Agent-Based Modelling of Burglary

Nicolas Malleson
The University of Leeds
School of Geography

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

January 2010

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated overleaf. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.
Section 6.4 is based on the joint publication:


This work appears as published in Appendix C for reference. I declare that the research for this publication was solely my own work and that I am the lead author. The contribution of the other named authors, Alison Heppenstall and Linda See, were purely editorial and advisory.
Acknowledgements

It is thanks to the help and support of many people that I was able to complete this research, and to everyone I am very grateful. I would first like to thank my three supervisors: Andrew Evans, Alison Heppenstall and Linda See. Their wide-ranging expertise and constant availability for support and advice have been invaluable. I would also like to thank the experts at Safer Leeds and in particular Simon Hodgson and Fiona McLaughlin. They provided essential crime data (without which the project would not have been possible), useful contacts in the “industry” and expert knowledge on crime in Leeds which helped to ground the research to the real world.

As part of the research I was given the opportunity to visit Patricia and Paul Brantingham, the “grandparents” of environmental criminology, and their research group at ICURS in Simon Fraser University, Vancouver. I would like to thank everyone at SFU for the help they offered me and for their support and encouragement. I would also like to thank the ESRC for funding the overseas visit and for the project funding as a whole.

The School of Geography at Leeds is an extremely friendly and encouraging place to work and I would like to thank everyone in the school who helped me, particularly the other students and the people I have had the pleasure to share an office with.

I would also like to thank my dad and Francina for their encouragement over the course of my research career and for the Champagne when it was going well, and my mum for her helpful messages on my research blog and for pretending that she knew what I was doing for three years.

And finally my biggest thank you is to Megan for putting up with me for so long. You gave me the space to be miserable and grumpy when I needed to be and you were there when I needed you, I would not have made it through the project without your help.
Abstract

Understanding the processes and drivers behind crime is an important research area in criminology which has major implications for both improving policies and developing effective crime prevention strategies. Of all crime types, residential burglary is one of the most common and can be extremely traumatic for victims. The ability to more accurately represent, simulate and predict residential burglary will be invaluable to the design of crime reduction practice as well as for investigating criminological theory.

The central challenges of modelling a system as complicated as that of residential burglary lie in accurately representing the urban environment and simulating human behaviour. Crime modelling research that does not account for unique individuals who are located in a highly detailed (i.e. household- or street-level) environment will struggle to represent the processes and dynamics that drive the crime system and ultimately lead to city-wide crime rates. One technique that shows considerable promise for overcoming these limitations is agent-based modelling. An agent-based model is comprised of autonomous, decision making entities called agents that have the ability to interact with each other and with their environment. Through this mechanism, the methodology is able to account for the low-level interactions that drive a system directly, rather than having to predict system behaviour at an aggregate level.

This thesis will document the process of building an advanced agent-based residential burglary model. The model operates at the level of the individual; simulating individual potential burglars as they navigate a realistic virtual environment and decide, at the household level, where they will commit burglary. In this manner, the model is better suited to account for the dynamics and processes behind the burglary system than all other documented models. To demonstrate the advantages of this type of modelling approach, case studies will explore the effects of a large regeneration scheme in the city of Leeds and also how the public transport network drives burglary in Vancouver, Canada.
Publications

The following table presents the peer-reviewed papers, working papers and conference presentations that have emerged as part of the thesis and the chapter that they correspond to.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Chapter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peer-reviewed papers</strong></td>
<td></td>
</tr>
</tbody>
</table>

| **Working papers**                                                          |            |

Working papers are available from the School of Geography website at: http://www.geog.leeds.ac.uk/research/wpapers

<p>| <strong>Conference presentations</strong>                                                |            |
| Agent-Based Modelling of UK Crime. WUN Global GIS Academy Virtual Seminar, 11th November 2009 |            |
| Using Simulation to Predict Prospective Burglary Rates in Leeds and Vancouver Paper presented at the 7th National Crime Mapping Conference, Manchester, UK, 7th-8th May 2009 |            |</p>
<table>
<thead>
<tr>
<th>Publication</th>
<th>Chapter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving Burglary Reduction Initiatives through a Realistic Agent-Based</td>
<td></td>
</tr>
<tr>
<td>Model. Paper presented at the British Society of Criminology Conference,</td>
<td></td>
</tr>
<tr>
<td>University of Huddersfield, UK, 9-11 June 2008.</td>
<td></td>
</tr>
<tr>
<td>Realistic Crime Simulation: Predicting crime rates through a sound model of</td>
<td></td>
</tr>
<tr>
<td>offender behaviour coupled with an accurate virtual environment. Paper</td>
<td></td>
</tr>
<tr>
<td>presented at the NCeSS Fourth International Conference on e-Social Science,</td>
<td></td>
</tr>
<tr>
<td>Human Behaviour in an Agent-Based Model of Burglary. Paper presented at</td>
<td></td>
</tr>
<tr>
<td>GeoComputation 2007, 3-5 September, NUI Maynooth, Ireland.</td>
<td></td>
</tr>
<tr>
<td>Human Behaviour in an Agent-Based Model of Burglary. Paper presented at</td>
<td></td>
</tr>
<tr>
<td>Agent-Based Modelling and Burglary in Leeds. Paper presented at the 4th Lake</td>
<td></td>
</tr>
<tr>
<td>Arrowhead Conference on Human Complex Systems, 25-29 April 2007, Lake</td>
<td></td>
</tr>
<tr>
<td>Arrowhead, Los Angeles.</td>
<td></td>
</tr>
</tbody>
</table>
# Contents

1 Introduction 1
   1.1 Introduction to the Research .......................... 1
   1.2 Aims and Objectives ................................. 3
   1.3 Organisation of the Thesis ........................... 3

2 Understanding Burglary 7
   2.1 Introduction ........................................... 7
      2.1.1 Environmental Criminology ....................... 7
      2.1.2 The Geography of Crime .......................... 9
   2.2 Crime and the Environment ............................. 10
      2.2.1 Physical Factors ................................. 11
      2.2.2 The Social Environment ......................... 14
      2.2.3 Summary of Crime Attractors .................... 16
   2.3 Crime and the Offender ................................ 18
   2.4 Crime Theories ........................................ 19
      2.4.1 Routine Activity Theory ......................... 19
      2.4.2 Crime Pattern Theory ............................. 21
      2.4.3 The Rational Choice Perspective ................. 22
      2.4.4 The Journey to Crime ............................. 23
   2.5 Summary – Understanding Burglary ..................... 24

3 Modelling Burglary 27
   3.1 Introduction ........................................... 27
   3.2 The Difficulties with Modelling Social Systems ....... 28
      3.2.1 Chaos, Complexity and Complex Systems .......... 28
      3.2.2 Complex Social Systems ......................... 29
   3.3 Traditional Methods for Modelling Crime .............. 30
   3.4 Computer Modelling of Crime .......................... 33
      3.4.1 Computer Modelling and Simulation ............... 33
      3.4.2 Spatial Interaction Modelling and Spatial Microsimulation ........ 33
      3.4.3 Agent-Based Modelling ......................... 36
      3.4.4 A Critique of Agent-Based Modelling ............. 41
## Data Analysis: Crime and the Environment

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>49</td>
</tr>
<tr>
<td>4.2 The EASEL Regeneration Scheme</td>
<td>50</td>
</tr>
<tr>
<td>4.3 Geographical Data</td>
<td>51</td>
</tr>
<tr>
<td>4.3.1 UK Administrative Boundaries</td>
<td>51</td>
</tr>
<tr>
<td>4.3.2 The 2001 Census of Population</td>
<td>51</td>
</tr>
<tr>
<td>4.3.3 The Output Area Classification (OAC)</td>
<td>52</td>
</tr>
<tr>
<td>4.3.4 The Indices of Deprivation</td>
<td>52</td>
</tr>
<tr>
<td>4.3.5 Ordnance Survey MasterMap</td>
<td>53</td>
</tr>
<tr>
<td>4.3.6 Crime Data</td>
<td>55</td>
</tr>
<tr>
<td>4.4 The Victims of Crime</td>
<td>57</td>
</tr>
<tr>
<td>4.4.1 City-wide Burglary Patterns</td>
<td>60</td>
</tr>
<tr>
<td>4.4.2 Students and Burglary</td>
<td>63</td>
</tr>
<tr>
<td>4.4.3 Temporal Patterns</td>
<td>65</td>
</tr>
<tr>
<td>4.4.4 Crime and the Urban Environment</td>
<td>69</td>
</tr>
<tr>
<td>4.4.5 Crime and Demographic Variables</td>
<td>71</td>
</tr>
<tr>
<td>4.5 The Offenders</td>
<td>74</td>
</tr>
<tr>
<td>4.5.1 City-wide Offender Locations</td>
<td>74</td>
</tr>
<tr>
<td>4.5.2 Travel-to-crime</td>
<td>75</td>
</tr>
<tr>
<td>4.5.3 Offenders and Demographic Variables</td>
<td>77</td>
</tr>
<tr>
<td>4.6 Regression Models of Burglary</td>
<td>79</td>
</tr>
<tr>
<td>4.6.1 A Linear Regression Model</td>
<td>80</td>
</tr>
<tr>
<td>4.6.2 A Geographically Weighted Regression (GWR) Model</td>
<td>83</td>
</tr>
<tr>
<td>4.7 Data Analysis – Conclusion</td>
<td>83</td>
</tr>
</tbody>
</table>

## Creating Virtual People and their Virtual Environment

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>85</td>
</tr>
<tr>
<td>5.2 An Overview of the Virtual Environment</td>
<td>86</td>
</tr>
<tr>
<td>5.3 The Road Network</td>
<td>90</td>
</tr>
<tr>
<td>5.3.1 Road Accessibility Attributes</td>
<td>90</td>
</tr>
<tr>
<td>5.3.2 Estimating Road Traffic Volume: Space Syntax</td>
<td>91</td>
</tr>
<tr>
<td>5.4 Buildings</td>
<td>93</td>
</tr>
<tr>
<td>5.4.1 Type</td>
<td>93</td>
</tr>
<tr>
<td>5.4.2 Accessibility</td>
<td>96</td>
</tr>
<tr>
<td>5.4.3 Visibility</td>
<td>97</td>
</tr>
<tr>
<td>5.5 Communities</td>
<td>98</td>
</tr>
<tr>
<td>5.5.1 Collective Efficacy</td>
<td>100</td>
</tr>
<tr>
<td>5.5.2 The Sociotype and Community Similarity</td>
<td>102</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Attractiveness</td>
</tr>
<tr>
<td>5.5.4</td>
<td>Occupancy</td>
</tr>
<tr>
<td>5.6</td>
<td>An Overview of the Virtual Burglars</td>
</tr>
<tr>
<td>5.7</td>
<td>Agent Architectures / Cognitive Frameworks</td>
</tr>
<tr>
<td>5.7.1</td>
<td>Beliefs Desires Intentions (BDI)</td>
</tr>
<tr>
<td>5.7.2</td>
<td>Behaviour Based Artificial Intelligence (BBAI)</td>
</tr>
<tr>
<td>5.7.3</td>
<td>PECS</td>
</tr>
<tr>
<td>5.8</td>
<td>State Variables and Motives</td>
</tr>
<tr>
<td>5.8.1</td>
<td>Drugs</td>
</tr>
<tr>
<td>5.8.2</td>
<td>Sleep</td>
</tr>
<tr>
<td>5.8.3</td>
<td>Social</td>
</tr>
<tr>
<td>5.9</td>
<td>The Repertoire of Actions</td>
</tr>
<tr>
<td>5.10</td>
<td>The Process of Burglary</td>
</tr>
<tr>
<td>5.10.1</td>
<td>Decide where to start searching</td>
</tr>
<tr>
<td>5.10.2</td>
<td>Search for a victim</td>
</tr>
<tr>
<td>5.10.3</td>
<td>Choose a suitable victim</td>
</tr>
<tr>
<td>5.10.4</td>
<td>The Effects of Burglary on Surrounding Properties</td>
</tr>
<tr>
<td>5.11</td>
<td>Different Types of Burglar</td>
</tr>
<tr>
<td>5.12</td>
<td>Burglary Summary</td>
</tr>
<tr>
<td>5.13</td>
<td>Summary</td>
</tr>
<tr>
<td>6</td>
<td>Model Development and Testing</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>6.2</td>
<td>Tools for Building Agent-Based Models</td>
</tr>
<tr>
<td>6.3</td>
<td>Programming Overview</td>
</tr>
<tr>
<td>6.4</td>
<td>The Prototype Model</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Prototype Overview</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Proof-of-Concept Experiments</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Summary - The NetLogo Prototype</td>
</tr>
<tr>
<td>6.5</td>
<td>The Repast Simphony Model</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Contexts and Projections</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Burglar Agents</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Constructing the city: Environments, Layers and Projections</td>
</tr>
<tr>
<td>6.5.4</td>
<td>Moving agents around the Environment: Routing</td>
</tr>
<tr>
<td>6.5.5</td>
<td>Outputting Data</td>
</tr>
<tr>
<td>6.5.6</td>
<td>Utilising A High-Performance Grid</td>
</tr>
<tr>
<td>6.6</td>
<td>Verifying the Model – Experiments with Idealised Data</td>
</tr>
<tr>
<td>6.6.1</td>
<td>The Process of Verification</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Tests Using The Null Environment</td>
</tr>
<tr>
<td>6.6.3</td>
<td>Tests Using The Grid Environment</td>
</tr>
<tr>
<td>6.6.4</td>
<td>Tests Using The GIS Environment</td>
</tr>
</tbody>
</table>
# CONTENTS

6.7 Model Development – Summary .............................................. 168

7 Evaluating the Model – Experimenting with Real Data .............................................. 171

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>171</td>
</tr>
<tr>
<td>7.1.1 The Process of Calibration</td>
<td>172</td>
</tr>
<tr>
<td>7.1.2 The Process of Validation</td>
<td>172</td>
</tr>
<tr>
<td>7.2 Comparing Spatial Data</td>
<td>172</td>
</tr>
<tr>
<td>7.2.1 Comparing Point Patterns Visually</td>
<td>173</td>
</tr>
<tr>
<td>7.2.2 Comparing Point Patterns Mathematically</td>
<td>177</td>
</tr>
<tr>
<td>7.2.3 Goodness-of-Fit Statistics</td>
<td>181</td>
</tr>
<tr>
<td>7.2.4 Expanding cell validation method</td>
<td>182</td>
</tr>
<tr>
<td>7.2.5 Summary - methods used to validate models</td>
<td>184</td>
</tr>
<tr>
<td>7.3 Calibration</td>
<td>184</td>
</tr>
<tr>
<td>7.3.1 Calibration (&quot;Base&quot;) Scenarios</td>
<td>186</td>
</tr>
<tr>
<td>7.3.2 Base 1 – Default Conditions</td>
<td>187</td>
</tr>
<tr>
<td>7.3.3 Base 2 – Faster Security Decreases</td>
<td>193</td>
</tr>
<tr>
<td>7.3.4 Base 3 – Agents Dislike Security</td>
<td>197</td>
</tr>
<tr>
<td>7.3.5 Summary – Model Calibration</td>
<td>198</td>
</tr>
<tr>
<td>7.3.6 Comparing the Results to Regression Models</td>
<td>200</td>
</tr>
<tr>
<td>7.4 Validation</td>
<td>201</td>
</tr>
<tr>
<td>7.4.1 Preparing the Crime Data</td>
<td>202</td>
</tr>
<tr>
<td>7.4.2 Validation Results</td>
<td>202</td>
</tr>
<tr>
<td>7.5 Summary</td>
<td>204</td>
</tr>
</tbody>
</table>

8 EASEL Experiments and Forecasts .............................................. 207

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Introduction</td>
<td>207</td>
</tr>
<tr>
<td>8.2 The Optimistic Scenario – GS1</td>
<td>208</td>
</tr>
<tr>
<td>8.3 The Pessimistic Scenario – GS2</td>
<td>212</td>
</tr>
<tr>
<td>8.4 The “Insecure Buildings” Scenario – GS3</td>
<td>213</td>
</tr>
<tr>
<td>8.5 Crime Displacement and Burglar Travel Patterns</td>
<td>213</td>
</tr>
<tr>
<td>8.5.1 Crime displacement after the “optimistic” scenario</td>
<td>218</td>
</tr>
<tr>
<td>8.5.2 Displacement in the Low-Security Scenario</td>
<td>224</td>
</tr>
<tr>
<td>8.6 Summary – EASEL Scenarios</td>
<td>225</td>
</tr>
</tbody>
</table>

9 Experiments in Vancouver .............................................. 227

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1 Introduction</td>
<td>227</td>
</tr>
<tr>
<td>9.2 Context: Vancouver</td>
<td>228</td>
</tr>
<tr>
<td>9.3 Preparing the Environmental Data</td>
<td>230</td>
</tr>
<tr>
<td>9.3.1 Roads and Transport Routes</td>
<td>230</td>
</tr>
<tr>
<td>9.3.2 Buildings</td>
<td>233</td>
</tr>
<tr>
<td>9.3.3 Communities / Census Parameters</td>
<td>237</td>
</tr>
<tr>
<td>9.4 Preparing the Offender Agents</td>
<td>239</td>
</tr>
</tbody>
</table>
9.5 Real (Expected) Burglary Data .............................................. 242
9.6 Vancouver SkyTrain Scenarios ................................................ 243
  9.6.1 SkyTrain 1 – Normal Stations ......................................... 243
  9.6.2 SkyTrain 2 – No Millennium Line .................................... 245
  9.6.3 SkyTrain 3 – No Public Transport .................................... 245
9.7 Summary ................................................................. 246

10 Conclusion ............................................................... 249
  10.1 Introduction ............................................................... 249
  10.2 A Summary of the Research Findings ................................. 250
  10.3 A Critique of the Methodology ......................................... 253
  10.4 Recommendations for Future Work .................................... 255
  10.5 Concluding Remarks .................................................... 256

A The Output Area Classification (OAC) .................................... 257

B Demographic Variables and Crime Correlations ......................... 261

C Computer Environment and Urban Systems Paper ...................... 265

D Regional Review Paper .................................................... 281
# List of Figures

2.1 The “crime triangle”, as proposed by Clarke and Eck (2005) .......................... 20
2.2 Notes, paths, edges and activity/opportunity spaces ................................. 22
2.3 Diagram of crime theories ................................................................. 23

3.1 Situating simulation among other approaches to investigating systems ........ 34
3.2 An agent-based model of segregation .................................................. 39

4.1 Photographs illustrating the types of houses in the EASEL area ..................... 50
4.2 The Index of Multiple Deprivation in Leeds ......................................... 53
4.3 An example of the OS MasterMap Topography layer ................................. 55
4.4 Yearly densities of burglaries from recorded crime data ............................ 59
4.5 Aggregated burglary counts at the OA and SOA level and point densities ...... 60
4.6 Contour percentages ............................................................................. 61
4.7 Burglary hotspots near the city centre .................................................... 62
4.8 Hotspot contours compared to the IMD and OAC .................................... 63
4.9 Burglary rates in and out of term time ................................................... 65
4.10 Example time clock ............................................................................. 66
4.11 Burglary times in Leeds and the Hyde Park hotspot ................................. 67
4.12 Monthly burglary counts in Leeds in 2001 ............................................. 67
4.13 Time clocks for burglaries by month ..................................................... 68
4.14 Accommodation types in Leeds ........................................................... 69
4.15 Scatter plots of household type and the number of burglaries ..................... 70
4.16 Histograms of burglary counts at OA and SOA levels ............................. 72
4.17 OAC variables mapped at the SOA level .............................................. 73
4.18 Density map of offender home locations .............................................. 75
4.19 Hotspots of nominal locations compared to the burglary hotspots ............ 76
4.20 The commuter and marauder hypotheses .............................................. 77
4.21 Examples of offender journey to crime ................................................. 78
4.22 Scatterplot of IMD score and offender counts at the SOA level ............... 80
4.23 The spatial variation of GWR coefficients ............................................ 84

5.1 The layers that make up the virtual environment ...................................... 87
5.2 Parameters used in the household and community environment layers . . . . . . 89
5.3 Road accessibility attributes. ............................................ 92
5.4 Space syntax integration values for the entire city and a local area ............. 93
5.5 Weights which are used to estimate traffic volume at different times. .......... 94
5.6 Histograms of the square area of different MasterMap objects. ................. 95
5.7 Histograms of garden sizes before and after normalisation. ..................... 98
5.8 Degree of isolation histograms before and after normalisation ................. 99
5.9 Number of neighbours, size of garden and degree of isolation ................. 99
5.10 The measure of collective efficacy and its constituents ....................... 103
5.11 The similarity of a selection of output areas ................................ 104
5.12 Functions to calculate occupancy for different variables depending on time. . 106
5.13 Motives and motive selection, adapted from Schmidt and Schneider (2004) .... 109
5.14 How motive intensity varies with state variable value ......................... 113
5.15 Determining the action-guiding motive .................................. 114
5.16 How sleep and social motives vary with time. ................................ 115
5.17 Actions to satisfy goals .................................................. 116
5.18 Function used to determine household suitability to a burglar. ................ 120
5.19 Example of security increases in response to a burglary ....................... 121
5.20 The burglars’ decision process for burglary ................................ 124
5.21 Offender burglary space/time visualisation .................................. 126

6.1 How agent’s needs can lead to actions. Taken from Malleson et al. (2010b). . . 132
6.2 The environment used in the NetLogo prototype ............................. 133
6.3 Determining building suitability in the prototype model ......................... 135
6.4 The burglary locations in the prototype control experiment .................... 136
6.5 How wealth and sleep can vary over simulated time for a typical burglar and citizen. 136
6.6 The spatial location of different types of community ........................... 138
6.7 Results of the target hardening experiment using the prototype model .......... 139
6.8 Example organisation of Repast Simphony contexts and projections .......... 140
6.9 The organisation of Burglar objects ..................................... 142
6.10 The different environment layers which make up the virtual city ............. 143
6.11 Example GIS route between A and C ................................... 144
6.12 The database schema used to store a model history ........................... 146
6.13 Distributing models on the NGS ....................................... 147
6.14 The layout of the environment to test PECS variables. ......................... 152
6.15 PECS state variable and motive values in the null environment ................. 155
6.16 Null environment house/community configurations. .......................... 157
6.17 The layout of the environment used in the Grid sensitivity tests ............... 161
6.18 The distribution of burglaries with the default model configuration ........... 161
6.19 The evolution of a burglar’s awareness space in a single example model run... 162
6.20 Results when altering the Distance_W agent parameter ...................... 163
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.21</td>
<td>The layout of the environment used in GIS sensitivity tests.</td>
</tr>
<tr>
<td>6.22</td>
<td>The GIS environment verification test results.</td>
</tr>
<tr>
<td>6.23</td>
<td>A public transport network test in the GIS environment.</td>
</tr>
<tr>
<td>7.1</td>
<td>Three point data sets which will be used to experiment with spatial methods.</td>
</tr>
<tr>
<td>7.2</td>
<td>Counts of points aggregated up to (Super) Output Area administrative boundaries.</td>
</tr>
<tr>
<td>7.3</td>
<td>Changing the kernel and grid cell size used in the KDE algorithm.</td>
</tr>
<tr>
<td>7.4</td>
<td>Comparing data using the Fuzzy Kappa statistic.</td>
</tr>
<tr>
<td>7.5</td>
<td>SRMSE and $R^2$ errors comparing model data and random data to expected data.</td>
</tr>
<tr>
<td>7.6</td>
<td>Relative percentage error comparing model data to randomly generated data.</td>
</tr>
<tr>
<td>7.7</td>
<td>The locations of different building types.</td>
</tr>
<tr>
<td>7.8</td>
<td>L Functions for a number of Base 1 models compared to expected data.</td>
</tr>
<tr>
<td>7.9</td>
<td>Graph of the difference in Base 1 L functions at different time points.</td>
</tr>
<tr>
<td>7.10</td>
<td>Map comparing Base 1 model results hotspots to expected data.</td>
</tr>
<tr>
<td>7.11</td>
<td>Expanding cell maps for a Base 1 Scenario.</td>
</tr>
<tr>
<td>7.12</td>
<td>A histogram of all expanding cell Base 1 errors.</td>
</tr>
<tr>
<td>7.13</td>
<td>Real crime hotspots occurring in the Halton Moor area.</td>
</tr>
<tr>
<td>7.14</td>
<td>Journey-to-crime in the Halton Moor area.</td>
</tr>
<tr>
<td>7.15</td>
<td>The Output Area Classification (OAC) groups in the EASEL area.</td>
</tr>
<tr>
<td>7.16</td>
<td>Frequency distribution of building security at different time points.</td>
</tr>
<tr>
<td>7.17</td>
<td>Comparing Base2 results to Base 1 and expected data.</td>
</tr>
<tr>
<td>7.18</td>
<td>L Functions for a number of Base 3 models compared to Base 2 and expected data.</td>
</tr>
<tr>
<td>7.19</td>
<td>Comparing Base 3 Hotspots to expected data.</td>
</tr>
<tr>
<td>7.20</td>
<td>Expanding cell maps for a Base 3 Scenario.</td>
</tr>
<tr>
<td>7.21</td>
<td>Comparing all Base results at different resolutions.</td>
</tr>
<tr>
<td>7.22</td>
<td>EASEL crime hotspots in the 2004 data.</td>
</tr>
<tr>
<td>7.23</td>
<td>Maps comparing validation result hotspots to 2004 expected data.</td>
</tr>
<tr>
<td>7.24</td>
<td>Comparing the fitness of the validation results (using 2004 crime field data) to base results (using 2001 field data) at different resolutions.</td>
</tr>
<tr>
<td>8.1</td>
<td>Output Area Classification groups after Gipton/Seacroft changes.</td>
</tr>
<tr>
<td>8.2</td>
<td>Changes to the Gipton/Seacroft areas.</td>
</tr>
<tr>
<td>8.3</td>
<td>Collective efficacy levels before and after Gipton/Seacroft changes.</td>
</tr>
<tr>
<td>8.4</td>
<td>Comparing Gipton/Seacroft and Base 3 results.</td>
</tr>
<tr>
<td>8.5</td>
<td>Comparing results from the “optimistic” and “pessimistic” scenarios.</td>
</tr>
<tr>
<td>8.6</td>
<td>Comparing the “insecure buildings” scenario to the previous Gipton/Seacroft scenarios.</td>
</tr>
<tr>
<td>8.7</td>
<td>Comparing the “bad buildings” scenario to the previous Gipton/Seacroft scenarios.</td>
</tr>
<tr>
<td>8.8</td>
<td>Comparing the “bad buildings” scenario to the previous Gipton/Seacroft scenarios.</td>
</tr>
<tr>
<td>8.9</td>
<td>Displacement into buffer regions surrounding Gipton and Seacroft.</td>
</tr>
<tr>
<td>8.10</td>
<td>Comparing Base3 and GS1 results using the expanding cell algorithm.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>8.11</td>
<td>Offender movement patterns in the GS1 scenario</td>
</tr>
<tr>
<td>8.12</td>
<td>Visualising the journey to and from a burglary in Seacroft</td>
</tr>
<tr>
<td>8.13</td>
<td>Burglary risk for buildings in Seacroft</td>
</tr>
<tr>
<td>8.14</td>
<td>Comparing GS1 and GS3 results using the expanding cell algorithm</td>
</tr>
<tr>
<td>9.1</td>
<td>The Greater Vancouver Regional District</td>
</tr>
<tr>
<td>9.2</td>
<td>Vancouver viewed northwards as displayed using Google Earth</td>
</tr>
<tr>
<td>9.3</td>
<td>Vancouver transport routes used in the simulation</td>
</tr>
<tr>
<td>9.4</td>
<td>Roads types and estimated traffic volume in Vancouver</td>
</tr>
<tr>
<td>9.5</td>
<td>Buildings used in the Vancouver simulations</td>
</tr>
<tr>
<td>9.6</td>
<td>Vancouver housing density</td>
</tr>
<tr>
<td>9.7</td>
<td>The VANDIX measure of deprivation in Vancouver</td>
</tr>
<tr>
<td>9.8</td>
<td>Estimated levels of collective efficacy in Vancouver</td>
</tr>
<tr>
<td>9.9</td>
<td>Levels of attractiveness and occupancy in Vancouver</td>
</tr>
<tr>
<td>9.10</td>
<td>An example of the measure of similarity of different communities in Vancouver</td>
</tr>
<tr>
<td>9.11</td>
<td>The number of offenders created in dissemination areas in Vancouver</td>
</tr>
<tr>
<td>9.12</td>
<td>Expected burglary rates in Vancouver, 2001</td>
</tr>
<tr>
<td>9.13</td>
<td>Vancouver results with and without Leeds calibration</td>
</tr>
<tr>
<td>9.14</td>
<td>Comparing burglary patterns with and without the Millennium Line SkyTrain</td>
</tr>
<tr>
<td>9.15</td>
<td>Results of the SkyTrain 3 scenario: no public transport routes</td>
</tr>
<tr>
<td>A.1</td>
<td>The Output Area Classification supergroups in Leeds</td>
</tr>
</tbody>
</table>
# List of Tables

1.1 The organisation of the thesis with respect to the research objectives. .......... 4

2.1 Environmental factors found to attract or deter crime. ...................... 17

3.1 Drawbacks with spatial interaction modelling. ............................... 35

3.2 Drawbacks with microsimulation. ............................................. 36

4.1 Summary of common administrative boundaries in increasing size. .......... 51

4.2 Relevant National Land Use Group Codes (Harrison, 2006) .................. 54

4.3 The crime data attributes. ..................................................... 57

4.4 The number of crimes and burglaries in the victim data ...................... 58

4.5 Comparing census data in OAs in Hyde Park, Harehills and Leeds. ........ 64

4.6 Proportions of burglaries by accommodation type. .......................... 70

4.7 Correlation between census variables and burglary. .......................... 72

4.8 Details of the offenders whose offending behaviour is illustrated in Figure 4.21 77

4.9 Correlation between census variables and offender counts. .................... 79

4.10 Results of the burglary regression model at SOA level. ...................... 82

5.1 Road types specified in the ITN layer and their mapping to model road types. 91

5.2 Variables used for the Index of Heterogeneity, after Shepherd (2006). ....... 101

5.3 OAC variables that represent daily habits. ....................................... 105

5.4 Different types of PECS behaviour. ............................................ 109

5.5 Goals which can be accomplished to satisfy motives. .......................... 116

5.6 Variables that determine a burglar’s assessment of household burglary suitability. 120

5.7 A classification system for burglar types. ...................................... 123

5.8 How the motives of potential burglars and their responses to environmental cues will be implemented in the model. ................................. 124

6.1 The change from the default value for attractiveness and security variables associated with different community types. .............................. 138

6.2 All model parameters and the environment which will be used to test the sensitivity of the model to their values. .............................. 149

6.3 Default values and range for PECS variables to be tested. .................... 154
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>Summary results for tests of the WealthGain parameter. As the parameter increases, the agent can spend less time working and more time doing nothing.</td>
<td>156</td>
</tr>
<tr>
<td>6.5</td>
<td>The proportion of burglaries committed in houses/communities with given parameter values</td>
<td>158</td>
</tr>
<tr>
<td>6.6</td>
<td>The number of burglaries committed by different agents in different houses</td>
<td>159</td>
</tr>
<tr>
<td>6.7</td>
<td>The average distances that crime were committed away from the agent’s home and all anchor points</td>
<td>167</td>
</tr>
<tr>
<td>7.1</td>
<td>A summary of spatial statistics that can be used to describe and compare the spatial structure of point patterns</td>
<td>177</td>
</tr>
<tr>
<td>7.3</td>
<td>Goodness-of-Fit at different resolutions.</td>
<td>182</td>
</tr>
<tr>
<td>7.4</td>
<td>Base 1 model errors at different cellular resolutions</td>
<td>189</td>
</tr>
<tr>
<td>7.5</td>
<td>Model errors at different cellular resolutions. The errors that represent the greatest goodness-of-fit are bold.</td>
<td>198</td>
</tr>
<tr>
<td>8.1</td>
<td>Changes to household parameters for the Gipton/Seacroft scenario.</td>
<td>211</td>
</tr>
<tr>
<td>9.1</td>
<td>Different classes of road available in the GIS Innovations product.</td>
<td>231</td>
</tr>
<tr>
<td>9.2</td>
<td>Relevant BCAA land-use codes used to derive property types.</td>
<td>233</td>
</tr>
<tr>
<td>9.3</td>
<td>BCAA land use codes to represent social locations.</td>
<td>234</td>
</tr>
<tr>
<td>9.4</td>
<td>The mapping of Canadian and British ethnicities to groups used in the model.</td>
<td>239</td>
</tr>
<tr>
<td>A.1</td>
<td>Variables used to construct the Output Area Classification.</td>
<td>257</td>
</tr>
<tr>
<td>A.2</td>
<td>Output area classification group and super group descriptions.</td>
<td>260</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Contents

1.1 Introduction to the Research ............................................. 1
1.2 Aims and Objectives ..................................................... 3
1.3 Organisation of the Thesis .............................................. 3

1.1 Introduction to the Research

Understanding the processes and drivers behind crime is an important research area in criminology which has major implications for both improving policies and developing effective crime prevention strategies (Brantingham and Brantingham, 2004; Groff, 2007a). Recent advances in environmental criminology, such as routine activities theory (Cohen and Felson, 1979), the rational choice perspective (Clarke, 1995) and crime pattern theory (Brantingham and Brantingham, 1993) have highlighted a shift from the study of the motivation of offenders to understanding the social and environmental contexts in which crimes occur. However, in order to test these “opportunity theories” and to predict crime accurately, it is essential to be able to account for the complex, dynamic interactions of the individuals involved in each crime event as well as their interactions with others and with their environment.

Of all crime types, residential burglary is one of the most common and can be extremely traumatic for victims because it involves an invasion of otherwise safe personal space (the home). In Leeds, UK, residential burglary rates have been consistently the highest when compared to any other local authority in England and Wales (Shepherd et al., 2004). To combat this, the public body responsible for implementing and evaluating crime reduction strategies, Safer Leeds, are involved in developing strategies to reduce residential burglary. Although they are usually highly successful, burglary reduction schemes are often reactive, due to difficulties in predicting emerging burglary hot spots. Therefore the ability to more accurately represent, simulate and predict burglary will be invaluable to the design of crime reduction practice as well as for investigating criminological theory.

The central challenges of modelling a system as complicated as that of residential burglary
lie in accurately representing the urban environment and simulating human behaviour. With respect to the environment, modern criminology research pinpoints the importance of studying the “micro-places” that frame an individual crime event (Eck and Weisburd, 1995). Therefore studies that work at levels above the individual house or street obscure important information about the real dynamics of crime (Andresen and Malleson, 2010). With respect to human behaviour, criminology also teaches us that an individual crime event depends largely on the unique circumstances of the individual(s) involved. Therefore, crime modelling research that does not account for unique individuals who are located in a highly detailed (i.e. household- or street-level) environment will struggle to represent the processes and dynamics that drive the crime system and ultimately lead to city-wide crime rates. However, the majority of crime modelling research to date works with relatively large aggregate areas and therefore struggles to accurately represent the underlying dynamics of the crime system. Brantingham and Brantingham (1993) described that, in their opinion:

“The most productive model in environmental criminology is one that places both the actual criminal events at a specific site, situation and time and the individual committing the crime while in a specific motivational state on (or in) an environmental backdrop, that may itself be mostly stable, regular and predictable or may instead be irregular, rapidly changing and unpredictable” (Brantingham and Brantingham, 1993).

One technique that shows considerable promise for overcoming these limitations is agent-based modelling (ABM). ABM represents a shift in the social sciences away from aggregate models towards those that work at the level of the individual. Agent-based models are comprised of autonomous, decision making entities called agents that have the ability to interact with each other and their environment (Bonabeau, 2002). Agents can represent individuals, groups of individuals or, if appropriate, objects such as houses, businesses or organisms. As the model iterates, each agent has the ability to assess its circumstances and, based on a set of probabilistic rules, make an informed/educated decision about its future course of action (Bonabeau, 2002). Through this mechanism, more realistic human behaviour can be incorporated into models (Moss and Edmonds, 2005).

The development and application of an agent-based model for simulating residential burglary provides a unique opportunity to both further our understanding of the processes and dynamics of this system as well as providing a platform for testing out crime reduction policies. Indeed, one of the most attractive elements of ABM is the ability to experiment with different crime theories and crime reduction policies before implementation in the real system. This thesis will document the process of building an advanced agent-based residential burglary model. The model operates at the level of the individual; simulating individual potential burglars as they navigate a realistic virtual environment and determine, at the household level, where they will commit their individual burglaries. In this manner, the model is better suited to account for the dynamics and processes behind the burglary system than all other documented models. To demonstrate the advantages of this type of modelling approach, case studies will explore the effects of a large regeneration scheme
in the city of Leeds and also how the public transport network drives burglary in Vancouver, Canada.

1.2 Aims and Objectives

The overall aim of this research is to explore the use of agent-based modelling in the context of the residential burglary system\(^1\). To achieve this aim, the following objectives have been formalised:

1. Review and discuss the crime literature and available data to establish which factors drive the residential burglary system and therefore need to be included in a model.

2. Review and discuss the modelling techniques that have been used to model crime in order to highlight potential areas for research, inform the methodology used for this research and to guide the model development process.

3. Design and build an agent-based model that is able to account for the dynamics of the residential burglary system and create accurate predictions of crime rates.

4. Evaluate the model by assessing its response to varying parameter values and by comparing the results to known field conditions in order to gauge its predictive accuracy.

5. Apply the model to real-world scenarios in order to predict the potential impacts on residential burglary rates.

1.3 Organisation of the Thesis

Table 1.1 summarises the structure of the thesis, relating the research objectives to relevant chapters.

Chapter 2 introduces the relevant crime literature to find the most important factors that need to be included in a burglary model. It begins with an introduction to environmental criminology and the history behind studying the geographical element of crime. The chapter then proceeds to critique research that discusses the environmental factors that relate to burglary. To follow, research that discusses how offenders behave is reviewed. To conclude, theories in environmental criminology that have been developed to explain the empirical findings are discussed as these will be implemented in the model.

Chapter 3 extends the literature review with an examination of the types of modelling techniques that can be used to simulate crime as well as some of their associated difficulties. It is noted that most crime studies use mathematical techniques such as regression and these fail to account for the low-level interactions and processes that drive the crime system. Agent-based modelling is presented as an improvement over traditional techniques and the few published agent-based crime studies are critiqued.

\(^1\)The residential burglary system refers to the interactions, behaviours and environmental characteristics that lead to a crime event taking place and, from which, city-wide crime rates ultimately emerge.
Table 1.1: The organisation of the thesis with respect to the research objectives.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Review and discuss the crime literature and available data to establish which factors drive the residential burglary system and therefore need to be included in a model.</td>
<td>2: Understanding Burglary</td>
</tr>
<tr>
<td>2: Review and discuss the modelling techniques that have been used to model crime in order to highlight potential areas for research, inform the methodology used for this research and to guide the model development process.</td>
<td>4: Data Analysis: Crime and the Environment</td>
</tr>
<tr>
<td>3: Design and build an agent-based model that is able to account for the dynamics of the residential burglary system and create accurate predictions of crime rates.</td>
<td>3: Modelling Burglary</td>
</tr>
<tr>
<td>4: Evaluate the model by assessing its response to varying parameter values and by comparing the results to known field conditions in order to gauge its predictive accuracy.</td>
<td>5: Creating Virtual People and their Virtual Environment</td>
</tr>
<tr>
<td>5: Apply the model to real-world scenarios in order to predict the potential impacts on residential burglary rates.</td>
<td>6: Model Development and Testing</td>
</tr>
<tr>
<td></td>
<td>6: Model Development and Testing</td>
</tr>
<tr>
<td></td>
<td>7: Evaluating the Model – Experimenting with Real Data</td>
</tr>
<tr>
<td></td>
<td>8: EASEL Experiments and Forecasts</td>
</tr>
<tr>
<td></td>
<td>9: Experiments in Vancouver</td>
</tr>
</tbody>
</table>

Chapter 4 outlines the data sources that provide essential data required to build the model. An overview of the crime system in Leeds is presented to aid understanding of the dynamics of the system that the simulation must model. Following this, the chapter analyses the data to determine the extent to which it supports the findings from the literature and also to establish if there are any other important variables that should be included in a model. To conclude, simple regression models are utilised in order to establish their predictive accuracy and set a benchmark on which to compare the accuracy of the final agent-based model produced by this research.

To tie the findings from the literature to later model development chapters, Chapter 5 discusses how the model can be constructed from the available data. It describes how the individual elements of the virtual environment – roads, buildings and communities – can be formalised and how the burglar agents have been designed. It includes a description of the use of a cognitive framework that controls the agents to simulate realistic human behaviour and also how the burglar construction relates to the act of burglary.

Building upon the theoretical model design, Chapter 6 outlines the process of formally building the model. It outlines the available tools, provides details of a model prototype and then
supplies the formal model design. The chapter concludes by experimenting with the model using idealised data; rigorously testing it to ensure that it behaves as it is expected to before applying it to a real-world urban environment.

Chapter 7 furthers the evaluation of the model that began in Chapter 6 through experiments with real data. Two procedures are documented: *calibration* refers to the process of adjusting parameter values in order to generate results that match field conditions and *validation* applies the model to a system other than the one it was calibrated on. This allows the researcher to establish whether or not the model is transferable to other systems and, therefore, how reliable its predictions are likely to be.

Having fully evaluated the performance of the model, Chapter 8 documents the use of the model to create crime forecasts in the city of Leeds, UK. The case study is a large urban development scheme termed EASEL and the model is used to predict how the physical urban changes are likely to impact local burglary rates. The model results reveal interesting patterns of crime displacement into the local area and suggest whether or not individual household variables are more important to burglars than community-wide factors.

Following the Leeds experiments, Chapter 9 presents the second forecasting scenario which demonstrates the flexibility of the model by applying it to a significantly different urban environment; that of Vancouver, Canada. The model is used to predict what effect the high-speed rail system has on crime rates.

The final chapter, Chapter 10, draws concluding remarks about the project as a whole. It documents the successes and the drawbacks of the research and suggests how the drawbacks could be mediated. Finally it suggests fruitful avenues for future work.
Chapter 2

Understanding Burglary

Contents

2.1 Introduction .......................................................... 7
2.2 Crime and the Environment ........................................ 10
2.3 Crime and the Offender ............................................. 18
2.4 Crime Theories ...................................................... 19
2.5 Summary – Understanding Burglary ............................ 24

2.1 Introduction

A key aim of this research is to develop an agent-based model that can be used to explore the dynamics of residential burglary. To build an accurate model, however, a thorough understanding of the burglary system is required. This chapter will explore the different aspects of environmental criminology, drawing upon research that will identify the key features of the burglary system that should be included in a crime model. To begin, Sections 2.1.1 and 2.1.2 will provide an introduction to environmental criminology and discuss the evolution of geography of crime research. To determine the important factors that should be included in the model’s virtual environment, Section 2.2 will follow with an outline of research that investigates how crime is influenced by the environment. Then Section 2.3 will discuss research that explores how offenders behave in order to gain a thorough understanding of the components that are required to drive the behaviour of the virtual burglars. To complete the review, Section 2.4 will demonstrate how the concepts discussed have shaped environmental criminology theory. These theories can then be directly implemented in the model.

2.1.1 Environmental Criminology

An occurrence of crime is the result of a culmination of a very large number of factors that include the motivations and behaviour of the criminal, the influence of the physical surroundings and the behaviour of the victim and others. This makes the crime system extremely complex and therefore
difficult to model. However, occurrences of crime are not random and a large body of literature has evolved to explain some of the patterns exhibited by criminal occurrences. This section will briefly outline the evolution of the field of environmental criminology, setting the context for the remainder of the chapter. Criminology is an interdisciplinary research field that studies the nature and extent of crime in a society and will be drawn on to build the burglary model. Andresen (2010) notes that pre-1970 criminology research has largely been dominated by research into victims (what makes some people more susceptible to crime than others), the law (how laws affect crime) and offenders (what makes some people commit crime). Although these research directions are essential for understanding the drivers of crime, they omit a vital element: the place in which the crime occurs. It was to this end that the field of “environmental criminology” arose in the 1970s as a discipline to study the spatial variations of crime and the underlying reasons for these variations (Johnson et al., 2002).

That is not to say that the link between crime and the environment had not been made before 1970. In the thirteenth century, for example, Edward I tried to reduce highway robbery by clearing highway verges in order to remove hiding places for robbers (Brantingham and Brantingham, 1993). Similarly, long before the term “environmental criminology” was coined, the effects of the urban environment on crime were commented on by sociologist Enrico Ferri in 1899:

“High-roads, railways, and tramways disperse predatory bands in rural districts, just as wide streets and large and airy dwellings, with public lighting and the destruction of slums, prevent robbery with violence, concealment of stolen goods, and indecent assaults” (Ferri, 1899).

Nevertheless, it was not until the 1940s that a theory linking crime to its environment was developed. In their seminal work at the University of Chicago, Shaw and McKay used Burgess’s concentric zone model (Burgess, 1925) and concepts of social disorganisation to explain city-wide juvenile delinquency (Shaw and McKay, 1942). Their Social Disorganisation Theory stipulated that crime was a normal response by individuals to an environment which lacked the possibility of policing itself (or being adequately policed from outside) due to the transient nature of the residents and the subsequent lack of community cohesion. The importance of Shaw and McKay’s research cannot be overstated as it forms the base on which many later environment criminology theories stand.

“The term “environmental criminology” itself was coined by Jeffery (1971) when he called for a new avenue of research in criminology to focus on the environment in which criminality can occur. Although Jeffery’s ideas were largely ignored by policy makers, they laid the path from Shaw and McKay’s Social Disorganisation Theory to the first published theory in environmental criminology (Andresen, 2010) called Routine Activity Theory (Cohen and Felson, 1979). This, along with other major theories such as the Geometric Theory of Crime (Brantingham and Brantingham, 1981a), the Rational Choice Perspective (Clarke and Cornish, 1985) and the Pattern Theory of Crime (Brantingham and Brantingham, 1993) form the groundwork upon which most environmental criminology research is based. Although published separately, the theories are largely in agreement about the causes and drivers of crime; they represent different parts of the larger system.
This agreement is made evident by Ekblom (2001) who developed the “Conjunction of Criminal Opportunity” framework which creates a union of the major environmental criminology theories.

The theories themselves are discussed in more detail in Section 2.4 and their integration into the model is addressed explicitly in Section 5.12. For a fuller account of the history and origins of environmental criminology, the interested reader can start with Wortley and Mazerolle (2008).

2.1.2 The Geography of Crime

Alongside the theoretical work, empirical research in environment criminology has greatly increased our understanding of how crime varies spatially. Since the pioneering work of Quetelet (1831) and Mayhew (1861), research has shown that the location of crimes can explain much about their underlying dynamics (Messner et al., 1999). In particular, criminologists are becoming increasingly aware that crime must be analysed at very large spatial scales; otherwise important patterns that are present at finer units of analysis can be missed (a point that will be discussed more fully below).

The study of the geography of crime first originated in France through the beginnings of the Cartographic School of Criminology (Kongmuang, 2006). Early studies that explored the spatial distribution of crime commonly linked crime or offender rates to socio-demographic data at regional scales; see Quetelet (1831) for example. In England, Glyde (1856) provides an excellent example of this type of research. The author examined data on the home addresses of convicted offenders in the county of Suffolk and found that there were very large differences in offender rates even when towns had similar populations and were relatively close together. For example, Southwold and Halesworth had populations of 2,109 and 2,509 respectively, were only a few miles apart and had offender:population ratios of 1:309 and 1:1,320 respectively. This research was perhaps the first to make explicit the notion that crime varies considerably depending on the spatial unit and that using geography at too small a scale can be extremely misleading.

Since the work of Glyde, geography of crime research has moved to even smaller units of analysis. In their seminal work on juvenile delinquency, Shaw and McKay (1942) used the census tract (an American administrative boundary of approximately one square mile). This is roughly the unit of analysis that most modern crime research has continued to use (Weisburd et al., 2009), with the exception of some more recent studies that work at smaller census area boundaries of approximately 100–200 households (called the “Output Area” or “Dissemination Area”) such as Herbert and Hyde (1985). Taylor (1997) and Weisburd et al. (2004, p 290) argue that the street segment is the best unit of analysis, but part of the argument is based on the premise of data quality, so if accurate data are available at an even lower level of aggregation then this might be preferable. Indeed, modern criminology theories, such as Routine Activities Theory (Cohen and Felson, 1979), suggest that individual-level data are required. Cohen and Felson theorised that individual crime occurrences are dependent on individual units being present (such as a victim and an offender) and therefore crime should be analysed in the context of these micro-units that cause it to occur (Weisburd et al., 2004).

It has been found that burglars choose individual homes based on their individual charac-
CHAPTER 2. UNDERSTANDING BURGLARY

characteristics (Rengert and Wasilchick, 1985) so it cannot normally be assumed that a community or neighbourhood is homogeneous with respect to burglary risk. Eck (1995) began the trend of studying the “micro places” that crime occurs in, which is a theme that is increasingly evident in the literature. For example, when analysing the success of a burglary reduction scheme in Liverpool, Bowers et al. (2003) showed that aggregated data can be extremely misleading. Similarly, more recent research shows that street-level crime varies considerably so that aggregation to even the smallest administrative boundaries would hide important information (Weisburd et al., 2004; Groff et al., 2009; Andresen and Malleson, 2010).

The main problem with administratively-defined boundaries is that they are not built to be homogeneous. For example, output areas in the UK are designed to hold a particular number of households or people, regardless of how similar those people are. Therefore it is difficult to use such data in statistics because a researcher cannot justify that aggregate relationships hold at the individual level unless the area is known to be homogeneous (Rengert and Lockwood, 2009). Although using smaller areas reduces the reliability of statistics because there are smaller numbers of units in each area, Oberwittler and Wikström (2009) show that this is preferable to using larger areas that are less homogeneous.

However, the problems with administrative boundaries to study crime are not easily overcome as much data used in crime research (such as censuses) are released using these boundaries. Rengert and Lockwood (2009) suggest a method of creating separate regions from a combination of administrative boundaries and buffers around points of interest (such as bars) but, as the authors note, this has the effect of skewing the data towards zero which breaks the assumptions of normality required by many commonly used statistical models (such as ordinary least squares regression). As Chapter 5 will discuss, through the use of “micro” data this project avoids many of the pitfalls associated with using administratively-defined spatial data because large parts of the model work with individual place data (such as roads and buildings) and depend on individual offender decisions.

2.2 Crime and the Environment

Criminologists are becoming increasingly aware of the important role that “place” plays in crime occurrences (Eck, 1995). However, the relationship between crime and the physical environment in which it occurs is complex. Brantingham and Brantingham (1993) express this idea when they discuss the concept of an “environmental backcloth” that is so detailed as to have an “uncountable” number of elements and is also temporally dynamic. The backcloth includes physical features such as street networks, buildings and land-use types (Brantingham and Brantingham, 2008) and also social elements that will affect the way in which residents or passers-by respond to a (potential) crime event. In fact, a recent special issue of *Built Environment* is devoted to the relationship between crime and its physical/social urban surroundings (Hirschfield, 2008). Each of the “layers” of the environment must be researched in detail in order to establish which are the vital elements that must be explored further with the available data (Chapter 4) and which must be included in a burglary model (Chapter 5). Therefore the subsequent section is divided as follows: Section 2.2.1
outlines the physical environmental factors that influence a burglary; Section 2.2.2 outlines the important social elements of the environment; and Section 2.2.3 concludes with a summary of all the important factors that have been identified as likely to influence burglary.

2.2.1 Physical Factors

It has been shown that the physical form of an area, including natural features and the design of the built environment (Jeffery, 1971; Newman, 1972), has a significant impact on local communities and also on crime in those communities (Bottoms et al., 1992). The complex relationship between features of the built environment and the choice of burglary target can be simplified by using the categorisation scheme introduced by Cromwell et al. (1991):

- **Accessibility** refers to building features that affect the difficulty of actually entering the property;
- **Visibility** alludes to features that will influence how visible the property is to passers by (Cromwell et al. (1991) call this “surveillability”);
- **Occupancy** refers to cues that indicate whether or not someone is at home.

Individually or together, each of these categories have been found to be extremely important factors that influence burglary by a wealth of criminology research studies including Newman (1972); Cohen and Felson (1979); Bennett and Wright (1984); Rengert and Wasilchick (1985); Wright and Decker (1996); Cromwell et al. (1991); Robinson and Robinson (1997); Clarke (1995); Wright and Decker (1996); Felson (2002); Weisel (2002); Bernasco and Luykx (2003); Hirschfield (2004); Cromwell and Olson (2005); Nicholas et al. (2005) and Kent (2006). The remainder of this section discusses literature that addresses each of these three categories with the aim of elucidating the most important factors that need to be included in a model.

Accessibility relates to how easy it is to actually enter a property. A detached house, for example, has more walls that are open to the outside that might provide a potential entry point (a window or a door) making it a higher burglary risk than a semi-detached or terraced house (Felson, 2002). Similarly, ground floor corner flats have been found to be particularly vulnerable because they are very accessible (Robinson and Robinson, 1997). One method of reducing accessibility, therefore, is to control access to certain areas (Clarke, 1995), in this case the grounds surrounding the property. This type of approach is advocated by Newman (1972) in his “defensible space” concepts (although for reasons of increasing community ownership rather than reducing offender access). Although access control has been shown to reduce certain types of crime such as vandalism and theft in public housing (Poyner and Webb, 1987) and burglary in student accommodation (Barberet et al., 2004), its usefulness in reducing residential burglary is unclear as many of the proposed methods (security guards or pin-code access to communal areas, for example) are not relevant to residential properties.

A more common approach to reducing the accessibility of residences is a technique that involves increasing the physical security of potential entry points, commonly known as “target hardening”. Measures can include fitting new locks on doors or windows, the installation of burglar
alarms and fitting chains to doors (Hamilton-Smith and Kent, 2005). Non-security related measures are also useful, such as the fitting of movement detection lighting (Newton et al., 2008). In addition to physical improvements, Weisel (2002) note that factors such as victim education can be as valuable as physical prevention measures and this is reflected in many modern crime reduction practices. For example, students have been found to routinely leave doors unlocked (Fisher et al., 1997) so in Leeds the police have put false “hairy arms” through open doors or windows to educate residents about the risks of leaving their residence unlocked (Marston, 2007).

Target hardening is part of a larger set of methods that come under the umbrella of “Situational Crime Prevention” (Clarke, 1995). Situational crime prevention arose in the UK as a means of directly tackling crime by increasing the effort required and reducing the potential gains (Clarke, 1997). Many authors note that these situational measures have utility. Through direct interviews with burglars, Cromwell et al. (1991) found that most burglars targeted homes that were easy to enter. In the UK, Hirschfield (2004) found that situational crime prevention measures, such as target hardening, were strongly associated with a reduction in burglary across a range of “high crime” communities. Furthermore, Nicholas et al. (2005) found that, from British Crime Survey data, homes with security measures had a significantly lower burglary risk than those without. The Reducing Burglary Initiative (Hamilton-Smith, 2004; Home Office, 2009) invested £25 million between 1999-2002, of which target hardening played an important part in reducing burglary (Hamilton-Smith and Kent, 2005). Support for target hardening also extends to the United States; Weisel (2002), for example, advocates many types of target hardening methods as ways of reducing burglary. Similarly Mayhew (1984) notes that although it is unlikely that professional burglars will be deterred by security measures, the majority will.

Along with the houses themselves, it is also necessary to consider the accessibility of individual items. Cohen and Felson (1979) suggest that reducing the ease of access or removal of a target will also reduce its risk. Alternatively, if the object is an unnecessary source of attraction it could be removed entirely. For example, after reviewing numerous practical crime reduction projects, Poyner (1993) found that removing coin-operated gas or electricity meters from houses – see Pease (1991) for example – was one of the most successful methods of reducing burglary as houses were commonly burgled by offenders attempting to gain access to the meters. Clarke (1983) agrees with some of these suggestions, implying that target hardening and environmental management (removing the target altogether or at least reducing its visibility) will reduce the chances of victimisation. Although coin-operated gas meters as less common these days, the approach is still relevant as an emerging motivation for burglary is the theft of car keys as new car designs have made them very difficult to steal otherwise (Levesley et al., 2004).

There are some authors, however, who doubt the effectiveness of target hardening methods (Robinson and Robinson, 1997). In a large study in Kent, Winchester and Jackson (1982) noticed that security levels were the same for burgled and non-burgled houses. Furthermore, the authors found that occupancy rates, the surrounding environment and the possible rewards were more significant factors in determining victims than security levels. Similarly through interviews with burglars, Wright and Decker (1996) found that many were not deterred by security measures (although the authors admit that this could be discounted as bravado). One the whole, however,
accessibility appears to be a significant factor in determining whether or not a property is likely to be burgled. Therefore this feature must be included in any model of burglary.

The second category of physical factor that might influence burglary is “visibility”. This is described by Cromwell et al. (1991) as a measure of how well a residence can be seen by neighbours or passers-by. There is an abundance of research that suggests visibility has an important influence. For example, Felson (2002) notes that one of the reasons that detached properties in low-density areas can be a higher burglary risk is because vegetation often obstructs them from neighbours or passers-by. Weisel (2002) also recommends that residents try to improve the visibility of their properties by removing objects that might otherwise obscure them. Brown and Bentley (1993) found that effective surveillance from neighbours provided a significant deterrent for burglars. Robinson and Robinson (1997) found that properties that are partially hidden (due to poor street lighting, quiet streets or physical obstructions) and easily accessible were the most heavily victimised. Mayhew (1984) notes that the most vulnerable houses can be approached unseen by neighbours or passers by, regardless of security.

Although obviously popular, Clarke (1995) notes that it is not clear whether the benefits of improving natural surveillance (by cutting back hedges or improving street lighting for example) may have been overstated. This is echoed by Poyner (1993) who found that the results of improved lighting and neighbourhood watch schemes were, in general, uncertain. However, as with accessibility, visibility appears to be an important influence and should be included in any model of burglary.

The final category that relates to the victimisation risk of individual properties is occupancy. All studies that interview potential burglars have found that most will not intentionally enter occupied houses (Cromwell et al., 1991; Wright and Decker, 1996; Brown and Bentley, 1993). These findings are reflected in British Crime Survey data (Kent, 2006) and therefore increasing signs of occupancy have been recommended to residents as a method of reducing their burglary risk (Weisel, 2002).

In summary, the following are considered to be the critical factors that determine whether or not a target is more likely to be burgled (Cromwell and Olson, 2005):

- unoccupied;
- access and entry points not easily seen from the street or from neighbours’ houses;
- the offender will not stand out and be noticed in the neighbourhood;
- the target is accessible and relatively easy to break in to;
- there are valuable items.

Before continuing to the next section, a note of caution must be employed. Many studies referred to here, particularly those that interview burglars directly, were based in the United States. It is not clear, therefore, how relevant these studies are to British culture and housing tenure. For example, Wright and Decker (1996) note that maintaining the “street” lifestyle was the main reason that offenders needed to burglar. It is not clear that statements such as “to be seen as hip on the street, one must be able to keep the party going” (Wright and Decker, 1996, page 201)
could be made by British offenders. This is a point that the authors themselves address, noting that the offenders in their study displayed a much greater tendency towards the use of illicit drugs than that of British burglars. Although this does not mean that non-British studies should be disregarded, it is important to elucidate local knowledge about the crime system in Leeds through close collaboration with Safer Leeds and careful data analysis. Subsequent chapters will document how this is accomplished.

2.2.2 The Social Environment

It has been shown that the physical environment has a significant effect on burglar behaviour. However, the complexity of the “environmental backcloth” (Brantingham and Brantingham, 1993) extends well beyond simple physical factors such as household security. It is also important to consider the social factors that surround a crime event. As Bottoms et al. (1992, page 118) comment, “communities, like individuals, can have careers in crime”. However, the role of the social environment on crime is one of the least understood aspects (Oberwittler and Wikström, 2009) and, as this section will demonstrate, includes a multitude of possible interacting features that might affect crime.

Crime and Socioeconomic Status

Like crime and the built environment, the relationship between crime and the community is complex. For example, areas with a “low” socioeconomic status have often been found to suffer disproportionately high crime rates (Baldwin and Bottoms, 1976a; Sampson et al., 1997) but, particularly with burglary, the reverse has also been found (Wilkström, 1991; Bowers and Hirschfield, 1999). From these findings a fair assumption, and one that fits with burglar interviews (Cromwell et al., 1991; Wright and Decker, 1996) is that burglars predominantly live in areas with low socioeconomic status but travel to more affluent areas to burgle if possible. This hypothesis is supported by modern criminology theory (Section 2.4 will expand on this) and also by empirical research (Wilkström, 1991; Bowers and Hirschfield, 1999; Bernasco and Luykx, 2003; Snook, 2004).

However, the socioeconomic status of a community is not the only factor that will influence how attractive an area is to potential burglars. Students, for example, are seen as an easy target (Deakin et al., 2007) and often suffer disproportionately high levels of victimisation (Tilley et al., 1999; Barberet et al., 2004). This is particularly true in Leeds, where student communities are very heavily victimised largely from people who live outside the area (Shepherd, 2006). Section 4.4.2 will present real crime data that supports this. However, burglars often base individual assessments regarding the quality of goods contained within a property on the general affluence of an area (Cromwell and Olson, 2005) so other indicators of affluence – such as a large car parked outside (Wright and Decker, 1996) – can also be important and might not be covered in measures of socioeconomic status. For this reason, careful data analysis will be required to supplement the literature in order to determine which factors are the most important to burglars.
Crime and Community Cohesion

It is clear that a measure of socioeconomic status or affluence is important to establish how attractive a target is to a burglar. However, there is a crucial element that these types of measures miss: community cohesion. Shaw and McKay (1942) present the first published example of a theory that directly links crime to community stability. Their Social Disorganisation Theory suggests that crime is the natural response by an individual to a community’s inability to police itself. This can occur when an area has a largely transient population; people do not make close social ties and a feeling of “community” never develops.

In the 1970s, these ideas were revisited by Jeffery (1971) who presented a new approach to crime reduction entitled “Crime Prevention Through Environmental Design”. Jeffery’s theory promoted making crime less attractive by changing the physical environment and also making long-term societal changes to reduce crime (such as reducing unemployment and providing communities with the tools to police themselves). Newman (1972)’s theory of “Defensible Space” is similar to Jeffery’s ideas. Newman’s approach to crime reduction involved increasing a community’s sense of ownership over its property, therefore allowing it to deter criminals by policing itself. These ideas, in particular, have been embraced in recent years and form a large part of modern crime reduction initiatives (Newman, 1996). A similar theory, which has also received a considerable amount of attention is Wilson and Kelling (1982)’s “Broken Windows Theory”. The theory stipulates that “one unrepaired broken window is a signal that no one cares, and so breaking more windows costs nothing” (Wilson and Kelling, 1982). This process can lead to a spiral of decay which decimates previously safe, respectable areas.

There is a large body of criminological research that supports these theories, strengthening the link between community cohesion and crime. For example, many researchers have found that offenders were less likely to target neighbourhoods in which the residents would “look out for each other” and might notice them or intervene (Hope, 1984; Brown and Bentley, 1993; Wright and Decker, 1996; Sampson et al., 1997). Similarly, Sampson et al. (1997) has shown that violence can be linked to factors such as the social and organisational structure of a neighbourhood more strongly than socioeconomic variables. The authors found that neighbourhoods with strong social ties were more likely to act against crime and therefore experienced lower crime rates. In a large survey of the crime modelling literature, Marris (2001) finds this work “compelling”. Kawachi et al. (1999) also drew very similar conclusions; they attempted to correlate the levels of different types of crime with variables that represent the quality of social relationships and conclude that:

“Increasing evidence points to social cohesion as a vital ingredient for the maintenance of collective wellbeing, and crime is the mirror of the quality of social relationships among citizens”(Kawachi et al., 1999).

Although the “broken windows” theory – and other theories of social disorganisation – have been accepted and incorporated into policy in a number of American (Brook, 2006) and also British (Hamilton-Smith and Kent, 2005) cities, some researchers question their validity. Harcourt and Ludwig (2006) cite an experiment in which poor families living in a high-crime public hous-
ing area were given vouchers to move to safer, less disorganised neighbourhoods. Once moved, however, the families continued offending at the same rate as in the inner-city (Brook, 2006). Also, Sampson and Raudenbush (2004) found that perceived disorder was more closely linked to racial and economic factors rather than cues from the physical environment.

As previously noted, a word of caution must follow the literature discussed here because, again, American studies might not relate well in the UK. Here the work of Sampson et al. (1997), for example, is based on neighbourhoods in the city of Chicago but previous models of Chicago, such as the concentric ring model (Burgess, 1925), do not correspond to council housing policies in Britain (Bottoms and Wiles, 1986) who build wherever they own land (Baldwin and Bottoms, 1976a). Another unrelated, but important problem, with many of the qualitative studies is that they can potentially suffer from problems associated with sampling of a small population and lack of rigorous empirical testing. Quantitative studies are not free from potential problems either. Section 2.1.2 illustrated how important it is to analyse crime at micro-levels but many studies discussed here use highly aggregated data. This has the effect of smoothing out natural variation in the environment, makes them susceptible to the ecological fallacy (Robinson, 1950) and modifiable areal unit problem (Openshaw, 1984a,b) and means that they are unable to represent the micro-level human and environmental factors that dictate whether or not an individual crime event will occur. There is no solution to these problems, but if affected research were disregarded there would be no base on which to build the model. However it is important to be aware of these potential problems and mediate for them where possible.

In summary, this section has illustrated that the relationship between crime and the community in which it is committed is extremely complex. One must consider whether high crime rates are a result of an abundance of potential burglars, due to features of the community itself, or as a result of the physical cues that dictate how the community is perceived. Even if the effects of the community can be isolated, establishing which community features are important is problematic. Although high socioeconomic status has been linked to burglary due to the abundance of attractive goods, community cohesion has also been shown to be important. Combining all these elements into a single theory that can be implemented in a model is non-trivial. Fortunately, criminologists have a number of tested theories that help to explain some of the complexity of the system and these will be reviewed in Section 2.4.

### 2.2.3 Summary of Crime Attractors

Table 2.1 summarises some of the major socioeconomic and environmental factors that have been found to be associated with crime (or specifically burglary) events. These are important because they will feed into the model, both from the perspective of the offender and also from the perspective of knowing what to include in the environment. Integration of the factors into the model is the subject of Chapter 5.
Table 2.1: Environmental factors found to attract or deter crime.

