UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE INFORMÁTICA

Departamento de Ingeniería del Software e Inteligencia Artificial

TESIS DOCTORAL

Towards a data-driven approach for Agent-Based Modelling: Simulating Spanish postmodernisation

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Samer Hassan Collado

Directores

Juan Pavón Mestras
Millán Arroyo Menéndez

Madrid, 2017

UNIVERSIDAD COMPLUTENSE DE MADRID

Towards a Data-driven Approach for Agent-Based Modelling: Simulating Spanish Postmodernisation

by

Samer Hassan Collado

Directed by

Juan Pavón Mestras
Millán Arroyo Menéndez

A thesis submitted in candidature for the degree of PhD
Facultad de Informática
Departamento de Ingeniería del Software e Inteligencia Artificial

September 2009

"In many systems, the situation is such that under some conditions chaotic events take place. That means that given a particular starting point it is impossible to predict outcomes. This is true even in some quite simple systems, but the more complex a system, the more likely it is to become chaotic. It has always been assumed that anything as complicated as human society would quickly become chaotic and, therefore, unpredictable. What I have done, however, is to show that, in studying human society, it is possible to choose a starting point and to make appropriate assumption that will suppress the chaos, and will make it possible to predict the future, not in full detail, of course, but in broad sweeps; not with certainty, but with calculable probabilities."

Hari Seldon

Isaac Asimov, *Prelude to Foundation*, 1988

Abstract

In the lasts decades, computer simulation in general, and agent based modelling (ABM) in particular, has become one of the mainstream modelling techniques in many scientific fields, especially in Social Sciences such as Sociology or Economics. Social simulation allows the study of the complexity inherent to social phenomena and it is attracting multidisciplinary research teams in order to manage this complexity.

There are different methodologies for ABM that, after compiling experience in processes, methods and tools, attempt to provide a systematic way to tackle new problems. Both the Multi-Agent Systems field and ABM have tried to provide robust methodologies to guide researchers in the modelling process.

However, there is an important epistemological distinction among agent-based models that these methodologies do not consider. Models can be classified depending on their research aim, and this classification can have methodological implications. Sometimes researchers seek a generic model to explain a social phenomena from a high degree of abstraction, and one that is simple enough to be used as an illustration of a specific theory or hypothesis. On the other hand, researchers may prefer to focus on the expressiveness of the model, together with the empirical descriptiveness of a specific case study. The first case corresponds to Theoretical Research, while the second one would be Data-driven Research.

Nowadays, most of the models are conceived from the theoretical approach, and thus methodologies are frequently biased towards them. However, without disregarding the role of theory, models can also seek expressiveness. In order to do that, they may have needs that are not met in general methodologies. For instance, issues such as the empirical initialisation, the limitations of data collection, the throughout empirical validation or the role of data in the design are not usually considered in those methodologies. Thus, there is a lack of a complete ABM methodology that, assuming data-driven research has a different approach and aims, provides a specific flow of data-driven model development. Such methodology should consider the key role of empirical data throughout all the modelling stages. This lack has caused most data-driven models to be constructed without a common frame.

This work attempts to fill this gap and build a complete methodology to guide data-driven agent-based modelling. Therefore, it can be advocated that when there are available data from the observation of the real phenomenon, the modelling and simulation process involves additional stages. This methodology attempts to guide the injection of empirical data into the simulation, bringing them closer to the real phenomenon under study, while acknowledging the important role of theory in the whole process. Therefore, the approach is complemented with a systematic method for the exploration of the

model space in order to achieve comprehensible but descriptive models. Such a method was coined 'Deepening KISS', as it is exposed in the methodological chapter.

This methodology is supported technically by the specification and implementation of a social agent framework. Such framework is structured in modules which can be enabled at will in order to facilitate the exploration of the model space and its incremental construction, both in the frame of the data-driven approach. Instead of attempting to build a general-purpose framework, this agent framework focuses on a family of problems which can be best tackled within it.

Moreover, an in-depth case study was developed to test and validate the application of the methodology and proposed framework. This case study addresses the complex issue of social values evolution, together with the friendship emergence and the demographic dynamics involved.

The construction of this agent-based model, coined Mentat, can be summarised in a series of key milestones. The proposed data-driven methodology is applied intensively through the course of its development. The modelling process has been realised as (a) bottom-up and (b) top-down. (a) is represented by the social network arising from the micro behaviour and friendship dynamics. (b) relies on the elaborated demographic model. The conceptualisation and specification of (a) and (b) has been justified theoretically in order to support its development. They have been implemented within the modular agent framework, designed in incremental layers. Mentat features are structured in modules which can be enabled or disabled in order to explore the model space following the stages defined in the methodology. The model is validated from a quantitative macro perspective (empirical validation), from a qualitative micro perspective (social dynamics matching the theoretical assumptions) and from a theoretical perspective (discussing its sociological consistency). Different techniques of Artificial Intelligence are applied and combined in the model, testing the framework adaptability and their use for social simulation. Mentat serves as a case study of the methodology and framework, but it also provides some sociological insight of the problem under study, by giving new support to specific theories. The ABM specifically stresses the key significance of demographic dynamics in the case study: the evolution of social values in Spain during the end of 20th Century. This implies that intergenerational changes are considerably more important than intragenerational ones in this Spanish context, and supports Inglehart's theories of values evolution.

Keywords: agent-based model, data-driven modelling, demography, friendship, fuzzy logic, microsimulation, social values, social simulation, surveys.

Resumen

Durante las últimas décadas, la simulación computacional y el modelado basado en agentes (agent-based modelling, ABM) en particular, se han convertido en una de las principales tecnologías de modelado en múltiples campos científicos, especialmente en ciencias sociales como la sociología y la economía. La simulación social permite el estudio de la complejidad propia de los fenómenos sociales, y está atrayendo equipos de investigación multidisciplinares para poder manejar dicha complejidad.

Existen distintas metodologías para ABM que, después de compilar suficiente experiencia en procesos, métodos y herramientas, ofrecen una forma sistemática para estudiar problemas nuevos. Tanto el campo de Sistemas Multi-Agente como el de ABM han tratado de ofrecer metodologías robustas que puedan guiar a los investigadores en el proceso de modelado.

Sin embargo, hay una distinción epistemológica importante entre los modelos basados en agentes que tiene implicaciones metodológicas y que dichas metodologías no consideran. Los modelos pueden ser clasificados en función de su objetivo de investigación. En ocasiones, el investigador persigue un modelo genérico para explicar el fenómeno social, desde un alto grado de abstracción y de forma suficientemente simplificada para ilustrar fácilmente una teoría o hipótesis concreta. Por otro lado, el investigador puede preferir centrarse en la expresividad del modelo, en la que se haga una extensa descripción empírica de un caso de estudio concreto. El primer caso corresponde a la llamada "investigación dirigida por teoría", mientras que la segunda es la "investigación dirigida por datos".

Hoy día, la mayor parte de los modelos son concebidos desde el punto de vista de la investigación dirigida por la teoría, y por ello las metodologías sueles estar sesgadas hacia ese enfoque. Sin embargo, y sin ignorar el importante papel de la teoría, los modelos pueden buscar principalmente expresividad y descripción. Y para ello, pueden tener requisitos que no son abordados por dichas metodologías genéricas. Por ejemplo, temas como la inicialización empírica, las limitaciones de la recolección de datos, la validación empírica intensiva, o el papel de los datos en el diseño no son considerados normalmente en estas metodologías. Así, existe una carencia de una completa metodología en ABM que, asumiendo que la investigación dirigida por datos tiene un enfoque y objetivos sensiblemente distintos, ofrezca el flujo de desarrollo de modelos dirigidos por datos. Esta metodología debería considerar el papel clave de los datos empíricos a lo largo de las fases de modelado. Esta carencia ha provocado que los modelos dirigidos por datos existentes hayan sido construidos sin un marco común.

Este trabajo pretende cubrir ese vacío y construir una metodología completa para guiar el modelado basado en agentes dirigido por datos (data-driven ABM). Así, puede afirmarse que cuando existen datos empíricos disponibles de la observación del caso de

estudio, el proceso de modelado y simulación implica nuevas fases. Esta metodología pretende guiar la introducción de datos empíricos en la simulación, acercándola al fenómeno real, pero reconociendo el papel fundamental de la teoría en todo el proceso. Este enfoque es complementado con un método sistemático de exploración del espacio de modelos para obtener modelos comprensibles pero a la vez descriptivos. Este método ha sido denominado "Deepening KISS" ("Profundizando en el KISS[1]") y es explicado en profundidad en el capítulo metodológico.

Esta metodología es apoyada técnicamente por la especificación e implementación de un framework de agentes sociales. Este framework está estructurado en módulos que pueden ser activados a voluntad, para facilitar la exploración del espacio de modelos y su construcción incremental, ambos en el marco del enfoque dirigido por datos. En lugar de tratar de construir un framework para uso general, éste se centra en una familia de problemas que puede abordar cómodamente.

Además, se ha desarrollado en profundidad un caso de estudio para probar y validar la aplicación de esta metodología en el marco del framework propuesto. Este caso de estudio se enfrenta al problema de la evolución de valores sociales, junto con los procesos simultáneos de emergencia de la amistad y dinámicas demográficas asociadas.

La construcción de este modelo basado en agentes, denominado Mentat, puede ser resumido en una serie de aspectos. La metodología dirigida por datos propuesta es aplicada intensamente a lo largo de su desarrollo. El proceso de modelado ha sido realizado (a) bottom-up y (b) top-down. (a) es representado por la red social que surge del comportamiento micro y las dinámicas de amistad. (b) se sostiene en el elaborado modelo demográfico. Todo ello ha sido implementado en el marco del framework de agentes, diseñado de forma modular y en capas incrementales. Las capacidades de Mentat han sido estructuradas en módules desactivables para poder explorar distintas combinaciones de modelos, siguiendo con la metodología definida. El modelo es validado desde un enfoque macro cuantitativo (validación empírica), desde un enfoque micro cualitativo (correspondencia de la dinámica social con los supuestos teóricos) y desde el enfoque teórico (discutiendo su consistencia sociológica). Además, distintas tecnologías de Inteligencia Artificial han sido incorporadas al modelo, probando la adaptabilidad del framework y la utilidad de éstas en simulación social. Mentat ha servido de caso de estudio para la metodología y el framework, pero a su vez ofrece un alto grado de comprensión sobre el problema, otorgando nuevo apoyo a determinadas teorías sociológicas. En concreto, este modelo enfatiza la importancia de la dinámica demográfica en el caso de estudio elegido: la evolución de valores sociales en la España de fin de siglo. Esto implica que

[1]KISS es el acrónimo de "Keep It Simple, Stupid", es decir, "Mantenlo simple, estúpido", principio equivalente a la navaja de Occam que aboga por la simplicidad como objetivo por sí sólo.

los cambios intergeneracionales son considerablemente más importantes que los intrageneracionales, al menos en el contexto español, reforzando así las teorías de Ronald Inglehart al respecto.

Palabras clave: amistad, demografía, encuestas, lógica fuzzy, microsimulación, modelado dirigido por datos, modelo basado en agentes, simulación social, valores.

Acknowledgements

First of all, I would like to give my sincere thanks to my supervisor Dr Juan Pavón Mestras. He gave me the opportunity, when I was still a lost undergraduate, to get involved in research and in the exciting social simulation field. He has provided me, as an expert advisor, with constant encouragement and thoughtful guidance for all these years. However, he has become much more than a supervisor. The passionate political discussions, the cheerful evenings sharing beers, and his company on plenty of trips and conferences have meant more to me than what any supervisor could offer. I know that my thesis would have never become a reality without his tireless efforts. Honestly, I cannot thoroughly express my gratitude towards him.

I would like also to thank my co-supervisor Dr Millán Arroyo Menéndez, who was involved in this weird adventure from the beginning. Thanks to his open mind and flexibility, and because of his effort for providing this work with a strong sociological support, this multidisciplinary thesis can be coined so.

