</th>
<th>&lt;90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>295</td>
<td>17.0</td>
<td>14.0</td>
<td>11.5</td>
<td>13.5</td>
</tr>
<tr>
<td>296</td>
<td>17.5</td>
<td>13.5</td>
<td>8.5</td>
<td>6.5</td>
</tr>
<tr>
<td>297</td>
<td>20.0</td>
<td>12.0</td>
<td>11.0</td>
<td>10.5</td>
</tr>
<tr>
<td>298</td>
<td>17.5</td>
<td>11.0</td>
<td>10.0</td>
<td>6.5</td>
</tr>
<tr>
<td>299</td>
<td>17.0</td>
<td>13.5</td>
<td>9.5</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Table 11: Percentage of high-risk vendors at the last five steps of the simulation in different scenarios. The policy option influenced the interaction of the vendor agents with new comers as well as those participant agents currently involving in fraud between Time 100 and Time 150.

Figure 27 presents the effect of four policy options as above. The overall percentage of high-risk vendors significantly decreased at Time 100, but increased again at Time 150, due to the implementation of the policy option during the times. The percentage of high-risk vendors in the WIC system varied after the policy option was removed, depending upon the options. The gap was relatively large between warning letters to <30% and warning letters to <90% or 100%.

Figure 27 shows discontinuity during the period of the policy intervention.
Figure 27: Percentage of high-risk vendors as a response to a warning letter when the letter was sent to randomly selected high-risk vendors at Time 100. The policy option was removed at Time 150. The option influenced not only vendor risk propensity, but also the behavior of vendors in terms of recording actual sales for those participants involving in fraud.
Future Work

These are very primitive simulations to test the effects of a policy intervention in the simulated WIC system. Several, but more realistic, scenarios for policy intervention can be tested. For example, the difference between the non-linear introduction and removal of the policy option and the proportional increase of sanctions depending upon the number of violations can be examined. High-risk vendors can be withdrawn, and new vendors can be introduced based on their performance. Policy options for fraud among participant agents can be designed along with those for vendor fraud. There are many possible scenarios to test and discuss in order to have some sense of the dynamics of the WIC program, based on different policy options.

4.6. Discussion

4.6.1. Future Trends

When crime is approached from a management perspective, traditional social science models add limited value to understanding the dynamic nature of crime in the real world. Focusing on the nature of criminals and calibrating the causes of crime do not tell us when, how, and why such criminal activities occur in public delivery programs. In particular, when the discussion is based on disciplinary assumptions, there is a high chance that only partial information regarding the complex social issue will be provided. Therefore, as more recent studies suggest, the issue needs to be approached using dynamic frameworks from multidisciplinary or transdisciplinary perspectives.

Crime is difficult to study because the process is hidden in many cases. Traditional research methods and tools have limitations in studying such hidden processes and understanding
the underlying mechanisms. Social simulation models have some promising aspects for studying such an issue of crime, public health, and emergency management. The models allow us to investigate the interactions and interdependency, which is key in understanding the occurrence and diffusion of social issues. Therefore, the applicability of simulation models is more relevant in such areas as crime and public health.

From a policy point of view, social simulation models also have advantages, in that policy analysts can test the consequences of certain policy options into the whole system. A current approach to test the effect of policy interventions is to implement pilot studies. This approach has significant weaknesses and is more resource-intensive compared to social simulation techniques. Public managers cannot implement the same pilot project repeatedly while carefully observing consequences for a long time due to political and administrative constraints. Studying human subjects has ethical and confidential issues. If human and social systems can be studied using agents in a virtual laboratory, it will help research and practice in the areas of policy and management.

4.6.2. Summary

Despite a long history in social thought and the considerable attention in recent years due to events, such as terrorist attacks and hurricane disasters, theoretical progress in the area of crime and deviance has been fragmented (Vaughan, 1999). Traditional studies do not provide useful guidance in understanding such dynamic issues. Crime is an issue that has multiple dimensions that need to be considered. It is not easily understood when a simple and static approach is taken, but requires knowledge on individual decision-making, behavior, context, and the connection to macro patterns (Vaughan, 1998).
Several methodological constraints have added burdens to overdue theoretical progress. The conceptual interest was in investigating the interactions and interdependency in a complex policy system so as to understand the link between the micro behavior of adaptive agents and the macro patterns. I aimed to contribute to the area by linking individual behaviors to emerging statistical and spatial patterns associated with fraud, using a comprehensive framework and implementing an agent-based model.

I framed fraud in public delivery programs as a complex system and explored the underlying mechanisms of such fraud within the framework of rational choice and its modifications. I replicated the distribution of vendors with risk status in the empirical data. This provides me a framework for a further study in the simulated policy system. This effort has obvious applications in the newly developing areas of homeland security and coordinated emergency management (Nunn, 2005; Choi & Brower, 2006), as well as in the more traditional areas of law enforcement and public health.
CHAPTER 5

DISCUSSION

Electric computers, game-theoretic models, and statistical formulas are but instruments after all; it is not they that produce scientific results but the investigator who uses them scientifically.

A. Kaplan (1964)
J. Bronowski (1965)

5.1. Contributions

5.1.1. Theoretical Contributions

There have been a plethora of studies on complexity and management (Stacey, 1996, 2000; Amin & Hausner, 1997; Rosenhead, 1998; Battram, 1999; Sanderson, 2002; Baets, 2005), but not much on how to model complexity for policy analysis and decision support. Therefore, this discussion remained largely abstract for a long time. This dissertation contributes to the body of knowledge by actually implementing these ideas in the study of fraud in a public delivery system using a developing modeling technique, an agent-based model. While there is still much to be done to produce fruitful outcomes from this trial, this dissertation initiated the first step by building a framework to investigate several interesting concepts and questions that have been discussed in the complexity and management literatures.
To provide a practical problem area, this dissertation examined the policy context of Ohio WIC, aiming to identify fraud in practice using a classical store choice model. Store choice models have traditionally been utilized for “good” purposes. The applications of such models include securing spatial accessibility, allocating locations, and calibrating trading areas. This dissertation proposes to use the models for “bad”, undesirable, or stigma-attached activity associated with public goods and services. This study shows that the traditional choice model can serve as an effective tool for policy problems, even in the disadvantaged population, which could potentially undermine the underlying assumptions behind the model.

Chapter 3 also extends the current application of the proximity concept and the use of GIS in practice. This concept and GIS have been used as a secondary component in monitoring fraud in states such as Louisiana and Missouri. Most analyses are done using business intelligence systems. Later, outliers are displayed using GIS. The notion of proximity deterministically confirms the result of the analysis. This dissertation proposed to model the concept of proximity with other attributes, such as size, in a probabilistic manner to describe participants’ store choice behavior and its use for monitoring fraud. This approach allowed me to identify statistical and spatial patterns of fraud in a public delivery program. Therefore, this study presents a different use of the proximity concept and tool by enhancing the current uses.

The current practice of fraud prevention and detection treats and focuses on players in the program separately. For example, public programs independently understand vendor schemes, internal employee schemes, and client schemes. They do not however consider fraud in public programs as an outcome of interactions among the different entities. Even with all its obvious limitations, the modeling in Chapter 3 showed that capturing static interactions using a store choice model can be useful for identifying abnormal spatial patterns of fraudulent vendors and participants.
In line with this understanding, I argued that the traditional crime literature does not provide an appropriate framework to guide the study of fraud, because fraud in public programs is a phenomenon that emerges from dynamic interactions among heterogeneous actors. Chapter 4 suggests framing fraud as a complex system and studying the interactions and interdependency among the actors. Therefore, this study opens a door for understanding and modeling fraud from the perspective of complex interactions in a policy system, thus designing a more realistic and robust research framework for policy analysis.

In other words, traditional social science models have focused on properties, factors, and characteristics of complex phenomena, rather than on the interactions and interdependency leading people to behave in certain ways. While this approach has contributed to the body of knowledge on human behavior, it is also well-acknowledged that it has provided only a partial understanding of the complexity of human behavior. The context in which human behavior occurs and the interactions among humans are frequently assumed to be homogeneous, deterministic, fixed, or linear. Therefore, this approach has serious limitations in providing a useful framework for policy analysis dealing with complex problems among heterogeneous actors who are loosely or tightly interrelated and interact dynamically.