<table>
<thead>
<tr>
<th>Factor</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accessibility</strong></td>
<td></td>
</tr>
<tr>
<td>Highly accessible/visible object</td>
<td>Cohen and Felson (1979)</td>
</tr>
<tr>
<td>Detached house</td>
<td>Felson (2002)</td>
</tr>
<tr>
<td>Ground floor flat</td>
<td>Robinson and Robinson (1997)</td>
</tr>
<tr>
<td>Low security</td>
<td>Clarke (1983, 1995); Weisel (2002); Hirschfield (2004); Hamilton-Smith and Kent (2005); Nicholas et al. (2005); Newton et al. (2008)</td>
</tr>
<tr>
<td><strong>Visibility</strong></td>
<td></td>
</tr>
<tr>
<td>Poor street lighting, quiet streets</td>
<td>Robinson and Robinson (1997); Rengert et al. (1999)</td>
</tr>
<tr>
<td>Property obstructed from view</td>
<td>Robinson and Robinson (1997)</td>
</tr>
<tr>
<td>No surveillance from neighbours</td>
<td>Brown and Bentley (1993)</td>
</tr>
<tr>
<td><strong>Occupancy</strong></td>
<td></td>
</tr>
<tr>
<td>No signs of occupancy</td>
<td>Cromwell et al. (1991); Brown and Bentley (1993); Wright and Decker (1996); Kent et al. (2006); Cromwell and Olson (2005)</td>
</tr>
<tr>
<td><strong>Attractiveness</strong></td>
<td></td>
</tr>
<tr>
<td>“Good outer appearance”</td>
<td>Wright and Decker (1996)</td>
</tr>
<tr>
<td>High affluence</td>
<td>Cromwell and Olson (2005)</td>
</tr>
<tr>
<td>High perceived gain</td>
<td>Rengert et al. (1999)</td>
</tr>
<tr>
<td><strong>Socioeconomic Status / Deprivation</strong></td>
<td></td>
</tr>
<tr>
<td>Low socioeconomic status</td>
<td>Baldwin and Bottoms (1976a); Wilkström (1991); Sampson et al. (1997)</td>
</tr>
<tr>
<td>High level of deprivation</td>
<td>Baldwin and Bottoms (1976a); Sampson et al. (1997)</td>
</tr>
<tr>
<td>Large wealth disparity</td>
<td>Wilkström (1991); Bowers and Hirschfield (1999); Kawachi et al. (1999)</td>
</tr>
<tr>
<td>High population density</td>
<td>Craglia et al. (2001)</td>
</tr>
<tr>
<td>High unemployment</td>
<td>Brantingham and Brantingham (1993).</td>
</tr>
<tr>
<td>Level of education</td>
<td>Schmid (1960); Carr (2003)</td>
</tr>
<tr>
<td>Presence of students</td>
<td>Shepherd (2006); Deakin et al. (2007)</td>
</tr>
<tr>
<td><strong>Community Cohesion</strong></td>
<td></td>
</tr>
<tr>
<td>Residential mobility</td>
<td>Shaw and McKay (1942); Sampson et al. (1997); Craglia et al. (2001)</td>
</tr>
<tr>
<td>Racial segregation</td>
<td>Sampson et al. (1997)</td>
</tr>
</tbody>
</table>
2.3 Crime and the Offender

The previous sections have discussed how potential burglars are influenced by the built environment and the community that surround a potential burglary target. This purpose of this section is to summarise these behavioural features to make it clear what needs to be included in a model. These are also closely related to the environmental criminology theories that are discussed in the following section. There are two main branches of research into understanding how potential burglars behave, their motivations and their responses to environmental cues. These can be broadly classified as qualitative using interview data and quantitative using large data sets and statistical models to establish trends and patterns of potential burglar behaviour. Although different in methodology, these studies draw very similar conclusions. In general, the motivations that can lead to a person committing a crime are wide ranging and depend upon individual circumstances and the type of crime. However, with regards to burglary the literature is largely in agreement (Cromwell and Olson, 2005). The main findings can be summarised as follows:

- It is a common finding that most burglary is motivated by the need for money (Repetto, 1974; Bennett and Wright, 1984; Rengert and Wasilchick, 1985; Cromwell et al., 1991; Wright and Decker, 1996; Bernasco and Luykx, 2003; Nee and Meenaghan, 2006).
- It follows that the expected “yield” is the most important consideration when selecting a target (Hearnden and Magill, 2003; Nee and Meenaghan, 2006) which can range from $0 – $12,950 (Snook, 2004).
- However, it is also widely accepted that most burglars are drug addicts and burgle in order to satisfy a drug habit (Scarr, 1973; Cromwell et al., 1991; Wiles and Costello, 2000; Hearnden and Magill, 2003) or to maintain a “high living”, e.g. displaying status symbols such as clothes, cars or jewellery (Scarr, 1973; Wright and Decker, 1996, 2005).
- A burglary is generally in response to a “pressing” demand rather than longer term commitments (Wright and Decker, 2005). It follows that when assessing the risk associated with a crime, potential offenders rarely consider long term criminal-justice sanctions, concentrating more closely on immediate problems such as occupancy or the amount of surveillance (Cromwell and Olson, 2005; Wright and Decker, 2005).
- To guarantee a good yield, most burglars will return to previously burgled properties, usually because they know what goods are available and how to enter the property (Wright and Decker, 1996; Hearnden and Magill, 2003). Also, they use “templates” (Brantingham and

<table>
<thead>
<tr>
<th>Factor</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low social cohesion</td>
<td>Brown and Bentley (1993); Wright and Decker (1996); Sampson et al. (1997); Kawachi et al. (1999)</td>
</tr>
<tr>
<td>Transient population</td>
<td>Shaw and McKay (1942); Craglia et al. (2001)</td>
</tr>
<tr>
<td>Low feeling of community “ownership”</td>
<td>Newman (1972, 1996)</td>
</tr>
</tbody>
</table>
2.4. CRIME THEORIES

Brantingham, 1993) and they seek familiarity in their targets which can best be obtained by revisiting previous victims or, at least, similar properties (Johnson et al., 2009).

- Few burglars are purely opportunistic (Nee and Meenaghan, 2006); the decision to burgle is often made away from the actual crime scene and the potential offender then travels to a target noted previously (Wright and Decker, 1996; Hearnden and Magill, 2003; Nee and Meenaghan, 2006). These targets have usually been found by passing them on their routine activities (Wright and Decker, 1996; Cromwell and Olson, 2005) so properties close to a burglar’s home are more likely to become victims (Snook, 2004; Bernasco and Nieuwbeerta, 2005).

- However, situational factors are often the most important so there is usually a degree of opportunity to a burglary (Cromwell et al., 1991). For example, burglars will not usually enter occupied properties (Cromwell et al., 1991; Wright and Decker, 1996; Nee and Meenaghan, 2006) or properties that are highly secure (Mayhew, 1984; Cromwell et al., 1991).

- Due to better knowledge of their victim’s daily habits and the potential items available for theft, many burglars admit burgling people they know personally (Wright and Decker, 1996; Hearnden and Magill, 2003).

- This is partly because the offender knows the area well and does not need to carry stolen objects too far (Hearnden and Magill, 2003) and also because the potential burglar chooses targets from within their cognitive awareness space (Bernasco and Nieuwbeerta, 2005).

- Burglars can be classified as experts in their field (Nee and Meenaghan, 2006); they possess a range of both behavioural characteristics and specific knowledge that is unique to them.

As with previous sections, some caution must be employed when discussing North American studies as there might be limited correspondence with British offenders.

2.4 Crime Theories

In previous sections, features of the built environment and the social environment that surround a crime event have been discussed along with various behaviours and motivations of potential burglars. However, combining these disparate concepts into a single model is extremely difficult. Fortunately, environmental criminology has made great strides towards integrating all of these effects into sound theories. These will be outlined here, collating all the aforementioned ideas into theories that can be directly implemented an a burglary model. The section will conclude with a brief discussion about the offender’s journey to crime and, based on the current theories, some considerations that guide the choice of spatial scale used to analyse crime.

2.4.1 Routine Activity Theory

Credited with being the first published environmental criminology theory (Andresen, 2010), routine activity theory was developed by Cohen and Felson (1979). Simply put, the theory states that for a crime to occur three elements must be present: a target, an offender and the absence
of a capable guardian. With respect to burglary, the target is a house or a particular object in the house and suitable guardians could be the owner, neighbours or passers-by. Whether or not these elements converge in the same space at the same time depends on their routine daily activities. It is particularly relevant for environmental criminology because the theory represents a break from social disorganisation theory, focusing on the behaviour of individuals rather than areas and inherently considers the temporal/spatial importance for crime. The theory was later adapted to include the concepts of a suitable “place” for the crime to occur (the neighbourhood), a “manager” of the place (police or citizens) and a “handler” who watches over the offender (such as the parent of a juvenile) as illustrated by Figure 2.1. To summarise the key elements, Felson notes:

“A crime occurs when the offender escapes handlers, finds targets free from guardians in settings not watched by managers.” (Felson, 2008, page 74)

![Figure 2.1: The “crime triangle”, as proposed by Clarke and Eck (2005)](image)

Although relatively simple, the theory offers an explanation for the increased rates of property crime in a period of increasing prosperity in America during 1947–1974. Cohen and Felson (1979) note that during this time, people increasingly left homes unoccupied during the day, providing greater opportunities for offending in a society that might otherwise expect to see declining crime rates due to increases in prosperity. Clearly the theory provides an explanation for many of the results from studies outlined previously, particularly with respect to one of the most significant deterrents of burglary: occupancy.

Also, the theory explains some of the temporal changes that crime data exhibit. For example, Chainey et al. (2004) found a peak in crimes committed by the 11-14 age group at 4pm which corresponds to the approximate end of the school day. Robinson and Robinson (1997) found that student apartment complexes were usually burgled between 8am-5pm, when students were attending lectures, and between 6pm-7am on weekends when students were usually out for recreational
2.4. CRIME THEORIES

activities. While investigating the methods and processes of street robbery, Deakin et al. (2007) found that most robberies were committed during daylight hours. Deakin et al. (2007) also found that the time at which an offence will occur is also dependant on the offender: drug users, for example, often committed street robbery in response to the immediate need for drugs which can be strongest upon waking up. Furthermore, burglaries often correspond to times when parents are travelling to schools with children in the mornings and afternoons (Cromwell and Olson, 2005; Rengert and Wasilchick, 1985).

Routine activity theory provides a considerable amount of information that could be used to create realistic burglar behaviour in a model. For more information about the theory, the reader is directed to the original paper (Cohen and Felson, 1979) or a more recent reflection (Felson, 2008).

2.4.2 Crime Pattern Theory

Initially conceived by Brantingham and Brantingham (1981a), the theory is ultimately concerned with how the “environmental backcloth” influences crime. For example, how people move around an urban environment, how they perceive it and how it thus influences their behaviour will determine whether or not a crime takes place (Felson and Clarke, 1998). It complements routine activity theory well, but rather than considering macro-effects (such as how societal changes are influencing peoples’ daily habits) it focusses on a smaller unit of analysis, considering the actual routes that people use to travel around a city and when these might combine with other environmental factors to facilitate a crime. This is particularly important for establishing where burglary might be likely to occur. Burglars do not search for targets at random, instead they look for targets near important “nodes” such as friends’ houses and leisure or work places (Brantingham and Brantingham, 1993). Crime pattern theory, therefore, examines these important nodes, the paths between them and the edges which mark social boundaries and are often associated with higher crime rates.

Also important are the ideas of activity and awareness spaces that portray the different levels of awareness that individuals have of their environment. These spaces will vary by age and by socioeconomic status: the poor and the young will have less developed spaces (Brantingham and Brantingham, 1993). The spaces will also be influenced by (public) transport routes because they affect the routes that people use to navigate their environment (Beavon et al., 1994; Brantingham et al., 2009). Within awareness spaces are areas of which an individual has greater knowledge; these are called activity spaces and are usually where criminal opportunities are chosen from (Brantingham and Brantingham, 1993). Where these areas overlap with crime opportunities are where the person might commit an offence. Figure 2.2 illustrates how important nodes (also termed “anchor points”) and the paths between them build up awareness and activity spaces.

Along with offender behaviour, crime pattern theory can be used to explain how the characteristics of particular areas can shape city-wide patterns of crime. Brantingham and Brantingham (1995) introduce the concept of crime generators, crime attractors, crime detractors and crime neutral places:

- Generators are locations whose activity attracts large numbers of people. Opportunities then
arise as a result of the mixture of different types of people and objects in a setting that is propitious to a particular type of crime. Typical crime-generating areas include shopping centres, sports stadiums, transport hubs, office complexes etc.

- **Attractors** are areas that are well known for the crime opportunities they offer and therefore attract people specifically for the purposes of participating in crime. Crime attractors include areas with high rates of drug use (for buying/selling drugs), area known for prostitution, shopping centres (for shoplifting) and car parks (for theft of/from a motor vehicle).

- **Detractors** are places or buildings that discourage crime. These include a stable business, abundant natural surveillance, high security and the presence of particular people (i.e. middle-aged women) (Felson and Clarke, 1998).

- **Neutral areas** neither attract potential offenders nor generate opportunities and are likely to suffer only occasional crimes committed by locals (Brantingham and Brantingham, 1995).

### 2.4.3 The Rational Choice Perspective

Routine activity theory addresses the locations of offenders, victims and guardians and observes how these change as a side-effect of the routine activities of each individual. Crime pattern theory then examines these travel patterns more closely, linking them to the potential opportunities offered by the environmental backcloth. Although these two theories can determine whether or not it is likely for a crime to occur, neither considers the decisions which must be made by the offender (Groff, 2006). This is the subject of the rational choice perspective (Clarke and Cor-
2.4. CRIME THEORIES

nish, 1985) as illustrated by Figure 2.3. The perspective suggests that offenders’ thoughts can be modelled as a formal decision, weighing up potential gains of a successful crime with the potential losses if apprehended. In this manner, a crime will only be committed if it is perceived as profitable. Rather than a traditional criminological theory, the rational choice perspective is a “conceptual tool” that provides a novel method of looking at offending and in particular the influence of the environment over behaviour (Cornish and Clarke, 2008).

The idea of offenders being entirely “rational” has led to criticisms of the approach (Cornish and Clarke, 2008), but this appears to stem largely from a misunderstanding of the term “rationality”. With the rational choice perspective, rationality is bounded, so that a decision is optimal only in the sense that the person is bounded by their own perceptions and the information available to them at the time (Andresen, 2010). Therefore a crime that appears completely irrational might be the most attractive choice to a particular person situated in a particular environment at a particular time, the decision being the optimal one for that person.

![Diagram of routine activity theory, crime pattern theory and rational choice perspective in relation to scale.](image)

2.4.4 The Journey to Crime

As this research will focus on individual-level actions it is imperative to properly understand a burglar’s journey to a burglary. Routine activity theory and crime pattern theory tell us a lot about where a burglar might be expected to travel to in order to commit a burglary. For example, a person’s awareness space develops as they travel around their environment on potentially legitimate business such as travelling between home and work. So a burglar is likely to choose a target (or start a search) from one of the houses on these routes. However, offender research also points to an absence of daily structure in offenders’ lives (Baldwin and Bottoms, 1976b; Carr, 2003; Cohen and Felson, 1979) so their routine activities do not follow “legitimate” patterns such as a daily commute to work. Instead, movements are driven by opportunities, needs and temptations as they present themselves (Wiles and Costello, 2000). Section 6.4 will demonstrate how this factor can be included in a burglary model.
With respect to the distance travelled, it might be assumed that, as travel has increased over the years, this trend is reflected in the distance that offenders will travel to commit crime (Wiles and Costello, 2000). There are many authors, however, who indicate this is not the case and describe offender movements as local in nature (Baldwin and Bottoms, 1976b; Bottoms et al., 1992; van Nes, 2006; Bennell et al., 2007), i.e. less than two miles from home (Wiles and Costello, 2000). Wright and Decker (1996) found that most preferred to burgle areas relatively close to their homes, as there are logistical problems associated with travelling to distant targets. It has also been found that areas of high socioeconomic class often suffer disproportionately high levels of crime, particularly if they are close to poorer areas (Wilkström, 1991) which further suggests limited offender travel. In addition, it has been found that the distance travelled increases with the age of the offender (Wiles and Costello, 2000). These studies all offer support for the crime theories insomuch as the area someone knows best is most likely to surround their home.

Although it has been suggested that many crimes are opportunistic in nature, some authors suggest that potential offenders can perform specific search patterns to look for an opportunity. The *commuter hypothesis* suggests that an offender travels from their base to a certain area and then commits a crime in that area (Canter and Larkin, 1993). The *marauder hypothesis*, on the other hand, suggests that the base acts as the focus point for an offender’s crimes: the offender travels from home to commit a crime in their local area and then returns back to their base (Canter and Larkin, 1993). Offender search patterns have even been compared to those exhibited by foraging insects (Brantingham, 2006). These findings will be explored further using the available Leeds data in Section 4.5 and will inform offender searching behaviour discussed in Section 5.10.

### 2.5 Summary – Understanding Burglary

The previous sections have discussed the relationship between crime and its environment, setting the groundwork on which a burglary model can be based. Section 2.1 provided a brief introduction to environmental criminology and the geography of crime. The section demonstrated that criminologists are realising how important it is to consider the environment in which a crime occurs (rather than victims and offenders in isolation) and, in particular, that crime is a location-specific event that should be analysed at sufficiently high resolutions (aggregating to community boundaries often hides important patterns).

Sections 2.2 followed with a discussion about the important physical and social characteristics that will influence whether or not a crime is likely to occur. In particular, the influence that a community and the design of the built environment can have over crime was established. Section 2.3 then discussed the behaviour of the offender, summarising their motivations and how they respond to environmental cues.

To follow, Section 2.4 outlined the main theories in environmental criminology: routine activity theory (Cohen and Felson, 1979); crime pattern theory (Brantingham and Brantingham, 1981a) and the rational choice perspective (Clarke and Cornish, 1985). These theories combine information at different scales (ranging from large societal changes to individual offender decisions) and coordinate much of the otherwise disparate evidence provided in previous sections. They can be
used in conjunction with the results of empirical studies to build a sound burglary model. The interested reader in search of more information about the theories might try a recent edited collection that contains papers by many of the theories’ authors (Wortley and Maxerolle, 2008) or an excellent introductory article in a collection of seminal papers that places the theories in their wider criminological contexts (Andresen, 2010).

Having established the important crime theories that should be included in a model, the following chapter will review methodological literature that discuss how the burglary system could be modelled.
Chapter 3

Modelling Burglary

Contents

3.1 Introduction ................................................................. 27
3.2 The Difficulties with Modelling Social Systems ......................... 28
3.3 Traditional Methods for Modelling Crime .............................. 30
3.4 Computer Modelling of Crime ........................................... 33
3.5 Agent-Based Models of Crime ............................................ 43
3.6 Summary – Modelling Burglary ........................................... 47

3.1 Introduction

Explaining crime patterns is an exercise that has taxed policy-makers, criminologists, social reformers and the police ever since the first patterns were recorded (Marris et al., 2003). The previous chapter provided a review of the relevant crime literature in order to better understand the burglary system and to ascertain which factors need to be included in a burglary model. To understand the requirements and pitfalls associated with building the model itself, this chapter will outline how crime has been modelled previously. This will situate the research within the wider crime-modelling literature and inform the design of the methodology. To build a model that can replicate the processes and dynamics that contribute to the occurrence of burglary, it is necessary to understand the difficulties of modelling the burglary system that must be overcome and this is the subject of Section 3.2. Then to understand how this research can improve upon other crime modelling work, Section 3.3 reviews some of the “traditional” statistical techniques that have been used to model crime. Section 3.4 then details some more advanced modelling techniques that commonly rely on the use of computers. In particular, agent-based modelling is introduced and put forward as the most appropriate modelling technique. Finally, Section 3.5 introduces current research that applies agent-based modelling to crime in order to clarify how this research improves upon others to date.
3.2 The Difficulties with Modelling Social Systems

Social systems (including the burglary system) belong to a class of system known as “complex systems”. These types of systems have a number of defining characteristics that will influence how successfully they can be modelled and the types of modelling techniques that are likely to hold promise. Before discussing social systems in general, a short introduction to complex systems will follow.

3.2.1 Chaos, Complexity and Complex Systems

Reductionism is the principle that a system can be explained by observing and understanding its constituent parts. This approach has been guiding the hard sciences for millennia and is one of the principles driving the segregation of scientific activity into distinct hierarchical disciplines: physics, chemistry, biology, psychology, sociology and so on. Each discipline studies problems at a certain level of complexity, building upon the knowledge of the level before it. However, this approach does not necessarily lead to a better understanding of some systems; the ability to deconstruct a system into its constituents does not necessarily mean that it can be re-constructed from those parts (Anderson, 1972). For example, understanding sub-atomic interactions does not help us to explain the social/environmental interactions that lead to a person committing a burglary.

However, having established that these systems cannot be explained through traditional reductionist science, how should they be? It is on this basis that the chaos theory and the study of complex systems arose. Chaos theory is commonly studied in mathematics and relates to a class of systems that are deterministic but highly dependent on initial starting conditions. They are chaotic in the sense that it is impossible to predict the state of the system a long way into the future, but possible to predict for shorter time periods with very high accuracy (Flake, 1998). This effect means that in one respect it is fruitless to attempt to predict chaotic system behaviour (it is impossible to make accurate long-term predictions) but conversely it enables enquiry into systems that would otherwise be assumed to be random (O’Sullivan, 2004). Lorenz (1963, 1993) was possibly the first to realise chaos theory with his simulated weather system; he found that two simulations would quickly diverge even if the initial conditions were almost identical.

But why is this relevant to burglary modelling research? Chaos theory is closely related to the idea of complex systems and we shall show the burglary system to fall into this category. Complex systems contain a large number of relatively simple components that behave in response to simple rules. Although the individual components are simple, the global patterns that result from their interactions cannot be understood by examining the rules which govern a single unit in isolation (Darley, 1994; Mihata, 1997), i.e. ‘the whole is greater than the sum of its parts’. In fact, a single entity in one complex system, might itself be a complex system (Batty, 2009). Ant systems are an excellent example; although the rules that govern individual ants are few and simple, examining a single ant does not help us to understand how a colony emerges or how it behaves. It is the relationships between the ants that are important; in general, complex systems are usually defined more by the interactions between components than by the constituent parts.
3.2. THE DIFFICULTIES WITH MODELLING SOCIAL SYSTEMS

Complex systems are chaotic in the sense that their behaviour is non-linear but it can be predicted, in the short term, if we know the rules that drive the individual components and can simulate them correctly. The most fruitful approach to modelling these types of systems, therefore, is to model the individual units, allowing them to interact as they would do in reality (Bonabeau, 2002). In this sense research still follows reductionism by first breaking down a system to understand its constituents, but then rebuilding the system to understand it holistically (O’Sullivan, 2004), often though computer simulation.

Before continuing, it is worth noting that the definitions of chaos and complex systems are much less clean in the literature than put forward here (O’Sullivan, 2004). Manson (2001) suggests that deterministic complexity is used to refer to complex mathematical systems (and is more closely related to chaos theory) whereas aggregate complexity is used to refer to systems where the complexity arises from interactions of numerous individuals. However, in general there are a number of theories concerned with complex systems that all describe themselves as “complexity research” (Manson, 2001) and often the terms are used interchangeably (Johnson and Burton, 1994). As we will see, it is the “aggregate” form of complexity that is relevant to this research.

3.2.2 Complex Social Systems

The term “complex system” can be used to describe a broad range of systems, including the stock market, termite mounds (and the termites themselves), weather systems, human beings and, importantly, human systems. The burglary system is one such complex system. It consists of large numbers of diverse components (offenders, victims, guardians and the urban environment itself), varied and complex interactions (which are often non-linear and unknown) and emergent properties (e.g. city- or neighbourhood-wide burglary rates) that cannot be attributed to any individual part of the system. Coupled with this inherent system complexity are “human” factors, such as complex psychology, subjective choices and seemingly irrational behaviour that further compound the rules that drive the individuals (Bonabeau, 2002).

To complicate the situation further, the geography of the system itself introduces additional complexity. For example, the characteristics of spatial complex systems vary depending on the scale (O’Sullivan, 2004). A city that appears in a state of equilibrium is most likely in a constant state of flux if observed at smaller (neighbourhood / street / household) scales. This is particularly true with crime as Section 2.1.2 indicated (Eck, 1995; Weisburd et al., 2004; Bowers et al., 2003; Andresen and Malleson, 2010).

For these reasons, scientists struggle to create realistic aggregate behaviour from the individual components (Costanza et al., 1993). This chapter will continue by outlining the common approaches that have been used to model crime, discussing their advantages and disadvantages in the context of modelling complex systems.
CHAPTER 3. MODELLING BURGLARY

3.3 Traditional Methods for Modelling Crime

The crime modelling literature is “voluminous” (Marris, 2001). This section will not attempt to review all crime studies in detail, indeed many have already been discussed in Sections 2.2 and 2.3, but will instead discuss a small number of studies that are methodologically representative of common crime modelling work to provide a sufficient overview of how crime has been traditionally modelled.

Since the earliest explorations (e.g. Glyde, 1856), researchers have found that crime exhibits clear spatial variation. For this reason, crime mapping techniques are a popular method for describing or modelling crime. Shaw and McKay (1942), for example, utilised mapping techniques in their seminal work on juvenile delinquency in Chicago. Similarly Baldwin and Bottoms (1976a) used maps to describe the spatial distribution of crime and offenders in their influential Sheffield study and found that housing tenure was particularly important. However, the techniques employed by Shaw and McKay and Baldwin and Bottoms were, by modern standards, primitive. Advances in geographical information systems (GIS) and the availability of individual-level data have catalysed the development of more advanced mapping analysis techniques such as “hotspot” detection (Grubesic and Murray, 2001). These methods use spatial mathematical techniques such as kernel density estimation to gauge where the highest crime areas are and how these relate to resource deployment (Chainey and Ratcliffe, 2005). For example, Pain et al. (2006) overlaid crime hotspot maps with streetlight location maps to investigate the impact that street lighting had on crime and fear-of-crime levels; the results were used to inform existing street lighting policies. Advancing the field further, Brantingham and Brantingham (1998) use location quotients (a statistical technique more commonly used in regional economics/planning) to help identify whether or not crime is particularly high in a given area, based on its surroundings. A similar technique, the weighted displacement quotient (Bowers and Johnson, 2003), has also been used in practice to mathematically search for crime displacement after a crime-reduction initiative (Bowers et al., 2003; Hirschfield, 2004). Although these techniques are invaluable for crime prevention practitioners (due in part to their ability to highlight areas with unusually high crime rates), they fail to provide insights into the dynamics and processes that govern crime systems. Furthermore they are limited in their predictive power.

More common than using mapping techniques directly, however, are mathematical (statistical) models. Crime analysis in a “traditional” sense (following the pioneering work of Guerry (1831) and Quetelet (1842)) use aggregate crime rates or counts as the dependent variable in some statistic (e.g. a regression equation) (Brantingham and Brantingham, 1998; Groff, 2007a). The methods used are numerous and constantly changing, many being adapted from other fields. Early examples include the use of principal components analysis to investigate the factors related to social deprivation (Giggs, 1970) and cluster analysis to search for associations between crime and environmental factors (Brown et al., 1972), although these approaches are strongly criticised by Baldwin (1975) who describes them as “unilluminating”. Instead, Baldwin and Bottoms (1976a)

---

1 A GIS is a piece of software designed to analyse and manipulate geographical data and they are becoming increasingly popular in many applications.
performed a linear regression analysis of offender rates in relation to census data and found that the patterns emerging were very different in areas of different tenure type.

More recent statistical modelling has been centred around the use of regression models. It is not necessary to differentiate between different regression models in detail because, as we will see, they share similar drawbacks. However some approaches will be briefly introduced; for a review of different regression techniques used in crime modelling the reader is directed to Kongmuang (2006). Craglia et al. (2001) compared high intensity crime areas to census data and Dahlbäck (1998) found high population density and weak social bonds to be associated with high theft rates through application of longitudinal multivariate regression. Tseloni et al. (2004) used negative binomial regression based on routine activity theory to compare cross-national burglary victimisation in England and Wales, the United States and the Netherlands and found that the variables used to represent the proximity to potential offenders homes and the level of guardianship were both significant in predicting burglary. These offer support for the routine activity theory (see Section 2.4.1) on a cross-national level. Other studies using regression include Gaviria and Pagès (2002) who linked the chance of being a victim to individual and city-wide variables, and Meera and Jayakumar (1995) who attempted to explain the relationship between rising levels of crime and different demographic and economic variables. The British Crime Survey has also encouraged many regression analyses. For example, Hope et al. (2001) applied a multivariate statistical model to investigate multiple victimisation and Tseloni et al. (2002) generated statistical models (negative binomial regression) to investigate individual household property victimisation (using hypothetical houses). For the interested reader, Tseloni et al. (2002) also write extensively on the use of the British Crime Survey for modelling property crime.

An adaption of a standard regression model that deserves note is geographically weighted regression (GWR: Brunsdon et al., 1996). Using this advancement, instead of applying a constant, global weight to each explanatory (independent) parameter, the weight is varied spatially for each area. This should increase the accuracy of the overall model by not suppressing spatial variation in parameter values and also allow for visualisation of the values to indicate where the parameters have a greater or lesser affect. Examples of the use of this algorithm include Cahill and Mulligan (2007) who compared GWR to global regression for crime in Portland, Oregon and Lee et al. (2009) who used GWR to explore crime in Seoul. Although obviously useful, the approach suffers from similar drawbacks to other regression approaches (outlined below).

The research outlined here study wide-ranging crime effects using diverse methods, but they all share some similar characteristics and also some drawbacks. These can be summarised as follows:

- **Inherent complexity of the system.** Section 3.2 noted that the burglary system was a complex system and that the best way to address the inherent complexity is to model the individual system components themselves and not through the use of global equations or rules.

- **The importance of “place”**. Environmental criminology has demonstrated that the micro crime setting is extremely important in determining whether or not an individual crime will
occur (see Section 2.2). Statistical techniques commonly use aggregated data and therefore struggle to account for the micro-effects that may result in a significant variation in crimes on a street by street basis (Andresen and Malleson, 2010).

- **Travel patterns.** Traditional techniques often use Euclidean distance measures which do not take road networks or impassable barriers (such as major roads or rivers) into account. This will have a strong influence on where people travel, their internal awareness spaces and subsequently where they are likely to commit crime.

- **Temporal variations.** Crime hotspots vary temporally (Chainey and Ratcliffe, 2005). By aggregating data temporally (such as on an yearly basis) this important variation is lost.

- **The modifiable areal unit problem and the ecological fallacy** (Openshaw, 1984a,b) are always present when data are aggregated spatially. This can give rise to misleading assumptions about crime in an area and subsequently lead to a poor model.

- **Limited number of variables.** Traditional statistical techniques are generally limited to a small number of variables (Brantingham and Brantingham, 1993) so cannot account for the complexity of the environmental backcloth and the non-linear human-human or human-environment interactions that drive the system.

- **Human heterogeneity and behaviour.** Crime studies have demonstrated that, although there are some common themes, the techniques used by burglars vary from person to person. Furthermore, there are other important characteristics that are unique to each person (such as their individual awareness spaces) which are not homogeneous across the population and statistical studies will struggle to account for these variations or for the unique effects of human behaviour which drive their actions.

In summary, the central drawback common to each of the approaches discussed above is that they fail to address the importance of individual incidents located in a specific time and space involving individual people. Instead, findings are concerned with general, aggregate patterns; this makes it difficult to draw conclusions regarding how the individual behaviour of victims or offenders maybe be affecting the occurrence and rate of crime. Brantingham and Brantingham (1993, pg 6) sum this up as follows:

“Potentially, the most productive model in environmental criminology is one that places both the actual criminal events at a specific site, situation and time and the individual committing the crime while in a specific motivational state on (or in) an environmental backcloth, that may itself be mostly stable, regular and predictable or may instead be irregular, rapidly changing and unpredictable.”

The following section will discuss a number of modern computer modelling techniques that attempt to address some of these shortcomings.
3.4 Computer Modelling of Crime

Although most crime modelling utilises statistical techniques, Section 3.3 illustrated that this approach might not always be the most suitable. Advances in software engineering catalysed by increases in computer data storage and processing power has precipitated an uptake in computational approaches to the modelling of crime. This section will outline some more recently developed computer modelling techniques that might hold promise. To start, the meaning of the terms modelling and simulation will be clarified as they are important in order to discuss later modelling approaches.

3.4.1 Computer Modelling and Simulation

For our purposes, a model is defined as a “simplified representation of reality” (Castle and Crooks, 2006). Some systems are simple enough to use mathematical models to obtain analytical (exact) solutions to problems (Law and Kelton, 2000). Some of the models outlined in Section 3.3 are examples of these types of models; they usually use statistics whose behaviour can be proven mathematically. However, the complex interactions which govern real-world systems mean that analytical solutions are usually not possible. In these cases, models are studied by using simulation whereby a computer evaluates the model numerically and produces data in order to estimate the characteristics of the model (Law and Kelton, 2000). Axelrod (1997) describes simulation as a “third way of doing science”:

“Like deduction, [simulation] starts with a set of explicit assumptions. But unlike deduction, it does not prove theorems. Instead, a simulation generates data that can be analyzed inductively. Unlike typical induction, however, the simulated data comes from a rigorously specified set of rules rather than direct measurement of the real world” (Axelrod, 1997, page 24).

Simulation is one of many possible approaches to modelling systems. Figure 3.1 situates simulation among other approaches including direct experimentation, the construction of physical models or the use of analytical solutions. None of these approaches, however, are applicable to this project: it is not ethically acceptable to experiment directly because the experiment itself might lead to individuals becoming victims of crime; it is not possible to create a physical model of the system; and, as Section 3.2 discussed, the system is complex so cannot be analysed analytically. Therefore computer simulation methods are better suited to handling this complexity and potentially offer improvements over the techniques outlined in Section 3.3.

3.4.2 Spatial Interaction Modelling and Spatial Microsimulation

Two relatively modern approaches that come under the umbrella term “computer modelling” are spatial interaction modelling and microsimulation. They can address some of the criticisms of other approaches that have been mentioned and have occasionally been applied to crime.
Spatial interaction (SI) models are used to analyse flows between various origins and destinations. They are often known as “gravity models” because the flows are attracted to destinations in a similar manner to that of objects in space. The principle determinants of the spatial interaction are: the distance/time taken to travel from an origin to a destination; the attractiveness of the destination; any intervening opportunities; and demand for the destination (Birkin and Clarke, 1991). It is this trade off between attractiveness and accessibility which is at the heart of SI modelling (Fotheringham and O’Kelly, 1989). Generally models are very accurate, often within 10% of reality (Birkin et al., 2002). The flows usually represent people, i.e. migration, recreation, shopping, travel to work or school, hospital visits, buying petrol, etc. and they are most commonly used in the retail market (Fotheringham and O’Kelly, 1989). Traditionally, distance decay has been calculated by using the straight line distance between an origin and a destination but models now often incorporate road networks and concepts such as travel time and different types of trip (including multi-purpose trips and single “home-shops-home” trips) (Birkin et al., 2002). Kongmuang (2006) applied an SI model to crime in Leeds; this model will be discussed shortly.

SI modelling provides some benefits over more traditional statistical modelling. The spatial variation of parameter values is an inherent part of the model but one that is often missed in statistical models (with the exception of geographically weighted regression). Furthermore the addition of a road network can improve upon Euclidean distance measures that misrepresent the actual journeys taken by people travelling around an environment. Time can also be more easily included in a model, for example the attraction of objects can be varied during the course of a simulation. However, there are still some drawbacks as outlined in Table 3.1.

Another useful modelling technique is spatial microsimulation. The procedure is used to investigate the impacts of social and economic policies on individual micro units (Ballas and Clarke, 2001). The micro units in a simulation can include individuals, households, firms and organisations (Merz, 1991). They are identified by particular characteristics, for example the annual income of an individual or the number of children in a household. These characteristics are modified by the microsimulation model depending on the behaviour of the individual units and their rela-
Table 3.1: Drawbacks with spatial interaction modelling.

<table>
<thead>
<tr>
<th>Drawback</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal variation</td>
<td>Although time can be represented in a SI model, the model structure is inherently static. Therefore future predictions assume that the basic properties of the system do not change (Sayer, 1976). Clarke et al. (1998) suggest that dynamic SI models can be used and there are examples of hybrid models (Heppenstall et al., 2006) although these are scarce.</td>
</tr>
<tr>
<td>Complex environments</td>
<td>Environmental factors can be incorporated into the attraction term of destinations, although it is not possible to account for small-scale, detailed environmental components such as individual houses.</td>
</tr>
<tr>
<td>Data aggregation</td>
<td>Even if detailed data are available at the individual level (as is the case in this project) it must be aggregated to build attractors.</td>
</tr>
<tr>
<td>Human behaviour</td>
<td>As SI modelling is an aggregate technique it is not possible to incorporate heterogeneous behaviour in individuals.</td>
</tr>
</tbody>
</table>

tionships with the institutions in which they operate (Merz, 1991). The technique is particularly useful for combining and disaggregating spatial data sets in order to synthesise a population of individuals (Smith et al., 2009). Kongmuang et al. (2005), for example, used microsimulation to combine the British Crime Survey and the UK census (two aggregate data sets) to estimate crime victimisation at the household level. There are also a number of examples of the combined use of microsimulation and SI modelling (Ballas et al., 1999; Rephann et al., 2005; Kongmuang, 2006) and the technique can be used as a forecasting tool in its own right (Ballas and Clarke, 2001). It is particularly relevant to national policy analysis because it enables the investigator to examine the effects that national policies will have upon the local people of which their policies are aimed.

A study particularly relevant to this project is that of Kongmuang (2006) in which microsimulation and spatial interaction techniques are used to investigate the dynamics of crime in Leeds. Kongmuang used a microsimulation model to spatially disaggregate the British Crime Survey from the regional to the, much smaller-scale, ward and output area levels. This makes it possible to perform detailed analysis of crime at the local level and to make policy predictions. However, as the author notes, the model failed to capture some important spatial elements of crime that are closely related to modern environmental criminology theory. For example, the model could not account for the increases in crime that are often found when affluent areas are surrounded by deprived areas or when an area is in close proximity to the homes of many potential offenders. Although the approach was able to incorporate the importance of individual places, it was not advanced enough to link the places with the other effects of urban dynamics. To further investigate the movement of offenders, Kongmuang (2006) applied a SI model in order to explore the burglars’ travel patterns. The model showed that neighbourhood renewal in one area could have a significant, yet unforeseen, effect on surrounding areas.

Microsimulation has a number of advantages over other techniques. By definition, the approach is aimed at resolving data aggregation problems by disaggregating data from aggregate to individual-levels. Therefore it does not suffer the same aggregation problems associated with
SI models and statistical techniques. Although the literature here illustrates that microsimulation models have been usefully applied to a number of scenarios, there are still drawbacks to the methodology which makes it unsuitable for this project in isolation. Table 3.2 outlines these drawbacks.

<table>
<thead>
<tr>
<th>Drawback</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handling complexity</td>
<td>Although the technique operates at the micro level, it is the interactions between micro units that are important to determine behaviour in complex systems and this is not addressed by microsimulation.</td>
</tr>
<tr>
<td>Temporal aspects</td>
<td>As with SI techniques, microsimulation methodologies face problems with regards to the dynamic elements of the real systems; most models are designed to represent a single point in time and prediction is enacted by tracking the ramifications of altering otherwise static variables (Clarke et al., 1998).</td>
</tr>
<tr>
<td>Complex environments</td>
<td>Detailed environmental components such as quiet streets or new urban developments cannot be incorporated.</td>
</tr>
<tr>
<td>Human behaviour</td>
<td>Although microsimulation works at the micro-level it is not possible to incorporate behaviour or intelligence into the micro-units.</td>
</tr>
</tbody>
</table>

Spatial interaction and microsimulation approaches have offered improvements over traditional statistical techniques. However, as Tables 3.1 and 3.2 illustrate, there are still significant drawbacks which render these approaches inappropriate for this research. The following section will present agent-based modelling as a more appropriate methodology.

### 3.4.3 Agent-Based Modelling

“There was a time when I rather arrogantly believed I had read all the key papers in the multi-agent systems field, and had a basic working knowledge of all the main research problems and techniques. Well, if that was ever true, then it certainly isn’t any more, and hasn’t been for nearly two decades: the time has long since passed when any one individual could have a deep understanding of the entire multi-agent systems research area” (Wooldridge, 2009, page xix).

Bearing Wooldridge’s comments in mind, this review will not attempt to provide a full account of all agent-based modelling concepts, practices and applications. For this, the reader is directed to Wooldridge (2009) or Paredes and Hernández (2008). Instead it will provide a brief introduction to the methodology and focus on its application to crime in Section 3.5.

Described as a “breakthrough in computational modelling in the social sciences” (Gilbert and Terna, 2000, page 60) and “one of the most exciting practical developments in modeling since the invention of the relational database” (Macal and North, 2005, page 2), agent-based modelling (ABM) is a new method of modelling systems. An agent-based model is comprised of autonomous, decision making entities called agents who can interact with each other and their
3.4. COMPUTER MODELLING OF CRIME

environment (Bonabeau, 2002). As the model iterates, each agent has the ability to assess its circumstances, and based on a set of probabilistic rules, makes an informed/educated decision about its future course of action (Bonabeau, 2002). Through this mechanism, more realistic human behaviour can be incorporated (Moss and Edmonds, 2005) and models can be used to create systems which mimic real scenarios and produce a dynamic history of the system under investigation (Axtell, 2000). This approach is particularly relevant to social science studies because it is possible to build models in order to carry out experiments that would be impossible or unethical to perform otherwise (Gilbert, 2004).

There are numerous definitions of the term “agent” (O’Sullivan and Haklay, 2000). However, from the point of view of modelling social systems (and the crime system in particular), many authors note that the following are considered to be important characteristics of an agent:

- **Autonomy**: free to control its own state, interact with other agents and its environment and make decisions without direct control from some central source (Macal and North, 2005; Castle and Crooks, 2006). This seems to be an ideal mechanism for modelling people, including potential burglars.

- **Heterogeneity**: agents need not be identical and where groups of agents are formed they can be generated from the bottom-up (Castle and Crooks, 2006). Heterogeneous burglars can be created in this manner to reflect the variety of different techniques and behaviours that burglars have been found to display.

- **Reactivity**: agents are able to respond to changes in their environment and the response should be proactive, indicating goal-directed behaviour (Wooldridge and Jennings, 1995). This is particularly useful for a burglary model because the environment will change as a result of burglary (residents might become aware of the risk) and this in turn will influence the behaviour of potential burglars.

- **Bounded Rationality**: particularly with modelling in the social sciences, it is important that agents do not always act perfectly rationally as this is generally not considered to be an accurate human characteristic (Green and Shapiro, 1996; Axelrod, 1997). To circumvent this pitfall, agents can be programmed with “bounded” rationality by limiting their perception of their environment so that choices are not always perfectly optimal (Castle and Crooks, 2006). This is directly relevant to the rational choice perspective (Clarke and Cornish, 1985).

Agent-based modelling has been applied to a vast number of subject areas, including computer systems which assist car drivers (Miller et al., 2003), models of pedestrian movements (Castle, 2006; Turner and Penn, 2002), simulations of human immune systems (Jacob et al., 2004), models of insect behaviour (Parry et al., 2006), models of land use (Parker et al., 2003) and models of petrol station prices (Heppenstall et al., 2005, 2006). Some examples of the general uses of agent-based models will follow to provide examples of the methodological considerations that are important to this project. First cellular automata will be introduced because they are closely related to ABM and play a large part in many modern crime models.
Cellular Automata and ABM

Cellular automata (CA) have been employed in many crime studies and have influenced this research. A CA consists of a regular grid of cells which have certain properties that vary across the population. Rules determine how cells interact with each other and how one cell’s state affects another. Models are “executed” in a similar manner to ABM; at each iteration cells can inspect their neighbours and alter their state accordingly.

Although in theory they are different methodologies, in practice the boundary between cellular automata and agent-based modelling is indistinct. For example, Schelling (1969, 1971) created a well known model of racial segregation that has been described as both a cellular automata (Hegselmann and Flache, 1998) and also one of the first agent-based models in the social sciences (Vag, 2004; Castle and Crooks, 2006; Iannaccone and Makowsky, 2006). Fortunately a clean definition of the boundary between CA and ABM is unimportant so long as it is understood that many agent-based studies that are based on cellular grid environments could equally be classified as cellular automata.

Regardless of its classification, Schelling’s model is extremely interesting and goes a long way to illustrate how useful CA/ABM can be for modelling social systems. The model consists of a 2D environment populated by households that belong to one of two types of racial group. A global parameter dictates the percentage of the same group that each household wants to live next to. Households are able to move if they are unsatisfied (i.e. they live near to too many of the opposite racial type). The novel research finding was that, even with a relatively low preference for the same racial type, the environment can become highly segregated as illustrated by Figure 3.2. Methodologically this is interesting because Schelling showed that the dynamics of the segregation could not have been predicted by examining the rules of the individual cells in isolation; emergence was present and a simulation was required to illustrate this. Also from a practical position it provides an insight into the dynamics of human residential segregation, suggesting that extreme segregation can emerge from individuals who actually have low preferences for segregation (Batty, 2009).

It is common to build hybrid agent-based / cellular automata models, using agents to represent mobile individuals who interact in an environment that is itself represented by a CA. Examples abound in the literature such as the well known “sugarscape” simulation (Epstein and Axtell, 1996), which explores how a society of agents interact with each other and their environment to grow. Also using a grid-based environment, Iannaccone and Makowsky (2006) investigated the link between peoples’ personal religious preferences and those of the culture that surrounds them, Doran (1998) experimented with the effects of collective misbelief in agent societies and illustrated how misbeliefs can spread and Epstein (2002) outlined a model that can be used to investigate civil violence, finding that long-term drops in legitimacy are less likely to spark a civil reaction.

Although the studies cited use relatively simple rules and environments, they illustrate a point made by Axelrod (1997): simple models can provide important insights to some general process. This point has been echoed recently in the crime modelling literature (Elffers and van Baal, 2008).
as researchers are starting to use computer simulations to experiment with environmental criminology theory or with the dynamics of crime. However, care must be taken when applying the results of these very simple studies to real systems. Human systems are vastly more complex than the models outlined here and it is often very difficult to verify that human systems would respond as these models do. Schelling (1971), for example, indicates that his models are “too abstract and artificial” to present a picture of how real people interact in a city. Thus it is possible that the models outlined here might be improved if they were extended to include a more realistic environment (Benenson, 1998). The following discusses some agent-based models that have embraced more representative environments.

GIS, ABM and A Suitable Level of Complexity

A major objective of this research is to be able to make predictive analyses. To this end, it is essential that the environment is a realistic representation of the physical area under study. To include realistic environments, some studies have begun to incorporate Geographical Information Systems (GIS). These are computer systems designed to analyse spatial data. By integrating a GIS into a model it is possible to create an environment that is an accurate spatial representation of the real world (assuming, of course, that high quality spatial data are available). Gimblett (2002) discusses the use of GIS in agent-based modelling, noting that they can be a powerful tool for decision makers.

GIS are inherently static (Gimblett, 2002) so it is difficult to perceive how they can be adapted to suit the needs of dynamic models (Goodchild, 2005). Also, they are designed with only basic modelling capabilities (Maguire, 2005). Even if GIS were better suited to coping with temporal data, continuous data over a period of time rarely exists (Castle and Crooks, 2006). As the dif-
ferring views of Date et al. (2003) and Snodgrass et al. (1998) make evident, even the concept of storing temporal data is relatively new and lacks standardisation. Nevertheless, some studies have successfully linked models with GIS. The extent to which the two components can be considered integral is described as coupling. Loose coupling is a common approach and involves passing files or objects between the model and the GIS; see Clarke and Gaydos (1998) for an example. Such a system takes advantage of the analysis and visualisation of spatial data by the GIS, although the two components cannot communicate directly. Tight coupling, on the other hand, involves linking the model and the GIS directly. With this approach, the GIS is used to prepare input and display output and also control the spatial component of the model directly. Although a more complex approach, Brown et al. (2005) note that tight coupling can “reap benefits” in terms of efficiency (computation time) and functionality.

There is, however, a potential drawback of incorporating GIS into a model. When modelling complex systems, it is important to determine the suitable level of complexity of the model. Too much complexity can detract from understanding the dynamics of the component interactions that are at the heart of a complex system. A proponent of simpler models, Axelrod (1997) notes that with simple models the subtle effects of its hypothesised mechanisms are easier to understand or discover and that the complexity should be found in the results, not in the assumptions of the model. If the goal of a simulation is to attain a greater degree of understanding of some fundamental process, then it is the simplicity of the assumptions which is important, not the accuracy of the surrounding environment (Axelrod, 1997). This is a position also followed by Elffers and van Baal (2008) who note that including a complex environment makes crime models difficult to understand and detracts from experimenting with the pure dynamics of theory.

However, some researchers are moving away from the “keep it simple, stupid” (KISS) slogan. Edmonds and Moss (2005) note that simplicity should only be a goal of a model if it is supported by evidence and understanding of the model and the target process. They propose a new slogan, “keep it descriptive, stupid” (KIDS) that follows a different approach as is similar to that which will be taken by this research:

1. Start with a model which relates to the target phenomena in the most straightforward manner possible. These “descriptive” models allow a wide range of evidence (anecdotal as well as statistical) to be taken into account.

2. Develop the system with support from evidence and a greater degree of understanding of the model. This development could lead to model simplifications, if appropriate.

For this research a degree of environmental complexity is essential. If an abstract environment were used it would not be possible to make predictions of future crime rates in the real world. However, the trade off between practicality and descriptive adequacy is complicated (Edmonds and Moss, 2005) so the important variables that need to be included will be chosen carefully. Schmidt (2000) summarises the situation nicely.

“A model cannot be identical to the original system. The construction of a model always involves elements of idealisation and abstraction. It follows from these two
3.4. COMPUTER MODELLING OF CRIME

aspects, that the model never reproduces the whole colourful spectrum of the real
world but only a very small number of chosen state variables and only a very small
number of chosen modes of behaviour.” (Schmidt, 2000, pg 11).

3.4.4 A Critique of Agent-Based Modelling

This section will summarise many of the advantages that agent-based modelling holds over other
techniques for modelling social systems along with the drawbacks.

Capturing Emergent Phenomena

“A phenomenon is emergent when it can only be described and characterised using
terms and measurements that are inappropriate or impossible to apply to the compo-
nent units” (Gilbert, 2004, page 3).

By definition then, emergent phenomena (such as city-wide burglary rates) cannot be reduced
to the individual components of a system (Bonabeau, 2002). This makes it impossible to analyse
how emergent phenomena arise using top-down approaches. Also, aggregate differential equations
“smooth out” important individual interactions (Batty, 2005; Castle and Crooks, 2006). Because
agent-based models work at the micro level, individuals or groups (which often cause stable macro
patterns) can be examined explicitly.

Flexibility

Although agent-based models must be created to serve a specific purpose (Martinez-Miranda and
Aldea, 2005), they still offer considerable flexibility. The complexity of models can be controlled
by adjusting the rules that determine individual agent behaviour or by simply adding more agents
into the model (Bonabeau, 2002). Also there are many possibilities for altering the behaviour of
the agents, including the degree of rationality, types of interactions and the ability to learn (Castle
and Crooks, 2006). In addition, Axtell (2000) notes that it is trivial to create purely rational or
homogeneous agents depending on the requirements. They are also environmentally flexible: it is
easy to apply the same model to different environments (i.e. different cities) as chapters 8 and 9
will demonstrate.

Natural Description of a System

Perhaps the most significant advantage of ABM is the “natural” description of a system which it
provides. This is illustrated by examples:

- There are many systems that cannot sensibly be described by mathematical equations (Axt-
tell, 2000). Although mathematics provides a good basis for describing unexplained phe-
nomena in the natural sciences, this experience is not echoed in the social sciences (O’Sullivan,
CHAPTER 3. MODELLING BURGLARY

2004; Moss and Edmonds, 2005). Often simplified assumptions are required if mathematical models become too complex (Groff, 2006) and these assumptions are often implausible (Gilbert and Troitzsch, 1999).

- Social networks and physical environments are an essential part of a human social system. These are difficult to generate mathematically, but are relatively simple using ABM (Axtell, 2000).

- To understand geographic human systems it is necessary to understand the reasoning behind individual decisions (Crooks, 2006). Modelling these individuals directly is more natural than trying to build equations to control them (Bonabeau, 2002). Both Sayer (1976) and Lee (1973) were devastatingly critical of aggregate mathematical geographical models, although this is a critique that can potentially be addressed by models that sufficiently handle complexity (O’Sullivan, 2004).

- Individual cognitive models of a local area are as important as the physical characteristics of the area as offenders will only commit crimes within their routine activity spaces (Beavon et al., 1994). These cognitive models will be very difficult to incorporate into models which do not examine individuals directly.

- Axtell (2000) notes that many agent-based models have developed from a dissatisfaction with the rationality which cannot be avoided when using purely mathematical equations. It is easy to limit rationality in agent-based models (described as “bounded rationality”) because each agent is explicitly represented.

Not only does ABM offer a more natural description, but the ability of the methodology to represent individuals directly offers new opportunities which are not possible with macro-methodologies. This is particularly evident when modelling micro-theories such as routine activities theory or crime pattern theory. Brantingham and Brantingham (2004) note that with ABM it is possible to treat offenders in a similar manner to non-offenders and explore the effects that non-criminal activities will have on crime. In this manner, the “natural variety” of cities becomes part of the model, rather than smoothed out by aggregate methods (Brantingham and Brantingham, 2004).

Disadvantages

Although agent-based modelling (ABM) has been described by many authors cited here as the most appropriate approach to modelling social systems, there are nevertheless certain drawbacks which must be addressed.

Section 3.2 outlined some of the difficulties associated with modelling social systems. In particular the “soft factors” exhibited by humans, such as seemingly irrational behaviours and complex psychology (Bonabeau, 2002) can be problematic. These factors must be defined explicitly in models which work at the micro-level, leading to a strong commitment to minimal behavioural complexity (O’Sullivan and Haklay, 2000). However, Edmonds and Moss (2005) note that phrases such as “for the sake of simplicity” are not well-founded and simplicity should only be a target
if this is justified by the underlying system. As Section 3.2 suggested, human systems (such as the one to be modelled in this project) are far from simple. The conclusion to be drawn from the works of O’Sullivan and Haklay (2000) and Edmonds and Moss (2005), therefore, is that realistically modelling the behaviour of individual people is a daunting task and one that must be tackled explicitly using ABM. This task is further aggravated by difficulties in gathering appropriate data to understand complex, dynamic societies (Gilbert, 2004).

A product of complex systems is the concept of emergence. Although central to the use of ABM, it is important to notice that emergence is predominantly one-way (O’Sullivan and Haklay, 2000). Thus macro behaviours can emerge from the actions of individuals but not the other way round which leads to an individualist view on society (O’Sullivan and Haklay, 2000). Society and the individual should both have an effect on each other which must be expressed in the model.

There are also difficulties that relate to the implementation of the model. Because relatively small pieces of computer code can form integral parts of hundreds of agents, it is possible that small errors in the logic of the code can have huge effects on the outcome of the model. This is further compounded when attempting to ensure that work is repeatable because when programmers implement models, they unavoidably include many assumptions which are not documented (O’Sullivan and Haklay, 2000).

Although computational power is increasing at vast rates, agent-based modelling is still an extremely computationally expensive technique as it must deal with individual units (Bonabeau, 2002). The probabilistic nature of ABMs means that in order to assess the robustness of a result it is necessary to execute a model a number of times with varying parameters (Axelrod, 1997; Axtell, 2000). This places a further drain on already stretched computing power.

Although agent-based modelling appears the most appropriate technique for the system under investigation, this section has illustrated that there are many factors which must be addressed. To conclude the chapter, the following section will summarise the applications of ABM to crime.

### 3.5 Agent-Based Models of Crime

Around the start of the new millennium, early agent-based crime simulation work began to show promise. For example, Winoto (2003) developed a multi-agent model of human crime that assumes the model of rational choice theory (Becker, 1968; Clarke and Cornish, 1985), whereby criminals are driven purely by a rational assessment of the perceived losses and gains of their crime. Rational choice theory fits well with agent models because each agent can be programmed to maximise the expected gains from crime (Winoto, 2003). By finding optimal policies to govern the agent society, Winoto suggests that such policies could be applied to human systems. The author concludes that simply raising the probability of conviction (i.e. investing heavily in law enforcement) is not sufficient to significantly reduce crime rates and that rational choice theory can only really explain certain types of economically driven crimes (such as “white-collar crimes”).

Melo et al. (2005) investigated different strategies of the physical reorganisation of police patrol routes in an agent-based environment. Criminal agents were used who travelled to targets and decided whether or not to commit a crime there, being apprehended if a police office was present.
The model makes some very general assumptions, for example it is assumed that criminal agents simply choose a target and then travel there. Modern theories (such as routine activity theory and crime pattern theory) as well as qualitative research (such as Bennett and Wright (1984) and Wright and Decker (1996)) suggest that this is not an accurate representation of human behaviour. However, the simple model was still able to demonstrate that regularly re-organising police routes lead to a greater number of arrests which provides obvious advice for on police force managers.

Gunderson and Brown (2000) present a multi-agent methodology for predicting both physical and cyber crime. Using overlaid opportunity, guardianship and deterrence layers the authors created a dynamic environment in which agents can interact. Although the framework appears promising it has not been tested in a practical application. Brantingham et al. (2005a,b) cite the work of Gunderson and Brown (2000) as a “novel research direction” (Brantingham et al., 2005c). Also developing a model of crime, they used an abstract state machine (ASM) to provide a precise mathematical foundation to an agent-based model. Agents and other environmental objects are mapped to static and dynamic functions in the ASM. In this manner they are able to create a dynamic environment in which agents can move through time and space interacting with each other and the environment. Agents are able to learn, and they have “preferences” which translate to forms of behaviour. A “person” agent is itself an ASM which uses various modules to control its behaviour and perception of the environment. Along with experimenting with criminology theory, the resulting simulation can be used as an interdisciplinary tool to assist criminologists in investigating the dynamics of urban crime (Brantingham and Brantingham, 2004; Brantingham et al., 2005a,b, 2008). In a similar vein, Wang et al. (2008) outlined a tool to study the interactions between actors involved in a crime event.

A study of direct relevance to this research, because it focusses on burglary in Leeds, is that of Malleson (2006) and Malleson et al. (2009). The authors developed a simple agent-based model which was used to predict rates of residential burglary using administratively-defined census boundaries. Although the model was able to demonstrate utility by indicating the importance of using a high degree of spatial accuracy when investigating crime or experimenting with criminology theory, the authors note a number of areas in which the model could be improved. These include allowing agents to travel freely around their environment and incorporating a more realistic model of human behaviour; both of these areas will be addressed by the model produced for this research.

Although these types of applications can provide insights into theory, they are not sufficiently detailed to allow for predictions of actual crime rates. Recently, more advanced models have begun to emerge. Notable sources are the recent book entitled “Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems” (Liu and Eck, 2008) and a special issue of the Journal of Experimental Criminology (Groff and Mazerolle, 2008).

An example of a more advanced approach, Dray et al. (2008a) used agent-based modelling to explore drug market dynamics in Melbourne. The authors examined the relationship between local actors (such as police, addicts, dealers and outreach workers) and macro effects (such as national drug supply and governmental drug policies). The environment is abstract; environmental objects are “spatial metaphors” rather than actual places but the authors justify this decision by noting...
3.5. AGENT-BASED MODELS OF CRIME

that complexity should be increased gradually as recommended by Liu and Eck (2008). Most findings are unsurprising; e.g. increasing the effectiveness of outreach workers leads to more users in treatment and targeting hotspot areas rather than random patrols leads to more dealer arrests. However, Dray et al. (2008b) showed that micro-level interventions are more successful at disrupting drug markets than macro-level efforts which demonstrates the potential of these types of individual-level models.

Extending earlier work (Liang, 2001; Liang et al., 2001), Liu et al. (2005) present an early example of an agent-based/cellular-automata model of street robbery applied to a real area (Cincinnati). The model contains three types of agents: places; targets; and offenders. Places are static and represent the locations that crimes occur in. Offender agents are responsible for robbing targets, based on the security of the place, their own level of motivation (which increases after a successful crime) and the desirability of the targets. The offenders’ locations are chosen randomly and are based on the inverse distance from their homes. The targets also move randomly and are more likely to be placed on the most accessible streets. The model appears to re-create crime rates in Cincinnati relatively well, although the authors do not apply statistical analysis to verify this. The main drawback with the model is that it is based on routine activity theory but does not attempt to reflect realistic agent movements between anchor points. Although the model is an excellent example of the potential of agent-based modelling in crime, it requires substantial improvements to make it more realistic and to more fully realise routine activity theory.

Research is also underway that uses ABM to simulate burglary and improves upon a number of the limitations of Liu et al. (2005)’s model. Birks (2005, 2007) is developing a general simulation that incorporates criminology theory to provide a bottom-up description of crime dynamics and has applied the model to burglary specifically (Birks et al., 2008). In the model, offenders travel around an environment and a burglary occurs if they come into contact with a suitable target and the potential rewards outweigh the probable costs. The environment consists of a transport network and a regular grid is used to represent different types of targets (e.g. groups of commercial buildings or residential houses). Agent behaviour is driven by the Behaviour-Based Artificial Intelligence (BBAI) framework which will be discussed in greater detail in Section 5.7.2. Agents have their own routine, an awareness space which they build up as they travel around the environment and criminal propensity which dictates the rate that they will offend. Routines are included in the model through “routine activity nodes” which must be visited by particular offenders at particular times (e.g. simulating the requirements of being at work between 9am and 5pm). In its current published form, real road network data is used to create the transport network but all other environmental and agent parameters are given random values. The only model evaluation related to repeat victimisation; the authors utilised a metric that measures spatio-temporal clustering and note that this is similar to real-world crime data (although statistical evidence for this is not provided).

Although the work of Birks et al. (2008) represents a much more advanced attempt at integrating important individual-level factors from the criminology literature, there are a number of areas on which it can be improved. For example, the behaviour of the agents is static in the sense that they cannot change their behaviour based on dynamic changes in circumstance. Therefore
the agents cannot decide not to work between 9am and 5pm if it does not suit their needs on a particular day. Also, there are a number of important factors that have not been accounted for that will affect burglary such as differing household occupancy habits depending on the community, public transport and community cohesion. However, the research is more focussed on a general crime model so it is possible that these attributes could be included if the focus was more specific.

Also researching burglary, Hayslett-McCall et al. (2008) built an ABM/CA model based on social disorganisation and routine activity theories. The model contained immobile offender agents with variables that relate to their motivation (such as age and gender) and houses with variables that relate to their desirability (such as income). Offender characteristics are generated from police data on known offenders and offender motivation decays exponentially with the distance from their home. To determine household attractiveness to each offender, a likelihood surface is generated based on the offender’s motivation, the target desirability and presence/lack of guardianship. Many model features relate closely to this research. For example, the homes of offender agents are determined from crime data and household properties are assigned values based on the socio-economic status of the area (although, unlike in this research, houses do not have heterogeneous levels of desirability/vulnerability). Also similar is that target affluence is relative to the race and income of each agent and agents prefer areas that are similar to their own. One way in which this research can improve upon the model, however, is by including a comprehensive cognitive framework such as that implemented by Birks et al. (2008). In this manner agent behaviour can be made more dynamic (it can change to reflect unforeseen external or internal changes) and a greater number of behavioural characteristics could be included. The main drawback with what is otherwise a comprehensive model, however, relates to agent movements. Agents do not travel around the environment, they simply decide where to commit a crime based on the Euclidean distance of the target from their home (as well as considering the environmental factors already mentioned). Therefore it is not possible to simulate “commuting” offenders (see Section 2.4.4) who travel some distance from their home to another area which they nevertheless know well. This relates closely to the formation of awareness spaces which the model does not include.

Although not applied to burglary, Groff (2006, 2007a,b) and Groff and Mazerolle (2008) present a particularly relevant model because it incorporates a realistic street network and mobile intelligent agents. Groff (2007a) noted that due to a lack of dynamic, individual-level interactions, previous studies had failed to effectively test routine activity theory. To avoid these pitfalls, the authors constructed an ABM tightly coupled to a GIS in order to test the applicability of routine activity theory to street robbery. The model was based on the city of Seattle, Washington and used a graph-based environment to realistically represent the city’s road network. Two types of agent were created: citizens (offenders, victims and guardians) and police. The civilian agents were randomly assigned a particular home location and, in the model’s most advanced form, the civilian agents spent time away from home by visiting randomly assigned work and activity nodes following pre-defined routes. The offender’s decision to offend is stochastic and based on levels of guardianship and the wealth of the potential target at their current location. Groff’s model found that the number of street robberies increased with the amount of time spent away from home because citizens had the chance to meet more potential offenders. Interestingly, some street inter-
sections exhibited significant clusters of events even though the travel patterns of the agents were random (Groff, 2007a).

This research can improve upon Groff’s excellent model by enhancing the behaviour of the offenders and the realism of the environment. With respect to the burglars, global rules dictate how much time agents should spend away from home and they cannot create routes dynamically (Groff, 2007b). This is similar to the assumptions made by Birks et al. (2008) and prohibits the inclusion of variable/dynamic travel patterns such as using public transport or altering routes depending on internal or external conditions. As Chapter 5 will demonstrate, in the model produced for this research the need to travel is always decided upon dynamically (based on satisfying a greater goal) and routes are calculated dynamically based on the transport available to the agent and on other factors. In Groff’s model, agents are not controlled by individual cognitive models so the richness of their behaviour is limited. For example, agents cannot engage in goal directed behaviour to satisfy some internal need or change their behaviour depending on the influence of the environment (e.g. having to travel elsewhere because a usual drug dealer is not present). This model will improve upon this situation by including a comprehensive cognitive framework that allows much more rich, realistic human behaviour in a similar manner to Birks et al. (2008). The environment can also be improved to bring it more in line with current criminology theory in a similar manner to that of Hayslett-McCall et al. (2008). For example, no account is made for the effects of community guardianship or the extent to which a potential offender feels out-of-place in a particular area. The virtual environment used in this research will include a socio-economic classification of the type of the environment (so that different types of agents will feel more or less comfortable in different types of environments), a measure of community cohesion and estimations of occupancy levels and affluence. Also, each agent’s cognitive map recalls the number of times they have visited an area so it is known which areas the agent is most familiar with. By incorporating all of these features, principles from many crime theories can be incorporated and the environment to be portrayed realistically. Chapter 5 will discuss how this is achieved.

3.6 Summary – Modelling Burglary

Burglary is characterised by a large number of diverse components, a magnitude of interactions between components, emergent properties that cannot be attributed to any of the rules that drive the individual components and system-wide behaviour that is generally extremely difficult to predict. Although modelling complex systems seems daunting, another lesson we learn from complexity theory is that even the most complex systems can exhibit regular patterns built up from the interactions of relatively simple components that behave in an understandable, predictable way. For example, in transportation it is possible to predict complex city-wide patterns (such as rush-hour traffic rates) from the interactions of relatively simple “car drivers”. This also holds for crime systems; if we are able to model the behaviour of individual offenders, victims and guardians in enough detail so that they perform normal, realistic daily tasks, then it might be possible to model their convergence in space/time and then accurately predict crime patterns and crime rates.

On this premise, the chapter has reviewed different methods that can be used to model complex
systems, and the burglary system in particular. Following an introduction to modelling complex social systems in Section 3.2, Section 3.3 reviewed the “traditional” methods of modelling crime that follow the pioneering work of Guerry (1831) and Quetelet (1842). These use statistical techniques to compare crime rates to other variables, typically at aggregate levels. It was shown that these methods have a number of drawbacks; in particular they fail to capture the complex properties of the crime system or account for the importance of individual people / objects situated in a complex urban backcloth. To follow, Section 3.4 outlined some computer modelling techniques that potentially offer improvements. In particular, it was shown that agent-based modelling holds the most potential for successfully modelling crime. Some agent-based crime studies were outlined in Section 3.5 and it was shown that this research will be able to improve upon published techniques by more accurately representing human behaviour and the urban environment. The information discussed in this chapter will be used to build the model as outlined in Chapters 5 and 6.
Chapter 4

Data Analysis: Crime and the Environment

Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>The EASEL Regeneration Scheme</td>
<td>50</td>
</tr>
<tr>
<td>4.3</td>
<td>Geographical Data</td>
<td>51</td>
</tr>
<tr>
<td>4.4</td>
<td>The Victims of Crime</td>
<td>57</td>
</tr>
<tr>
<td>4.5</td>
<td>The Offenders</td>
<td>74</td>
</tr>
<tr>
<td>4.6</td>
<td>Regression Models of Burglary</td>
<td>79</td>
</tr>
<tr>
<td>4.7</td>
<td>Data Analysis – Conclusion</td>
<td>83</td>
</tr>
</tbody>
</table>

4.1 Introduction

Understanding the main processes and dynamics that govern the burglary system is a central precursor to modelling it. Chapter 2 reviewed the criminology literature and discussed which environmental factors were the most important determinants of individual crime events and which behavioural influences characterise the behaviour of burglars. This chapter will extend our understanding of the system through a thorough analysis of the available data, in search of new insights as well as evidence in support of the findings from the literature. Also, the available geographical data will be discussed as this will determine the level of detail possible for modelling.

As the burglary system is inherently geographical, Section 4.3 will begin with an outline of the available geographic data that will be used throughout the research. The study area is Leeds, a large UK city in the north of England. Then Section 4.4 will explore the locations of recorded burglaries to enhance our understanding of the temporal/spatial dynamics of burglary occurrences and identify which variables need to be included in a model. The time period for the analysis is 31st March 2000 – 1st April 2002. Following this line of enquiry, Section 4.5 will explore the available offender data to better understand how burglars behave and how this behaviour might be implemented in a model. To conclude, Section 4.6 will build a “traditional” crime model using
regression in order to provide a benchmark accuracy measure to which the agent-based model can be compared. As the model will be used to simulate the effects of a regeneration scheme in Leeds, called EASEL, this scheme will be briefly introduced in the following section.

4.2 The EASEL Regeneration Scheme

Parts of the south-east of Leeds, UK, contain some of the most deprived neighbourhoods in the country. This is illustrated in Section 4.3.4 as Figure 4.2 maps levels of deprivation in the city. Although there are large areas of green space and public parks, the area has extremely high rates of economic inactivity and low educational attainment. To address these problems, Leeds City Council has instigated an ambitious urban renewal scheme valued at approximately £1 billion that covers an area of 1700 hectares which is home to 79,000 people living in 36,500 households (EASEL Team, 2007). The scheme aims to sell council land to private developers who will develop different types of houses to attract buyers from outside the area and include social housing for local residents. Money from the house sales should be re-invested by the council in the area. Figure 4.1 illustrates the typical types of housing in the area and the new buildings that are being built. As Beavon et al. (1994) note: “cities create the backdrop for crime through their control of roads, commercial development, housing, building costs and transportation networks.” Therefore the EASEL regeneration scheme, which addresses all of those factors, is an ideal candidate to test the forecasting potential of the model to predict the affects of these types of changes on burglary rates.

![Image](a) A new EASEL home  ![Image](b) Typical EASEL urban landscape

Figure 4.1: Photographs illustrating the types of houses in the EASEL area.

The project has been organised into a number of distinct stages that will be underway at different times. At present, work has begun in two of the EASEL areas called Gipton and Seacroft. Large areas of land have been demolished and houses are already under construction. Chapter 8 will explore the predicted effects of the EASEL regeneration scheme on burglary rates and describe the current building work in more detail. For more information about the scheme in general, the reader is directed to the Area Action Plan (EASEL Team, 2007) or internet resources (Leeds City Council, 2009).
4.3 Geographical Data

The model produced by this research will simulate at the level of the individual house. Therefore highly detailed geographical data are required to build an accurate picture of the environment. This section will discuss the available data sources that will be used throughout the remainder of the research, including the crime data in Section 4.3.6.