But this would be a different piece of work and I would be a different person if Dr Nigel Gilbert had not accepted me in his research group for my long visiting period in the University of Surrey (UK). The multidisciplinary and multicultural environment of the Centre for Research in Social Simulation (CRESS) was the fertile ground which provided me with the deep insight of my field which I needed. I want to express my gratitude to Dr Gilbert for his advice and enlightening suggestions. I would also like to extend it to the members of CRESS and the Department of Sociology of Surrey for their enriching discussions and for never making me feel like an outsider.

Another key person in my work has been Dr Luis Antunes (Universidade de Lisboa, Portugal), who turned from a collaborator into a friend in the course of our research together. Dr Antunes' help meant a great improvement in both the model and the methodology of this thesis. Moreover, he has made a special effort in introducing me to the international community of social simulation.

I would like to express my gratitude to the 16 co-authors of the published papers along this period, as this work mainly advanced with each new publication that was written.

I also want to thank my colleagues in the department, especially Dr Pablo Gervás, who allowed me to get involved in my first large project during my undergraduate degree, and who has been constantly providing me with advice since then.

I must also thank the following friends for their numerous ideas, contributions and support: Jorge Carrillo, Mauricio Salgado, Rana Hassan, Sara Lafuente, Dario Taraborelli, Bastien Guerry, Laura Plaza, Ruba Mashtoub, Maciej Latek, José Manuel Galán, Jakob Grazzini, Pablo Rivière, Vicente J. Ruíz, Xavier Castelló, Javier Bustamante, Maite Delgado, Luis Garmendia, Miguel Ángel López, Carlos Jaurrieta, Carlos García and Stella Chatzitheochari.

I acknowledge financial support from the projects *Methods and tools for modelling multi-agent systems* (TIN2005-08501-C03-01) and *SICOSSYS: Agent-based tools for the modelling and simulation of complex social systems* (TIN2008-06464-C03-01), both supported by Spanish Council for Science and Technology (nowadays Spanish Council for Science and Innovation).

Special thanks are devoted to my family, with a special mention of my brother Tarek and of course my parents, for their understanding, endless patience, and continuous support.

Finally, I would like to express my appreciation and gratitude to Belén, who has indeed been my main support for these years of work. She encouraged me when it was most required, advised me when I was lost, and specially listened to me during the course of this work. Without her help and energy, this thesis would not have been completed.

FIGURE 1: The logos of the three research groups which supported this work: GRASIA (Universidad Complutense de Madrid, Spain), CRESS (University of Surrey, UK) and GUESS (Universidade de Lisboa, Portugal).

Contents

Abstract **iv**

Resumen **vi**

Acknowledgements **ix**

List of Figures **xv**

List of Tables **xix**

Abbreviations **xxi**

1 Introduction **1**
1.1 Context . . . 1
1.2 Motivation . . . 2
1.3 Objectives . . . 7
1.4 Document Structure . . . 9

2 Methodology: Injecting Empirical Data into Social Simulation **15**
2.1 Introduction . . . 15
2.1.1 Reviewing the MAS Literature . . . 15
2.1.2 The ABM Methodologies . . . 17
2.1.3 Theoretical vs. Data-driven Research . . . 19
2.1.4 Managing the Abstraction Level . . . 20
2.2 Towards a Data-driven Approach . . . 21
2.2.1 Why Data . . . 21
2.2.2 Data-driven Modelling: a Review . . . 23
2.3 The Logic of Simulation under the Data-driven Light . . . 24
2.3.1 The Classical Logic of Simulation . . . 24
2.3.2 Abstraction . . . 25
2.3.3 Data Gathering . . . 26
2.3.4 Simulation . . . 27
2.3.5 Verification & Validation . . . 27
2.3.6 Final Remarks . . . 28
2.4 Modelling Approaches . . . 28
2.4.1 Simple is Beautiful: The KISS paradigm . . . 28
2.4.2 Breaking the Rules: KIDS . . . 29

2.4.3 A New Perspective: Deepening KISS . . . 29
2.5 Filling the holes: new Stages . . . 31
2.5.1 Uprising Collected Data . . . 31
2.5.2 Data-driven Design . . . 31
2.5.3 Data Initialisation . . . 32
2.6 A Step Forward: Introducing Surveys into ABM . . . 34
2.6.1 From Microsimulation to Agent-Based Simulation . . . 34
2.6.2 Simulation & Surveys: a Tortuous Relationship . . . 35
2.6.3 Handling Data Sources: some Guidelines . . . 36
2.7 A Data-driven Agent-based Modelling Approach . . . 39
2.7.1 The Data-driven Cycle . . . 39
2.7.2 Merging Artificial Intelligence Technologies . . . 47
2.8 Discussion and Difficulties . . . 49
2.9 Concluding Remarks . . . 51

3 Mentat, the Evolution of Social Values in the Postmodern Spain 53
3.1 Introduction . . . 53
3.1.1 Introducing Simulation . . . 53
3.1.2 Some Contributions of ABM to the Quantitative Research . . . 54
3.1.3 Objectives of Mentat from the Sociological Perspective . . . 57
3.2 Theoretical Framework . . . 58
3.2.1 Data Input: The European Values Study and other sources . . . 58
3.2.2 Theoretical Background of the Study on Religiosity Evolution . . . 60
3.2.3 Theoretical Background of the Cultural Change . . . 63
3.2.4 Social Influence vs. Demography . . . 66
3.2.5 The Key Role of Modelling Demography . . . 68
3.2.6 Size Matters: Simulating Survey Individuals . . . 70
3.3 Conclusions: the Sociological Model Principles . . . 71

4 The Micro View: the Mentat Agent 75
4.1 Introduction . . . 75
4.2 Agent Models: A Review . . . 76
4.3 A General Architecture for the M-Agent . . . 78
4.4 The Agent Architecture Applied to this Case Study . . . 83
4.4.1 Contextual Agent Architecture . . . 83
4.4.2 Agent Attributes . . . 86
4.4.3 Life Cycle . . . 87
4.5 Concluding Remarks . . . 89

5 Social Dynamics 91
5.1 Introduction . . . 91
5.2 Overview of Mentat Social Dynamics . . . 93
5.3 Introducing the Fuzzy Mentat . . . 96
5.4 Preliminary: Fuzzy logic in a fuzzy environment . . . 96
5.4.1 Framing the Issue . . . 96
5.4.2 Reviewing the Importance of Fuzzy Logic in ABM . . . 97
5.4.3 Introducing Fuzzy Sets . . . 98

5.5 Communication . . . 99
5.5.1 Neighbourhood . . . 99
5.5.2 Similarity Functions . . . 100
5.5.2.1 Crisp Gratification Similarity . . . 100
5.5.2.2 Building a Fuzzy Similarity . . . 101
5.6 Friendship Dynamics . . . 103
5.6.1 Understanding Friendship . . . 103
5.6.2 Strangers → Acquaintances → Friends: a Fuzzy Logistic Function 103
5.6.3 Friends → Partner . . . 105
5.6.4 Family . . . 107
5.7 Results and Discussion . . . 108
5.8 Concluding Remarks . . . 110

6 The Macro View: the Global Conditions 113
6.1 Introduction . . . 113
6.2 Environment: A World. A Space. A Grid. . . . 114
6.3 Insight of the Model . . . 116
6.3.1 The Model Design . . . 116
6.3.2 Configuration: Main Parameters . . . 119
6.3.3 The Demographic Model: Equations . . . 119
6.4 Concluding Remarks . . . 124

7 Experimentation & Evolution of Mentat 127
7.1 Introduction . . . 127
7.2 The Model Evolution . . . 128
7.2.1 The First Prototype . . . 128
7.2.2 Mentat 0.1 - First Results . . . 130
7.2.3 Mentat 0.3 - Deepening: Need of Children . . . 133
7.2.4 Mentat 0.4 - Deepening: Warming-Up . . . 135
7.2.5 Mentat 0.5 - Deepening: Complex Demographic Equations . . . 136
7.2.6 Mentat 0.9 - Deepening: Complex Relationships . . . 138
7.2.7 Mentat 1.0: The Complete ABM . . . 139
7.3 Analysis of the Results . . . 139
7.3.1 Random Initialisation vs. Empirical Initialisation . . . 141
7.3.2 Validation: Discussing the Structural Modifications . . . 142
7.3.3 Validation: Discussing the Fuzzification Process . . . 146
7.3.4 Theoretical Validation: Discussing its Sociological Consistency . . . 148
7.4 Concluding Remarks . . . 149

8 Conclusions 151
8.1 Summing-up . . . 151
8.1.1 Summary of Achievements . . . 151
8.1.2 Contributions to the Field . . . 153
8.1.3 Published Works . . . 154
8.2 Future Work . . . 158
8.2.1 Natural Language Processing . . . 158
8.2.2 Other Fuzzy Extensions . . . 160

8.2.3 Data Mining 162
8.2.4 Mentat Extensions 163

A Apéndice: Síntesis del trabajo 165
A.1 Introducción 165
A.1.1 Contexto 165
A.1.2 Motivación 166
A.1.3 Objetivos 172
A.1.4 Estructura del documento 174
A.2 Conclusiones 177
A.2.1 Resumen del trabajo 177
A.2.2 Contribuciones al campo 178

Bibliography 181

List of Figures

1 The logos of the three research groups which supported this work: GRASIA (Universidad Complutense de Madrid, Spain), CRESS (University of Surrey, UK) and GUESS (Universidade de Lisboa, Portugal). x

1.1 Data-driven modelling and theoretical modelling can be understood as two opposite trends in ABM. 4
1.2 Reading flow diagram of the document structure, recommended for social simulators. 10
1.3 Reading flow diagram of the document structure, recommended for computer scientists. 11
1.4 Reading flow diagram of the document structure, recommended for sociologists. 13

2.1 Conception of the Axiom-theory-model-phenomena relationship, from the perspective of Natural Sciences. Source: Goldspink [109] and McKelvey [182]. 18
2.2 Conception of the Axiom-theory-model-phenomena relationship, from the perspective of Social Sciences. Source: Goldspink [109] and McKelvey [182]. 18
2.3 Semantic conception of the Axiom-theory-model-phenomena relationship, proposed for complex systems by McKelvey. Source: Goldspink [109] and McKelvey [182]. 19
2.4 Diagram showing the classical logic of simulation. Source: Gilbert, N., Troitzsch, K.G. 'Simulation for the Social Scientist'. Open University Press (1999) [107]. 25
2.5 Diagram showing a modified data-driven logic of simulation. 31
2.6 Proposed flow diagram for the data-driven agent-based modelling. 38
2.7 Sub-diagram concerning the Data Collection stage. 40
2.8 Sub-diagram concerning the Abstraction stage. 42
2.9 Sub-diagram concerning the Data-driven Design stage. 43
2.10 Sub-diagram concerning the Implementation stage. 44
2.11 Sub-diagram concerning the Simulation stage. 45
2.12 Sub-diagram concerning the Validation stage. 46

3.1 Snapshot showing emergent clusters in an agent-based simulation. 54
3.2 EVS question sample. Source: European Values Study Master Questionnaire 2008 Final Version [84]. 59
3.3 An example of the EVS comparative potential, according to a social values question: the importance of God. Source: Halman, L., Luijkx, R., van Zundert, M. 'Atlas of European values'. Brill Academic Pub (2005) [114]. 61