In line with these critiques, I showed the feasibility of agent-based models for policy analysis. I built a simulation model for a policy delivery system. The interdependency and interactions were specified based on the actual WIC business mechanism in Ohio. The behavior of agents was modeled using the store choice and fraud negotiation rules. Chapter 4 shows that agent-based models can successfully implement what the traditional analytic approaches do and can also go beyond. For example, the Huff spatial interaction model was modeled and visualized using the agent-based model. Also, the process of developing as a high-risk vendor due to fraudulent activities has also been visualized by modeling such an interaction and interdependency among players in the WIC system.
This provides another theoretical implication. There have been two distinct approaches in social science disciplines for studying complex social phenomena. One approach focuses primarily on micro behavior, whereas the other approach pays more attention to macro patterns. The link between the two has been conjectured or arbitrarily argued. Complexity and nonlinearity were assumed away in many analyses. Agent-based models provide solid theoretical and methodological foundations for exploring the intriguing links between micro and macro contexts. Chapter 4 explores a methodological bridge between models that explain micro behavior and the macro patterns emerging from the adaptive interactions at the micro level.

The simulation model was validated by comparing the statistics and spatial patterns of fraud with those of the empirical data. If the simulation is accurately specified, the patterns may reasonably replicate what is found in the empirical data. The statistical and spatial patterns that emerge from the interactions among corrupt agents were captured by implementing the integrated framework. This successful replication provides a framework to test some other theories, assumptions, or policy options in the simulated reality.

As an example, I simulated the effect of a policy intervention called a “warning letter”. I presented the results of a computer experiment for the policy intervention. It shows the possibility that much richer experiments can be done within the framework. If the future is plural, one can adaptively respond to the upcoming turbulent future by developing scenarios and continuously modifying the mental model. Simulation does not prove anything; however, simulation can help one understand the reality that they experience and prepare the future with flexible options. Chapter 4 is a trial to enhance such an aspect of policy analysis.

In the field of policy analysis, traditional research has extended existing knowledge by testing policy variables or focusing on specific populations, as other disciplines in social science have practiced. This dissertation introduced an alternative way to extend the body of knowledge by building models of human interaction and decision-making from a systems perspective. This
extension can enhance the understanding of the complexity of policy problems. This study shows that some rules can reasonably explain complex systematic patterns in a policy system, also allowing for simulating the effect of a policy option.

5.1.2. Practical Contributions

The vendor propensity measure based on the spatial interaction model is expected to have immediate practical use in monitoring the performance of vendors participating in the WIC program (Chapter 3). This method can provide an effective vendor management tool. The current procedures for identifying potentially problematic vendors yield a large number of vendors that need to be targeted for field investigation. The field investigations tend to be expensive, reactive, and ad hoc. Implementing a monitoring system based on store choice models can yield a systematic approach to identifying vendors that are likely to commit fraud and allow proactive intervention. This approach can help determine when and where to launch field investigations and can provide evidence, in the form of anomalous behavior, for potential legal action.

Senge (1990) argued that “new insights fail to get put into practice because they conflict with deeply held internal images of how the world works, images that limit us to familiar ways of thinking and acting” (p. 174). These internal images are mental models, which are the picture of one’s reality. Senge (1990) further argues that all we ever have are assumptions, never “truths”, and that we always see the world through our mental models, which are always incomplete due to the inevitable biases in our own ways of thinking. Simulations help researchers and managers bring their mental models to the table by allowing them to build, visualize, and discuss their realities in the virtual laboratory.

Therefore, I expect that the process modeling and assumptions can be reviewed, discussed, modified, and tested by program managers. This evaluation process will provide an
opportunity for improving the understanding of the underlying mechanisms of fraud and facilitate the discussion for further action. In other words, the flexibility of the modeling tool can help include and test various implicit and explicit theories based on practical experience from the field. New assumptions, theories, and paths can be modeled and tested and their consequences can be observed in the simulated policy system, leading to enhanced decision-making processes in organizations.

Policy analysts have long been looking for a way to contribute to the process of decision-making in organizations. There is no shortage of documents indicating that analytic answers based on retrospective evidence have had difficulty in effectively informing decision-making, especially in the area of policy analysis, program evaluation, and organization. Therefore, Weiss (1999) redefined the role of evaluations and analyses as “enlightenment” rather than prompt input to decision-making. Weiss may be right in some sense. However, the assumption behind this argument is one-way communication from academics to practice. Sometimes “enlightenment” comes from practice to academics, because, as I discussed in Chapter 2, the quality of understanding and description can be significantly different whether one is inside or outside the system.

Simulations aim to replicate constructed reality and assist decision-making. Building simulations reveal one’s understanding of reality along with one’s biases. It is difficult to share one’s way of thinking with others in a group. One’s mental model may only be known to those who have had long and deep interactions with the individual. Simulations allow managers and practitioners to indirectly experience the consequences of their mental model. One may not need to wait until others are “enlightened” if more effective decision support tools can be provided to those who make decisions in practice.
5.2. Policy Implications

Two alternative ways of addressing the complexity surrounding fraud in public service delivery programs were introduced in this dissertation. Statistical measures based on store choice models are effective in capturing static interactions at certain points in time; however, the agent-based model provides a dynamic framework for testing several theories and assumptions in the policy system. At least two policy implications resulting from this study can be discussed.

First, Ohio WIC effectively utilized the propensity measure in order to monitor high-risk vendors. This method can be transferred to other state WIC programs or other public service delivery programs. For example, the states of Louisiana and Missouri are building a framework based on GIS and Business Intelligence System for the purpose of monitoring fraud. While, in their framework, proximity serves as secondary component for monitoring fraud, this dissertation proposed to directly incorporate spatial interactions for identifying fraud. This can enhance accountability in public programs.

Second, in practice, many problems are difficult to deal with. The problems can be temporarily mitigated with some coping mechanisms. However, complete elimination may not be feasible. Public agencies need to explore several creative mechanisms to deal with the adaptive behavior of intelligent agents and understand the underlying mechanism to develop more effective monitoring systems. The agent-based model built in this dissertation can serve as a virtual laboratory for testing several policy intervention mechanisms. This can provide some insight into the consequences of interventions for the whole system. Therefore, it maybe a resource-saving approach to current mechanisms that public agencies utilize in practice, such as pilot studies and randomized studies for program evaluations.
5.3. Limitations & Future Research

Many research limitations still remain. Here, I summarized the limitations by separating studies based on store choice and agent-based models.

5.3.1. Vendor Propensity Measures

When fraud is seen as the interactions among players within a system, traditional store choice models have a limitation in that they do not include complex interactions. As this dissertation identified in Chapters 1 and 4, there are several other mechanisms of fraud. It is not clear how to introduce the interactions among participants or vendors to specify the utility function in order to acquire participants’ store choice probabilities, when it is utilized for monitoring fraud. Therefore, while the current models reasonably predict the possibility of fraudulent behavior, it is a very primitive and crude approximation that is based on limited assumptions.

In the Huff model, both factors have a nonlinear nature. As some critiques have pointed out, people do not consider stores beyond a certain distance. The threshold of distance for store choice may remain an empirical question. In Ohio WIC, the basic unit of operation is the county level. Participants are informed of all possible vendors within the administrative boundary. However, I still suspect that there could be a certain threshold of distance, especially in metropolitan areas with many WIC vendors, and the threshold would be different depending on the participants’ preferences. Size measure, based on the number of lanes, may also not have a linear influence on participants’ store choice. This was not addressed when the vendor propensity measure was developed.
In addition, the WIC program does not limit the area that participants can use the voucher as long as benefit exchanges occur in Ohio. Therefore, participants in Cincinnati can show up in Cleveland or in adjoining counties. Some participants can have a job in a neighbor county or the opposite side of the residential county. Thus, they can use their voucher after work while this study used residential addresses to calculate distance. Further, the structure of transportation systems in a county can influence on the flow of the participants. Euclidean distance can produce multiple methodological issues depending upon the environment in which the studies are conducted.

Certainly, there are more advanced models than the Huff spatial interaction models that can be used to explain consumers’ store choice decisions. This research used the Huff model for two reasons. One was data availability. Ohio WIC vendor specialists, as a team, spent more than a year to collect, clean, and continuously update the two variables related to the 1,400 vendors, location and number of lanes. Without extra resources and support, other variables cannot be designed or collected. The other was the purpose of utilizing models in policy analysis. It is certainly attractive for academics to develop more accurate and comprehensive models. However, if the program goal can be reasonably achieved using a simple model, there is no benefit to finding one or more high-risk vendors by developing a sophisticated model in practice. The Huff spatial interaction model provides sufficient utility for the program at the moment.