It must be noted that the word “house” can have a number of different meanings. A household, for example, refers to a group of people who live together and can move to a new building whilst still remaining a single unit (e.g. a family). A dwelling refers to a place in which people live and can mean an entire building or a part of it, i.e. a self-contained living area within a larger building. For this research, however, a house is defined as a property which can contain one or more separate living areas but is not a block of flats (this distinction will be discussed in Section 4.4.4). Therefore, houses can be identified from the geographical boundaries of buildings.

4.3.1 UK Administrative Boundaries

Boundary datasets define geographical areas and are essential for mapping data that are not released as individual points. For this research, these data include the UK census and government measures of deprivation. Table 4.1 illustrates the common administrative boundaries used in England (sizes vary in Wales, Scotland and Northern Ireland). The boundaries share borders so that the smaller areas can be aggregated into larger ones. Leeds consists of 33 wards, 108 MSOAs, 477 SOAs and 2440 OAs.

<table>
<thead>
<tr>
<th>Boundary</th>
<th>Minimum Size</th>
<th>Recommended or Mean Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Households</td>
<td>Persons</td>
</tr>
<tr>
<td>Output Area (OA)</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>Lower Super Output Area (SOA)</td>
<td>-</td>
<td>1000</td>
</tr>
<tr>
<td>Medium Super Output Area (MSOA)</td>
<td>-</td>
<td>5000</td>
</tr>
<tr>
<td>Ward</td>
<td>400</td>
<td>-</td>
</tr>
</tbody>
</table>

4.3.2 The 2001 Census of Population

Every decade in the UK since 1801 (with the exception of 1941), a census of population has been conducted and collated by the Office for National Statistics (ONS) in England and Wales. Similar censuses are also administered by the General Register Office in Scotland and the Northern Ireland Statistics and Research Agency but these are not relevant in this research as it will focus exclusively on Leeds. The most recent census was conducted on 29th April 2001 and consisted of 36 individual and 10 household questions (Rees et al., 2002a). The attributes collected by the census form a comprehensive view of the population, comprising numerous demographic, social and economic characteristics. The data are geographically referenced and the finest area at which
they are released is the output area (see Table 4.1). For more information about the census in general, the interested reader is directed to Rees et al. (2002b).

There are a huge number of census variables that could potentially be useful for this research. Fortunately, this search can be restricted by the use of variables selected by other research. Relevant sources include Vickers and Rees (2006, 2007) who built a national classification based on census data, Shepherd (2006) who build a classification for Leeds neighbourhoods in a crime context and Kongmuang (2006) who also explored the relationship between census variables and crime levels in Leeds. The work of Vickers and Rees (2006, 2007) includes the largest number of census variables; every census variable analysed by Kongmuang (2006) and Shepherd (2006) are included in the classification directly or in a similar form. Therefore, this research will explore the relationship between crime and the 41 variables chosen by Vickers and Rees’ classification which is the subject of the next section.

4.3.3 The Output Area Classification (OAC)

The Output Area Classification (OAC) was produced jointly by the Office for National Statistics (ONS) and researchers at the University of Leeds (Vickers and Rees, 2006, 2007). It is a classification scheme based on the 2001 UK census and produced at the output area boundary level. Unlike other classification schemes such as Mosaic (Experian, 2007), the methodology is well documented and all the data used in the classification are available from the census. This is important because it is possible that the classification might need to be re-calculated in order to use the model to make future predictions. This is only possible if the methodology is published in detail and all data are available.

The classification is the result of an extensive geographical k-means cluster analysis that identified 41 important variables. These variables were chosen because they were the most successful at creating distinct clusters of people. The OAC partitions each output area into one of seven super groups, which are then broken down into a number of groups and sub groups. Appendix A provides a description of each variable used in the classification, the classification group names and a map of the supergroups in output areas in Leeds. The city centre largely consists of the “city living” type with sections of relatively deprived “constrained by circumstances” communities between the centre and the more prosperous “prospering suburbs” and “countryside” communities on the outskirts of the city. Later sections will discuss, in more detail, how the OAC is to be used and how it compares to crime statistics.

4.3.4 The Indices of Deprivation

Chapter 2 made reference to the strong link between crime and deprivation, so a specific measure of deprivation will be useful for a crime model. The English Indices of Deprivation have been developed by the government as a tool for identifying the most deprived areas nationally. Unlike the census, deprivation data are collected annually. There are seven distinct indices, but these are also combined into a single Index of Multiple Deprivation (IMD) which summarises the overall deprivation of each lower super output area (SOA) in England. The IMD scores for Leeds SOAs
are illustrated in Figure 4.2, where high values indicate high deprivation. It is clear that areas surrounding the city-centre are considerably more deprived than most other SOAs in Leeds. These appear to correspond to the “constrained by circumstances” communities which is to be expected and supported by other research (Vickers, 2006).

![The Index of Multiple Deprivation (IMD)](image)

Figure 4.2: The Index of Multiple Deprivation in Leeds.

It must be recognised that although the output area classification is based on the 2001 census, the deprivation data used here were collected in 2004. This year was chosen because prior to this time the IMD data were released at the ward level (Vickers, 2006) which is a considerably larger geography, hiding much underlying geographical variation. Therefore the drawback of using data sources that differ by three years is offset by the considerable additional spatial accuracy offered by the SOA geography. Mixing data time periods is not unique to this study; it is an often unavoidable necessity in any crime research (Brantingham et al., 2009).

### 4.3.5 Ordnance Survey MasterMap

Along with administrative boundaries for demographic data, detailed geographical data are required to model roads and buildings. MasterMap (Ordnance Survey, 2009) is a mapping product
produced by the Ordnance Survey (OS) and is ideal for these purposes. It is a highly accurate and detailed dataset that contains information on over 450 million geographic features in the UK. The product is divided into separate layers, including:

- The **Topographic Layer** which contains land area classifications such as buildings, railways, roads and water features. As Figure 4.3 illustrates, the data are highly detailed and will support the inclusion of a comprehensive virtual environment in a model;
- The **Transport Layer** which is a detailed network of transport paths including minor roads, motorways and alleyways along with information about width restrictions and one-way roads;
- The **Address Layer** which contains approximately 29 million postal features (such as residential or commercial properties derived from Royal Mail’s Postal Address File) and non-postal miscellaneous premises (including churches and car parks) (MasterMap, 2006). The address layer is particularly important because, using the National Land Use Database code (Harrison, 2006), it is possible to differentiate between the different types of building that can subsequently be simulated in a model. Relevant land use codes are provided in Table 4.2.

<table>
<thead>
<tr>
<th>Order</th>
<th>Group</th>
<th>Group Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>U040</td>
<td>Recreation and Leisure</td>
<td>U041 Outdoor amenity and open spaces</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U042 Amusement and show places</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U043 Libraries, museums and galleries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U044 Sports facilities and grounds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U045 Holiday parks and camps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U046 Allotments and city farms</td>
</tr>
<tr>
<td>U050</td>
<td>Transport</td>
<td>U051 Transport tracks and ways</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U052 Transport terminals and interchanges</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U053 Car parks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U054 Vehicle storage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U055 Goods and freight terminals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U056 Waterways</td>
</tr>
<tr>
<td>U070</td>
<td>Residential</td>
<td>U071 Dwellings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U072 Hotels, boarding and guest houses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U073 Residential institutions</td>
</tr>
<tr>
<td>U080</td>
<td>Community and Services</td>
<td>U081 Medical and health care services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U082 Places of worship</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U083 Education</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U084 Community services</td>
</tr>
<tr>
<td>U090</td>
<td>Retail</td>
<td>U091 Shops</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U092 Financial and professional services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U093 Restaurants and cafes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U094 Public houses and bars</td>
</tr>
<tr>
<td>U100</td>
<td>Industry and Business</td>
<td>U101 Manufacturing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U102 Offices</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U102 Storage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U103 Wholesale distribution</td>
</tr>
</tbody>
</table>

MasterMap data are comprehensive; in Leeds alone there are 351,892 address points and 2,234,189 topographic area features. The data are also extremely accurate. For example, of the 317,493 address points that are registered as residential properties (code U071), only 3042 (0.1%)
4.3. GEOGRAPHICAL DATA

Figure 4.3: An example of the OS MasterMap Topography layer. The data are highly detailed and include many features that could be useful for a crime model such as sports stadiums (which will affect peoples’ awareness spaces) and individual gardens (which could help or hinder a burglar’s access to the property).

are not within topographic areas that are registered as type “building”. Section 5.4 will detail some other inaccuracies that are found as the data are being prepared to input to the model, but on the whole the data are ready for use without any required cleaning.

4.3.6 Crime Data

Recorded Crime

In addition to geographic and demographic data, essential crime data collected by West Yorkshire Police data have been provided by Safer Leeds (Safer Leeds, 2009). The main dataset consists of all the crimes that have been reported to the police for the period 1st April 2000 – 31st March 2004 and can be linked with an offender dataset which is discussed in the following section. Recorded crime data are also available for more recent years (1st April 2006 – 31st March 2008) but these data have no associated offender information. More information about the data (such as a breakdown of the individual table attributes) can be found in Kongmuang (2006).

There are some serious implications for using recorded crime data, particularly the extent
that the data are an accurate reflection of real crime. An obvious drawback with recorded crime data is that not all crime that takes place is actually reported to the police. A common means of determining the extent of under-reporting/recording is to use the British Crime Survey (BCS). The BCS is a large social survey conducted annually that surveys peoples’ personal experiences of crime and, in particular, asks them if they reported crime to the police. In 2004 it was estimated that only 42% of crime was reported to the police. However, the reporting rate is very different depending on the crime type: although only 30% of common assaults were reported (the lowest of all crime types), approximately 80% of burglaries were reported if something was stolen (Dodd et al., 2004) due to the requirements of insurance claims (Brantingham and Brantingham, 1998; Shepherd, 2006).

Even if it is reported, however, it is not guaranteed that the crime will actually be recorded by the police. To standardise recording practices across the country, the National Crime Recording Standard (NCRS) was introduced in 2002. The NCRS promoted a more “victim centred” approach, meaning that a crime should be recorded based on the victim’s perception of the crime, rather than on an evidential basis (Simmons et al., 2003). This led to an increase in many types of crime which also varied by police force (Shepherd, 2006). After the introduction of the NCRS it is estimated that the police record approximately 70% of crime that is reported to them (Chainey and Ratcliffe, 2005). Taking all this information into account, Dodd et al. (2004) estimates that only 30% of all crime is actually represented by recorded crime data. Fortunately this figure will likely be higher for burglary because it is more commonly reported. Another encouraging factor is that unreported crime is expected to cluster in the same places as reported crime (Chainey and Ratcliffe, 2005), thereby effectively increasing the association between reported crime figures and actual amounts of crime.

Another potential problem with crime data relates to the spatial accuracy of the recorded crime. There are two potential sources of error: the actual address recorded and the way in which spatial coordinates are generated from the address (geocoding). The address itself is usually manually entered so there is potential for human error, such as recording street and house information correctly but using the wrong postcode. Shepherd (2006) notes that this type of error was prevalent in the Leeds crime dataset but thanks to a data audit that was instigated by Safer Leeds and performed by Shepherd et al. (2004), the accuracy of the recorded addresses was greatly increased. The other potential source of spatial error relates to geocoding. Geocoding is the process of calculating spatial coordinates from other information (addresses in this case) and is usually conducted automatically by computer software. Accurate geocoding is particularly important for this research because it is being conducted at the individual level, rather than at an aggregate level where exact address matches might be less important. Geocoding with 100% accuracy is usually not possible – again due to address errors – but Ratcliffe (2004) suggests that 85% is an acceptable geocoding rate that will not lead to distortions in the data. The accuracy of the provided recorded crime data set is unknown, but due to the cleaning conducted by Shepherd et al. (2004), it can be assumed to be sufficiently high (Kongmuang, 2006).
4.4. THE VICTIMS OF CRIME

Nominal (Offender) Dataset

Alongside recorded crime data, information about the people who have been somehow involved in the crime is also available for the same time period. The dataset holds information about people who were suspected or otherwise involved in the crime, but not necessarily convicted (termed a “nominal”). Clearly, therefore, the dataset will contain people who were not actually involved in the crime at all (Shepherd, 2006). Nonetheless, the dataset is the best empirical evidence for offending activities available. Again, more information about the attributes of data can be found in Kongmuang (2006).

Nominal information includes the person’s age, gender and ethnicity as well as their home postcode. Through a crime number, each nominal in the dataset can be linked to the associated entry in the recorded crime data. Also, each nominal has a unique ID so their behaviour can be tracked through the data. There is a many-to-many relationship between the two datasets: many nominals might be involved in a single crime and a nominal might be involved in many crimes.

Unlike the recorded crime data which has been geocoded to individual houses, the nominal data contains only the postcode so cannot be geocoded more accurately than this. Fortunately this still enables research to investigate aggregate flows from one neighbourhood to another, just not between individual houses. It must be noted that the home address is the place where the nominal was recorded to be staying at the time and might be an institution (such as a prison) or a temporary address (such as a friend’s house) so might not be a good representation of their “home” (if indeed such a place exists for a population who often live extremely chaotic lives).

4.4 The Victims of Crime

This section will explore burglary patterns in Leeds and assess the extent to which the available data support the crime literature discussed in Chapter 2. As Section 4.3.6 noted, information about victims and nominals (people who are potentially linked to a crime) is available for the period 1st April 2000 – 31st March 2004. A more recent crime data set is also available for 2006 – 2008 but these data have no associated nominal information and are less rich, as illustrated by Table 4.3.

Table 4.3: The crime data attributes that are available for different years (all periods are 1st April – 31st March).

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All crime types</td>
<td>Only burglaries (no other crimes)</td>
</tr>
<tr>
<td>House address and ((x, y)) coordinates</td>
<td>Only ((x, y)) coordinates</td>
</tr>
<tr>
<td>Date/time of offence</td>
<td>Date/time of offence</td>
</tr>
<tr>
<td>Victim information: age, gender, ethnicity</td>
<td>No victim information</td>
</tr>
<tr>
<td>Nominal (offender) data</td>
<td>No associated nominal data</td>
</tr>
</tbody>
</table>

Although the crime data have been cleaned extensively (Shepherd et al., 2004), some crimes (approximately 1%) actually occurred outside the Leeds area (this can happen if a crime occurred
outside Leeds but was reported inside) and were removed. Table 4.4 summarises the number of crimes in each dataset for every available year (after removing the outside-Leeds crimes). The most striking characteristic is that the number of burglaries drops considerably from a high of 16,350 in the period 2002/03 to less than 8,500 in 2006/07 and 2007/08. This pattern corresponds well with both British Crime Survey (BCS) results and national reported crime levels which have both been showing substantial falls in crime and burglary from 2001 to 2008 (Simmons, 2002; Simmons and Dodd, 2003; Dodd et al., 2004; Nicholas et al., 2005; Walker et al., 2006; Nicholas et al., 2007; Kershaw et al., 2007), although the decline started in Leeds some years after the national trend which began in approximately 1997.

Table 4.4: The number of recorded crimes and burglaries in Leeds. (Recall that Table 4.3 made note that only burglary data are available for the years 2006 – 2008).

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Number of Crimes</th>
<th>Number of Burglaries</th>
<th>Percentage of Burglaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000 - 2001</td>
<td>104813</td>
<td>13594</td>
<td>13.0</td>
</tr>
<tr>
<td>2001 - 2002</td>
<td>119909</td>
<td>15676</td>
<td>13.1</td>
</tr>
<tr>
<td>2002 - 2003</td>
<td>128484</td>
<td>16350</td>
<td>12.7</td>
</tr>
<tr>
<td>2003 - 2004</td>
<td>126407</td>
<td>13793</td>
<td>10.9</td>
</tr>
<tr>
<td>2006 - 2007</td>
<td>-</td>
<td>8302</td>
<td>-</td>
</tr>
<tr>
<td>2007 - 2008</td>
<td>-</td>
<td>8417</td>
<td>-</td>
</tr>
</tbody>
</table>

Before starting to analyse the crime data in detail the question must be asked: which years of crime data should be analysed? It is possible to explore crime for every available year in less detail – an approach employed by Kongmuang (2006) – but this would not reveal as much information about local- or individual-level crime patterns that are particularly relevant for this project. Data around April 2001 appear to be the most appropriate for two reasons: they correspond to the time of the UK census, therefore representing the point in time that will be modelled, and also have associated nominal information. However, the later data will be a better representation of current crime patterns. To compare yearly crime patterns spatially, Figure 4.4 maps the burglary densities using the Kernel Density Estimation (KDE) algorithm (a routine which is discussed in more detail in Section 7.2.1). Each map has the same thematic range so the reduction in the absolute number of burglaries is apparent. However, even though the number of crimes has dropped substantially, the distributions are very similar from one year to the next. It appears, from the perspective of these global maps, that burglary patterns do not differ substantially year-on-year. Therefore the benefit of having associated nominal information and of being collected at the same time as the census outweigh the risks of being out-of-date, and crime data covering the period around April 2001 will therefore be used. To limit small-number problems that might occur when crime is explored at a fine level of detail in subsequent chapters, data covering the period 1st April 2000 – 31st March 2002 will be used rather than using a single year. The census (April 2001) falls in the middle of this dataset.
Figure 4.4: Density of burglaries from recorded crime data. Each map shares the same thematic legend. Generated using the KDE algorithm (1km radius and 100m cell size).
4.4.1 City-wide Burglary Patterns

This section will illustrate the city-wide burglary patterns across Leeds, comparing them to other data sources and in relation to the literature discussed in Chapter 2. Although crime data are available as individual points, some non-crime data sources (such as the census and the indices of deprivation) are released using administrative boundaries such as the output area (OA) and super output area (SOA). Therefore crime must also occasionally be aggregated to these levels (to search for correlated variables for example). Figure 4.5 illustrates the crime counts when aggregated up to the OA and SOA boundaries along with the KDE densities. The OA-aggregated data do not appear to capture crime rates well when compared to the KDE map which, as will be discussed in detail in Section 7.2.1, is the preferred method of visualising crime rates (Chainey and Ratcliffe, 2005). Therefore in the following analyses, point densities will be used over aggregated crime counts where possible and, failing this, SOA-aggregated data will be used.

Figure 4.5: Aggregated burglary counts at the OA and SOA level and point densities (using KDE with a 350m kernel). Data are from 31st March 2000 – 1st April 2002.

From Figure 4.5 it is apparent that crime is heaviest around the city centre. To explore these hotspots in more detail it is necessary to draw boundaries around them so that they can be distinguished, this will allow them to be named and will allow for comparisons between years. Distinguishing the hotspots can be accomplished by drawing contours at a given density value in the same way that contour maps are generated from raster elevation data. However, using absolute density values for contours will make it difficult to compare different hotspot maps (recall that absolute crime numbers have dropped dramatically so the density will have fallen as well). Therefore a percentage of the total density can be used instead, as illustrated by Figure 4.6.

Figure 4.7 displays the hotspots around the city centre and their associated contours in more detail. Astonishingly, the 50% contour (the median value of the total density) covers a small area (2.17km$^2$) in the Headingley ward known locally as Hyde Park. This represents 8.5% of all the crime in Leeds. Using census data from all OAs that intersect the Hyde Park hotspot, it appears
4.4. THE VICTIMS OF CRIME

Figure 4.6: How percentages of the total density can be used to draw contours through crime density (hotspot) maps. Here, a hypothetical density surface is viewed as a cross-section and contours are drawn containing 50%, 70% and 80% of the most dense areas. These percentages were chosen because they accurately distinguish the hotspots in Figure 4.5.

that the hotspot has a burglary rate of 2,494 burglaries per 10,000 households\(^1\) compared to the city average of 936\(^2\), a Yorkshire and the Humber regional average of 646 (Dodd et al., 2004) and a national average of 422 (Dodd et al., 2004) (although the national Figures are from the 2003 British Crime Survey so are probably smaller than the national/regional figures in 2001). The most likely explanation for the hotspot is because the area is made up predominantly of students: 13% of people in the area are students compared to a mean of 3% across the city. It is possible that, due to the large number of multiple-occupancy houses in the area, the burglary rate has been inflated because a single burglary can be recorded once for each person in the house. However, this does not seem to affect the results: only 93 events (a total of 307 burglaries) of the 2496 recorded incidents that occurred in the Hyde Park hotspot had the same date/time of occurrence and \((x, y)\) coordinates (thereby most likely representing a burglary reported numerous times). Nor is the hotspot unique to the 2000/02 data set; it still exists in a similar form in the 2007/08 recorded crime data. As this hotspot clearly has an important influence on crime patterns in the city, Section 4.4.2 will explore the effect that students have on burglary in more detail.

It is informative, before conducting mathematical analysis in Section 4.4.5, to compare burglary hotspots to deprivation and census data visually. Figure 4.8 maps the burglary hotspot contours used in Figure 4.7 with the Index of Multiple Deprivation (IMD) and the Output Area Classification (OAC). It is apparent that, on the whole, burglary hotspots are in areas with high deprivation. This supports the literature which made explicit the link between crime and deprivation (see Section 2.2.2). The exception is the Hyde Park area which has low deprivation. As discussed, this is likely due to the student population and will be explored in more detail in Section 4.4.2.

Observing the OAC data in Figure 4.8, most crimes appear to be committed in ‘multicultural’ and ‘constrained by circumstances’ communities (with the exception of the Hyde Park area). There are several possible explanations for this from the crime literature. The groups are the most deprived in the classification (Vickers, 2006) so the crime rate might be simply linked to deprivation rather than other community factors. However, both groups also have high rates

\(^1\)2,691 burglaries in 10,787 households
\(^2\)29,270 burglaries in 312,571 households
Figure 4.7: Burglary hotspots near the city centre in (2000-2002) and contours that contain 50%, 70% and 80% of the total density (see Figure 4.6).
4.4. THE VICTIMS OF CRIME

Figure 4.8: Hotspot contours from Figure 4.7 with the Index of Multiple Deprivation and the Output Area Classification.

of rented accommodation (both public and private) (Vickers, 2006) which often suggests a transient community and therefore lower levels of community cohesion and guardianship (discussed in Section 2.2.2). Conversely, they have high levels of unemployment (Vickers, 2006) which might actually lead to an increase in guardianship as people are more likely to be at home during the day – recall from Section 2.4 that routine activity theory (Cohen and Felson, 1979) attributes the increasing crime rates in the 1970s to an increase in time spent away from home and the subsequent reduction in guardianship. It is also possible, however, that unemployment might increase the number of potential burglars in an area as people have limited legitimate means to acquire wealth (Wright and Decker, 1996) thereby leading to an increase in crime. The relationship between these demographic factors and burglary is undoubtedly complex and numerous contradictory conclusions can be drawn from these findings. Section 4.6 will attempt to clarify the situation by using statistical models to explore the relationships between these variables further.

4.4.2 Students and Burglary

Figure 4.7 showed that there is a considerable crime hotspot around the Hyde Park area and this can most likely be attributed to the concentrated student population who live there. It was noted that although the hotspot only covers a very small area it accounts for a considerable portion of all Leeds crime and therefore deserves a closer examination. From the literature it is clear that the high burglary victimisation rate among student populations is not a unique characteristic of Leeds; other studies have found this (Fisher et al., 1997; Barberet et al., 2004; Deakin et al., 2007). The reasons, however, are multifaceted. Tilley et al. (1999) notes that typical student demographic characteristics (young, low income), household behaviour (living in privately rented
Table 4.5: Comparing census data in OAs in Hyde Park, Harehills and Leeds. See Figure 4.7 for a definition of these areas.

<table>
<thead>
<tr>
<th>Census Variable</th>
<th>Mean OA value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Leeds</td>
</tr>
<tr>
<td>Age – percentage people aged 15–24</td>
<td>14.1</td>
</tr>
<tr>
<td>Tenure – percentage people who own their property</td>
<td>62.8</td>
</tr>
<tr>
<td>Income – number cars/vans per household.</td>
<td>0.95</td>
</tr>
<tr>
<td>Accommodation type - percent terraced houses(^1)</td>
<td>29.4</td>
</tr>
</tbody>
</table>

\(^1\) Income information was not collected in the 2001 census so the number of cars/vans per household is used as a proxy.

terraced houses or flats and regularly leaving the property empty), property ownership (an abundance of portable, high-value electronic equipment) and housing locations (typically in poorer areas) put them at risk. Furthermore, the transient nature of student communities can reduce the effects of community cohesion and guardianship (see Section 2.2.2), further increasing the risks of burglary. Most of these observations can be verified by the census; Table 4.5 compares average demographic values for OAs that intersect with the Hyde Park hotspot to the city average. As expected, Hyde Park has a younger population, fewer property owners, a lower income and a larger number of terraced houses.

Along with practical research findings, the Hyde Park hotspot also offers support for crime theories. Figure 4.9 illustrates hotspots for burglaries committed during different months in 2001 and suggests that when students leave the city the burglary rate in Hyde Park decreases considerably. In August (during the university summer holiday) the Hyde Park hotspot (measured using 50% density contours as in Section 4.4.1) covers an area of 2.1km\(^2\) whereas in November (during term time) it is approximately 4.5km\(^2\). In August there are fewer burglaries in total (1,104 compared to 1,356) which suggests that some burglars are not active at all (although it cannot be discounted that this decrease is a natural fluctuation). This provides strong support for routine activity theory: when the victim is not present (i.e. students have left the city) a crime will not occur. However, it is important to note a potential contradiction: routine activity theory (and practical research) suggest that unoccupied properties lack guardianship and are therefore at a higher risk of burglary. The reason that a decrease is seen in this case is most likely because, unlike when a property is left empty for a short period, during the holidays students take their valuable possessions with them, thereby lowering the attractiveness of the unoccupied properties.

This section has demonstrated that, in Leeds, the student presence has a substantial influence on city-wide burglary rates. Even if the Hyde Park area is not included in a model, this section makes it clear that student-like factors must be accounted for; including the attractiveness of available goods and probable levels of guardianship. Section 5.2 illustrates how factors such as building security, area attractiveness and community cohesion can be used to simulate the presence of student-like communities.
4.4. THE VICTIMS OF CRIME

4.4.3 Temporal Patterns

The crime theories discussed in Section 2.4 indicate that crime patterns must be considered within their spatial and temporal contexts. Deakin et al. (2007), for example, found that robberies were often committed in the morning because that is when respondents’ drugs needs were highest. To explore temporal burglary patterns in Leeds, the data clock can be used as implemented in ArcGIS (ESRI, 2006). Figure 4.10 displays a data clock illustrating the time and day of all burglaries in 2001. Each cell indicates the number of crimes that were committed on a particular day at a given time on an hour-by-hour basis. Caution must always be used with burglary times because they are often inaccurate. For example, if a burglary occurred while the residents were on holiday the time of the burglary could potentially cover a period of weeks. The data clock accounts for this by giving a lower weighting to events for which the exact time is not known. For example, if a burglary might have occurred over a four-hour period, each appropriate cell in the clock will be incremented by 0.25.

Figure 4.10 suggests that the largest proportion of burglaries are committed between 6pm and 11pm on Friday and Saturday evenings. This is consistent with criminology theory as it might be expected that more people to leave the house on Friday and Saturday evenings than on other days. In general, few burglaries are committed at night (midnight – 6am) and even fewer between 6am and midday the following day. This possibly corresponds to the time that potential burglars are asleep or that houses are occupied. Burglary prevalence starts to increase again after midday when many residents will be at work. There are some contradictions between the data clock and the literature, however. Burglary rates are high between 5pm and 9pm on week days which is unusual because this is when levels of guardianship might be expected to be highest. Also, Sorensen (2004) reported a peak in burglary at 11am which is not evident from these data.

It is possible that this global view hides important local effects; by disaggregating the data spatially and temporally the patterns revealed might be more in line with other research. To this end, Figure 4.11 illustrates burglary times in Leeds and the Hyde Park hotspot in August (outside...
university term time) and November (during term time). The times of burglaries are quite different in each which further adds to the evidence from Section 4.4.2 that students have a considerable influence over city-wide burglary patterns. In the city as a whole, burglaries are much more distributed in August (when the students are absent) compared to November where they occur largely in the afternoons and evenings regardless of the day of the week. As the patterns in August are more in-line with other research (there are large numbers of burglaries at night when people are asleep or in the afternoon when they are at work) it is possible that students are the reason that the times found in Figure 4.10 are not consistent with other research. The burglary times in the Hyde Park hotspot in November support this presumption. Burglaries are most common in the evening; presumably this corresponds to times when students are socialising. In particular, a large number of crimes occur on Wednesday afternoons which is a common time set aside for sporting activities. In August, burglaries in the hotspot are evenly distributed, suggesting the timing of students is not influencing burglary occurrences. Caution must be taken with these analyses as it is likely that small number problems play a part in the Hyde Park hotspot; each diagram is based on less than 150 crimes compared to over 1,000 in the city-wide analysis.

Seasonal trends can also be found by exploring the temporal data. Figure 4.12 graphs the monthly burglary counts in 2001. It appears that burglary counts are lower in the summer months with the exception of July (which probably corresponds to the time that many students in Leeds move house). Lower burglary rates in the summer have also been found by other research. Sorensen (2004), for example, attributes the finding to the greater guardianship found in summer months as people spend more time outside and in their gardens. Another note of caution must be presented, relating to the cutoff points chosen for temporal boundaries. In the same way that the modifiable areal unit problem affects spatial data, it is highly likely that choosing different...
4.4. THE VICTIMS OF CRIME

Figure 4.11: Burglary times in Leeds and the Hyde Park hotspot (from Figure 4.7) during university holidays (August) and term time (November) 2001.

Temporal boundaries will affect these results although it is beyond the scope of this research to explore this further.

Figure 4.12: Monthly burglary counts in Leeds in 2001

Figure 4.13 presents the data clocks for each month in 2001. It appears that in the winter months burglaries generally occur earlier with the times getting later in the summer months. Again this coincides with the literature and most likely corresponds to the length of daylight hours.

This section has illustrated that burglary exhibits considerable temporal variation, but the patterns are far from obvious. It is not clear to what extent the fluctuations are due to the presence of students, the number of daylight hours or anomalies in the data such as a modifiable temporal unit boundary problem or the inherent uncertainty of temporal burglary data. It does appear, however, that burglary in student households occurs at different times to burglary of other types of properties due in part to the differing daily routines of the residents. This feature must be in-
Figure 4.13: Time clocks for burglaries by month.
4.4. THE VICTIMS OF CRIME

Included in a burglary model and Section 5.5.4 will demonstrate how the virtual environment can be configured to simulate different occupancy levels at different times of day depending on the types of people who live in the community. Also, it will demonstrate how it is possible to alter the risks of burglary depending on daylight levels (although this avenue of research is left for future work). Furthermore, Section 5.8 will illustrate how agents can be created with realistic temporal behaviour to reflect the temporal patterns this section has found.

4.4.4 Crime and the Urban Environment

Literature reviewed in Section 2.2.1 referred to a link between physical household characteristics and burglary risk. One factor that has been identified as affecting burglary and can be explored using the MasterMap data is the type of the accommodation (i.e. detached, semi-detached, terraced or a block of flats). Although this information is not available directly in the data, it can be calculated through spatial queries using the MasterMap Topographic Area and Address Point data sets. In short, residential properties can be distinguished from other types of building using the Address Point NLUD code (see Section 4.3.5). Flats can be identified by those with more than eight addresses in them and different house types can be distinguished by calculating how many houses each building is adjacent to (e.g. semi-detached houses are only adjacent to one other house). Section 5.4 will document the process in detail and explain assumptions. The steps appeared to have broadly captured housing trends as illustrated by Figure 4.14.

Once each building has been classified it is possible to count the number of burglaries that have occurred in them and calculate the proportion of all burglaries by accommodation type. Table 4.6 illustrates this. Using the total number of addresses rather than the total number of buildings as the denominator in burglary rate calculations is more reliable because it takes the number of potential victims in a building into account. The most striking feature from Table 4.6 is that flats receive very few burglaries per address even though they are often in the most deprived areas. This is to be expected because there are commonly fewer entry points such as windows or back doors (Budd, 1999). There is little to distinguish between the non-flat accommodation types although terraced houses suffer slightly more than semi-detached and detached. However, terraced houses are in much more deprived areas than detached houses so it might be expected that they should suffer considerably higher victimisation rates.

It is also interesting to count the number of each type of property in each SOA and compare this to the number of burglaries in the SOA. Figure 4.15 presents scatter plots of this data. As might
Table 4.6: Proportions of burglaries by accommodation type. Measured using the number of buildings and the number of addresses (e.g. flats contain multiple separate addresses in a single building).

<table>
<thead>
<tr>
<th>Accommodation Type</th>
<th>Detached</th>
<th>Semi-detached</th>
<th>Terraced</th>
<th>Flats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of burglaries</td>
<td>4219</td>
<td>12732</td>
<td>9247</td>
<td>821</td>
</tr>
<tr>
<td>Total number of buildings</td>
<td>46123</td>
<td>145054</td>
<td>77884</td>
<td>847</td>
</tr>
<tr>
<td>Burglaries per building</td>
<td>0.09</td>
<td>0.09</td>
<td>0.19</td>
<td>0.97</td>
</tr>
<tr>
<td>Total number of addresses</td>
<td>53250</td>
<td>145054</td>
<td>94733</td>
<td>24456</td>
</tr>
<tr>
<td>Burglaries per address</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.03</td>
</tr>
<tr>
<td>Mean deprivation (IMD)</td>
<td>15.8</td>
<td>27.2</td>
<td>35.0</td>
<td>33.8</td>
</tr>
</tbody>
</table>

be expected, terraced houses are positively correlated to the number of burglaries ($\rho = 0.22$) and the opposite is true of detached houses ($\rho = -0.39$). Both these correlations are significant at the 95% level (two-tailed). However, Table 4.6 suggested that the number of burglaries per detached house is relatively high so the number of crimes should be expected to increase with the number of detached houses. As this is not the case it is apparent that the houses with the highest burglary risk must be those detached houses that are in areas with small numbers of detached houses. This offers additional support for the literature; detached houses are at a greater risk but this is not immediately apparent because they are usually in affluent areas which are outside the awareness spaces / routine activities of potential burglars. This information is also invaluable to the design of the model and Section 5.4 will demonstrate how it can be used to alter the burglary risks associated with individual properties.

![Correlation between the number of detached houses and burglary](image1)

![Correlation between the number of terraced houses and burglary](image2)

Figure 4.15: Scatter plots and lines of best fit for the number of each household type and the number of burglaries in each SOA in Leeds.

A serious drawback with this type of analysis is that it does not take account of important features from environmental criminology. For example, no consideration is made of the likelihood that a building is within a potential offender’s awareness space, whether or not the building is regularly left unoccupied and the extent to which the local community might deter a burglar. Therefore establishing actual risk is non-trivial and points to the need for accurate individual-level
models that can take these types of factors into account when assessing accommodation-related burglary risks.

4.4.5 Crime and Demographic Variables

Along with the physical factors that can affect burglary, Chapter 2 also discussed the link between burglary and the community within which it occurs. In Section 4.4.1, Figure 4.8 compared burglary rates with community data visually, but more concrete evidence can be sought by comparing community variables and burglary rates mathematically. This information will be used to determine which factors must be included in the virtual environment for a model. Recall from Section 4.3.3 that the Output Area Classification (OAC) is a classification scheme based on the 2001 UK census which is made of 41 census variables (Vickers and Rees, 2006). As these variables have been shown to be important determinants of a community cluster they will be compared to burglary rates. It is important to note that this is an ecological correlation rather than an individual correlation and is therefore subject to the ecological fallacy (Robinson, 1950), whereby assumptions made about an area might not correspond to all the individuals in the area. Unfortunately there is no individual-level demographic data available so this is an unavoidable issue. Baldwin and Bottoms (1976a), for example, recognise the problem but do not take any steps to alleviate it, other than by taking care in the interpretation of their results.

Table 4.7 presents the correlation, $\rho$, between the OAC variables and burglary at the OA and SOA levels for the top and bottom eight most correlated variables. The full list is available in Appendix B and descriptions of the variables are available in Appendix A. The Index of Multiple Deprivation (IMD) is also included in the analysis. For completeness, both absolute burglary counts and rates per household are included and, for clarity, the data are sorted by ascending burglary rate correlation at the OA level.

In general the results at OA and SOA levels are in agreement, with SOA being more extreme (both positive and negative). This occurs because the number of burglaries in OAs are skewed towards zero, an effect that is not apparent with the larger SOAs. Figure 4.16 evidences this with a histogram of burglary count frequencies at OA and SOA levels. Also, the results are similar regardless of whether burglary counts or rates are used, although this is not entirely unexpected because OAs are designed to contain similar numbers of households.

The most highly correlated variable to burglary is ‘students (full-time)’, further supporting findings from the literature and from Section 4.4.2 that students are a strong determinant of crime in Leeds. Recall that section 4.4.1 made reference to the high burglary rates in communities that have low community cohesion. The ‘rent (private)’ variable is the 3rd highest predictor which provides additional evidence for the importance of community cohesion (although it should be noted that private renting is positively correlated to the number of students, $\rho = 0.71$). Surprisingly, indicators of the presence of large ethnic groups (‘born outside the UK’, ‘Indian, Pakistani or Bangladeshi’ and ‘black African, black Caribbean or other black’) are 2nd, 4th and 5th highest respectively. The reasons for this are most likely because these communities exist in the more deprived areas in the centre of the city as illustrated by Figure 4.17 (this point will be discussed...
Table 4.7: Correlation using Pearson’s correlation coefficient, $\rho$, between OAC census variables and burglary (counts and rates per household) in each OA or SOA. All variables are significant at the 95% level (two-tailed).

<table>
<thead>
<tr>
<th>Variable Name and OAC Number</th>
<th>Burglary count OA</th>
<th>Burglary rate OA</th>
<th>Burglary count SOA</th>
<th>Burglary rate SOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Age 45-64</td>
<td>-0.42</td>
<td>-0.58</td>
<td>-0.44</td>
<td>-0.6</td>
</tr>
<tr>
<td>37 Manufacturing employment</td>
<td>-0.36</td>
<td>-0.55</td>
<td>-0.38</td>
<td>-0.55</td>
</tr>
<tr>
<td>14 Two adults no children</td>
<td>-0.34</td>
<td>-0.59</td>
<td>-0.35</td>
<td>-0.57</td>
</tr>
<tr>
<td>33 Working part-time</td>
<td>-0.34</td>
<td>-0.53</td>
<td>-0.34</td>
<td>-0.54</td>
</tr>
<tr>
<td>30 Provide unpaid care</td>
<td>-0.31</td>
<td>-0.49</td>
<td>-0.33</td>
<td>-0.49</td>
</tr>
<tr>
<td>5 Age 65+</td>
<td>-0.27</td>
<td>-0.31</td>
<td>-0.3</td>
<td>-0.38</td>
</tr>
<tr>
<td>36 Mining/Quarrying/Construction employment</td>
<td>-0.24</td>
<td>-0.51</td>
<td>-0.28</td>
<td>-0.52</td>
</tr>
<tr>
<td>15 Households with non-dependant children</td>
<td>-0.26</td>
<td>-0.52</td>
<td>-0.25</td>
<td>-0.48</td>
</tr>
<tr>
<td>23 People per room</td>
<td>0.26</td>
<td>0.33</td>
<td>0.31</td>
<td>0.42</td>
</tr>
<tr>
<td>- IMD Score</td>
<td>0.34</td>
<td>0.48</td>
<td>0.31</td>
<td>0.4</td>
</tr>
<tr>
<td>27 Public Transport to work</td>
<td>0.35</td>
<td>0.54</td>
<td>0.34</td>
<td>0.5</td>
</tr>
<tr>
<td>7 Black african, Black Caribbean or Other Black</td>
<td>0.34</td>
<td>0.55</td>
<td>0.35</td>
<td>0.5</td>
</tr>
<tr>
<td>6 Indian, Pakistani or Bangladeshi</td>
<td>0.34</td>
<td>0.48</td>
<td>0.36</td>
<td>0.51</td>
</tr>
<tr>
<td>17 Rent (Private)</td>
<td>0.34</td>
<td>0.5</td>
<td>0.36</td>
<td>0.51</td>
</tr>
<tr>
<td>8 Born Outside the UK</td>
<td>0.37</td>
<td>0.57</td>
<td>0.39</td>
<td>0.55</td>
</tr>
<tr>
<td>31 Students (full-time)</td>
<td>0.45</td>
<td>0.59</td>
<td>0.49</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Figure 4.16: Histograms of burglary counts at OA and SOA levels.
4.4. THE VICTIMS OF CRIME

in more detail below). With respect to deprivation, measures of non-affluence (‘public transport to work’, ‘IMD score’ and ‘people per room’) are also important (6th, 7th and 8th respectively). These further highlight the important link between deprivation and crime and will be used to influence the design of the model (see Section 5.5.1).

At the other end of the scale, older communities appear to be negatively correlated with burglary; ‘age 45–64’ is the strongest negative determinant of burglary and ‘age 65+’ is the 6th. Also, the following employment-related variables are also negatively correlated to burglary (2nd, 3rd and 4th most important respectively): ‘manufacturing employment’, ‘working part-time’ and ‘mining / quarrying / construction employment’. It is difficult to see why these employment variables relate to burglary risk; however, once the variables are mapped a pattern becomes apparent. Figure 4.17 compares some of the largest positively- and negative-correlated variables spatially. On the whole, variables that are positively correlated to burglary are highest in the city centre and vice versa for negatively correlated variables. Therefore there is a clear spatial element to the correlation that simple statistics do not account for and might explain why some variables, that appear unrelated to burglary, have strong correlations. Section 4.6 will explore the relationship between these variables in more detail and discuss the extent to which collinearity might explain some of the more unexpected findings.

Figure 4.17 compares some of the largest positively- and negative-correlated variables spatially. On the whole, variables that are positively correlated to burglary are highest in the city centre and vice versa for negatively correlated variables. Therefore there is a clear spatial element to the correlation that simple statistics do not account for and might explain why some variables, that appear unrelated to burglary, have strong correlations. Section 4.6 will explore the relationship between these variables in more detail and discuss the extent to which collinearity might explain some of the more unexpected findings.

The correlation analysis in the section has assessed the relationship between key demographic variables and crime rates. Among other things, students populations were shown to be strongly positively correlated to burglary while older communities were negatively correlated. Indicators of employment type were also strongly related to burglary (both positively and negatively depending
on the type). These findings offer local empirical support for the choice of variables that need to be included in a burglary model.

4.5 The Offenders

Section 4.3.6 introduced the second available crime data set which lists the home postcodes of people who have been linked to a crime in the victims dataset (so called “nominals”). Recall that this is not a dataset of convicted offenders, rather those who are suspects, potential accomplices or simply wanted for questioning (Kongmuang, 2006). Inevitably some people in the database are innocent and will never be convicted in relation to the crime but unfortunately it is not possible to distinguish such cases and in the following analyses each nominal will be referred to as an “offender”.

The data are available for the period 1st April 2000 – 31st March 2004 and contain approximately 70,000 records. Of these, only 5,478 are “burglary dwelling”, 5022 had correctly recorded home postcodes and 4,728 of those had offender home addresses in Leeds. Usually it would be appropriate to analyse offender data covering the period 1st April 2000 – 31st March 2002 in order to be consistent with the analysis of recorded crime data. However this would lead to problems with small numbers particularly when analysing the data in small geographical areas so instead the entire dataset will be used. The drawbacks of inconsistency between data are mediated by the benefits of the more comprehensive patterns that will be produced by using a fuller data set.

4.5.1 City-wide Offender Locations

Figure 4.18 depicts the locations of nominal home addresses as a density surface. Two approaches are used to visualise the data. The first, 4.18(a), creates a point for every entry in the dataset, i.e. if an offender is associated with ten crimes, there will be ten points created at their home address. This approach can be used to explore the existence of prolific offenders. The second approach, 4.18(b), only creates one point for each unique nominal and home address so shows the distributions of offenders without taking into account the number of crimes for which they are responsible. It appears that the hotspots in Hyde Park and Kirkstall only exist because they house a small number of highly prolific offenders. For example: in Kirkstall there is a nominal associated with 72 crimes and one with 40; in Hyde Park there is a nominal associated with 78 crimes and two with 17.

It must be decided whether it is more appropriate to measure offending behaviour by volume (Figure 4.18(a)) or by unique home locations (Figure 4.18(b)). The vast majority of nominals are only associated with one crime but a small number have an extremely large crime count. However, because prolific offenders are likely to have a much greater effect on an area because they commit such a large number of crimes, it is more appropriate to look at the volume. Using home addresses without any measure of the number of crimes committed from that address fails to capture a vital element of the system. Ultimately the amount of burglary in an area is the same regardless of whether or not it is committed by a number of people or by a single person.
4.5. THE OFFENDERS

(a) One point for every data entry (i.e. multiple points for prolific offenders).
(b) One point per nominal home address (i.e. a single point for prolific offenders).

Figure 4.18: Density maps of offender home locations. KDE parameters: kernel 100m, cell size 20m$^2$.

Figure 4.19 visualises the city centre nominal hotspots (many points per address) and compares them to burglary hotspot contours from Figure 4.7 (page 62). The offender/offence hotspots in Beeston are very similar which suggests that most offenders in the area are locals. This supports findings from the literature in Section 2.4.4; most offender movements are short as they generally burglar areas they know well surrounding their home. The same is true, to a lesser extent, in Bramley, Wortley and Harehills. Hyde Park, however, demonstrates a slightly different pattern. A large part of the area has a very low offender count, but there are hotspots surrounding it to the west and south-east. It could be assumed, therefore, that offenders who live in these areas are travelling into Hyde Park to burglar students. The following section will explore this hypothesis in more detail.

4.5.2 Travel-to-crime

As this research focuses on individual burglar actions it is important to understand where offenders are likely to travel to in order to commit a crime. Section 2.4.4 discussed the findings from the literature and outlined the commuter and marauder hypotheses (Canter and Larkin, 1993). The commuter hypothesis stipulates that an offender travels to a particular area to commit crime and then travels home whereas the marauder hypothesis suggests that an offender is more likely to commit crimes in their home areas. This is illustrated by Figure 4.20. Along with literature findings, there is evidence in the crime data to support these theories which this section will outline.

Figure 4.21 presents maps that illustrate where a selection of prolific offenders have committed their crimes. These maps were chosen for this research because they are illustrative of particular patterns and are not necessarily a fair summary of all the offender data (more on this point shortly). Images A – C show “commuter” type offenders who travel some distance (over one mile) but burglar in a relatively small area. Although the distribution of offences in Image B are reasonably spread out, all burglaries are committed in the Headingley / Hyde Park / Burley areas
which house large numbers of students. The offender in image C appears to choose two target areas, although crimes within these areas are distributed relatively closely together. There are two possible reasons why the offenders in images A – C chose those particular areas to burgle in. They might know about the areas because they have legitimate reasons to travel there (visiting friends or for employment as stipulated by crime theories outlined in Section 2.4) or the areas might have reputations as easy targets (which is likely the case in the student areas).

Images D – E in Figure 4.21 illustrate “marauder” behaviour; the offenders burgle within a relatively small area around their home, within $\frac{1}{2}$ mile. This also supports crime theories because it suggests that the offenders are burgling from within their awareness spaces that surround their homes. It might be expected that these “marauders” are younger than the commuters because they do not travel as far. Table 4.8 presents some evidence to support this. Two of the offenders are the youngest of all those examined (14 and 19 years) although the shortest mean distance is actually attributed to a 31-year-old who is the third oldest of all nine nominals. Drawing any firm conclusions of this sort, however, is difficult when examination of offender patterns is a manual and time consuming operation and subsequently only a small number of nominals are examined. A computer algorithm, on the other hand, could classify offenders into different types and would be able to compare these classifications to other offender attributes (such as age, home location etc) to search for patterns. In this manner the entire data set could be analysed in a much more structured way, unaffected by human inference. This is beyond the scope of the project, however,
4.5. THE OFFENDERS

Figure 4.20: The commuter and marauder hypotheses, adapted from Canter and Larkin (1993).

Table 4.8: Details of the offenders whose offending behaviour is illustrated in Figure 4.21

<table>
<thead>
<tr>
<th>Image</th>
<th>Commuters</th>
<th>Marauders</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Age</td>
<td>26</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>Gender</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Ethnicity&lt;sup&gt;1&lt;/sup&gt;</td>
<td>WE</td>
<td>WE</td>
<td>WE</td>
</tr>
<tr>
<td>Mean travel dist. (km)</td>
<td>2.10</td>
<td>3.13</td>
<td>5.09</td>
</tr>
</tbody>
</table>

<sup>1</sup> For ethnicity, WE stands for “western european” and AC stands for “afro-caribbean”.

Finally, images G – I illustrate three behaviours which cannot be classified as marauders or commuters. Each nominical travels relatively large distances (up to 6km in one case) and there are no discernible patterns. Unfortunately this type of distribution was the most commonly found by this researcher which makes it difficult to draw any firm conclusions about the travel patterns of the offenders in the data.

4.5.3 Offenders and Demographic Variables

As with the recorded crime data in Section 4.4.5, it is possible to compare demographic variables to offender home locations in order to establish which variables are the most likely predictors of offender populations. Table 4.9 illustrates the Pearson’s correlation coefficient values, $\rho$ for the 8 most highly positively- and negatively-correlated variables. The full list of variables can be found in Appendix B.

As with the recorded crime analysis, there is a correspondence between OAs and SOAs and also between counts and rates. The most interesting feature is that, at both ends of the scale, measures of deprivation appear to be the most powerful correlates with offender rates. The ‘IMD Score’ is very strongly positively-correlated correlated ($\rho = 0.56$, see Figure 4.22) as well as ‘Un-
Figure 4.21: Examples of offender journey to crime. Illustrates patterns that fit with commuter and marauder hypotheses as well as less categorisable patterns.
### Table 4.9: Correlation using Pearson’s correlation coefficient, $\rho$, between OAC census variables and offender home locations (counts and rates per household) in each OA or SOA. All values are significant at the 95% level (two-tailed)

<table>
<thead>
<tr>
<th>Variable Name and OAC Number</th>
<th>Offender count</th>
<th>Offender rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OA</td>
<td>SOA</td>
</tr>
<tr>
<td>26  2+ Car household</td>
<td>-0.26</td>
<td>-0.5</td>
</tr>
<tr>
<td>14  Two adults no children</td>
<td>-0.18</td>
<td>-0.39</td>
</tr>
<tr>
<td>22  Rooms per household</td>
<td>-0.15</td>
<td>-0.34</td>
</tr>
<tr>
<td>19  Detached Housing</td>
<td>-0.14</td>
<td>-0.34</td>
</tr>
<tr>
<td>30  Provide unpaid care</td>
<td>-0.13</td>
<td>-0.21</td>
</tr>
<tr>
<td>24  HE Qualification</td>
<td>-0.14</td>
<td>-0.24</td>
</tr>
<tr>
<td>4   Age 45-64</td>
<td>-0.13</td>
<td>-0.25</td>
</tr>
<tr>
<td>28  Work from home</td>
<td>-0.13</td>
<td>-0.35</td>
</tr>
<tr>
<td>13  Lone Parent household</td>
<td>0.22</td>
<td>0.41</td>
</tr>
<tr>
<td>10  Divorced</td>
<td>0.19</td>
<td>0.33</td>
</tr>
<tr>
<td>27  Public Transport to work</td>
<td>0.22</td>
<td>0.41</td>
</tr>
<tr>
<td>16  Rent (Public)</td>
<td>0.23</td>
<td>0.47</td>
</tr>
<tr>
<td>-   Burglary count</td>
<td>0.28</td>
<td>0.44</td>
</tr>
<tr>
<td>29  Llti (SIR)</td>
<td>0.26</td>
<td>0.49</td>
</tr>
<tr>
<td>32  Unemployed</td>
<td>0.28</td>
<td>0.52</td>
</tr>
<tr>
<td>-   IMD Score</td>
<td>0.3</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Furthermore, ‘2+ car household’, ‘Rooms per household’ and ‘Detached housing’ are strongly negatively correlated. This adds a considerable amount of empirical evidence for the link between crime and deprivation and suggests that more burglars are likely to live in the more deprived areas of the city. This evidence will be invaluable when determining where (and how many) potential burglars live in the model.

### 4.6 Regression Models of Burglary

As discussed in Section 3.3, crime is commonly modelled using regression. This section will utilise the technique in order to further explore the relationship between burglary and its potential determinants. This will also provide the basis on which to compare the accuracy of the agent-based model produced by this research to “traditional” modelling (for examples see Section 7.3.6). Two types of models will be employed: a typical linear regression (e.g. Cohn, 1993; Beavon et al., 1994) and a geographically weighted regression (e.g. Cahill and Mulligan, 2007), the latter being better suited to modelling spatially variable data.
CHAPTER 4. DATA ANALYSIS: CRIME AND THE ENVIRONMENT

Figure 4.22: Scatterplot of IMD score and offender counts at the SOA level. As deprivation increases so does the expected number of offenders in an SOA.

4.6.1 A Linear Regression Model

A multiple linear regression model expresses a dependent variable, $y$, on the basis of the linear combination of a number of predictor variables, $x_i$:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$  \hspace{1cm} (4.1)

where $\alpha$ and $\beta_i$ are the coefficients that must be calculated and $n$ is the total number of predictor variables. Typically regression models are built by finding the combination of coefficients that result in the smallest size of the sum of square residuals (called ordinary least squares, OLS, models). In order to make results comparable across different models and data, however, standard methods can be used to assess model accuracy rather than reporting the sum of square residuals.

The most common of these is the coefficient of determination, $R^2$, which is the proportion of the variance in the dependent variable that can be accounted for by the model (Faraway, 2002):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$  \hspace{1cm} (4.2)

where $\hat{y}_i$ is the predicted value of the observation $i$, $y_i$ is the actual value of $i$ and $\bar{y}$ is the mean of the predicted values ($\hat{y}$). The statistic ranges from 0 (no variance in the dependant variable is explained by the model) to 1 (which only occurs when model predictions are exactly the same as the real observations, i.e. a perfect model). Section 7.2.3 will explore the statistic in more detail and also discuss alternative measures.

It is possible to use all the OAC variables in a model, but some might not have an effect and should therefore not be included. Such parameters will have $\beta$ coefficients that are not significantly different from zero and will therefore have the same effect on the dependant variable, $y$, regardless of their value. To account for this, a stepwise variable selection method will be used to select, from the set of all variables, only those that have an effect on the model. Selecting variables to
add or remove is non-trivial, however, as p-values should not be treated “too literally” (Faraway, 2002); they can increase or decrease as variables are added or removed. Preferred methods observe *all* possible models (with \( m \) parameters there are \( 2^m \) possible models) and keep the one with the optimal balance between a large \( R^2 \) value and a small \( m \). Here the Akaike Information Criterion (AIC) is used, as implemented in the R statistical package. For more information about AIC see Akaike (1974).

The overall accuracy of the model, \( R^2 \), is 0.6908 when applied to SOAs. This is lower than the work of Kongmuang (2006) who used similar data although this is most likely attributable to the lower spatial resolution of Kongmuang’s model which operated at the ward level. However, this value of \( R^2 \) is still acceptable when compared to other models. Cahill and Mulligan (2007), for example, note that an \( R^2 \) value of only 0.36 is generally on par with other criminal justice studies.

Table 4.10 summarises the model coefficients. It is ordered by the descending t-value (so that the variables with the strongest positive and negative effects are at the top and bottom of the table respectively) and includes the following columns:

- Predictor variable – the variables used to explain crime rates, \( x \);
- Estimate – the value of the predictor’s \( \beta \) coefficient;
- Standard error – the standard error of the \( \beta \) estimate;
- t-value – the t statistic (to test whether or not the predictor is significantly different from zero);
- p-value: the probability that t value is not significant (i.e. if \( p < 0.025 \) then the \( \beta \) is significant at the 95% level).

In general, the results agree with the previous analysis of correlation in Section 4.4. Indicators of deprivation are strong influences of burglary (both positively and negatively). Interestingly, the *Students (full time)* variable is not the most significant predictor of burglary \( (p = 0.039) \) but *HE Qualifications* is \( (p < 0.001) \) which probably accounts for the student populations. Section 4.4.1 highlighted the confusing relationship between unemployment and crime, suggesting that unemployment could lead to higher rates of offending because people no longer have legitimate means to earn money. However it might equally lead to increased guardianship and therefore reduced crime. From this model, the ‘*Work from home*’ variable is a strong negative predictor of burglary, suggesting that if people are at home during the day there is indeed increased property guardianship and reduced burglary. Unemployment does not feature in this model but it does feature in a similar model that estimates offender rates rather than victimisation rates (not shown here). This suggests that unemployment leads to increased numbers of offenders but not necessarily increased crime. This supports evidence from Section 2.4.4 which discussed the journey to crime: although burglars are unlikely to travel far, they will not commit crime too close to their homes for fear of being seen by someone they know.

Another interesting feature is that ‘*Terraced housing*’ is a strong \( (p < 0.01) \) negative predictor of burglary but Pearson’s correlation coefficient, \( \rho = 0.16 \), suggests positive correlation (see Table 4.7 on page 72). This unusual effect is most likely caused by collinearity in the explanatory
Table 4.10: Results of the burglary regression model at SOA level, ordered by the t-value (descending) which indicates how significant the predictor is. $R^2=0.6908$.

<table>
<thead>
<tr>
<th>Predictor variable, $x$</th>
<th>$\beta$ Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>1.26</td>
<td>0.19</td>
<td>6.59</td>
<td>0.000</td>
</tr>
<tr>
<td>IMD Score</td>
<td>111.99</td>
<td>22.43</td>
<td>4.99</td>
<td>0.000</td>
</tr>
<tr>
<td>Public Transport to work</td>
<td>100.85</td>
<td>22.87</td>
<td>4.41</td>
<td>0.000</td>
</tr>
<tr>
<td>Two adults no children</td>
<td>114.38</td>
<td>29.24</td>
<td>3.91</td>
<td>0.000</td>
</tr>
<tr>
<td>Rooms per household</td>
<td>166.16</td>
<td>43.52</td>
<td>3.82</td>
<td>0.000</td>
</tr>
<tr>
<td>Lone Parent household</td>
<td>63.72</td>
<td>21.45</td>
<td>2.97</td>
<td>0.003</td>
</tr>
<tr>
<td>Number of people</td>
<td>0.03</td>
<td>0.01</td>
<td>2.38</td>
<td>0.018</td>
</tr>
<tr>
<td>Students (full time)</td>
<td>43.3</td>
<td>20.93</td>
<td>2.07</td>
<td>0.039</td>
</tr>
<tr>
<td>Llti SIR</td>
<td>156.07</td>
<td>76.26</td>
<td>2.05</td>
<td>0.041</td>
</tr>
<tr>
<td>All Flats</td>
<td>16.49</td>
<td>9.47</td>
<td>1.74</td>
<td>0.082</td>
</tr>
<tr>
<td>No central heating</td>
<td>15.62</td>
<td>9.56</td>
<td>1.63</td>
<td>0.103</td>
</tr>
<tr>
<td>Finance 5–14</td>
<td>-39.41</td>
<td>26.64</td>
<td>-1.48</td>
<td>0.140</td>
</tr>
<tr>
<td>Financial intermediation employment</td>
<td>-34.24</td>
<td>19.52</td>
<td>-1.75</td>
<td>0.080</td>
</tr>
<tr>
<td>Divorced</td>
<td>-61.31</td>
<td>33.37</td>
<td>-1.84</td>
<td>0.067</td>
</tr>
<tr>
<td>Age 25–44</td>
<td>-94.12</td>
<td>48.9</td>
<td>-1.92</td>
<td>0.055</td>
</tr>
<tr>
<td>Households with non-dependant children</td>
<td>-46.24</td>
<td>19.25</td>
<td>-2.4</td>
<td>0.017</td>
</tr>
<tr>
<td>Terraced Housing</td>
<td>-17.87</td>
<td>6.67</td>
<td>-2.68</td>
<td>0.008</td>
</tr>
<tr>
<td>Work from home</td>
<td>-87.48</td>
<td>26.69</td>
<td>-3.28</td>
<td>0.001</td>
</tr>
<tr>
<td>Manufacturing employment</td>
<td>-102.06</td>
<td>30.15</td>
<td>-3.39</td>
<td>0.001</td>
</tr>
</tbody>
</table>

$\alpha$                     |-222.54          | 75.11          | -2.96   | 0.003   |

variables. Faraway (2002) notes that collinearity can cause, among other things, imprecise estimates of the $\beta$ coefficients including misleading signs. Removing a collinear variable does not necessarily reduce the accuracy of the model and can affect the new coefficients. For example, the ‘IMD Score’ is strongly positively correlated with ‘Public Transport to work’ ($\rho = 0.72$), ‘Llti SIR’ ($\rho = 0.89$) and ‘Lone parent household’ ($\rho = 0.74$) and after removing these three parameters the model explains only 1.4% less of the variance ($R^2 = 0.6766$). Furthermore, some parameters in this reduced model are no longer significant so could potentially be removed and the process repeated. However, it is not necessary to explore this regression model further because the overall accuracy is unlikely to increase much above 69% by tweaking the choice of parameters (the AIC test has been used in a stepwise algorithm to search the space of all possible models). Importantly a benchmark has been found with which the agent-based model can be compared.

However, there is an unrelated but nevertheless serious drawback with this type of model that should be discussed: it is “global”. It assumes that the effect that the $\beta$ values will have on the dependent variable does not vary geographically. This is extremely unlikely to be the case (all the research thus far suggests that geographical variation is present in the data) and empirical evidence for this can provided. Moran’s I is a measure of spatial autocorrelation and can be used to estimate the extent that the data depart from spatial randomness. Values of the statistic range from -1 (dispersed data) to +1 (clustered data) with values near to 0 indicating spatial randomness. ArcGIS (ESRI, 2006) was used to calculate the statistic and all variables tested had positive I values with Z scores above 1.96 which suggests they are statistically significant (95% level, two
tailed). As spatial variation in the parameters is clearly present, the following section will apply an alternative regression technique that allows coefficients to vary spatially.

### 4.6.2 A Geographically Weighted Regression (GWR) Model

An alternative to standard linear regression, geographically weighted regression (GWR: Brunsdon et al., 1996) determines the $\beta$ values as a function of their spatial location, $l$:

$$y = \alpha(l) + \beta_1(l)x_1 + \beta_2(l)x_2 + \ldots + \beta_n(l)x_n$$  \hspace{1cm} (4.3)

The spatial variation is introduced by using a weighting scheme such that the $\beta$ values at a particular location are more heavily influenced by regression points that are closer than those that are farther away. The result is that spatial variation can be included in the model to explain the relationship between the $x_i$ and $y$ that is otherwise not included in the $x_i$ (Heppenstall, 2004). Also, the values of the $\beta_i$ can be mapped to explore how the variables are changing spatially.

A GWR model was constructed using an implementation in the ArcGIS (ESRI, 2009) software. Ideally the GWR model would replicate the previous OLS model exactly but, due to the implementation used, only a subset of all the variables could be included because the model failed to run in the presence of collinearity between variables. Including only the three most important variables (the ‘IMD Score’, ‘HE Qualifications’ and ‘Public Transport to Work’) nevertheless results in a slightly more accurate model with $R^2 = 0.70$.

Figure 4.23 maps the values of the coefficients for the three explanatory variables. It is interesting to note that the ‘IMD Score’ and ‘Public Transport to Work’ variables vary relative to the city centre whereas the coefficient for ‘HE Qualifications’ demonstrates a decrease from the north-west to south-east corners of the city. Clearly there is a wealth of further investigation that could be conducted, but this is beyond the scope of the research. This section has demonstrated that although a regression model that can account for spatial variation in the coefficients is able to improve upon a typical OLS regression model, the increase in accuracy is limited. Therefore a base of approximately $R^2 = 0.70$ is a suitable level of accuracy on which the agent-based model that will described in subsequent chapters should aim for.

### 4.7 Data Analysis – Conclusion

An important part of this research is to determine which factors must be included in a burglary model. To this end, this chapter has expanded on the findings from Chapter 2 in search of empirical evidence to support the crime literature. The chapter began by identifying important data sources and discussing some of their limitations and drawbacks. A systematic analysis of the recorded crime data as well as the available offender data followed in order to both explore crime patterns in Leeds and to identify important variables that should be included in a model.

With respect to the recorded crime data, deprivation appears to have a strong influence over where burglary occurs; often the most deprived areas are also the most dense burglary hotspots. Students were also shown to have a significant influence on the distribution of burglary in the city.
There is a dense burglary hotspot that consistently covers a large predominantly student area called Headingley / Hyde Park. Therefore a measure of student populations as well as deprivation must be included in a model. In order to estimate household burglary risk based on accommodation type, burglary rates were compared to the abundance of terraced, detached and semi-detached houses in an area. It was shown that, as expected, detached houses have a high burglary risk so this is another factors that will be included in the model.

Hotspots of offender home locations were often found to overlap with burglary hotspots which suggests, as noted in the literature, that burglar movements are relatively short. However, in some areas (such as Headingley / Hyde Park) there was clear evidence that burglars were travelling to the area from elsewhere (“commuters”). This information will be used to build realistic burglar behaviour. Also, deprivation was shown to be strongly correlated to the number of offenders in an area which further strengthens the link between crime and deprivation.

To conclude the chapter, regression techniques were employed to estimate burglary based on census variables. These models provide a benchmark on which to compare the accuracy of the agent-based model produced by this research (see Section 7.3.6). However, it was shown that even when the model parameters were able to vary spatially using geographically weighted regression, the model could only account for 70% of the variation in burglary rates. The following chapter will outline how the data discussed here can be used to build a more realistic model that is able to account for these factors.
Chapter 5

Creating Virtual People and their Virtual Environment

5.1 Introduction

Previous chapters have identified some of the important factors that drive the residential burglary system, both from a review of crime literature and through the analysis of Leeds data. This chapter will tie the reviews to the model development chapters through a discussion of how the two components of a model – the virtual environment and virtual burglars – can be conceptualised using the available theories and data. This will set the groundwork for Chapter 6, which will document, technically, how the model has been developed. This chapter begins with an overview of the environment (Section 5.2), indicating how it is able to represent the “environmental backcloth” that strongly indicates where a crime is likely to occur. Then an outline of how the geographical and
demographic data presented in Chapter 4 can be used to create roads, buildings and communities will follow.

The second part of the chapter will discuss how the virtual burglars can be conceptualised and how burglar behaviour theories can be incorporated into the virtual agents. The chapter will finish with a discussion of the cognitive framework that will be used to drive their behaviour, the set of actions available to the agents and the actual process of burglary.

5.2 An Overview of the Virtual Environment

The virtual environment is used to represent the space that the agents inhabit. Although it must allow the agents to perform normal, every-day activities (such as travelling from one place to another) the most important function of the virtual environment is to incorporate many of the factors that form the “environmental backcloth” (Brantingham and Brantingham, 1993) which will influence where a potential burglar is likely to commit a crime. This section will re-visit the important features identified from the literature and discuss how they can be conceptualised in an agent-based model. It presents an overview of the structure of the environment; the technical details regarding how the demographic/geographic data can be used to establish parameter values in the environment are the subject of sections 5.3, 5.4 and 5.5. Figure 5.1 presents the general overview which will be explained throughout the remainder of this chapter.

With respect to burglary risk, Chapter 2 identified features that correspond to both individual houses (e.g. security) as well as community-wide factors (e.g. community cohesion). Therefore the environment will consist of two distinct layers: one containing individual house objects and the other containing community objects. Establishing house boundaries is relatively simple as MasterMap data (see Section 4.3.5) contains the individual boundaries of every building in Leeds. Therefore all that is required to create virtual houses is to distinguish residential properties from other buildings; this will be discussed in Section 5.4. Determining community boundaries, on the other hand, is much more difficult. Using output areas appears to be the obvious choice because they are the smallest administrative boundary used in the 2001 census. However, Section 2.1.2 noted that using administratively-bounded data is fraught with problems, largely because the boundaries are designed for administrative purposes rather than for distinguishing homogeneous communities. Ideally, this research would create new boundaries based on community heterogeneity, thus enabling a much more accurate definition of a “community”. However, this is beyond the scope of the project and, therefore, communities will be identified by output area boundaries.

Many of the problems with traditional crime models that were identified in Section 3.3 relate to their inability to model individual-level objects such as houses, victims, guardians and offenders etc. To truly capture the dynamics of the environmental criminology theories discussed in Section 2.4, it is necessary to model the individual behaviour of all the people who live in the city and could, in theory, be involved with a crime event. Although this research will model the behaviour of individual offenders directly (through the use of intelligent “burglar agents” as discussed in Section 5.6), it is beyond the scope of the project to model every person in Leeds – although this
5.2. AN OVERVIEW OF THE VIRTUAL ENVIRONMENT

Environment layers

Community

1. Collective Efficacy
   a. Concentrated Disadvantage
   b. Residential Stability
   c. Ethnic Heterogeneity

2. Attractiveness
   a. Deprivation Disparity (IMD)

3. Sociotype
   a. Output Area Classification

4. Occupancy

Individual Houses

1. Accessibility
   a. Number of Possible Entrances
   b. Security

2. Visibility
   a. Size of Garden
   b. Degree of Isolation

3. Traffic Volume
   a. Space Syntax

Figure 5.1: The layers that make up the virtual environment.
is recommended as future work. Therefore the communities layer of the virtual environment will be used to simulate the existence of other people whose presence might deter a potential burglar. Sections 5.3.2 and 5.5.4 will discuss how this can be accomplished by altering traffic volume and occupancy levels dynamically; keeping the backcloth detailed and dynamic without actually simulating individual daily habits.