3.4 The Inglehart-Welzer Cultural Map of the World, where countries are located depending on their aggregated values of the two dimensions of Cultural Modernisation. Spanish values are located close to the centre. Source: Ronald Inglehart and Christian Welzel, 'Modernization, Cultural Change and Democracy'. New York, Cambridge University Press, 2005: p. 64 [151] based on the World Values Surveys [150]. 65
3.5 Scatter chart of Authority and Survival or Well Being. Source: R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148]. 66
3.6 The shift toward Postmaterialist Values among nine Western Societies, 1970-1994. Source: European Community Surveys (February 1970, Fall 1994), and U.S. National Election Surveys from 1972 and 1992. Appeared in R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148]. 67
3.7 Value type by age group, among the Publics of Britain, France, West Germany, Italy, Belgium, and The Netherlands in 1970. Source: European Economic Community Survey Feb. 1970, appeared in R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148]. 68
3.8 Societies with fast-growing economies have relatively large differences between the values of young and old. Note the high position of Spanish society. Source: World Values Survey 1993, appeared in R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148]. 69

4.1 Diagram showing the global M-agent architecture. Note that each layer inherits the behaviour of the previous one. *Basic Node* may inherit the behaviour of the basic node class of the agent library used, such as Repast, Mason, or others. 80
4.2 Sub-diagram showing the architecture of the Basic Node layer. 81
4.3 Sub-diagram showing the architecture of the Advanced Node layer. 82
4.4 Sub-diagram showing the architecture of the Cognition layer. 83
4.5 Sub-diagram showing the architecture of the Context layer. 84
4.6 Sub-diagram showing the architecture of the Social layer. 84
4.7 UML diagram of the agent hierarchy classes. 85
4.8 UML diagram of the main attributes aggregation class. This belongs to the *AttributedIndiv* class, and thus it is part of the context-based layer. . 87
4.9 UML diagram of the main attributes. 88
4.10 UML diagram of the tolerance attributes. 88
4.11 Finite state automaton showing the agent life cycle. 89

5.1 Flow Diagram summarising the process of emergence of a friendship link. 94
5.2 Flow Diagram summarising the process of partner finding (matchmaking). 95
5.3 A graphical example of concept definitions in fuzzy terms. 98
5.4 Different neighbourhoods: a) Von Neumann; b) Moore; c) Extended Moore with radius=2. 100
5.5 Graphical representation of the agent communication radius. 101
5.6 UML diagram of the different similarity strategies. 102
5.7 The logistic function chosen for the evolution of friendship, for several one-to-one similarity values. 105

5.8 Exponential growth of parameter J, in relation with the similarity S. 106
5.9 UML diagram of the possible agent groups. 108
5.10 UML diagram of the agent class extensions with the social relationships related methods. 109

6.1 A global view of Mentat in action. 114
6.2 UML diagram of the world classes. 117
6.3 The Model Architecture from the Model-View-Controller (MVC) pattern perspective. 117
6.4 UML diagram of the model hierarchy classes. 119
6.5 UML diagram of the auxiliary model classes, for the statistical measures, parameter configuration, initial population and demographic dynamics. . 121
6.6 Representation of the female life expectancy, according to the disperse empirical data from the INE [152] (before the linearisation of Figure 6.7). 124
6.7 Linear approximation to the female life expectancy, with a mean error of 2.8% compared to Figure 6.6. 124

7.1 For the implementation of Mentat, Java 5.0, the Eclipse IDE and Repast J 3.1 were used. 128
7.2 A view of the preliminary prototype with a few agents. Agent colours depend on their age and gender. 129
7.3 Mentat with 500 agents showing friendships (yellow links) and family nuclei (green links). 131
7.4 Density testing: changing space size *ceteris paribus*. 500 agents in a 60x60 space, showing the evolution of the religiosity patterns (1980-1999). Compare with Table 7.2. 132
7.5 Density testing: changing space size *ceteris paribus*. 500 agents in a 40x40 space, showing the evolution of the religiosity patterns (1980-1999). Compare with Table 7.2. 132
7.6 Mentat 0.5 space view, with three thousand agents and Boolean friendship. 137

8.1 Diagram showing the architecture of Mentat double output: stats and biographies. 159
8.2 Sample of an XML log of an agent turned into a piece of life story of a representative individual. 160
8.3 Screenshot of Repast Simphony. Source: Repast agent simulation toolkit web page: http://repast.sourceforge.net [212]. 164

A.1 La investigación dirigida por datos y dirigida por teoría pueden ser entendidas como dos tendencias opuestas en el modelado basado en agentes. 168
A.2 El diagrama de flujo para la lectura del documento recomendado para los especialistas en simulación social. 176

List of Tables

3.1 The cultural modernisation process: a comparison of traditional, modern and postmodern societies, according to Inglehart. 64

4.1 Classification of agent attributes. 86
4.2 Different behaviour according to the life cycle. 89

5.1 Comparison among the different ABM, in increasing order of fuzzification. 110

6.1 Main parameters of Mentat. 120
6.2 Activable mechanisms of Mentat. 122

7.1 Summary of the first prototype characteristics. 130
7.2 Evolution of the religious patterns of the Spanish society, according to the European Values Study (1980, 1990, 1999) [84]. 131
7.3 Summary of the characteristics of Mentat 0.1. 133
7.4 Summary of the characteristics of Mentat 0.3. 135
7.5 Summary of the characteristics of Mentat 0.4. 136
7.6 Summary of the characteristics of Mentat 0.5. 139
7.7 Summary of the characteristics of Mentat 0.9. 140
7.8 Summary of the characteristics of Mentat 1.0. 140
7.9 Comparison between EVS, the random initialised version and the data-driven version of Mentat 0.5. 141
7.10 Validation: comparison between EVS, a slightly improved Mentat 0.1 (as it has 2303 agents instead of 500) and Mentat 0.5 (with its 3019 agents). . 143
7.11 Comparison among different Mentat versions, in increasing order of fuzzification. 147

Abbreviations

ABM	Agent-Based Modelling
ABS	Agent Based Simulation
ABSS	Agent Based Social Simulation
AI	Artificial Intelligence
AL	Alternatives (religious pattern)
API	Application Programming Interface
AOSE	Agent Oriented Software Engineering
BDI	Beliefs-Desires-Intentions
CBR	Case-Based Reasoning
CA	Cellular Automata
CM	Cultural Modernisation
CSV	Comma-Separated Values file format
DM	Data Mining
EC	Ecclesiastical (religious pattern)
EVS	European Values Study
GIS	Geographic Information System
JADE	Java Agent DEvelopment framework
KIDS	Keep It Descriptive, Stupid
KISS	Keep It Simple, Stupid
IDE	Integrated Development Environment
IDK	INGENIAS Development Kit
JDK	Java Development Kit
LI	Low-intensity (religious pattern)
MABS	Multi-Agent Based Simulation
MAS	Multi-Agent System
MVC	Model-View-Controller architectural pattern

NLP	Natural Language Processing
NR	Non-religious (religious pattern)
OD	Opinion Dynamics
OWL	Web Ontology Language
RND	Random
RCT	Rational Choice Theory
REPAST	Recursive Porous Agent Simulation Toolkit
SA	Sensitivity Analysis
SNA	Social Network Analysis
SPSS	Statistical Package for the Social Sciences
UCM	Universidad Complutense de Madrid
UML	Unified Modelling Language
XML	eXtensible Markup Language
WVS	World Values Survey

For Belén

Chapter 1

Introduction

Any sufficiently advanced technology is indistinguishable from magic

Clarke's Third Law
Arthur C. Clarke, 1973

Any technology distinguishable from magic is insufficiently advanced

Gehm's Corollary to Clarke's Third Law
Barry Gehm, 1991

1.1 Context

A social system is a collection of individuals that interact among them, evolving autonomously and motivated by their own beliefs and personal goals together with the circumstances of their social environment. Social systems are an example of complex non-linear systems, as their behaviour can not be expressed as a sum of the behaviours of their parts. Thus, they usually exhibit emergent phenomena whose properties need new categories at a higher abstraction level in order to be studied. Such complex systems are difficult to study because most cannot be understood analytically: it is frequent the case where there is no set of equations that can be devised to describe the system [197]. However, such non-linear behaviour can be studied through the construction of a model and its simulation [107].

The idea beneath Agent-Based Social Simulation (ABSS) is that the researcher may be able to understand this complexity not by trying to model it at the global level but analysing emergent properties resulting from local interactions between autonomous agents that affect each other in response to the influences they receive [178]. Therefore, the specification of characteristics and behaviour of each agent is critical, in what it can affect the dimensions of the studied problem.

Such method facilitates a bottom-up approach for the analysis of macro behaviour in societies of interacting entities [20]. This aggregate behaviour is called emergent, as the collective and even individual behaviours could not be predicted or expected from the initial settings of the simulation [54].

There are, however, some limitations when trying to simulate social systems. The main issue is that the individual, with regard to a software agent, is by itself a complex system, whose behaviour is unpredictable and less determined than for an agent, which has behaviour and perception capabilities that can be designed with relative simplicity. Moreover, it is not possible in practice to consider the simulation of countless nuances that can be found in a real social system concerning to agent interaction, characterisation of the environment, etc. For this reason, it is impractical to intend the simulation of a social system in all its dimensions [107]. Thus, in order to have a feasible approach, a model should be constructed as an abstraction of the phenomenon under study. On the other hand, the researcher should limit to simulate concrete social processes in a systemic and interactive context. Therefore, the simulation of social systems should be considered in terms of focus on a concrete process [77].

In spite of these limitations, the agent paradigm offers many advantages to express the nature and peculiarities of social phenomena. However, most agent-based simulation models are simple, similar to cellular automata [262]. This may be valid to study emergent behaviour that results from deterministic behaviour of agents. Nevertheless, when the individuals are considered as complex mental entities with high cognition capacities as beliefs and values, that approach is rather limited. This is the reason why this work tries to promote an approach encouraging complexity against simplicity, in all dimensions: technical (with the specification of a modular framework which integrates several Artificial Intelligence technologies), epistemological (moving from an abstract view to a data-driven one, without losing generality) and methodological (defining a methodology to adapt a data-driven approach for ABM).

1.2 Motivation

Research methods usually follow one of the two traditional ways of doing science: induction or deduction. Induction is the discovery of patterns in empirical data, while deduction implies the specification of some axioms in order to prove logical consequences that are derived from them. However, according to several authors such as Axelrod [20] and Gilbert [101, 102], simulation can be defined as the '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 analysed inductively. Unlike typical induction, however, the*

> *simulated data comes from a rigourously specified set of rules rather than direct measurement of the real world. While induction can be used to find patterns in data, and deduction can be used to find consequences of assumptions, simulation modeling can be used as an aid [to] intuition.'* [20]

Thus, computer modelling can be defined as the computer-aided construction of an abstraction of an observed system for a specific purpose [235], and computer simulation can be understood as a particular type of modelling [107] with the purpose of 'driving a model of a system with suitable inputs and observing the corresponding outputs' [20, 33].

There are multiple types of simulation useful for Social Sciences [107]. Agent-based models (ABM[1]) present several features which turn them into a significantly attractive option to simulate complex social phenomena. As it was explained in the previous section, they facilitate the study of non-linearity of complex systems (such as human societies). Moreover, several advantages can be found for their use in Social Sciences. ABM attempts to establish a direct correspondence among entities in the social phenomena and computational agents, together with the relationships among them. This fact develops a potential of natural but formal representation of environments and social phenomena [80]. Besides, ABM is naturally prepared to model the emergence and self-organisation processes found in many complex systems. This is because the agents are autonomous entities whose local interaction in complex social networks develop a bottom-up global behaviour which is emergent.

There are multiple methodological proposals concerning the 'how to' build agent-based models. A methodology compiles experience in processes, methods and tools, providing a robust and systematic way to tackle new problems. Section 2.1 reviews different methodologies, together with their associated tools, coming from the Multi-Agent Systems (MAS) and Agent Oriented Software Engineering (AOSE) field. However, those methodologies do not focus on the simulation of 'artificial societies of autonomous agents' [52] and thus they usually lack several important considerations. For instance, the emergence and self-organisation processes, the conception of the individual as more than the rational goal-driven agent, or the role of the theory in the methodology (see section 2.1 for an in-depth review).