The assumption behind the vendor propensity measure was that WIC participants’ vendor choices will not be different from the store choice behavior of general consumers, on which the store choice models originally aimed to explain. Therefore, I suspect that abnormal store choice behaviors can signal the possibility of fraudulent behaviors of vendors and participants. The prediction has been a major objective in this study. As a result, the study did not show how such abnormal behavior has been developed or how it will be progressed. It only shows an equilibrium pattern at a certain point in time. The study provided only partial explanations of the findings.
5.3.2. Fraud Simulation

A major limitation of this study is that the agent-based model only captures the interaction between vendors and participants at the current stage. In practice, the complicity of agents in fraud is more complicated than what has been tested in both models. As described in Chapter 4, participants can be involved in fraud among themselves. A group of vendors can be assembled to commit organized crimes. An outside entity can also be involved in the WIC business process. In this dissertation, I hope to have developed a foundation on which to build future studies by focusing on the most common fraud mechanism in the public delivery program. Therefore, this research is still in progress.

While this simulation stepped forward in terms of introducing dynamics in the choice model, the actual WIC program is much more dynamic than the simulation. For example, in my agent-based model, participant agents do not exit or get replaced; whereas in the WIC program, participants frequently enroll and exit as their status changes. However, I simplified this dynamic and made current agents continue to be served once they are randomly generated. I also simplified voucher usage by participant agents. I programmed that every voucher is used by participant agents each iteration. In reality, this may not be true. Some participants may not use their vouchers at all during certain months.

The simulation also has limitations in terms of decision-making. In the current simulation, participant and vendor agents make store choices and fraud decisions right away when they confront each other. In reality, both decisions may take longer and also go through the process of trial and error. Other personal characteristics or situations can influence how decisions are made. The current simulation does not consider many of these complexities in the process of decision-making.
This simulation implements those ideas that I discussed throughout this dissertation within the framework of a utility function. This has advantages and weaknesses. One advantage is that I can utilize some of the previous models and evidence to facilitate the simulation. A weakness is that the simulation still holds limiting assumptions regarding human behavior. A good example of this weakness appears when modeling the situation where a high-risk vendor is replaced by a new vendor. Once this decision is made by a public agency agent, all participant agents recalculate their store choice probabilities without hesitation. In reality, it may take longer to inform participants and make them visit the new vendor.

Agent-based models are developing as an alternative tool for studying complex natural and human phenomena. Compared to statistical or optimization procedures, the process of building and testing the models has not been firmly established. Several parts of my agent-based models are based on heuristics, rather than on parameters that are based on empirical evidence. This aspect of the current models provides an opportunity, as well as a barrier, for extending this study. For example, the process of building a simulation highlights the area on which most research has not paid attention (i.e. we do not know how to model and measure dynamic changes or adjustments of certain propensity due to certain events in human behavior).

This dissertation aimed to go beyond the traditional positivistic approach in policy analysis. Although this does not necessarily mean that research will be solely based on the post-positivistic approach, my models are still under the heavy influence of the positivistic paradigm and rational choice traditions. For example, the dissertation has focused on replicating empirical patterns rather than on building a constructive model for learning. This research needs to be improved in many ways in order to fully achieve the original objective.
REFERENCES

Cited References


140


Influential readings not cited in the dissertation


Technical References


Other Websites

USDA. http://www.usda.gov

Ohio WIC. http://www.odh.ohio.gov/odhPrograms/ns/wicn/wic1.aspx

GAO. http://www.gao.gov/


VENDOR PROPENSITY MEASURES

APPENDIX A

cd D:\VendorAnalysis\Data\Cuyahoga
clear all;
msg = sprintf('data reading...'); disp(msg)
load org.txt; % The record should have ID, X, Y coordinates (e.g. customer)
load tgt.txt; % The record should have ID, X, Y coordinates, and weight (e.g. vendor)
msg = sprintf('End of data reading...'); disp(msg)
[row_org col_org] = size(org);
[row_tgt col_tgt] = size(tgt);
probVal = zeros(row_org + 1, row_tgt + 1);
result = fopen('result.txt','w');
tempfile = fopen('TempWorkforCalProb.txt','w');
fprintf(result, '%s', 'Origin');
fprintf(tempfile, '%s', '0000');
for j = 1:row_tgt
fprintf(result, ', %d', tgt(j,1));
fprintf(tempfile, ', %d', tgt(j,1));
probVal(1,j+1)=tgt(j,1);
end
fprintf(result,', TotalNum
');
fprintf(tempfile,', -9999
');
count = 0;
for i = 1:row_org
fprintf(result, '%s', num2str(org(i,1),'%10.0f'));
fprintf(tempfile, '%s', num2str(org(i,1),'%10.0f'));
probVal(i+1,1)=org(i,1);
totnumer = 0;
for j = 1:row_tgt
dist(1,j) = sqrt(((org(i,2)-tgt(j,2))^2) + ((org(i,3)-tgt(j,3))^2));
%For the parameter values (1)
totnumer = totnumer + ((1/(dist(1,j)^2)) * (tgt(j,4)^1));
% fprintf(result, ', %d, %d, %f', org(i,1), tgt(j,1), dist);
% fprintf(result, ', %f', totnumer);% count = count + 1;
if (mod(count, row_tgt)==0)
for jj = 1:row_tgt
probVal(i+1,jj+1) = ((1/(dist(1,jj)^1))*tgt(jj,4))/totnumer
end
end
fprintf(result, ', %f', probVal(i+1,jj+1));
%For the parameter values (2)
fprintf(result, ', %f', ((1/(dist(1,jj)^2))*(tgt(jj,4)^1))/totnumer);
fprintf(result, ', %f', dist(1,jj));
%For the parameter values (3)
fprintf(tempfile, ', %f', ((1/(dist(1,jj)^2))*(tgt(jj,4)^1))/totnumer);
fprintf(tempfile, ', %f', dist(1,jj));
end
if (totnumer == 0)
msg = sprintf('**************************'); disp(msg)
msg = sprintf('Error data is %10.0f!', org(i,1)); disp(msg)
msg = sprintf('**************************'); disp(msg)
fprintf(result, ', %f', -9999);
fprintf(tempfile, ', %f', -9999);
continue
end
fprintf(result, ', %f 
', totnumer);
fprintf(tempfile, ', %f 
', totnumer);
%            fprintf(result, '
');
end
end
end
fprintf(result, '
End');
% fprintf(tempfile, '
End');
close(result);
close(tempfile);
msg = sprintf('End of the distance calculation process...'); disp(msg)
clear org;
clear i;
clear j;
load tgt.txt;
load lnk.txt;  % PartID, VenderID
probVal = load('TempWorkforCalProb.txt');  % PartID, # of row_tgt's, totnumer
msg = sprintf('End of link & prob file loading...'); disp(msg)
statresult = fopen('stat.txt','w');
staterror = fopen('error.txt','w');
fprintf(statresult, '************************
');
fprintf(statresult, '****** Statistics ******
');
fprintf(statresult, '************************
');
fprintf(statresult, '
');
fprintf(statresult, 'VenderID');
fprintf(statresult, ', Mean');
fprintf(statresult, ', Variance');
fprintf(statresult, ', Std.Dev');
fprintf(statresult, ', Max');
fprintf(statresult, ', Min');
fprintf(statresult, ', Total Part''s');
fprintf(statresult, ', Excluded Part''s
');
flgerror = 0;
for i=1:row_tgt
%Find Participants' locations in the Link corresponding Vender ID
[parts_IDXs]=find(lnk(:,2)==tgt(i,1));
%Find the number of Participant corresponding Vender
nPartperVend = size(parts_IDXs,1);
%The Location Index of the Vender in the Prob Table
[probVal_ID]=find(probVal(1,:)==tgt(i,1));
clear ProbPerVend;
numError = 0;
for j=1:nPartperVend
%Get the ID of the j-th participant
part_ID = lnk(parts_IDXs(j,1),1);
pID = find(probVal(1,:)==part_ID(1,1));
VID = find(probVal(1,:)==tgt(i,1));
[pID_row, pID_col] = size(pID);
if (pID_col == 0) | (pID_row == 0)
    if (flgerror == 0)
        fprintf(staterror, '%d: %s', tgt(i,1),num2str(part_ID(1,1),'%10.0f'));
        flgerror = 1;
    else
        fprintf(staterror, ', %s', num2str(part_ID(1,1), '%10.0f'));
    end
else
    %...
numError = numError + 1;
else
    ProbPerVend(1,j-numError) = probVal(pID,vID);
end