House parameters

It was established from the literature in Section 2.2.1 that the following household-level factors were particularly important at determining burglary risk:

- **Accessibility** – how easy it is to gain entry to the house (e.g. the number of windows or doors);
- **Visibility** – how visible the house is to neighbours and passers-by who can act as suitable guardians and deter a potential burglar;
- **Occupancy** – burglars are less likely to enter properties that are occupied;
- **Security** – effective physical security can deter burglars;
- **Attractiveness** – an abundance of valuable goods can encourage burglary;
- **Traffic volume** – high levels of pedestrian or vehicle traffic can make it difficult to gain access to a property without being seen.

Parameter values for *accessibility* and *visibility* can be calculated through an analysis of the geographic household data (see Sections 5.4.2 and 5.4.3 respectively). Unfortunately there is limited data relating to household *security* so while this parameter will be included because it is important, at this stage all houses will be given the same security value and it will therefore not influence the burglars’ burglary decision. The *traffic volume* parameter can be calculated through an analysis of the geographic roads data which will be described in Section 5.3.2. Each building is assigned the traffic volume of the road it is nearest to. The two remaining parameters, *occupancy* and *attractiveness*, cannot be estimated at the household level due to insufficient data. They depend on the behaviour of individual people (e.g. the times that a family go to work or school and the abundance of goods they own) but proxies for these attributes are only available through the census. It is possible to create an individual-level population using microsimulation (discussed in Section 3.4) but this is beyond the scope of the project. Instead, each house in a community will be assigned the same attractiveness and occupancy values, the calculation of which are discussed in Sections 5.5.3 and 5.5.4 respectively.

Community parameters

With respect to community elements of the environmental backcloth, Section 2.2 discussed two important factors: *collective efficacy* and *community similarity*. Collective efficacy is a measure of community cohesion; communities that are highly cohesive are more likely to act as a deterrent to potential burglars. This deterrent can take the form of greater guardianship as residents are more
mindful of their property (and that of their neighbours) and also from physical cues that suggest to a potential burglar that they will be noticed committing a crime. Both of these elements will be simulated in a single measure which is documented in Section 5.5.1.

Community similarity is a measure of how comfortable an offender feels in a neighbourhood. Literature discussed in Section 2.3 found that some offenders will avoid areas in which they do not “fit in” because they are worried about being noticed as an outsider. This value, therefore, depends on where an agent lives and it will be used as a comparison between their home community and the community they could potentially commit a burglary in. The **sociotype** variable is defined to allow community comparisons as discussed in Section 5.5.2.

### Parameter Summary

Figure 5.2 presents the organisation of the different parameters, indicating which environmental layer they are applied to. Although there are relatively few parameters, they are able to incorporate a large degree of the complexity of the environmental backcloth. For example, Chapters 2 and 4 were consistent in their support for the hypothesis that students suffer disproportionately high levels of burglary. This type of community could be simulated as follows:

1. low **security** values – students were shown to rarely utilise security and/or leave doors or windows unlocked;
2. high **attractiveness** – student houses often have an abundance of attractive goods;
3. low **collective efficacy** – communities are often highly transient;
4. different **occupancy** levels – student communities are likely to have empty houses at different times when compared to other communities (e.g. families or professionals).

![Virtual Environment Parameter Organisation](image)

Figure 5.2: Parameters used in the household and community environment layers. Although they are strictly household parameters, there is insufficient data to estimate occupancy and attractiveness at the household level so they must be estimated from census data at the community level.

This example indicates that the virtual environment suitably represents theory, is highly flexible and allows for the incorporation of many crime concepts. The following sections will provide
the details of how environmental parameter values are calculated before Section 5.6 presents the other major model component: the virtual burglars.

Normalising Data

Before continuing, it is important to note that raw data in the following sections will all be normalised to the range $0–1$. This operation is performed so that the influence a data source has on the model is not biased by its range of values. The normalised value $x_{\text{norm}}$ of a parameter $x$ in a dataset $S$ will be normalised such that:

$$x_{\text{norm}} = \frac{S_{\text{max}} - x}{S_{\text{max}} - S_{\text{min}}}$$  \hspace{1cm} (5.1)

where $S_{\text{max}}$ and $S_{\text{min}}$ are the minimum and maximum data values in $S$ respectively.

5.3 The Road Network

Roads are required in the model because they form the paths that the agents travel along. Also, as discussed in Section 5.2, analysis of the road network will be used to estimate levels of traffic volume. Section 4.3.5 introduced the Ordnance Survey MasterMap product which includes the Integrated Transport Network (ITN) layer. The ITN layer consists of 45,815 separate line objects in Leeds that represent different types of roads. Each road is assigned a type which can be used to identify how the road should be used by agents in the model, as discussed in the following section.

5.3.1 Road Accessibility Attributes

Before analysing the road data further, it must be decided which transport methods are available to agents to travel around the city. Although there is limited data regarding how burglars actually travel to burglary sites, Wiles and Costello (2000) suggest that most have access to cars – whether owned personally (49%), owned by a friend (77%) or stolen (51%) – often use public transport (86%) and also walk (77%). Bicycles are by far the least common method of transport for burglars (29%) so these will not be included. Unlike car or foot journeys, public transport routes depend on the location of bus stops or rail stations so the public transport network is created independently from the road network and will be described in Section 6.5.4. To allow for foot and car journeys, each road object in the ITN layer can be assigned an attribute dictating which method(s) of transportation it provides. These are called “road accessibility attributes”. Furthermore, an account must also be made for the speeds of transport which will depend on the type of road. For example, average speeds by cars on motorways (69mph) are considerably faster than those on roads on built up areas (30mph) (Department for Transport, 2009). It is extremely important to include a measure of these different speeds because it is possible that a route is shorter when travelled using minor roads but will almost certainly be quicker if major roads are used. This means that

---

1The question posed by Wiles and Costello (2000) was multiple-choice so the responses do not sum to 100%.

2Speed figures from Department for Transport (2009) are “free flow” speeds so do not take congestion into account. It is likely that actual average journey speeds will be much lower in practice.
agents with cars will have very different awareness spaces to those without cars, as stipulated by the crime literature (Brantingham and Brantingham, 1993), and this feature can only be included if the use of major and minor roads is appropriate for the transport available to the agents.

It follows that roads which are specified as A-roads, B-roads or Motorways (these are the road categories with the highest speed limits) will be designated “major roads” and cars will be able to move much more quickly along them in the model (the specific speed increases can be specified within individual scenarios). Table 5.1 provides the different road categories in the ITN layer and the attributes that apply to them in the model. Agents can walk along all roads except motorways, cannot drive down pedestrian or private roads, and have a speed increase when driving on major roads. Ideally additional information could be incorporated (e.g. to isolate A-roads that do not have a pedestrian footpath and cannot therefore be walked along) but these data are not available.

Table 5.1: Road types specified in the ITN layer and their mapping to model road types.

<table>
<thead>
<tr>
<th>ITM Road Type</th>
<th>Model attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>walk</td>
</tr>
<tr>
<td>Motorway</td>
<td>X</td>
</tr>
<tr>
<td>A-road</td>
<td>X</td>
</tr>
<tr>
<td>B-road</td>
<td>X</td>
</tr>
<tr>
<td>Local Street</td>
<td>X</td>
</tr>
<tr>
<td>Private Road - Publicly Accessible</td>
<td>X</td>
</tr>
<tr>
<td>Minor road</td>
<td>X</td>
</tr>
<tr>
<td>Pedestrianised Street</td>
<td>X</td>
</tr>
<tr>
<td>Private Road - Restricted Access</td>
<td>X</td>
</tr>
<tr>
<td>Alley</td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 5.3 maps the different attributes that each road has in Leeds. Motorways and arterial routes are clearly represented, along with roads that are inaccessible to cars such as alleyways.

### 5.3.2 Estimating Road Traffic Volume: Space Syntax

In Section 2.2 it was noted that houses which are obscured from the view of neighbours or passers-by are easier targets for burglars than those that are highly visible. Road traffic volume might be related to this as it dictates the number of passers-by. Although most evidence suggests that houses next to busy roads are a higher burglary risk because they are more likely to be known by potential burglars (Brantingham and Brantingham, 1993; Beavon et al., 1994), it is also possible that houses on busy roads are less of a risk at certain times of day because there are many passers-by to deter a potential burglar.

Estimating vehicle and pedestrian traffic volume can be accomplished through space syntax, which is a set of theories and techniques that explore the relationships between societies and the space they inhabit (Bafna, 2003). A space is converted into a graph consisting of interconnected nodes and edges, and a significant amount of information can be gathered by analysing these graphs (Bafna, 2003). The level of connectivity or “integration”, \( i \), of an edge, \( e \), is a measure of
Figure 5.3: Road accessibility attributes in city as a whole and the local area. Arterial routes and motorways that cars can drive quickly along are clearly distinguished (left) as well as alleyways and back streets that can only be walked on (right).

the number of edges that must be used to get from $e$ to every other edge in the graph. The final integration value is typically expressed as an inverse (Bafna, 2003) such that edges with higher integration are therefore the most accessible because, on average, they are the easiest to reach from any point in the graph. Integration can be calculated globally (where $i_e$ is relative to every other edge) or locally (where $i_e$ is relative to all edges within $n$ turns of $i_e$). For a full definition of integration see Hiller and Hanson (1984) or, for a more recent reflection, see Park (2005). Integration is relevant to this research because, interestingly, integration values in street networks have been found to correlate with the amount of pedestrian or vehicle traffic (van Nes, 2006) and therefore space syntax analysis of Leeds roads can be used as a proxy for the volume of traffic on the roads. Global integration is not appropriate as $i_e$ will be influenced by boundary effects; roads that are closer to the Leeds boundary will appear to be less connected simply as a result of the spatial limits of the roads data. Therefore local measures must be used and an appropriate value of $n$ must be found. Through trial-and-error, $n = 11$ appears to best distinguish between arterial routes and minor roads without biasing roads based their geographic location. Levels of integration for Leeds are illustrated both for a small area and the city as a whole in Figure 5.4. As one might expect, areas towards the city centre have the highest integration values and can expect the highest volume of traffic. Also, by observing the local neighbourhood, the integration value for a cul de sac is lower that for the main road it is connected to.

Once integration values are calculated it is possible to apply weights to each value depending on the time of day so that all roads will be busier during rush-hour, for example, than at other times. Figure 5.5 illustrates the generic relationship between time and traffic volume for the city. Ideally empirical data would be used to simulate traffic volume at different times but these data are not
Figure 5.4: Space syntax integration values for the entire city and a local area. High scores indicate that a road is well connected to the rest of the network and can therefore expect higher volumes of pedestrian and vehicle traffic.

available. Although commuter journeys are included in the census, the times of the journeys are not. Furthermore, Annual Average Daily Traffic Flows (AADF) data collected by the Department for Transport were also obtained and although these show the type and volume of vehicles on surveyed roads, again the time-of-day was not included. Although these values are not based on empirical evidence they are trivial to change in the model if opposing evidence becomes available. The final traffic volume on a road is calculated by multiplying the road’s normalised integration value by the current global rush hour value (i.e. 0.5 at midday).

5.4 Buildings

As discussed in Section 5.2, buildings in the virtual environment form the places that agents live and where they will burgle. Section 5.6 will present the virtual burglars and will note that burglars are also able to work, socialise and take drugs, so appropriate buildings for these activities must be identified as well as accessibility and visibility values for burglary. The following sections will outline how the above attributes can be calculated.

5.4.1 Type

To fulfil the needs of the agents, as will be outlined in Section 5.6, the following building types must be established from analysis of the MasterMap Topographic Area data:

- House – the buildings where agents live and can also burgle;
- Workplace – a building that provides employment for agents;
Social place – a building that can be used by agents to socialise;

Drug dealer – a place in which agents can purchase illicit drugs.

The model will focus on residential burglary from houses specifically. Blocks of flats are discounted because they are the scene of very few burglaries (see Section 4.4.4) and the process of burglary in a flat is very different to that of a house (e.g. normal measures of visibility such as trees or hedges obstructing views of ground-floor doors or windows do not apply). Therefore the first task is to distinguish houses from all other types of building.

Initially the MasterMap Topographic layer contains 2,234,189 objects in Leeds and 980,723 of these are classified as “buildings”. However, this description applies to any building structure such as garages, public toilets etc. To differentiate between different types of building the National Land Use Database (NLUD) code available in the MasterMap Address Layer can be used. For example, NLUD code U071 describes residential properties. The full list of NLUD codes was given in Table 4.2, Section 4.3.5. Extracting all buildings with codes U071, however, includes a large number of buildings that are not houses (some “houses” are less than 1m² in area) so further cleaning is necessary. A coarse cleaning approach would be to remove all properties which have extraordinary values for the square area. Following Shepherd (2006), Figure 5.6(a) illustrates a histogram of the log of the square area of all buildings. The natural logarithm is used to account for the wide range of values (Shepherd, 2006). Also, Figure 5.6(b) illustrates the same histogram just for the buildings which have been classified as “houses” from using Address Point data.

The two peaks in Figure 5.6(a) are likely to correspond to garages and houses respectively (Shepherd, 2006). Figure 5.6(b) shows that although small objects such as garages were successfully removed using the Address data, there are still some very large properties which are clearly not houses. This includes multi-use building types such as shopping centres with flats above them. Therefore a range of \( \exp(2.85) \approx 17m^2 \) to \( \exp(5) \approx 150m^2 \) (chosen because it represents 90% of the data around the median) should safely exclude all objects which are too large or too small to be
5.4. BUILDINGS

Figure 5.6: Histograms of the square area of different MasterMap objects (measure using the building footprint not floor-space) using a log scale x axis.

(a) All MasterMap Topographic Area objects classified as “buildings”.

(b) The subset of all buildings which can be classified as “houses” using MasterMap Address Point data.

preparing workplaces, social addresses and drug dealer addresses

A major advantage of the agent-based modelling approach as employed by this research is that it is able to simulate individual offender actions and therefore cleanly model the micro-level interactions that modern environmental criminology and “crime at place” theories are concerned with. However, this also leads to one of the most difficult aspects of this research – which is common to agent-based modelling in general (Torrens and Nara, 2007) – that of determining what these individual level-actions should be. For example, a powerful feature of this model is that it is able to simulate the individual awareness space of a burglar as it is built up over their daily routine activities and, therefore, accurately predict where the most likely place for a burglary is. The disadvantage with this approach, however, is that the researcher must first predict what these individual daily routines are and where they will be performed. Although some research suggests which industry a “typical burglar” might engage in – generally those that are low-paid and low-skilled (Wright and Decker, 1996) – there is no evidence that can support the assumption regarding where an agent will go to socialise. Furthermore, this will certainly vary across the agent population and will include friends’ houses as well as establishments like pubs or cinemas. Again, however, there is no data to support these decisions.

Therefore, in light of the lack of empirical evidence to guide our choice of workplaces and social addresses, broad assumptions must be made. At this stage, the National Land Use Database (NLUD) code is used to estimate where burglar agents might go to work or to socialise. It is estimated that workplaces include buildings with codes U091 (“Shops”) and U0101 – U103 (“Industry and Business”). Social places are likely to be U093 (“Restaurants and cafes”) and U094
chapter 5. creating virtual people and their virtual environment

("Public houses and bars"). It is important to note that none of these decisions are fixed, it is trivial to change work- and social-place addresses in the future as new data becomes available. The effect of these assumptions is mediated somewhat, however, because the actual addresses themselves are of limited importance. Their main influence over the model will be to affect which areas the agent is aware of and therefore where they might choose to burglar. If the individual address itself is inaccurate but the area is correct (i.e. in the real world the agent would work in the locality but not at the specific address) then the affect of the error is minimal as the agents’ awareness space would be similar regardless.

With drug dealer addresses a similar situation arises; it is difficult to estimate where a potential burglar will go to service a drug addiction. Here, drug dealer addresses will be estimated directly from the offender data as it is known where nominals who have been associated with a drug crime live. This will change depending on the scenario under investigation so full details will be provided in Section 7.3.

5.4.2 Accessibility

Burglary research discussed in Section 2.2.1 and evidence from Leeds provided in Section 4.4.4 suggests a link between accommodation type (detached, semi-detached, flats or terraced) and burglary risk. Detached houses tend to exhibit a higher burglary risk because (along with reduced visibility to neighbours which will be analysed in Section 5.4.3) they have a larger number of possible entry points. The buildings data can be analysed to determine how many other buildings a house is adjacent to and therefore how many possible entrances it is likely to have. This figure was calculated using the MapInfo GIS.3

From these data, it is possible to estimate how many entrances there are to the property. For example, a property which touches one other building can be assumed to be a semi-detached house with three possible entrances (one at each of the outer walls) whereas a building which touches two others can be assumed to be terraced with only two possible entrances. This measure will also pinpoint buildings which are on the corners of residential blocks as these have been identified as suffering a higher burglary risk than others (Brantingham and Brantingham, 1993; Robinson and Robinson, 1997).

As a result of the geographic analysis, some buildings have more than four neighbours (one house has 80!) which can be caused by unusual arrangements of buildings and when buildings such as garages or outhouses are adjacent. This can be mediated when the data are normalised by setting the upper bound of the normalisation to 4 (i.e. a maximum of four neighbours) and assigning all buildings with more than that to a normalised value of 1. Finally the data are “reversed” (transformed) such that new values are derived from old values using $x_{new} = (1 - x_{old}) + 1$. This

---

3The following SQL query was used to estimate the number of neighbours to each building:

```
SELECT tab1.TOID, COUNT(tab2.TOID) FROM buildings tab1, buildings tab2
WHERE tab1.geometry INTERSECTS tab2.geometry AND tab1.TOID <>
  tab2.TOID GROUP BY tab1.TOID
```

where TOID is a unique feature identifier in MasterMap.
makes the houses with the most neighbours the least accessible. A map of some example normalised accessibility values will be provided at the end of this section in Figure 5.9.

5.4.3 Visibility

The literature suggests that houses which are hidden from the view of their neighbours and passers-by suffer a higher burglary risk because a burglar can access the house without being seen. Although there is no available data on aspects such as the height of hedges surrounding properties or the presence of trees, it is hypothesised here that the size of the property’s garden and its isolation from nearby neighbours will form a proxy for visibility.

Size of Garden

Houses with large gardens might offer offenders an opportunity to gain access to the property without being seen, whereas small terraced houses with no gardens do not offer such hidden access. In the MasterMap Topographic Area data, gardens are described as “land” (although this is the same description as applied to parks and other types of area which causes some problems as discussed below). To estimate the size of a garden adjacent to a property, again using MapInfo, the square area of any “land” object that touches a building was calculated.\footnote{The following SQL query was used to estimate square area of land adjacent to a house:

```
SELECT tab1.TOID, Sum(Area(tab2.geometry)) FROM buildings tab1,
    TopographicArea tab2 WHERE tab1.geometry INTERSECTS tab2.geometry
AND tab2.Theme=land
```
}

Although this method is not entirely accurate because it can include a neighbour’s garden it is sufficient to summarise the large differences between detached houses with large gardens and terraced houses with little or no gardens. When normalising the data, it is, again, important to choose a suitable upper threshold. Without this threshold the data would be considerably skewed due to a small number of properties that have very large garden sizes which can occur when buildings border parks or other large areas of land. For example, two buildings next to Leeds Bradford Airport have apparent garden areas of more than $3 \times 10^6 \text{m}^2$! This would result in visibility values close to 0 for the vast majority of buildings in the model, with a small proportion having values of 1. Setting the upper bound for normalisation to three standard deviations from the mean (i.e. $\mu + 3\sigma$) and assigning every building with a garden size of greater than this to 1 prevents outliers from dominating the range of the measure. This is illustrated by histograms before and after normalisation in Figure 5.7

Degree of Isolation

For the second component of visibility, it is hypothesised that houses which are isolated from other buildings will be less visible and therefore easier to gain safe access to. To calculate the number of properties within a certain distance of a particular building, a buffer was constructed around it and the number of properties within the buffer was counted.\footnote{The following SQL query will count the number of buildings within a given buffer:

```sql
SELECT Count(*) FROM buildings WHERE geometry INTERSECTS buffer
```
}
CHAPTER 5. CREATING VIRTUAL PEOPLE AND THEIR VIRTUAL ENVIRONMENT

Buffers of sizes ranging from 50m to 200m were experimented with. Ultimately a buffer of 50m was chosen because this was found to best illustrate the large difference between densely-clustered houses in terraced-housing neighbourhoods and the sparsely distributed housing common to (semi) detached housing neighbourhoods.

Again normalisation must be performed with care so that a small number of buildings are not responsible for the entire range of the statistic. As with the size of garden, outliers with isolation values above $\mu + 3\sigma$ have a normalised value of 1. The distributions of the data before and after normalisation are illustrated by Figure 5.8. The final visibility value for each building is simply mean of its garden size and degree of isolation.

Figure 5.9 presents the normalised values for the number of neighbours, size of the garden and degree of isolation of each property. Although the geographical techniques are coarse and there are some errors (for example some terraced houses towards the north of the map have a larger number of neighbours than than should be expected) they are able to broadly distinguish between the different physical house attributes that will influence burglary.

5.5 Communities

Whereas the previous section outlined how to create individual virtual buildings, this section will discuss how the virtual communities can be formalised from the literature outlined in Section 2.2.2. As summarised in Section 5.2, there are four parameters for which values must be calculated. These are:

\[
\text{SELECT } \text{buffer.TOID, Count(buildings.TOID) FROM buffer, buildings WHERE buildings.geomery WITHIN buffer.geomery AND buffer.TOID <> buildings.TOID}
\]
5.5. COMMUNITIES

Figure 5.8: Degree of isolation (the number of buildings within 50m) histograms before and after normalisation.

Figure 5.9: Number of adjacent neighbours, size of garden and the number of neighbours within 50m (inverse of the ‘degree of isolation’). All normalised.
• **Collective efficacy** – a measure of how cohesive the community is;

• **Attractiveness** – a measure of how attractive the community is to a burglar, i.e. the availability of attractive goods;

• **Sociotype** – a description of the type of the community; potential burglars will be less comfortable in communities that are very different to their own;

• **Occupancy** – a measure of when houses are likely to be (un)occupied.

The following sections will outline how each of these values has been calculated using the available data outlined in Section 4.3.

### 5.5.1 Collective Efficacy

Collective efficacy can be defined as the collective belief among a group of people that they are able to succeed at a given task (Bandura, 1986). In a crime setting, Sampson et al. (1997, pg 918) defines collective efficacy as “social cohesion among neighbours combined with their willingness to intervene on behalf of the common good” and has found the phenomenon to correlate with low levels of violence.

The link between crime and a lack of community cohesion has been well established so community cohesion / collective efficacy (the terms will be given joint meaning here) must form part of the model. The value will indicate how likely it is that an offender thinks they will be confronted or noticed by passers-by or neighbours. For example, burglars have expressed a preference for communities where neighbours will not “look out for each other” (Wright and Decker, 1996). There are a number of different measures of community cohesion which can be used and the following are consistent in the literature: concentrated disadvantage; residential stability; and ethnic heterogeneity (Shaw and McKay, 1969; Sampson et al., 1997; Bernasco and Luykx, 2003; Browning et al., 2004). The following sections will outline how each variable can be calculated.

#### Concentrated Disadvantage

The Index of Multiple Deprivation (IMD) was outlined in Section 4.3.4 and is a comprehensive measure of deprivation. The only drawback with the index for this research is that it is published at the super output area (SOA) level, which is a larger geography than the output area (OA). Fortunately, however, SOAs and OAs share boundaries so it is possible, using a GIS, to estimate the deprivation of individual output areas by disaggregating SOAs. As with other data, the resulting values are normalised to the range 0 – 1.

#### Residential Stability

Measures of residential stability vary between authors. For example, Shepherd (2006) uses the number of people who have lived at the same address for one year and the proportion of working residents who work less than 2km from home whereas Bernasco and Luykx (2003) sum the number of people moving in and out of an area over the course of a year. Here, the percentage of
people who own their own homes is used as the measure of residential stability which corresponds with the work of Browning et al. (2004) and Sampson et al. (1997).

**Ethnic Heterogeneity**

Although Bernasco and Luykx (2003) use the index for qualitative variation (Agresti and Agresti, 1978), Hirschfield and Bowers (1997) and Shepherd (2006) have used the Index of Heterogeneity (Blau, 1977) which is also chosen for this study. Shepherd (2006) notes that an advantage of this approach is its ability to consider different ethnic groups rather than just two (e.g. white/non-white); see Massey and Denton (1988) for a comprehensive review. As defined by Hirschfield and Bowers (1997), the index or heterogeneity, $H$, is calculated as follows:

$$H = 1 - \sum P_i^2$$  \hspace{1cm} (5.2)

where $P_i$ is the proportion of the population in group $i$. The range of the statistic is 0 to 1 where high values indicate highly heterogeneous areas. Following Shepherd (2006), the 11 ethnic groups defined in the 2001 census were aggregated into five groups, as illustrated by Table 5.2 before calculating the proportions.

<table>
<thead>
<tr>
<th>Supergroup</th>
<th>Ethnicity (from 2001 Census)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>British, Irish</td>
</tr>
<tr>
<td>White Other</td>
<td>Other White</td>
</tr>
<tr>
<td>Black</td>
<td>Caribbean, African, Other Black, Mixed White and Caribbean, Mixed White and African</td>
</tr>
<tr>
<td>Asian</td>
<td>Indian, Pakistani, Bangladeshi, Other Asian, Mixed White and Asian</td>
</tr>
<tr>
<td>Chinese and Others</td>
<td>Chinese, Other Ethnic Group, Other Mixed</td>
</tr>
</tbody>
</table>

Table 5.2: Variables used for the Index of Heterogeneity, after Shepherd (2006).

**Overall Collective Efficacy**

As with other input data, each collective efficacy component must be normalised to the range 0 – 1 and correctly aligned (see Section 5.4.2) so that high values are less attractive to burglars (i.e. in this case have high collective efficacy). The final collective efficacy value is calculated as the mean of the three inputs. These, along with the overall collective efficacy value in all OAs in Leeds are
illustrated by Figure 5.10. As expected, the more deprived areas in the centre of the city are less cohesive (and more attractive to burglars) than the relatively affluent suburbs and rural areas on the outskirts of the city.

5.5.2 The Sociotype and Community Similarity

For the following sections, it is necessary to be able to summarise a community numerically in order to distinguish between communities that are demographically (dis)similar. As discussed in Section 4.3.3, the Output Area Classification (OAC) can be used to classify different areas using a variety of census variables. This classification makes it possible to refer to communities as a particular type, subsequently referred to as a sociotype. The name sociotype is used because it does not refer to a particular classification scheme; the model will be designed to work seamlessly with different classification systems such as Mosaic (Experian, 2007) for example. This is essential when the model is applied to Vancouver in Canada as Chapter 9 will discuss.

Using the OAC, a sociotype is not represented by a single variable but rather the combination of all 41 input variables chosen by Vickers (2006) to form the classification, i.e. a vector. In this manner, different aspects of a sociotype can be used to calculate different measures such as the attractiveness of a community or its expected occupancy levels. The manner in which this is accomplished will be outlined in the sections 5.5.3 and 5.5.4. An advantage of using a vector to represent sociotypes is that it makes it possible to compare one community to another. This can be accomplished by simply by calculating the Euclidean distance between the two sociotype vectors. Figure 5.11 illustrates this measure for four different output areas. These areas are chosen because they represent communities that are very different (based on the OAC and the crime data discussed in Section 4.4.1).

5.5.3 Attractiveness

For this research, household attractiveness is defined as a measure of the abundance of valuable goods in a property. Section 2.2.1 indicated that the crime literature is consistent in its support for the importance of attractiveness. Ideally this measure would be unique to each individual property, but in the absence of individual-level data census variables at the OA level must be used as a proxy. To measure the attractiveness of areas to burglars, Bernasco and Luykx (2003) use the percentage of owner-occupied dwellings and the average real estate value. Kongmuang (2006), on the other hand, found that (among other factors) the number of cars and the number of students in an area to be good indicators of attractiveness. Further evidence for variable selection comes in the form of the regression models utilised in Section 4.6 and analysis of the correlation of demographic variables in Section 4.4.5. On the whole the following four variables from the OAC represent the range of factors deemed important: the percentage of full time students; the mean number of rooms per household; the percentage of houses with more than two cars; and the percentage of people with higher education qualifications. Each of these values are normalised to the range 0 – 1 and the total area attractiveness is calculated as the mean of the four inputs.
5.5. COMMUNITIES

Figure 5.10: The measure of collective efficacy and its constituents: index of heterogeneity; the percentage of people who own their own home; the Index of Multiple Deprivation (IMD) score. Low values represent areas that are attractive to burglars (i.e. have low collective efficacy).
Figure 5.11: Using the Euclidean distance between OAC areas to illustrate community similarity.
5.5. COMMUNITIES

5.5.4 Occupancy

Many studies discussed in Section 2.2.1 have found that a burglar will not target a property if they are aware that the occupants are at home. Therefore it is essential to estimate the daily habits of residents to assess which houses are likely to be (un)occupied. As with the measure of attractiveness, the OAC can be used to estimate occupancy levels of resident populations at different times. Table 5.3 lists the relevant variables which will be used to estimate peoples’ daily habits. These variables were chosen because they broadly represent households with people who have work habits other than normal “9–5” jobs and those who have families. These are important for occupancy levels because, as noted in Section 2.2.1 burglary times often correspond to those that people are out at work or collecting children from school.

Table 5.3: Relevant variables from the Output Area Classification (Vickers et al., 2005) which can be used to estimate people’s daily work habits.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>v31 (f)</td>
<td>Students (full-time): Percentage of people aged 16-74 who are students</td>
<td>Students are more likely to be at home during the day and socialising in the evening (Robinson and Robinson, 1997).</td>
</tr>
<tr>
<td>v32 (g)</td>
<td>Unemployed: Percentage of economically active people aged 16-74 who are unemployed</td>
<td>We assume that unemployed people are more likely to be at home during the day.</td>
</tr>
<tr>
<td>v33 (h)</td>
<td>Working Part-time: Percentage of economically active people aged 16-74 who work part time</td>
<td>It is assumed that people who have part-time employed are more likely to be absent during work hours, but less likely than people who are employed full time.</td>
</tr>
<tr>
<td>v34 (i)</td>
<td>Economically Inactive Looking after Family: Percentage of economically inactive people aged 16-74 who are looking after the home</td>
<td>People who are looking after families and not working are more likely to have occupied properties during the day but unoccupied properties around school opening/closing times.</td>
</tr>
</tbody>
</table>

The next step in determining occupancy is establishing how the separate elements will be used to estimate the probability that a property is occupied. At this stage no account is taken for different week days although the model will be built with this extension in mind. The “probability of occupancy”, $P_{oc}$ can be calculated for each output area at time, $t$, as follows:

$$P_{oc} \propto N_{v31}f(t) + N_{v32}g(t) + N_{v33}h(t) + N_{v34}i(t)$$ (5.3)

where $N_x$ is the percentage of variable $x$ in the output area and the functions $f - i$ vary temporally and are defined in Figure 5.12. The shape of the functions $f - i$ are based on anecdotal evidence because, as with commuting times, there is no census data on which to support their design.
5.6 An Overview of the Virtual Burglars

A drawback with traditional modelling approaches, as discussed in Section 3.3, is that they struggle to account for the micro-interactions that lead to an individual crime event occurring and, from which, city-wide burglary rates ultimately emerge. Important concepts from environmental criminology – such as individual offender awareness spaces or the convergence in space/time of offenders and victims – cannot be included directly because the models work at an aggregate level, rather than at the level of the individual. This research addresses this drawback by incorporating individual house data where possible (as discussed earlier in this chapter) and by simulating the behaviour of individual burglars directly. The remainder of this chapter will discuss how the burglars can be simulated in a computer model, discussing how the literature and the available data have guided the agent design process.

5.7 Agent Architectures / Cognitive Frameworks

An agent’s architecture determines how the functionality of the agent is organised and how the agent replicates human or biological traits (Singh, 2005). A number of architectures (or cognitive frameworks) have been proposed to address how these traits should be mimicked; three relevant ones are outlined in the following sections.

It is possible not to use a published architecture at all and to, instead, program behaviour manually. This is the approach that has been taken by some of the agent-based crime studies outlined in Section 3.5 (e.g. Groff and Mazerolle, 2008; Hayslett-McCall et al., 2008). However, unless considerable care is taken over the design of the agents’ behaviour, a manually-defined approach is unlikely to offer the level of flexibility or the behavioural realism of a published framework.
For example, none of the research outlined in Section 3.5 allows truly dynamic agent behaviour; agents often have set routines and cannot change their behaviour as a result of changing internal/external conditions. To avoid these drawbacks and provide a more accurate representation of human behaviour, this research will build upon a published cognitive architecture. An interesting experiment therefore, would be to test the model with and without the presence of the cognitive framework. However, this would effectively require a complete re-design of the agents’ behaviour (if the framework were removed from the model there would be nothing to control the agents) so must be left for future work.

5.7.1 Beliefs Desires Intentions (BDI)

The BDI architecture (Bratman et al., 1988) is perhaps the most popular architecture and is centred around equipping agents with the mental components of beliefs, desires and intentions. The actor loop is a process which is used to determine how the agent will react to some input from the environment. This approach follows rational choice ideas because no action is performed without some form of deliberation (Balzer, 2000). The behaviour of a BDI agent is characterised by “practical reasoning”: goals are decided upon and then a plan is formed in order to satisfy the goals (Singh, 2005).

Beliefs represent the agent’s internal knowledge of the world. The agent has a “memory” of past experiences and the state of the environment as it was last seen. Desires are all the goals which the agent is trying to achieve. These can include short term goals such as “eat food” or more complex, long term goals such as “raise children”. As some goals might be contradictory, intentions represent the most important goals which the agent chooses to achieve first. Intentions are sometimes viewed as a subset of goals, while at other times they are viewed as the set of plans which will achieve the desired goals (Singh, 2005). Goals (and therefore intentions as well) will change throughout time depending on external inputs and the agent’s internal state. A level of caution can be integrated into a BDI agent by specifying how eager the agent is to change its intentions.

Although the BDI architecture has been widely used (Rao and Georgeff, 1995; Müller, 1998; Taylor et al., 2004; Brantingham et al., 2005b,a), it has also suffered some criticism. Fundamentally the architecture assumes rational decision making which is difficult to justify because people rarely meet the requirements of rational choice models (Axelrod, 1997). Also, Balzer (2000) notes that the core human elements (beliefs, desires and intentions) are difficult to observe directly. Access to them can only be achieved in a laboratory setting which might not relate to real situations (Balzer, 2000). Some criticise the three attitudes which form the core of the architecture (beliefs, desires and intentions) with being too restrictive and others with being overly complicated (Rao and Georgeff, 1995).

5.7.2 Behaviour Based Artificial Intelligence (BBAI)

BBAI is a modular behaviour architecture developed by Brooks (1986). It will be briefly outlined here because, as Section 3.5 discussed, it was used by another agent-based crime model (Birks
et al., 2008). The architecture was designed to control autonomous robots, but can equally be applied to software agents. Its basic structure consists of a number of hierarchical layers of increasing behavioural complexity. All layers act as individual controllers of the agent, they operate independently and simultaneously. Therefore it is the purpose of a ‘suppression mechanism’ to determine which layer should have overall control at a particular time. The advantage with this approach is that the agent can work towards different goals simultaneously, no early decision needs to be made about which goal to pursue (Brooks, 1986). Although having separate and autonomous layers provides robustness (if a high level fails the lower behavioural levels will take over the robot so that it continues to do something) and efficiency (there is no communication overhead between layers) a drawback is that a new layer must re-implement basic functionality that would otherwise be provided by lower layers. Therefore attempts to implement intelligence using BBAI have not proved as successful as alternative, hand-designed systems (Bryson, 2002).

Although the architecture might be appropriate for implementing simple behaviour in physical robots, virtual agents do not need such robust behaviour because the model can be fully tested and the agents’ environment has been wholly specified by the researcher. Therefore an agent will not encounter an unexpected object in the virtual environment which might cause a fault in a behavioural routine. The advantages of robustness offered by BBAI are thus counteracted by the difficulties required to implement complex “human-like” behaviour. As a consequence, architectures such as BDI which are specifically designed to model high levels of agent intelligence are more appropriate for this research.

5.7.3 PECS

Proposed by Schmidt (2000) and Urban (2000), the PECS architecture states that human behaviour can be modelled by taking into account a person’s physical conditions, emotional states, cognitive capabilities and social status. Personality is incorporated into the agents by adjusting the rate that internal state variables change and also how these changes are reflected in agent behaviour (Schmidt, 2002). The framework is modular, so that separate components control each aspect of the agent’s behaviour (Martinez-Miranda and Aldea, 2005). Proponents of PECS cite that as rational decision making is not required and the framework is not restricted to the factors of beliefs, desires and intentions (Schmidt, 2000), it is an improvement on the BDI architecture. Also, the framework is highly modular which makes it trivial to add or remove different types of behaviour as appropriate. For these reasons the framework appears to be the most appropriate for this research and the remainder of this section will outline PECS in detail, later illustrating how it can be used to create realistic burglar behaviour.

To illustrate the PECS features, an example proposed by Urban (2000) will be adapted. Consider a person in a shop who is thinking about purchasing some goods. They might experience physical needs (such as hunger), emotional states (such as surprise at the available goods), cognition (such as information about current prices) and social status (which will affect how the agent reacts to the shop assistant). Schmidt and Urban believe that all aspects of human behaviour can modelled using these components. In order to compare the strength of all the different types of
behaviour which might be acting upon a person simultaneously (such as the shopper), PECS uses the concept of “motives”. Intensity functions make it possible to compare motives from different behavioural systems (such as comparing the drive to eat food with the act of will of studying for an exam). The motive with the highest intensity becomes the “action guiding motive” as depicted in Figure 5.13. Once the action guiding motive is known the agent can behave accordingly, whether this is to instinctively react to a stimulus or create a complex action plan to pursue a constructive goal.

![Figure 5.13: Motives and motive selection, adapted from Schmidt and Schneider (2004)](image)

The PECS framework distributes all behaviours in to two major categories: reactive and deliberative. Reactive behaviour classifies actions which are largely instinctive, it can be modelled using a set of rules without deliberation on the part of the organism. The organism does not consider why it is behaving the way it is. For example, a reactive being is not aware that looking for food is a task which ultimately ensures survival. Deliberative behaviour, on the other hand, involves the conscious pursuit of goals. The organism is able to deliberate over its current goal(s), form action plans to satisfy a goal and break a larger goal into smaller sub-goals. Table 5.4 summarises the different types of behaviour as stipulated by Schmidt (2005) and also provides their intensity functions.

Table 5.4: Different types of PECS behaviour.

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reactive Behaviour</strong></td>
<td></td>
</tr>
<tr>
<td>Instinctive behaviour</td>
<td>An automatic reaction to a stimulus such as a parent reacting instinctively to a child’s cry. Instinctive behaviour can be modelled relatively easily using pre-defined rules which are called up in certain circumstances.</td>
</tr>
<tr>
<td>Learned behaviour</td>
<td>Similar to instinctive behaviour but with rules that are learnt dynamically. Schmidt (2000) cites the example of a car driver who will instinctively brake if they see a child running across the road.</td>
</tr>
</tbody>
</table>
CHAPTER 5. CREATING VIRTUAL PEOPLE AND THEIR VIRTUAL ENVIRONMENT

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Description</th>
</tr>
</thead>
</table>
| Drive controlled behaviour | This type of behaviour is directed by internal drives to satisfy needs. These range from basic needs required to preserve life (such as the need for food or safety) to social needs and finally to intellectual needs. Schmidt defines the function, $f$, to determine drive intensity, $T$, as  

$$ T = f(N, E, X) $$

(5.4)  

where: $N$ is the agent’s personal preference for the need; $E$ represents environmental influences; and $X$ denotes other influences. For example, a drug addict will have a strong drive to take drugs if the need, $N$, is high because they have gone without drugs for some time. However, the environment, $E$, must also be taken into account: the drive might be strong if they are surrounded by other addicts who are also using drugs even if the need, $N$, is not great. |

| Emotionally controlled behaviour | Emotions are similar to drives because, if they are strong enough, they will affect the behaviour of the agent. Unlike drives, however, they are stimulated externally, not by internal state changes. Schmidt notes that the intensity of emotions, $E$, are very hard to model, but defines the following formula:  

$$ E = g(I, A, X) $$

(5.5)  

where $I$ represents the importance of the event which has generated the emotion, $A$ is the agent’s personal assessment of the event and $X$ represents other influences. |

<table>
<thead>
<tr>
<th>Deliberative Behaviour</th>
<th></th>
</tr>
</thead>
</table>
| Constructive behaviour      | Schmidt (2000) discusses how an organism which is able to perform constructive behaviour is able to build an internal representation of its environment and also construct and deliberate over plans of action which should allow it to satisfy goals. Goals assembled in this manner are associated with acts of will, the organism “wants” to achieve the goal (Schmidt, 2000). In a similar fashion to reactive forms of behaviour which have a “need” associated with them, constructive behaviours have an “importance” attached to them by the agent which will influence their intensity. For example, one agent might attach a higher importance to the pursuit of gaining knowledge than another. In addition, the closer a goal is to completion, the higher the will associated with the goal. Schmidt defines the following function, $h$, to calculate will intensity, $W$:  

$$ W = h(I, D, X) $$

(5.6)  

where $I$ is the importance of the goal, $D$ is the distance from completing the goal and $X$ are other influences. |

| Reflective behaviour        | Representing the highest level of behaviour, the ability to exhibit reflective behaviour is reserved for human beings above all other organisms. Reflective action relates to the ability to monitor and control one’s own thought processes. Also, in addition to a model of their environment, reflective organisms have a model of self which can lead to the most advanced forms of emotion such as inferiority complex and jealousy (Schmidt, 2000). To model this type of behaviour Schmidt states that, within its cognitive module, the PECS agent will have another entire PECS model of itself. |

At the time of writing, documented use of the PECS framework is limited, especially when compared to other behavioural models such as BDI. However, the few studies that were found originate from diverse fields. For example, PECS has been used by Ammar et al. (2006) and Neji and Ammar (2007) to build emotions into a virtual learning environment. The authors incorporate non-verbal communication in the form of emotional facial expressions to improve the relationship between a human learner and a computer-controlled tutor. In the field of health care, Brailsford and Schmidt (2003) used the framework to improve a simulation of disease screening. The au-
Adapting PECS for this research

As already discussed, Schmidt (2000) notes that it not necessary to model the entire spectrum of human behaviour. However, a fine balance must be reached. The greater the degree of abstraction of a model, the greater the differences will be between the model and the real system (Schmidt, 2000). For this study, it is decided that modelling deliberative behaviour brings additional complexity which is unnecessary. At this stage, the behavioural focus is on creating realistic daily behaviour patterns that cause the potential burglars to travel around the city at sensible times, going to realistic places and building up accurate awareness spaces. It is not necessary to build a more complex behavioural model that includes forms of deliberative behaviour because, at this stage, the outcomes would be the same; the agents would still travel to the same addresses and perform burglary in the same way, although they will be more aware of themselves and the motivations behind their actions. Only when the model itself is much more complex and includes a greater variety of behaviour (such as developing long-term life plans for example) does deliberative behaviour have to be modelled. To simulate the level of human behaviour required for this model, therefore, the focus is on drive controlled behaviour as relatively simple drives can be used to determine agents’ current actions.

However, the following sections will demonstrate that the agent behaviour is actually more complex than simple reactive behaviours stipulated by PECS allow for so the framework has been adapted somewhat. For example, the use of awareness spaces are consistent with learned behaviour (agents remember the neighbourhoods that they have visited which influences future choices for where to look for burglary targets) and, when burgling, agents are involved in the “conscious pursuit of goals” (Schmidt, 2000) (i.e. obtaining money to satisfy a drug addiction) which is deliberative behaviour. So whilst the agents do not know why they need to satisfy their goals (a reactive trait) the methods they use to satisfy them are complex and involve the conscious pursuit of goals with intermediate stages (a deliberative trait).

5.8 State Variables and Motives

As the previous section discussed, the basis for the agent behaviour is built around the PECS concept of drive controlled behaviour. Agents have a number of motives which vary in strength; the strongest motive at a particular time drives the agents’ behaviour. The intensity of the motives can be calculated using intensity functions which are based on some internal factor as well as external influences. The internal factors which influence motive intensities are termed state variables. For example, an agent’s level of energy could be a state variable that affects their hunger motive.

The first decision to make, therefore, is which state variables should be included. As suggested by crime theories in Section 2.4 and empirical evidence in Section 4.5, burglars commonly
become aware of potential burglary victims either by actively searching, or by passing them on journeys that are otherwise unrelated to burglary. Therefore it must be decided what the legitimate (non-burglary related) behaviours should be (the search itself will form part of the burglary process discussed in Section 5.10). Again from the crime literature, the most prevalent requirements put upon potential burglars are the need to sustain a drug addiction or to maintain “high living” (i.e. socialising). Also, most burglars had, at some time, legitimate employment. Therefore, the following activities are the most important to a virtual burglar: drug taking, socialising and working. Clearly this is a vast simplification of the needs that affect a real person, but at this stage they are sufficient to create realistic daily behaviour. This is still an improvement upon the existing agent-based crime models outlined in Section 3.5; none can account for such a variety of agent behaviour. In addition, Chapter 6, which outlines model development, will demonstrate that the model will be designed with flexibility in mind to make it trivial to add additional behaviours in the future. With these behaviours in mind, the following state variables are sufficient:

- **Drugs** – the level of drugs in an agent’s system;
- **Sleep** – a measure of the amount of sleep an agent has had;
- **Social** – a measure of how much the agent has socialised.

These state variables can be considered to be internal levels, such as an energy level. For example, in the same way that a person with a low energy level will be hungry, an agent with a low drugs/sleep/social level will want to behave in such a way as to increase the value of the state variable by taking drugs, sleeping, or socialising respectively. The value of state variables will theoretically be in the range $0 – \infty$ but, in practice, will range from $0 – 1$ because values above 1 will lead to very low motive intensities and the agent will not attempt to increase the state variable value further (discussed below).

PECS (and indeed criminology literature) indicate that internal variables alone are insufficient as external influences will affect behaviour. To account for this, the motive associated with each state variable can include external influences and, through the use of intensity functions, motives can be compared to each other thus establishing what the agent’s current action should be. The external influences are specific to each motive and will be discussed in detail below. In general, the intensity of a motive, $m$, is inversely proportional to the size of its state variable, $s$. Also, the population of agents does not need to be homogeneous (this is one of the benefits of agent-based modelling) so different agents can be affected by state variables and motives differently. This feature is incorporated into the model by including a **personal parameter**, $p$, that affects motive intensities such that:

$$m \propto p \frac{1}{s}$$

For example, an agent with a large value for $p$ will be more strongly affected by a particular motive than an agent with a low value for $p$ even if both agents have the same state variable level. This can be used to change the importance that agents place on particular behaviours, indeed if
\( p = 0 \) then \( m \) will always be zero and the agent will be unaffected by \( s \). In the absence of external influences or personal preferences, Figure 5.14 depicts how motive intensity varies with state variable size. An exponential function is used so that as \( s \to 0 \), \( m \to \infty \) and subsequently motives with the lowest associated state variable value are likely to be the strongest.

The levels of various state variables will deteriorate over time. Schmidt (2000) provides an example of an agent whose “energy” state variable deteriorates at different rates depending on the action which the agent is performing (some actions require greater energy expenditure than others). At present, this feature is not deemed appropriate for the model because the types of needs being modelled (drugs, socialising and sleeping) will not deteriorate at noticeably different rates. Therefore each need will deteriorate by one unit over the course of a simulated day. Section 7.3 will explore this assumption in greater detail by varying the rate at which state variables deteriorate. The following sections will discuss each behaviour (state variable and motive) in more detail.

Also, Figure 5.15 graphically illustrates how state variable levels are combined with personal preferences and external factors to determine the strongest motive (termed the action-guiding motive).

### 5.8.1 Drugs

Burglary is commonly a response to a drug addiction (see Section 2.3), so drug addiction must form part of the model. In particular, the process of going to purchase drugs will have important influences on the agent’s cognitive map, making areas with large numbers of drug dealers more susceptible to burglary.

Drug taking, in this instance, is clearly a drive-regulated behaviour (see Section 5.7.3). The drive is dependant on the current drug level in the agent’s system. In contrast to the general form of drive-regulated behaviour, the motive associated with drug use does not depend on other factors such as the surrounding environment (more on this point below). The strength of an agent’s drug motive, \( d_m \), can be calculated from their personal preference for drugs, \( p_d \) and the durgs state variable level, \( s_d \):

\[
d_m = p_d \times s_d
\]
Figure 5.15: How state variables, $s$, personal preferences, $p$ and external factors (e.g. the time of day, $t$) are used in intensity functions to determine the action-guiding motive. In this example, the agent’s social level is very low (the agent has not socialised in some time) and this is the strongest motive.

$$m_d = \frac{P_d}{s_d} \quad (5.8)$$

A clear avenue for future work is to include environmental influences, i.e. the effects of drug addiction might be more strongly felt if drug use is taking place nearby. Schmidt (2000), for example, notes that a person might feel hunger more strongly if they smell cooking food.

5.8.2 Sleep

All agents must sleep for, on average, eight hours per day which is a generally accepted healthy amount (Tune, 1968; National Sleep Foundation, 2009). The time of day also affects the strength of the motive so that the desire to sleep is stronger at night than during the day. The shape of this function is illustrated by Figure 5.16(a). Although this type of sleep pattern might reflect most people’s habits, it should be noted that the chaotic lifestyles of many potential burglars might result in very different sleep patterns. However, Section 6.4.2 will demonstrate that the irregularity of burglary results in much more chaotic behavioural patterns for burglar agents than those of non-burglars (who can make money with regular employment) even though the effects of the time of day on the sleep motive are not irregular.

In general, an agent’s sleep motive intensity, $m_s$, can be calculated from the sleep state variable level, $s_s$, the time of day function, $f(t)$ and their personal preference for sleep, $p_s$, as follows:
5.8. STATE VARIABLES AND MOTIVES

5.8.3 Social

All agents have the need to socialise which includes visiting friend’s houses, pubs etc. The choice of areas which the agent is likely to visit is very important because it will strongly influence the size and shape of the agent’s cognitive map. As with sleeping, the social motive \( m_{soc} \) is dependant on the time of day, \( g(t) \), the agent’s personal preference for socialising, \( p_{soc} \), and the size of the social state variable, \( s_{soc} \):

\[
m_{soc} = \left( \frac{1}{s_{soc}} \right) \left( \frac{g(t) + p_{soc}}{2} \right)
\]

(5.10)

The shape of \( g(t) \) is illustrated in Figure 5.16(b). Agents will desire to socialise in the evenings.
more so than during the day.

## 5.9 The Repertoire of Actions

Each motive has an associated goal which the agent can accomplish to lower the intensity of the motive. These goals are also actions (i.e. an agent satisfies their sleep goal by performing a sleep action) and are outlined in Table 5.5

<table>
<thead>
<tr>
<th>Motive</th>
<th>Goal</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drugs</td>
<td>Purchase Drugs</td>
<td>Drugs can be purchased from specific locations assuming the agent has enough wealth.</td>
</tr>
<tr>
<td>Sleep</td>
<td>Go To Sleep</td>
<td>The agent can go to sleep if they are at home.</td>
</tr>
<tr>
<td>Social</td>
<td>Socialise</td>
<td>The agent can socialise at certain locations assuming they have enough wealth.</td>
</tr>
</tbody>
</table>

To satisfy a goal, there are often numerous actions (or “sub-goals”) that must be accomplished first. For example, to sleep, an agent first needs to go home. Also with socialising and buying drugs, wealth is required that can either be sought through legitimate employment or through burglary. Flowcharts illustrating how goals can be accomplished via the use of sub-actions are presented in Figure 5.17.

Figure 5.17: Actions to satisfy goals. Sleeping simply requires the agent to go home first, whereas purchasing drugs or socialising require wealth to be generated first.
5.10 The Process of Burglary

Finding a property to burgle is the most intricate of all the agents’ actions (it consists of numerous goals and sub-goals) and is also the most important as it will have the greatest effect on final city-wide burglary patterns. This section will discuss how findings from criminology research have been formalised in the burglary action and how the elements of the environment discussed at the beginning of this chapter affect the burglars’ decisions. The burglary process is broken into three distinct parts:

1. Deciding where to start looking for victims;
2. Searching for a victim;
3. Deciding upon a suitable target.

The approach is designed to be modular so that different types of burglary can be simulated simply by replacing the relevant part (e.g. replacing one target search routine for another). Section 5.11 will discuss this in more detail. The main drawback with this approach is that it does not allow for purely opportunistic burglary which, as Section 4.5 noted, is common. For example, using the above scheme it is not possible for an agent to notice an open door or window and choose to burgle imminently even if they were not otherwise considering burglary (i.e. travelling to work). They have to make a conscious decision to start the burglary process regardless of their surrounding environment. Including more opportunistic types of burglary is recommended as an avenue for future research. The remainder of this section will discuss the three parts of the “burglary template” in more detail.

5.10.1 Decide where to start searching

Burglars act as “optimal foragers” (Johnson and Bowers, 2004; Bernasco and Nieuwbeerta, 2005) when they choose target areas because their decision is based on an analysis of potential rewards against risks. The model here works in a similar manner. When deciding where to start searching, agents consider the communities that they are aware of and assign a likelihood, \( l \), to every area, \( a \), in their cognitive map, relative to their home, \( h \), and current location, \( c \):

\[
l_a = w_1 \cdot \frac{1}{\text{dist}(c,a)} + w_2 \cdot \text{attractiveness}(h,a) + w_3 \cdot \text{socialDiff}(h,a) + w_4 \cdot \text{prevSucc}(a)
\]  

(5.11)

where

- \( \text{dist}(c,a) \) represents the distance (in travel time) to the target from the agent’s current position. Research has shown that agents are unlikely to travel far from their homes, so farther areas are less attractive.\(^6\)

\(^6\text{When calculating distance, if } c = a \text{ (the agent is examining the area in which they are currently situated) the distance, } d, \text{ is set to the average distance from the centre of the community to any point assuming the community is a}\)
- **attractiveness** represents the attractiveness (i.e., the abundance of attractive goods) of the potential target relative to the agent’s home. Using a relative attractiveness allows for the finding that affluent communities are at most risk of burglary when they are close to deprived communities. Determining community attractiveness was the subject of Section 5.5.3.

- **socialDiff** represents the difference between the sociotypes of the agent and the potential target (where values of 1 indicate similarity and 0 indicates dissimilarity) as offenders are more likely to target areas which they know well and feel safe in. Calculating sociotype difference was discussed in Section 5.5.2.

- **preSucc** is the number of previous successful burglaries which the agent has committed in a: numerous successful burglaries are likely to encourage the offender to return to the same area. This “near repeat” phenomenon has been found by numerous crime research (Townesley et al., 2003; Johnson, 2007) and also informs burglary-reduction policies in Leeds.

An extremely important element of Equation 5.11 are the variables $w_1 - w_4$. These are weights which will influence how important the different factors are to the different burglars. In this manner, different types of burglar can be created based on the values of the weights – a subject that will be discussed in more detail in Section 5.11. For example, a “professional” might be unconcerned with the distance that they need to travel (a low value for $w_1$) compared to a “chaotic” burglar who is desperate to satisfy a serious drug addiction (a high value for $w_1$).

Before applying the weights and calculating $l$, each variable must be normalised so that they all have a fair affect on the overall likelihood. Then to determine the area which the offender will travel to, a roulette wheel selection is used (following Heppenstall, 2004) so that the decision is probabilistic. An agent calculates $l$ for every area, $a$, in their memory and then chooses an area such that the probability of being chosen, $P(a)$, is proportional to its share of the sum of all $l$:

$$P(a) = l_a / \sum_{i=0}^{n} l_i$$

(5.14)

This probabilistic component makes the decision more realistic; it is unlikely that a human will always reach an identical decision even when faced with similar input information.

---

Perfect circle with area, $A$:

$$d = \frac{1}{2} \sqrt{\frac{A}{\pi}}$$

(5.12)

This is to take account of the fact that houses in larger areas are effectively further away than those in smaller areas. Although the measure is not entirely accurate ($d$ is equivalent to half the radius of the circle and there are a larger number of points with $d > \frac{1}{2}r$ than $d < \frac{1}{2}r$ which will increase the average distance) it is close enough to encapsulate the fact that houses in larger communities (as measured by community area) are effectively further away from the agent than those in smaller communities.

7Relative attractiveness is 0 if the target area is less attractive than the agent’s home area or the difference in attractiveness between the two if the target area is more attractive. I.e.:

$$\text{attractiveness}(h, a) = \begin{cases} 0 & \text{if } \text{attract}(h) \geq \text{attract}(a) \\ \text{attract}(a) - \text{attract}(h) & \text{if } \text{attract}(a) > \text{attract}(h) \end{cases}$$

(5.13)
5.10. **THE PROCESS OF BURGLARY**

5.10.2 **Search for a victim**

Research has shown that burglars do not search randomly for burglary targets, they exhibit identifiable search patterns (Johnson and Bowers, 2004; Brantingham and Tita, 2006). For example, Rengert (1996) identifies, among others, the tear-drop and bulls-eye patterns. The bulls-eye pattern suggests that the burglar’s search expands outwards from their home (or another anchor point) such that houses close to the start of the search have the largest risk. This is comparable to marauder behaviour (see Section 4.5.2). The tear-drop pattern, on the other hand, stipulates that the journey to the start of the search is included so that the overall search pattern looks like a tear-drop connecting the home and the location that marks the start of the search. This is comparable to commuter behaviour. The search implemented in the model combines both of these ideas. As the previous section discussed, the offender chooses an area to start their search. They then travel to their chosen location and start a bulls-eye like search. On the route to their chosen start location the offender observes properties for burglary suitability (discussed in the following section) so that if the chosen area is close to their starting position then the final search pattern will resemble a bulls-eye, otherwise it will resemble a tear-drop.

The actual bulls eye search is accomplished by orbiting the start location. When choosing a direction of travel at a road junction, roads which form angles of 90° and 270° to the original target are more likely to be chosen than those which move the agent directly towards (0°) or away from (180°) the target. Once the offender reaches a suitable area and begins a search, the amount of time spent on the search depends on the characteristics of the offender, with desperate offenders usually finding a target relatively quickly (Rengert, 1992). If an agent has not found a target within a set amount of time (specified by the *search time* variable) the burglary process is repeated; the agent chooses a new start location, travels there and begins the search again. The *search time* variable can be altered in the model.

5.10.3 **Choose a suitable victim**

Once the agent decides to commit a burglary, they start to examine all the houses they pass in order to determine their suitability for burglary. This happens on the way to the target location as well as while they are actually performing the search, as stipulated in the literature. Table 5.6 illustrates the variables that determine the *suitability* of each property.

Overall household suitability is then calculated by summing the individual components and dividing by the sum of the weights:

\[
suitability = \frac{w_5 \times CE + w_6 \times TV + w_7 \times Occ + w_8 \times Acc + w_9 \times Vis + w_{10} \times Sec}{\sum_{i=5}^{10} w_i}
\]  

(5.15)

The weights \(w_5 - w_{10}\) are used to distinguish between different types of burglar behaviour in a similar manner to those applied to the decision regarding where to start the search (see Section 5.10.1). For example, a “professional” burglar might be less deterred by security than an
Table 5.6: Variables that determine a burglar’s assessment of household burglary suitability.

<table>
<thead>
<tr>
<th>Environment variable</th>
<th>How it affects burglars’ behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collective efficacy ((CE))</td>
<td>High levels of (CE) make the area less attractive to burglars because the community appears cohesive and neighbours / passers-by are likely to notice someone acting suspiciously.</td>
</tr>
<tr>
<td>Traffic volume ((TV))</td>
<td>High levels of traffic volume make the houses on a road less attractive because it is difficult to access a property without being seen by passers-by.</td>
</tr>
<tr>
<td>Occupancy ((Occ))</td>
<td>Houses which are likely to be occupied are less attractive.</td>
</tr>
<tr>
<td>Accessibility ((Acc))</td>
<td>Houses with few possible entrances are more difficult to enter.</td>
</tr>
<tr>
<td>Visibility ((Vis))</td>
<td>Houses which are highly visible to neighbours / passers-by are more difficult to enter without being seen by others.</td>
</tr>
<tr>
<td>Security ((Sec))</td>
<td>High levels of security are often a deterrent.</td>
</tr>
</tbody>
</table>

“amateur” and will subsequently have a lower value for \(w_{10}\) (this discussed in more detail in the following section). The suitability is normalised in the range 0 – 1 by dividing by the sum of the weights \(w_i\) (this is possible because the input variables are guaranteed to fall in the range 0 – 1 as explained in the following section). Therefore the most suitable properties will have a value near 0 and the least suitable near 1.

Equation 5.15 will return an absolute value for the suitability of a house for a particular agent. The final step is to determine whether or not the agent is desperate enough to commit the burglary. “Desperateness” is based on the intensity of the motive which is currently driving the agent’s behaviour: if the suitability value of a property is higher (i.e. less suitable) than the intensity of the motive then the agent will not attempt a burglary.

If, however, the suitability value is lower than the agent’s motive intensity then the agent might attempt a burglary. A random element is included such that the agent is more likely to burgle if the difference between the suitability of the house and the agent’s motive intensity is large. This is important because otherwise the agent’s decision becomes somewhat deterministic and they are also likely to burgle the first house they find once they become active (i.e. their next-door neighbour) which is not realistic. Crime studies have shown that although burglars are unlikely to travel far, they will not burgle too close to home for fear of
being recognised (Rengert et al., 1999). An exponential gradient \( y = x^2 \) is chosen so that with only small differences between an agent’s motive intensity and house suitability they are unlikely to burglar.

5.10.4 The Effects of Burglary on Surrounding Properties

When a burglary occurs, the security of the surrounding buildings will change in response to the burglary. In Leeds, for example, there are crime-reduction initiatives that provide local residents with information following a burglary in their neighbourhood. Thus it might be expected that near residents will become aware of the risks of burglary and increase their security following a burglary in their locality. This can be simulated by varying security levels dynamically as illustrated by Figure 5.19. Also, over time residents will become complacent of the risk and security levels will decrease to base levels. The rate at which this degradation occurs is specific to individual scenarios and will be explored in more detail in Chapter 7.

5.11 Different Types of Burglar

The previous sections have illustrated that although each burglar uses the same “burglary template”, by altering certain weight variables it is possible to represent a wide range of burglar
behaviour and, ultimately, simulate different “types” of burglar and burglary behaviour. Although this feature of the model is not used in later experiments, it is described briefly here because it provides strong evidence in support for the choice of methodology and is a very powerful feature of the model. Aggregate techniques traditionally employed in crime, as discussed in Section 3.3, could not incorporate this level of behavioural complexity and simulate, at an individual level, the effects of heterogeneous burglar behaviour on city-wide crime rates. A brief literature review that provides a background to the different types of burglar behaviour will precede the technical details regarding how the different types can be accounted for.

Although there is a large body of literature that attempts to classify criminal propensity such as Foxe (1939), there is less research that attempts to classify burglar behaviour. Herbert and Hyde (1985) found that most burglaries were committed by low income, amateur offenders, defined as “offenders originating from low-income homes themselves who typically offend in an opportunist way against properties of a similar category for limited returns.” This is consistent with the literature already discussed in Section 2.3, the data provided in Section 4.5 and is also echoed by other authors – e.g. Maguire and Bennett (1982) and Davidson (1981) who found that the majority of crimes were not committed by in a planned, systematic way by “professionals”.

Therefore an obvious classification of burglary behaviour divides it into two classes: amateurs and professionals. However, being an “amateur” does not necessarily mean that no planning goes into a burglary. Mawby (2001) suggests that most qualitative studies have found that burglaries are not usually purely opportunistic. For example, Bennett and Wright (1984) differentiate between “opportunistic”, “search” (an offender becomes active and then searches for an immediate victim) and “planned” (the offender returns to a previously found opportunity at a later time/date) offences and argue that over half the offences were planned. Similarly, Cromwell et al. (1991) proposes the labels “professional”, “journeyman” and “novice” when referring to burglar types. It is worth noting, however, that it is highly probable that when interviewed after the event, a burglar might rationalise the burglary to a much greater extent than they did at the time (Mawby, 2001).

To support the crime literature, and gain an insight into the specific behaviour found in Leeds, local crime-reduction practitioners at Safer Leeds were interviewed. They suggested that the different types of burglar behaviour were very similar to those already outlined in the literature. Table 5.7 provides the types recommended by Safer Leeds – and supported by the literature – and indicates how they can be implemented using the burglary template. The behaviour classifications provide an example of the type of behaviour that can be modelled, but clearly they could be improved with additional empirical evidence which is recommended for future work.

Note that with the chaotic and travelling chaotic types, many of the weight values that determine household suitability are low (indicated by a ‘↓’ symbol). Therefore these types are more likely to find a given house suitable than the professional and opportunist types which reflects the influence of drugs (the chaotics are made more desperate by their drug addiction and are therefore more likely to burgle houses that would otherwise seem unsuitable).
### Table 5.7: A classification system for burglar types. Illustrates how the different environmental variables affect an agent’s burglary decisions (where to start searching and what makes an individual property suitable).

**Key:**
- ↑ High deterrent: large weight value (will strongly influence agent’s decision).
- → Not important for the agent: average weight value (will affect decision but not strongly)
- ↓ Low deterrent: small weight value (unlikely to influence agent’s decision)

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Start Search</th>
<th>Determining suitability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chaotic</td>
<td>An opportunist who is desperate to generate wealth by any means possible (i.e. through burglary or other types of crime) in order to fund a drug addiction. This high level of drug addiction is reflected by chaotic lifestyle. Often unemployed and with limited means of transport, they are unlikely to travel far to search for a victim. Also will want to avoid confrontation.</td>
<td>↑ ↑ ↑ ↓</td>
<td>↓ ↑ ↓ ↓ ↓ - ↓</td>
</tr>
<tr>
<td>Local Opportunist</td>
<td>Again an opportunist but will not commit crime if it is too difficult (they will be deterred by security and guardianship). Not necessarily a “full time” burglar, they are often younger, do not need to support a drug habit, occasionally have legitimate employment and will only burgle if a good opportunity presents itself. Their limited access to transport means that they will only travel short distances.</td>
<td>↑ - ↑ ↑</td>
<td>↑ ↑ ↑ ↑ ↑ ↑</td>
</tr>
<tr>
<td>Travelling chaotic</td>
<td>Similar to the chaotic type but more aware of the best opportunities and will travel further to reach them. For example, travelling to an insecure car park to commit theft from a motor vehicle. Otherwise very similar to the chaotic type.</td>
<td>- - - -</td>
<td>↓ ↑ ↓ ↓ - ↓</td>
</tr>
<tr>
<td>Organised / Professional</td>
<td>The most organised type, there is a considerable amount of planning involved in a burglary. The type is characterised by low levels of drug addiction, occasional legitimate employment, skills to evade security precautions and will travel the furthest distances for the greatest rewards. Confident in ability to “blend in” with community and tackle security precautions.</td>
<td>↑ - - ↓</td>
<td>↓ - - - ↓ -</td>
</tr>
</tbody>
</table>
5.12 Burglary Summary

This section will summarise the burglary process and clearly illustrate how the findings from the criminology literature have driven the design of the agents, the environment, and the activity of burglary. For clarity, Figure 5.20 illustrates diagrammatically how the environmental variables influence the different aspects of an agent’s burglary decision process. Also, Figure 5.21 illustrates how the overall search commences in time and space for a real agent in a real model.

**Agents’ Burglary Decision Process**

1. Attractiveness
2. Social difference
3. Previous successes
4. Distance

1. Collective Efficacy (community)
2. Occupancy levels (community)
3. Accessibility
4. Visibility
5. Security
6. Traffic volume (road)

**Agents’ Thought Process**

Communities in the Agent’s Cognitive Map

Objects in the Environment

Figure 5.20: The burglars’ decision process for burglary and the environmental variables which affect it. Note that weights can be applied to each environmental variable in order to account for different types of burglar behaviour.

To show how the quantitative and qualitative criminology research has been accounted for in the model, Table 5.8 outlines the findings from the literature and notes, explicitly, how these will be incorporated into the model to provide a sound theoretical foundation.

Table 5.8: How the motives of potential burglars and their responses to environmental cues will be implemented in the model.