Several attempts of building a methodology for the modelling process have taken place from the ABM perspective. Again, section 2.1 reviews the main works concerning this issue, such as McKelvey's three ways of modelling phenomena [182] (positivist,

---

[1] ABM is the terminology used for referring to both the agent-based models and the process of building them: the agent-based modelling. Note that an ABM is different than the Multi-Agent System (MAS) from Computer Science. An ABM refers to a computational model in the frame of social simulation, which focuses on the modelling and simulation of societies. Such societies are composed by individuals that are represented by autonomous agents, and thus by a multi-agent system. But both the background and implications of these two acronyms diverge substantially.

theoretical and for complex systems) and Goldspink [109] and Gilbert [207] efforts to improve them. These efforts approach the classical logic of simulation [107] stages in different ways, exploring its sub-stages and the relationships among them.

However, there is an important epistemological distinction among agent-based models that can have methodological implications, and that all those methodologies do not consider. Models can be classified depending on the aim of the research [75]. Sometimes researchers seek a generic model to explain a social phenomena from a high degree of abstraction, and one that is simple enough to be used as an illustration of a specific theory or hypothesis. On the other hand, researchers may prefer to focus on the expressiveness of the model, together with the empirical descriptiveness of a specific case study [75, 223]. The first case corresponds to Theoretical Research, while the second one would be Data-driven Research, as shown in Figure 1.1.

FIGURE 1.1: Data-driven modelling and theoretical modelling can be understood as two opposite trends in ABM.

Most of the agent-based models are included in the Theoretical Research approach. Although they can be in different points of the right side in the continuum expressed in Figure 1.1, they frequently see the model mainly as an illustration of a theory.

However, without disregarding the role of theory, models can also seek expressiveness. In order to do that, they may have needs not met in general methodologies, frequently focused on theoretical models. For instance, issues such as the empirical initialisation, the limitations of data collection, the throughout empirical validation or the role of data in the design are not usually considered in those methodologies. Thus, there is a lack of a complete ABM methodology which, assuming data-driven research has different approach and aims, provides a specific flow of data-driven model development. Such methodology should consider the key role of empirical data throughout all the modelling stages. This lack has caused most data-driven models to be constructed without a common frame (as shown in the review of section 2.2.2).

Therefore, it can be advocated that when there are available data from the observation of the real phenomenon, the simulation process involves additional stages. For instance, empirical data (or at least empirically grounded distributions) should be considered instead of using abstract random distributions, in an effort to bring the model closer to the real phenomenon. Moreover, it can be used to inform the design and calibration of the model (cf. [122] and section 2.5).

This work attempts to cope with the relevance of empirical data in the agent-based modelling process. Moreover, following the example of AOSE methodologies, together with this methodology an agent framework is provided. This framework builds upon the data-driven approach which is exposed, and facilitates the development of ABM for a family of problems. Next, a specification of those problems is detailed (see section 4.3 for additional explanations):

- From a bird's-eye view, this agent framework facilitates the exploration [5] of complex data-driven models, as the framework allows the isolation of certain layers and modules. Thus, it is feasible to analyse the weight of different factors in the resulting aggregated effect. E.g. opinion dynamics vs. demographical evolution, media effect, second order emergence, norms emergence, cognitive capabilities, learning mechanisms and others.

- More specifically, this framework is optimised for studying the evolution of multiple individual characteristics in a given society and period, especially in contexts with abundant quantitative sociological data. Depending on the model, the focus can move from the micro behaviours (such as values or relationships) to the macro indicators (e.g. unemployment rates [134]).

- Besides, the framework is also prepared to study any problem from the whole agent-based computational demography spectrum, such as population pyramids, migration patterns, marriage and family dynamics, mortality crises effects, fertility decline and ageing consequences, etc [29].

Collateral to the case study, it is possible to explore the possibilities of different techniques in social simulation. Thus, some specific problems that may require specific technologies can be easily integrated as new modules inside the system. For instance: GIS[2], Artificial Intelligence (as explored in this work), statistical, analytical tools, visual tools, etc.

Moreover, the choice of a complex problem in the Sociology domain as a case study, such as the evolution of social values in the postmodern Spanish society, was considered essential to support the feasibility of the approach for Social Sciences. This problem is modelled following the proposed methodology and using the given agent framework. This ABM should provide an insight of the social phenomena studied.

---

[2] *Geographic information system* which includes mapping software for creating, storing, analysing and managing spatial data and associated attributes

Therefore, the case under study carries out an analysis of the evolution of several indicators in Spain between 1980 and 2000, focusing on social values and mental attitudes. This Spanish period is considerably interesting for research, due to the big shift on values and attitudes that the society bore then. The almost 40 years of dictatorship finished on 1975, when the country was far from Europe in all the progress indicators, including the dominant values and modernisation level (more traditional, conservative and religious). However, in 1999, the quantitative data (e.g. surveys) show not only a strong convergence with European values, but also positioning itself as one of the most modern countries [8, 148]. Thus, the change in Spain has been developed with a special speed and intensity during the studied period. This work studies this problem: the shift in the people's mentalities in this society and in this period.

In the last decades, Social Sciences are tackling issues that, due to their internal and subjective nature, were traditionally difficult to study if not avoided. This is the case of happiness studies [168], emotions [57, 90], religion dynamics [17, 18, 158] and values [142, 214].

Thus, social research has tackled the complex issue of values (not economic value, but moral or social values) from several perspectives. However, the process of values change and evolution has been qualitatively different since the studies of Ronald Inglehart in the subject [2, 147, 148].

His efforts are mainly framed in the field of Quantitative Sociology. The European Values Study (EVS) was first conducted by Jan Kerkhofs and Ruud de Moor [84], but its extension and further study must be attributed to Inglehart [147]. The EVS is a survey realised since 1981 in most of Western Europe, in different waves every 9-10 years. It aggregates the data by country, asking a representative sample of each of them a long set of questions mostly related to social values and complemented with useful secondary information [149] (see section 3.2.1 for further information related to EVS). This method provides possibilities of comparing data of different countries in the same period, or of the same country in different periods. Later on, Inglehart extended the EVS procedure to a world-scale with the World Values Survey (WVS) [150].

Inglehart research focuses on the existence of the Cultural Modernisation macro-trend [147, 148]. His early works were extended and supported by other researchers such as Halman [83] and Flanagan [88]. Besides, they were studied in-depth for the Spanish case by Arroyo [9], who contrasted it with additional qualitative and quantitative sources. These authors defend the hypothesis of a process of modernisation of societies from a Traditional perspective to a Modern and later Postmodern one. For an analysis of its different dimensions and consequences, together with other research in the area, see section 3.2.3.

According to Inglehart's works [148], this process of cultural modernisation would be caused mainly by intergenerational dynamics and only secondarily by intragenerational ones. Intergenerational dynamics take into account the changes across generations, which are socialised in different values. Intragenerational dynamics are considered the internal changes in a person course of life, that is, evolution inside each generational group. This perspective is justified arguing that values are acquired and shaped during an early life stage (the socialisation process), and then these values will keep rather stable the rest of the person's life. That is, later changes are possible, but sensibly more moderated. In this context, it is expected an essential influence of parents' values in their children's.

Assuming Inglehart's theory and Arroyo studies on religiosity evolution [10, 11, 14], values change is mainly driven by demographic inertia (responsible of intergenerational dynamics) and the socialisation process of the children in their families. Thus, an ABM which takes into account mainly these points should provide rather accurate results on modernisation indicators (such as religiosity, political ideology, or tolerance of controversial issues). The matching of a similar emergence of values change to the observed in the empirical data would provide additional support to Inglehart's hypotheses. See Chapter 3 for a support of the reasons for using ABM for this problem and section 7.3.4 for the sociological conclusions of the model.

The evolution of values is affected by a large number of interrelated and dynamic factors: gender, age, education, economic level, political ideology, religiosity, family, friendship social network, matchmaking and reproduction patterns, life cycles, tolerance regarding several subjects, etc. Thus, this case study suits well the high complexity degree typical for data-driven ABM.

Thus, large amounts of empirical data are included in the model, together with the supporting theory outlined here. For instance, the EVS will be loaded to initialise the agent population, qualitative research and advise from a domain expert will be used for the agent behaviour design, additional social research helps in the design of the friendship social network, empirically grounded equations will reflect the Spanish demographic dynamics, and an empirical validation together with theoretical validation will take place. See section 2.6.2 for a review of other ABM using surveys and section 5.1 for a review of social dynamics in simulation.

1.3 Objectives

The main goal of this work is the development of a methodology and appropriate tools to facilitate the study and analysis of a family of social systems. The approach considered is focused on expressiveness and descriptiveness, and thus has been strongly driven by data. Then, the subset of problems where it can be applied is determined by their proximity to the actual social phenomena[3]. The family of problems (see section 4.3 for an in-depth explanation) cover the whole agent-based computational demography spectrum [29], together with the study of multiple data-driven specific social phenomena. This framework facilitates the exploration [5] of those complex models, as the framework allows the isolation of certain layers and modules. Thus, it is feasible to analyse the weight of different factors in the resulting aggregated effect.

The accomplishment of this will result on two main contributions:

1. *A new methodology for data-driven agent-based social simulation.* In order to be systematic and be able to apply the data-driven approach in agent-based models (ABM), a methodology is needed. Thus, this work covers the elaboration of a methodology for building data-driven agent-based models, using the empirical data not only for the validation but also for the design and initialisation. Besides, this methodology provides guidelines to import data from surveys (Quantitative Sociology main tool) and to integrate other Artificial Intelligence techniques into ABM. See Chapter 2 for a detailed explanation.

2. *A modular and flexible agent framework* which supports the application of the defined methodology. From a software engineering approach, ABM researchers should encourage the construction of full context-based extensible frameworks for collections of problems instead of a myriad of isolated models. Such an agent framework would consist of a modular social agent architecture (in Chapter 4) together with the implementation of a whole flexible and extensible data-driven system composed of independent activable modules. This framework should be able to handle a family of sociological problems which would be validated through empirical data.

Thus, it involves the following tasks:

1. *Study, as a case study, the evolution of social values*, together with other related social parameters. This problem has a high level of complexity due to the amount of interrelated factors involved. The use of a specific real case allows an empirical assessment of the methodology and framework. Thus, the data-driven ABM coined Mentat tackles the study of social values change in the postmodern Spanish society during the period 1980–2000. Such model supports the hypothesis of the key importance of demography in this process. For an in-depth theoretical view, see Chapter 3, while for implementation details check Chapter 7.

[3] An abstract and theoretical problem would be to model 'the immigration patterns', while a problem close to the actual social phenomena would be 'the immigration flow of Spanish people to Germany in the sixties'

2. *Specify the agent social dynamics* which would provide the ABM with the capabilities of demographic evolution, match-making of similar individuals and friendship dynamics. The decisions taken should be both empirically and theoretically grounded. See Chapters 5 and 6 for details.

3. *Integration of multiple technologies in Agent-Based Modelling*: Mentat's modularity allows the model exploration with new modules integrating Artificial Intelligence (AI) techniques inside a framework of social agents. The methodology proposed recommends some possible uses of AI in different parts of the modelling process. Therefore, this model has been a field of experimentation for the merging of ABSS with fuzzy logic, natural language processing (NLP) or data mining (DM):

   (a) *Promote the use of fuzzy software agents.* This model encourages the exploration of fuzzy logic potentials specially in the context of social simulation. Mentat uses fuzzy agents and fuzzy homophily relationships among them. The whole evolving process of fuzzification is explained step by step, defining general procedures to fuzzify a crisp agent-based social simulation. Due to fuzzy logic natural adaptation to human concepts, it suits Social Sciences categories appropriately, allowing an improvement in the model quality (see Chapter 5 for a deeper explanation).