gerror = 0;
fprintf(staterror,'
');
fprintf(statresult, '%d', tgt(i,1));
fprintf(statresult, ', %f', mean(ProbPerVend));
fprintf(statresult, ', %f', var(ProbPerVend));
fprintf(statresult, ', %f', std(ProbPerVend));
fprintf(statresult, ', %f', max(ProbPerVend));
fprintf(statresult, ', %f', min(ProbPerVend));
fprintf(statresult, ', %d', nPartperVend);
fprintf(statresult, ', %d', numError);
fprintf(statresult, '\n');
end

fprintf(statresult, '\nEND');
fclose(staterror);
fclose(statresult);
delete('D:\VendorAnalysis\Data\Cuyahoga\TempWorkforCalProb.txt');
mesg = sprintf('End of whole processes...'); disp(mesg)
APPENDIX B

FRAUD SIMULATION USER INTERFACE

This simulation presents spatial patterns emerged from the interaction between fraudulent agents in a policy system. The policy system consists of key artificial agents such as participants, service providers, and a public agency.

Participants visit service providers to redeem their benefits. The choices is made by optimizing store choice probability based on the Huff model at the beginning. Some agents offer fraud to other agents who was randomly selected. Depending upon their negotiation on fraud, agents' risk propensity changes. I expect that this process will form certain spatial pattern at macro level and reveal underlying mechanisms.
APPENDIX C

CLASS DIAGRAM OF FRAUD SIMULATION
APPENDIX D

FRAUD SIMULATION

[Agent.class]

```java
package sim.app.fraudsim;

import java.lang.Math;
import java.awt.geom.*;
import java.awt.*;
import sim.portrayal.*;
import sim.engine.*;
import sim.util.*;

public abstract /*strictfp*/ class Agent extends SimplePortrayal2D implements Steppable
{
    public String id;
    public Double2D agentLocation;
    public int intID = -1;

    public Agent( String id, Double2D location ) {
        this.id = id;
        this.agentLocation = location;
    }

    double distanceSquared( final Double2D loc1, Double2D loc2 )
    {
        return( (loc1.x-loc2.x)*(loc1.x-loc2.x)+(loc1.y-loc2.y)*(loc1.y-loc2.y) );
    }

    public abstract String getType();
}
```

[Coupon.class]

```java
package sim.app.fraudsim;

import java.lang.*;

public class Coupon {
    int CID;
    double value;
    boolean used;

    public Coupon( ) {
        this.CID = CID;
        this.value = value;
        this.used = used;
    }
```
package sim.app.fraudsim;
import java.util.*;
import java.io.*;
import java.lang.*;
import java.awt.*;
import sim.portrayal.*;
import sim.util.*;
import sim.engine.*;
import java.text.DecimalFormat;
public /*strictfp*/ class Participant extends Agent implements Steppable
{
static final long serialVersionUID = 3L;
public double prp;
public int race;
public int category;
public int edu;
public int vendorVisited;
public double randomChoice;
public Coupon[] myCoupons;
public int[] selectionOrder;   // vendor ID's sorted according to the huff's model
public int clidx;   // current index in the above selection order[]
protected Color partColor;
public boolean change_vendor;
public int myVendor;
public int myVendor_type;  // no. of lanes
private int behavior_type;
public PublicAgency myagency = null;
public void set_public_agency(PublicAgency _a) {
    myagency = _a;
}

public Participant ( String id, Double2D location, double prp, double randomChoice, int num_active_vendor ) {
    super(id, location);
    try {
        intID = Integer.parseInt( id.substring(4) );    // "Part"
    } catch( IndexOutOfBoundsException e )   {
        System.out.println( "Exception generated: " + e );
        System.exit(1);
    } catch( NumberFormatException e )   {
        System.out.println( "Exception generated: " + e );
        System.exit(1);
    }
    this.prp = prp;
    this.randomChoice = randomChoice;
    this.partColor = new Color (100,100,100);
    selectionOrder = new int [num_active_vendor];
    myCoupons = new Coupon[3];
    for (int i=0; i<3; i++)
        myCoupons[i] = new Coupon();
    this.change_vendor=true;
}

public Object domRandomChoice() { return new Interval(0.0, 1.0); }
public double getRandomChoice() { return randomChoice; }
public void setRandomChoice( double x ) { if (x >= 0 && x <= 1) randomChoice = x; }

public void vendorSelectionProbability ( Vendor[] v )
{
    Vector myprobl = new Vector();
    duoIntDouble aprob;
    double n;
    double probSum = 0;
    double locDistance;
    for (int i=0; i<v.length; i++)
    {
        if (myagency.vendor_status[i] != 1)
continue;
locDistance = distanceSquared(v[i].agentLocation, this.agentLocation);

n = (v[i].size * Math.pow(locDistance, -2.0));  // Huff
aprob = new duoIntDouble(v[i].intID, n);
myprob1.add(aprob);
probSum = probSum + n;
}

for (int i = 0; i < myprob1.size(); i++)
{
aprob = (duoIntDouble) myprob1.get(i);
aprob.dDouble = aprob.dDouble / probSum;
myprob1.set(i, aprob);
}

Collections.sort(myprob1);

for (int i = 0; i < myprob1.size(); i++)
{
aprob = (duoIntDouble) myprob1.get(i);
selectionOrder[i] = aprob.iInt;
}

myVendor = selectionOrder[0];
partColor = v[myVendor].activeVendorColor;
cidx = 0;

public int fraud_negotiation ( Vendor[] v )
{
int i = 0;
for (i = 0; i < myCoupons.length; i++)
{
if (myCoupons[i].used == false)
break;
}

if (change_vendor == true)
{
for (int j = cidx; j < selectionOrder.length; j++)
{
myVendor_type = v[selectionOrder[cidx]].size;
boolean Vendor_start_fraud = v[selectionOrder[cidx]].fraud_neg_v();
boolean part_start_fraud = fraud_neg_p();
double usage_amount = myCoupons[i].value * usage();

if (vendor_start_fraud)
{
if (!part_start_fraud)
{
updatePrp(-0.0001);
v[selectionOrder[cidx]].vrp_change(true);
behavior_type = 1;
}
else if (part_start_fraud)
{
updatePrp(0.0001);
v[selectionOrder[cidx]].vrp_change(true);
v[selectionOrder[cidx]].updateActualSales(myCoupons[i].value);
v[selectionOrder[cidx]].updateTotalSales(myCoupons[i].value);
v[selectionOrder[cidx]].collect_part(true);
behavior_type = 3;
}
}
else if (!vendor_start_fraud)
{
if (part_start_fraud)
{
updatePrp(0.0001);
v[selectionOrder[cidx]].vrp_change(false);
behavior_type = 2;
}
else if (!part_start_fraud)
{
v[selectionOrder[cidx]].updateActualSales(usage_amount);
v[selectionOrder[cidx]].updateTotalSales(myCoupons[i].value);
v[selectionOrder[cidx]].collect_part(true);
behavior_type = 4;
}
}

if (behavior_type <= 2) {  // decision on vendor change
if (cidx == selectionOrder.length) {
cidx = 0;
}
}
else {
    cidx++;
    if (cidx == selectionOrder.length)
        cidx = 0;
    change_vendor=true;
} else {
    change_vendor=false;
    break;
}

myVendor=v[selectionOrder[cidx]].intID;
}

else {
    double usage_amount = myCoupons[i].value * usage();
    if (behavior_type == 3) {
        updatePrp(0.0001);
        v[myVendor].vrp_change(true);
        v[myVendor].updateActualSales(myCoupons[i].value);
        v[myVendor].updateTotalSales(myCoupons[i].value);
        v[myVendor].collect_part(true);
    }
    if (behavior_type == 4) {
        v[myVendor].updateActualSales(usage_amount);
        v[myVendor].updateTotalSales(myCoupons[i].value);
        v[myVendor].collect_part(true);
    }
}
return myVendor;
}