<table>
<thead>
<tr>
<th>Behaviour / Motive</th>
<th>Implementation in model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for money is the primary reason for burglary (Repetto, 1974; Bennett and Wright, 1984; Rengert and Wasilchick, 1985; Cromwell et al., 1991; Wright and Decker, 1996; Bernasco and Luykx, 2003; Nee and Meenaghan, 2006) and usually to buy drugs (Scarr, 1973; Cromwell et al., 1991; Hearnden and Magill, 2003) or maintain “high living” (Scarr, 1973; Wright and Decker, 1996, 2005).</td>
<td>Agents in the model burgle to satisfy the desire to socialise or take drugs; these cost money. High desires to socialise or high drug addictions can be modelled by increasing the associated motive’s personal preference or by increasing the rate at which the state variables decay.</td>
</tr>
<tr>
<td>The decision to burgle is made away from the actual crime scene and the potential offender then travels to a target noted previously (Wright and Decker, 1996; Hearnden and Magill, 2003; Nee and Meenaghan, 2006).</td>
<td>Agents build a cognitive map of their environment as they travel regardless of their reason for traveling. Targets are later chosen from these known areas.</td>
</tr>
</tbody>
</table>
5.12. BURGLARY SUMMARY

<table>
<thead>
<tr>
<th>Behaviour / Motive</th>
<th>Implementation in model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Few burglars can be classed as “opportunistic” although most interviewees will</td>
<td>During the journey to their chosen target, agents examine properties which they pass and</td>
</tr>
<tr>
<td>alter their usual routine if a particularly attractive target presents itself</td>
<td>will commit a burglary if the target is deemed suitable.</td>
</tr>
<tr>
<td>(Nee and Meenaghan, 2006).</td>
<td></td>
</tr>
<tr>
<td>The expected “yield” is the most important consideration when selecting a target</td>
<td>Potential burglars choose to travel to the area with the largest abundance of attractive</td>
</tr>
<tr>
<td>(Hearden and Magill, 2003; Nee and Meenaghan, 2006).</td>
<td>goods that they are aware of.</td>
</tr>
<tr>
<td>Burglars will not usually enter occupied properties (Cromwell et al., 1991;</td>
<td>Occupancy has been estimated at the neighbourhood level and acts as a deterrent.</td>
</tr>
<tr>
<td>Wright and Decker, 1996; Nee and Meenaghan, 2006).</td>
<td></td>
</tr>
<tr>
<td>Most burglars will return to previously burgled properties, usually because they</td>
<td>Previous success is remembered by the agent and makes them more likely to travel back to</td>
</tr>
<tr>
<td>know what goods are available and how to enter the property. (Wright and Decker,</td>
<td>that area in the future.</td>
</tr>
<tr>
<td>Properties close to the burglar’s home are more likely to become victims (Snook,</td>
<td>Properties close to a burglar agent’s home are more likely to form part of the agent’s</td>
</tr>
<tr>
<td>2004; Bernasco and Nieuwbeerta, 2005).</td>
<td>cognitive map and are therefore a higher burglary risk. Also, the distance from the</td>
</tr>
<tr>
<td>This is partly because the offender knows the area well and does not need to</td>
<td>agents current location is taken into account when deciding where to start a burglary</td>
</tr>
<tr>
<td>carry stolen objects too far (Hearden and Magill, 2003) and also because the</td>
<td>search.</td>
</tr>
<tr>
<td>potential burglar chooses targets from within their cognitive awareness space</td>
<td></td>
</tr>
<tr>
<td>(Bernasco and Nieuwbeerta, 2005).</td>
<td></td>
</tr>
<tr>
<td>Suitable targets are often found by passing them on their routine activities</td>
<td>The agent’s cognitive map is built up from their routine activities and a target is</td>
</tr>
<tr>
<td>(Wright and Decker, 1996; Cromwell and Olson, 2005)</td>
<td>chosen from these known properties.</td>
</tr>
</tbody>
</table>

To conclude the chapter, the following briefly summarises how the crime theories, outlined in Section 2.4, have been accounted for in the model design:

- **Routine activity theory** (Cohen and Felson, 1979) emphasises that, for a crime to occur, victims converge with offenders in a time and place that is void of capable guardians. Furthermore, these convergences come about as a result of individual people’s routine activities. The model accounts for this firstly by modelling individual burglars and allowing them to perform typical daily routines. The impacts of capable guardians have been estimated at the level of the community by predicting whether or not the occupants of a property are home and how likely neighbours are to intervene if they see suspicious behaviour (i.e. acting as capable guardians of their neighbour’s property).

- **Crime Pattern Theory** (Brantingham and Brantingham, 1981a) stipulate that crimes are most likely to occur in places in which an offender’s awareness space (the areas that they know well) overlap with crime opportunities. In the model, agents use cognitive maps to remember the areas that they have passed through, whether on legitimate or illegitimate business. They then choose burglary targets from within this awareness space and can favour areas that they have visited more often. Also, awareness spaces can develop differently depending on the method of travel that an agent is using. For example, if an agent is on a train it could be hypothesised that they will be less likely to build up a clear, detailed picture of an area compared to when they walk through it (see, for example, Brantingham and
Figure 5.21: An offender burglary space/time visualisation. Time is in the vertical axis and space is represented by the horizontal axes. The burglar starts at home (yellow events), then searches for a burglary location (green events), travelling home once the burglary has been completed (blue events).

GeoTime software used courtesy of Oculus Info Inc. All GeoTime rights reserved.
Another aspect of crime pattern theory is the concept of anchor points or nodes which are important addresses to a person that influence where they travel to and subsequently the areas that they know well. In the model, anchor points are represented by the agents’ home locations, their drug dealers, their places of employment and the places in which they go to socialise. Finally, crime attractors (areas that attract offenders for the purposes of committing crime) are also accounted for in the model design. These will include areas with large numbers of drug dealers (burglar agents are attracted to such areas to buy drugs) and areas with concentrations of highly attractive houses (for the purposes of burglary).

- The Rational Choice Perspective (Clarke and Cornish, 1985) views burglars as rational beings who make a cost-benefit analysis whether or not to commit a crime. In the model this is accounted for directly because agents make an explicit decision whether or not to commit a burglary, weighing up the suitability of the target house against the strength of their desire for the money that will come from the burglary. The bounded nature of the burglars’ rationality is incorporated through the absence of global knowledge of the environment (the agent might not know, for example, that there is a much more suitable house a short distance away) and through a random element to their burglary decision.

Furthermore, all crime theories are concerned with an individual offender, located at a particular place, with a specific motivation and in the presence of other individual people. Of all the methodologies reviewed in Chapters 2 and 3, agent-based modelling is the only one that is able to truly capture these individual-level dynamics which are ultimately responsible for the city-wide crime patterns. The design of the model exemplifies this approach.

### 5.13 Summary

This chapter has outlined the design of a model that is able to incorporate many of the features from the environmental criminology literature outlined in Chapter 2 and the results of data analysis in Chapter 4. It uses the agent-based modelling methodology which, as the result of an analysis of current modelling practice in Chapter 3, was found to be the most accurate method of representing the dynamics of the residential burglary system. The chapter began by discussing how the virtual environment would be constructed from the available data. There are three layers included that represent the environmental backcloth for the potential burglars, these are buildings, communities and the road network. Wherever possible, individual-level data has been used to build up a picture of the environment, as demonstrated by the analysis of household accessibility and visibility in Section 5.4. Where individual-level data is not available (such as estimations of household occupancy patterns to represent the presence of capable guardians as stipulated by Routine Activity Theory), the community layer would be used to estimate these factors from census data.

The second half of the chapter then discussed the design of the burglar agents. An advanced cognitive framework called PECS was used to improve the behavioural realism of the burglars compared to the other agent-based crime models reviewed in Chapter 2. The design of the agents
allows them to behave dynamically by setting goals in response to changing internal or external conditions and subsequently plan behaviour that will ultimately lead to the satisfaction of these goals. The chapter concluded by discussing the process of burglary. It was shown that by using internal agent variables it is possible to implement different types of burglar as discussed in the crime literature and following conversations with crime-reduction experts at Safer Leeds. For all these reasons, the model design represents the most comprehensive burglary model published to date.
Chapter 6

Model Development and Testing

Contents

6.1 Introduction .................................................. 129
6.2 Tools for Building Agent-Based Models ......................... 130
6.3 Programming Overview ....................................... 130
6.4 The Prototype Model ....................................... 131
6.5 The Repast Simphony Model ................................. 139
6.6 Verifying the Model – Experiments with Idealised Data .......... 147
6.7 Model Development – Summary .............................. 168

6.1 Introduction

Building on a review of criminological and modelling literature in Chapters 2 and 3, and an exploratory data analysis in Chapter 4, Chapter 5 outlined a theoretical agent-based model that could be used to simulate the residential burglary system. The model was designed to incorporate current criminological research using a methodology that is relatively unexplored in the field but is the most adequately suited to modelling the dynamic, individual-level interactions that characterise the burglary system. This chapter will forward the research by describing the process of building the model and testing it with experiments on idealised data in an idealised environment. It will begin by outlining the different tools and platforms that can assist with building an agent-based model in Section 6.2, introduce the concept of object-oriented computer programming in Section 6.3 and then discuss a prototype model and the full model in Sections 6.4 and 6.5. The final part of the chapter, Section 6.6, presents the first sets of experiments, using idealised data, that are designed to ensure the model has been constructed correctly and test the effects of the numerous parameters on the behaviour of the agents. This paves the way for the model to perform experiments using real data in Chapter 7 and finally make real-world crime forecasts in Chapters 8 and 9.
6.2 Tools for Building Agent-Based Models

Early agent-based models were often developed independently using conventional programming languages (Gilbert and Bankes, 2002). This approach has the disadvantage that common elements such as graphic libraries, algorithms and analysis tools must be repeatedly re-implemented by developers working on different models. Furthermore, this approach creates a boundary for researchers who want to build ABMs but do not possess the necessary programming skills (Railsback et al., 2006). To address these problems, pieces of computer software have emerged that can provide common functions for model developers or entire graphical model building interfaces that require no programming at all. These software systems, which aid the building of ABMs, can be referred to as tools, platforms or frameworks. Possible tools that could have been chosen for this research include Ascape (Parker, 2001; Inchiosa and Parker, 2002), the Agent Building and Learning Environment (ABLE) (Alphaworks, 2009), MASON (Luke et al., 2004), Modelling4All (University of Oxford, 2009), Swarm (Minar et al., 1996; Macal and North, 2005, 2007) and Agent Analyst (The Redlands Institute, 2009) (which was the tool used by Groff and Mazerolle (2008) to build an agent-based crime model as discussed in Section 3.5). Although the tools noted above are very powerful, they are often difficult to learn to use. For early prototyping, a tool is required that allows for quick model development and has a shallow learning curve. NetLogo (Wilensky, 1999) is an example of such a tool which is highly recommended for prototyping (Railsback et al., 2006) and has been chosen for the early stages of the project which are outlined in Section 6.4.

Although well suited for prototyping, NetLogo is too restrictive to be used to develop the final model. At the time of development the only popular tool which is capable of working with spatial data, other than Agent Analyst, is the Recursive Porous Agent Simulation Toolkit (Repast). Tobias and Hofmann (2004) reviewed four toolkits (Repast, Swarm, Quicksilver and VSEit) and found that Repast was the most suitable framework for “the applied modeling of social interventions based on theories and data”. Also, unlike Agent Analyst, Repast is not tied to any proprietary software and does not require the Windows operating system which are essential characteristics for the model to be able to utilise a high-performance computer network (as discussed in Section 6.5.6). Therefore Repast Simphony, the latest version of the software, will be used for the majority of the model development.

6.3 Programming Overview

Before outlining the model development process in more detail, an introduction to computer programming is required to explain some fundamental concepts that have guided the methodology. A computer program is a set of instructions which a computer follows to perform a given task. The idea of storing instructions to be read by a machine actually predates the first computers; textile looms in the 18th century were controlled by instructions on punched cards. But it was not until much later that this concept was applied to computing by Charles Babbage (1791 – 1871) with the design of his analytical engine. Unfortunately his design was never fully constructed and it was not until the 1940s that digital machines were built, recognisable as the ancestors of modern
6.4 THE PROTOTYPE MODEL

Computers.

The programming languages used by early computer programs are known as “1st generation” or “machine” languages. They are instructions which are executed by the computer processor directly and are written in the lowest form of language a computer understands: a stream of 0s and 1s (also termed “bits”). Obviously these programs were extremely difficult for humans to create and understand so, in the 1950s, 2nd generation “assembly” languages began to emerge that were easier for humans to use. They were less verbose, providing simple functions for programmers that would otherwise require large amounts of machine code. In the late 1950s, efforts to further improve the flexibility and reusability of computer code lead to the implementation of 3rd generation languages such as FORTRAN. Most commonly used applications today are written using 3rd generation languages. Along with making languages less verbose, an aim of 3rd generation languages was to make programming more accessible. The use of “methods” or “functions” in computer languages made it possible to design small collections of code that performed specific tasks and could be re-used. Rather than simply running from the first line of a program to the last, the flow of control was able to jump into functions and back again. This makes it easier to organise a program and reduces the need to repeat sections of identical code thus reducing program size, potential errors and increasing usability.

A further advancement, which has become the dominant paradigm in use today (Horstmann and Cornell, 2007), is object-oriented programming. Object-orientation means that program code is grouped into separate “objects”, each of which can contain any number of separate methods. Each object can be considered a separate entity; it has its own state. For example, a number of objects could be created to represent different people in a model, each will have similar characteristics (such as age or gender) but different values for their parameters. As such, objects of the same type can be heterogeneous although their underlying structure is the same. In complex systems (such as the burglary system) it is the interactions between the individual components that are of particular importance. Object-oriented programming forces the researcher to explicitly think about and define the relations that will be present in the model (O’Sullivan, 2004), making it particularly suited to the modelling of complex systems. For a full description of Java (the programming language used in this research) and object-oriented programming in general, the reader is directed to Horstmann and Cornell (2007).

6.4 The Prototype Model

Chapter 5 outlined a framework for the model, describing how the burglar agents and the environment could be implemented. Importantly, the PECS behaviour framework was introduced and the intensity functions which will be used to drive the agent’s behaviour were also presented. Before the main model development takes place using Repast Simphony, however, a prototype model was created for two main reasons:

1. To determine whether or not the PECS framework is able to control the agents appropriately;
2. To provide a practical demonstration to stakeholders of the general aims of the project (it can be difficult to visualise what an agent-based burglary model will look like).

Although this section will outline the model in sufficient detail, more information and additional experiments can be found in Malleson et al. (2010b) (which is included in Appendix C) and Malleson and Brantingham (2009).

### 6.4.1 Prototype Overview

The prototype consists of PECS agents who are placed in a regular grid environment (more details about the environment will follow the description of the agents). The agents use roads to travel between different addresses and can traverse one square per model iteration. With regards to timing, one model iteration is defined as three minutes so that it will take agents between 10 and 60 minutes to travel to work depending on where an agent lives. Therefore there are 20 iterations per hour and 480 in a day.

**Agents**

In the prototype model, the agents have two fundamental needs: the need to generate wealth and the need to sleep. Wealth represents commitments that require money for satisfaction, such as the need to buy food, socialise or sustain a drug addiction. There are two different types of agent who have the same basic structure but differing access to legitimate employment: “non-burglars” (also called “citizens” for clarity) are always able to generate wealth through working whereas “potential burglars” do not usually have sufficient employment to generate the wealth they need and they must subsequently turn to burglary. The burglars are assigned random amounts of work each day from a random uniform distribution between 0 (no work) and 8 hours (a full day’s work). The mechanisms that agents use to determine actions from the levels of need are illustrated in Figure 6.1. Each need has a level associated with it (i.e. the amount of wealth an agent owns). These levels are fed into PECS intensity functions which, along with other parameters such as the time of day, determine how intense the need is. The most intense need leads to an appropriate goal and the “action planner” is used to plan how to achieve this goal.

![Figure 6.1: How agent’s needs can lead to actions. Taken from Malleson et al. (2010b).](image-url)
6.4. THE PROTOTYPE MODEL

Levels of wealth and sleep deteriorate at a constant rate such that, on average, an agent needs to work for approximately 8 hours per day (or commit a single burglary) and sleep for approximately 10 hours per day. Using this work/burglary/sleep configuration it is possible to create behaviour which can be generally found in the daily patterns of employed people in most cities. Obviously this is a vast simplification of real human behaviour but it will suffice for the prototype whose aim is largely to act as a proof-of-concept for the PECS structures.

An important feature of the model is the inclusion of agents’ “cognitive maps”. These maps are unique to each agent and contain each property the agent has travelled past at some point in the simulation. The agents use their cognitive maps as the basis for their decision about where to commit a burglary (i.e. they travel to the most attractive targets first). Houses on routes that are travelled by many burglar agents are therefore at a higher risk of burglary because they exist in the activity spaces of a large number of burglars. This aspect is prevalent in criminology theory (Brantingham and Brantingham, 1993) but difficult to account for using traditional modelling methods where individual daily routines cannot be incorporated.

Environment

The environment is constructed using a 41*31 cell grid, illustrated in Figure 6.2. It consists of a central business district surrounded by residential areas. In this respect the model imitates part of the concentric ring model (Burgess, 1925), although it is possible to incorporate communities that are distributed in a less orderly fashion (see Malleson et al. (2010b) for examples).

![Figure 6.2: The environment used in the NetLogo prototype.](image)

Some squares are “empty” and play no part in the model because agents do not move diagonally, only horizontally or vertically. There are three types of cell in the environment: the commercial district, roads and residential properties. Each cell in the commercial district represents a single office that provides employment for an unlimited number of agents. The residential properties house the agents (maximum of one agent per house) and also act as burglary targets (one cell represents one house). Upon initialisation, each agent is randomly assigned a home address.
and also the address of their work place which will be within the commercial district.

The houses have two defining characteristics when it comes to burglary: security and attractiveness. Both these variables have default values of 4 although they can be varied to generate different types of community. The levels are also dynamic: if a burglary is committed the attractiveness of the victimised property increases by 5 and surrounding properties also receive an increase equal to 5 minus their distance\(^1\) from the burgled property. The security of the burgled property also increases by 10. The additional levels of security and attractiveness degrade at a rate of 1 unit per day. These number have been established through trial-and-error and provide the right balance between stopping the agents from repeatedly burgling the same house but also allowing high-crime areas to develop. Full sensitivity testing of the prototype model parameters is not necessary because the prototype is not designed to be used as a complete tool; just as a proof-of-concept and a demonstration of what the complete model will provide. Once the model has progressed from prototyping stages thorough experiments will be used to determine appropriate parameter values and how they effect the model in Section 6.6.

The process of burglary in the prototype is similar to that outlined in Chapter 5 although there are fewer variables to influence the agents’ decision. The process is as follows:

1. The agents use a roulette wheel selection to decide which building, from their cognitive map, they will travel to in search of a burglary target. This mechanism works by assigning a probability of being chosen based on the attractiveness of the buildings whereby the larger the attractiveness the larger the probability.

2. While travelling they observe every building they pass and determine how suitable it is for burglary. Suitability is based on the sum of two functions which use the agent’s level of wealth and the security of the building. Figure 6.3 illustrates these functions. If a building is deemed suitable they will commit a burglary. The only exception to this rule is that burglars will never burgle occupied properties.

3. If the agent reaches the target building and still has not found a suitable property the process is repeated.

### 6.4.2 Proof-of-Concept Experiments

To illustrate the potential of the prototype (and therefore the advanced model that will follow), two experiments will be performed. These experiments have been published in Malleson et al. (2010b) so the following will serve as a summary. In both experiments 300 agents are created, 5% of which are burglars and the rest citizens.

**Control Experiment**

The aim of the control experiment is to determine whether or not the model is robust and that the results are sensible. Default values for security and attractiveness of properties are used for every

---

\(^1\)In the prototype model one unit of distance is defined as the length of a cell so adjacent cells are 1 unit apart.
6.4. THE PROTOTYPE MODEL

Function used to determine building suitability for burglary

Figure 6.3: Determining building suitability in the prototype model. The final decision is based on a combination of the agent’s level of wealth and the security of the potential target building. As the agent’s level of wealth and the security of the properties they pass decrease, the suitability of the properties increases. Appropriate values for these parameters have been established through trial and error but will undergo proper systematic testing when the main model has been developed.

The first task is to determine how long it takes the simulation to reach equilibrium. Here equilibrium is defined as the model state when the overall crime distribution is constant. There will be small local fluctuations in crime because, while the model is running, burglaries will still be occurring. But once equilibrium has been reached the overall distribution of crime will not change if the model was allowed to continue indefinitely.

Figure 6.4 illustrates the burglary locations and their mean centre at different time points in a typical simulation. There is evidence that dynamic equilibrium has been reached by day 50 because the mean centre of the burglary distributions is the same on days 40–50 (the end of the simulation) as it is for the sum of all burglaries. Usually more evidence should be provided for the basis of equilibrium, but this is not necessary yet; Section 7.3 will discuss how to determine whether or not equilibrium has been reached in more detail. It is more important, at this prototyping stage, to determine whether or not the PECS functions are stable. Figure 6.5 graphs the levels of wealth and sleep along with the size of their associated needs over time for a typical burglar and citizen. Because the citizen is able to generate sufficient wealth through legitimate employment their needs for wealth and sleep follow a regular pattern which is largely dictated
by the time of day. The burglar, on the other hand, produces a considerably less regular pattern, owing to the irregular availability of wealth resultant on the randomisation of working patterns. This is, however, consistent with our understanding that the most burglars are thought to live extremely chaotic lifestyles that driven by needs and opportunities as they present themselves (see Section 2.4.4). This is also a finding that is recognised by practitioners at Safer Leeds. These graphs continue with similar patterns for an indeterminate period of time (the range of the x axes on the graphs could be grossly increased but the pattern would remain the same). Therefore it is safe to conclude that the PECS behaviours for citizens and burglars are stable.

Figure 6.4: The burglary locations in the prototype control experiment and their mean centre. Taken from Malleson et al. (2010b).

![Figure 6.4: The burglary locations in the prototype control experiment and their mean centre.](image)

Figure 6.5: How wealth and sleep can vary over simulated time for a typical burglar and citizen.

![Figure 6.5: How wealth and sleep can vary over simulated time for a typical burglar and citizen.](image)

Interestingly, burglaries are generally concentrated in the residential areas that border the com-
6.4. **THE PROTOTYPE MODEL**

Commercial area. This is a finding supported by Crime Pattern Theory (Brantingham and Brantingham, 1993), where the authors discuss that crime is likely to occur along the borders between areas that are noticeably different.

The results of the control experiment are not surprising. Burglaries are committed close to the commercial area because the potential burglars are more likely to have passed through the area on their route to work. They are therefore more aware of the burglary possibilities offered by those houses and are subsequently more likely to burgle there. This is a successful result because the aim of the control experiment was to establish that the model performs as would be expected. More importantly, the model demonstrates that PECS is able to successfully control simple agents, producing different daily patterns depending on the abundance of wealth they are able to generate. Further experiments can now be performed.

**Target Hardening**

The following will illustrate the power of this type of model by performing a simple experiment which has practical, real world applications. Again the details can be found in Malleson et al. (2010b) and will therefore only be summarised here. Target hardening is a burglary reduction scheme whereby houses that are deemed to have a high risk of burglary are offered additional protection. This protection often takes the form of free security hardware such as door/window gates or new locks. In reality, there are generally two methods that agencies use to determine which properties should be offered the extra security. In the first, which is commonly used in practice (Byron, 2003), the most “vulnerable” properties are targeted. Vulnerable people can include new and repeat burglary victims, the elderly, single parents and those renting private houses. To simulate this strategy in the model, vulnerable properties are defined as those with the highest number of burglaries. An alternative strategy, which is less commonly used in practice, offers extra security to all properties in a given area or street, not just individual ones. To simulate this in the model, all buildings in a given area will have their security increased. The aim of the target hardening experiment, therefore, is to establish which of these methods is better at removing a crime hotspot.

To prepare the environment for the experiment, the attractiveness and security of some properties in particular areas were altered to create different types of community as illustrated by Table 6.1. The examples here include a student area, a deprived area and a wealthy area. Malleson et al. (2010b) describe how a range of different environment configurations were used but all led to the same result: a crime hotspot develops in the student area. Therefore, when simulating the effects of target hardening an entire area rather than single properties, houses in the student area will be hardened. Figure 6.6 demonstrates the spatial locations of the different types of community that will be used in the target hardening experiment.

The two target hardening strategies begin on day 20 (so that crime hotspots have established first) and hardening consists of increasing a property’s security value by 4 units. In the block hardening scheme most of the student area is hardened simultaneously on day 20. With the individual targeting scheme the most heavily victimised property (or a randomly chosen one if there
Table 6.1: The change from the default value for attractiveness and security variables associated with different community types.

<table>
<thead>
<tr>
<th>Type of Area</th>
<th>Percentage Change from default value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attractiveness</td>
</tr>
<tr>
<td>Normal</td>
<td>-</td>
</tr>
<tr>
<td>Wealthy</td>
<td>150%</td>
</tr>
<tr>
<td>Deprived</td>
<td>50%</td>
</tr>
<tr>
<td>Student</td>
<td>150%</td>
</tr>
</tbody>
</table>

Figure 6.6: The spatial location of different types of community, created by altering the security and attractiveness of properties. Taken from Malleson et al. (2010b).

Figure 6.7 illustrates the results of the two strategies, showing burglary counts and clusters at different points in the simulation. For more information about the clustering algorithm used (Nearest Neighbour Hierarchical) see Malleson et al. (2010a). It is clear that the individual targeting scheme fails to remove the hotspot around the student area, a high crime rate develops early in the simulation and remains constant throughout. In the second simulation, however, target hardening the entire student area is able to create a large drop in crime. Although a cluster remains around the student area, it is smaller and has moved south. Also, a new cluster has begun to develop outside the student area. These facts suggest that the student area is considerably less heavily burgled and that, if the clustering algorithm were configured differently, the cluster might be removed entirely.

6.4.3 Summary - The NetLogo Prototype

Before work began on the main model, a prototype was created. The aims of the prototype were to test whether or not the PECS behavioural framework could suitably control virtual agents and to provide a demonstration of the final model to stakeholders. The prototype was successful in that the agents’ behaviour was sensible and it was able to demonstrate predictive power via experiments. This has laid the groundwork for the full model which will be introduced in the following
6.5. THE REPAST SIMPHONY MODEL

The prototype demonstrated that the PECS behavioural framework was able to correctly control simple agents, both “burglars” and “citizens”. This has paved the way for development to begin on the main burglary model. As outlined in Section 6.2, the model will be developed using Repast Simphony. This section will discuss the Simphony platform in more detail and outline how it will be used to build the model.

6.5.1 Contexts and Projections

In previous versions of Repast (and as is common with other toolkits) it was the responsibility of the programmer to create and organise groups of agents manually. Furthermore, if agents inhabited
a space they usually required internal variables to record their current location. Arguably the most significant development made by Repast Simphony is a new conceptualism for how agents and their space should be organised: using “contexts” and “projections”. As Howe et al. (2006) discuss, a “context” is used to store a population of agents and can be thought of as a “soup” of agents. It allows a population to be defined but does not provide a mechanism to give agents the concepts of space or relationships. “Projections” are then used to define the relationships between the agents. Many projections can be defined for the same context to represent the different types of relations that might exist between agents. For example, to describe agents in a city a spatial projection could be used to define each agent’s location is space and a network projection could be used to define social relationships between the agents. The objects which represent agents can therefore be built without any knowledge of the contexts or projections in which they might later be situated. This is conceptually a fundamental shift in how agents are related to their space. As opposed to accessing or changing an agent’s location through methods provided by the agent (effectively asking the agent where it is or who its friends are), now an agent’s location is queried or changed through methods provided by the projection. Figure 6.8 illustrates how contexts and sub-contexts could be used to hold different agents types in a virtual city.

Figure 6.8: Example organisation of Repast Simphony contexts and projections. A main context contains all the agents who live in the city and two sub-contexts hold different types of agent (victims and burglars in this case). The main context has a spatial projection which gives the agents a spatial location and a network projection which defines a relationship between the agents. There is also a separate projection applied to the burglars’ sub-context which could, for example, represent the people they choose to socialise with.
6.5.2 Burglar Agents

Recall that Section 6.3 introduced the idea of an “object” in computer programming terms and Section 5.6 discussed how burglar agents would be created utilising criminology theory and the PECS behavioural framework. To realise the theory, therefore, a Burglar object is used to represent the agents themselves. Each agent has a number of internal StateVariable objects which lead to Motives and then to Actions as discussed in Section 5.7.3. Therefore, heterogeneity can be introduced into the population through the creation of agents who have different state variables (e.g. some agents can suffer drug addiction whilst others do not) or through the strength of the state variables (e.g. some agents suffer drug addiction more strongly than others). Burglary (and all other actions specified in Section 5.9) are represented as Action objects. Again this allows for heterogeneity as different burglary templates (Brantingham and Brantingham, 1993) can be implemented as different Burglary action objects. Figure 6.9 illustrates the organisation of these different types of object graphically.

6.5.3 Constructing the city: Environments, Layers and Projections

Fully understanding the rules which drive an agent-based model is essential. As Elffers and van Baal (2008) note, these efforts might be hampered by the presence of a complex geography because the environment is likely to be an integral part of the rules which drive the behaviour. For example, the physical locations of houses will influence where an agent travels and, subsequently, where they later choose to burgle. Therefore it is not sufficient to experiment with a single geography; the attractiveness and density of potential burglary victims must be varied spatially by using multiple different geographies. Ideally, in fact, the rules should be tested first in the absence of any geography at all so that the non-spatial elements of the model can be tested in isolation of a spatial component. To this end, an Environment object is used which the agents can interact with to move around and query their environment. As long as different Environment contain the same methods, the agents in the model do not require any information about the type of environment that they exist in.

Three different environment types of increasing complexity are implemented. These include a one-dimensional null environment, a regular grid environment (similar to a cellular automata) and a realistic spatial GIS environment. These will be discussed in greater detail when they are used to test the model in Section 6.6.

Although the agents are unaware of the intricacies of Simphony contexts and projections (discussed in Section 6.5.1) due to the use of the Environment object, there is nonetheless a complex hierarchy underpinning the different types of environment. Figure 6.10 illustrates the different layers which make up the virtual city as a whole. Using these layers, the agents are able to find out everything they need to know about the environment they are in, without actually having to know which type the environment is. For example, if an agent wants to travel to a particular destination, they do not need to know whether they are travelling along a realistic vector road network or a hypothetical road on a regular grid, they simply ask the environment how to travel.
Two Burglar objects. Both have State Variable objects that represent Sleeping and Socialising, but Burglar 2 also has a drug addiction.

State Variable objects that have associated motives

Motive objects. Each motive will have different actions that might need to be performed before they can be satisfied. E.g. the Socialise and Drugs motive both require burglary if the agent does not have enough money whereas the Sleep motive only requires the agent to travel home first.

Figure 6.9: Example of the organisation of Burglar objects with State Variables, Motives and Actions.
6.5. **THE REPAST SIMPHONY MODEL**

Figure 6.10: The different environment layers which make up the virtual city and the Repast Simphony contexts/projections that underpin them. The network projection of junctions is used by the agents to plan routes from one location to another along roads (as discussed in Section 6.5.4).

### 6.5.4 Moving agents around the Environment: Routing

To maintain the layer of abstraction between Simphony Projections and the agents, moving agents around an environment is accomplished through the use of `Route` objects which can work with any type of environment. However, the means of calculating an agent’s route varies depending on the type of environment as follows.

#### Null Routes

When the “null” environment is being used, the agents do not actually move around the environment. The environment simply records the amount of time an agent has to spend travelling and reports the agent as having reached its destination after an allotted amount of time. As Section 6.6.2 will illustrate, this variable can be varied to see how it will affect the behaviour of the agents. This allows greater flexibility for testing the behaviour of agents without the complexities of differing journey times or the spatial clustering of burglary opportunities.
Grid Routes

Routing in a grid environment is relatively simple as the agent’s movements are restricted to a set of discrete cells: those defined as roads. The Repast ShortestPath functions, which use Dijkstra’s algorithm (Dijkstra, 1959), can be used to find the list of road cells which make up the path and the agent simply moves along one cell per iteration.

GIS Routes

Routing in a GIS environment is somewhat more complicated because, unlike on a regular grid where roads are made up of adjacent cells, roads vary in length and are unlikely to be straight lines. For example, Figure 6.11 illustrates a potential route between two points A and C. Moving between points A and B is relatively straightforward; the agent can move directly towards the next junction on the route. However, to move between points B and C the agent must pass over a number of individual road segments. Furthermore, unlike grid environments where an agent will always finish an iteration on a cell, it is extremely unlikely that an agent will finish a move in a GIS environment exactly on a road vertex. Therefore if an agent cannot not move all the way to the next vertex on their route they must travel as far as they can towards it, stopping before they reach the vertex. To further complicate matters, it is extremely unlikely that the agent will start their location directly on a junction or end on one. Therefore it is necessary to first find the closest junctions to the start and end of the route, and include allowances for the agents to get there from their initial and final positions.

![Figure 6.11: Example GIS route between A and C](image)

Road Accessibility and Transport Routes

Section 5.3 showed how different types of roads can be used by agents who have access to different types of transport. For example, some roads (such as motorways) can only be used by motor
vehicle drivers while others are available only to pedestrians. Also, it was shown how the speed that an agent is able to travel at will vary depending on both the type of road and the availability of different types of transport. This functionality is incorporated into the model with a slight adjustment to the routing algorithm. When calculating the shortest path, each road is assigned a weight relative to its length and these weights are passed to Dijkstra’s shortest-path algorithm. In the cases where an agent is able to travel faster than walking (the slowest form of transport), the weights are reduced accordingly so that a long road appears considerably shorter if it allows the agent fast movement across it. In this manner, agents with cars are more likely to use fast roads (i.e. motorways or A-roads) even if the physical distance is longer than the equivalent route using minor roads. As Section 5.3 discussed, this is important for the generation of realistic awareness spaces.

A further advancement is the inclusion of transport routes, such as buses or trains. These routes are slightly different to normal routes because they must begin and end at predefined stations (i.e. bus stops). Otherwise the routing algorithm is similar; the distances between stations (and the weights passed to Dijkstra’s algorithm) are calculated based on the improved speed of travel between stations compared to walking. Therefore it might be quicker for an agent to go out of their way to start a train or bus journey and this will affect their awareness spaces (Section 6.6.4 explicitly experiments with this feature).

6.5.5 Outputting Data

Simulation has been described as the “third way of doing science” (as apposed to induction or deduction) because it starts with a set of explicit assumptions (deduction) but then rather than proving theorems it creates data that can be analysed inductively (Axelrod, 1997). The consequence of this is that the type and manner of the data that are output from an agent-based model will directly affect the analyses that can be conducted and therefore the conclusions that can be drawn from the research. A relational database is ideal for the purposes of storing model data, both because modern databases allow quick, efficient access to the data they store and also because tables can be related which add structure to the underlying data. This makes it possible to ask more complex questions of the results than might be possible if results were output to other forms of data storage such as simple spreadsheets. Figure 6.12 illustrates the database schema which can be used to store all the necessary data from a model run. Using this schema it is also possible to output model results into the same database simultaneously which is essential when using a high-performance computing grid as outlined in the following section.

6.5.6 Utilising A High-Performance Grid

As with many agent-based models (e.g. Parry et al., 2006), the model produced by this research is extremely computationally expensive. A scenario outlined later (see Section 7.3) that covers an urban area of approximately 39 km² and contains 270 agents can require as much as twenty hours of computation time on a modern desktop PC. Although this is not an unreasonable amount of time to expect a complex model to run, and would not be a problem if the model needed to be
executed only once, the probabilistic nature of the model requires that it be executed a number of times to ensure results are robust. Running the model 100 times, for example, becomes impossible on a typical computer.

The “Grid” is a relatively modern concept in computing which aims to provide high computing power by linking typical computers (called “nodes”) and allowing them to work on a single problem simultaneously. The technology that underpins the paradigm is complex and rapidly evolving so will not be addressed here; the interested reader might start with Magoules et al. (2008) or see Parry and Evans (2008) for another agent-based modelling example. Following this paradigm, the National Grid Service (NGS: Geddes, 2006) is a service for UK researchers that provides compute power and data services to assist research. Although each individual computer node on a Grid does not have significantly more computing power than a typical desktop computer, over 100 can be utilised simultaneously to run models. Therefore running 100 models, which could take up to 2,000 hours on a desktop computer, could be completed on the grid in 20 hours using 100 nodes.

To make use of this computing power provided by the NGS, the model was adapted so that individual simulations could be executed, simultaneously, on different nodes. Figure 6.13 illustrates how this was accomplished; a “master” node controls a number of “slave” nodes by generating scenarios and informing the slaves to run them. Once a slave has finished a model it notifies the master and can have a new scenario allocated to it. In this manner hundreds of simulations can be executed in a single run, even though there are fewer nodes than models.
6.6 Verifying the Model – Experiments with Idealised Data

Thus far, this chapter has outlined the development of a prototype model and the important aspects of the full burglary simulation. The next stage, before testing and configuring the model with real data or using it to make real-world predictions, is to verify that the model works correctly and ensure that the dynamics are fully understood. Following Castle and Crooks (2006), the process of evaluating the model will be segregated into three distinct activities: verification, calibration and validation. Calibration and validation are addressed in Chapter 7 so the remainder of this chapter will outline the process of verification by experimenting with the model using idealised data. For the reader interested in the behaviour of the model in a realistic GIS environment rather than its behaviour in abstract environments, Section 6.6.4 (page 164) concludes the chapter with experiments using real Leeds environmental data.

6.6.1 The Process of Verification

Verification refers to the process of ensuring that the model has been programmed correctly and behaves as it is expected to. This can be thought of as “inner validity” (Axelrod, 1997; Brown, 2005). The model produced is highly complicated; it contains numerous environmental processes, detailed human behaviour, a large number of agents and intricate interactions between the agents and the environment. Also, from a programming perspective, it contains more than 10,000 lines of computer code; for all of these reasons verification is non-trivial. A reliable method of confirming that a simulation has been programmed correctly is to re-implement the same model separately, ideally using a different programming language. This process can be termed “docking” (Axtell
et al., 1996). However, this is far from feasible for this project because re-implementing such a large model would be extremely time consuming. Instead, Section 6.5.3 demonstrated that the model has been developed with the capacity for different types of environment to be “plugged-in” without having to change the implementation of other parts of the model (such as the agents, houses, communities etc.). The main aim of this novel feature is to allow environmental complexity to be added incrementally, making it easier to isolate any errors. Although this approach is not as comprehensive as complete docking, it is an effective substitute. Equally importantly, the approach addresses a worry in environmental criminology that models with complex environments can actually detract from the purposes of modelling; which is to better understand the dynamics of the system (Elffers and van Baal, 2008). By increasing environmental complexity gradually the researcher is able to both fully understand the dynamics of the model before incrementing the complexity of the environment and to isolate parts of the model from the effects of a geographical space.

At the same time as verifying that the model is error-free, the sensitivity of the model to different parameters will be tested to further our understanding of the dynamics of the model and to ensure that all parameters are affecting the model as they should. Ideally, every parameter would be tested but, as Table 6.2 will illustrate, this model is highly complicated and includes a very large number of parameters. In the absence of sufficient resources to systematically test every parameter, therefore, if a parameter is known to have little effect over the model it will not be tested directly. One might argue that if a parameter has little effect over the model it does not need to be included. The argument against this is that the model was built to be highly flexible and able to account for a broad range of burglary factors. For example, Section 2.2.1 noted that some authors disagree about the extent to which household visibility relates to burglary risk, but including visibility in the model provides the flexibility to include or omit the factor as new evidence arises or as a new theory is to be tested. Therefore even if a parameter is not used at this stage, it is available for future use if required without the necessity to rebuild the model.

To experiment with the effects of increasing geometric complexity on the model, three types of environment have been implemented:

1. An a-spatial or “null” environment in which agents do not actually move. The only parameter which can be set for this environment type is the length of time (number of iterations) that it will take for an agent to travel to a destination and all journeys take the same amount of time. For example, if an agent needs to travel to a house to purchase drugs the environment simply records the amount of time the agent has spent travelling, notifying the agent that the journey has ended after an allotted amount of time.

2. A simple Euclidean grid environment which is designed to loosely represent part of a real city, similar to that used in the prototype experiments of Section 6.4. Running the model on this type of environment offers fine-grained control over the geographical locations of different buildings and community types.

3. A complex GIS environment which accurately represents a real city. This is the most de-
6.6. VERIFYING THE MODEL – EXPERIMENTS WITH IDEALISED DATA

tailed and accurate environment type and will ultimately be used to make real-world crime predictions.

Table 6.2 documents each model parameter, how it affects the model and which environment will be used to test it. The criteria used to determine which environment should be used to test the parameter is based on the principle of being able to isolate the variable from as much extra complexity as possible. Therefore if the variable can be tested in the “spaceless” null environment it will be, otherwise it must be tested using the grid. Furthermore, the effects of some spatial variables can be subtle and wide ranging; in these cases the parameters require the size and variability provided in a GIS space. The GIS environment, therefore, utilises every variable but the grid and null environments do not use all (for example they do not have transport networks so transport-related variables are not relevant). Groups of variables in Table 6.2 are segregated into different categories depending on whether they affect the burglars (i.e. behaviour parameters), the environment (i.e. house, road and community parameters) or other parts of the model (i.e. the starting locations of the burglars). Variables in parentheses are included here for clarity but do not need to be tested directly either because they are dependant on other factors (such as an agent’s level of wealth) or because they are not expected to have a large effect on the model (if this is the case it will be justified).

Table 6.2: All model parameters and the environment which will be used to test the sensitivity of the model to their values.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AGENT VARIABLES</strong></td>
<td></td>
</tr>
<tr>
<td><strong>PECS Agent Parameters.</strong></td>
<td></td>
</tr>
<tr>
<td>Internal agent parameters can be isolated from space and tested in the <strong>NULL</strong> Environment</td>
<td></td>
</tr>
<tr>
<td><strong>(Wealth)</strong></td>
<td>The amount of wealth an agent has, required to socialise or buy drugs.</td>
</tr>
<tr>
<td>Work Gain</td>
<td>The amount of wealth gained per iteration of working.</td>
</tr>
<tr>
<td>(Sleep)</td>
<td>The level of sleep an agent has (a PECS state variable).</td>
</tr>
<tr>
<td>SleepGain</td>
<td>Amount of sleep gained per iteration when sleeping (configured by default so that agents want to sleep for 8 hours per day).</td>
</tr>
<tr>
<td>(Social)</td>
<td>The agent’s current level of the social PECS state variable.</td>
</tr>
<tr>
<td>SocialGain</td>
<td>Amount social gained per iteration of socialising (configured by default so that agents want to socialise for 2 hours per day).</td>
</tr>
<tr>
<td>CostSocialise</td>
<td>The cost of socialising per iteration.</td>
</tr>
<tr>
<td><strong>(Drugs)</strong></td>
<td>The level of drugs in the agent (a PECS state variable).</td>
</tr>
<tr>
<td>DrugsGain</td>
<td>Level of drug increase gained by taking drugs (configured by default so that agents want to take drugs once per day).</td>
</tr>
<tr>
<td>CostDrugs</td>
<td>Amount of wealth required to buy drugs.</td>
</tr>
<tr>
<td>DeteriorateAmount</td>
<td>The rate that internal state variables (e.g. sleep, drugs) will decrease.</td>
</tr>
</tbody>
</table>
### Burglary Agent Parameters.

Affect the agents’ decision regardless of spatial location of the buildings; tested in the **NULL Environment**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(BurgleGain)</td>
<td>The amount of wealth generated by a burglary. This parameter is not tested because it will simply dictate how often an agent needs to burgle, an effect that can be produced by varying other parameters (such as the cost of drugs for example).</td>
</tr>
<tr>
<td>Target weights</td>
<td>The weights that affect the area in which the agent will start searching for a burglary target. Based on the distance to the area, its attractiveness, the similarity to the agent’s home community and the number of previous successes (see Equation 5.11 in Section 5.10.1).</td>
</tr>
<tr>
<td>Victim weights</td>
<td>The weights which affect whether or not the agent will burgle a specific house. Based on the accessibility, visibility and security of the house, the traffic volume of the nearest road and the collective efficacy and occupancy levels of the community the house is in (see Equation 5.15 in Section 5.10.3).</td>
</tr>
<tr>
<td>(SearchTime)</td>
<td>The amount of time to spend searching before choosing a new target to travel to. This is unlikely to have a large effect and will not be tested.</td>
</tr>
</tbody>
</table>

### Spatial Agent Parameters.

Inherently spatial so will be tested in the **GRID Environment**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home locations</td>
<td>The locations of agents’ homes.</td>
</tr>
<tr>
<td>Social locations</td>
<td>Locations that agents go to in order to socialise.</td>
</tr>
<tr>
<td>Work locations</td>
<td>Locations that agents go to in order to work.</td>
</tr>
<tr>
<td>Drug dealers</td>
<td>Locations of drug dealers.</td>
</tr>
<tr>
<td>(Cognitive map)</td>
<td>The buildings and communities which the agent knows about and the number of crimes they have committed there.</td>
</tr>
</tbody>
</table>

### Environment Variables

#### Building Parameters.

Internal building parameters can be isolated from space and tested in the **NULL Environment**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>How easy a house is to break into. Based on the number of possible entrances (i.e. doors or windows).</td>
</tr>
<tr>
<td>Visibility</td>
<td>How visible the house is to neighbours or passers-by. Based on the size of the garden and the degree of isolation (the number of neighbours within a given buffer region).</td>
</tr>
<tr>
<td>Security</td>
<td>Security level of the house. No empirical data to set to a definite value, but included to increase scope of the model for future work.</td>
</tr>
<tr>
<td>TrafficVolume</td>
<td>The amount of traffic outside a house, based on the space syntax integration analysis of the road that the house is nearest to (this was discussed in Section 5.3.2).</td>
</tr>
</tbody>
</table>
### Community Parameters.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CollectiveEfficacy</td>
<td>The level of community cohesion. See Section 5.5.1 for a discussion of how the variable is derived.</td>
</tr>
<tr>
<td>Sociotype</td>
<td>The Output Area Classification (OAC) (Vickers and Rees, 2006) super-group of the community (see Section 5.5.2).</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>The general level of attractiveness of the community as calculated from OAC variables (see Section 5.5.3).</td>
</tr>
<tr>
<td>Occupancy</td>
<td>The expected levels of occupancy of the community at different times of day (see Section 5.5.4).</td>
</tr>
</tbody>
</table>

### Burglary Environment Parameters.

Inherently spatial because security increases depend on the locations of the buildings in relation to each other. These variables are used during the calibration phase (Section 7.3) so are not tested here.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SecurityIncrease)</td>
<td>The amount that security will increase in a house following a burglary</td>
</tr>
<tr>
<td>(SecurityDistance)</td>
<td>The amount that security will increase after a burglary for properties 1 unit away from the burglary (this deteriorates so that further properties are also affected but not by as much).</td>
</tr>
<tr>
<td>(SecurityDeteriorate)</td>
<td>The amount that security will deteriorate per day before returning to its base level.</td>
</tr>
</tbody>
</table>

### OTHER PARAMETERS

#### Other Spatial Variables.

Require a relatively large and varied environment so will be tested in a GIS Environment.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(IterPerDay)</td>
<td>The number of iterations per day. This will affect the number of properties that the agent will pass per iteration while they are travelling. This is included for flexibility (to find a balance between realism and computational complexity) but is unlikely to have a significant impact so will not be tested.</td>
</tr>
<tr>
<td>Transport availability</td>
<td>The different types of transport available to the agent (i.e. buses, car, train etc.).</td>
</tr>
<tr>
<td>(Transport Speeds)</td>
<td>The speed gains provided by different transportation methods. These are applicable only to the Vancouver case studies in Chapter 9 at this stage but can be readily included for Leeds as future research.</td>
</tr>
</tbody>
</table>

#### Other Non-Spatial Variables.

These parameters can be isolated from space and therefore tested in a NULL Environment.

---

2 A “unit” of length here is equal is the length of a cell in the grid environment or 20 meters in the GIS environment.
### Extracted Text

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConstTravelTime</td>
<td>The constant value used in the NULL environment to determine how many iterations it will take for an agent to reach their destination. Testing this value in the NULL environment will provide an indication of the maximum distances that agents are able to travel before they are unable to satisfy their motives in spatial environments. This variable does not apply to other environments, where travel time is determined by geography.</td>
</tr>
<tr>
<td>(TravelPerTurn)</td>
<td>The distance the agent can travel when walking. Currently 100m/min (4mi/h) in GIS or 1 unit in grid. This parameter is not tested because the distances that agents can travel each turn are experimented with by varying the effects of transportation.</td>
</tr>
</tbody>
</table>

### 6.6.2 Tests Using The Null Environment

There are two categories of tests which will be performed using the Null environment. The first category encompasses all the PECS variables. These variables are used to drive the behaviour of the agents and are based on the PECS cognitive framework as discussed in Section 5.7.3. It must be determined how agents’ internal state variables are influenced by external actions. For example, what will happen to an agent’s state variables if the cost of socialising increases and how will this affect their behaviour? The second test category relates to burglary; how will external building or community parameters and internal agent burglary parameters influence where an agent chooses to burgle?

The simulation will contain only a single agent for clarity, as agents do not interact directly this will not distort results. The environment will be configured slightly differently for each category of test. For the first category (PECS variables) there will only be a single building of each type (a house, work place, drug dealer and social place) to limit the complexity. All buildings will be contained in the same community and will have default parameter values for all of their internal variables (0.5). This is illustrated by Figure 6.14. In the second category (burglary variables) there will be different numbers of houses and communities in order to compare the effects of changing house and community parameter values.

With regards to timing, a single iteration is defined as 1 minute of virtual time; this equates to 60 * 24 = 1440 iterations in each virtual day. This is sufficient to provide a high temporal resolution and has the elegance of linking simulated minutes to iterations directly. The simulation will be run for 43,200 iterations (30 days). This was chosen because it is at least sufficient for all the simulations to reach equilibrium (as Section 7.3 will illustrate).

Although there is no space in the Null environment, agents still *think* that they are travelling...
between objects. The Null environment, therefore, controls how long each hypothetical journey takes. For these experiments the constant travel time will be set at 30 minutes/iterations but the travel time variable will also be tested in the first experiment with the PECS variables.

**Testing PECS Variables**

The following describes how the parameters have been configured initially. These follow decisions that were based on the initial experimentation with the prototype model (see Section 6.4). Assumptions will be justified afterwards.

- State variable levels deteriorate at a constant rate of 1 unit per day. The approximate amount of time that agents must spend increasing their state variables can be altered by changing the return gained by working, sleeping, socialising or taking drugs. For example, decreasing the amount of sleep gained from one hour of the “sleep” action effectively increases the amount of time the agents must spend sleeping.
- With default gains from the appropriate actions, agents want to spend approximately 8 hours sleeping, 2 hours socialising and want to take drugs once per day to maintain comfortable state variable levels.
- Working for 8 hours or committing a single burglary provides 1 unit of wealth. Socialising for two hours costs 0.25 units and taking drugs costs 0.5 units. Therefore agents will need to work for approximately 6–8 hours per day or commit slightly less than one burglary per day on average to make enough wealth to satisfy their needs for socialising and drug addictions.

These parameters are global at this stage, making agents homogeneous. However, it is trivial to vary parameters for individual agents which will lead to a heterogeneous population and allow researchers to experiment with different burglar behaviour and motivations – this is one of the major benefits of agent-based modelling (Axtell, 2000; Castle and Crooks, 2006). With regards to the assumptions made for these parameter values, they have initially been chosen because they will simulate, approximately, the daily patterns that might be exhibited by people employed in typical “9–5” jobs (following initial experimentation with the prototype in Section 6.4). The largest effect that the variable values will have on the model will be to develop the burglars’ awareness spaces and subsequently influence where they will look for burglary targets. For example, it is important that the need to socialise expands an agent’s awareness space to parts of the city that they would otherwise not explore. But it is less important whether or not they spend one or two hours actually socialising when they get there. Therefore, although the effect that the variables have on the model will be thoroughly tested here, seeking additional evidence to help choose their initial values is not necessary.

Table 6.3 outlines the variables that will be tested and their default values (that are set to reflect the previous assumptions based on the number of iterations per day). Using the gain from work as an example: if agents should spend 8 hours working and there are 1440 iterations per day, or \((1440/24) \times 8 = 480\) iterations in an 8 hour period, then each iteration must provide \(1/480 = 0.002083\) units of wealth per iteration. If the number of iterations per day changes these values
CHAPTER 6. MODEL DEVELOPMENT AND TESTING

must obviously be updated to reflect these changes. The ConstTravelTime variable is the time that each journey will take in the null environment and is also tested here to determine the maximum length of journeys before the agents are unable to satisfy their motives because they must spend so much time travelling. As discussed in the previous section, its initial value is set to 30 minutes (this is also consistent with the prototype simulations in Section 6.4).

Table 6.3: Default values and range for PECS variables to be tested.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Initial Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WorkGain</td>
<td>0.002083</td>
</tr>
<tr>
<td>SleepGain</td>
<td>0.002083</td>
</tr>
<tr>
<td>SocialGain</td>
<td>0.008333</td>
</tr>
<tr>
<td>DrugsGain</td>
<td>1</td>
</tr>
<tr>
<td>CostSocialise</td>
<td>0.002083</td>
</tr>
<tr>
<td>CostDrugs</td>
<td>0.5</td>
</tr>
<tr>
<td>ConstTravelTime</td>
<td>30</td>
</tr>
<tr>
<td>DeteriorateAmount</td>
<td>0.000694</td>
</tr>
</tbody>
</table>

Recall that a state variable is a PECS variable that represents the internal level of some need. For example, the sleep state variable represents an agent’s “sleepiness”. The state variable has an associated motive which competes with other motives to drive the agent’s behaviour (the strongest motive wins). So with sleeping, as the sleep state variable approaches zero the associated motive becomes very large and will take control of the agent who will subsequently go home to sleep. If a state variable value reaches zero, however, the model cannot function as there is no way to calculate the associated motive intensity (recall that Graph 5.14 demonstrated that motive intensity is infinite if its state variable value is zero). The model therefore becomes inoperative. This occurrence could represent a significant turning point for the agent which causes a dramatic change in their behaviour (such as choosing to break a drug habit) but exploring this is beyond the scope of the research. Instead, it is important to establish the range of parameter values that will cause a state variable to drop to zero as this is the effective minimum/maximum value of the parameter and therefore indicates how sensitive the model is to the parameter.

Figure 6.15 graphs the values of all state variables and motives for a single burglar when each parameter has been set to the initial value outlined in Table 6.3. Only iterations 0 – 5,000 are shown for clarity, although the same pattern is repeated for the duration of the simulation. After an initial period of disturbance the agent forms a routine and both state variable levels and motive intensities stabilise. All state variables deteriorate over time unless the agent is performing an activity which increases their value. The state variable levels for sleep and socialising increase steadily as the agent sleeps or socialises respectively. The drugs state variable, on the other hand, increases instantaneously when the agent takes a “hit”. The pattern for motive intensities is less orderly because the intensity is driven by the time of day as well as the size of its associated state variable, but motive intensities also reach equilibrium nonetheless. Further work demonstrated that the behaviour of the agent is also as expected if they have to burgle to make money rather
than work (Malleson et al., 2010a). This supports work on the prototype model (Section 6.4) and shows that PECS has been reimplemented using Repast Simphony successfully.

This proof-of-concept experiment establishes that the simulation is performing as expected under default conditions and in the absence of a geographical space, it is possible to begin to vary the values of parameters to establish what affect they will have on the model. Table 6.4 illustrates the results of varying the WealthGain parameter which represents the amount of wealth given, per iteration, to agents who are working. It should be noted that the parameter was varied in regular intervals between zero and ten times its default value but, for clarity, only some results are included in the table (these are representative of the pattern shown by all).

The first two simulations fail: agents are not able to generate enough wealth to satisfy their needs and the simulations terminate after 2,880 and 10,344 iterations respectively as shown by the Max Iteration field. As discussed, this occurs because the value of the state variable reaches zero. The remaining simulations illustrate firstly that the amount of time agents spend working (Avg Work Time) reduces significantly as the gain from work increases. Furthermore, the mean levels of state variables are generally larger and motive intensities generally lower as the return from working increases. This is to be expected because the agents do not need to spend as long working and can, instead, spend time satisfying other PECS motives. The agents are effectively more comfortable; their state variables never reach a particularly low point and therefore associated motives are never excessively strong. This is also evident by the increases in time spent doing nothing (AvgDoNothingTime) which occurs when there is no motive strong enough to take control.

Figure 6.15: The intensity of motives and values of state variables for a single agent under initial parameter values. The agent has full-time employment.
Table 6.4: Summary results for tests of the WealthGain parameter. As the parameter increases, the agent can spend less time working and more time doing nothing.

<table>
<thead>
<tr>
<th>WorkGain value</th>
<th>0.0</th>
<th>0.001</th>
<th>0.002</th>
<th>0.01</th>
<th>0.02</th>
<th>0.02083</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Iteration</td>
<td>2880</td>
<td>10344</td>
<td>43200</td>
<td>43200</td>
<td>43200</td>
<td>43200</td>
</tr>
<tr>
<td>Avg Sleep SV</td>
<td>-</td>
<td>-</td>
<td>0.99</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Avg Sleep M</td>
<td>-</td>
<td>-</td>
<td>0.48</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>Avg Social SV</td>
<td>-</td>
<td>-</td>
<td>1.22</td>
<td>1.41</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Avg Social M</td>
<td>-</td>
<td>-</td>
<td>0.41</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Avg Drugs SV</td>
<td>-</td>
<td>-</td>
<td>1.14</td>
<td>1.23</td>
<td>1.24</td>
<td>1.24</td>
</tr>
<tr>
<td>Avg Drugs M</td>
<td>-</td>
<td>-</td>
<td>0.47</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
</tr>
<tr>
<td>Avg Travel Time</td>
<td>-</td>
<td>-</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Avg Work Time</td>
<td>-</td>
<td>-</td>
<td>0.25</td>
<td>0.05</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Avg Social Time</td>
<td>-</td>
<td>-</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Avg Sleep Time</td>
<td>-</td>
<td>-</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Avg DoNothing Time</td>
<td>-</td>
<td>-</td>
<td>0.18</td>
<td>0.38</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Avg Drug Take</td>
<td>-</td>
<td>-</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

of the agent. This type of test was repeated for all the other parameters outlined in Table 6.3. Summary results are presented here and the tables themselves can be found in Malleson et al. (2010a).

- Varying SleepGain and SocialGain parameter values are similar to WealthGain as would be expected, affecting the amount of time that agents spend sleeping and socialising respectively but not strongly influencing other parts of the model.
- Changing the DrugsGain parameter to less than 0.5 results in a simulation failure as the agent cannot consume enough drugs to maintain a state variable level above zero. As the parameter increases above 0.5 the number of times the burglar must take drugs decreases. Again more time is spent doing nothing as the agent does not need to make money to buy drugs as often.
- Increasing the cost of socialising (CostSocialise) and taking drugs (CostDrugs) results in the agent spending more time working. Once the variables become too large (values of 0.006 and 1.25 respectively) the simulation breaks down as the agent is not able to generate sufficient wealth to satisfy their needs.
- Varying the amount of time each simulated journey takes to complete (the ConstTravelTime variable) directly influences how much time the agent is able to spend doing nothing, as expected due to longer travel times. At approximately ConstTravelTime $\approx 100$ the agent is not able to satisfy their needs because each journey takes such a considerable amount of time and the simulation breaks. This information will be useful when configuring the speed at which the agents are able travel around the GIS environment.
- As the rate of state variables deterioration (DeteriorateAmount) increases, the amount of time the agents spend doing nothing decreases. This is because they need to spend longer
and longer working to satisfy all their needs and therefore do not have time to do nothing. This continues until state variables deteriorate at such a rate that the agents are no longer able to satisfy their needs and the simulation fails.

As a result of these tests it has been established that the PECS behaviours are controlling the agent as they should do in the absence of a spatial model component and without the need to burgle. Furthermore, all parameters have similar affects on the model which is encouraging as it is not the intention of the model design (at this stage) that one variable should have a stronger affect on agents than another. Therefore it is now possible to systematically test burglary parameters in order to examine the effect that they will have on agent behaviour.

**Testing Burglary Variables**

These tests are necessary to determine how attractive a house or community is to a burglar (and subsequently whether or not they decide to burgle there) based on the values of relevant parameters. Parameters are divided into three separate groups: those of buildings; those of communities; and those internal to agents. The agent parameters are denoted with a following “_W” to indicate that they are weights which the burglars apply to an associated house or community variable when contemplating burglary. For example, the Attractiveness_W parameter represents how the attractiveness of a community affects an agent when they are deciding where to travel to in order to search for a burglary target; an agent with a high value for the Attractiveness_W weight is more likely to travel to a community which is highly attractive and vice versa. (In Equation 5.11, Section 5.10.1, this variable was denoted $w_2$). The environment will be configured for the different categories of tests as illustrated by Figure 6.16.

![Figure 6.16](image)

Figure 6.16: The environmental configurations (including the value of the parameter being tested) when testing house parameters and community parameters using the Null environment.

As Section 5.10 discussed, the burglary process incorporates random components into the agents’ decision such that they are more likely to burgle if conditions are favourable. The probabilistic nature of the decision means that the robustness of model results must be established by executing the model a number of times so that the expected number of burglaries in a house is
approximately equal to the mean of infinitely many executions. Following Voas and Williamson (2000), Malleson et al. (2010a) shows that below 50 separate model runs the standard error is unacceptable. But the improvement in error between 50 and 100 iterations are offset by the considerable extra execution times and therefore 50 iterations is deemed sufficient for these tests.

Table 6.5 illustrates the proportion of burglaries committed in houses or communities with a given parameter value over 50 model runs. Each parameter was tested individually so the table displays the results of seven individual experiments. Most results are as to be expected. The lower the accessibility (Acc), visibility (Vis), security (Sec) and traffic volume (TV) house parameters, the more burglaries the houses can expect. Similarly with communities, large values for the collective efficacy (CE) and occupancy (Occ) parameters lead to more burglaries in the community. The only exception is the attractiveness (Att) community variable which does not appear to effect the number of burglaries committed in the community; the proportion is consistent regardless of the parameter value. This is because the attractiveness parameter is used by agents in a different way: instead of directly influencing the likelihood that they will commit a burglary, attractiveness is used to determine where they will begin to search for a burglary target. As there are only 11 houses and communities in the environment it is likely that, although the agents start their search in the most attractive community, the final burglary actually occurs elsewhere. In a large spatial environment where each community contains a large number of houses it is more likely that attractiveness will more directly influence where burglaries occur. Sections 6.6.3 and 6.6.4 will experiment with this premise.

Table 6.5: The proportion of burglaries committed in houses or communities with given parameter values. Each parameter was tested separately so the table contains the results of seven separate experiments. For example, the “Exp. 1” (experiment 1) column shows that as the Acc variable (household accessibility) increases from 0 to 1 the proportion of burglaries in that house decreases.

<table>
<thead>
<tr>
<th>Parameter Value</th>
<th>Exp. 1: Acc</th>
<th>Exp. 1: Vis</th>
<th>Exp. 1: Sec</th>
<th>Exp. 1: TV</th>
<th>Exp. 2: CE</th>
<th>Exp. 2: Att</th>
<th>Exp. 2: Occ</th>
<th>Exp. 3: CE</th>
<th>Exp. 3: Att</th>
<th>Exp. 3: Occ</th>
<th>Exp. 4: CE</th>
<th>Exp. 4: Att</th>
<th>Exp. 4: Occ</th>
<th>Exp. 5: CE</th>
<th>Exp. 5: Att</th>
<th>Exp. 5: Occ</th>
<th>Exp. 6: CE</th>
<th>Exp. 6: Att</th>
<th>Exp. 6: Occ</th>
<th>Exp. 7: CE</th>
<th>Exp. 7: Att</th>
<th>Exp. 7: Occ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.27</td>
<td>0.26</td>
<td>0.28</td>
<td>0.26</td>
<td>0.08</td>
<td>0.25</td>
<td>0.21</td>
<td>0.09</td>
<td>0.2</td>
<td>0.16</td>
<td>0.17</td>
<td>0.14</td>
<td>0.13</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>0.1</td>
<td>0.22</td>
<td>0.2</td>
<td>0.23</td>
<td>0.22</td>
<td>0.21</td>
<td>0.09</td>
<td>0.2</td>
<td>0.16</td>
<td>0.1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>0.2</td>
<td>0.16</td>
<td>0.17</td>
<td>0.15</td>
<td>0.14</td>
<td>0.16</td>
<td>0.1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>0.3</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td>0.1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.1</td>
<td>0.17</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>0.4</td>
<td>0.08</td>
<td>0.1</td>
<td>0.09</td>
<td>0.09</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.11</td>
<td>0.09</td>
<td>0.09</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.11</td>
<td>0.09</td>
<td>0.1</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>0.5</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
<td>0.09</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.1</td>
<td>0.05</td>
<td>0.04</td>
<td>0.1</td>
<td>0.05</td>
<td>0.04</td>
<td>0.1</td>
<td>0.05</td>
<td>0.04</td>
<td>0.1</td>
<td>0.05</td>
<td>0.04</td>
<td>0.1</td>
<td>0.05</td>
<td>0.04</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.1</td>
<td>0.03</td>
<td>0.03</td>
<td>0.1</td>
<td>0.03</td>
<td>0.03</td>
<td>0.1</td>
<td>0.03</td>
<td>0.03</td>
<td>0.1</td>
<td>0.03</td>
<td>0.03</td>
<td>0.1</td>
<td>0.03</td>
<td>0.03</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

The following tests will experiment with internal burglar parameters. Recall from Section 5.6 that each environmental parameter (such as household accessibility) has an associated burglar
parameter which determines how strongly the environmental parameter influences the agent’s decision whether or not to burglar. For example, a burglar with a large value for the internal Accessibility weight will look more favourably on houses that are easily accessible than a burglar with a low value for Accessibility. In the following experiment, each burglar is assigned a different value for the Accessibility weight and each house has a different value of accessibility. All other parameters are kept constant. Table 6.6 illustrates the results and is representative of the patterns produced by other variables. It shows that, as expected, burglars with a low Accessibility weight do not take the accessibility of the house into account when looking for a burglary target. Burglar 0, for example, has a weight of 0.0 and commits approximately the same number of burglaries in all houses, regardless of their accessibility. Burglar 10, on the other hand, has a weight value of 1.0 and commits many more burglaries in houses which are highly accessible. Results for the remaining weights (Visibility, Security, TrafficVolume, CollectiveEfficacy and Occupancy) are similar (Malleson et al., 2010a). As with the previous tests, the results of Attractiveness (how important community attractiveness is to an agent when they decide where to look for a burglary target) do not affect where the ultimate burglary happens (which is to be expected for the reasons stated earlier – community attractiveness dictates where the agent starts their search, not where they ultimately burglar).

Table 6.6: The number of burglaries committed by different agents in different houses by changing the Accessibility agent behaviour parameter (how important house accessibility is to the agent). It is apparent that agents with a high Accessibility value (e.g. burglar 10) choose houses that are easily accessible whereas those with a low weight (e.g. burglar 0) are not influenced by household accessibility, they are equally likely to burglar any house regardless of its accessibility.

<table>
<thead>
<tr>
<th>Burglar number</th>
<th>B0</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
<th>B7</th>
<th>B8</th>
<th>B9</th>
<th>B10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Household accessibility (acre)</td>
<td>0</td>
<td>93</td>
<td>126</td>
<td>154</td>
<td>178</td>
<td>181</td>
<td>206</td>
<td>246</td>
<td>240</td>
<td>264</td>
<td>297</td>
</tr>
<tr>
<td>0.1</td>
<td>111</td>
<td>106</td>
<td>126</td>
<td>151</td>
<td>163</td>
<td>184</td>
<td>187</td>
<td>222</td>
<td>206</td>
<td>226</td>
<td>234</td>
</tr>
<tr>
<td>0.2</td>
<td>102</td>
<td>129</td>
<td>122</td>
<td>139</td>
<td>119</td>
<td>163</td>
<td>152</td>
<td>145</td>
<td>169</td>
<td>171</td>
<td>178</td>
</tr>
<tr>
<td>0.3</td>
<td>114</td>
<td>112</td>
<td>108</td>
<td>125</td>
<td>146</td>
<td>128</td>
<td>135</td>
<td>121</td>
<td>148</td>
<td>127</td>
<td>132</td>
</tr>
<tr>
<td>0.4</td>
<td>96</td>
<td>107</td>
<td>113</td>
<td>88</td>
<td>103</td>
<td>117</td>
<td>107</td>
<td>120</td>
<td>113</td>
<td>103</td>
<td>96</td>
</tr>
<tr>
<td>0.5</td>
<td>102</td>
<td>104</td>
<td>91</td>
<td>100</td>
<td>95</td>
<td>78</td>
<td>81</td>
<td>67</td>
<td>76</td>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>0.6</td>
<td>96</td>
<td>98</td>
<td>109</td>
<td>81</td>
<td>92</td>
<td>63</td>
<td>70</td>
<td>66</td>
<td>46</td>
<td>46</td>
<td>40</td>
</tr>
<tr>
<td>0.7</td>
<td>90</td>
<td>83</td>
<td>77</td>
<td>77</td>
<td>56</td>
<td>53</td>
<td>45</td>
<td>46</td>
<td>34</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>0.8</td>
<td>87</td>
<td>85</td>
<td>81</td>
<td>66</td>
<td>56</td>
<td>59</td>
<td>32</td>
<td>28</td>
<td>27</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>0.9</td>
<td>112</td>
<td>66</td>
<td>60</td>
<td>51</td>
<td>49</td>
<td>25</td>
<td>20</td>
<td>30</td>
<td>10</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>97</td>
<td>84</td>
<td>59</td>
<td>44</td>
<td>40</td>
<td>24</td>
<td>25</td>
<td>15</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

The tests carried out so far have systematically shown that the PECS variables are controlling the agents’ behaviour correctly and the burglary parameters are suitably influencing where the agents decide to burglar in the absence of a spatial environment. Changing parameter values does not result in any unexpected effects and model behaviour is not overly (or underly) sensitive to parameter changes. The next stage in model verification is to add a simple grid-based environment to the model and examine how it behaves.
6.6.3 Tests Using The Grid Environment

The previous section demonstrated that the model performs as expected in the absence of a spatial environment. The next stage in the verification of the model is to experiment with the addition of space: how will the agents behave when they actually have to travel around? The following tests will be performed:

1. Exploration of the spatial distributions of burglary under default conditions.
2. Varying the parameters that affect where an agent chooses to start their search: the Distance_W (how far to travel) and PreviousSuccess_W (how much more attractive areas are that the agent has already successfully burgled in) agent parameters.