   (b) *Exploration of new output types from ABM.* Typically, ABM has only a macro output in the form of statistics (maybe including Social Network Analysis metrics), graphs and the visualisation of the simulation space. Mentat's modularity has allowed the exploration of other alternative outputs of the simulation taking place. One possibility is to trace the agent actions and export their micro behaviour in XML, in order to be processed by an external tool that would select the most representative individual for the studied phenomenon. Afterwards, through NLP techniques it is possible to generate a biography of the selected agent life, which would provide an insight of the ABM micro level, presented in a way easy to understand. Another possibility would consist in the extraction of clusters from the input and output of the system through DM. This could be useful for a second order validation of the ABM. Both technologies and others are explored in the future work, in Chapter 8.

1.4 Document Structure

First of all, note that there is not a specific chapter of state of the art. This is because each chapter tackles a review of main works concerning the scope of the subject. Thus, each chapter is somehow self-contained. Besides, note this work has three different reader profiles in mind: computer scientists, social simulation specialists and pure sociologists. The vocabulary and contents are focused on these three groups, with a bias towards formal and technical concepts. However, certain chapters are more focused on one or another profile, allowing different readings of the same work, as it is explained further on.

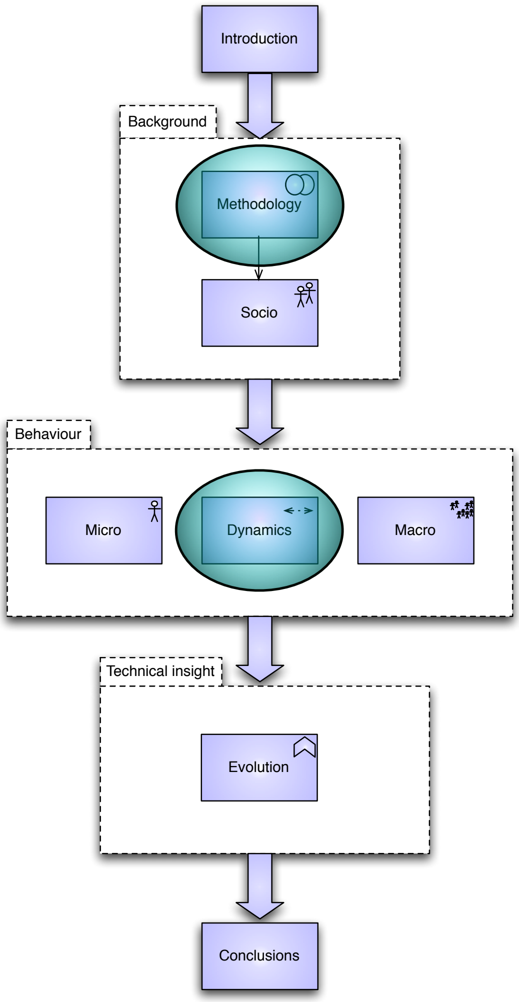

FIGURE 1.2: Reading flow diagram of the document structure, recommended for social simulators.

The following chapters constitute this PhD thesis:

1. **Introduction**. This chapter realises a brief synthesis of the discipline main facts, together with a description of the motivations and objectives of the whole work. This should be enough as a frame for the reader.

2. **Methodology: Injecting Empirical Data into Social Simulation**. This chapter presents the main contribution of the thesis: a methodology for data-driven

agent-based modelling. The classical agent-based modelling process is described, and each of its stages analysed. The issues found concerning the data-driven approach are discussed in a review of the different existing methodologies. This will lead to the exposition and discussion of the proposed methodology.

3. **Sociological Case Study: Mentat, the evolution of values in the post-modern Spain**. This is the case study chosen for the validation of the method and framework proposed in this work. This chapter is mainly theoretically focused, and though it may be of low interest for the computer scientist, it constitutes the basis to understand the sociological implications of this case study.

4. **The Micro View: the Mentat Agent**. It describes the social agent built following the proposed methodology and framework, for a specific family of problems, after discussing some of the existing alternatives. Following the chosen approach, it analyses the agent life cycle and characteristics from both behavioural and technical points of view.

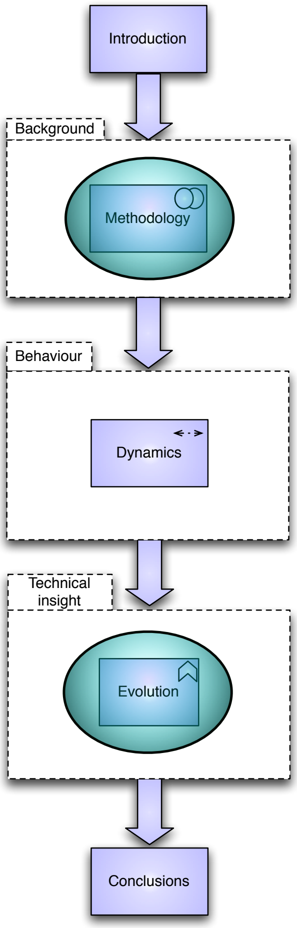

FIGURE 1.3: Reading flow diagram of the document structure, recommended for computer scientists.

5. **Social Dynamics**. This chapter describes one of the major contributions of this work. It handles a series of key concepts: the process of Mentat fuzzification (and its justification), the agents communication frame, the similarity strategies tested and, most important, the social dynamics. These dynamics are presented formally and explain the interrelated behaviour of the agent social network.

6. **The Macro View: the Global Conditions**. From a bird's-eye view, the whole agent-based system together with its environment is explained, taking into account every important choice taken, and the main system parameters. Besides, the demographic model, with a critical weight in the ABM, is detailed.

7. **Experimentation and Evolution of Mentat**. This chapter describes the evolution of the implementation, explaining every stage of Mentat and each key decision that made it evolve. Always following the data-driven 'Deepening KISS' approach (explained in chapter 2), the ABM increased in complexity with each iteration. Finally, the output is analysed, together technically and sociologically.

8. **Conclusions**. In the concluding chapter, the main achievements are detailed and discussed. Besides, multiple lines of future work (mainly merging other Artificial Intelligence technologies to improve certain aspects) are exposed and studied in-depth, with the results already achieved during the experimentation.

Depending on each of the three reader profiles, a different reading-flow can be suggested. The social simulation specialists will be able to extract the most benefit from this work, as they are the main target audience. Thus, the whole thesis can be interesting for them, and they should follow the main flow reading diagram 1.2. However, two chapters that can be stressed are the 2: Methodology and 5: Social Dynamics.

On the other hand, and as it can be expected, sociologists and computer scientists will be mainly interested in partial contents of this work. Computer scientists and Artificial Intelligence experts should focus on the methodological chapter and specially on the Experimentation & Evolution chapter. Besides, for interesting applications of AI techniques, Social Dynamics and Conclusions (with special attention to the Future Work sections) can be studied, as the diagram 1.3 shows.

Finally, pure sociologists with just basic skills in Computer Science (or none) still may find interesting this work. The introductory chapter should give the needed framework to be able to approach the Sociological chapter easily. The last sections of the Experimentation chapter and the beginning of the Conclusions cover some other aspects of interest. Finally, for the brave sociologist who wants to deepen the field, the Social Dynamics chapter is recommended, as drawn by the diagram 1.4.

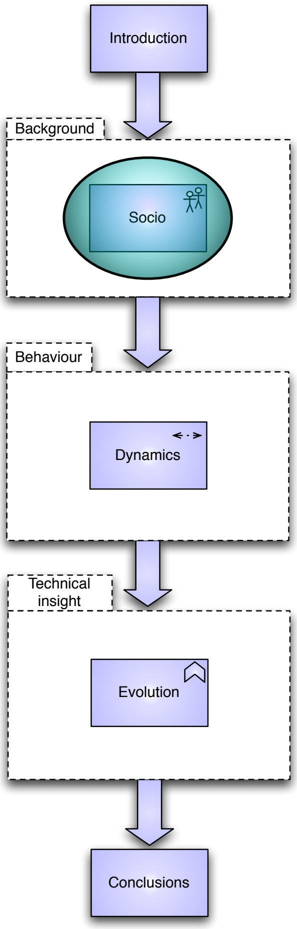

FIGURE 1.4: Reading flow diagram of the document structure, recommended for sociologists.

Chapter 2

Methodology: Injecting Empirical Data into Social Simulation

In Oceania at the present day, Science, in the old sense, has almost ceased to exist. In Newspeak there is no word for 'Science'. The empirical method of thought, on which all the scientific achievements of the past were founded, is opposed to the most fundamental principles of Ingsoc

George Orwell, *1984*, 1948

2.1 Introduction

2.1.1 Reviewing the MAS Literature

The discussion on which is 'the scientific way' of building models is extensively tackled in Philosophy of Science [46]. However, this should not be an obstacle to propose methodologies to help researchers and provide methods and tools that could guide the modelling efforts.

A methodology compiles experience in processes, methods and tools, providing a robust and systematic way to tackle new problems. A methodology for agent-based modelling should take care of: identification of relevant data, definition of the model, execution of the simulation, interpretation of the simulation, refining the model or the simulation, in-depth testing and validation. The search through the space of possible designs of the agents involved in the simulations, as well as their organisation (society) and the experimental set up (exploration) must be conducted with some guiding principles in mind. Thus, this section attempts to outline some of the guidelines that will be proposed further in this chapter, framing them within their historical context.

One of the main precursors of agent-based models (ABM) are the multi-agent systems (MAS), a technology situated in the field of Artificial Intelligence. From this discipline, multiple methodological proposals have arose [145]. However, they usually have different aims than the Social Sciences or Philosophy approaches, focusing on technical issues instead of epistemological ones.

MAS methodologies usually tackle issues related to the formalisation of the different stages, the technical implementation, software engineering conceptualisations that facilitate replication and flexibility, architectures to organise the different modules, efficiency, etc.

The rational choice approach is rather extended, frequently using BDI agents[1], but not necessarily. This fact of assuming individuals as rational self-motivated entities constrains the possibilities of application of these methodologies and tools to Social Sciences (a discussion on this issue can be found in section 4.2). Moreover, most of these methodologies (but not all, as it will be exposed) do not consider the simulation of complex system and thus the processes of emergence that take place in ABM. Finally, ABM tend to use a more intensive use of social theory than MAS, and this is reflected in the methodologies used, where the role of theory is usually not as critical as in the first.

However, several of the concepts developed in the MAS theory can be (and have been) useful for the agent-based modelling of social systems. For instance, this is the case of the formal definition of roles, tasks, scheduling, goals, scenarios and agent-oriented architectures.

From the wide spectrum of methodologies available [145, 202, 202] some representative examples can be outlined.

An example of interesting methodology is CoMoMAS [108] (or its extension MAS–CommonKADS [146]), based on the CommonKADS standard [227]. Even though it is biased towards the definition of goal-driven rational agents, it formally describes a methodology and architecture for the cooperation of agents with a knowledge base and cognitive and social capabilities.

Tropos [35] is a popular methodology, focused on resource management and formalisation of diagrams. Even though focused on BDI agents, it has a robust architecture which takes into account the different dependencies among resources, actors and goals.

An interesting methodology which approaches the issue in a different way is INGENIAS [202], which is supported by a whole collection of tools (the INGENIAS Development Kit, IDK). It proposes a model-driven development [16, 199], which means that the researcher must specify in detail the model to be simulated, aided by graphical tools, instead of programming it (as the IDK generates the programming code). Even

[1] The Beliefs-Desires-Intentions model (BDI), proposed by Bratman [34] have had a tremendous success in the field. However, it is strongly based on the rational choice theory, which implies several issues for modelling social systems, as tackled in section 4.2.

though recent efforts of application to ABM are taking place [221, 222], nowadays the IDK focuses on goal-driven JADE agents.

The last methodology reviewed here comes from MAS but approaches the ABM field: ADELFE [28]. It aims to model adaptive MAS from a software engineering approach, which can be considered an advancement towards the convergence with the ABM field. It does not tackle adaptive agents: it is the whole system which is adaptive, that is, shows an emergent behaviour the same way as ABM usually does. ADELFE tackles successfully the dynamics of these systems but it still encourages the use of classical goal-driven agents of MAS and is biased towards them.