public double usage() {
    double usage = 0.0;
    Random x = new Random();
    do {
        usage = x.nextGaussian() * 0.10 + 0.75;
    } while (usage < 0.04 || usage > 1.00);
    return usage;
}

public void updatePrp ( double t_prp ) {
    prp = prp + t_prp;
    if (prp <= 0.0001)
        prp = 0.01;
    else if (prp >= 0.9999)
        prp = 1.00;
}

public boolean fraud_neg_p() {
    boolean fraud = false;
    Random pRand = new Random();
    double x = pRand.nextDouble();
    if (prp < 0.25) {
        if (x < 0.01) fraud = true;
    } else if (prp >= 0.25 && prp < 0.5) {
        if (x < 0.1) fraud = true;
    } else if (prp >= 0.5 && prp < 0.75) {
        if (x < 0.2) fraud = true;
    } else if (prp >= 0.75 && prp < 0.9) {
        if (x < 0.5) fraud = true;
    } else
        fraud = true;
    return fraud;
}
public void tracePartHistory ( String stringOutput )
{
    String filename = "sim/app/fraudsim/data/Part_History.txt";
    File aFile = new File (filename);
    try
    {
        FileOutputStream myFile = new FileOutputStream ( aFile, true );
        PrintWriter pw = new PrintWriter (myFile);
        pw.println (stringOutput );
        pw.close();
    }
    catch (IOException e)
    {
        System.out.println(e);
    }
}

DecimalFormat my_df = new DecimalFormat("###0.00");

int steps = 0;  // steps
public void step( final SimState state )  {
    FraudSim fs = (FraudSim) state;
    long current_step = fs.schedule.getSteps();
    int i;
    for (i = 0; i < 3; i++) {
        if (myCoupons[i].used == false)
            break;
    }
    if (i <= 2 ) {
        if (state.random.nextBoolean(randomChoice)) {
            if (behavior_type != 3 && prp > 0.6 ) {
                Random random = new Random();
                myVendor = random.nextInt(fs.Num_Vend);
                while (myagency.vendor_status[myVendor] != 1);
                partColor = fs.vends[myVendor].activeVendorColor;
                myVendor_type = fs.vends[myVendor].size;
                boolean vendor_start_fraud = fs.vends[myVendor].fraud_neg_v();
                boolean part_start_fraud = fraud_neg_p();
                double usage_amount = myCoupons[i].value * usage();

                if (vendor_start_fraud) {
                    if (!part_start_fraud) {
                        updatePrp(-0.0001);
                        fs.vends[myVendor].vrp_change(true);
                        behavior_type = 1;
                        change_vendor=true;
                    }
                    else if (part_start_fraud) {
                        updatePrp(0.0001);
                        fs.vends[myVendor].vrp_change(true);
                        fs.vends[myVendor].updateActualSales(myCoupons[i].value);
                        fs.vends[myVendor].updateTotalSales(myCoupons[i].value);
                        fs.vends[myVendor].collect_part(true);
                        behavior_type = 3;
                        change_vendor=false;
                    }
                }
                if (!vendor_start_fraud) {
                    if (part_start_fraud) {
                        updatePrp(0.0001);
                        fs.vends[myVendor].vrp_change(true);
                        updateActualSales(usage_amount);
                        fs.vends[myVendor].updateActualSales(myCoupons[i].value);
                        fs.vends[myVendor].updateTotalSales(myCoupons[i].value);
                        fs.vends[myVendor].collect_part(true);
                        behavior_type = 4;
                        change_vendor=true;
                    }
                    else if (!part_start_fraud) {
                        updatePrp(0.0001);
                        fs.vends[myVendor].updateActualSales(usage_amount);
                        fs.vends[myVendor].updateActualSales(myCoupons[i].value);
                        fs.vends[myVendor].updateTotalSales(myCoupons[i].value);
                        fs.vends[myVendor].collect_part(true);
                        behavior_type = 4;
                        change_vendor=true;
                    }
                }
            }
        }
    }
}
public final void draw(Object object, Graphics2D graphics, DrawInfo2D info) {
    double diamx = info.draw.width*FraudSim.DIAMETER;
    double diamy = info.draw.height*FraudSim.DIAMETER;
    double diamx_p = diamx * this.prp;
    double diamy_p = diamy * this.prp;
    graphics.setColor(partColor);
    graphics.fillOval((int)(info.draw.x-diamx_p/2),(int)(info.draw.y-diamy_p/2),
        (int)(diamx_p),(int)(diamy_p));
}

public String getType () { return "Participant"; }

* This 'Participant' class includes two interaction rules, the store choice rule and the fraud negotiation rule.
package sim.app.fraudsim;
import java.io.*;
import java.util.*;
import java.lang.*;
import java.util.Random;
import java.awt.*;
import java.text.DecimalFormat;
import sim.portrayal.*;
import sim.util.*;
import sim.engine.*;

public /*strictfp*/ class Vendor extends Agent implements Steppable {
  static final long serialVersionUID = 3L;
  public double vrp;
  public int size;
  public int tiptype;
  public int contract;
  public int peer;
  public double hr;
  public boolean active; // true = active, false = inactive
  public double actualSales; // $ actual sales
  public double totalSales; // $ state paid
  public double no_participants;
  public boolean Display;
  protected Color inactiveVendorColor;
  protected Color activeVendorColor;
  public static Color[] vendorColors = {
      new Color(227, 26, 28),
      new Color(255, 127, 0),
      new Color(197, 27, 125),
      new Color(166, 206, 227),
      new Color(31, 120, 180),
      new Color(178, 223, 138),
      new Color(51, 160, 44),
      new Color(253, 191, 111),
      new Color(202, 178, 214),
      new Color(106, 61, 154)};

  public Vendor ( String id, Double2D location, int size, double vrp )
  {
    super( id, location );
    try {
      intID = Integer.parseInt( id.substring(4) ); // Vend
    } catch ( IndexOutOfBoundsException e )
    {
      System.out.println( "Exception generated: " + e );
      System.exit(1);
    }
    catch ( NumberFormatException e )
    {
      System.out.println( "Exception generated: " + e );
      System.exit(1);
    }
    this.size = size;
    this.vrp = vrp;
    this.Display = Display;
    this.actualSales = actualSales;
    this.totalSales = totalSales;
    this.no_participants = no_participants;
    this.inactiveVendorColor = new Color(255,255,255);
    Random randomColor = new Random(); // drawing vendors
    if (intID < 10) 
      activeVendorColor = vendorColors[intID];
    else 
      activeVendorColor = new Color (randomColor.nextInt(255),
      randomColor.nextInt(255),
      randomColor.nextInt(255));
    active = false;

    /* // create a file for output

    // FileOutputStream myFile;
    
    public Vendor ( String id, Double2D location, int size, double vrp )
    {
      super( id, location );
      try {
        intID = Integer.parseInt( id.substring(4) ); // Vend
      } catch ( IndexOutOfBoundsException e )
      {
        System.out.println( "Exception generated: " + e );
        System.exit(1);
      }
      catch ( NumberFormatException e )
      {
        System.out.println( "Exception generated: " + e );
        System.exit(1);
      }
      this.size = size;
      this.vrp = vrp;
      this.Display = Display;
      this.actualSales = actualSales;
      this.totalSales = totalSales;
      this.no_participants = no_participants;
      this.inactiveVendorColor = new Color(255,255,255);
      Random randomColor = new Random(); // drawing vendors
      if (intID < 10) 
        activeVendorColor = vendorColors[intID];
      else 
        activeVendorColor = new Color (randomColor.nextInt(255),
        randomColor.nextInt(255),
        randomColor.nextInt(255));
      active = false;
    }
*/
String filename = "sim/app/fraudsim/data/Vend_History.txt";
File aFile = new File (filename);
try
{
    myFile = new FileOutputStream ( aFile );
}
catch (IOException e)
{
    System.out.println(e);
}
*/

public final boolean isActive() { return active; }
public final void setActive( boolean x ) { active = x; }
public boolean setDisplay() { return Display; }
public void setDisplay ( boolean x ) { if(x == true | x == false) Display = x; }