Figure 6.17 depicts the layout of the environment. It has been kept simple to isolate the features being tested from spatial complexity (with the exception of the road network which will not be a simple regular grid in order to ensure the routing algorithms are working correctly, i.e. some roads are dead ends). It will consist of a single social location, drug dealer and workplace and contain one agent who always lives at the same home address. These locations are in opposite corners so that the agent must travel between them and build up an awareness of the environment (their “cognitive map”) as stipulated by routine activities theory and crime pattern theory (Cohen and Felson, 1979; Brantingham and Brantingham, 1981b). The different communities can be assigned different parameter values but are otherwise indistinct.

The first test is a general test of the addition of space to the model: all parameters have default values and all communities are identical. A single agent lives in the north-west corner of the environment and must burgle to satisfy their needs. As with previous tests the simulation was run for 30 simulated days and was run 50 separate times. Figure 6.18 illustrates the distribution of burglaries committed across all simulations. On the whole, burglaries are spread throughout the environment but are more dense around the agent’s home.

Observing the distributions of burglaries, it is interesting that there do not appear to be clusters of burglary surrounding the social and drug dealer locations. This might be expected because these areas are very likely to feature in the agents’ awareness spaces. Figure 6.19 illustrates the evolution of a burglar’s awareness space in a single example model run. It becomes clear that approximately a third of the way into the simulation (10,087 iterations) the agent has nearly explored the environment in its entirety, so the awareness space in this case actually makes little difference to the choice of where to begin searching for a burglary target (the agent has global environmental knowledge rather than isolated local knowledge). As Section 6.6.4 will show, this is not the case in larger environments that more closely represent real urban configurations such as the GIS environment. Nonetheless, Figure 6.19 is able to illustrate that the agent’s cognitive map is working correctly.

The reason that burglaries are more clustered around the agents’ home, therefore, is because although all communities are perceived as identical the agent is still more likely to pass nearby properties at the start of the journey regardless of where they are actually going so the potential to burgle is always present. This is consistent with the literature, where the proximity to potential
6.6. VERIFYING THE MODEL – EXPERIMENTS WITH IDEALISED DATA

Figure 6.17: The layout of roads, buildings and communities used in the Grid environment sensitivity tests. The different types of communities can be assigned different parameter values depending on the experiment, although there are no inherent differences between them.

Figure 6.18: The distribution of burglaries with the default configuration (50 model runs). The absolute burglary count (left) and interpolated density using KDE (right). The KDE algorithm estimates the density of burglaries and is introduced in Section 7.2.1.
162  CHAPTER 6. MODEL DEVELOPMENT AND TESTING

The evolution of a burglar’s awareness space

Figure 6.19: The evolution of a burglar’s awareness space in a single example model run.

offenders is commonly associated with burglary risk (Snook, 2004; Bernasco and Nieuwbeerta, 2005).

The following experiment will alter this behaviour by changing the Distance\_W parameter in order to determine where burglars are more likely to burgle when they are strongly/weakly influenced by the distance that a target community is away from their home. Figure 6.20 illustrates the results in the form of hotspots and absolute number of burglaries. However, it appears that altering the parameter has made little difference to the results: the patterns produced by the model when the parameter was altered between 0.0, 0.5 and 1.0 are similar. The reason for this is non-trivial. Recall from Section 5.10.1 that, when deciding where to search for a burglary target, the agent assigns a value, \( l_a \), to each community, \( a \), based partly on its distance \(^3\):

\[
l_a \propto \text{Distance\_W} \cdot \frac{1}{\text{dist}(c,a)}
\]

where \( \text{dist}(c,a) \) represents the distance between the agent’s current location and the target community. It is not unsurprising, therefore, that changing the value from 0.5 to 1.0 has no effect;

\(^3\)Equation 6.1 (see Section 5.10.1) in full is:

\[
l_a = w_1 \cdot \frac{1}{\text{dist}(c,a)} + w_2 \cdot \text{attract}(h,a) + w_3 \cdot \text{socialDiff}(h,a) + w_4 \cdot \text{prevSucc}(a)
\]

where \( \text{dist}(c,a) \) represents the travel distance to the community \( a \), \( \text{attract}(h,a) \) is the relative attractiveness of \( a \), \( \text{socialDiff}(h,a) \) represents the social difference between their home community and \( a \), \( \text{prevSucc}(a) \) is the number of times they have successfully burgled in \( a \) and \( w_1 - w_4 \) are internal weights that the burglar applies to each parameter.
as there are no other parameters influencing the agents’ decision, the attractiveness of the communities, relative to each other, will be the same. A value of 0.0 might be expected to alter the pattern, however, because that would discount the influence of distance completely and thus increase the spread of burglaries away from the agent’s home. The fact that this does not happen is indicative of the size of the environment: the agent is able to search most of the environment without becoming desperate for burglary so where they actually start the search is of little importance.

To ensure that allowing the model to run for a longer period of time does not change the effect that altering the Distance W parameter will have, further research showed that running the model for sixty days instead of only thirty does not have an effect. Furthermore, research showed that with other parameters that affect where an agent is going to start their search (PreviousSuccess W and Sociotype W) the results are similar and burglary patterns seem unaffected (Malleson et al., 2010a).

Overall, the results show that the experiments demonstrate utility. Importantly they are able to show the model can work in an environment other than the null environment without any changes to the agents. Furthermore, they have demonstrated that the burglars’ awareness space is working properly. In particular, notice that in the map of the agent’s awareness space (Figure 6.19) there is a dead-end road in the centre of the environment which is the last area that the agent discovers.
This demonstrates the advantages of an agent-based model and burglars with an awareness space: houses on quiet roads without through-traffic might be less of a burglary risk simply because potential burglars are unaware of them. Other types of models (such as the statistical ones outlined in Section 3.3) will find it extremely difficult to include these types of effects that are fundamental to modern criminology theory (Cohen and Felson, 1979; Brantingham and Brantingham, 1981b).

Another advantage of the tests is that it has become clear that there is a risk that where the agents start their search is immaterial if they are able to search a very large area before deciding to burgle. This risk can be mitigated in future experiments and provides evidence for the benefits of increasing model complexity gradually. Without exploring model dynamics in a simple environment first, this effect might not have become apparent.

6.6.4 Tests Using The GIS Environment

The previous sections have illustrated that in the absence of a Euclidean space the model performs as expected to. The addition of a simple Euclidean space illustrated that the model is able to work with different environment configurations without changing the behaviour of the agents or other environment objects (such as buildings, communities etc.). The following section will increase complexity further and introduce a space that closely represents a real city: a GIS environment.

The most important aims of this section are:

- to establish that the model still works properly using the new GIS environment;
- to determine whether or not the assumptions made in the previous section regarding the effects of community parameters on burglary locations are indeed accurate (burglars would have been influenced by the community parameters but the grid environment was too small);
- to test an important feature of the model that is otherwise untested: transport routes.

There are numerous additional variables that would, ideally, be tested also (such as the number of iterations per day or the time agents spend searching before travelling to a new community) but these will not have a great effect on burglary and so are not experimented with further. Some variables (such as how the security of houses respond to burglary) are tested indirectly when the model is calibrated later.

To meet the aims listed above, three experiments are conducted. The first uses the default model configuration (described below) to determine how an offender behaves in the GIS environment. The second changes the Distance_W parameter so that the agent prefers to stay near their home rather than travelling to distant communities. In the final experiment the effects of the transport network are tested. Figure 6.21 summarises the environment, showing the locations of different types of buildings and the transport route that will be used in the second experiment; Malleson et al. (2010a) discuss the experimental setup in more detail.

Figure 6.22 illustrates the results of all experiments. All models were run 50 times; the figure shows the aggregated burglary counts and the hotspots that were produced using the kernel-density algorithm (cell size 10m and bandwidth 100m). Altering the Distance_W agent parameter now has the desired effect on the model, although it is not pronounced. With low values the agent
6.6. VERIFYING THE MODEL – EXPERIMENTS WITH IDEALISED DATA

Figure 6.21: The layout of the environment used in GIS sensitivity tests. Notice that the environment includes a public transport route (green).

The most striking result, however, is the difference that the public transport route has on burglary patterns. When the agent has access to public transport they use this to travel between their home, their drug dealer and the social location. During travel on public transport, the agents do not add the houses they pass to their awareness space which explains why the locations of burglaries are so much more heavily clustered. The spatial locations of the agents’ burglaries are illustrated...
Figure 6.22: The GIS environment test results (total of 50 model runs): normal conditions (upper-left), with public transport (upper-right) and altering the Distance_W parameter (bottom). As the Distance_W parameter increases the agents are more likely to burgle close to their homes.
Table 6.7: The average distances that crime were committed away from the agent’s home and all anchor points. Distances increase with increasing values of the \( \text{Distance}_W \) parameter.

<table>
<thead>
<tr>
<th>Distance ( W ) Value</th>
<th>Mean Distance (m) From Home</th>
<th>Mean Distance (m) From Nearest Anchor Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1121</td>
<td>755</td>
</tr>
<tr>
<td>5.0</td>
<td>995</td>
<td>709</td>
</tr>
</tbody>
</table>

in detail by Figure 6.23. This seemingly simple result is significant because it illustrates the power and novelty of the model. No other published crime models are able to account so cleanly for criminology theory such as crime pattern theory (Brantingham and Brantingham, 1981a). Using the model in this way will certainly be useful for the forthcoming crime forecasts.

Figure 6.23: The GIS environment test: results produced when changing whether or not the agents have access to public transport (total of 50 model runs).

**Verification – Conclusion**

This section has systematically tested numerous model parameters. It was established in Section 6.6.2 at what point parameters could “break” the model by causing a PECS state variable value to drop to zero. Others were tested to determine whether or not they had the desired effect on the model and also how sensitive the model is to their value. Furthermore, complexity was added to the model systematically through the use of different types of environment. The experiments have shown that the model performs as expected given its definition and configuration and is not overly sensitive to any of the parameters. Confident that the model has been programmed
correctly and is performing as expected, it is possible to further evaluate the model as discussed in Chapter 7.

6.7 Model Development – Summary

This chapter has outlined how the model has been developed and has documented the process of testing the model. In summary, the model was developed in the Java language using the Repast Simphony agent-based modelling toolkit. The PECS behavioural framework controls the agents and has been implemented using state variables, motives and actions. A state variable is an internal “need” which an agent must maintain at a suitable level (such as drugs for addicted agents). Motives are related to state variables (and other factors such as the time of day) such that the lower a state variable, the higher the intensity of its associated motive. An agent’s most intense motive drives their behaviour by creating a list of actions which the agent must complete, in order, to increase the level of the state variable. These lists of actions range from the simple (go home and go to sleep) to something much more complex (such as searching for a burglary target in order to then travel to a bar to socialise). The burglars themselves are heterogeneous: they can be implemented with different state variables (e.g. being addicted to drugs or not), can use different actions to perform tasks (e.g. the manner in which they search for burglary targets) and, with respect to burglary, can view environmental cues very differently.

With regards to the environment, an important feature of the model is the ability to use the same agents and the same rules in different types of environment. For example, agents could exist in a realistic urban GIS environment, a simple cellular grid environment or an abstract environment which controls everything by simple rules and bears no relation to a Euclidean space. In this manner it is possible to isolate spatial complexity from other aspects of the model and provide a fuller understanding of model dynamics. Routing around the environment must therefore be managed differently depending on the environment so alternative routing algorithms have been developed for the GIS and the cellular environments (the abstract environment has no virtual space so there is no need for routing). These include the ability for agents to move around using public transport as well as with cars and on foot.

The model is too computationally expensive to run on a normal desktop computer. Therefore it was adapted to make use of the individual computers on the National Grid Service hosted at the University of Leeds. In this manner numerous individual models can be run simultaneously, allowing robust results to be generated in an acceptable amount of time.

As the model developed for this research uses a very large number of parameters it is accordingly extremely difficult to evaluate. The main evaluation began with Section 6.6 which verified that the model had been programmed correctly by experimenting with hypothetical data. It was decided that re-implementing the model in a different programming language was infeasible so, instead, the model was executed on different types of environment that increased in complexity gradually. This novel approach also had the advantage that it allowed for sensitivity tests to be performed accurately as environmental complexity could be limited. This is in response to a concern among some environmental criminologists (e.g. Elffers and van Baal, 2008) that models with
complicated environments can detract from understanding the underlying processes which is the main advantage of this type of model.

In the next chapter the process of evaluating the model outlined by Castle and Crooks (2006) will continue by calibrating the model and then validating it. This is accomplished through experiments with real data using a realistic virtual environment as apposed to the hypothetical data used to test the model in this chapter.
Chapter 7

Evaluating the Model – Experimenting with Real Data

Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>171</td>
</tr>
<tr>
<td>7.2 Comparing Spatial Data</td>
<td>172</td>
</tr>
<tr>
<td>7.3 Calibration</td>
<td>184</td>
</tr>
<tr>
<td>7.4 Validation</td>
<td>201</td>
</tr>
<tr>
<td>7.5 Summary</td>
<td>204</td>
</tr>
</tbody>
</table>

7.1 Introduction

Evaluating how a model performs is an essential aspect of any modelling research. Criminologists have commented that complex models can be difficult to understand which can detract from the interpretation of the underlying theories that drive them (Elffers and van Baal, 2008). To address this concern, Chapter 6 concluded by testing the model on idealised data and in varying types of geographical space to ensure that it had been programmed correctly and that the dynamics of the model were fully understood. Having determined that the model was behaving as expected, this chapter will further the evaluation of the model in preparation for it to perform experiments in Chapters 8 and 9. Following Castle and Crooks (2006), the process of evaluating the model has been segregated into three distinct activities: verification, calibration and validation. As verification was accomplished in Chapter 6, this chapter will discuss the process of calibrating the model to match field conditions and then validating it using data other than that it was calibrated on. These stages are accomplished by performing further sets of experiments, this time using realistic environmental and crime data rather than the idealised data used in Chapter 6. Comparing spatial data (i.e. real data and model results) is non-trivial so the chapter will begin with Section 7.2 which will investigate into how sets of spatial data can be compared. This will be followed by Section 7.3 which will discuss how the model was calibrated and finally Section 7.4 will validate the model.
7.1.1 The Process of Calibration

Calibrating a model is the process of configuring the model parameters with the aim of producing output that matches some field conditions. Often this is accomplished by comparing model results to an expected data set (recorded burglaries in this case) and configuring the parameters to provide the lowest error between the simulated and expected data. Automatic methods which vary numerical parameters systematically in search of the optimal configuration are used regularly, see Heppenstall et al. (2006), Malleson (2006) or Malleson et al. (2009) for examples. However, the complexity of this model does not allow for such neat calibration techniques for three reasons: firstly, several parameters cannot be translated to a ranged numerical scale (e.g. the starting locations of the agents or their behavioural characteristics); secondly (as Section 7.2 will illustrate) determining the goodness-of-fit between simulated and expected data is non-trivial for spatial models; and thirdly, the considerable computation time of the model deems some methods impossible to run on available hardware (even when utilising a high-performance computing grid). Therefore in the absence of formal calibration methods, model calibration will be performed manually through analysis of results and varying parameters based on the theory and knowledge of the dynamics of the model.

7.1.2 The Process of Validation

Model validity is the extent to which the model is able to represent the system it is attempting to simulate (Casti, 1997). Commonly, validation refers to the process of applying the model to a new system other than the one it was calibrated on. If the model is still able to perform reasonably then it indicates that the model has not been over-fitted to the original data set and can be safely applied to different scenarios. Here, validation poses somewhat of a problem because although there is sufficient crime data available, this is not the case for environmental data. Large parts of the model are based on the UK census which is only available for 2001. Section 7.4 will discuss how this problem can be mitigated.

7.2 Comparing Spatial Data

A significant part of this chapter and Chapters 8 and 9 will be spent comparing simulation results to each other and to expected data. It is therefore essential to fully understand the advantages and disadvantages of the methods that might be used. Both simulated and expected data are in the form of point patterns representing occurrences of burglary. To test the different point comparison methods available throughout this section, three point patterns will be used:

- **Model1**: Sample model data. Some results from a preliminary scenario.
- **Model2**: Sample model data produced using an identical simulation to the one used to generate the Model1 data set. The model is probabilistic so although there are small differences in the data they generally similar (as illustrated by Figure 7.1).
7.2. COMPARING SPATIAL DATA

- **Random**: A completely random data set, generated by a simulation of a random spatial process. Point patterns such as these are said to exhibit complete spatial randomness (O’Sullivan and Unwin, 2003).

The three data sets are illustrated by Figure 7.1 using the kernel-density algorithm to generate the raster surfaces (the algorithm is discussed in more detail in Section 7.2.1). The ArcGIS “Create Random Points” tool (ESRI, 2006) was used to generate the random points within the simulation boundary. Each data set contains the same number of points. For most of the methods used in this section, changing the number of points does not influence the results so long as the distribution of the data is the same. Where different numbers of points will affect results this will be discussed explicitly.

It must be noted that choosing suitable thematic thresholds for maps is non-trivial. Malleson et al. (2010a) compared a number of threshold schemes (equal interval, natural breaks, equal count and geometric interval) and found that the equal interval scheme was the most reliable at presenting the underlying data and was the easiest to interpret. Therefore this scheme, whereby each thematic interval has the same length, will be used in all maps unless stated otherwise.

### 7.2.1 Comparing Point Patterns Visually

Simply displaying point data on a map is not generally informative of the overall pattern, particularly for patterns which contain large numbers of points. For example, in Figure 7.1 it is extremely difficult to discern how dense the point patterns are by purely looking at the locations of the individual points. It is unclear where points overlap and identifying the density relies on human objectivity. Another problem, which is unrelated but no less significant, is that data protection or privacy rules often forbid point locations from being displayed because they will identify individuals. In crime studies such as this one, this is often a problem because points regularly centre on individual houses. Therefore there are alternative methods which can be used to describe point patterns that are both more informative and will not identify individual locations.

### Aggregating up to Administrative Areas

An alternative to displaying or analysing “raw” point data is to first aggregate the points up to a larger area. Then it is possible to analyse the aggregated data and produce thematic maps of the counts of points in each area or calculate rates. This is the approach taken in similar research (Kongmuang, 2006; Shepherd, 2006) and also by numerous police services who produce on-line crime maps for the public, e.g. Metropolitan Police (2009). However, as illustrated by the test data in Figure 7.2, this approach clearly suffers from the modifiable areal unit problem which is evident by the degree that the maps differ depending on the scale of the boundaries used. Furthermore, choosing too large a boundary hides important intra-area differences, seriously misrepresented the underlying pattern. Experimenting with using rates rather than absolute counts was no more promising.
Figure 7.1: The three point data sets which will be used to experiment with spatial methods throughout the remainder of this section.
Figure 7.2: Counts of points aggregated up to (Super) Output Area administrative boundaries.

For the reasons given above, maps of point patterns aggregated to administrative boundaries are generally considered extremely misleading by crime analysts and should not be used to illustrate crime patterns (Chainey, 2009a). A solution to these problems can be sought by aggregating up to arbitrary areas rather than those defined for specific administrative purposes. Aggregation to a square grid, for example, provides a better representation of the structure of the point pattern and is also considerably less susceptible to the modifiable areal unit problem (Malleson et al., 2010a). This is consistent with the crime mapping literature (Chainey and Ratcliffe, 2005; Eck et al., 2005) and is a technique that is explored in greater detail in Section 7.2.4.

Creating density surfaces

The previous section illustrated that aggregating points to administrative boundaries is not a reliable method for crime clusters. Grid thematic mapping was introduced as an alternative, but the “blockiness” of grid thematic maps makes them less visually appealing which can be a problem if
crime maps need to be shown to stakeholders (Chainey and Ratcliffe, 2005).

An alternative approach, which is becoming the most commonly used by crime analysts and is seen as the most reliable (Chainey, 2009b), is the use of density estimation algorithms. Kernel density estimation (KDE) methods are commonly used to estimate the density of points at every location in a study region. They operate by counting the number of points within a given distance (or 'kernel'). O’Sullivan and Unwin (2003) describe a “naive” method to estimate the density at a point, \( p \), as:

\[
\text{density}_p = \frac{\#(S \in C(p,d))}{\pi d^2}
\]  

(7.1)

where \( d \) is the kernel distance, \( C(p,d) \) is a circle of radius \( d \) centred at point \( p \), \( S \) is the set of all points and \# means the “number of” (as in Bailey and Gatrell (1995)). It is common to place a square grid over the study region and then estimate the density at the centre of every cell. Figure 7.3 illustrates that the size of the kernel (\( d \)) is an extremely important determinant of the density at each cell whereas the cell size only changes the resolution of the map. Values of \( d \) that are too large can result in ‘over-smoothing’ where two separate clusters appear as one. Similarly, values that are too small result in a single cluster around each point, also failing to capture the locations of clusters or the overall point density. In extreme cases, large \( d \) values result in similar density estimates for all cells (which will be close to the average point density) and small values result in zero density for all cells but those that have a point in them (O’Sullivan and Unwin, 2003). It is therefore advisable to experiment with bandwidths and to choose appropriately for the phenomenon under investigation. There are various methods which can be used to improve density estimations and the most common is to weight closer events more highly than distant events. In most analyses performed here, a quadratic kernel function is used as described in Silverman (1986) and implemented in the ArcGIS software package (ESRI, 2006).

Along with KDE, there are other methods which can be used to pinpoint hotspots visually. One with great potential that is still under-used in the field of crime mapping (Chainey, 2009a) is the GI* statistic (Getis and Ord, 1992, 1996). The main advantage is that it is is able to provide
a significance test which indicates how statistically significant a potential hotspot is. Using this information, thematic thresholds can be generated from the significance levels which allows the thematic thresholds to remain consistent throughout all maps. Although GI* appears to be a better predictor of cluster locations it is less visually pleasing and makes it more difficult to interpret the underlying point pattern when compared to KDE (Malleson et al., 2010a). As density maps in this research will not be used to perform any numerical analysis – they will only be used for visual data comparisons – KDE is a better choice of method.

### 7.2.2 Comparing Point Patterns Mathematically

Whilst methods to compare point patterns visually are essential, there is still much which remains susceptible to human objectivity. The maps used previously are able to indicate which patterns look similar, but this on their own does not provide sufficient evidence for (dis)similarity. Table 7.1 summarises a number of useful spatial statistics that can describe the distributions of point patterns and can be used to make mathematical comparisons.

Table 7.1: A summary of spatial statistics that can be used to describe and compare the spatial structure of point patterns

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Pros / Cons</th>
<th>Usage with Example Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest Neighbour Index (NNI)¹</td>
<td>• Gives a concise general picture of whether or not clustering is present (compared to random data)</td>
<td>Malleson et al. (2010a) showed that the model data are significantly clustered at the 99.99% significance level so there is utility in performing hotspot analysis later.</td>
</tr>
<tr>
<td></td>
<td>• Useful as a preliminary procedure (Bailey and Gatrell, 1995).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• It is difficult to account for edge effects³.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Too simplistic to be really useful.</td>
<td></td>
</tr>
</tbody>
</table>

¹Also known as the Clark and Evans $R$ statistic (Clark and Evans, 1954)
²The nearest-neighbour distance for a point $i$ is the distance to the closest neighbouring point.
³There are some solutions (e.g. circular or rectangular corrections (Levine, 2006)) but these are not ideal (Chainey and Ratcliffe, 2005).
<table>
<thead>
<tr>
<th>Statistic</th>
<th>Pros / Cons</th>
<th>Usage with Example Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>The $G$ function.</td>
<td><strong>Pros / Cons</strong></td>
<td>The following illustrates the $G$ functions for the Model1, Model2 and Random</td>
</tr>
<tr>
<td></td>
<td>- Describe clustering in more detail than the NNI index by providing a</td>
<td>data sets (graphs of $F$ are similar). Model1 and Model2 functions rise more</td>
</tr>
<tr>
<td></td>
<td>measure at different distances.</td>
<td>sharply than that of the Random points which implies that more points have</td>
</tr>
<tr>
<td></td>
<td>- Can be used to differentiate between clustered and uniform data (see</td>
<td>close nearest neighbours in the model data than the Random data. This is</td>
</tr>
<tr>
<td></td>
<td>O’Sullivan and Unwin (2003) for a fuller discussion).</td>
<td>indicative of clustering and supports the NNI test. Also, it shows that clustering in</td>
</tr>
<tr>
<td></td>
<td>- Only consider a single nearest neighbour distance in their calculations so</td>
<td>the Model1 and 2 data are more similar to that of the random data.</td>
</tr>
<tr>
<td></td>
<td>disregard a considerable amount of information.</td>
<td></td>
</tr>
<tr>
<td>$G = \frac{#(d_{\text{min}}(s_i) &lt; d)}{n}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Ripley’s $K$ Function</strong> counts, for each point $i$, the number of points</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$s_i$ that are within a distance $d$. $K(d)$ is then the mean of all counts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>divided by the overall point density, defined by O’Sullivan and Unwin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2003) as:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sum_{i=1}^{n} \frac{#(S \in C(s_i, d))}{A/n}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Takes all the neighbours that are within a given distance into account so</td>
<td>As with the $G$ and $F$ functions, graphs of $K(d)$ reveal information about the</td>
</tr>
<tr>
<td></td>
<td>is a more descriptive statistic.</td>
<td>clustering of the points but can be hard to interpret. The graph illustrates that, at</td>
</tr>
<tr>
<td></td>
<td>- Graphs for non-hypothetical situations (such as this) are very difficult</td>
<td>greater distances, there is a lower point density in the Model data than the random</td>
</tr>
<tr>
<td></td>
<td>to interpret.</td>
<td>data.</td>
</tr>
<tr>
<td></td>
<td>- The statistic is informal; it suggests how the data are clustered but does</td>
<td></td>
</tr>
<tr>
<td></td>
<td>for whether or not this clustering is more or less than would be expected</td>
<td></td>
</tr>
<tr>
<td></td>
<td>provide any formal evidence for whether or not this clustering is more or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>less than would be expected under randomness.</td>
<td></td>
</tr>
<tr>
<td>$K(d)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 7.2. COMPARING SPATIAL DATA

#### Statistic

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Pros / Cons</th>
<th>Usage with Example Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>The <em>L</em> Function is a transformation of <em>K</em> that provides evidence for whether or not clustering is more or less than would be expected under complete spatial randomness (O’Sullivan and Unwin, 2003):</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Using CSR as a baseline often provides little further insight because almost all natural processes do not exhibit spatial randomness. • A better approach, which is commonly used in the field of crime mapping, is to compare graphs of <em>L</em>(d) to determine how similar their clustering is.</td>
<td></td>
</tr>
<tr>
<td>[ L(d) = \sqrt{\frac{K(d)}{\pi}} - d ] (7.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values of <em>L</em>(d) &lt; 0 suggest that there are fewer events in the space than would be expected under CSR and that the data are therefore less clustered. The reverse is true for <em>L</em>(d) &gt; 0.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The following graph illustrates that the Model1 and Model2 data are more clustered than the random dataset (which equates to approximately <em>L</em>(d) = 0 for low d values). Above d ≈ 3000 <em>L</em> begins to fall due to boundary effects.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the statistics outlined in Table 7.1, the *L* function appears to be the most appropriate mathematical technique for concisely comparing two spatial data sets. The differences in the degree of clustering between the Model and Random datasets are apparent using the function and it has the added benefit of suggesting whether or not the clustering is greater than that expected under spatial randomness. For these reasons the statistic will be used in later sections as one means of comparing two spatial data sets.

#### Raster Map comparisons

Other than comparing mathematical descriptions of clustering in point patterns, it is also possible to generate raster density maps from the point patterns (using the KDE algorithm) and compare these mathematically. This approach is common in the field of spatial modelling for comparing simulated and real land use. For a review of recent approaches, the reader is directed to Kuhnert et al. (2005).

Figure 7.4 illustrates a density surface produced using the “Fuzzy Kappa” statistic (Hagen, 2003) to compare Model1 and Model2 data. The stages in the analysis used to calculate Fuzzy Kappa and the places where error can arise are as follows:

1. **Creating a raster surface.** Using the KDE algorithm, both the size of the kernel and the weighting method can be varied which will significantly influence the density surface produced.

---

4For example, comparing graphs of burglary and known drug dealer locations provides evidence for or against the hypothesis that burglary clusters around known drug dealers. Similarly, a comparison of *L*(d) graphs for knife crime and stop-and-search occurrences can provide the police with evidence for how well their crime reduction initiatives are being targeted at crime hotspots (Chainey, 2009a).

5This is because many of the large circles produced by the underlying *K* function are nearly empty at large distances because they cover areas outside the simulation boundary with no points (O’Sullivan and Unwin, 2003)
Figure 7.4: Using the Fuzzy Kappa statistic (Hagen, 2003) to compare Model1 and Model2 data. The overall kappa value is 0.616 which some suggests similarity (a value of 1 indicates identical maps). The results map illustrates the areas of (dis)similarity.
2. **Defining categorical values.** To convert the density surface to a categorical raster map it is necessary to define numerical ranges for the categories (i.e. define “high” and “low” crime). However, numerous threshold methods could be used to set the range values (such as equal interval, equal count, standard deviations etc.) and the number of categories can be varied: both of these will affect the categorical map created.

3. **Raster comparison statistics.** The Fuzzy Kappa statistic itself also uses numerous parameters which will influence how it behaves. These include the size of the neighbourhood, the distance decay function (exponential, linear or constant), parameters within the decay function itself and a definition of the similarity of categories (Hagen-Zanker et al., 2005).

In summary, because there are so many arbitrary parameters it is theoretically possible to completely change the result of the Fuzzy Kappa statistic. Furthermore, as the point patterns themselves are available to analyse, it seems unnecessary to lose a great deal of information by converting them to categorical maps. For these reasons, categorical raster map comparison methods will not be used further. The following sections will outline the development of a more appropriate technique for making mathematical and visual spatial data comparisons.

### 7.2.3 Goodness-of-Fit Statistics

Goodness-of-fit (GoF) statistics are used to describe how well a model matches a set of observations. Knudsen and Fotheringham (1986) experimented with a number of goodness-of-fit statistics and found the standardised root mean square error (SRMSE) to be the best performing. A drawback with SRMSE, however, is that it is difficult to interpret. An alternative statistic, $R^2$, solves this problem because it represents the percentage of agreement between the model and the expected data. However, $R^2$ is insensitive to the overall amount of error, predicting a good fit in some circumstances where the SRMSE would not (Harland, 2008).

The SRMSE is defined by Heppenstall (2004) as:

$$SRMSE = \sqrt{\frac{\sum (y'_i - y_i)^2}{\bar{y} n}}$$ (7.7)

where $y'_i$ is the predicted value at matrix point $i$, $y_i$ is the actual value at $i$, $\bar{y}$ is the mean value of the predicted values ($y'$) and $n$ is the total number of values. The lower limit of the statistic is 0 which indicates no difference between the predicted values ($y'_i$) and the observed values ($y_i$). The upper limit is usually 1 (Knudsen and Fotheringham, 1986) but can be greater, particularly when matrices are sparse (Harland, 2008).

Using the same notation, $R^2$ can be defined as:

$$R^2 = 1 - \frac{\sum (y_i - y'_i)^2}{\sum (y_i - \bar{y})^2}$$ (7.8)

A value of 1 indicates identical data sets and the lower limit of the statistic is 0. To use GoF statistics, the point patterns must first be aggregated in order to create comparable matrices. It
is common to aggregate up to a particular administrative boundary to perform the statistics (e.g. Kongmuang, 2006) and Table 7.3 illustrates the GoF values for the test datasets at different resolutions. The accuracy of the statistics vary considerably depending on the administrative boundaries used but both statistics reveal more similar data at larger resolutions (which is to be expected).

Table 7.3: Goodness-of-Fit at different resolutions.

<table>
<thead>
<tr>
<th>Administrative Boundary</th>
<th>Model1 / Model2</th>
<th>Model1 / Random</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SRMSE $R^2$</td>
<td>SRMSE $R^2$</td>
</tr>
<tr>
<td>Output Area</td>
<td>0.52 0.64</td>
<td>0.98 0.24</td>
</tr>
<tr>
<td>Lower Super Output Area</td>
<td>0.23 0.88</td>
<td>0.59 0.47</td>
</tr>
<tr>
<td>Medium Super Output Area</td>
<td>0.12 0.98</td>
<td>0.35 0.83</td>
</tr>
</tbody>
</table>

The main problem with this approach, as discussed in Section 7.2.1, is that the process of aggregating to administrative areas leads to the modifiable areal unit problem, whereas aggregating to a regular grid reduces these problems. Therefore, as an alternative to aggregating to administrative boundaries, a new approach is proposed based on that defined by Costanza (1989). As this technique will form the bulk of the model validation it is discussed in greater detail in the following section.

7.2.4 Expanding cell validation method

As an alternative to aggregating a point pattern up to administrative boundaries, the expanding cell method is an adaption of a technique proposed by Costanza (1989). The method works by aggregating simulated and expected data to various regular grids of decreasing cell size. Traditional goodness-of-fit statistics can then be applied as normal by comparing the counts of points from the two data sets in each of the grids. Both $R^2$ and SRMSE can be used; SRMSE because it provides a more suitable measure of fit for this application and $R^2$ because it is easier to interpret. In this manner, GoF results for a range of resolutions can be presented to provide a more comprehensive assessment of the spatial accuracy of the model. This approach has the advantage of reducing the effects of the ecological fallacy. To reduce the influence of the MAUP, at each resolution the grid will be shifted slightly in the northerly, southerly, easterly and westerly directions to provide a total of five different similarity estimates for the data at a given resolution.

Graphs of the global errors ($R^2$ and SRMSE) reveal information about how well a model is performing. There are various parameters which could be used on the x axis (such as cell length or the total number of cells in the grid) but the most appropriate here is the square area of the cells. This will allow us to investigate the spatial scale at which the model is able to make accurate predictions. In the following analyses, an expected dataset will be utilised along with the Model and Random data that consists of all the burglaries that occurred in the simulation area in 2000 – 2002 (see Section 4.4 for a full description of the data). Figure 7.5 illustrates the SRMSE and $R^2$ errors that arise when comparing both the model to expected data and random to expected data. In both cases, as the cell size increases (up to the point that a single cell covers the entire point
7.2. COMPARING SPATIAL DATA

pattern) the errors decrease. This is to be expected (Gehlke and Biehl, 1934; Robinson, 1950). The graphs also show that at all but the smallest cell sizes the model data are a better approximation of the expected data than the random data set. This is also to be expected but is nevertheless comforting that the model is a better predictor of crime than a random process!

Figure 7.5: SRMSE and $R^2$ errors comparing model data and random data to expected data.

Along with generating reliable GoF values from well-known statistics, the expanding cell method is also able to produce spatial estimates of error. Once the counts of points within each cell have been calculated, it is possible to derive the error associated with each individual cell and map this. Various definitions of “error” can be used, but it is important to recognise that in some cases the total counts of points might not be equal (the model might not simulate the total number of burglaries exactly) so proportional values are used. Here, the relative percentage error between the cells is appropriate because it is easy to interpret and accounts for variable numbers of observations. The relative percentage error between two cells, $y_i$ and $y'_i$, is defined as the difference between the proportions that the cells contribute to the total observation count:

$$100 \times \left( \frac{y_i}{\sum y} \right) - 100 \times \left( \frac{y'_i}{\sum y'} \right)$$

(7.9)

Figure 7.6 illustrates these errors spatially using different resolution grids. At the finest resolution, a 41\*41 cell grid, the cells appear too small to properly capture the differences between the datasets, and the results look like random noise. However, at coarser resolutions the expanding cell method is clearly able to distinguish between similar data sets (Model1 compared to Model2) and dissimilar ones (Model1 compared to Random). The figure also demonstrates a further benefit of shifting grids: that the results from all grids can be overlaid to give a comprehensive, “fuzzy”
interpretation of the difference between the two input data sets. This method of summarising the results appears to be the most concise and the most informative and will be used when examining model results in later sections. For the interested reader, further details of the expanding cell method and the influence of shifting grids on the MAUP and ecological fallacy can be found in Malleson et al. (2010a).

### 7.2.5 Summary - methods used to validate models

This section has experimented with a number of different methods for both summarising and comparing point patterns. On the basis of this work it was decided that the the most reliable method of visualising a point pattern is to produce a map of the density of the points using the kernel density estimation (KDE) algorithm. To compare point patterns there are two methods that show promise. Graphs of the $L$ functions of two point patterns can be compared to determine how similar the clustering of the patterns is and the expanding cell method can be used to produce maps and graphs describing the difference between two point patterns at different resolutions. These methods will be used in subsequent sections to compare results from different model configurations and to expected data.

### 7.3 Calibration

Calibration refers to the process of configuring a model’s parameters to match some observed (also termed “expected”) historical data (O’Sullivan, 2004). Usually this consists of searching for a combination of parameters that cause the model to produce data which is similar to that collected from the real system under investigation. In many cases, a single value can be found which summarises how closely the simulated and field data match, this is commonly called “fitness” (Gilbert and Terna, 2000). Section 7.2.3 reviewed statistics that can be used to establish fitness in this application. Algorithms can then be used to search the model parameter space looking for the combination of parameters which produce the greatest fitness (i.e. the combination that produces outputs that are the closest match to field data). Examples of the use of a genetic algorithm to optimise an agent-based model exist in crime (Malleson et al., 2009) and retail (Heppenstall et al., 2006) among others. Unfortunately, using an automatic calibration routine is beyond the scope of this research for two reasons. Firstly some parameters that will have a strong influence over model results do not translate to a numerical scale (such as the starting locations of the burglars or the addresses of drug dealers). More importantly, however, is that a conservative estimate for the run time of a single automated calibration routine running on a high-performance computing grid is more than thirty days\(^6\).

---

6 Assuming that the automated routine requires 100 iterations to reach optimal fitness (Malleson, 2006) and each model requires 50 separate runs to ensure the results are stable (discussed in Section 6.6.2) that is $100 \times 50 = 5000$ individual runs. If a node on the high-performance computing grid requires 20 hours to complete a model run (discussed in Section 6.5.6) and there are a total of 128 nodes available the total time is $(5000 \times 20)/128 = 781$ hours, or 32.6 days. Furthermore, this is conservative because the compute resource will be required by other users at the same time, reducing the number of nodes available to this researcher.
Figure 7.6: Relative percentage error comparing model data to randomly generated data
Instead, the model will be calibrated manually, finding the best fit to known data by visually and mathematically comparing results and adjusting parameters accordingly. Here, model fitness will be determined by producing graphs of the SRMSE and $R^2$ errors at different spatial resolutions as discussed in Section 7.2.4.

7.3.1 Calibration (“Base”) Scenarios

The following scenarios are termed “base” because they represent the foundation on which later forecasting experiments will be based. They incorporate criminology theory and real data analysis and will simulate based on input parameters from the 2001 census. The scenarios are based in the area of the EASEL urban regeneration scheme which was discussed in Section 4.2 plus a 1km buffer around the area. The purpose of the buffer is to limit boundary effects such that all crimes committed in the buffer area will be disregarded when results are analysed. Ideally the buffer would be larger, e.g. 3km is more than the average distance travelled by burglars found in many studies (Baldwin and Bottoms, 1976b; Bottoms et al., 1992; Wiles and Costello, 2000; van Nes, 2006; Bennell et al., 2007). However, computing requirements increase exponentially with the radius of the environment (due to the increasing number of offenders and environmental objects) and 1km is an acceptable compromise. The buffer area is considered the same as the inner EASEL area by agents, but when results are collected only crimes committed within the EASEL area will be counted. All environment variables will have the values outlined in Chapter 5, as calculated from the census and the geographical data. Figure 7.7 summarises where the different types of buildings in the environment are located. The drug dealer locations were established directly from the crime data, creating a virtual dealer address for every point in the data set where a dealing-related crime had been recorded. The starting locations and number of offender agents were also determined directly from the crime data, creating a point for every known burglar in the offender data for the period 2001 – 2002. This results in 273 burglar agents. An immediate area for future work is to analyse the offender data in more detail with the aim of creating offenders of different types (as discussed in Section 5.11) but this is beyond the scope of this research.

It is also necessary to decide, from the set of all available locations, where an agent chooses to travel if they need to visit a drug dealer or to socialise. This is one of the most difficult features to estimate as there is very limited data which can assist with the assumptions. With regards to drug dealers, it is decided that the agent is assigned a drug dealer at random and always uses the same one. It is likely that in reality a person builds a preference for a certain dealers but often travels to different addresses depending on the abundance of supply, but this avenue of exploration is not in the scope of this work. With regards to social locations, it is assumed that an agent is more likely to travel to a social location that is in a community of a similar type to their own. Again this is likely to be too simplistic but can be investigated further if required. Section 10.3 will discuss the implications of having to make such broad assumptions in the context of the overall choice of methodology.

---

7The crimes defined as related to drug dealing (and their Home Office crime codes) are: 77/2 – Supply / Offer to supply drugs; 77/4 – Possess with intent to supply drugs; 92/3* – Supply drugs (Class A); 92/4* – Supply drugs (Class B or C).
7.3. CALIBRATION

Figure 7.7: The locations of different building types, established from crime, census and land-use data (for more details refer to Chapter 5).

7.3.2 Base 1 – Default Conditions

This section will present the results of the first scenario, i.e. executing the model under default conditions. As with previous experiments, the probabilistic nature of the model means it was executed 50 times. To ensure that results are consistent, Figure 7.8 graphs the spatial distribution of all models in the form of L functions (recall that Section 7.2.2 explained that the L function provides a measure of clustering at a given distance). The graph suggests that the spatial distribution of the point patterns in the models are similar as the lines representing the L functions are close together. Notice that the degree of clustering is greater than a random dataset but less than the expected data; the implications of this will be discussed shortly. It is possible (although unlikely) that different point patterns could produce similar L functions if the degree of clustering was the same but the clusters were in different places. To guard against this possibility, maps of the density of four randomly chosen models were generated and show that the point densities are visually as well as mathematically similar (for details see Malleson et al., 2010a). Therefore it is safe to continue analysing the model results assuming that all models are similar.

As with most agent-based models, the model produced for this research can run indefinitely; there is no limit to the maximum number of iterations. Therefore it is necessary to determine when the simulation should terminate so that the results can be analysed. For the purposes of
this research, *equilibrium* can be defined as the iteration after which the model results will not change even if the model were allowed to continue to run indefinitely. This was also discussed when the prototype was introduced in Section 6.4. After some trial and error tests, 30 simulated days (43,200 iterations) was found to be sufficient to allow the model to reach equilibrium. This is illustrated by Figure 7.9 which graphs the difference in the L functions for a single randomly chosen model at four different simulation times. The small difference between the third and final time points on the graph indicates that the simulation has reached equilibrium in that time. This is supported by density maps of the corresponding time periods that are visually similar (Malleson et al., 2010a). Therefore it is safe to assume that the model reaches equilibrium after 30 simulated days and does not need to be run for a longer time.

Having determined that the simulation results are consistent and have reached equilibrium, Figure 7.10 compares the burglary density of a single model to that of the expected data. In both (and all subsequent) cases, the same parameters were used in the kernel-density algorithm (a 350m bandwidth with a cell size of 20m – chosen through trial-and-error because they produced the most appropriate visual summary). To supplement this, the expected and simulated point patterns are compared directly using the expanding cell algorithm. Figure 7.11 maps the results of the analysis at two different resolutions and Table 7.4 illustrates the global errors found by the expanding cell algorithm at the two mapped resolutions.

Observing Figure 7.10, the model does not seem to be simulating burglary accurately (although this is to be expected with an un-calibrated model). Simulated burglaries are distributed throughout the environment whereas in the expected data they are more concentrated in hotspots. In observing the expanding cell map (Figure 7.11) another discrepancy becomes apparent: there
7.3. CALIBRATION

Figure 7.9: Graph of the difference in L functions at different time points (T1=25%, T2=50%, T3=75% and T4=100% total time) in a Base 1 simulation.

Table 7.4: Base 1 model errors at different cellular resolutions

<table>
<thead>
<tr>
<th>Number of Cells</th>
<th>Resolution (Cell Area, km²)</th>
<th>SRMSE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>0.65</td>
<td>0.84</td>
<td>0.61</td>
</tr>
<tr>
<td>441</td>
<td>0.18</td>
<td>1.05</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Figure 7.10: KDE maps comparing Base 1 model results hotspots to expected data. Both maps have the same thematic ranges; the model results are notably less clustered.
Expanding Cell Results for a Single Model at Different Resolutions

Figure 7.11: Expanding cell maps for a Base 1 Scenario
are a small number of cells in which the model is significantly under-predicting the number of crimes. This is further illustrated by the frequency distribution of cell errors in Figure 7.12. The large model under-estimations relate mainly to a single neighbourhood in the south-east of the EASEL simulation area called ‘Halton Moor’.

![Histogram of individual cell errors](image)

Figure 7.12: A histogram of all expanding cell Base 1 errors at the 0.85km\(^2\) resolution. The peak between 0 and -0.25 is a result of the greater dispersion of burglary in the simulated data compared to the expected data. The largest under-estimates mostly relate to an area called Halton Moor.

Before continuing with the calibration process, it is important to determine why the errors are so large in this particular area. It is possible that the crime data are actually misrepresenting the number of crimes committed in Halton Moor. For example, a geocoding error might have placed a number of crimes in the area that do not belong there. A manual inspection of the data suggest this is not the case and, as Figure 7.13 illustrates, the crimes appear to be realistically distributed (there is not a single address that accounts for all crimes in the area for example). Therefore it appears that the hotspot in Halton Moor is a realistic portrayal of crime in the area and not a result of data errors.

It is also possible that a large number of offenders are travelling to the area from a long distance away. As discussed in Section 2.2.2, this has been found to occur if an area is particularly attractive to burglars. If this is the case, and the offenders in the real world are travelling from outside the area that is being simulated, then they will not be accounted for in the simulation. This would explain why the model was not correctly simulating crime in the area. However, Figure 7.14 visualises the offender dataset and demonstrates that most of the crimes are committed by offenders who live within the area of the hotspot. Although the offender data are extremely sparse and probably a poor indicator of offending activity, it nevertheless gives no indication that crimes are being committed by people from outside the EASEL area.

As the crime data fail to provide a reason for why the hotspot should exist in Halton Moor, it is possible to search for one in the community and census data instead. Recall that the Output Area Classification (outlined in Section 4.3.3) classifies communities into distinct groups based on the 2001 UK census. Examining these groups in EASEL and Halton Moor (Figure 7.15) does not
Figure 7.13: Real crime hotspots occurring in the Halton Moor area.
7.3. CALIBRATION

Figure 7.14: Journey-to-crime in the Halton Moor area. The starting locations of nominals who were associated with a crime that occurred in the Halton Moor hotspot, from the offender dataset (see Section 4.5).

reveal any reasons for the hotspot. At the supergroup level (the highest level in the classification hierarchy) there are no particular differences between Halton Moor and other parts of EASEL. For example, most of Halton Moor is made up of “blue collar” and “constrained by circumstances” groups but these are prevalent throughout the EASEL area.

In summary, the available data provides no explanation for why the model is unable to predict the existence of a crime hotspot in the Halton Moor area. To try to gather further information, experts at Safer Leeds were consulted regarding the hotspot. It was suggested that, at the time, the area had some severe social problems and burglary was thought to be used often as a form of intimidation. However, the usual assumption (as implemented in the model) is that burglars are largely motivated by monetary gains. In this case, the model is able to demonstrate where, specifically, common assumptions about burglary fail. This is an extremely interesting result and will be revisited when the results are discussed in Section 7.5. In the meantime, the process of calibrating the model will proceed, disregarding the inability to simulate the Halton Moor hotspot.

7.3.3 Base 2 – Faster Security Decreases

Section 7.3.2 found that the model was unable simulate a particular crime hotspot and that this was possibly a result of incorrect assumptions regarding burglar motivation. Although the model has
been designed to incorporate different types of motivation, there are other means of calibrating the model more generally first. Therefore the absence of a hotspot in Halton Moor will be disregarded for the time being and alternative methods will be used to improve the accuracy of the model. The first of these, which will be tested in this scenario, relates to the dispersion of burglaries.

Figure 7.8 provided a graph of the L functions for all the Base 1 simulations and compared these to the L function produced by the expected data. As the L values for the simulated data are less than the expected data at all distances, there is less clustering in the model than suggested by the expected data. This is also apparent by visually comparing the density maps (see Figure 7.10).

One reason for the greater dispersion might be due to the effects of household security. When a property is burgled, it has been assumed that both the burgled house and its neighbours increase their security levels in response to the new threat. Even though agents are subsequently more strongly attracted to the area (because they remember that they have been successful there), the high security levels make it more difficult to actually commit a burglary in the area. A closer inspection of security changes indicates that, in this manner, security in the simulation increases indefinitely. Figure 7.16 demonstrates this with a frequency distribution of the security levels in all buildings at different time points in the simulation. Initially most buildings have very low levels of security, but over time security increases indefinitely.

By making security decrease more rapidly, it should be possible to address this error and reduce the dispersion (as there will no longer be areas with extremely high security that are effectively impossible to burgle). This can be accomplished by changing two parameters. Firstly the rate at which security decreases is adjusted so that it halves over the course of a week (reducing
by \( \left( \frac{1}{2} \right)^k \) per day) rather than simply by decreasing by 0.5. Secondly, the amount that security will increase in surrounding properties following a burglary is reduced. Now, instead of an increase of 2 units, houses within 1 distance unit of a burglary will only receive an increase of 0.5 units (and this value decays as normal, see Section 5.10.4). These numbers were established through trial and error as empirical evidence to support their value is limited. For example, quantifying how quickly a household becomes complacent following a burglary, if they do at all, is non trivial and beyond the scope of this research. However this is the advantage of these types of models; they provide a foundation for expert-driven experimentation on such parameters.

As with the previous simulation, it was confirmed that scenario results were consistent across model runs and that the simulation reached equilibrium (see Malleson et al., 2010a). Figure 7.17(a) contrasts the burglary density of a typical model to the expected data. The data still appear distinct but the new simulation results appear to be a closer match to expected data than the previous scenario (Figure 7.10). To provide mathematical evidence for this, Figure 7.17(b) graphs the overall error of the scenarios and shows that, interestingly, the changes appear to have actually lowered the accuracy of the model slightly. It is possible, however, that this affect might simply be an artefact of the way in which the goodness-of-fit statistics are calculated. In the Base2 scenario crimes are less evenly distributed about the environment. This is illustrated in Figure 7.17(c) which shows that the L functions for the scenario suggest much higher levels of clustering. This means there are more cells with no crimes in them (the data are more sparse) which can influence the goodness-of-fit statistics (Knudsen and Fotheringham, 1986).

Regardless of whether or not Base2 is a poorer simulation based on the distribution of burglary, it is a better simulation in the sense that it has brought the behaviour of the environment in line with
CHAPTER 7. EVALUATING THE MODEL – EXPERIMENTING WITH REAL DATA

(a) Maps comparing Base 2 model hotspots to expected data (using the same thematic ranges.

(b) Expanding cell graphs for a Base 2 scenario compared to Base 1

(c) L Functions for a number of Base 2 models compared to expected data.

Figure 7.17: Comparing Base2 results to Base 1 and expected data
our expectations. In Base1, house security increased indefinitely which is clearly inaccurate, this
problem was solved by the changes made to the second scenario. Therefore if we assume that the
environment is now behaving correctly in response to burglary (with respect to house security at
least), there might be utility in continuing the calibration even though it appears more appropriate
to disregard the Base2 changes because they did not improve the model fitness.

7.3.4 Base 3 – Agents Dislike Security

The previous experiment changed the default behaviour so that the dynamics of household security
changes after burglary were more appropriate. However, now it appears that burglary is too heavily
spatially clustered (when compared to expected data). An alternative to experimenting further
with security values is to change how the burglar agents perceive security precautions. At present,
burglars view all household parameters equally (they do not consider security to be a greater
deterrent than occupancy for example). This experiment will alter this so that, when deciding
whether or not to burgle, agents give a greater weight to security precautions, i.e. they are more
deterred by security than by other environmental factors. This can be accomplished by changing
the Security_W parameter which influences how strongly household security deters a potential
burglar. Deciding on a new weight, however, is non-trivial as there is no previous research that
can help to quantify a new weight value. Previous experiments suggest that a value of 5 creates
the appropriate behaviour: making crime moderately less heavily clustered.

As with all previous results, the model reaches equilibrium and is consistent across separate
runs (Malleson et al., 2010a). To begin analysing the results, Figure 7.18 graphs the L functions.
From the perspective of crime concentration, the results are extremely encouraging. The degree of
clustering in the Base3 model is very similar to that of the expected data and much more similar
than the previous calibration experiments.

![L functions: Model results compared to expected and random data.](image)

Figure 7.18: L Functions for a number of Base 3 models compared to Base 2 and expected data.
To explore the burglary distributions spatially, Figure 7.19 compares the burglary rates to expected data. There are still some difference between the two data sets but they are clearly more similar than the Base 1 (Figure 7.10) and Base 2 (Figure 7.17(a)) results.

Figure 7.19: Map comparing model Base 3 hotspots to expected data. Both maps have the same thematic range.

The improvements of the Base 3 calibration are further illustrated by the expanding cell analysis; Figure 7.20 maps the model errors at two different resolutions. Comparing this map to the original (Base 1) expanding cell map (Figure 7.11) it is clear that a much larger area has similar crime levels (the thematic ranges are the same on both maps). Table 7.5 provides the global errors at the two mapped resolutions and Figure 7.21 graphs the errors. Again the global goodness-of-fit statistics support the visual analysis and suggest that the calibration has improved the accuracy of the model (with the exception of the 0.18km\(^2\) resolution SRMSE measure which is slightly larger than the Base 1 equivalent).

Table 7.5: Model errors at different cellular resolutions. The errors that represent the greatest goodness-of-fit are bold.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Number of Cells</th>
<th>Resolution (Cell Area, km(^2))</th>
<th>SRMSE</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base 3 Scenario</td>
<td>121</td>
<td>0.65</td>
<td>0.76</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>441</td>
<td>0.18</td>
<td>1.10</td>
<td>0.54</td>
</tr>
<tr>
<td>Base 2 Scenario</td>
<td>121</td>
<td>0.65</td>
<td>0.86</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>441</td>
<td>0.18</td>
<td>1.27</td>
<td>0.44</td>
</tr>
<tr>
<td>Base 1 Scenario</td>
<td>121</td>
<td>0.65</td>
<td>0.84</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>441</td>
<td>0.18</td>
<td>1.05</td>
<td>0.50</td>
</tr>
</tbody>
</table>

7.3.5 Summary – Model Calibration

The calibration process has improved the global accuracy of the model (based on goodness-of-fit analysis) and has brought the spatial distribution of the resulting burglaries more in line with that
7.3. CALIBRATION

Figure 7.20: Expanding cell maps for a Base 3 Scenario.
of the expected data. It was noted that calibrating the model using an automated routine, such as a genetic algorithm, was infeasible. Instead the model was calibrated manually. There are both advantages and disadvantages to this approach to calibration. An advantage, as illustrated by the second scenario (Base 2), is that the researcher can use their knowledge of the model to apply changes that do not necessarily increase the fitness of a model and might be disregarded by an automatic routine. A disadvantage, however, is that the parameter space is not explored nearly as comprehensively as it would be by a computer algorithm. There are therefore various routes which could potentially be taken in order to calibrate the model further. The behaviour of the burglars could be improved in order to better simulate crime rates in the Halton Moor area which was very poorly simulated. It would also be interesting to analyse the SRMSE and $R^2$ error statistics in more detail to establish why they are occasionally in disagreement regarding the most accurate result. Or the error graphs could be analysed to estimate the resolution at which the model is able to make reliable predictions to. However, at this stage the model is deemed to perform adequately on the calibration dataset and will now be used to make predictions. Model calibration will continue to be an area for further research in the future.

### 7.3.6 Comparing the Results to Regression Models

Having calibrated the model, it is possible to compare the results to the regression models introduced as part of the data analysis stages in Section 4.6. As it was found that most crime models typically use regression (see Section 3.3) this will provide an insight into how this agent-based model compares to others in the field. Two regression models were presented, a typical OLS model and a more advanced GWR model that was better suited to handling spatial variation. Both models were generated using data at the super output area (SOA) level and produced $R^2$ values of approximately 0.70 (i.e. they were able to account for 70% of the variation in the dependant variable).
To compare this model to the regression models, the $R^2$ value was calculated at the SOA level by aggregating the number of burglaries in the expected and simulated (Base 3) data sets. Interestingly, the $R^2$ value for the model was extremely low at $R^2 = 0.078$. This suggests that the model was significantly outperformed by the regressions. However, the simulation was only applied to the EASEL area, whereas the regression models were calculated for the entire city. Therefore the OLS regression model was generated again using the same variables as the global model but only the SOAs that were used to calculate the agent-based $R^2$ value (those that make up the EASEL area). This resulted in $R^2 = 0.57$ which is lower than the global model value ($R^2 = 0.70$) but still considerably higher than agent-based value ($R^2 = 0.078$).

However, aggregating to administrative boundaries is fraught with problems (see Section 2.1.2). To avoid this arbitrary areal aggregation, the model results were instead aggregated to a square grid (as discussed in Section 7.2.4) so it is preferable to publish these $R^2$ values than those produced by SOA aggregation. Unfortunately it is not possible to use the cellular aggregation method for the regression models because the input data is only available at the SOA level (whereas the agent-based model generates point data which can be aggregated to any boundary). However, it is possible to generate the cellular-aggregated $R^2$ values for the agent-based model using cell sizes that are comparable to the average SOA cell size. This effectively allows the two models to be compared at a similar resolution but using different aggregation areas. The mean square area of the SOAs used in the previous calculations is 0.42km$^2$. Using the expanding cell method, with a cell size of 0.42km$^2$, results in $R^2 = 0.64$. Even after excluding two large outliers (there are two SOAs that extend a long way beyond the EASEL boundary) and therefore reducing the mean SOA square area to 0.32km$^2$, the error is still similar; $R^2 = 0.61$. Therefore, when aggregating to a square grid (which was shown to be a more accurate method of generating error statistics) the agent-based model ($R^2 = 0.61$) has outperformed a local OLS model ($R^2 = 0.57$), although the difference is only slight. The repercussions of this finding are discussed in greater detail in Chapter 10.

### 7.4 Validation

The previous sections have verified that the model works as expected (Section 6.6) and calibrated it to reflect the main trends and behavioural processes using 2001 field data (Section 7.3). The final task before the model can be used to perform experiments is to ensure that it has not been over-fitted to the available data and is able to forecast in situations outside that on which it was calibrated. This process is often termed validation.

Validating this research, however, poses some problems. The most significant is that most of the data used to create the virtual environment are based on the 2001 UK census and no data for any other years is available. Therefore validating the model on subsequent years will be inherently inaccurate because the environment cannot be updated to reflect the changes that will have occurred to the demographics since 2001. Furthermore, offender data are only available up to 2004 so for any other years there is no way to seed the offender home locations. Therefore the model will be validated on 2004 data. This is the most suitable year because it is close enough to
2001 so that the demographics of the EASEL area should be similar but far enough from 2001 so that crime patterns will be different.

7.4.1 Preparing the Crime Data

For consistency with the 2001 data, the 2004 victim and offender data sets would ideally cover the period 1st April 2003 to 31st March 2005 to make the data more spatially reliable. However, as offender data is not available past 31st March 2004 (discussed in Section 4.3.6) the period will be 1st April 2003 to 31st March 2004 instead. Figure 7.22 illustrates the crime hotspots generated from the victim data. For the interested reader, maps of offender locations can be found in Malleson et al. (2010a).

Figure 7.22: EASEL crime hotspots in the 2004 data.

7.4.2 Validation Results

The model was configured using the final calibration configuration (Base 3) and, as with all other experiments, was run for 30 days and executed 50 times. As there is no 2004 environmental data to use, the only difference in input data are the locations of the offenders. As with other experiments, the results are consistent and they reach equilibrium. Figure 7.23 compares the burglary rates to expected data. There are clear differences; in particular crime in Halton Moor has been significantly under predicted again. This is to be expected however because the motivations that drive agent behaviour have not been changed. As Section 7.3.2 discussed, it is possible that the
burglary hotspot in Halton Moor is not a result of burgling for financial gain (as suggested by the crime literature) but rather for intimidation or other reasons (Safer Leeds, personal communication). Therefore, the model (in its current configuration) is not able to account for this hotspot. Interesting future work would be to experiment with the behaviour of offender agents to try to replicate the hotspot more closely but this is beyond the scope of this research.

Figure 7.23: Maps comparing validation result hotspots to 2004 expected data. Both maps have the same thematic ranges.

Figure 7.24 graphs the expanding cell errors of all models including this validation scenario. Ideally, the validation experiment should produce a similar fitness to the final (Base 3) calibration experiments. Unfortunately, the fitness of the validation experiment, measured using $R^2$ and SRMSE, is lower than the calibration. Usually this would suggest that the model has been over-fitted to the 2001 field data. However, at the start of this section it was noted that no environmental data for 2004 were available (the UK census only occurs every decade). Therefore it is probable that the performance of the validation model is a result of inadequate input data, rather than a being over-fitted to 2001 data.

In summary, the validation process suggests that the model does not perform as well as it did for the calibration (“base”) experiments. Under normal circumstances this would indicate that the model was over-fitted to the calibration data and needed to be improved to make it useful in other situations. However, in this case there are a number of factors that might lead to poorer than expected validation results. The most likely reason is that there was no updated census data to use as input for the model. Others include missing model parameters (e.g. crime reduction scenarios in 2004 have changed the way in which people respond to burglary, rendering some of the calibration changes unrealistic) or poor assumptions about how offenders respond to environmental cues. Unfortunately it is difficult to verify whether or not these are the case in the absence of accurate data on which to validate the model.

The validation process is not without its uses, however. It has demonstrated the risks associated with applying the model to scenarios for which sufficient data are lacking. Therefore when the next chapter explores the power of the model as a forecasting tool it will do so with care, particularly
7.5 Summary

Evaluating a model is an essential aspect of any modelling endeavour. To this end, Chapter 7 has continued the evaluation process that started in Chapter 6 by calibrating and validating the model through experimentation with real data. Section 7.2 began with an in-depth review of methods that can be used to compare point patterns and it was decided that no method in isolation was sufficient for providing a sound process for evaluating the model. Therefore a detailed analysis framework was devised that would allow for model results to be evaluated both by comparing the spatial locations of points to expected data (the “expanding cell” method) as well as comparing the spatial structure of the data through the use of L functions. In combination, these methods provide a complete and robust means of evaluating the accuracy of the model at different resolutions. This method has significant advantages over other methods, such as avoiding the requirement of aggregating up to an area boundary before applying goodness-of-fit tests.

Section 7.3 then outlined the process of calibration. The advantages of using multiple methods to compare simulated and result data became clear when the model began to perform more accurately, even though this initially led to a drop in the global goodness-of-fit. Finally Section 7.4 attempted to validate the model by executing a new scenario using crime data from an alternative year (2004). The validation results were poorer than the calibration results which might suggest that the model has been over-fitted to the calibration data. However, there is insufficient input data to build a reliable virtual city that represents the situation in 2004 and it is likely that this accounts somewhat for the poor validation performance.

One of the most interesting findings from the calibration process (see Section 7.3.2) was the failure of the model to simulate a hotspot in the Halton Moor area. Both the environmental data
and the crime data failed to indicate why there should be such a large hotspot in the area that was not simulated by the model. Through discussions with experts at Safer Leeds, it was found that burglary in Halton Moor was often motivated by reasons other than wealth generation, such as the intimidation of new residents. As the model has been built upon common assumptions about burglary – where burglars are motivated primarily by the need for money – the failure of the model to simulate the Halton Moor hotspot indicates that the common assumptions about burglary might be incorrect in this case. This has significant implications for crime reduction policies; those in Halton Moor might need to take account of the potential different burglar motivations or risk being poorly targeted.

Now that the model has been evaluated, the following chapter will use it to perform crime forecasting experiments in Leeds. Then Chapter 9 will extend the use of the model across the Atlantic and perform forecasting experiments in Vancouver, Canada.
Chapter 8

EASEL Experiments and Forecasts

Contents

8.1 Introduction ......................................................... 207
8.2 The Optimistic Scenario – GS1 ................................. 208
8.3 The Pessimistic Scenario – GS2 ................................. 212
8.4 The “Insecure Buildings” Scenario – GS3 .................. 213
8.5 Crime Displacement and Burglar Travel Patterns ........ 213
8.6 Summary – EASEL Scenarios ................................. 225

8.1 Introduction

Up to this point, the agent-based burglary model has been extensively evaluated through the use of experiments on idealised data in Chapter 6 and with experiments using real data in Chapter 7. Following the thorough evaluation, the simulation is at a stage that it can be used to fulfil its ultimate goal: to provide crime forecasts and reflections on the burglary system. The experiments in this chapter will simulate some of the effects that an urban regeneration scheme in Leeds might have on burglary rates and Chapter 9 will extend the experiments to an alternative urban area; that of Vancouver, Canada. The urban development being experimented with here, which was outlined in detail in Section 4.2 is known as EASEL. At the time of writing, no work had been fully completed but development was heavily underway at two of the main EASEL sites and, therefore, these sites have been chosen for the scenario. The sites are called “Gipton” and “South Seacroft”. The experiments will attempt to answer the following questions:

- What effect will the EASEL regeneration project have on burglary rates in Leeds?
- With respect to burglary is it more important that houses are secure or that communities are cohesive?
- Will the EASEL project lead to unpredicted crime displacement or diffusion of benefit into the surrounding areas?
Three experiments will be conducted to answer these questions. The first (an “optimistic” scenario) assumes that the plans for the area are a success and that a secure, cohesive community is created. The second simulation (the “pessimistic” scenario) determines what might happen if the results of the regeneration are less successful and the new community is not as mixed and cohesive as decision makers plan for. The final scenario will assume that although the community is cohesive and mixed (as the council plans for) the houses themselves are poor quality and not secure from burglary.

The validation work in Section 7.4 demonstrated that with imprecise input data the model is, understandably, inaccurate. Therefore the results of the following forecasts must be treated with caution because there is no demographic (community) data more recent than that provided by the 2001 UK census. Therefore the experiments will actually be simulating the situation in 2001 if the developments had been completed then. Although this limits the potential of the model to predict actual future crime rates it does not detract from its strong ability to provide reflections on our understanding of the burglary system. As Section 8.5 will show, these reflections are still illuminating.

8.2 The Optimistic Scenario – GS1

This section will outline how the environment can be configured assuming that the EASEL changes have the desired effect on the community. In this sense it is an “optimistic” scenario. The EASEL Area Action Plan (EASEL Team, 2007), includes the following proposed changes to the EASEL areas:

- New houses will be a mixture of sizes / types and more privately owned tenure than exists at the moment (mostly council) to attract a variety of people.
- Greater density of houses to support shops / services.

These changes can be implemented in the model by making the proposed communities more mixed, more cohesive and by using actual building plans to construct new properties to the correct densities. Creating a more mixed community can be reflected in the model by changing the community Output Area Classification (OAC: Vickers and Rees, 2006) type. Recall that an OAC subgroup (the finest grouping available in the hierarchy) consists of a single group classification as well as individual values for all the 41 input variables which make up the classification. To reflect the most “typical” type of community that the EASEL plans are hoping to create, all input variables for the new Gipton and Seacroft communities are assigned normalised values of 0.5. Having set values for the input variables, it is also necessary to classify this new type of community into an individual subgroup type but re-building the classification to find a grouping for this new type of community is beyond the scope of the project. Instead, to find the most “typical” subgroup the variance of the mean standardised values of the 41 variables for all subgroups was calculated and the lowest, 0.002409, belonged to subgroup 6b2. This is part of the “Typical Traits” supergroup which is a group characterised by its “averageness” (Vickers et al., 2005) and therefore
appears ideal for these purposes. As such the new communities will be assigned to subgroup 6b2. Figure 8.1 maps these OAC changes. For a list of all OAC types, refer to Appendix A.

With respect to the new buildings being built, the area action plan (EASEL Team, 2007) stipulates that the density must be high enough to support shops and services but is otherwise unspecific. Fortunately, the actual plans used by the developers are also available and can be used directly to create the new buildings and roads. Figure 8.2 illustrates the layout of the new buildings and roads created manually from the paper plans. The plans have been recreated as closely as is practical. The correct number of buildings have been simulated and, although it is not identical, the layout of the new buildings and roads is similar. Values for the new household variables are outlined in Table 8.1.

The entire EASEL project is aimed at creating highly cohesive communities. This can be implemented in the model by increasing the level of collective efficacy. Figure 8.3 illustrates that levels of collective efficacy were relatively low before the changes, but in the forthcoming scenario they are set to the highest level (1.0).