2.1.2 The ABM Methodologies

However, there are fundamental differences between the MAS and ABM fields. As Conte et al. state [52]:

> *'If the MAS field can be characterised as the study of societies of artificial autonomous agents, agent-based social simulation can be defined as the study of artificial societies of autonomous agents.'*

As any scientific field, agent-based social simulation needs a scientific method to be used as main guidelines. The starting point can be considered the classical 'logic of simulation' [107], in which a target phenomenon is modelled, and simulation data are validated against collected data, as it is described in section 2.3. Such general scheme follows similar patterns of classical sciences such as Physics, in which a real phenomenon (e.g. a block falling on a ramp) is modelled (with formulas for gravity and friction), an experiment is carried out (in the lab, with a small block and a ramp), and validated against the collected data (measurements from the block falling).

There have been several efforts in defining a methodology for the design of ABM, approaching the logic of simulation from different points of view.

The early proposals such as Fishwick's [87] are significantly generic and thus difficult to be used as guidelines for modelling. Fishwick just considers three bidirectionally linked stages (design, analysis, execution), which do not clarify enough neither the subtasks that those stages imply (such as validation, abstraction or role of theory), nor the rest of the stages (such as the formalisation, construction of the computational model or exploration).

A subsequent conception is McKelvey's [182], analysing the relationship among Axiom-Theory-Model-Phenomena in Complexity Theory. His approach is rather interesting as it differentiates three ways of approaching the model. First, the sequence 'Axiomatic Conception', frequently used in Natural Sciences: the Theory, which is derived from an Axiomatic Base, is used to build a Model, that will be tested against the

Phenomena (and thus all the links shown in Figure 2.1 are unidirectional, in a sequential approach). If the Model's predictions are consistent with the empirical observation of the Phenomena, the Theory is considered correct [259]. This strongly positivist approach has been classically disregarded by social scientists (except Economics, which frequently follows many positivist assumptions), especially in those cases were emergent phenomena take place.

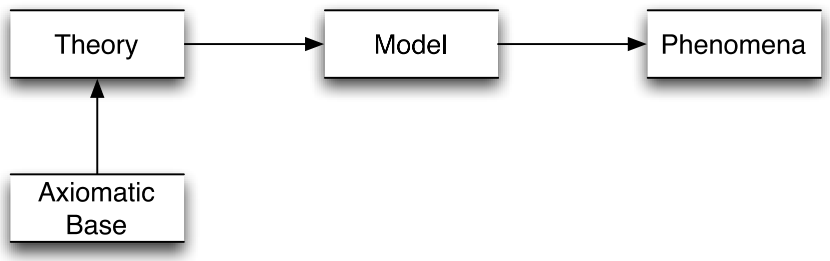

FIGURE 2.1: Conception of the Axiom-theory-model-phenomena relationship, from the perspective of Natural Sciences. Source: Goldspink [109] and McKelvey [182].

The 'natural' way of modelling from Social Sciences is the second McKelvey's relationship: Theory has a bidirectional and direct relationship with the Phenomena, and the Model is derived from the Theory without interaction with the Phenomena. It is the case where models are just illustrative from the proposed Theory, and several models are possible and even necessary. Those models usually show a high degree of abstraction and informality, rather difficult to falsify in Popper's sense[2] [205].

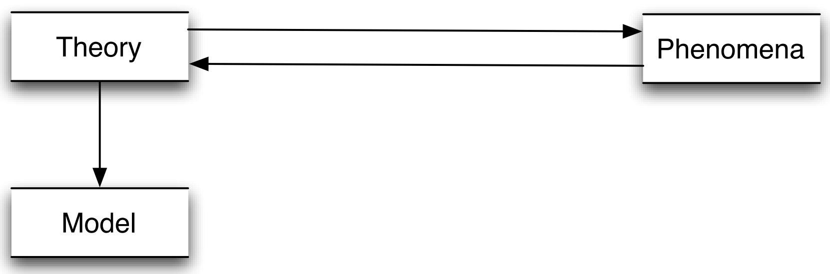

FIGURE 2.2: Conception of the Axiom-theory-model-phenomena relationship, from the perspective of Social Sciences. Source: Goldspink [109] and McKelvey [182].

The third McKelvey's relationship is the one proposed for simulation of complex systems (and agent-based social simulation): the Theory is derived from an Axiomatic Base; a Model is built from the Theory, and the Model is tested against the Phenomena.

[2]Falsifiability is the logical possibility that an assertion can be shown false by an observation or a physical experiment. According to Karl Popper, a hypothesis, proposition, or theory is scientific only if it is falsifiable [205].

However, the Model-Phenomena link is bidirectional, and so it is the Theory-Model. This means that this method proposes a two-way test, where the model is formal enough to be validated empirically (as in the positivist approach) but must be consistent with the theory as well, and could lead to modify it in this exploratory cycle.

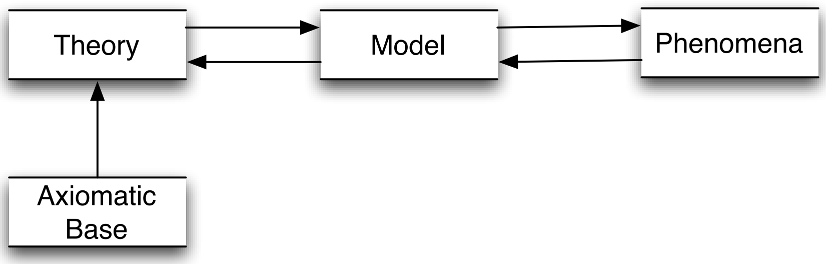

Figure 2.3: Semantic conception of the Axiom-theory-model-phenomena relationship, proposed for complex systems by McKelvey. Source: Goldspink [109] and McKelvey [182].

Goldspink [109] extends the third relationship formalising and extending the cycle to include the Axiomatic Base (Hypotheses), in a way it could be considered part of the theoretical corpus in McKelvey's cycle. This formal approach allows an improved systematic exploration of agent-based models.

Gilbert and Ramanath [207] review different contributions including Goldspink's advancements and define a global cycle of generic stages for the modelling and simulation research. This work includes a separation among model conceptualisation, model design, and model implementation. Besides, it considers a Verification and Validation stage. The addition of Publication and Replication completes the approach, although it could be considered apart from the original modelling sequence (as a model could be complete even without any of them).

Other proposals of the model development process take another perspective: the roles participating throughout such process. Independently the roles are played by a single researcher or several of them, they can be identified in the different stages. Thus, Edmonds [76], Drogoul [74] and Galán [94] differentiate the roles of the thematician (the domain expert), the modeller (the 'bridge' among theory and computing) and the computer scientist (and even programmer in Galán's work). The stages considered are similar to those of Gilbert and Ramanath, even though they are coined in a different way.

2.1.3 Theoretical vs. Data-driven Research

There is an important epistemological distinction among agent-based models that can have methodological implications, and that all those methodologies do not consider.

Models can be classified depending on the aim of the research [75]. Sometimes researchers seek a generic model to explain a social phenomena from a high degree of abstraction and one that is simple enough to be used as an illustration of a specific theory or hypothesis. On the other hand, researchers may prefer to focus on the model's expressiveness, together with the empirical descriptiveness of a specific case study [75]. The first case corresponds to Theoretical Research, while the second one would be Data-driven Research, as shown in Figure 1.1 (whose differences are better tackled in section 2.2).

Most agent-based models are conceived from the Theoretical Research approach. Although they can be in different points of the right side in the continuum expressed in Figure 1.1, they frequently seek a similar aim to the second way of modelling in the McKelvey's conception [182] previously exposed: the model is mainly an illustration of a theory.

However, without disregarding the role of theory, models can also seek expressiveness. In order to do that, they may have needs not met in general methodologies frequently focused on theoretical models. Thus, there is a lack of a complete ABM methodology which provides a specific flow of data-driven model development, assuming data-driven research has different approach and aims. Such methodology should consider the key role of empirical data throughout all the modelling stages. This lack has caused most data-driven models to be constructed without a common frame (as it is shown in the review of section 2.2.2), following different approaches usually not specific for their needs.

This work attempts to fill this gap in the field and build a complete methodology to model data-driven ABM. It supports the hypothesis that when there are available data from the observation of the real phenomenon, the simulation process involves additional stages. For instance, empirical data should be used to initialise the simulation instead of typical random distributions, in an effort to bring the model closer to the real phenomenon. Moreover, data can be used to inform the design and calibration of the model (cf. [122] and section 2.5).

2.1.4 Managing the Abstraction Level

The debate between abstraction and descriptiveness has been going on in the philosophy of science for quite some time [242]. In the specific case of social simulation, this debate moves towards theoretical models and data-driven ones. Within this issue around the abstraction level and accuracy to the target phenomenon, the ruling paradigm in the discipline is KISS [20], typical from theoretical models. KISS stands for 'Keep It Simple, Stupid!' and stems from Occam's razor: the idea that things should be kept as simple as possible and made as little more complex as explanation purposes demand. Applied to social simulation, KISS ideally seeks simple and abstract models that are general enough to be explanatory for multiple specific cases (it is further explained in section 2.4.1). This approach has been criticised by data-driven modelling supporters, even proposing an opposite approach: KIDS [77]. This alternative, 'Keep It Descriptive, Stupid', claims to be an appropriate perspective for data-driven modelling, in which the model must be constructed 'around the case study'. That is, instead of building it from abstract theoretical assessments, it should be build from empirical data, the closer to the real phenomenon as possible, seeking complexity instead of simplicity, descriptiveness instead of abstractness. However, such approach rises new issues associated to modelling with a high complexity level that may be sometimes unnecessary, and usually difficult to correct later on.

Starting from the classical logic, and inspired by the iterative development of process models in current Software Engineering, Antunes and colleagues have proposed a methodology for multi-agent-based exploratory simulation that proposes series of models designed to progressively tackle different aspects of the phenomena to be studied [5]. This new approach, coined 'Deepening KISS' during its development in this project, starts from a KISS model, following Sloman's prescription of a 'broad but shallow' design [232]. Then, through the use of evidence and especially data, a collection of models can be developed and explored, allowing for the designer to follow the KIDS prescription without really aiming at more simplicity or abstraction. This exploration of design space allows to pick the best features of each model in the collection to design a stronger model, and the process iterated. Deepening KISS with intensive use of data can be placed middle way in terms of simplicity versus descriptiveness, whilst it acknowledges the role of the experimenter as guide the search for the adequate models to face the stake-holders purposes, as it will be further explained in 2.4.3. This work emerges in order to fulfil the need of the definition of a methodology, that takes into account this iterative approach and the consideration of the role of observed data in ABM. Thus, such methodology is exposed throughout this chapter, while a framework which should facilitate the application of these ideas is explained in the following chapters.

2.2 Towards a Data-driven Approach

2.2.1 Why Data

'The best model of a real phenomenon is the phenomenon itself' [215], but of course such a model would have the unmanageable level of complexity that wants to be avoided by the process of modelling. Thus, the 'art' of modelling is the art of simplification, as every model is, by definition, a simplification of reality. A simpler model would often be expected to be easier to understand, to contain fewer errors and be quicker and easier to build, run and analyse.

Then, the problem turns to be the selection of the optimal model, among all the possible simplifications of the phenomenon. The collected empirical data from the phenomenon can guide such selection, with the aim of finding a good balance between simplification and realistic behaviour. Such data can be useful during the modelling of the different objects or relations, in the initialisation and the validation. But raw data must be pre-processed to adapt it to the model, in order to provide a structure that a program can process. Therefore, modellers often have to resort to techniques that filter away unnecessary data complexity.