public boolean fraud_neg_v()
{
    Random vRand = new Random();
    double x = vRand.nextDouble();
    boolean fraud = false;
    if (size < 3) {
        if (vrp < 0.3) {
            if (x < 0.1) fraud = true;
        } else if (vrp >= 0.3 && vrp < 0.4) {
            if (x < 0.3) fraud = true;
        } else if (vrp >= 0.4 && vrp < 0.5) {
            if (x < 0.4) fraud = true;
        } else if (vrp >= 0.5 && vrp < 0.6) {
            if (x < 0.5) fraud = true;
        } else if (vrp >= 0.6 && vrp < 0.7) {
            if (x < 0.6) fraud = true;
        } else if (vrp >= 0.7 && vrp < 0.8) {
            if (x < 0.7) fraud = true;
        } else if (vrp >= 0.8 && vrp < 0.9) {
            if (x < 0.8) fraud = true;
        } else {
            if (x < 0.95) fraud = true;
        }
    } else if (size <6) {
        if (vrp < 0.3) {
            if (x < 0.1) fraud = true;
        } else if (vrp >= 0.3 && vrp < 0.5) {
            if (x < 0.3) fraud = true;
        } else if (vrp >= 0.5 && vrp < 0.8) {
            if (x < 0.5) fraud = true;
        } else {
            if (x < 0.6) fraud = true;
        }
    } else {
        if (vrp < 0.3) {
            if (x < 0.0001) fraud = true;
        } else if (vrp >= 0.3 && vrp < 0.5) {
            if (x < 0.001) fraud = true;
        } else if (vrp >= 0.5 && vrp < 0.9) {
            if (x < 0.01) fraud = true;
        } else {
            if (x < 0.1) fraud = true;
        }
    }
    return fraud;
public void updateVrp ()
{
    if (vrp <= 0.0001)
        vrp = 0.001;
    else if (vrp >= 0.9999)
        vrp = 1.0000;
}

public void vrp_change( boolean change ) // due to fraud
{
    double _change;
    Random x = new Random();
    if (change == true) {
        if (size < 3) {
            do {
                _change = x.nextGaussian() * 0.00001 + 0.00001;
            } while (_change < 0.000001 || _change > 0.01);
            vrp = vrp + _change;
        }
        else if (size < 10) {
            do {
                _change = x.nextGaussian() * 0.00001 + 0.000001;
            } while (_change < 0.000001 || _change > 0.01);
            vrp = vrp + _change;
        }
        else {
            do {
                _change = x.nextGaussian() * 0.00001 + 0.0000001;
            } while (_change < 0.0000001 || _change > 0.01);
            vrp = vrp + _change;
        }
    }
    else {
        if (size < 3) {
            do {
                _change = (x.nextGaussian() * 0.00001) + 0.000001;
            } while (_change < 0.0000001 || _change > 0.01);
            vrp = vrp - _change;
        }
        else if (size < 10) {
            do {
                _change = (x.nextGaussian() * 0.00001) + 0.00001;
            } while (_change < 0.000001 || _change > 0.01);
            vrp = vrp - _change;
        }
        else {
            do {
                _change = (x.nextGaussian() * 0.00001) + 0.0001;
            } while (_change < 0.000001 || _change > 0.01);
            vrp = vrp - _change;
        }
    }
    updateVrp();
}

public void updateActualSales (double salesvalue)
{
    actualSales = actualSales + salesvalue;
}

public void updateTotalSales (double salesvalue)
{
    totalSales = totalSales + salesvalue;
}

public void collect_part (boolean exchange)
{
    if (exchange == true)
        no_participants = no_participants+1;
}

public void get_warning_letter() // behavir on warning letter
{
    double _rx;
    Random x = new Random();
    if (vrp < 0.40)
do {
    \_rx = x.nextGaussian() \* 0.001 + 0.001;
    } while (_rx < 0.000001 || _rx > 0.001);
vrp = vrp - \_rx;
}
else if (vrp >= 0.40 && vrp < 0.80)
{
    do {
        \_rx = x.nextGaussian() \* 0.001 + 0.00001;
        } while (_rx < 0.000001 || _rx > 0.001);
    vrp = vrp - _rx;
}
else
{
    do {
        \_rx = x.nextGaussian() \* 0.001 + 0.001;
        } while (_rx < 0.000001 || _rx > 0.001);
    vrp = vrp - _rx;
}

public void traceVendHistory ( String stringOutput )
{
    String filename = "sim/app/fraudsim/data/Vend_History.txt";
    File aFile = new File (filename);
    try
    {
        FileOutputStream myFile = new FileOutputStream ( aFile, true );
        PrintWriter pw = new PrintWriter (myFile);
        pw.println (stringOutput );
        pw.close();
    }
    catch (IOException e)
    {
        System.out.println(e);
    }
}

public int drawVendorSize = 0;
public int drawVendorSize ()
{
    if (active == true)
    {
        if (size <= 3)
            drawVendorSize = size;
        else if (size > 3)
            drawVendorSize = 4;
    }
    else
        drawVendorSize = 0;
    return drawVendorSize;
}

DecimalFormat my_df = new DecimalFormat("###0.00");
long steps = 0;  // steps
public void step( final SimState state )
{
    FraudSim fs = (FraudSim) state;
    long current_step = fs.schedule.getSteps();
    steps = current_step;
    /*
    if (active) {
        String StrOutput = current_step + ", " + this.intID + ", " + size + ", " + my_df.format(vrp) + ", " + my_df.format(actualSales) + ", " + my_df.format(totalSales) + ", " + no_participants;
        this.traceVendHistory(StrOutput);
    }
    */
}

public Font myFont = new Font("SansSerif", Font.BOLD, 12);
public final void draw(Object object, Graphics2D graphics, DrawInfo2D info)
{
    if (Display)
if (steps > 50)
{
    if (vrp < 0.80)
        activeVendorColor = activeVendorColor.white;
}

if (steps > 10)
{
    if (vrp > 0.80)
        myFont = new Font("SansSerif", Font.BOLD, 30);
}

double diamx = info.draw.width*FraudSim.DIAMETER;
double diamy = info.draw.height*FraudSim.DIAMETER;
this.drawVendorSize();
double diamx_v = diamx + (diamx*this.drawVendorSize);
double diamy_v = diamy + (diamy*this.drawVendorSize);

if (isActive())
    graphics.setColor( activeVendorColor );
else
    graphics.setColor( inactiveVendorColor );

graphics.fillRect((int)(info.draw.x-diamx_v/2),
    (int)(info.draw.y-diamy_v/2),(int)(diamx_v),(int)(diamy_v));

if (isActive())
{
    graphics.setFont( myFont );
    graphics.setColor( Color.black );
    String vrp_str = Double.toString(vrp);
    if (vrp_str.length() >= 4)
        graphics.drawString( vrp_str.substring(0, 4),
            (int)(info.draw.x-diamx/2), (int)(info.draw.y-diamy/2) );
    else
        graphics.drawString(vrp_str, (int)(info.draw.x-diamx/2),
            (int)(info.draw.y-diamy/2) );
}

public String getType () { return "Vendor"; } // type
public class PublicAgency implements Steppable
{
  static final long serialVersionUID = 3L;
  public int lastCID;
  public int num_vend;
  public int num_inactive_vend;

  public int[] vendor_status; // 0 - inactive, 1 - active, 2 - to be active
  public int[] vendor_risk;
  public int[] vendor_warning_letter;

  // FileOutputStream myFile;

  public PublicAgency (int _nv, int _ninactive)
  {
    num_vend = _nv;
    num_inactive_vend = _ninactive;
    vendor_status = new int [num_vend];
    vendor_risk = new int [num_vend];
    vendor_warning_letter = new int [num_vend];
    int i;
    for (i=0; i<num_vend; i++) {
      if (i < num_vend-num_inactive_vend)
        vendor_status[i] = 1;
      else
        vendor_status[i] = 0;
      vendor_risk[i] = 0;
      vendor_warning_letter[i] = 0;
    }
    this.lastCID = lastCID;