Finally, it is assumed that because the new communities will be more cohesive, they will house no burglars or drug dealers. Although this is a broad assumption, estimating the potential number of burglars who might live in the area after regeneration is highly problematic and will involve deep sociological, geographical and criminal research that is well beyond the scope of this project.
CHAPTER 8. EASEL EXPERIMENTS AND FORECASTS

Results of the Optimistic Scenario

On the whole, the results of the Gipton / Seacroft compared to previous experiments (Base 3) are very similar. This is to be expected though as the Gipton / Seacroft areas are very small so, on the whole, the environments are very similar. There are some small differences, however. Figure 8.4 illustrates burglary hotspots in the Gipton area produced by the Base 3 scenario and the optimistic Gipton / Seacroft scenario. Because the focus is on a very small area, the individual houses that are subject to burglary will vary slightly across model runs so four different randomly chosen results are illustrated. On the whole, however, the patterns are similar. The kernel density used in the KDE algorithm is only 50m (compared to 350m in other maps) in order to illustrate the density more finely. Although not shown, results for the Seacroft area are similar to those in Gipton. In the Base 3 scenario there were some burglaries committed in the Gipton area whereas in the Gipton
### Table 8.1: Changes to household parameters for the Gipton/Seacroft scenario.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>The accessibility (how easy the house is to break into). This is based on the number of possible entrances and as most buildings are semi-detached (three possible entrances) this value is normalised to 0.25 (refer to Section 5.4.2 for more information).</td>
<td>0.25</td>
</tr>
<tr>
<td>VIS</td>
<td>The visibility of the house to neighbours. This is derived from the degree of isolation and the size of the garden. As the paper plans have not been copied exactly and there is not information in the AAP on which to set these values, particularly for garden size, this will be set to the default value.</td>
<td>0.5</td>
</tr>
<tr>
<td>SEC</td>
<td>The security. Assume high “secured by design” standards.</td>
<td>1</td>
</tr>
<tr>
<td>TV</td>
<td>Traffic Volume. It is possible to recalculate the this value using space syntax analysis but because the road layout in the simulation is not an exact copy of the paper plans this might not be accurate. Also, the AAP includes no information that might guide the estimate of traffic volume. Instead assume that traffic volume will have the default normalised value.</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Figure 8.3: Collective efficacy levels before and after Gipton/Seacroft changes.
/ Seacroft scenario there were none. Judging from this scenario, therefore, it can be concluded that the EASEL regeneration scheme would not affect burglary patterns greatly but, if anything, it would be expected to reduce the number of burglaries in the redeveloped areas. This is not as intuitive as it sounds, for it might be expected that richer households would draw burglars and subsequently suffer higher victimisation rates as illustrated by alternative Leeds research (Kong-muang, 2006). However, it does appear that the houses surrounding the Gipton/Seacroft area might experience a slight increase in burglary (a hypothesis that will be explored in greater detail in Section 8.5). This can be explained by considering the security of the properties: non-redeveloped areas have security levels of 0.5 whereas the new houses have higher values of 1.0. Therefore the attractiveness of the new communities does appear to attract burglars, but the security of the new houses forces them to look in the surrounding areas for targets. This is something that decision makers might need to take into account.

Figure 8.4: Comparing randomly chosen results of the optimistic Gipton / Seacroft scenario area to Base 3 results (in the Gipton area).

8.3 The Pessimistic Scenario – GS2

It is possible that against the efforts of planners and decision makers, the EASEL changes do not lead to the pleasant, cohesive communities they are aimed at creating. To test this hypothesis a scenario can be initiated by simply resetting the demographic OAC parameters to the values that they held before the hypothetical changes (see Figure 8.1). All other household variables will retain their GS1 scenario values (from Table 8.1) because the physical changes will be the same
regardless of the people who eventually decide to live in the area. Figure 8.5(a) compares the optimistic and pessimistic scenarios across the entire Gipton / Seacroft area and Figure 8.5(b) examines Gipton specifically. Although the differences are not large, in the pessimistic scenario there are more crimes committed in the affected areas (only Gipton is shown here but the effects in Seacroft are similar). Interestingly the household attributes (e.g. security) in both cases are similar so this effect is entirely attributable to the changing community parameters. Nevertheless the increase in burglary rates is not dramatic so at this point the model is suggesting that making a cohesive community is not absolutely essential to reduce burglary.

8.4 The “Insecure Buildings” Scenario – GS3

Finally, to consolidate the previous two experiments and better understand the risks of regenerating an area poorly this scenario will hypothesise that although the planners are successful in attracting a range of people to the regenerated area the houses themselves are poorly designed and as a result insecure. The communities will have the same demographics as the first “optimistic” scenario (GS1) but the houses will only have a security value of 0.1 (poor security) as apposed to 1.0 (high security). This value is lower than all other properties in the simulation (0.5) but above 0.0 because zero security implies that the residents have left the door open which is not accounted for in this model.

Figures 8.6, 8.7 and 8.8 illustrate the results, mapping the burglary density in the regeneration areas at different levels of detail. It is apparent in all figures that burglary rates are very high in both the Gipton and the Seacroft areas. This result suggests that even if a regenerated area is highly cohesive it can still suffer a large amount of burglary if security standards are poor. However, the reasons for the large difference are non-trivial. The following sections will explore these results in greater detail by looking for evidence of crime displacement and showing how the burglars’ travel patterns are influenced by the environmental changes.

8.5 Crime Displacement and Burglar Travel Patterns

This section will explore the results in more detail by assessing the extent to which crimes might be displaced into the surrounding areas from the regeneration areas. This will answer the third question posed in the chapter introduction: will the EASEL project lead to unpredicted crime displacement of diffusion of benefit into the surrounding areas? For the following analyses, all the results for each particular scenario (50 individual model runs) will be combined into single data sets, rather than comparing a subset of the results as was done previously. This is appropriate because rather than observing density maps visually, quantitative comparisons will be made. These are more accurate once all of the results for a particular scenario have been merged because they will not be influenced by random model fluctuations or problems with small numbers.
(a) Comparing randomly chosen results from the Gipton / Seacroft “optimistic” and “pessimistic” scenarios.

(b) Comparing results and focussing on Gipton. (Results for the Seacroft area are similar).

Figure 8.5: Comparing results from the “optimistic” and “pessimistic” scenarios.
Figure 8.6: Comparing the “insecure buildings” scenario to the previous Gipton / Seacroft scenarios.
Figure 8.7: Comparing the “bad buildings” scenario to the previous Gipton / Seacroft scenarios, focussing on the Gipton area.
Figure 8.8: Comparing the “bad buildings” scenario to the previous Gipton / Seacroft scenarios, focusing on the Seacroft area.
8.5.1 Crime displacement after the “optimistic” scenario

From the results of Section 8.2, the model predicts that if the council is successful in their plans for the Gipton/Seacroft area then it will have low burglary rates (the GS1 “optimistic” scenario). However, it is not clear whether or not the optimistic scenario will lead to the displacement of burglaries into the surrounding area, or even a diffusion of benefit (i.e. lower burglaries in the surrounding areas). One method of quantifying the presence of crime displacement that has been used in practice (Bowers et al., 2003; Hirschfield, 2004) is the weighted displacement quotient (WDQ: Bowers and Johnson, 2003). This technique will be used to compare GS1 results to the Base3 results (i.e. results before and after the EASEL implementation). The WDQ is the ratio of two measures: a “success” measure that quantifies the change in crime rates in the target area compared to a control area and a “buffer displacement” measure which looks for evidence of displacement into the surrounding area. Thus the WDQ provides an estimate for whether or not the change in crime in the target area was similar to that in the buffer zones, relative to a control area. Formally:

\[ WDQ = \frac{\text{buffer displacement measure}}{\text{success measure}} = \frac{B_{t_1}/C_{t_1} - B_{t_0}/C_{t_0}}{X_{t_1}/C_{t_1} - X_{t_0}/C_{t_0}} \]  

(8.1)

where \( X \) is the crime rate in the target area, \( B \) is the crime rate in the buffer zone, \( C \) is the crime rate in the control area, \( t_0 \) represents the crime rates before the intervention and \( t_1 \) represents crime rates after the intervention. Values of WDQ > 0 are indicative of diffusion of benefit (a reduction of crime in the areas surrounding a target area) and values of WDQ < 0 suggest that there has been displacement of crime.

One of the benefits of using simulation models, such as the one developed here, is that the effects of an intervention (the EASEL changes in this case) are isolated completely from any other effects because in all other respects the scenarios are identical. Therefore there is no need for the control area, \( C \) and Equation 8.1 can be rewritten as:

\[ WDQ = \frac{B_{t_1} - B_{t_0}}{X_{t_1} - X_{t_0}} \]  

(8.2)

Furthermore, instead of calculating a single WDQ value, it is possible to use a number of concentric expanding buffer zone areas to measure displacement at different distances away from the target area. Bowers and Johnson (2003) use concentric ring zones increasing in 400m increments around the target area up to a distance of 2km (the authors note that 2km is the maximum distance that a burglar is likely to travel). However, in this case the target area is much smaller and we are looking for displacement into the immediate area (the following section will consider the possibility of wider displacement) therefore smaller buffer zones will be used: 50m rings up to a maximum of 250m. These buffer regions are illustrated by Figure 8.9(a).

The WDQ value for each of the concentric rings was calculated and Figure 8.9(b) graphs the change in the WDQ value in both the Gipton and Seacroft areas for each buffer zone. In Gipton, there appears to be considerable displacement into the most immediate buffer zone, some displacement into the second and a slight diffusion of benefit into the farther buffer regions. In Seacroft, most buffer zones show slight crime displacement into the surrounding area. Therefore,
8.5. CRIME DISPLACEMENT AND BURGLAR TRAVEL PATTERNS

(a) Buffers used in the calculation of the weighted displacement quotient (WDQ).

(b) Weighted displacement quotient graphs. A line at $WDQ = 0$ is drawn to illustrate the difference between crime displacement ($WDQ < 0$) and diffusion of benefit ($WDQ > 0$). In Gipton there appears to be crime displacement in the area immediately surrounding the target area, with slight diffusion of benefit at greater distances. In Seacroft, there appears to be limited displacement in all buffers regions.

Figure 8.9: Displacement into buffer regions surrounding Gipton and Seacroft
it is possible to conclude that if the Gipton and Seacroft areas are regenerated successfully it is likely that crime rates in the immediate surrounding areas will increase as a direct result of the regeneration.

Use of the WDQ is interesting here because it is a technique that is commonly used in practice to look for crime displacement or diffusion of benefit. An advantage of the technique is that it attempts to distinguish the effects of the crime reduction initiative from other effects that might have influenced the crime rate. However, one of the benefits with this type of modelling approach is that the model acts like a virtual laboratory. With the exception of the Gipton and Seacroft areas, the environment remains completely unchanged so that all crime variation can be attributed solely to the changes implemented by the EASEL scenario. Therefore it is also possible to use the expanding cell algorithm in order to search for displacement or diffusion. This would not be possible using real-world data as the fluctuations caused by EASEL could not be isolated from those caused by other factors. Using the expanding cell approach has the advantage over the WDQ that it will provide evidence for where displacement is occurring, not just at what distance. To explore crime displacement in Base3 and GS1 further, therefore, Figure 8.10 maps the relative percentage difference (see Section 7.2.4) between the Base3 and GS1 scenario results using the expanding cell algorithm.

From Figure 8.10 it is apparent that, on the whole, the results of the Base3 and GS1 scenarios
are very similar. This is to be expected as the experiments are identical with the exception of the relatively small Gipton and Seacroft areas. The green areas are indicative of fewer crimes being committed in Gipton/Seacroft after the EASEL changes which is to be expected because, as Section 8.2 demonstrated, the regeneration scheme was successful and burglary rates within the regenerated areas were low. However, the orange and red areas surrounding the regeneration zones indicate that there are some houses which show a substantially higher risk of burglary than others. The effect is highly localised which is unusual because it might be expected that burglaries would be more evenly distributed in the surrounding area (for example see Malleson et al., 2009).

The largest increase in burglaries can be found in a small number of houses to the north of the Seacroft development site so burglary in these houses will be explored in more detail. A preliminary procedure to assess why certain individual houses have considerably higher burglary rates is to observe the travel patterns of some of the agents in order to determine whether or not the urban configuration of the Seacroft area is the likely candidate (e.g. the layout of the road network might lead to a large number of agents passing the highly burgled houses but not the others). To this end, Figure 8.11 plots the movements of four agents chosen at random; two who did not commit crimes in the highly burgled area and two that did. For the agents that did commit crimes in the highly burgled area, large parts of Seacroft were relatively unexplored. However, the highly-burgled houses are situated on a main road that runs along the northern boundary of the Seacroft area which was regularly used by agents. Furthermore, a close inspection of Figure 8.11 indicates that the agents passed the houses whilst looking for a burglary target, not during legitimate travels on some other business (such as travelling to a social location). Figure 8.12 illustrates this in more detail.

Therefore it appears that the EASEL changes attracted the agents to the area specifically for burglary purposes and the location of the houses on the main road meant that they were certain to be noticed by the agents whereas other houses were not. Subsequently it is apparent that the houses which had been highly victimised attain part of their risk from their location in space. These findings are strongly supported by the crime theories outlined in Section 2.4. In the cases where the houses do not feature in a burglar’s awareness space – because they have not been passed on burglars’ routine activities – they have a relatively low burglary risk. Once a burglar becomes aware of the houses near the regeneration area, however, their risk increases. Nevertheless, the theories in isolation could not have predicted which individual houses in the regeneration area might be susceptible to burglary. Only when the theories have been implemented in a model that is able to account for the low-level dynamics of the burglary system can specific real-world predictions such as this be made.

Although spatial location is clearly a factor in determining why a small number of houses have been disproportionately burgled, in isolation it does not fully account for the high burglary rates. Clearly there are many houses along the main road in Seacroft that do not suffer high levels of victimisation after the EASEL changes. Therefore some of the risk might be attributed to the physical attributes of the houses (i.e. their visibility or accessibility). Figure 8.13 illustrates the
Figure 8.11: Some offender movement patterns in the GS1 scenario, illustrating the difference between those who did and did not burgle in on the outskirts of the Seacroft development.
8.5. CRIME DISPLACEMENT AND BURGLAR TRAVEL PATTERNS

Searching for a burglary target
Committed burglary
Travelling to a drug dealer

A Burglary in the Seacroft area

Figure 8.12: Visualising the journey to and from a burglary close to the outskirts of the Seacroft regeneration area. The agent travels to the area specifically for burglary. For clarity, both images illustrate the same journey but from different angles.

GeoTime software used courtesy of Oculus Info Inc. All GeoTime rights reserved.
estimated burglary risk\textsuperscript{1} for houses in the area and compares this to the total number of burglaries committed in each house. The two houses indicated have a slightly higher burglary risk than their neighbours which also explains why they have higher burglary rates. Again this finding is consistent with the crime literature: the rational choice perspective stipulates that an offender makes a (bounded) rational decision based on a cost-benefit analysis of potential burglary targets. When examining houses in the Seacroft area, some houses are more attractive than others and, on average, it is these houses that will receive the largest number of burglaries.

Figure 8.13: The burglary risk for heavily-burgled houses in Seacroft. The histogram shows the risk of the indicated houses compared to all houses in the environment.

In conclusion, it is apparent that the effects of having a slightly higher burglary risk, coupled with their location on a main road, mean that on average the indicated houses have received a large number of burglaries after the EASEL changes. But only after examining the routine activities of the burglar agents as well as the household attributes (at the individual level) does this become apparent. This result demonstrates the power of this modelling approach because without considering the behaviour of the burglars and the attributes of individual houses it would not be possible to pinpoint which individual houses might suffer a high burglary risk as an direct but unintended consequence of urban regeneration. This also leads to a specific policy implication: the houses identified surrounding the Seacroft area (as well as some in the Gipton area) should be target hardened.

8.5.2 Displacement in the Low-Security Scenario

The final analysis of the simulation results will compare the change in burglary rates of two EASEL scenarios: the optimistic scenario (GS1) and the low-security buildings scenario (GS3). In the latter case, Section 8.4 demonstrated that if the buildings in the EASEL areas were not built securely then they would suffer large increases in rates of residential burglary. This is illustrated quantitatively by Figure 8.14 which maps the relative percentage difference in the number

\textsuperscript{1}The burglary risk is calculated as the mean of the variables that feature in an agents burglary decision: accessibility, security, traffic volume, occupancy and collective efficacy.
of burglaries between the GS3 and GS1 scenarios. As expected, the regeneration areas have high burglary rates after the GS3 regeneration when compared to the GS1 regeneration. What is interesting, however, is that, as with the previous analysis, a small number of houses experience a disproportionately large decrease in burglaries in GS3 when compared to GS1. Again the change in burglary rates is not shared across the area as a whole as might be expected. In particular, notice that the individual houses that were highly victimised in the GS1 scenario (see the previous section) receive many fewer burglaries in the GS3 scenario as a consequence of the opportunities offered by the newly regenerated buildings. Although this does not provide any further policy implications, it is an interesting example of where crimes might be displaced from if the regeneration fails and becomes a target for burglars.

Figure 8.14: Comparing GS1 and GS3 results using the expanding cell algorithm

8.6 Summary – EASEL Scenarios

The first half of this chapter outlined experiments in the EASEL area that altered the effects of community cohesion and household security on burglary rates. The first two simulations suggested that creating mixed and cohesive communities, as EASEL intends to, is not particularly important as it only has a small effect on the subsequent burglary rates. However, the final scenario demonstrated that if household security is low then the area is likely to suffer a dramatic burglary problem. Therefore it appears that, for burglars, the differences between individual houses at a local level are much more important than the wider community-level factors. This finding leads to
two important recommendations:

- The field of criminology might benefit from researching this avenue further, with a particular consideration for the impacts this might have on theories that emphasise the importance of an offender’s daily habits or their activity spaces. For example, although an offender will never burgle a house that they are unaware of (hence the importance of routine-activities or crime pattern theories) the individual house security is also extremely important (as stipulated by rational choice theory). Therefore although this research does not argue against these theories it provides further evidence for the importance of individual-level considerations over larger community-wide factors.

- EASEL policy makers should be aware of the importance of the features of individual properties. These appear to be considerably more important, with respect to burglary, than the communities.

The second half of the chapter examined the results in more detail, noting where crime displacement could be a problem. The analysis was able to pinpoint a small number of buildings close to the areas being regenerated under EASEL that would be likely to suffer large numbers of burglaries. The increase in burglary was a direct consequence of the new behaviours of the burglar agents in response to the urban regeneration. This lead to another policy implication: the houses directly adjacent to the regeneration sites (particularly a small number of houses in the north-east of the Seacroft site) should be target hardened to reduce their risk.

Of course, some caution must always be taken when providing recommendations from modelling results. A pessimist might criticise the results by declaring that the model has so many variables that it could be used to make any predictions simply by varying parameter values. For example, it might be possible that the previous EASEL experiments could be reproduced by changing a different combination of variables. If this were the case the results would not tell us anything about the relationship between the community attributes and the security of individual properties, just about how the model behaves under in a particular parameter space.

However, we can argue against this hypothesis as follows. Calibration shows us that the real data constrains the values of house security and the community variables. If these values are changed then the model no longer matches expected data. So, importantly, there is only a small range of values that lead to predictions that are similar to real data. Therefore the original community and security values are reliable and can be varied from their real-world values to evaluate scenarios without the risk that varying any other parameters could lead to the same result. Ideally this argument would be further strengthened by systematically testing the model’s entire parameter space. For example a genetic algorithm could be employed following Malleson et al. (2009). This would then provide concrete evidence that the model could only produce the EASEL results under very particular parameter configurations. Unfortunately, due largely to the computational requirements, this approach is infeasible at this time and must be instead recommended for future work.
Chapter 9

Experiments in Vancouver

Contents

9.1 Introduction ............................................. 227
9.2 Context: Vancouver ..................................... 228
9.3 Preparing the Environmental Data ................. 230
9.4 Preparing the Offender Agents ....................... 239
9.5 Real (Expected) Burglary Data ...................... 242
9.6 Vancouver SkyTrain Scenarios ..................... 243
9.7 Summary ................................................. 246

9.1 Introduction

As a further demonstration of the flexibility and the forecasting potential of the model, this chapter will discuss the process of applying the simulation to an environment a long way removed from the one that it was created for in Leeds. Vancouver, British Columbia, Canada, was chosen as an appropriate case study area because it follows the traditional North American ‘grid’ road system whereby the city is broken up into blocks and is therefore very different to UK cities like Leeds. Furthermore, part of the research project involved an overseas visit to the Institute for Canadian Urban Research Studies (ICURS) at Simon Fraser University in Vancouver. ICURS work closely with the Royal Canadian Mounted Police and have access to a range of crime data which are stored in a highly detailed crime database. Therefore the visit allowed for the collection of essential crime and environmental data that are required as inputs into the model.

As a consequence of the 2010 Winter Olympics which were held in Vancouver, at the time of writing the city was in the process of building a new high-speed public rail line. These “SkyTrain” lines are seen as high crime attractors by the public and are therefore very contentious. Thus the Vancouver scenario will be configured to answer the following question: what is the effect of the SkyTrain on residential burglary patterns in Vancouver?
9.2 Context: Vancouver

The Greater Vancouver Regional District (also known as Metro Vancouver or Greater Vancouver) is a region in the province of British Columbia on the west coast of Canada. It has a population of approximately 2.1 million people at the last census (2006) making it the third largest city in Canada (after Toronto and Montreal). Figure 9.1 illustrates that Greater Vancouver actually consists of a number of independently-governed areas (including the city of Vancouver itself). It is also well known for its scenic beauty, being situated close to northerly mountains that are easily visible from the city (Figure 9.2). Other than being beautiful, the landscape actually has an influence on crime in the city. For example, Simon Fraser University is on a mountain surrounded by a wilderness park so is outside the awareness spaces of many potential offenders and subsequently suffers little crime (Brantingham and Brantingham, 1995).

![Figure 9.1: The Greater Vancouver Regional District.](image)

Vancouver has a well-developed transport infrastructure that consists of standard buses, express buses (limited stop, called “B-Line”) and an over-ground rail network called the SkyTrain. These are illustrated by Figure 9.3. There are two separate SkyTrain lines: the Expo line (constructed in the 1980s) and the Millennium line (opened in 2002).

The effect that public transport systems have on urban crime rates is non-trivial. Although a body of literature has evolved that investigates crime on the system, such as at stations or on
trains/buses (Smith and Clarke, 2000; Loukaitou-Sideris et al., 2002), there is limited research that looks at how the system might affect crime patterns off the system. With respect to crime in the surrounding areas, transport hubs are generally seen as crime generators. For example, planners must often confront the fears of suburban communities that a new public transport link will bring criminals from other parts of the city into their neighbourhoods (Smith and Clarke, 2000). However, there is limited empirical evidence to support these fears (Smith and Clarke, 2000). The SkyTrain is no exception to this, the stations are often seen as dangerous crime hotspots (Sinoski, 2008) although this might be unwarranted (South Coast British Columbia Transportation Authority, 2008). However, there is some empirical evidence for the effect of the SkyTrain on crime. Robinson (1998) found that SkyTrain stations attracted a larger than expected number of calls for service (Smith and Clarke, 2000) and Brantingham et al. (1991) suggest that the centre of a crime hotspot might have shifted from a public housing estate to a new SkyTrain station once it was opened.

With respect to offending behaviour, the situation is similarly contradictory. Theory suggests that public transport will alter peoples’ awareness spaces (Brantingham and Brantingham, 1993) although Smith and Clarke (2000) note that there is little direct evidence to suggest that public transport is used in crime journeys. However, in writing about the SkyTrain, Brantingham et al. (1991) note that the ways in which people move about the city will directly affect their awareness spaces. As public transport often limits the generation of awareness spaces (particularly if routes are elevated or go below ground as is the case in Vancouver) it is expected that offenders who use public transport will commit crimes in tighter areas than car drivers (Brantingham et al., 1991). Therefore, although it is unlikely that burglars will use public transport to actually commit burglary – due to the physical difficulties associated with carrying stolen goods (Smith and Clarke, 2000; Brantingham et al., 1991) – it is not unreasonable to expect that they might nevertheless return with alternative transportation to the area surrounding a station because it is well known to them.
This remainder of this chapter will explore the effect that the SkyTrain will have on rates of residential burglary in the city in an attempt to shed light on an otherwise confusing situation.

9.3 Preparing the Environmental Data

The following will sections outline how the available data can be configured to be used as input to the model. Although the SkyTrain and bus lines run into the neighbouring cities of Burnaby and Richmond, crime data is only available for the city of Vancouver so the simulations will be restricted to this area.

9.3.1 Roads and Transport Routes

In the model, certain roads can only be used by agents who have access to particular types of vehicle. As discussed in Section 5.3, these accessibility types are:

- **walk** – roads that are accessible only to walkers (such as alleyways);
- **car** – roads that are accessibly only to cars (such as motorways);
- **majorRoad** – major roads that can be driven along at higher speed than other roads (i.e. major arterial routes).
Combinations of accessibility are possible, for example a large arterial route with a pavement will be useable by walkers and drivers alike and will also offer a speed increase to drivers. In Leeds, this information is available via the MasterMap Integrated Transport Layer. For Vancouver, road network information is available through a product called GIS Innovations (GIS Innovations, 2009) that contains a large amount of information about all roads in the province of British Columbia. Each geometric object contains an attribute which details the classification type of the road and allows us to distinguish its accessibility. Table 9.1 outlines all the available road types and indicates how they can be mapped to accessibility in the model. Figure 9.4(a) illustrates the roads spatially.

Table 9.1: Different classes of road available in the GIS Innovations product.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Access value</th>
</tr>
</thead>
<tbody>
<tr>
<td>freeway</td>
<td>controlled access, typically divided carriageway</td>
<td>majorRoad car</td>
</tr>
<tr>
<td>highway major</td>
<td>a primary provincial highway</td>
<td>majorRoad car</td>
</tr>
<tr>
<td>highway minor</td>
<td>a secondary provincial highway</td>
<td>majorRoad car</td>
</tr>
<tr>
<td>arterial major</td>
<td>a major thoroughfare with a generally large traffic capacity, may have more than 1 lane each way</td>
<td>majorRoad car walk</td>
</tr>
<tr>
<td>arterial minor</td>
<td>a thoroughfare with medium traffic capacity, has one lane each way</td>
<td>majorRoad car walk</td>
</tr>
<tr>
<td>collector major</td>
<td>a road to feed traffic within town with right of way, may have more than 1 lane each way</td>
<td>car walk</td>
</tr>
<tr>
<td>collector minor</td>
<td>a road to feed areas of local traffic, has one lane each way</td>
<td>car walk</td>
</tr>
<tr>
<td>local</td>
<td>local and residential roads</td>
<td>car walk</td>
</tr>
<tr>
<td>strata</td>
<td>residential roads with potential public restriction, trailer parks, first nations and strata developments</td>
<td>car walk</td>
</tr>
<tr>
<td>lane</td>
<td>alleyways for access to the rear of properties</td>
<td>car walk</td>
</tr>
<tr>
<td>ramp</td>
<td>ramps for highway access or turning lanes</td>
<td>majorRoad car</td>
</tr>
<tr>
<td>restricted</td>
<td>a restricted road, generally not accessible to the general public</td>
<td>-</td>
</tr>
<tr>
<td>ferry</td>
<td>a crossing made by public or private ferry boat</td>
<td>-</td>
</tr>
<tr>
<td>recreation</td>
<td>a road to access back country or recreation sites</td>
<td>car walk</td>
</tr>
<tr>
<td>resource</td>
<td>a road for resource extraction</td>
<td>car</td>
</tr>
<tr>
<td>trail</td>
<td>a pathway for pedestrians or bikes, not for vehicles</td>
<td>walk</td>
</tr>
<tr>
<td>service</td>
<td>roads with no formal name that access facilities or places</td>
<td>car walk</td>
</tr>
</tbody>
</table>

It is also necessary to create estimates for the amount of traffic on each road in Vancouver. In Leeds these estimates were generated using space syntax analysis where the level of connectivity (or “integration”) of a road is used as a proxy for the amount of traffic. The algorithm was applied using the same parameter values and the results are illustrated in Figure 9.4.
CHAPTER 9. EXPERIMENTS IN VANCOUVER

(a) Road types used in the Vancouver simulations

(b) Integration levels as a result of the Vancouver space syntax analysis

Figure 9.4: Roads types and estimated traffic volume in Vancouver.
9.3. PREPARING THE ENVIRONMENTAL DATA

Public transport routes are created separately from normal roads using information about the locations of the individual stations for each line. Unfortunately this information was not available so had to be created manually from paper maps. Although Vancouver contains numerous bus routes, only two of these are express routes and it is these that are included in the simulation. It is not preferable to include all routes because this will over-complicate the model and make it more difficult to establish the effect that the SkyTrain line will have on burglary patterns (which is the ultimate aim of the simulation). The transport routes included in the model were illustrated in Figure 9.3.

9.3.2 Buildings

For Leeds, high quality MasterMap data is available that contains the geometry of different physical objects as well as information about their land use type. For the Vancouver scenario, such highly detailed data is not available. Instead, point data is available from the British Columbia Assessment Authority (BCAA: BC Assessment, 2009). The BCAA conduct yearly assessments of each property in British Columbia and although actual property geometries are not available the land-use attribute information is detailed. As with the roads data, the buildings data can be analysed in order to establish the following attributes that are required by the model:

- **Type** – the type of building. Can be either *house*, *workplace*, *social place* or *drug dealer*.
- **Accessibility** – how easy it is to enter the property (e.g. the number of possible entrances).
- **Visibility** – how visible the property is to neighbours / passers-by.
- **Security** – how secure the property is.

**Building Type**

The BCAA land-use data contains information about the actual use for the each property and these fields can be used to estimate what the type of land is in the model. The list of all possible land-uses is extensive and not included here, instead Table 9.2 summarises relevant groups of similar land use:

<table>
<thead>
<tr>
<th>Codes</th>
<th>Land Use Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–99</td>
<td>Residential</td>
</tr>
<tr>
<td>100–199</td>
<td>Farm</td>
</tr>
<tr>
<td>200–299</td>
<td>Commercial</td>
</tr>
<tr>
<td>400–499</td>
<td>Industrial</td>
</tr>
<tr>
<td>500–599</td>
<td>Transportation, Communication and Utility</td>
</tr>
<tr>
<td>600–699</td>
<td>Civic, Institutional and Recreational</td>
</tr>
</tbody>
</table>

Therefore, houses can be derived from residential land uses and workplaces can be derived
CHAPTER 9. EXPERIMENTS IN VANCOUVER

from commercial, industrial and utility land uses. Deriving social buildings, however, is slightly more complicated. By working at a finer level of clarification, Table 9.3 lists the land uses that seem appropriate to represent places that people might go to socialise.

Table 9.3: BCAA land use codes to represent social locations.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>Theatre Buildings</td>
</tr>
<tr>
<td>254</td>
<td>Neighbourhood Pub</td>
</tr>
<tr>
<td>256 – 258</td>
<td>Restaurants</td>
</tr>
<tr>
<td>266</td>
<td>Bowling Alley</td>
</tr>
<tr>
<td>270</td>
<td>Hall (Community, Lodge, Club, Etc.)</td>
</tr>
<tr>
<td>610</td>
<td>Parks and Playing fields</td>
</tr>
</tbody>
</table>

Establishing drug dealer locations is also problematic and should require some analysis of crime (drugs) data. Unfortunately, no drugs data for the scenario area was available so it was not possible to determine where drug dealers should be expected to live. Instead, therefore, it is hypothesised that all drug dealers live in an extremely deprived area called “Downtown East Side”. The area is well known for chronic and high levels of drug abuse and is the logical choice for drug dealer locations in the absence of more appropriate crime data. Figure 9.5 illustrates where the locations of the different types of buildings are.

**Building Accessibility**

Accessibility is a measure of how easy it is for someone to gain access to the property. In Leeds, using Mastermap data it is possible to estimate how many entrances a house will have by calculating how many sides of the house are open to the outside and assuming that each exposed wall will have at least one door or window in it. Unfortunately the geometries of individual buildings in Vancouver are not available so accessibility cannot be calculated. This is not a major problem, however, because the majority of houses in Vancouver are detached so will have more similar accessibility values than houses in the UK (which are a mixture of terraced and (semi) detached). For these case studies, therefore, accessibility is effectively disregarded and assigned a default value of 0.5.

**Building Visibility**

Visibility is used to indicate how visible a property is to its neighbours. Using Leeds data, visibility was calculated from both the size of a property’s garden (under the assumption that large gardens restrict the view of the house) and the degree of isolation (the number of surrounding properties within a given buffer). The data available for Vancouver does not contain information regarding the size of gardens so the degree of isolation is the only measure used. In Leeds, a 50m buffer was deemed appropriate because it was able to distinguish between tightly-packed terraced houses and more spread-out detached or semi-detached houses. Buildings in Vancouver,
however, are structured very differently. Traditional “back-to-back” terraces which are common in the UK are nonexistent, whereas large detached houses set back from the roads (which are rare in UK cities) are much more common. Figure 9.6 illustrates the equivalent of densely and loosely packed properties in Vancouver. Both areas contain largely detached houses, but the affluent areas outside the core of the city (Figure 9.6(a)) are much more spread out than the inner-city buildings (Figure 9.6(b)). An ideal buffer region size, therefore, is one that captures the difference between these types of areas. Through trial and error, 75m was found to be the most appropriate buffer radius to use. Figure 9.6(c) illustrates the final calculated building densities, aggregated to the enumeration district area for clarity.

**Building Security**

Estimating security is extremely difficult as there is no direct measure of security that could be used as input for the model. Thus for the following scenarios security is discounted and all properties are assigned the default value of 0.5.
(a) Loosely packed houses.

(b) Densely packed houses

(c) A measure of isolation: the number of buildings within a 75m buffer zone. For visual clarity the buildings have been aggregated; mean building counts are shown.

Figure 9.6: Vancouver housing density. Note that the images of densely packed and loosely packed houses have the same scale and were produced using Google Earth.
9.3. PREPARING THE ENVIRONMENTAL DATA

9.3.3 Communities / Census Parameters

Following the roads and buildings, values for community parameters must also be established. The required variables are:

- Collective efficacy – a measure of the amount of community cohesion
- Sociotype – the “type” of the community, including:
  - Attractiveness – the attractiveness of the sociotype as an area for burglars to look for burglary targets.
  - Occupancy – predicted levels of occupancy at different times of day.
  - Similarity – how to compare sociotypes and return a numerical similarity value.

In a similar manner to Leeds data, the 2001 Canadian census is available for use. As only partial results for the 2006 census were available at the time, they were disregarded. The smallest geographical boundary that census data is available at is the enumeration district (ED) which are similar in size to the UK output area (OA) with each one containing approximately 200 houses.

**Community Collective Efficacy**

In a similar vein to the Leeds calculation, collective efficacy will be calculated using three layers: concentrated disadvantage; residential stability; and ethnic heterogeneity. Concentrated disadvantage can be estimated using a similar index to that used in Leeds (the Index of Multiple Deprivation) called the Vancouver Area Neighborhood Deprivation Index (VANDIX: Bell et al., 2007; Schuurman et al., 2007). VANDIX has been calculated specifically for the Vancouver area and is therefore better able to account for the specific demographic features of the area such as the large student population around the University of British Columbia on the west coast. The index was calculated manually following instructions by Schuurman et al. (2007) and Figure 9.7 illustrates the normalised values of the index in Vancouver. The index does seem to capture the deprivation levels accurately, correctly distinguishing between the more affluent areas in the west of the city, the less affluent eastern city districts, the highly deprived “Downtown East Side” and an Musqueam Indian reservation that is clearly socioeconomically distinct from its surroundings.

As with Leeds analysis, residential stability can be estimated simply from the number of people who own their own homes which is available from the Canadian census.

Finally, ethnic heterogeneity can again be estimated using the Index of Heterogeneity (Blau, 1977). In the Canadian census, every ethnic country of origin is recorded separately but it is necessary to classify people into one of five distinct groups to remain consistent with the Leeds scenarios. Therefore, to simplify the process of classifying every unique country of origin the “Visible Minority Groups” table will be used and Table 9.4 outlines the mapping from Canadian ethnic groups to those used in the UK.

The overall collective efficacy value can be defined as the mean of the three layers, which will be in the range 0 – 1 because the three input layers have already been normalised to that range. Note that the index of heterogeneity and VANDIX must be “reversed” (using $x = 1 - x$) so
it has the same orientation as other layers where a low value to applies to a low collective efficacy. Figure 9.8 illustrates the overall estimated levels of collective efficacy in Vancouver.

Community Sociotypes

In Leeds the Output Area Classification was used to estimate the “type” of each community. Obviously this data is not available for Vancouver and it is beyond the scope of this project to build a similar classification. Instead, the individual aspects which are used to build up a sociotype (attractiveness, occupancy and a method of comparing sociotypes) will be constructed from census data manually as follows:

- **Attractiveness**: it is hypothesised that the median income of residents is sufficient to estimate a community’s attractiveness to burglars. Criminology research (e.g. Wright and Decker, 1996) suggests that burglars use complex sets of cues to determine the wealth of a property but as this type of information is not available the income will be used as a proxy.
- **Occupancy**: Leeds levels of occupancy are calculated dynamically during the simulation based on the number of people in the groups “full time students”, “unemployed”, “part-time workers” and “economically inactive (looking after family)”. It is not possible to re-
9.4 Preparing the Offender Agents

It is necessary to estimate how many offenders there are in the model and where they live. Unfortunately, as no offender data is available this is somewhat problematic. Instead of using actual

Table 9.4: The mapping of Canadian and British ethnicities to groups used in the model.

<table>
<thead>
<tr>
<th>Group</th>
<th>UK Census Ethnicity</th>
<th>Canadian Census Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>British Irish</td>
<td>Total population minus all minorities</td>
</tr>
<tr>
<td>White other</td>
<td>White other</td>
<td>n/a</td>
</tr>
<tr>
<td>Black</td>
<td>Caribbean</td>
<td>Black Latin American</td>
</tr>
<tr>
<td></td>
<td>African</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other Black</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed White and Caribbean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed White and African</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>Indian</td>
<td>South Asian</td>
</tr>
<tr>
<td></td>
<td>Pakistani</td>
<td>Southeast Asian</td>
</tr>
<tr>
<td></td>
<td>Bangladeshi</td>
<td>Asian</td>
</tr>
<tr>
<td></td>
<td>Other Asian</td>
<td>Korean</td>
</tr>
<tr>
<td></td>
<td>Mixed White and Asian</td>
<td>Japanese</td>
</tr>
<tr>
<td>Chinese and others</td>
<td>Chinese</td>
<td>Chinese</td>
</tr>
<tr>
<td></td>
<td>Other Ethnic Group</td>
<td>Arab</td>
</tr>
<tr>
<td></td>
<td>Other Mixed</td>
<td>All others</td>
</tr>
</tbody>
</table>

create these classes directly, but the following describes how they can be estimated from the Canadian census:

– Full time students: can be estimated from the percentage of people aged 15-24 who are in full time education.

– Unemployed: available in the Canadian census directly.

– Working part time: unfortunately the Canadian census does not have information about people who work part time, but it does have the number of people who work from home. This can be used instead and is possibly a more accurate measure of occupancy anyway.

– Economically inactive looking after family: a close approximation is the number of people who spend more than 15 hours a week looking after children (unpaid).

Figure 9.9 provides the final measures of attractiveness and the average levels of occupancy (obviously actual occupancy values depend on the time of day).

• Similarity: as with Leeds analysis, the similarity of two sociotypes can be calculated from the Euclidean distance between all the variables which make up a sociotype. Figure 9.10 provides the similarity levels for a choice of different areas.
offender instances, therefore, it will be hypothesised that more offenders live in the more deprived neighbourhoods. There is strong empirical evidence for the link between offender rates and deprivation in Leeds (see Section 4.5.3) as well as considerable evidence in the literature (see Section 2.2.3). However, the absence of Canadian offender data makes it impossible to test this assumption. Even if predicted burglary rates are not as accurate as they were in the Leeds scenarios as a result of poor offender home locations, it does not detract from the ability of the model to provide reflections on the burglary system and how burglars interact with public transport. Therefore the number of offenders will be set so that the model is able to generate useful patterns (not too few agents) but also execute in an acceptable time (not too many agents). Using the VANDIX value, $v$, for each area the number of offenders can be calculated as $\|3v^2\|$, where $\|x\|$ means the value $x$ rounded to the nearest whole number. This results in 337 offenders in total which is similar to the 271 used in Leeds. An exponential function is used to focus offenders in the more deprived areas, otherwise after rounding the VANDIX value actually makes little difference to the location of offenders. Offenders in the Vancouver model are otherwise the same as those used in the Leeds scenarios (they have the same behavioural functions). This is a questionable assumption but, in the absence of data, a necessary one. Future work should experiment with different types of burglary behaviour, in particular examining the difference between burglars in UK cities and North American cities such as Vancouver. Figure 9.11 presents the VANDIX levels and the number of offenders created in each dissemination area.
9.4. PREPARING THE OFFENDER AGENTS

Figure 9.9: Levels of attractiveness and occupancy used in the calculation of Vancouver socio-types.

Figure 9.10: An example of the measure of similarity of different communities in Vancouver.
9.5 Real (Expected) Burglary Data

Although no offender data is available, there is some victim data available for 2001 in Vancouver which can be used to validate the model results. Unlike the Leeds data, the provided data has been aggregated up to the DA level. Therefore in order to allow the expanding-cell validation routine to function in the same manner, points were created randomly in each DA to generate a point dataset. Although this will obviously introduce error it is not a serious problem because the general pattern will be unaffected at all but the finest resolutions. Figure 9.12 illustrates the number of burglaries spatially.

Figure 9.12: Expected burglary rates in Vancouver, 2001. Hotspots were generated by applying the KDE algorithm to the points that were randomly generated based on aggregate district counts.
9.6 Vancouver SkyTrain Scenarios

The following scenarios will experiment with the effects of adding or removing a SkyTrain line. Future work will be able to expand upon these types of experiments by testing the effects of the new Canada Line train (Trans Link, 2009) which is being built for the 2010 Winter Olympics. As the Canada Line was incomplete at the time of writing, it was difficult to establish the exact route and the Millennium Line will be removed instead. This section will outline three scenarios that experiment with Vancouver public transport.

9.6.1 SkyTrain 1 – Normal Stations

Before running the scenarios it is important to determine whether or not the Leeds calibration changes are effective in the Vancouver environment as well. Therefore the first scenario (which simulates burglary patterns in the presence of normal SkyTrain lines) will be run twice; with and without the calibration changes. Figure 9.13(a) illustrates the results of a single model compared to expected data and Figure 9.13(b) graphs the results of the expanding cell analysis for both configurations. Although the calibration is having the desired effect of increasing the amount of clustering both results are a long way from the expected data, missing important hotspots in the north of the city. For this reason it is unsurprising that the non-calibrated model actually performs better; if the burglary locations are more spread out then there is more chance of inadvertently matching crime locations “by accident”.

As the non-calibrated results are mathematically closer to the expected data it might make sense to disregard the calibration changes for Vancouver. However, the maps clearly demonstrate that without calibration the simulated burglary locations are less dense those of the real data. So, for this reason, the calibration will remain in future Vancouver scenarios: even if the locations of hotspots are not entirely accurate at least the relative crime density is. It should be noted that graphs of the L functions for simulated and expected data would provide a useful demonstration of the similarity in the degree of clustering between the data. However, as the expected data points were disaggregated from DA-level data the clustering will not be accurate at lower resolutions so the analysis of L functions is not reliable.

From Figure 9.13(b), it appears that most burglaries in the expected data occur in the north of the city near downtown and the major Hastings, Broadway and Nanaimo routes. These are all relatively deprived areas but, with the exception of Downtown, are not significantly more deprived than most of the eastern section of the city (Figure 9.7 illustrated the city-wide deprivation levels). Therefore it is not clear why the northern areas should have such a large crime rate compared to other areas, and this is illustrated by the difficulty that the model has in predicting burglary correctly. Therefore the most likely reason for the poor predictions relates to the estimation of where the virtual burglars live and the extent to which the model scales in the face of unknowns. It is possible that, in the Leeds experiments, the home locations of the virtual offenders were not correct. But because the experiments cover a much smaller geographical area, most burglary locations fall within the burglars potential range regardless of where they live. However, as the
CHAPTER 9. EXPERIMENTS IN VANCOUVER

(a) The SkyTrain1 scenario with and without Leeds calibration compared to expected data.

(b) Expanding cell graphs comparing the accuracy of the SkyTrain 1 scenario (normal stations) with and without calibration.

Figure 9.13: Vancouver results with and without Leeds calibration compared to expected data
geographical area covered by the Vancouver simulation is much larger the offenders are not willing

to travel the great distances required to burgle in the “correct” areas in the northern part of the city
and the model breaks down.

As the results are different from the expected data, predictive modelling is not appropriate. However, as we will see the model does allow us to reflect on the dynamics of crime in the area
more generally and, in particular, how it might be influenced by public transport.

9.6.2 SkyTrain 2 – No Millennium Line

Having determined the burglary distribution when burglars are able to use public transport, this
scenario investigates burglary patterns before the addition of the Millennium Line SkyTrain. Other-
wise the model was configured with the same parameters as the previous experiment (including
the calibration changes). Figure 9.14 compares a single result without the SkyTrain to a previous
randomly chosen result. City-wide patterns are very similar across the two simulations. The main
difference is that when the Millennium Line is removed, the small hot-spots surrounding the ef-
fected stations disappear. This is to be expected as the agents no longer have any reason to travel
to these stations when there is no SkyTrain, whereas before they could start a journey to downtown
drug dealers from there.

![Comparing SkyTrain scenario results]

Figure 9.14: Comparing burglary patterns with and without the Millennium Line SkyTrain.

This result is somewhat disappointing because it is uninformative; it teaches us little about the
potential spatial distribution of burglary in the presence or absence of a public transport system. To
explore the effects of public transport on burglary in Vancouver further, a final scenario will
examine what happens when the entire public transport network is removed.

9.6.3 SkyTrain 3 – No Public Transport

Removing the entire public transport network is not a completely hypothetical scenario because it
is possible that the public transport system could be made inaccessible to potential burglars. There
are currently no restrictions on boarding the high-speed bus or rail lines, ticket checking is performed only sporadically and there are no gates or other obstacles. There is evidence, anecdotally at least, that some people use these public transport lines to travel around who would otherwise struggle to afford a fare. If these are the same people who are likely to attempt burglary then introducing barriers or some other mechanism to stop boarding without fares might be the same as closing off the transport system to agents in the model.

Figure 9.15 illustrates the results when no public transport routes were available to agents and compares them to previous Vancouver simulations. As with prior simulations, the results are again similar and the safest conclusion to draw is that public transport in this simulation has little effect on where an agent chooses to burgle.

![Comparing SkyTrain scenario results](image)

Figure 9.15: Results of the SkyTrain 3 scenario: no public transport routes.

### 9.7 Summary

The Vancouver scenarios have experimented with the influence of Vancouver’s public transport network on burglary rates. The literature examining the relationship between urban crime and public transport systems is contradictory (Smith and Clarke, 2000; Brantingham et al., 1991) so it is hoped that this simulation will shed some light on it.

Unlike the Leeds scenarios, however, the initial simulated burglary rates do not match expected rates. The model fails to account for the extreme hotspots in the north of the city near the Downtown area. It is likely that this is the result of missing Vancouver input data or different behaviour
by burglars in Vancouver compared to Leeds. Because the model cannot match expected data it should not be used to make crime forecasts, although this is not an unreasonable conclusion for a model that is using poor input data. If the potential offender locations were better known, for example, the results might be very different.

Nevertheless, the model is still able to provide reflections on the burglary system and in particular how the system responds to public transport in Vancouver. Chapter 5 showed that the model encapsulates current theoretical (criminology) and professional (police / crime reduction practitioner) understanding about the crime system and Chapter 7 demonstrated that it was a good predictor of burglary in Leeds. However, as the Vancouver model does not successfully replicate crime patterns there must be important elements of the system that have not been included. The model does include a realistic transport system so this is unlikely to be the missing feature. In fact, the transport system itself has little effect on simulated burglary patterns which suggests it is not important in the Vancouver model. As it does not change the behaviour of the model, it is unlikely to change behaviour in the real world either, although it is difficult to confirm this absolutely until the model has been better calibrated. However, this research provides initial evidence that policy makers in Vancouver should look for factors other than the SkyTrain system if they are interested in influencing burglary. It would be extremely interesting to apply similar tests to Leeds and determine if the same could be said about Leeds’ public transport but this must be reserved for future work.
Chapter 10

Conclusion

Contents

10.1 Introduction .......................................................... 249
10.2 A Summary of the Research Findings .......................... 250
10.3 A Critique of the Methodology .................................. 253
10.4 Recommendations for Future Work ............................. 255
10.5 Concluding Remarks ................................................ 256

10.1 Introduction

The work within this thesis is novel. There are no other published examples of an agent-based model applied to the residential burglary system that is able to account for the intricacies of the urban environment or the depth of the human behaviour exhibited by the model produced for this research. The model includes detailed information about the burglary risk of every house in the study area and this information is used to model burglary at the most highly detailed resolution possible, that of the individual house. Furthermore, the burglary risks associated with different neighbourhoods or communities are accounted for through the estimation of the behaviour of individuals other than potential burglars (such as homeowners). Along with individual houses, the behaviour of individual burglars is also modelled directly. Autonomous, intelligent burglar agents are controlled by an advanced cognitive framework that allows them to participate in normal daily routine activities as well as the act of burglary.

This chapter will conclude the thesis by summarising the main research findings and highlighting the extent to which the project objectives, as outlined in Chapter 1, have been met. Following this, a critique of the methodology is included in Section 10.3. The chapter concludes with recommendations for future work and some concluding comments.
10.2 A Summary of the Research Findings

As stated in Chapter 1, the overall aim of this research is to explore the use of agent-based modelling in the context of residential burglary. To accomplish this aim, several research objectives were formulated. This section will review these research objectives, discussing the extent to which they have been accomplished by the main body of the research.

1: Review and discuss the crime literature and available data to establish which factors drive the residential burglary system and therefore need to be included in a model

This objective was accomplished through a review of the crime literature in Chapter 2 and analysis of the available data in Chapter 4. Chapter 2 demonstrated initially that it is extremely important to assess the environment in which a crime occurs and that crime is a location-specific event that should be analysed at sufficiently high resolutions; aggregating to community boundaries often hides important patterns. After this early observation, the remainder of Chapter 2 reviewed the crime literature with the aim of determining which factors were the most important to describe the low-level dynamics of the burglary system. With respect to individual houses, research pointed to the importance of a house being easy to enter, not visible to neighbours or passers-by and not occupied by the residents. The surrounding community was also found to have an effect in two ways: burglars choose not to visit areas where they might stand out and areas with low community cohesion often experienced higher crime rates because the residents were unlikely to intervene. With respect to the burglars themselves, research points to the importance of drug addiction as the main motivation for burglary. Also, burglars were found to often commit burglaries in houses that they became aware of on their legitimate (non-burglary) daily activities as well as through a targeted search of the area. This information is invaluable and was used to inform the design of the model. Chapter 2 also reviewed three prominent theories in environmental criminology: routine activities theory, crime pattern theory and the rational choice perspective. These theories are able to amalgamate the otherwise disparate research findings and, therefore, along with empirical research findings, form the theory on which the model is based.

Chapter 4 extended the findings from the literature review through analysis of the available data. Highly accurate burglary data were available which provided important insights into the residential burglary system in Leeds. There was found to be a strong correlation between burglary and deprivation so specific indicators of deprivation were included in the model. Also, it was found that, in Leeds, there is a strong link between crime and student communities (the relationship was much more strongly defined than suggested in the literature which is likely a result of the unusually high student concentration). This highlighted the importance of including the spatio-temporal behaviour of people other than the potential burglars in order to simulate the presence of “student-like” communities.
2: Review and discuss the modelling techniques that have been used to model crime in order to highlight potential areas for research, inform the methodology used for this research and to guide the model development process

One of the key findings from Chapter 2 was that crime should be analysed at a high resolution; aggregating to levels above the individual house or street obscure essential information. However, Chapter 3 noted that most crime modelling work to date utilised regression techniques and involved the use of data that were aggregated to relatively large areas. These models were not able to account for the low-level factors that determined whether or not an individual crime might occur, such as the individual household security levels or the individual behaviour of a potential burglar. Also, all of the criminology theories that were introduced in Chapter 2 are based on the individual interactions between people and their local environment and, as such, should not be modelled by aggregate techniques. Chapter 3 introduced some computer modelling techniques that could, potentially, overcome some of these problems but it was found that they were unable to fully account for the effect of individual human behaviour.

Agent-based modelling was found to be the most appropriate solution to these problems and the few published examples of agent-based crime models were critiqued. The models presented by Birks et al. (2008), Hayslett-McCall et al. (2008) and Groff and Mazerolle (2008) have been improved upon by this research as follows. Birks et al. (2008) use the behaviour-based artificial intelligence cognitive framework to control their agents, and this is an improvement upon the other models that do not incorporate a model of human cognition at all. However, Chapter 5 showed that the PECS framework used here has much greater potential for modelling rich, complex human behaviour. Similarly, agents in the models of Birks et al. (2008) and Groff and Mazerolle (2008) have static behaviour; they cannot change their behaviour in response to unforeseen circumstances. Agents in the model produced here are fully dynamic, they continuously monitor their motivations and can change their current goal at any time and for any reason. With respect to travel, this model includes an array of different transportation methods that accurately reflect the methods that people use to navigate their urban environment. No other models attempt to account for this, indeed in the Hayslett-McCall et al. (2008) model agents do not move at all. The environment implemented here is also the most accurate; through an analysis of the literature in Chapter 2 many of the important household- and community-level factors were identified and these have been included in the model. No other models are able to account for the environmental realism and this model is, therefore, the most suited for modelling the crime theories discussed in Chapter 2.

3: Design and build an agent-based model that is able to account for the dynamics of the residential burglary system and create accurate predictions of crime rates

Chapter 5 tied the literature review and data analysis chapters to the later model development chapter through a discussion of how the two components of a model – the virtual environment and virtual burglars – can be conceptualised using the available theories and data. The chapter began with an overview of the environment which indicated how the “environmental backcloth”, as discussed in the crime literature, could be represented in a computer model. Three separate objects were used to represent the environment – roads, buildings and communities – and it was
shown how further analysis of the data was used to create appropriate values for the objects’
parameters. Using this scheme it was possible to account for a large degree of the complexity
of the crime system which the literature pointed to. This degree of complexity surmounts that
of all other published crime models. The second half of Chapter 5 discussed how the burglar
agents themselves were conceptualised. To improve upon other agent-based crime models, the
agents were equipped with an advanced cognitive framework which allows them to exhibit a wide
range of different behaviours, a feature that was absent from most models. The agents exhibit
realistic daily behaviours such as partaking in legitimate employment and visiting friends as well
as burgling. To complete the design of the agents, the chapter concluded with an assessment of
the set of actions available to them and discussed the actual process of burglary in detail.

Building upon the planned model, Chapter 6 discussed the formal model development pro-
cess. To make best use of the available agent-based technologies, the chapter began by listing
the available tools that could potentially be used and noted that, due to its excellent support for
working with the geographical data necessary to accurately reflect the environmental backcloth,
Repast Simphony was chosen. Before starting the main model development stages, a prototype
was constructed to illustrate that the model design was sound. Then the chapter technically docu-
mented the final model. The build process was able to account for all of the desired features and
was therefore able to live up to the requirements of the designed model. On the whole, Chapters 5
and 6 combined the findings from other crime research to produce a simulation that is ideally
suited to modelling the dynamics of the residential burglary system.

4: Evaluate the model by assessing its response to varying parameter values and by compar-
ing the results to known field conditions in order to gauge its predictive accuracy

An essential part of any modelling endeavour is to determine the extent to which the model is
able to replicate the system it is attempting to copy. This is extremely important for highly com-
licated models, such as that developed in this research, because the numerous parameters and
interactions can potentially make them as difficult to understand as the real system itself. To this
end, Chapter 6 concluded by rigorously testing the built model with experiments using idealised
data. In order to separate the part of the model that was being evaluated from unnecessary spatial
complexity, a novel feature was to use different types of environment of increasing geographical
realism. In this manner it was possible to fully understand the model in the context of a simple
system before increasing its complexity. This addresses the concerns of some environmental criminologists who worry that complex environments can make models more difficult to understand and ultimately detract from our understanding of the dynamics of the underlying system (Elffers
and van Baal, 2008).

Following the hypothetical experiments, Chapter 7 evaluated the model by experimenting with
it using real data from the city of Leeds. The chapter began with a discussion of the different ways
in which point data (i.e. burglary locations) can be compared and developed the “expanding cell”
routine – following Costanza (1989) – to allow for spatial data comparisons without the need for
aggregation to administrative boundaries. The remainder of the chapter then calibrated the model
to field data and finally validated it with a new data set. One of the most interesting findings from
10.3. A CRITIQUE OF THE METHODOLOGY

the calibration process (see Section 7.3.2) was the failure of the model to simulate a hotspot in the Halton Moor area. As the model has been built upon common assumptions about burglary – where burglars are motivated primarily by the need for money – the failure of the model to simulate the Halton Moor hotspot indicates that, in this area, the common assumptions might be incorrect. This has significant implications for crime reduction policies; those in Halton Moor might need to take into account of the potential different burglar motivations or risk being poorly targeted.

Unfortunately the validation suggested that the model had, in fact, been over-fitted to the calibration data, although the reason for this was possibly due to insufficient validation data. Regardless, Chapter 7 evaluated the model extensively and was able to determine where the model performed well and also the areas in which it performed poorly. Furthermore, as regression modelling was found to be the most common form of crime model in Chapter 3, the regression model built in Chapter 4 was compared to the agent-based model and it was found that the agent-based model marginally outperformed the corresponding regression model.

5: Apply the model to real-world scenarios in order to predict the potential impacts on residential burglary rates

To fully assess the impacts of the agent-based methodology to crime, Chapters 8 and 9 outlined the process of using the model to predict crime rates after altering the urban environment. Two case studies, both of direct relevance to their local populations, were utilised. The first, discussed in Chapter 8, covered a larger urban development scheme in Leeds entitled EASEL. By changing the physical and social environments to reflect the changes proposed by EASEL, the model was able to estimate burglary at the level of the individual household. Interestingly it was found that household security was a much stronger determinant of burglary than community-wide levels of cohesion. Also, the model highlighted houses close to those being developed which might expect to have a higher burglary risk as a direct result of the regeneration scheme.

The second scenario, outlined in Chapter 9, was able to demonstrate the flexibility of the model by applying it to a very different urban system; that of Vancouver, Canada. The model explored the effect that the high-speed public transport rail line (called SkyTrain) has on burglary rates in preparation for the opening of a new line for the 2010 Winter Olympics. Interestingly the model suggested that public transport had little effect on burglary patterns which is a finding at odds with common perceptions of the rail system. Therefore, although the model was not able to accurately predict real crime rates in Vancouver (potentially due to an absence of essential input data) it was nevertheless able to explore the dynamics of the burglary system and provide reflections on the relationship between urban form (i.e. the layout of the road network) and the occurrences of burglary.

10.3 A Critique of the Methodology

This thesis has explored the use of agent-based modelling in the context of residential burglary. The research has been successful in that a model has been built which is able to account for the low-level dynamics of the residential burglary system (i.e. the interactions between individual burglars
and individual houses) and predict global burglary rates. However, there are some weaknesses to the research that will be discussed here.

The UK census and the ecological fallacy

Chapter 2 illustrated that using aggregated data in the study of crime often hides important intra-area effects. A major advantage of this type of model is that it uses individual household data and, therefore, does not assume that all houses within a given administrative area are homogeneous with respect to their burglary risk. However, census data and the Output Area Classification (OAC) are also necessary inputs and there are two distinct problems with using these data.

The first problem relates to the census and the OAC; at the time of writing the 2001 census is eight years out of date. Therefore the OAC groups are potentially highly inaccurate. There are alternative classifications that could have been used, such as ACORN and MOSAIC, which are updated annually and might provide more accurate data\(^1\). However, Harland (2008) compared ACORN to a highly accurate census of school pupils (PLASC), which is also up to date, and found that there were large discrepancies between PLASC and ACORN. Therefore, although the use of census data is likely to provide a source of error, it is not clear how much of an advantage other classifications would have offered.

The second problem relates to the boundaries that are used to release census data. Administrative boundaries are not designed for the purposes of creating demographically homogeneous areas so it is highly unlikely that all people within an area will be identical. This problem is known to geographers as the ecological fallacy. Although the problem is mediated somewhat through the use of relatively small areas it will still provide a source of error. A similar problem relates to the homogeneity of the behaviour of the communities within the model. It is assumed that a particular type of community will always behave in a particular way but this might not necessarily hold true. For example, a rural community with a pub might exhibit different occupancy patterns in the evenings compared to an identical community somewhere else which does not have a local pub.

The most obvious means of addressing this problem in the context of agent-based modelling is to create agents to model the behaviour of all people directly rather than estimating their behaviour at an aggregate level, although this would greatly increase the complexity of the model.

Methodological individualism

An advantage with agent-based modelling is the possibility of emergence; i.e. a society is able to emerge from the interactions and beliefs of collections of individual agents. However, O’Sullivan and Haklay (2000) discuss how this notion of emergence is one way; although society can emerge from individual agents it is not possible for society to subsequently change the behaviour of the agents. In the context of this research, burglary is likely to affect a community in a variety of ways and this will then influence the behaviour of the residents and the potential burglars. For example, although houses in the model change their security precautions in a response to burglary, burglary does not influence other factors such as community cohesion. Therefore it is

\(^1\)The OAC was chosen over other classification systems because the methodology and data are publicly available.
not possible for a community to actively tackle a burglary problem, effecting behaviour from the “top down” as well as from the “bottom up”.

**More realistic burglars and burglary**

Although the model of burglary used here is the most comprehensive and realistic published example to date, there are some important factors that have not been accounted for. For example, research in Chapter 2 found that burglars regularly worked together but this type of cooperation is not accounted for in the model. Although the agents are able to affect each other indirectly through their effects on the environment, they cannot interact directly, i.e. cooperating on a burglary. Also, the model does not take account of a particularly important aspect of burglary: what the burglar does with the goods once they have stolen them. Clearly this will have a strong effect on their awareness spaces and the areas they subsequently might choose to burgle in.

**The complexity of the model**

Chapter 3 made reference to the complexity of the burglary system and found that the most appropriate way to account for this complexity was to use agent-based modelling to simulate the interactions that define the system directly. The model is able to account for criminology theory and burglar behavioural traits in great detail, but the downside is that assumptions that are not present in simpler models must be made explicitly here. For example, it must be decided how much time agents should spend socialising, how much money they can make from a burglary, what they do during the day etc. The advantages of high model complexity and flexibility are tempered by the difficulty of finding suitable values for these parameters based on empirical evidence.

**The calibration process**

Automatic calibration algorithms offer advantages to the manual calibration process exhibited here. Not only are they much more likely to find the optimal model configuration, but they can also reveal details about the effects that individual parameters have on the overall model results. For example, Malleson et al. (2009) showed that an automatic calibration routine found that certain model parameters did not influence the model results and could therefore be removed from the model. However, as Chapter 7 discussed, automatic calibration was not possible here due to the computation time of the model. Even on the most high-powered computer systems available the time requirements were unacceptable. Therefore the manual calibration process used the researcher’s knowledge of the model dynamics to improve the results, but this process is time consuming and inaccurate and there are undoubtedly improved model configurations yet to be found.

### 10.4 Recommendations for Future Work

There are several research directions from which this project could be extended. The most obvious avenue for future research is to make the model accessible to the police or policy makers. This thesis has shown that the model has potential as a forecasting tool and making it available to
policy makers might encourage a greater degree of planning with consideration to the effects of new schemes on burglary rates. Predicting burglary rates is otherwise an extremely difficult task.

Another practical use for the model is as a virtual criminology laboratory. Simulation is seen as a means of offering social scientists the tools for experimentation that the hard sciences take advantage of by providing a laboratory for repeatable experiments. Other work, e.g. Brantingham et al. (2008), is currently exploring the potential of this type of software for crime but the model produced here includes a much more advanced behavioural framework and greater environmental accuracy. It therefore lends itself cleanly to experimentation with criminology theory. By altering the dynamics of the model it is possible to change the underlying assumptions and, therefore, test the extent to which the theories cause the behaviour they are expected to in the real world.

With respect to improving the model itself, the most profitable direction might be to create individual agents to represent the citizens who live in the city. At present, their behaviour is estimated through information about the community in which they live, but this nevertheless omits an important element of the crime theories on which the model is based; the direct influence of home owners, their neighbours and passers-by on burglary. Work is currently underway to create extremely large agent-based models of city growth, so coupling these types of model with the burglary simulation might fulfil these requirements. However, this approach must be followed with caution as the resulting model will be considerably more complex that the current one. This would make it both highly computationally expensive and also extremely difficult to understand.

10.5 Concluding Remarks

This thesis has explored the use of agent-based modelling for analysing and predicting occurrences of residential burglary. Through the use of an advanced cognitive framework and a detailed virtual environment the research has created a comprehensive model that can directly account for the interactions and dynamics that drive the system. Although it inevitably has some drawbacks, the agent-based modelling approach is the most appropriate currently available technique for modelling a system that is characterised by individual interactions and contains intelligent organisms that exhibit complex behaviour. It is hoped that with some aesthetic improvements the model could be made available to the police or to policy makers and thus form part of the planning system for new crime reduction initiatives or urban developments.
Appendix A

The Output Area Classification (OAC)

This appendix includes a summary of all the variables used to construct the Output Area Classification (Vickers and Rees, 2006, 2007), a summary of the classification names and a map of the supergroups in Leeds.

Table A.1: Variables used to construct the Output Area Classification.

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic</strong></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>Age 0 - 4</td>
</tr>
<tr>
<td>v2</td>
<td>Age 5 -14</td>
</tr>
<tr>
<td>v3</td>
<td>Age 25 - 44</td>
</tr>
<tr>
<td>v4</td>
<td>Age 45 - 64</td>
</tr>
<tr>
<td>v5</td>
<td>Age 65+</td>
</tr>
<tr>
<td>v6</td>
<td>Indian/Pakistani/Bangladeshi</td>
</tr>
<tr>
<td>v7</td>
<td>Black African, Black Caribbean or Black Other</td>
</tr>
<tr>
<td>v8</td>
<td>Born outside UK</td>
</tr>
<tr>
<td>v9</td>
<td>Population density</td>
</tr>
<tr>
<td><strong>Household composition</strong></td>
<td></td>
</tr>
<tr>
<td>v10</td>
<td>Divorced</td>
</tr>
<tr>
<td>v11</td>
<td>Single person household (not pensioner)</td>
</tr>
<tr>
<td>v12</td>
<td>Single pensioner household (pensioner)</td>
</tr>
<tr>
<td>v13</td>
<td>Lone parent household</td>
</tr>
<tr>
<td>v14</td>
<td>Two adult no children</td>
</tr>
<tr>
<td>v15</td>
<td>Households with non-dependent children</td>
</tr>
<tr>
<td><strong>Housing</strong></td>
<td></td>
</tr>
<tr>
<td>v16</td>
<td>Rent (public)</td>
</tr>
<tr>
<td>v17</td>
<td>Rent (private)</td>
</tr>
<tr>
<td>v18</td>
<td>Terraced Housing</td>
</tr>
</tbody>
</table>
## APPENDIX A. THE OUTPUT AREA CLASSIFICATION (OAC)

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>v19 Detached Housing</td>
<td>percent of all household spaces which are detached</td>
</tr>
<tr>
<td>v20 All Flats</td>
<td>percent of all household spaces which are purpose built, converted and communal building flats</td>
</tr>
<tr>
<td>v21 No central heating</td>
<td>percent of occupied household spaces without central heating</td>
</tr>
<tr>
<td>v22 Rooms per household</td>
<td>average household size</td>
</tr>
<tr>
<td>v23 People per room</td>
<td>average number of people per room</td>
</tr>
<tr>
<td><strong>Socio-economic</strong></td>
<td></td>
</tr>
<tr>
<td>v24 HE qualifications</td>
<td>percent of people aged between 16 - 74 with a higher education qualification</td>
</tr>
<tr>
<td>v25 Routine/Semi-Routine occupation</td>
<td>percent of people aged 16-74 in employment working in routine or semi-routine occupations</td>
</tr>
<tr>
<td>v26 2+ Car household</td>
<td>percent of households with 2 or more cars</td>
</tr>
<tr>
<td>v27 Public transport to work</td>
<td>percent of people aged 16-74 in employment usually travel to work by public transport</td>
</tr>
<tr>
<td>v28 Work from home</td>
<td>percent of people aged 16-74 in employment who work mainly from home</td>
</tr>
<tr>
<td>v29 Llti (SIR)</td>
<td>percentage of working age population with limiting long term illness</td>
</tr>
<tr>
<td>v30 Provide unpaid care</td>
<td>percent of people who provide unpaid care</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
</tr>
<tr>
<td>v31 Students (full time)</td>
<td>percent of people aged 16-74 who are students</td>
</tr>
<tr>
<td>v32 Unemployed</td>
<td>percent of economically active people aged 16-74 who are unemployed</td>
</tr>
<tr>
<td>v33 Working part-time</td>
<td>percentage of economically active people aged 16-74s who work part time</td>
</tr>
<tr>
<td>v34 Economically inactive looking after family</td>
<td>percentage of economically inactive women aged 16-74 who are looking after the home</td>
</tr>
<tr>
<td>v35 Agriculture/fishing employment</td>
<td>percent of all people aged 16-74 in employment working in agriculture and fishing</td>
</tr>
<tr>
<td>v36 Mining/quarrying/construction employment</td>
<td>percent of all people aged 16-74 in employment working in mining, quarrying and construction</td>
</tr>
<tr>
<td>v37 Manufacturing employment</td>
<td>percent of all people aged 16-74 in employment working in manufacturing</td>
</tr>
<tr>
<td>v38 Hotel &amp; catering employment</td>
<td>percent of all people aged 16-74 in employment working in hotel and catering</td>
</tr>
<tr>
<td>v39 Health/social work employment</td>
<td>percent of all people aged 16-74 in employment working in health and social work</td>
</tr>
<tr>
<td>v40 Financial intermediation employment</td>
<td>percent of all people aged 16-74 in employment working in financial intermediation</td>
</tr>
<tr>
<td>v41 Wholesale/retail employment</td>
<td>percent of all people aged 16-74 in employment working in wholesale/retail trade</td>
</tr>
</tbody>
</table>
Figure A.1: The Output Area Classification supergroups in Leeds
Table A.2: Output area classification *group* and *super group* descriptions.