One technique is to use statistical measures to provide parameters to probability distributions. For instance, 'let's assume that salaries follow a Gamma distribution'. This abstract description of a given quantity spares the researcher from deepening into real data, but it is by no means more general than a sensible use of data. The only way the use of statistical distributions can be considered more adapted to the problem at hand, and possibly more general to encompass other similar problems, is precisely by the use of several collections of data, carefully tested with statistical techniques against the distributions that are been advocated here. Each of these techniques will allow for a quantified error (confidence of fitness to the distribution). Also, those distributions are usually static, so not particularly adapted to the dynamical nature of computer simulations.

Not knowing usually what the correct statistical distribution is, it is probably better to use one or more empirical distributions, that is, a distribution based on real data. Or a typical set of data that could be followed could be preferable. The problem is that 'typical' is hard to define formally. The statistical methods aim to define that notion. Another fundamental problem with probability distributions is that they are good to describe static overall behaviours, especially from an *a posteriori* perspective. These distributions have many more problems in providing the *reasons* that may cause individual behaviour.

The power of statistical distributions to fit well (and quantifiably) a collection of data, and their mathematical elegance, give the researcher no reasons why models based on them can be more general than the scope of the data collection they are based on. That generality can only be achieved by the use of even more data, and more statistical fits.

The use of data is advocated here not only through statistical models, but in other phases of the simulation development. Thus, the researcher must pay a close watch to the universe they are coming from. Whatever form researchers are using to inject data into the model (and surely statistical measures are one), they must ensure that data are representative of the universe for which they are designing the model. Representativeness is again hard to define formally. However, by using data in a mediated manner, the representativeness problem arises twice. In 2.6.3 some procedures for handling data for the purposes of data-driven modelling are provided.

2.2.2 Data-driven Modelling: a Review

Most agent-based models follow the KISS principle, and therefore try to be rather abstract and general enough to be explanatory for plenty of specific cases. As the empirical data are specific to one site and time, in order to do that they often decide to use standard distributions in several steps of the design: configuring the initial conditions of simulations, distributing objects spatially, determining exogenous factors or aspects of the agents' behaviour [130].

However, an increasing number of agent-based models are appearing [32], specially in recent years, which follow different approaches that try to be more realistic by getting closer to the target. Increasing complexity of the models, against the KISS paradigm, is strongly linked to a more intense use of the available data. This view implies breaking with the modelling 'for the sake of simplicity' and can even slightly modify the classical logic of simulation. In this section two alternatives to KISS are presented, while in the subsection 2.5.1 an alternative logic is proposed. However, before that theoretical descriptions some examples of data-driven simulations will be outlined.

Multiple recent works have considered the introduction of empirical data into ABM. There are plenty of works encouraging the empirical validation [32] and calibration [186] of models, but this is not the only use for empirical data in data-driven research.

A well-known example is the model of the extinction of the Anasazi civilisation [21, 69]. In this model, large amounts of empirical historical data were used, including geographical, archaeological and historic documents. These empirical data are used for improving the fitness between the simulation and the observed history of the evolution of an extinct civilisation. In this example, the exogenous factors (environmental variables) are not randomised, although the initial conditions are [61].

An interesting case study due to its complexity are the different water demand models such as Edmond and Moss work [77], in which data about household location and composition, consumption habits and water management policies are used to steer the model. This ABM is empirically validated successfully against actual water usage patterns. In the case of the water demand model of Galan et al. [95], a more intensive and systematic use of data is used, crossing from several sources. Moreover, this model offers a multi-layer model were several dynamics merge, including opinion dynamics, urban dynamics and migration patterns.

A good example of how data-driven models do not need to be just based on quantitative data are the works of Geller [98–100]. Those works on the power structures in the Afghan society constitute a data-driven model strongly driven by qualitative sources.

The simulation of solidarity networks is based on agent behaviour informed by qualitative data collected in Afghanistan. Cross-validation of the ABM reveals small world characteristics [3] in those solidarity networks.

Another relevant case is Hedström's model of Stockholm youth unemployment [134]. In this example, quantitative data from surveys is imported in order to initialise the agent population, an uncommon practice in ABM as it is discussed in section 2.6. Besides, regression equations empirically grounded are used to calculate transition probabilities.

Another approach is taken by Deffuant et al. within the IMAGES project [62, 64], which explores opinion dynamics from a data-driven approach, using both quantitative and qualitative data. This complex model incorporates economic rational anticipation, decision function over several criteria and dynamics of information transmission. It was partly derived from interviews and experimental data, and provides an insight on political extremism, including economic interests, expectations, and past history.

Many other works do not make an intensive use of data like the ones presented, so they could be classified as partially data-driven. This is the case of several models using spatial data, such as pedestrian flow modelling [25] or simulations of markets like the electricity market [191]. For other data-driven works, focused on the use of surveys, see section 2.6.

These examples, and other available in the literature, show a lack of agreement and systematic modelling. Each example must follow the authors' particular view of the logic of simulation, applying it to their case instead of following an integrating methodology.

2.3 The Logic of Simulation under the Data-driven Light

2.3.1 The Classical Logic of Simulation

A generic view of the simulation process is represented in Figure 2.4, consisting in the execution of a model, which is an abstraction of the real phenomenon, together with the validation of the resulting data collected during the simulation against empirical data gathered from observations from the real world [107].

The Target is an observed phenomenon. As a result of a process of Abstraction, a Model, a simplified view of this phenomenon, can be obtained. This Model, in this case an Agent-Based Model, can be executed to obtain results, the Simulation data. A process of Data gathering (qualitative, quantitative or both) can be used to get the Collected data from the Target. The comparison of this data and the simulation output allows a process of validation. If there is structural similarity[3] between them, the ABM

[3]There is structural similarity when there is covariation in similar circumstances, and then this counts as evidence about the adequacy of the model, as a representation of the target [107]. Agent-based models are usually not deterministic because of their stochastic behaviour and thus they show different outputs

is validated and considered a good representation of the phenomenon. If there is not, the model should be modified and the simulation repeated until the output fits the gathered data.

These four steps, Abstraction, Data gathering, Simulation and Validation, are the basic steps of the classical logic of simulation. But, taking into account the ambivalence of theoretical and data-driven models, several questions arise from the detailed analysis of these stages, as the following subsections explain.

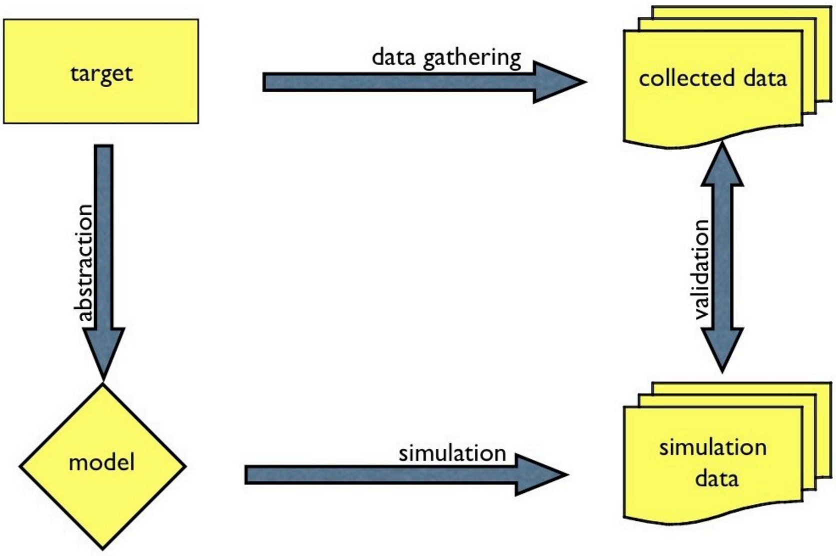

FIGURE 2.4: Diagram showing the classical logic of simulation. Source: Gilbert, N., Troitzsch, K.G. 'Simulation for the Social Scientist'. Open University Press (1999) [107].

2.3.2 Abstraction

As observed in subsection 2.2.1, the 'art' of modelling can be considered the art of simplification, that is, abstraction of reality. Thus, abstraction can be considered a mechanism for obtaining models. In the same section, the abstraction level is discussed: the aim of finding a good balance between simplification and realistic behaviour. The call for abstraction is justified in the fact that the increase in model complexity does not necessarily imply the increase in its accuracy. Thus, abstract models can behave better than some specific and detailed models, or at least give equivalent results with less complexity.

However, a high degree of abstraction is, by definition, far from the collected data. Frequently, theoretical models show a behaviour that is rather hard to validate against

in different executions. However, a model showing structural similarity would provide similar outputs (e.g. with the same trends and patterns) in different executions using the same initial conditions.

real data, due to this abstraction degree. Nevertheless, the modellers tend to assess that 'the real phenomenon behaves like this'. An example of such a model is Schelling's model of segregation (1971) [225], a very abstract model that, with only two parameters, claimed to model the human behaviour concerning segregation. Its abstractness is such, that it is impossible to falsify. However (or maybe because of that), a significant part of the scientific community admitted it as a correct model of this process. Attending to Popper's criteria of falsifiability, this model should not be considered even scientific (as it cannot be properly falsified against empirical data). In fact, recently several papers argue that income is more important than prejudice/tolerance in ethnic segregation, or that there is a complex interaction [38].

In the case of data-driven ABM, the abstraction degree that can be applied is sensibly constrained. Data brings the model closer to the real phenomenon. This has two immediate effects: usually it brings a closer matching with the real behaviour, and this comparison can be quite easily made through a proper empirical validation. As a drawback, a data-driven model is by nature more specific (with less generality) than an abstract one. However, the validity of this assessment can be questioned following Scott Moss [77], with the argument than an abstract model does not necessarily behaves as any real example, while a data-driven one at least behaves like one.

2.3.3 Data Gathering

The process of data gathering provides the empirical data needed for the final step of validation. Besides, this gathering process can be inspiring for the design of the model, as it will be further explored in section 2.5. However, the 'data' to be collected can be presented in many ways, immersed in the complexity of real world. Which data should be considered? How to collect such relevant data? Should modellers gather the data themselves? If they do not, how to adapt the available data collections to fit their needs? What if the data they need cannot be gathered?

Some of these issues bring other classical debates. In the process of collecting data the observer cannot be completely objective, so inherently by selecting the data that is considered relevant, the observer is 'affecting' reality with a particular theoretical framework (and prejudices), inserting an unintentional bias. This is an issue that appears in every discipline which requires a data collection process, not only in Social Sciences. Orthodox positivist assumptions of pure objectiveness are considered näive by critical theorists [139] and current trends, including neo-positivism [51]. However, even assuming everyone has a world view which no one can scape from, if some considerations and formal methodologies are followed in the process of collecting data, the collected data can be validated by other scientists (with different world views) which can confirm its validity. An alternative is to use existing data collections already available, standardised and with a public formal procedure. These collections (surveys, records or other research studies) are used by many scientists which make easier both their validation and their comparison with similar research. Other issues concerning data, and specially surveys, will be tackled in section 2.6.

2.3.4 Simulation

The actual process of simulation involves several steps: the configuration of the initial parameters, the simulation itself, the extraction of an understandable output, and the loop of this sequence. This loop means the simulation must be executed a reasonable number of times, first with the same parameters to check the structural similarity, and second an exploration of the parameter space through the multiple executions. This will be useful to guide the sensitivity analysis of the Validation step. Besides, during the simulation executions, some problems will arise and be located (e.g. implementation bugs, model inconsistencies). The model will be corrected as needed and the Simulation whole process recommenced. This clearly follows the 'Iterative Model' of the Software Development Process [243].

But the Simulation stage can be completed with the inclusion of another step: the exploration of the model state space [5]. Thus, the changes of the parameters and inputs would go further than what an initial sensitivity analysis would accomplish. A superficial exploration would test different types of inputs (e.g. different data sources), not just small changes in some parameters. A deep exploration would carry out the implementation of different model versions, modifying different modules. Any of these explorations would search for other alternatives to the current agent-based model, looking for the best option. This exploratory method is later explained in-depth in 2.4.3.