    // create a file for output
    // String filename = "sim/app/fraudsim/data/Pa_History.txt";
    // File aFile = new File (filename);
    // try
    // {
    //   myFile = new FileOutputStream ( aFile );
    // }
    // catch (IOException e)
    // {
    //   System.out.println(e);
    // }
  }

  public void issueCoupons ( Participant parts )
  {
    Coupon newCoupon = new Coupon();
    for (int i=0; i<3; i++)
    {
      parts.myCoupons[i] = newCoupon;
      newCoupon.CID = lastCID+1;
      newCoupon.value = randomCvalue();
      newCoupon.used = false;
    }
  }

  public double randomCvalue ()
  {
    double Cvalue;
    double temp;
    Random randomValue = new Random();
    do { temp = randomValue.nextGaussian() * 0.40 + 0.45; // (Gaussian * sigma) + mu }
    while ( temp < 0.02 || temp > 1);
    return Cvalue = temp * 100;
public int check_vendor (int i)
{
    return vendor_status[i];
}

public int get_random_inactive_vendor ()
{
    int temp;
    Random random = new Random();
    do {
        temp = random.nextInt(num_vend-1);
    } while ( vendor_status[temp] != 0 );
    return temp;
} /*
public void find_outliers_1 (Vendor[] vends)
{
    Vector myvalues = new Vector();
    duoIntDouble entry1;
    double value;
    int percent10;
    int i;
    for (i=0; i<vends.length; i++) {
        if (vendor_status[i] != 1)
            continue;
        value = vends[i].totalSales / vends[i].size;
        entry1 = new duoIntDouble(i, value);
        myvalues.add(entry1);
    }
    Collections.sort(myvalues);
    percent10 = (int) (myvalues.size() * 0.1);
    for (i=0; i<percent10; i++) {
        entry1 = (duoIntDouble) myvalues.get(i);
        vendor_risk[entry1.iInt] ++;
    }
}
*/
public void find_outliers_2 (Vendor[] vends)
{
    Vector myvalues = new Vector();
    duoIntDouble entry2;
    double value;
    int percent10;
    for (int i=0; i<vends.length; i++) {
        if (vendor_status[i] != 1)
            continue;
        value = vends[i].actualSales / vends[i].totalSales;
        if (value > 0.9)
            vendor_risk[i] +=1;
        entry2 = new duoIntDouble(i, value);
        myvalues.add(entry2);
    }
    Collections.sort(myvalues);
    percent10 = (int) (myvalues.size() * 0.1);
    for (i=0; i<percent10; i++) {
        entry2 = (duoIntDouble) myvalues.get(i);
        vendor_risk[entry2.iInt] ++;
    }
} /*
public void find_outliers_3 (Vendor[] vends)
{
    Vector myvalues = new Vector();
    duoIntDouble entry3;
    double value;
    int percent10;
    int i;
    for (i=0; i<vends.length; i++) {
        if (vendor_status[i] != 1)
            continue;
        value = vends[i].totalSales / vends[i].no_participants;
        entry3 = new duoIntDouble(i, value);
        myvalues.add(entry3);
    }
} */
public void tracePaHistory ( String stringOutput )
{
    String filename = "sim/app/fraudsim/data/Pa_History.txt";
    File aFile = new File (filename);
    try
    {
        FileOutputStream myFile = new FileOutputStream ( aFile, true );
        PrintWriter pw = new PrintWriter (myFile);
        pw.println (stringOutput);
        pw.close();
    }
    catch (IOException e)
    {
        System.out.println(e);
    }
}

public void set_huff_model(FraudSim fs)
{
    for (int i=0; i<num_vend; i++) {  // vendor color setup and printing outputs
        if (check_vendor(i) >0)
            fs.vends[i].setActive(true);
        else
            fs.vends[i].setActive(false);
    }
    for (int j=0; j< fs.parts.length; j++)  {
        fs.parts[j].vendorSelectionProbability(fs.vends);
    }
}

public void step ( final SimState state )
{
    FraudSim fs = (FraudSim) state;
    long current_step = fs.schedule.getSteps();
    int i, k;
    if ( (current_step % 3) == 0) {
        for (k=0; k< fs.parts.length; k++)
        {
            issueCoupons( fs.parts[k] );
        }
    }
    // find_outliers_1 (fs.vends);
    find_outliers_2 (fs.vends);
    // find_outliers_3 (fs.vends);
    for (i=0; i<num_vend; i++) {  // every step
        if (vendor_status[i] != 1)
            continue;
        fs.vends[i].totalSales = 0.0;
        fs.vends[i].actualSales = 0.0;
        fs.vends[i].no_participants = 0;
        if (vendor_risk[i] == 1)  {
            Random random = new Random();
            double x = random.nextDouble();
            if (x <0.9)
            {
                vendor_warning_letter[i] += 1;
                fs.vends[i].get_warning_letter();
            }
            vendor_risk[i] = 0;
        }
        String StrOutput = current_step +", " + fs.vends[i].intID +", " +
        vendor_risk[fs.vends[i].intID];
        this.tracePaHistory(StrOutput);
        vendor_risk[i] = 0;
}
if (vendor_warning_letter[i] >= 3) {
    vendor_status[i] = 0;
    vendor_status[get_random_inactive_vendor()] = 2;
}
}

/*
if ( (current_step % 6) == 0) { // every six steps
    int count_active_vend = num_vend - num_inactive_vend;
    double limit = count_active_vend * 0.2;
    double count = 0.0;
    while (count <= limit)
    {
        int temp;
        Random random = new Random();
        do
        {
            temp = random.nextInt(num_vend);
        } while ( vendor_status[temp] == 0 ); // being chance of same vendor selection
        if (vendor_risk[temp] >= 3)
        {
            vendor_status[temp] = 0;
            vendor_status[get_random_inactive_vendor()] = 2;
        }
        count++;
    }
    for (k=0; k< fs.parts.length; k++)
    {
        fs.parts[k].vendorSelectionProbability(fs.vends);
    }
    System.out.println( "6 month process is done." );
}
*/
for (i=0; i<num_vend; i++) {
    if (vendor_status[i] == 2)
        vendor_status[i] = 1;
}

* Public agency issues coupons, monitors vendors, and sends a warning letter for those who are identified as high-risk.
package sim.app.fraudsim;
import java.util.*;
import java.lang.*;
import sim.util.*;
import sim.engine.*;
import ec.util.*;

// a class for sorting an array of a tuple defined as <int, double>
public class duoIntDouble implements Comparable
{
    int iInt;
    double dDouble;

    public duoIntDouble()
    {
        iInt = 0;
        dDouble = 0.0;
    }

    public duoIntDouble(int _iInt, double _dDouble)
    {
        this.iInt = _iInt;
        this.dDouble = _dDouble;
    }

    public int compareTo(Object obj1)
    {
        int result;
        if (dDouble < ((duoIntDouble) obj1).dDouble)
            result = 1;
        else if (dDouble > ((duoIntDouble) obj1).dDouble)
            result = -1;
        else
            result = 0;
        return result;
    }
}

* This class is used for desorting a vector that includes an integer and an double.
package sim.app.fraudsim;
import java.io.*;
import java.util.*;
import java.awt.*;
import sim.field.continuous.*;
import sim.util.*;
import sim.engine.*;
import ec.util.*;

public class FraudSim extends SimState
{
    static final long serialVersionUID = 3L;
    public double XMIN = 0;
    public double XMAX = 1000;
    public double YMIN = 0;
    public double YMAX = 800;

    public int Num_Part = 1000;
    public int Num_ActiveVend = 20;
    public int Num_Vend = 20;

    public double randomChoice = 0.01;
    public Continuous2D environment = null;

    public static final double DIAMETER = 8;
    public static boolean Is_Empirical = false;
    public static boolean Display = false;