<table>
<thead>
<tr>
<th>Super group</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Blue Collar Communities</td>
<td>1a: Terraced Blue Collar</td>
</tr>
<tr>
<td></td>
<td>1b: Younger Blue Collar</td>
</tr>
<tr>
<td></td>
<td>1c: Older Blue Collar</td>
</tr>
<tr>
<td>2: City Living</td>
<td>2a: Transient Communities</td>
</tr>
<tr>
<td></td>
<td>2b: Settled in the City</td>
</tr>
<tr>
<td>3: Countryside</td>
<td>3a: Village Life</td>
</tr>
<tr>
<td></td>
<td>3b: Agricultural</td>
</tr>
<tr>
<td></td>
<td>3c: Accessible Countryside</td>
</tr>
<tr>
<td>4. Prospering Suburbs</td>
<td>4a: Prospering Younger Families</td>
</tr>
<tr>
<td></td>
<td>4b: Prospering Older Families</td>
</tr>
<tr>
<td></td>
<td>4c: Prospering Semis</td>
</tr>
<tr>
<td></td>
<td>4d: Thriving Suburbs</td>
</tr>
<tr>
<td>5: Constrained by Circumstances</td>
<td>5a: Senior Communities</td>
</tr>
<tr>
<td></td>
<td>5b: Older Workers</td>
</tr>
<tr>
<td></td>
<td>5c: Public Housing</td>
</tr>
<tr>
<td>6: Typical Traits</td>
<td>6a: Settled Households</td>
</tr>
<tr>
<td></td>
<td>6b: Least Divergent</td>
</tr>
<tr>
<td></td>
<td>6c: Young Families in Terraced Housing</td>
</tr>
<tr>
<td></td>
<td>6d: Aspiring Households</td>
</tr>
<tr>
<td>7: Multicultural</td>
<td>7a: Asian Communities</td>
</tr>
<tr>
<td></td>
<td>7b: Afro-Caribbean Communities</td>
</tr>
</tbody>
</table>
Appendix B

Demographic Variables and Crime Correlations

Burglary Victim Correlations

The following table illustrates the correlation using Pearson’s correlation coefficient, $\rho$, between OAC census variables and burglary (counts and rates per household) in each OA or SOA. All variables are significant at the 95% level (two-tailed). Crime data is from the period 1st April 2000 – 31st March 2002.

<table>
<thead>
<tr>
<th>Variable Name and OAC Number</th>
<th>Burglary count</th>
<th>Burglary rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OA</td>
<td>SOA</td>
</tr>
<tr>
<td>4 Age 45-64</td>
<td>-0.42</td>
<td>-0.58</td>
</tr>
<tr>
<td>37 Manufacturing employment</td>
<td>-0.36</td>
<td>-0.55</td>
</tr>
<tr>
<td>14 Two adults no children</td>
<td>-0.34</td>
<td>-0.59</td>
</tr>
<tr>
<td>33 Working part-time</td>
<td>-0.34</td>
<td>-0.53</td>
</tr>
<tr>
<td>30 Provide unpaid care</td>
<td>-0.31</td>
<td>-0.49</td>
</tr>
<tr>
<td>5 Age 65+</td>
<td>-0.27</td>
<td>-0.31</td>
</tr>
<tr>
<td>36 Mining/Quarrying/Construction employment</td>
<td>-0.24</td>
<td>-0.51</td>
</tr>
<tr>
<td>15 Households with non-dependant children</td>
<td>-0.26</td>
<td>-0.52</td>
</tr>
<tr>
<td>19 Detached Housing</td>
<td>-0.22</td>
<td>-0.45</td>
</tr>
<tr>
<td>26 2+ Car household</td>
<td>-0.24</td>
<td>-0.44</td>
</tr>
<tr>
<td>28 Work from home</td>
<td>-0.2</td>
<td>-0.42</td>
</tr>
<tr>
<td>12 Single pensioner household</td>
<td>-0.14</td>
<td>-0.12</td>
</tr>
<tr>
<td>25 Routine/Semi-Routine Occupation</td>
<td>-0.14</td>
<td>-0.15</td>
</tr>
<tr>
<td>3 Age 25-44</td>
<td>-0.1</td>
<td>-0.19</td>
</tr>
<tr>
<td>35 Agriculture/Fishing employment</td>
<td>-0.1</td>
<td>-0.26</td>
</tr>
<tr>
<td>10 Divorced</td>
<td>-0.06</td>
<td>-0.02</td>
</tr>
<tr>
<td>40 Financial intermediation employment</td>
<td>-0.09</td>
<td>-0.26</td>
</tr>
<tr>
<td>2 Age 5-14</td>
<td>-0.06</td>
<td>-0.23</td>
</tr>
<tr>
<td>41 Wholesale/retail trade employment</td>
<td>-0.04</td>
<td>-0.17</td>
</tr>
<tr>
<td>1 Age 0-4</td>
<td>-0.01</td>
<td>-0.14</td>
</tr>
<tr>
<td>22 Rooms per household</td>
<td>-0.04</td>
<td>-0.25</td>
</tr>
<tr>
<td>24 HE Qualification</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>34 Economically inactive looking after family</td>
<td>0.12</td>
<td>0.1</td>
</tr>
<tr>
<td>39 Health and Social work employment</td>
<td>0.12</td>
<td>0.26</td>
</tr>
</tbody>
</table>
### APPENDIX B. DEMOGRAPHIC VARIABLES AND CRIME CORRELATIONS

<table>
<thead>
<tr>
<th>Variable Name and OAC Number</th>
<th>Burglary count OA</th>
<th>Burglary rate OA</th>
<th>Burglary count SOA</th>
<th>Burglary rate SOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 Terraced Housing</td>
<td>0.12</td>
<td>0.13</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>13 Lone Parent household</td>
<td>0.14</td>
<td>0.14</td>
<td>0.2</td>
<td>0.15</td>
</tr>
<tr>
<td>21 No central heating</td>
<td>0.17</td>
<td>0.15</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>16 Rent (Public)</td>
<td>0.19</td>
<td>0.16</td>
<td>0.35</td>
<td>0.25</td>
</tr>
<tr>
<td>11 Single person household (not pensioner)</td>
<td>0.22</td>
<td>0.17</td>
<td>0.45</td>
<td>0.32</td>
</tr>
<tr>
<td>29 Llti (SIR)</td>
<td>0.21</td>
<td>0.18</td>
<td>0.36</td>
<td>0.28</td>
</tr>
<tr>
<td>9 Population Density</td>
<td>0.22</td>
<td>0.23</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>20 All Flats</td>
<td>0.25</td>
<td>0.23</td>
<td>0.52</td>
<td>0.42</td>
</tr>
<tr>
<td>32 Unemployed</td>
<td>0.26</td>
<td>0.24</td>
<td>0.42</td>
<td>0.34</td>
</tr>
<tr>
<td>38 Hotel Catering employment</td>
<td>0.26</td>
<td>0.27</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>23 People per room</td>
<td>0.26</td>
<td>0.31</td>
<td>0.33</td>
<td>0.42</td>
</tr>
<tr>
<td>- IMD Score</td>
<td>0.34</td>
<td>0.31</td>
<td>0.48</td>
<td>0.4</td>
</tr>
<tr>
<td>27 Public Transport to work</td>
<td>0.35</td>
<td>0.34</td>
<td>0.54</td>
<td>0.5</td>
</tr>
<tr>
<td>7 Black, Black Caribbean or Other Black</td>
<td>0.34</td>
<td>0.35</td>
<td>0.55</td>
<td>0.5</td>
</tr>
<tr>
<td>6 Indian, Pakistani or Bangladeshi</td>
<td>0.34</td>
<td>0.36</td>
<td>0.48</td>
<td>0.51</td>
</tr>
<tr>
<td>17 Rent (Private)</td>
<td>0.34</td>
<td>0.36</td>
<td>0.5</td>
<td>0.51</td>
</tr>
<tr>
<td>8 Born Outside the UK</td>
<td>0.37</td>
<td>0.39</td>
<td>0.57</td>
<td>0.55</td>
</tr>
<tr>
<td>31 Students (full-time)</td>
<td>0.45</td>
<td>0.49</td>
<td>0.59</td>
<td>0.64</td>
</tr>
</tbody>
</table>

### Burglar (Offender) Correlations

The following table illustrates the correlation using Pearson’s correlation coefficient, ρ, between OAC census variables and offender home locations (counts and rates per household) in each OA or SOA. All value are significant at the 95% level (two-tailed). Crime data is from the period 1st April 2000 – 31st March 2002.

<table>
<thead>
<tr>
<th>Variable Name and OAC Number</th>
<th>Offender count OA</th>
<th>Offender count SOA</th>
<th>Offender rate OA</th>
<th>Offender rate SOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>26 2+ Car household</td>
<td>-0.26</td>
<td>-0.26</td>
<td>-0.5</td>
<td>-0.49</td>
</tr>
<tr>
<td>14 Two adults no children</td>
<td>-0.18</td>
<td>-0.17</td>
<td>-0.39</td>
<td>-0.39</td>
</tr>
<tr>
<td>22 Rooms per household</td>
<td>-0.15</td>
<td>-0.17</td>
<td>-0.34</td>
<td>-0.34</td>
</tr>
<tr>
<td>19 Detached Housing</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.34</td>
<td>-0.34</td>
</tr>
<tr>
<td>30 Provide unpaid care</td>
<td>-0.13</td>
<td>-0.14</td>
<td>-0.21</td>
<td>-0.22</td>
</tr>
<tr>
<td>24 HE Qualification</td>
<td>-0.14</td>
<td>-0.12</td>
<td>-0.24</td>
<td>-0.23</td>
</tr>
<tr>
<td>4 Age 45-64</td>
<td>-0.13</td>
<td>-0.11</td>
<td>-0.25</td>
<td>-0.25</td>
</tr>
<tr>
<td>28 Work from home</td>
<td>-0.13</td>
<td>-0.11</td>
<td>-0.35</td>
<td>-0.34</td>
</tr>
<tr>
<td>40 Financial intermediation employment</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.32</td>
<td>-0.32</td>
</tr>
<tr>
<td>5 Age 65+</td>
<td>-0.11</td>
<td>-0.14</td>
<td>-0.08</td>
<td>-0.13</td>
</tr>
<tr>
<td>42 NumPeople</td>
<td>0.03</td>
<td>-0.07</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>15 Households with non-dependant children</td>
<td>-0.04</td>
<td>-0.06</td>
<td>-0.19</td>
<td>-0.19</td>
</tr>
<tr>
<td>33 Working part-time</td>
<td>-0.05</td>
<td>-0.05</td>
<td>-0.13</td>
<td>-0.13</td>
</tr>
<tr>
<td>12 Single pensioner household</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>35 Agriculture/Fishing employ</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.13</td>
<td>-0.13</td>
</tr>
<tr>
<td>37 Manufacturing employment</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.06</td>
</tr>
<tr>
<td>36 Mining/Quarrying/Construction employment</td>
<td>0.03</td>
<td>0.01</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>41 Wholesale/retail trade employment</td>
<td>0.05</td>
<td>0.03</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Variable Name and OAC Number</td>
<td>Offender count</td>
<td>Offender rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------</td>
<td>---------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OA SOA</td>
<td>OA SOA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Age 5-14</td>
<td>0.07 0.11</td>
<td>0.04 0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39 Health and Social work employment</td>
<td>0.05 0.1</td>
<td>0.04 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Age 25-44</td>
<td>0.05 0.06</td>
<td>0.05 0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 Students (full-time)</td>
<td>0.08 0.15</td>
<td>0.06 0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Age 0-4</td>
<td>0.11 0.19</td>
<td>0.08 0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Rent (Private)</td>
<td>0.08 0.13</td>
<td>0.08 0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Indian, Pakistani or Bangladeshi</td>
<td>0.11 0.22</td>
<td>0.1 0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Born Outside the UK</td>
<td>0.11 0.25</td>
<td>0.1 0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 Terraced Housing</td>
<td>0.12 0.23</td>
<td>0.1 0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Population Density</td>
<td>0.11 0.25</td>
<td>0.11 0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 People per room</td>
<td>0.16 0.31</td>
<td>0.11 0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 Hotel / Catering employment</td>
<td>0.11 0.29</td>
<td>0.11 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 No central heating</td>
<td>0.16 0.29</td>
<td>0.14 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 All Flats</td>
<td>0.13 0.33</td>
<td>0.15 0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Routine/Semi-Routine Occupation</td>
<td>0.16 0.26</td>
<td>0.15 0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Black african, Black Caribbean or Other Black</td>
<td>0.17 0.36</td>
<td>0.16 0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34 Economically inactive looking after family</td>
<td>0.2 0.34</td>
<td>0.17 0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Single person household (not pensioner)</td>
<td>0.16 0.35</td>
<td>0.18 0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Lone Parent household</td>
<td>0.22 0.41</td>
<td>0.19 0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Divorced</td>
<td>0.19 0.33</td>
<td>0.2 0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 Public Transport to work</td>
<td>0.22 0.41</td>
<td>0.21 0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Rent (Public)</td>
<td>0.23 0.47</td>
<td>0.23 0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Burglary count</td>
<td>0.28 0.44</td>
<td>0.24 0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29 Llti (SIR)</td>
<td>0.26 0.49</td>
<td>0.25 0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 Unemployed</td>
<td>0.28 0.52</td>
<td>0.27 0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- IMD Score</td>
<td>0.3 0.56</td>
<td>0.28 0.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

Computer Environment and Urban Systems Paper

The following paper (Malleson et al., 2010b), jointly authored by the writer of this thesis, has been published in the journal *Computers, Environment and Urban Systems*. Parts of the research have been included in Chapter 6.
Crime reduction through simulation: An agent-based model of burglary

Nick Malleson*, Alison Heppenstall, Linda See
School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

ARTICLE INFO

Article history:
Received 30 April 2009
Received in revised form 25 August 2009
Accepted 22 October 2009

Keywords:
Agent-based modelling
Burglary
Computer simulation
Crime reduction

ABSTRACT

Traditionally, researchers have employed statistical methods to model crime. However, these approaches are limited by being unable to model individual actions and behaviour. Brantingham and Brantingham (1993) described that in their opinion a useful and productive model for simulating crime would have the ability to model the occurrence of crime and the motivations behind it both temporally and spatially. This paper presents the construction and application of an agent-based model (ABM) for simulating occurrences of residential burglary at an individual level. It presents a novel framework that allows both human and environmental factors to be simulated. Although other agent-based models of crime do exist, this research represents the first working example of integrating a behavioural framework into an ABM for the simulation of crime. An artificial city, loosely based on the real city of Leeds, UK, and an artificial population were constructed, and experiments were run to explore the potential of the model to realistically simulate the main processes and drivers within this system. The results are highly promising, demonstrating the potential of this approach for both understanding processes behind crime and improving policies and developing effective crime prevention strategies.

1. Introduction

Understanding the processes behind crime is an important research area in criminology which has major implications for both improving policies and developing effective crime prevention strategies (Brantingham & Brantingham, 2004; Groff, 2007a). Recent advances in criminology, such as routine activities theory (Cohen & Felson, 1979), lifestyle exposure theory (Hindelang, Gottfredson, & Carafano, 1978) and crime pattern theory (Brantingham & Brantingham, 1993) have highlighted a shift from aggregated data and they are unable to represent the micro-level human and environmental factors to be simulated. Although other agent-based models of crime do exist, this research represents the first working example of integrating a behavioural framework into an ABM for the simulation of crime. An artificial city, loosely based on the real city of Leeds, UK, and an artificial population were constructed, and experiments were run to explore the potential of the model to realistically simulate the main processes and drivers within this system. The results are highly promising, demonstrating the potential of this approach for both understanding processes behind crime and improving policies and developing effective crime prevention strategies.

When compared to any other local authority in England and Wales (Shepherd, See, Kongmuang, & Clarke, 2004). However, one of the central challenges of modelling a system as complicated as that of residential burglary lies in simulating human behaviour within a computer environment. Humans exhibit soft factors such as seemingly irrational behaviour and complex psychology (Bona-beau, 2002); these characteristics are highly challenging to simulate in a computer model. This is compounded by the fact that burglars can be classified as experts in their field (Nee & Meenaghan, 2006); they possess a range of both behavioural characteristics and specific knowledge that is unique to them. Many studies have interviewed burglars (both incarcerated and active) to gather qualitative evidence regarding their behaviour and motives (Brown & Bentley, 1993; Cromwell & Olson, 2005, chap. 5; Hearnden & Magill, 2003; Nee & Meenaghan, 2006; Wright & Decker, 1996). While these studies have revealed valuable insights into the possible behaviour and motives of offenders (many of which have been incorporated into the design of the model presented here), they often suffer from problems associated with sampling of a small population and lack of rigorous empirical testing. Quantitative studies have helped to establish general trends in burglars’ behaviour (Bernasco & Luykx, 2003; Massey, Krohn, & Bonati, 1989; Snook, 2004). However, these approaches are limited due to the use of aggregated data and they are unable to represent the micro-level human and environmental factors that dictate whether or not an individual crime event will occur.

One technique that shows considerable promise for overcoming these limitations is agent-based modelling (ABM). ABM represents...
a shift in the social sciences towards the use of models that work at the level of the individual. For a recent overview of applications, see Paredes and Hernández (2008). ABMs are comprised of autonomous, decision making entities called agents that have the ability to interact with each other and their environment (Bonabeau, 2002). Agents can represent individuals, groups of individuals and, if appropriate, inanimate objects such as houses or cars. As the model iterates, each agent has the ability to assess its circumstances and, based on a set of probabilistic rules, makes an informed/educated decision about its future course of action (Bonabeau, 2002). Through this mechanism, more realistic human behaviour can be incorporated (Moss & Edmonds, 2005).

Simulating residential burglary is a particularly challenging problem, largely because the system can be regarded as highly complex. Not only does it contain potentially unlimited entities (broadly categorised as social, environmental and behavioural factors perceived by and unknown to the physical environment), the system is highly dynamic, changing both over time and space. For example, an occurrence of a residential burglary is affected by the time of day, a combination of spatial and environmental factors (e.g. low security, easily accessible property) and individual behaviour (opportunistic crime, individual motivation). One of the most attractive elements of ABM is the ability to experiment with different crime theories and reduction policies before implementation in the real system. Examples of this type of application can be found in the areas of urban planning (Al-Ahmad, Heppenstall, Hogg, & See, 2009) and education (Harland & Stillwell, in press, chap. 16). The development and application of an ABM for simulating residential burglary thus provides a unique opportunity to both further understanding of the processes and dynamics of this system as well as providing a platform for testing out crime reduction policies.

Brantingham and Brantingham (1993) described that, in their opinion, a useful and productive model for simulating crime would include the ability to model the occurrence of crime and the motivations behind it in a dynamic time and space. This paper presents the development and application of an ABM for simulating the occurrence of crime (specifically residential burglary) at an individual level. A particular focus of this paper is the modification and inclusion of the physical conditions, emotional states, cognitive capabilities and social status (PECS) framework for simulating more realistic human behaviour within a computational/artificial environment. The model is tested through the development of an artificial city loosely based on the real city of Leeds, UK. Previous approaches to crime modelling are discussed in Section 2, illustrating how this approach enhances existing work to date. The PECS framework along with details of how this was integrated into the ABM is outlined in Sections 3 and 4. A series of experiments testing the behaviour and robustness of the model are presented in Section 5. Sections 6 and 7 conclude with a discussion of the results and a critique of the methodology while future work is briefly outlined in Section 8.

2. Previous approaches to crime modelling

The crime system is driven by a large number of interrelating factors. These include, but are not limited to, an offender’s individual perceptions and knowledge of the suitability or attractiveness of the target, the offenders cognitive representation of the environment, the layout of the physical environment and other factors relating to the surrounding community (Brantingham & Brantingham, 1993).

Environmental criminology has employed numerous methods for understanding and examining the most important environmental factors that influence how criminals choose their targets (Brantingham & Brantingham, 1993). Early seminal work by Shaw and McKay (1969) utilized mapping techniques to investigate the link between juvenile delinquency and social or cultural characteristics. The authors found that juvenile delinquency rates were at their highest in city centres and exhibited similar spatial patterns to other indicators of social problems. Advances in geographical information systems (GIS) and the availability of individual-level data have catalysed the development of more advanced mapping analysis techniques such as “hotspot detection (Grubesic & Murray, 2001). For example, Pain, MacFarlane, Turner, and Gill (2006) overlaid crime hotspot maps with streetlight location maps to investigate the impact that street lighting had on crime and fear of crime levels; the results were used to inform existing street lighting policies. Although these techniques are invaluable for crime prevention practitioners (due in part to their ability to highlight areas with unusually high crime rates), they fail to provide insights into the dynamics and processes that generate these phenomena.

In addition to mapping techniques, statistical or mathematical models have also been widely used. Early examples include the use of principal components analysis to investigate the factors related to social deprivation (Giggs, 1970) and cluster analysis to search for associations between crime and environmental factors (Brown, Mcculloch, & Hiscox, 1972) (although these approaches are strongly criticised by Baldwin (1975) who describes them as “unilluminating”). More recent statistical modelling has been centred around the use of regression models. For example, Craglia, Haining, and Signoretta (2001) compared high intensity crime areas to census data whilst Dahlbäck (1998) found high population density and weak social bonds to be associated with high theft rates through application of longitudinal multivariate regression. Other studies using regression include Gaviria and Pagès (2002) research that linked the chance of being a victim to individual and city-wide variables, and Meera and Jayakumar (1995) who attempted to explain the relation between rising levels of crime and different demographic and economic variables. These approaches have revealed interesting links between crime and other variables, but they are unable to account for the motivations and impact of individual actions upon both other individuals and the environment.

Advances in software engineering catalysed by increases in computer data storage and processing power has precipitated an uptake in computational approaches to the modelling of crime. A recent example can be found in the work of Kongmuang, Clarke, Evans, and Ballas (2005) and Kongmuang (2006) who utilised spatial micro-simulation and spatial interaction models to investigate urban residential burglary rates. This research both successfully estimated offender flows within a city and predicted the risk of being a victim of residential burglary at the individual level. Despite the advances that this technique provided, this work was limited by the inherent inability of micro-simulation to model interactions between individual entities and most importantly cannot represent human behaviour.

A central drawback common to each of the approaches discussed above is that they fail to address the importance of individual incidents located in a specific time and space. Instead, findings are concerned with general, aggregate patterns; this makes it difficult to draw conclusions regarding how the individual behaviour of victims or offenders may be affecting the occurrence and rate of crime. Brantingham and Brantingham (1993) describe that, essentially, the most productive model in criminology will be the model that “places both the actual criminal events at a specific site, situation and time and the individual committing the crime while in a specific motivational state on (or in) an environmental backcloth, that may itself be mostly stable, regular and predictable or may instead be irregular, rapidly changing and unpredictable.” Due to their aggregate nature, traditional statistical modelling techniques
are limited in their ability to represent local variation present in the “environmental backcloth”. Factors such as the individual location of houses (e.g. corner blocks) (Taylor & Nee, 1988), their visibility to neighbours and passers-by (Robinson & Robinson, 1997) and the layout of the local street network (Bevis & Notter, 1977) will affect their propensity to be burgled; however, these factors cannot be incorporated into models which do not operate at the level of the individual.

To improve our understanding of the trends and characteristics of crime patterns, it is necessary to examine the individual actors who play important roles in discrete crime events. ABM has been applied to a vast number of subject areas including computer systems that assist car drivers (Miller, Hwang, Torkkola, & Massey, 2003), pedestrian movements (Castle & Crooks, 2006; Turner & Penn, 2002), human immune systems (Jacob, Litorco, & Lee, 2004) and simulating processes and dynamics in the retail petrol market (Heppenstall, Evans, & Bratman, 2005). Despite this uptake, the potential benefits of ABM are only just beginning to be realised in criminology; current work is briefly outlined below. For a detailed review of the theory and concepts behind ABM, the reader is directed to Axelrod (2000).

Early agent based crime models were relatively simple but began to show that the technique can hold promise in the field of criminology. For example, Winoto (2003) investigated rational choice and whether a society without crime is attainable, Gundersson and Brown (2000) presented a methodology for predicting both physical- and cyber-crime and Melo, Belchior, and Furtado (2005) modelled police patrol route reorganisation. More recently, even more advanced models have begun to emerge. Notable sources are the recent book entitled “Artificial Crime Analysis” (Dray, Mazerolle, Perez1, & Ritter, 2008), street robbery in Seattle (Groff, 2006, 2007a, 2007b), crime patterns in Cincinnati (Liu, Wang, Eck, & Liang, 2005) and burglary in an abstract environment (Haylett-McCAll et al., 2008, chap. 14). Agent-based crime models are also under development whose aim is not to actually predict crime rates but experiment with criminological ideas. For example, Brantingham, Glasser, Kinney, Singh, and Vajihollahi (2005b, chap. 13), have used the abstract state machine formalism to represent agents who have memory, behaviour and motivations who can be situated in an abstract environment. The resulting simulation can be used as an interdisciplinary tool to assist criminologists in investigating the dynamics of urban crime. In a similar vein, Wang, Liu, and Eck (2008, chap 11) outlined a tool to study the interactions between actors involved in a crime event.

To improve upon the previous models, this research will present a model that includes a larger number of factors and a more accurate model of human behaviour to better represent the real burglary system. These factors include the effect of drug addictions (Wright & Decker, 1996), offenders’ perceptions of their physical environment (Brantingham & Brantingham, 1993; Beaven, Brantingham, & Brantingham, 1994) and social status (which can be replicated (Singh, 2005a)). Despite its uptake, BDI has been widely criticised. Some authors criticise the three core components (beliefs, desires, and intentions) of the architecture as being too restrictive while others feel that they are overly complicated (Rao et al., 1995). Fundamentally, the architecture assumes rational decision making; this is difficult to justify because people rarely meet the assumptions of rational choice (Axelrod, 1997). Brailsford and Schmidt (2003) see the restriction of the architecture to cognitive processes as a limitation; BDI cannot integrate physical, emotional or social processes or the interactions between them. Balzer (2000, chap. 5) also notes that the core elements are difficult to observe directly: observation can only be achieved in a laboratory setting which is unlikely to relate to real situations.

An alternative, but rarely used, architecture is the PECS framework (Physical conditions, Emotional states, Cognitive capabilities and Social status). Proposed by Schmidt (2000) and Urban (2000, chap. 6), this architecture states that human behaviour can be modelled by taking into account physical conditions, emotional states, cognitive capabilities and social status. Personality is incorporated into the agents by adjusting the rate that internal state variables change and also how these changes are reflected in agent behaviour (Schmidt, 2002). The framework is modular, allowing separate components to control each aspect of the agent’s behaviour (Martinez-Miranda & Aldes, 2003). Proponents of PECS cite that as rational decision making is not required and the framework is not restricted to the factors of beliefs, desires, and intentions (Schmidt, 2000), it is an improvement on the BDI architecture.

To illustrate the PECS features, an example proposed by Urban (2000, chap. 6) is adapted here. Consider a person in a shop who is contemplating purchasing some goods. They might experience physical needs (such as hunger), emotional states (such as surprise at the available goods), cognition (such as information about current prices) and social status (which could affect how the agent reacts to the shop assistant). Schmidt (2000) and Urban (2000, chap. 6) argue that every aspect of human behaviour can be modelled using these components although, depending on the application, it might not be necessary to incorporate all of them (Schmidt, 2002).

Despite documented use of the framework being limited, the applications that have incorporated it are diverse. For example,
PECS has been used to build emotions into a virtual learning environment (Ammar, Neji, & Gourdel et al., 2006; Neji & Ammar, 2007). Here, non-verbal communication was incorporated in the form of emotional facial expressions with the aim of improving the relationship between a human learner and a computer-controlled tutor. In the field of health care, Bradlford and Schmidt (2003) used the framework to improve a simulation of disease screening. The authors noted that through the use of PECS they were able to incorporate individual behaviour; an important determinant of a patient’s attendance at a screening session, a factor that is absent from the majority of models in their field.

PECS places behaviour into two categories: reactive and deliberative. Reactive behaviour encompasses actions that are largely instinctive; no deliberation is required. Schmidt (2000) describes how reactive behaviour can be subdivided:

- **Instinctive behaviour.** An automatic reaction to stimulus depending on the internal state of the agent, for example, a parent reacting instinctively to a child’s cry. Instinctive behaviour can be easily incorporated using pre-defined rules.

- **Learned behaviour.** Here, rules are learnt dynamically, for example Schmidt (2000) cites the example of a car driver who instinctively brakes if a child runs in front of their car.

- **Drive controlled behaviour.** This behaviour is directed by internal drivers to satisfy needs. Needs range from basic, for example preserving life (such as the need for food or safety) to social needs and intellectual needs. The drivers determine an individual’s behaviour as they attempt to satisfy the drive with the greatest intensity. The following function is used to determine drive intensity:

  \[ T = f(N, V, X) \]

  Where \( N \) is the need, \( V \) represents environmental influences and \( X \) represents other influences. For example, if hungry, an individual will have a strong drive to eat if \( N \) is high. However, the environment also plays a part; the drive to eat may be stronger if the person can smell food, even if \( N \) is not great.

- **Emotionally controlled behaviour.** As with drives, if emotions are strong enough this will dictate the behaviour of the agent. However, the key difference is that they are stimulated externally, and not by internal state changes. Schmidt (2000) defines the intensity of emotions, \( E \); as:

  \[ E = g(I, A, X) \]

  Where \( I \) represents the importance of the event that has generated the emotion, \( A \) the agent’s personal assessment of the event and \( X \) represents other influences.

Schmidt (2000) also discusses deliberative behaviour. With reactive forms of behaviour the organism is not truly aware of the reasons that cause their behaviour. For example, they are not aware that looking for food is a task which ultimately ensures survival. Agents who engage in deliberative behaviour, however, do so in order to consciously pursue goals. These goals, such as take up a new hobby, can be complex and might involve numerous intermediate targets. As Section 4.1 will illustrate, the model presented here primarily uses reactive forms of behaviour to drive the agents, but the agents then use deliberative techniques to satisfy their goals.

The next section will outline how the PECS framework is implemented in an agent-based model (ABM) to introduce realistic behaviour into the agents (people).

### 4. An agent-based model of burglary

This section explains the framework of the agent-based model of burglary, in particular the characteristics, behaviours and cognitive maps of the offender agents and the model’s physical environment.

#### 4.1. The agents

The model is populated by “people” agents. All agents possess the same basic structure and fundamental needs: the need to generate wealth and the need to sleep (more details will follow). These people agents are further divided into two groups: those who can always generate sufficient wealth through legitimate work (termed ‘citizens’) and those who do not have sufficient employment and must burglar occasionally (the ‘potential burglars’). Potential burglars are assigned random amounts of work each day; however the amount of work does not always fully satisfy their need for wealth. This behaviour is consistent with the literature. For example, Wright and Decker (1996) found that burglars are often employed and this employment can lead to them recognising new, suitable targets that they would be otherwise unaware of.

As discussed later, we recognise that this is a vast simplification and do not support the notion that all unemployed people are burglars!

The roles of the agents do not change; a potential burglar agent cannot become a citizen and vice-versa. Although this is a simplification of real life, whereby external circumstances might drive people towards or away from burglary (effectively changing their “role”), the model is based on the burglar’s individual behaviour and their relationship with the physical environment rather than the social or political processes which drive people towards a life of crime. This area of research will, however, inform future work.

The number of agents in each model run can be varied, but for the experiments outlined here, was fixed at 300. This value was chosen to ensure that the majority of the houses in the environment are occupied by an agent. However, there are also unoccupied houses: this allows for future examination of the difference that citizen daily habits have on their burglary risk. When they are created, each agent has a 5% chance of being a burglar agent and a 95% chance of being a citizen. These percentages have been chosen under the assumption that the number of burglars in the population is small. As the model is not trying to predict actual numbers but spatial patterns, this weighting is felt to be reasonable. The total number of agents and the probabilities of being a burglar or citizen can be varied but they are kept constant for all experiments outlined here. This results in approximately 15 burglar agents in each model, although (due to the probabilistic nature of agent generation) the total number in each run will vary slightly.

Wealth is used to encompass factors that require money for satisfaction, for example, the need to buy food, socialise, support a family or sustain a drug addiction. All agents also require sleep which must be sought at home. Levels of wealth and sleep deteriorate at a constant rate throughout the simulation and can be replenished by working, burgling or sleeping. Using these two needs it is possible to create behaviour which can be generally found in the daily patterns of employed people in most cities. An avenue for future research is the transference of this work from a homogenised case-study to an application based on a real city.

**Fig. 1** and Table 1 illustrate how the needs of an agent drives their actions. PECS intensity functions are used to calculate which need is the greatest at each time step. For each agent, the intensity functions take into account the current levels of wealth and sleep, the agent’s personal preference for generating wealth or sleeping and the current time of day. It should be noted that not all the possible features of the PECS framework have been included in the model at this stage. For example, social variables do not play a part in the model. As Schmidt (2000) notes, it is important to choose the behavioural factors which are important in the chosen system, not to try to include all possible variables. At this stage the effects of...
social interactions are deemed too complex to be of use in the model.

Personal preferences allow for the inclusion of heterogeneous agents. For example, a drug addiction which requires considerable wealth to satisfy could be simulated by including agents whose personal preferences for generating wealth are higher than others. At this stage of research, however, personal preferences are not varied so the burglary agents are homogeneous.

Fig. 2 illustrates how $T$ influences the overall intensity of the wealth and sleep needs where time $t = 0$ is set to approximately 7am. The need to work is largest during the day whereas the need for sleep is the strongest during the night.

It is worth noting that, at a first glance, it appears that the model only includes the most basic of human behaviours as stipulated by PECS: that of reactive behaviour. The agents have simple needs that they must satisfy and the strongest of these needs drives their behaviour. However, the behaviour required to satisfy a need is more complex than simple reactive behaviours will allow for. The agents use learned behaviour (they remember where they have visited which influences future choices regarding where to look for burglary targets) and even deliberative behaviour when trying to find a burglary target as they are involved in “conscious pursuit of goals” (Schmidt, 2000). So whilst the agents do not know why they need to satisfy their goals (a reactive trait) the methods they use to satisfy them are complex and involve the conscious pursuit of goals with intermediate stages (a deliberative trait).

### 4.2. Cognitive maps and temporary employment

An important feature of the model is the inclusion of ‘cognitive maps’. These maps represent each agent’s internal representation of their environment. According to routine activities theory, crime pattern theory and qualitative studies (Cronwell & Olson, 2005, chap. 5; Wright & Decker, 1996), a potential offender is likely to find a suitable target by passing one on their routine travels. The model presented here uses this theory with the cognitive maps adapted from Brantingham and Brantingham (1993) activity spaces concepts. As the agents move around the environment (whether they are looking for a burglary target, or simply travelling to work or home) they remember each house that they have passed. These houses and their locations in the environment are stored internally by each agent as a list. The agents also remember the levels of security and attractiveness of the houses that they have stored in their map. These parameters will be explained in the following sections along with a detailed description of how the agents’ cognitive maps are used for burglary.

Assigning potential burglar agents temporary employment allows the agents to visit areas of the environment that they might not do otherwise. This helps them to build up their cognitive maps. Although employment is seen as an essential aspect of the burglary system the different types of employment that agents can engage in (for example industries that service houses such as delivery

---

**Table 1**

How actions are generated from PECS needs.

<table>
<thead>
<tr>
<th>Need</th>
<th>Generate wealth</th>
<th>Sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of the need</td>
<td>The need to generate wealth: a proxy for any activity which requires wealth to satisfy it.</td>
<td>The need to sleep.</td>
</tr>
<tr>
<td>PECS intensity value, I, at time, t</td>
<td>Based on the time of day, $T$, the agent's current level of wealth, $W$, and the agent's personal preference for generating wealth, $P$, $I_t = f(W_t, P, T_t)$</td>
<td>Based on the time of day, $T$, the agent's current level of sleep, $S$, and the amount of sleep the agent needs each day, $A$, $I_t = f(S, A, T_t)$</td>
</tr>
<tr>
<td>Resulting agent behaviour</td>
<td>Try to obtain wealth, either through employment (if available) or through burgling (if not employment is available).</td>
<td>Go home as soon as possible to sleep.</td>
</tr>
</tbody>
</table>

**Fig. 1.** Agent needs and behaviour. The agents’ needs and how their intensities determine the behaviour of the agents, adapted from Schmidt (2000).

**Fig. 2.** Need intensities over time. How the time of day affects an agents need to generate wealth or sleep where Time $t = 0$ is defined as approximately 7am.
companies) are not investigated in this iteration of the model development. This will be a future research stream.

4.3. The model environment

The agents populate an artificial environment that is designed to reflect many of the urban features found in modern cities. There is a commercial area in the centre of the environment and this is surrounded by residential properties. This environmental layout does not represent an entire city, rather a small "micro" town centre. This type of pattern is repeated throughout modern cities, where commercial and residential areas are fairly mixed (in Leeds, for example, only 30% of employment is found in the city centre (Unsworth & Stillwell, 2004). Fig. 3 illustrates the layout of the environment. Although simple and hypothetical, the model environment was designed to allow comparisons with real urban configurations. A central business district represents the centre of employment for city residents, which is a feature found in many modern cities. In this respect the model imitates part of the concentric ring model (Burgess, 1925), although later experiments incorporate communities that are distributed in a less orderly fashion. This corresponds better to British historical housing developments which are often initiated by local councils who build wherever they own land (Baldwin & Bottoms, 1976) and illustrates that the model is highly flexible because the environment can be adapted to reflect the type of city under examination.

The environment is constructed on a grid measuring 41 x 31 cells. Some squares are "empty" and play no part in the model because agents do not move diagonally, only horizontally or vertically. There are three types of cell in the environment: the commercial district, roads, and residential properties. Agents use roads to navigate the environment, always taking the shortest path from their current location to their destination. Each cell in the commercial district represents a single office that provides employment for an unlimited number of agents. The residential properties house the agents (maximum of one agent per house) and also act as burglary targets (one cell represents one house). The houses have two defining characteristics: security and attractiveness. The security variable is a measure of the level of security of a property, encompassing both physical security and security utilisation; attractiveness is a measure of the wealth of the property. As noted previously, agents remember properties that they utilise; attractiveness is a measure of the wealth of the property increasing along with smaller increases in the surrounding properties. The victimised property and those adjacent then remain at a higher risk for several days following the burglary. This "near repeat" phenomena has been found to exist in the criminology literature (Townsend, Homel, & Chaseling, 2003) and by police force managers (Johnson, 2007). In the UK city of Leeds, there are several proactive crime prevention initiatives which target properties in close proximity to a recent burglary. For this reason, the security levels of the victim and the surrounding properties are increased along with attractiveness after a burglary is committed. These levels of attractiveness and security gradually degrade as the residents become complacent of the risks (from anecdotal evidence); if no further burglaries are committed, the security returns to base levels.

Upon initialisation of the simulation, each agent is randomly assigned a home address and workplace (which will be within the commercial district). The experiments presented in this paper place the potential burglar agents both in low-income areas and also distributed evenly throughout the environment. The agents use roads to travel between different addresses and can traverse one square per model iteration. Agents always take the shortest (optimal) route between their origin and destination. Further research will focus on a more accurate representation of travel through the area. Time is measured in the simulation through model iterations; one iteration is classified as 3 min. This means that it will take agents between 10 and 60 min to travel to work depending on their origin. There are therefore 20 iterations per hour and 240 in a day.

4.4. Modelling offender behaviour

There are two main branches of research into understanding how potential burglars behave, their motivations and their responses to environmental cues. These can be broadly classified as qualitative using interview data and quantitative using large data sets and statistical models to establish trends and patterns of potential burglary behaviour. Although different in methodology, these studies draw very similar conclusions; these will be used in the design and implementation of behaviour in the agents. Table 2 outlines findings from studies in the criminology literature and how these will be incorporated into the model to provide a sound theoretical foundation.

The burglary process works as follows:

1. The burglar agent decides that they must commit a burglary to generate wealth because they do not have any temporary employment.
2. The agent chooses a house to visit from the list of all those they know about (the houses that are stored in their cognitive map). A "roulette wheel selection" process is used so that each house has a probability of being chosen based on its attractiveness.
3. The agent travels directly to the chosen house using the shortest path. As they pass houses they examine their security to determine whether or not they are suitable for burglary.
4. If the agent reaches their chosen house and has not found a suitable burglary target they choose another house from their cognitive map (using the same roulette-wheel procedure) and begin the process again.
Two criteria determine whether a target property will be burgled: occupancy and security. An occupied property will never be burgled and the potential burglar is less likely to burglar a secure property particularly if there are possible targets with lower levels of security. These elements are consistent with many findings, including Cromwell, Olson, and Avary (1991), Wright and Decker (1996). There are no “unsuccessful” burglars at present, the burglar either commits a burglary or does not, based solely on the security of the target property and whether or not it is occupied. Therefore if the security of all houses is increased there will be fewer burglaries and the agents’ levels of wealth will steadily decrease. In this sense, therefore, the number of burglaries in the model is essentially fixed and the model is only able to change changing patterns of burglary, not overall rates. Allowing for unsuccessful burglaries and other agent behaviours (such as choosing not to burglar at all) will form interesting avenues for future research.

Determining the amount of money which can be generated from a burglary is non-trivial. Snook (2004) found that the average amount was $900, but the range was $0–$12,950 and the value depended on the distance travelled. For simplicity, agents in the model are assumed to be capable of burgling any property in the area and the agents’ levels of wealth will steadily decrease. In this sense, therefore, the number of burglaries in the model is essentially fixed and the model is only able to change changing patterns of burglary, not overall rates. Allowing for unsuccessful burglaries and other agent behaviours (such as choosing not to burglar at all) will form interesting avenues for future research.

The behaviour that is being examined in this model is a simplification of offending behaviour; for example, it is obviously too simplistic to state that an individual will automatically turn to burglary if they have no money. However, there is a scientific basis for not oversimplifying a model. Schmidt (2000) for example, notes that a model does not need to replicate reality, if it did then it would cease to be a model. Furthermore, Elffers and van Baal (2008, chap. 2) note that crime models can become overcomplicated, making it difficult to understand and experiment with the rules that underpin the model. For this model, the inclusion of different types of crime and a complex cognitive framework which gives agents stronger control over how to behave when the drive to generate wealth (rather than always turning to burglary) are seen as unnecessary at this stage. The factors which have been chosen are deemed, from the criminological literature, the most important to the residential burglary system, not a general model of crime and offending.

5. Model experimentation

The model will be applied to testing out crime theories and the effectiveness of varying crime reduction strategies. As highlighted earlier, the environment is artificial, but designed using the UK city of Leeds as its template.

The following experiments will be performed:

- **Control experiment**: The default parameters of security and attractiveness of properties will be used to explore the basic behaviour of the model. The values of the defaults were chosen to coincide with the drive intensity functions which determine how potential offenders should behave (see Sections 3 and 4). The values were calibrated to allow, on average, an offender to commit one burglary per day. This coincides with the expected return of a single burglary which is the equivalent of a single day’s work (discussed in Section 4.4).

- **Different types of community**: To simulate the presence of different types of community, such as a deprived area, an affluent area, and an area occupied predominantly by students, the environment will be adapted by modifying the security and attractiveness of property values.

- **Target hardening strategies**: The model is used to test the effectiveness of the crime reduction strategy of target hardening. Target hardening is an intervention scheme whereby government agencies offer additional security protection in the form of physical hardware or verbal/written advice to residents. In the model, target hardening is simulated by increasing the security

Table 2

<table>
<thead>
<tr>
<th>Behaviour/motive</th>
<th>Implementation in model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for money is the primary reason for burglary (Bennett &amp; Wright, 1984; Bernasco &amp; Nieuwbeerta, 2005; Nee &amp; Meenaghan, 2006; Bennett &amp; Wright, 1996) and usually to buy drugs (Cromwell et al., 1991; Hearnden &amp; Magill, 2003; Scarr, 1973).</td>
<td>Agents in the model burglar to satisfy the desire for wealth. Drug addiction can be represented by increasing the personal preference for generating wealth. Agents with these characteristics will quickly become desperate to generate wealth as if they had a drug addiction to satisfy.</td>
</tr>
<tr>
<td>Most burglars will return to previously burgled properties, usually because they know what goods are available and how to enter the property (Hearnden &amp; Magill, 2003; Wright &amp; Decker, 1996).</td>
<td>During the journey to their chosen target, agents examine properties which they pass and will commit a burglary if the target is deemed suitable.</td>
</tr>
<tr>
<td>If an agent has no money or their current strategy is not working, they may choose to travel to a different area to burglar (Cromwell &amp; Olson, 2005).</td>
<td>Potential burglars choose to travel to the most attractive property they are aware of.</td>
</tr>
<tr>
<td>Potential burglars choose to travel to the most attractive property they are aware of.</td>
<td>When occupants are at home a burglar agent will not victimise the property. The burglars agents always “know” if a person is at home; they do not use environmental cues as they might do in the real world.</td>
</tr>
<tr>
<td>The agent’s cognitive map is built up from their routine activities (Cromwell &amp; Olson, 2005; Wright &amp; Decker, 1996).</td>
<td>Properties close to a burglar agent’s home are more likely to become part of the agent’s cognitive map and are therefore a higher burglary risk.</td>
</tr>
<tr>
<td>Suitable targets are often found by passing them on their routine activities (Cromwell &amp; Olson, 2005; Wright &amp; Decker, 1996)</td>
<td>The agent’s cognitive map is built up from their routine activities and a target is chosen from these known properties.</td>
</tr>
</tbody>
</table>
of a targeted property up to levels that match the most secure properties in the environment. Two different target hardening strategies are tested. The first is commonly used by local government agencies. The strategy involves targeting the most vulnerable people, which includes new and repeat burglary victims, the elderly, single parents, those renting private houses and people who have recently moved into new properties (see Byron (2003) for a practical example). In the model, vulnerable properties are identified by those that have the highest number of burglaries. The second strategy is an alternative method which is not commonly used in practice. Here, all the properties in a community that has been identified as a high-crime area simultaneously undergo target hardening. The aim of the experiments is to establish which strategy is the most effective at removing a crime hotspot.

- **Different routine activity patterns**: The addresses of potential burglars are altered to change their routine activity patterns. The model is used to generate new crime patterns. This allows us to examine routine activities theory and the effect that different offender daily patterns will have on crime rates.

6. Results

6.1. The control experiment

The aim of the control experiment is to check that the model is robust and that it produces sensible results, a process often termed “verification” (Gilbert & Troitzsch, 1999). Default values for security and attractiveness of properties are used throughout the environment and all agents (potential burglars and non-burglars) are assigned randomly to houses. The model is run until it reaches a dynamic equilibrium which refers to the state when aggregate crime patterns are stable although individual crimes are still occurring and, therefore, small local variations are present (van Baal, 2004). For this model, we define dynamic equilibrium as being reached when both the number of crimes committed each day and the mean centre (average) of all burglary locations does not change and we will show that 50 days is sufficient for the model to reach dynamic equilibrium. Fig. 4 illustrates the number of burglaries committed at different intervals. The model was executed 100 times which allows the robustness and sensitivity to initial starting conditions to be assessed. It was noted earlier that the size of the population of burglars varies slightly for each run, but the number of burglaries committed at different intervals remains fairly consistent suggesting that, with respect to the number of burglaries committed, that the model has reached equilibrium. Further evidence for equilibrium is presented by the spatial distribution of crimes illustrated in Fig. 5; the model reaches dynamic equilibrium in this time because the mean centre of the burglaries does not change. Fig. 5 suggests that, in fact, equilibrium might be reached earlier than day 50, between days 20 and 30 for example. However, later experiments will require time to allow the system to adapt to changes made during the course of a simulation so to ensure that all experiments reach equilibrium all simulations presented will run for 50 days.

Fig. 6 depicts the burglary rates at the end of a typical simulation run. Due to the probabilistic nature of the model, burglary patterns vary between runs. However, burglary levels are routinely highest in the areas closest to the commercial area. These findings are consistent with the principles of crime pattern theory. Brantingham and Brantingham (1993, page 18) note that crime clusters at high activity nodes, along major paths and along edges, where edges represent the boundary between areas that are noticeably
different such as the commercial and residential areas in our hypothetical environment.

Fig. 7 further illustrates that most crimes occur near the centre of the environment. The Pearson correlation coefficient was calculated between the number of burglaries a property received and its distance from the nearest commercial patch and resulted in a value of 0.38. This implies that as the distance from the commercial district increases the number of crimes committed decreases.

6.2. Different “Types” of community

Experimentation in Section 6.1 served to show that the model is stable, producing expected results under default conditions. The next stage is to increase the realism of the model by introducing environmental factors. This added realism is achieved by altering the attractiveness and security of each property to create different communities. Three different areas have been chosen: an affluent area, a deprived area and a student area. These sociotypes have been chosen to reflect the different crime patterns that are prevalent in each of these areas. Offenders travel different distances depending on the affluence of the target (Snook, 2004) and the community type from which an offender originates influences where they are likely to burgle. Shepherd (2006) also found evidence that burglary patterns depended on the type of community. The author discovered that offenders would travel considerable distances to burgle affluent areas, whereas burglaries in deprived areas were often committed by local residents travelling short distances. In addition, students were victimised by residents of nearby deprived areas but not from within student communities (Shepherd, 2006). The relative variable values associated with each area are shown in Table 3. These values have been chosen not on the basis of empirical evidence (determining how much more attractive a “student” area is compared to a “normal” area, for example, is non-trivial) but because they are different enough to sufficiently influence the burglars’ behaviour and lead to the creation of crime “hotspots”. The new areas created, therefore, are not designed to fully represent “student” or “deprived” communities but provide a method of experimenting with how the burglars respond to changes in their environment.

Using these different types of area it is possible to investigate how high-crime areas (often called hotspots) arise. Four different layouts for the cityscape were used to ensure that hotspots do not arise as a result of the arbitrary layout of the environment. Fig. 8 illustrates these environments and the burglary rates produced by day 50. Regardless of the layout of the environment or the initial starting positions of the agents, the student areas suffer the highest victimisation rates. This is still evident when there are multiple student areas as illustrated in environment 4.

Further evidence can be supplied through hotspot detection. The nearest neighbour hierarchical spatial clustering algorithm (NNH) is commonly used to search for clusters of points based on their spatial proximity. The CrimeStat application (Levine, 2006) was used on the case-study data. Fig. 9 illustrates the hotspots found by the algorithm when analysing the crimes committed near the end of the simulation (days 40–50). The last ten days are used here because this is the time at which the simulation is judged to have reached equilibrium. The results illustrate that, regardless of the physical layout of the environment, the student areas still suffer the highest levels of burglary victimisation. This is consistent with the criminology literature (Robinson & Robinson, 1997; Tilley, Pease, Hough, & Brown, 1999) and data from the city of Leeds. For example, in Leeds burglary hotspots are highly correlated with areas that house large numbers of students during term-time. In August, when the majority of the student population live outside the city, the burglary clusters move to the poorer areas to the east and west of the city centre.

6.3. Crime reduction: Target hardening strategies

As illustrated in Section 6.2, the layout of the environment does not appear to influence the burglary hotspots found in student areas. Target hardening was therefore applied to environment 1. The strategies were implemented on day 20 and, as with other experiments, the simulation was run to day 50. The area chosen for the block-targeting method covered 50% of the student community (to allow comparisons between the hardened and non-hardened sections). This consisted of 46 properties. In order to test both strategies fairly, it is essential that they increase the overall security of the environment by the same amount. If this is not done...
then the total number of crimes committed might differ between experiments simply because the overall security changes so results will not be comparable across different experiments. Equations in Appendix A demonstrate that the victim-targeting method will increase the security of approximately one house every 2 days to be comparable with the block-targeting method.

Fig. 10 illustrates the results of the victim targeting strategy. The advantage of ABM to view a dynamic history of the model (Axtell, 2000) rather than a single, final equilibrium is utilised here and crime hotspots and burglary rates are illustrated at different points in the simulation. By observing the crime patterns at different points during the simulation we can gain an insight into how crime hotspots arise. The results suggest that the strategy is ineffective at removing the crime hotspot found around the student area. A crime hotspot is established early in the simulation and remains fairly constant throughout.

Fig. 11 illustrates the results of the block targeting strategy. It appears that crimes are displaced south towards the remainder of the student area with no target hardening. Although the hotspot produced between days 40 and 50 still covers the target hardened area, only three crimes were committed in the area during that time period. This suggests that if the NNH algorithm was configured differently the hotspot would not cover the north area at all. Interestingly, towards the end of the simulation a new hotspot has started to develop close to the city centre as crimes are displaced away from the student area. This provides further evidence that the hotspot around the student area is less significant.

The patterns produced by the two experiments are very different. When individual houses are targeted, offenders are still attracted to the area because many appealing properties still exist, even if some are now less appealing due to the target-hardening initiative. This suggests that targeting single properties in isolation is unlikely to tackle burglary hotspots because many insecure properties remain in the area. This finding is consistent with the expectations of crime reduction practitioners. Shepherd (2006) notes that the administrators of the Burglary Reduction In Leeds (BRIL) scheme (Safer Leeds, 2007) believe that targeting blocks of properties rather than individuals might have an effect greater than the sum of the parts.

6.4. Different routine activity patterns

The final experiment increases the realism of the model further by investigating the effect that changing the addresses of the pe-
Fig. 10. Individual target-hardening results. Results of the individual victim target-hardening initiative: hotspots produced by the NNH algorithm and burglary rates.

Fig. 11. Block target-hardening results. Results of the entire community target-hardening initiative: hotspots produced by the NNH algorithm and burglary rates.

Fig. 12. Burglaries originating from “Constrained by Circumstances” communities. Graph illustrating the number of crimes committed in different community types which have originated from “constrained by circumstances” communities using 2006/07 crime data and the ONS Output Area Classification.
potential burglars has on crime rates. In the experiment it is hypothesized that most potential burglars live in the most deprived areas. This notion follows that of the literature: numerous studies have made reference to the link between crime and deprivation (Baldwin & Bottoms, 1976; Bowers & Hirschfield, 1999; Brantingham & Brantingham, 1993; Hesseling, 1992; Sampson, Raudenbush, & Earls, 1997; Shover, 1991; Wilkström, 1991). There is also supportive data from the city of Leeds. Using the Office of National Statistics Output Area Classification (Vickers & Rees, 2006) and over 700 pairs of the addresses of convicted burglars and their victims it was possible to estimate which "type" of community most burglars originated from. Fig. 12 illustrates that the most deprived communities ("constrained by circumstances") export the most crimes. We can hypothesise, therefore, that most burglars live in constrained by circumstances communities.

The locations of the potential burglars in the model were altered from living in randomly chosen patches to the poorest area. This represents a shift in the model towards the inclusion of real geographies and people. This change will obviously also impact on the routine activity patterns of potential burglars. We would therefore expect to observe higher burglary rates in the poorest neighbourhood and on the routes into the commercial district. Environment 2 (illustrated by Fig. 13) was chosen because in this environment the student area and the deprived area are a large distance apart. Therefore if the burglary hotspot still covers the student area we can conclude that the daily activities of the offenders does not influence the location of burglary hotspots in the model.

Fig. 13. Environment 2.

Fig. 14. Routine activities experiment results. Results of routine activities experiment: hotspots produced by the NNH algorithm and burglary rates.

The aim of this paper has been to demonstrate the strengths, flexibility and applicability of an individual-based model combined with a behavioural model (the PECS framework) for simulating residential burglary. Within the scope of modelling crime theory, there are few published examples of this type of work; the research presented here represents initial modelling attempts to capture the complex micro-level dynamics of this system using an advanced behavioural model. Alternative approaches to modelling crime were outlined, including some existing agent-based models of crime. However there are many factors that are absent in other approaches which this research is able to account for.

Incorporating a detailed behavioural framework into an individual-level model is a relatively new approach in criminology. The PECS framework (Schmidt, 2000; Urban, 2000, chap. 6) was chosen because it does not require rational decision making as an assumption, a drawback of the BDI approach (Schmidt, 2000), and can (theoretically) be extended to model the entire spectrum of human behaviour. PECS uses the concept of intensity functions to determine, in any given situation, which drive is the strongest and how the agent will behave. Two drives were used in this model: the need to generate wealth and the need to sleep. Although the range of drives is limited, they are adequate to loosely represent the daily behavioural patterns of people employed in British or American cities. In addition, the intensity functions can be enhanced to amalgamate different types of behaviour. For example, drug addiction can be simulated by increasing the desire to generate wealth in the model a burglar with a drug addiction will therefore be forced to commit more risky burglaries to satisfy their greater needs.

Findings from both qualitative and quantitative studies (outlined in Table 1) were utilised to ensure that the behaviour of...
offenders in the model reflected findings from the real world. One of the most interesting features of the model are the cognitive spaces which are individual to each agent and are built up dynamically during the simulation. Potential offenders do not have global knowledge of their environment and they must choose to victimise a property that they know about already, finding new properties as they travel around the environment (whether on legitimate business or not). This feature reflects modern thinking in criminology and has yet to be included in this type of model.

Four experiments were designed to test the validity of the model and then experiment further with it. In particular, two target hardening approaches were tested. The first, which is an approach commonly used in practice, targeted single properties that were deemed a high burglary risk and the second targeted an entire block of properties. Cluster analysis confirmed that targeting individual properties in isolation was insufficient at removing the hotspot as offenders in the model were simply able to burgle nearby houses that have not undergone target hardening. Targeting an entire block, however, successfully removed the hotspot because the entire area became unattractive to burglars. This demonstrates that the model, through matching empirical findings, is both robust and able to simulate the important processes and trends within the system.

The final experiment examined what happened if the home locations of potential burglars were altered. The effect was that the agents’ routine activities no longer took them through the student area. However, as the simulation progressed a hotspot nevertheless formed around the student area. This suggests, for criminoologists, that although routine activities are important we should not discount the pull of highly attractive areas which might drive offenders away from their routinely travelled routes. Obviously a greater investigation is required before making any firm conclusions, but the experiment nevertheless demonstrates the utility of using even simple types of agent-based models in criminology.

It should be noted that the total number of crimes in the environment remains unchanged. In other words, there is spatial crime displacement but no other types of displacement such as a change in modus operandi (MO), crime type (for example the offender could move from burglary to drug dealing) or indeed the decision to stop burgling altogether. Furthermore, the vast complexity of human behaviour and the urban environment are extremely difficult (or even impossible) to capture in a computer model, thus models such as these will never be able to account for everything that can affect the real system. These are not necessarily limits of a model, but a drawback of ABM in general. We do not subscribe to the notion that this renders the individual-level approach useless; rather we recognise the drawbacks of the approach and consider these when making conclusions regarding the applicability of the results to the real world.

8. Future work

One of the major benefits of the ABM approach is its flexibility. Incorporating additional needs, such as the need to socialise, will provide the agents with a greater range of behaviour and allow us to implement different types of citizen such as students, unemployed people, and family members. These different types of people could also influence the behaviour of the potential burglars, by acting as capable guardians for example. Anchor points could also be included (such as friends’ houses or the addresses of drug dealers) which would generate interesting cognitive environments. The model at this stage does not consider the possibility for citizens to become opportunist burglars, which could represent a significant portion of burglaries committed. Consideration of how the model framework could be modified to incorporate these types of behaviours and more heterogeneous agents is ongoing.

There are also a number of ways in which the environment itself could be enhanced. These range from including ideas regarding collective efficacy (Sampon et al., 1997) or “broken windows” theory (Wilson et al., 1982) to incorporating real GIS data similar to that of Groff (2007a, 2007b). In addition, the availability of vehicle transport (such as cars or public buses) could be included by adding different layers to the environment. There are a number of challenges associated with incorporating a more realistic environment, however. Initially, it is necessary to obtain low-level, accurate physical and social environmental data to act as inputs to the model. Furthermore, accurate individual-level crime data is necessary to use in evaluating the accuracy of the model (comparing simulation results to real data). Assuming these data are available, the next challenge is to adapt the model to function in a much more complex virtual environment. Routines to allow agents to travel on the transport system must be implemented and it is likely that the complexity of the agents must be increased to allow them to perceive their new environment correctly. Regardless of the difficulties, including a more realistic urban backcloth is an ultimate goal of this research (with the proviso that the simple model is fully understood first). This will allow for crime predictions in the real world which could ultimately influence policy.

This type of model has obvious benefits and has the potential to form an integral part of a tool for policy makers to test the impact of varying scenarios. The next stage is to translate the simple model into a more advanced framework and to incorporate a real environment.

Acknowledgements

We would like to thank Safer Leeds, the local Crime and Disorder Reduction Partnership, for providing invaluable crime data and their experiences in crime reduction.

Appendix A

The simulation is run for 50 days, so the overall increase in security produced by the block-targeting method (which covers 46 houses and is instigated on day 20, thus running for 30 days) can be calculated as:

\[ 46 \times 30 = 4 \times 5520 \]

where 4 is an arbitrary number of units which represents a 150% increase in security. Method 1 is also implemented on day 20, and will target the x most vulnerable properties every day. The strategy stops at day 40 so that the simulation is allowed 10 days to reach equilibrium. Therefore, to ensure that both methods lead to the same overall increase in security, the number of houses targeted each day by method 1, x, between days 20 and 40 is:

\[ x \left( \sum_{i=0}^{20} 4(i+1) \right) = 840x \]

and this security increase is applied for a further 10 days (between days 40 and 50), so:

\[ 840x \times 10 = 5520 \]

\[ x = 6.571 \]

so each day (between days 20 and 40) there is a 66% chance that a house will be targeted which will, on average, increase the overall security of the environment by the same amount as the block targeting strategy.
Appendix D

Regional Review Paper

The following paper (Malleson, 2008), authored by the writer of this thesis, has been published in the journal *The Yorkshire and Humber Regional Review*. 
Simulating burglary: what does the future hold for Leeds?

Nick Malleson, an ESRC-funded postgraduate student, discusses how computers can be used to simulate the behaviour of residential burglars in Leeds. The work is being conducted closely with crime-reduction practitioners at Safer Leeds and it is intended to help decision makers estimate future burglary rates as the city develops.

Despite popular belief among tabloid readers that crime has increased ‘a lot’ since 1995, British Crime Survey figures indicate that national crime levels have been consistently falling (Walker et al. 2006). Residential burglary, in particular, has fallen by more than 50 per cent nationally.

The picture in Leeds, however, is less optimistic as burglary levels are found to be consistently the highest when compared to other local authorities (Shepherd et al., 2004). In March 2008, the insurance company Endsleigh found Leeds to be the fifth-highest burglary risk, nearly 50 per cent above the national average.

There are a number of possible explanations for this. For example, Leeds has a disproportionately large student population which suffers extremely high rates of residential burglary. This is partly due to the abundance of expensive goods in student accommodation and partly due to a lack of security precautions on the parts of the students or landlords. In addition, Leeds houses some of the most deprived communities in the country and deprivation is regularly linked to high crime rates.

There are many crime-reduction measures which have been shown to reduce burglary and several of these are being successfully implemented by Safer Leeds, the local crime and disorder reduction partnership. However, in a rapidly developing city such as Leeds it is very difficult to predict where resources should be targeted in the future and what effect new environmental or crime-reduction initiatives will have on prospective populations. Crime simulation is a new and under-developed approach which might be able to offer insights into future crime rates.

Burglary is often an opportunistic crime which involves little conscious searching on the part of the offender. Most targets are known to the burglar beforehand because they have passed them on their routine activities or committed a successful burglary in the locality on a previous occasion. Therefore, regardless of individual security precautions, houses which are passed regularly by large numbers of potential burglars will be a higher risk than others, but this artefact can only be investigated by looking at the local urban configuration.

Accounting for these low-level influences is extremely difficult with models which aggregate human populations or houses up to a high level as is commonly done in crime models. Recently, however, a shift in the social sciences towards models which work at the level of the individual has sparked interest in crime models which also work at the individual level.

The individual-level approach (often called ‘agent-based modelling’) has been heralded as the “third way of doing science” (Axelrod, 1997) and holds promise as being the most fruitful method of simulating systems in which aggregate patterns (such as city-wide crime rates) emerge from the behaviour of individuals.

A burglary simulation

An agent-based model of burglary is currently under development which attempts to simulate the behaviour of individual burglars as they travel around...
the city. Using criminology theories and the experiences of Safer Leeds crime reduction practitioners to create realistic virtual burglars, the work aims to predict future burglary rates which might arise as a result of planned urban developments.

High-quality Ordnance Survey Mastermap data areas are being used (Figure 1) so the virtual environment will be highly realistic and include many of the low-level factors which are difficult to account for in traditional crime models. Although the model is being developed for Leeds, some of the benefits of the agent-based approach are flexibility and scalability. It is possible to apply the same simulation to different cities or different scales without making any significant changes to the underlying model.

The EASEL Regeneration Scheme

The main aim of the project is to be able to predict the affect that urban developments will have on burglary rates in the city before they are implemented. The East and South-East Leeds (EASEL) regeneration scheme is one of the largest in Europe and, over the coming years, will dramatically re-shape the physical environment and the demographics of Leeds. Figure 2 illustrates the large area covered by the planned developments.

The area covered by EASEL is one of the most deprived in the city. The regeneration scheme aims to address this imbalance by rebuilding many of the existing properties in the area and replacing them with a mixture of different tenure types aimed to attract a range of different tenants from varying socioeconomic backgrounds.

The new houses will be built alongside improved transport links, new green spaces, better educational and employment opportunities, and improved access to local services. The scheme will undoubtedly have a significant impact both on the physical urban environment and the demographics of the area.

However, little work has been conducted into how crime rates will be affected by the regeneration. This project will hope to predict how these demographic and environmental changes will influence burglary patterns.

Preliminary results

Although the model is still at a preliminary stage of development, it shows promise as an accurate predictor of crime rates. Figure 3 compares clusters of burglaries which were produced by running the model and those found by analysing real crime data. Although the clusters are not identical and there are some which are found in the real data but not reproduced by the model, as a preliminary result this is encouraging. It suggests that, after some improvements, the simulation will be able to accurately model the behaviour of burglars and successfully predict future crime rates.

Another advantage of this type of model is the ability to observe, in fine detail, how individuals in the model travel around the environment. It is very difficult to collect highly detailed information about what a real potential offender does during the day, but these types of models can generate unlimited amounts of useful data.

Figure 4 illustrates how it is possible to visualise the movements of simulated individuals through time and space. In the image, movement through time is represented on the vertical axis and movement through space is represented on the horizontal axes. The travel patterns of two different individuals are visualised...
as they travel east across the city on a particular activity (for example they might be travelling towards an area in which they will attempt a burglary). At one point, the two individuals meet in the same place at the same time. This can be useful to see if individuals in a simulation are working together (which will be a future research aim for the project). Also, using these types of analyses it is possible to observe which streets are regularly travelled by potential burglars who might be on legitimate routine activities. For example, part of the EASEL regeneration scheme will include building new transport links. Crime simulation allows researchers to examine the precise routes which simulated burglars are using and could offer an insight into how potential burglars are using new transport routes. This might suggest which houses will be a higher or lower burglary risk because of the improved transportation.

**Conclusion**

Predicting crime rates is an extremely challenging venture. Although crime reduction partnerships like Safer Leeds are implementing highly successful crime-reduction initiatives, it can be difficult to predict where to plan for future initiatives. Also, in a modern, dynamic city such as Leeds, urban developments will have a significant impact on crime patterns and, again, predicting these effects can be difficult.

Crime simulation at the level of the individual person or house is a relatively new field which holds promise as a method of predicting the effect that urban developments will have on crime rates, allowing policy makers to experiment with scenarios before their implementation.

**References**


References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Taylor, R. (1997). Social order and disorder of street blocks and neighborhoods: ecology, microe-
cology, and the systemic model of social disorganization. *Journal of Research in Crime and

www.spatial.redlands.edu/agentanalyst/.

from the Crime Reduction Programme. Policing and Reducing Crime Unit Crime Reduction

based simulation. *Journal of Artificial Societies and Social Simulation*, 7(1):online.


Canada-Line.aspx.


and Wales, the United State and the Netherlands: A cross-national comparative test of routine


tigation into human pedestrian behaviour in the built environment. *Environment and Planning

uk/.


Vag, A. (2004). First generation multi-agent models and their upgrades. *Interdisciplinary De-


REFERENCES