    Participant[] parts;
    Vendor[] vends;
    PublicAgency pa;

    public boolean getIs_Empirical() { return Is_Empirical; }
    public void setIs_Empirical( boolean x ) { Is_Empirical = x; }
    public int getNum_Part() { return Num_Part; }
    public void setNum_Part(int x) { Num_Part = x; }
    public int getNum_ActiveVend() { return Num_ActiveVend; }
    public void setNum_ActiveVend(int x) { Num_ActiveVend = x; }
    public Object domRandomChoice() { return new Interval(0.0, 1.0); }
    public double getRandomChoice() { return randomChoice; }
    public void setRandomChoice( double x ) {
        if (x >= 0 && x <= 1) {
            randomChoice = x;
            for( int i = 0 ; i < Num_Part; i++ )
                if (parts[i]!=null)
                    parts[i].setRandomChoice ( randomChoice );  
        }
    }
    public boolean getDisplay() { return Display; }
    public void setDisplay ( boolean x ) {
        if (x == true | x == false) {
            Display = x;
            for( int i = 0 ; i < Num_ActiveVend; i++ )
                if (vends[i]!=null)
                    vends[i].setDisplay ( Display );  }
    }

    public double prp()
    {
        double rp;
        Random x = new Random();
        do {
            rp = x.nextGaussian() * 0.4 + 0.4;
        } while (rp < 0 || rp > 1);
        return rp;
    }

    public double vrp()
    {
        double rp;
        Random x = new Random();
        do {
            rp = x.nextGaussian() * 0.3 + 0.4;
        } while (rp < 0 || rp > 1);
    }
}
return rp;
}

public int random_lane()
{
    Random Rand = new Random();
    double percent = Rand.nextDouble();
    int lane;
    if (percent < 0.25)
        lane = 1;
    else if (percent < 0.46)
        lane = 2; //1-2 47%
    else if (percent < 0.61)
        lane = Rand.nextInt(3) + 3; //3-5 6%
    else if (percent < 0.73)
        lane = Rand.nextInt(4) + 6; //6-9  15%
    else if (percent < 0.94)
        lane = Rand.nextInt(7) + 10; //10-16 12%
    else
        lane = Rand.nextInt(4) + 17; //17-20 26%
    return lane;
}

public double hr()
{
    Random Rand = new Random();
    double percent = Rand.nextDouble();
    double _rp;
    do { _rp = Rand.nextDouble();
    if (percent < 0.58) {     // 0 - 0.01
        _rp = 0.01 * _rp;
    }
    else if (percent < 0.84) {   // 0.01 - 0.25
        _rp = (0.25-0.01)*_rp + 0.01;
    }
    else if (percent < 0.94) {   // 0.25 - 0.5
        _rp = (0.5-0.25)*_rp + 0.25;
    }
    else {                      // > 0.5
        _rp = 0.5*_rp + 0.5;
    }
    } while (_rp < 0.001 || _rp > 1);
    return _rp;
}

public double hr_conversion ( double t_hr )
{
    double hr = t_hr;
    double _rp;
    Random Rand = new Random();
    do { _rp = Rand.nextDouble();
    if (hr == 0) {      // 0 - 0.01
        _rp = 0.01 * _rp;
    }
    else if (hr == 1) {    // 0.01 - 0.25
        _rp = (0.25-0.01)*_rp + 0.01;
    }
    else if (hr == 2) {    // 0.25 - 0.5
        _rp = (0.5-0.25)*_rp + 0.25;
    }
    else {     // > 0.5
        _rp = 0.5*_rp + 0.5;
    }
    } while (_rp < 0 || _rp > 1);
    return _rp;
}

/** Creates a Fraud simulation with the given random number seed. */
public FraudSim(long seed)
{
    super (new MersenneTwisterFast(seed), new Schedule(2));
    System.out.println("Good FraudSim");
}
private void start_hypothetical()
{
    for(int x=0; x<Num_Part+Num_Vend; x++)
    {
        Double2D loc = null;
        Int2D pos = null;
        Agent agent = null;
        loc = new Double2D( random.nextDouble()*(XMAX-XMIN-
            DIAMETER)+XMIN+DIAMETER/2,
            random.nextDouble()*(YMAX-YMIN-
            DIAMETER)+YMIN+DIAMETER/2);
        if( x < Num_Part )
        {
            agent = new Participant( "Part"+x, loc, prp(), randomChoice,
                Num_ActiveVend);
            parts[x] = (Participant) agent;
        }
        else if ( x < Num_Part+Num_Vend )
        {
            agent = new Vendor ("Vend"+(x-Num_Part), loc, random_lane(),
                vrp());
            vends[x-Num_Part] = (Vendor) agent;
        }
        environment.setObjectLocation(agent,loc);
        schedule.scheduleRepeating(agent, 1, 1);
    }
}
private void start_empirical()
{
    int i;
    String s, s2 = new String();
    int commaidx = -1;
    int firstidx = 0;
    String str_wid = new String();
    String str_x = new String();
    String str_y = new String();
    String str_size = new String();
    String str_tiptype = new String();
    String str_contract = new String();
    String str_peer = new String();
    String str_hr = new String();
    String str_wid = new String();
    String str_race = new String();
    String str_category = new String();
    String str_edu = new String();
    String str_vendorVisited = new String();
    String directory = "sim/app/fraudsim/data";
    String fileName = "RandSample_Part.txt";
    try
    {
        File input = new File (directory, fileName);
        BufferedReader in = new BufferedReader (new FileReader (input));
        commaidx = -1;
        firstidx = 0;
        for(i=0; i<Num_Part; i++) // create participants
        {
            s = in.readLine();
            commaidx = s.indexOf("",firstidx);
            s2=s.substring(firstidx,commaidx);
            str_wid = s2;
            firstidx = commaidx + 1;
            commaidx = s.indexOf("",firstidx);
            s2=s.substring(firstidx,commaidx);
            str_x = s2;
            firstidx = commaidx + 1;
            commaidx = s.indexOf("",firstidx);
s2=s.substring(firstidx,commaidx);
str_y = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_race = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_category = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_edu = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_vendorVisited = s2;
firstidx = 0;
commaidx = -1;
Double2D loc = null;
Int2D pos = null;
Agent agent = null;
loc = new Double2D( Double.parseDouble(str_x),
Double.parseDouble(str_y) );
agent = new Participant( "Part"+Integer.toString(Integer.parseInt(str_wid) - 1),
loc, prp(), randomChoice, Num_ActiveVend);
parts[i] = (Participant) agent;
parts[i].race = Integer.parseInt(str_race);
parts[i].category = Integer.parseInt(str_category);
parts[i].edu = Integer.parseInt(str_edu);
parts[i].vendorVisited = Integer.parseInt(str_vendorVisited) - 1;
environment.setObjectLocation(agent,loc);
schedule.scheduleRepeating(agent, 1, 1);
}
in.close();
fileName = "RandSample_Vend.txt";
File input2 = new File (directory, fileName);
BufferedReader in2 = new BufferedReader (new FileReader (input2));
for (i=0; i<Num_ActiveVend; i++) // read in active vendors
{
    s = in2.readLine();
    commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_vid = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_x = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_y = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_size = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_tiptype = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_contract = s2;
firstidx = commaidx + 1;
commaidx = s.indexOf("\",firstidx);
s2=s.substring(firstidx,commaidx);
str_peer = s2;
firstIdx = commaIdx + 1;
s2 = s.substring(firstIdx, s.length());
str_hr = s2;
firstIdx = 0;
commaIdx = -1;

Double2D loc = null;
Int2D pos = null;
Agent agent = null;

loc = new Double2D(Double.parseDouble(str_x),
                  Double.parseDouble(str_y));
agent = new Vendor("Vend"+Integer.toString(Integer.parseInt(str_vid)-1),
                   loc, Integer.parseInt(str_size), hr());
vends[i] = (Vendor) agent;

vends[i].tiptype = Integer.parseInt(str_tiptype);
vends[i].contract = Integer.parseInt(str_contract);
vends[i].peer = Integer.parseInt(str_peer);
vends[i].hr = Integer.parseInt(str_hr);
vends[i].vnr = hr_conversion(vends[i].hr);

environment.setObjectLocation(agent, loc);
schedule.scheduleRepeating(agent, 1, 1);
}
in2.close();
}
catch (FileNotFoundException e) {
    System.err.println(e);
    return;
}
catch (IOException e) {
    System.err.println("Error reading input file " + e);
    return;
}
}

public void start() { // start
    super.start();
    environment = new Continuous2D(25.0, (XMAX-XMIN), (YMAX-YMIN));
    if (Is_Empirical) {
        Num_Part = 3500;
        Num_ActiveVend = 20;
    }

    PublicAgency pa = new PublicAgency(Num_Vend, Num_Vend-Num_ActiveVend);
schedule.scheduleRepeating(pa, 0, 1);

    parts = new Participant[Num_Part];
vends = new Vendor[Num_Vend];
    if (Is_Empirical)
        start_empirical();
    else
        start_hypothetical();
    for (int i=0; i<parts.length; i++)
        parts[i].set_public_agency(pa);
    pa.set_huff_model(this);
    pa.issueCoupons(parts);
}

public static void main(String[] args) { // main method
    if (args.length == 1)
        Is_Empirical = true;
doLoop(FraudSim.class, args);
    System.exit(0);
}