2.3.5 Verification & Validation

Before Validation, the Verification process (sometimes called 'Internal Validation' [74, 221, 240]) should be carried out, in order to evaluate whether or not the system complies with its specifications, properties and conditions imposed. On the other hand, Validation can be defined as the process of establishing evidence that provides assurance that the system accomplishes its intended requirements, that is, that the program does what it was planned to do, reflecting the behaviour of the real phenomenon. However, the case of simulations are usually harder than common programs [107], due to their essential differences: common use of random number generators, non-deterministic behaviours, high level of entities, concurrent and interrelated activities, etc. Furthermore, validating simulations can rise complex epistemological problems, related to the status of the knowledge produced by a simulation, as discussed in-depth in [165] or turned out to be potentials in [58].

The notion of validity relies on the comparison of the output with collected data. If such a comparison is below an error threshold, the model is considered valid. However, this idea cannot be generally applied. Sometimes, it is not possible to check all parameters and output values against empirical data. This is characteristic of the theoretical ABM, or significantly abstract models like the ones coming from Physics and Complexity Theory [43]. If this is the case, the only validation that makes sense is a sensitivity analysis (SA). SA is the study of how the variation (uncertainty) in the output of a model (mathematical or computational) can be apportioned, qualitatively or quantitatively, to different sources of variation in the input of a model [219], investigating its robustness. However, through this exhaustive process, it is easy to reach a maybe too large number of runs needed: e.g. a model with 6 parameters, each taking 10 possible values, needs a million runs to explore the whole parameter space. This is the reason why it is relevant an exploration of different key scenarios, representative of the different cases and limit conditions.

2.3.6 Final Remarks

This section has reviewed the impact of data in the main stages of the classical logic of simulation. Several issues have arisen, specially in the debate of the level of abstractness and the theoretical models vs. data-driven models issue. Such logic underlies the development of most of the agent-based models. However, in the frame of this general logic, several modelling approaches can be defined. Each approach may hold a particular view of the world, focusing in certain aspects and guiding the modelling from it. The next section carries out a review of the possible modelling approaches that the modeller can choose.

2.4 Modelling Approaches

2.4.1 Simple is Beautiful: The KISS paradigm

In general, *abstraction* can be defined as a process of generalisation by removing some elements from an observable phenomenon in order to build a model that can be handled. The amount of information kept in this process is an issue for debate, but it usually depends on the purpose of the model.

As stated in the Introduction, agent-based social simulation usually follows the KISS principle ('Keep It Simple, Stupid') proposed by Axelrod [20] following the 'Occam's razor' argument. The justification to use it relies on the importance and practicality of simplicity in modelling. Simplicity is helpful for transmitting the model to the scientific community, promoting understanding and extensibility. Besides, making an abstract and simple model is often supported with the argument that such models are more general, and therefore have possible applications in many real cases. Another reason for its spreading is that building a simple model is easier than building a complex one. And furthermore, it is not only the design: it is easier to implement, analyse and check [77].

The typical KISS model is a model with a very high abstraction degree, which makes very difficult its validation against empirical data (so sensitivity analysis should be done), and thus rarely carries out a deep data gathering. Although it is possible to follow a KISS approach for data-driven modelling, it is difficult not to break it: the data-driven approach usually adds high levels of complexity that clearly clashes with simplicity.

2.4.2 Breaking the Rules: KIDS

The KIDS approach, formulated by Edmonds and Moss [77], opposes KISS and promotes the principle, 'Keep It Descriptive, Stupid'. This alternative has achieved some notoriety in the ABM community (e.g. [100, 160]) because, while KISS is attractive and understandable, it is not always realistic or useful. KIDS asks modellers to begin with the most similar model to the target, in spite of its complexity. Only afterwards, should the model be analysed to see which parts could be simplified while preserving the behaviour. With further simplifications, a KIDS ABM could be used in several contexts while being sure that it has solid foundations.

The KIDS modelling view especially takes into account the data-driven modelling, as it demands a great effort to get as close as possible to the real phenomenon, which obviously needs big amounts of data. It is completely against the KISS ideal, and very abstract models do not fit in this context. As it calls to collect huge amounts of data, it is possible to validate them against the empirical data.

However, the authors admit that 'Neither the KISS nor the KIDS approach will always be the best one, and complex mixtures of the two will be frequently appropriate'. Though this view limits the approaches to just these two, without considering the possible appearance of others, like the one presented in the next subsection.

2.4.3 A New Perspective: Deepening KISS

Whereas in KISS the models are designed as simple as possible and only made more complex when difficulties are met; and in KIDS the idea is to start with a model that is descriptive in face of evidence and made progressively more abstract and simpler as more evidence and understanding allows it; there is a third way that can be coined 'Deepening KISS'. This deepening phase is a part of a more comprehensive methodology described in [5]. The idea is to start from something close to a KISS model, but following Sloman's prescription of a 'broad but shallow' design [232]. Then, through the use of evidence and especially data, as it is prescribed in [130], a collection of models can be developed and explored, allowing the designer to follow the KIDS prescription without really aiming at more simplicity of abstraction. Once the design space (of agents, societies and experiments) has been reasonably explored, the best features of each model in the collection can be used to design a stronger model, and the process iterated. This perspective has some similarities to the Extreme Programming software engineering methodology [26].

This deepening principle allows models to be made as complex as necessary, but no more than the designer wants for the sake of control of the model and adaptiveness to the research questions raised. It is the exploration of the models themselves that will inspire further deepening, or allow the process to stabilise and other features to be addressed.

A possible sequence of deepening a concept, representing some agent feature, (say parameter c, standing for honesty, income, or whatever) could be to consider it initially a constant, then a variable, then assign it some random distribution, then some empirically validated random distribution, then include a dedicated mechanism for calculating c, then an adaptive mechanism for calculating c, then to substitute c altogether for a mechanism, and so on and so forth.

Deepening KISS and its underlying methodology allows for the iterative refinement and exploration of all the objects in the undertaken scientific questions. Hypotheses, theories, conjectures, programs, models, simulations are all situated in complex design spaces, which, together with the modeller (and even stake-holder, see participatory simulation [113]), are explored to find the best combinations to allow an in-depth understanding of the target phenomenon.

The 'Deepening KISS' approach suits the needs of data-driven modelling, as more and more data and complexity can be injected into the ABM, in an iterative process. The approach is not necessarily as specific (non-abstract) as KIDS, but it can reach the same level of complexity. However, whereas beginning with a simple model and increasing complexity (top-down) the researcher can always modify the model to cover a new aspect, to begin with a complex model which must be simplified (bottom-up, corresponding to KIDS) can be counter-productive: if the initial complex model missed a relevant aspect, all the derived models would miss it as well, with no expectations to fix it.

This approach plays a role between the two others, so both Data gathering and Validation can be done as in KIDS (intensively and validating against data), or as in the typical KISS (low-profile data collection and sensitivity analysis validation).

2.5 Filling the holes: new Stages

2.5.1 Uprising Collected Data

By refining and experimenting with data-driven modelling, an alternative logic of simulation is proposed. Its novelty relies on its focus on collected data, as it enriches it with new dimensions and applications not considered in the classical approach. In the classical logic of simulation presented in section 2.3.1 the data gathering could be done after building the model and the simulation, because it was used just for validation. However, in the reformulated diagram presented in Fig. 2.5 the new arrows represent a turn in the sequence. The new flow forces the data gathering to take place before the simulation. This is due to the two processes represented by the new arrows: the influence of collected data in the design of the model and the initialisation of the model based on some of these data. The model building is carried out by the Abstraction, Data-driven design and Initialisation activities. Then, the simulation can be executed and the output obtained. The last stage, the validation process, must be done with data not used previously in initialisation. There may be a need for feedback and modification of the model again (for example, the results of the Validation stage may force changes to the design of the model).

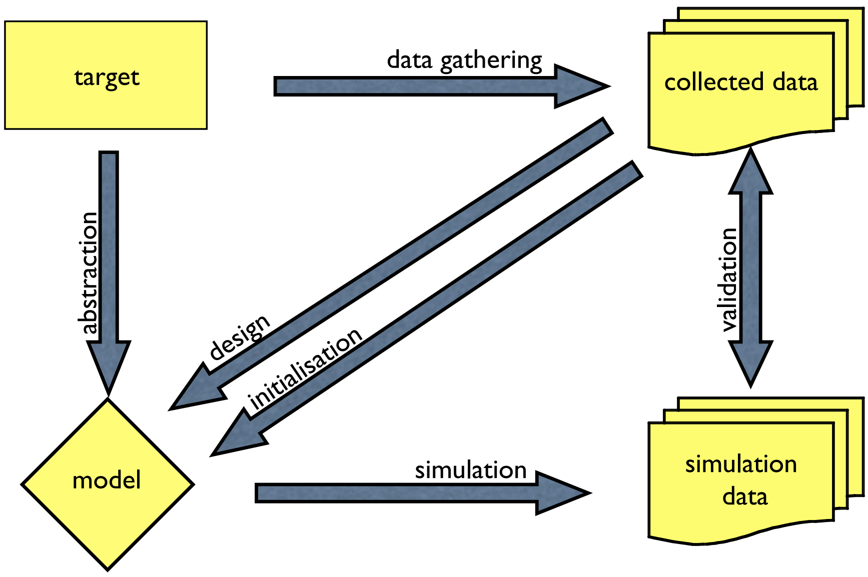

FIGURE 2.5: Diagram showing a modified data-driven logic of simulation.

2.5.2 Data-driven Design

Forcing the Data gathering activity to be previous to the Simulation execution allows it to participate in the model design, together with the Abstraction process, enriching it. Therefore, the collected empirical data will provide useful insights of the behaviour of the modelled phenomenon. For example, a systematic quantitative compilation can determine agents attributes, or qualitative process descriptions may be useful for their micro-behaviour. They could also allow to quantify and limit the range of the model parameters.

For instance, the Anasazi civilisation example [61] uses several data sources regarding the environment to model the exogenous factors successfully. The water demand model of Galán et al. [95] takes into account large quantitative databases to initialise and design the micro behaviour of the agents (households and families). Other more theoretical-type models use empirical data just to find a numerical correlation among some variables, such as [229].

Moreover, a pre-processing of the collected data can feed the model with interesting findings. For instance, clustering can be used for identifying some kind of relationships to implement, or point out patterns to study in depth [126]. This will be further explain in section 2.7.

2.5.3 Data Initialisation

The initial conditions of a simulation constitute a key element on simulations. After the model is carefully designed (through the Abstraction and Data-driven design activities), it must be initialised in some way before the actual execution of the Simulation. However, it rises some issues that should be taken into account. Should it be initialised with random distributions, or with data collections? Which are the benefits of each? Can the effort on collecting data lead to a significant qualitative improve in the simulation output?

Although agent-based models frequently aim to simulate some real world phenomenon, their initial conditions usually do not attempt to reproduce the real world. Most often, the simulation begins with values taken from an uniform random distribution. But there are many cases where the choice of initial conditions can affect the output of the model and where a uniform random distribution is a poor choice.

The common application of this uniform random distribution follows a typical procedure, in which a series of simulations (each with a different starting random seed value) is run and their outputs aggregated into a mean. This is an appropriate method to check the relationships among a set of parameters in a model. However, it does not ensure that the output cannot be improved with other initial conditions, especially when there is a need to compare with real systems and precise data.

To understand why, assume that at least some observable elements of the real world are stochastic. Then the instance of the real world that actually exists can be thought of as a random selection from a population of possible worlds. This means that, while the most probable case is that the real world has the same attribute values as the means of the values in all possible worlds, it is also quite likely that the real world value is not close to the mean and certainly possible that it is an outlier, far from the mean. Now suppose that due to a happy chance, the researcher has a model that accurately represents the real social processes. Then, the model is initialised with random conditions, executed many times and its mean behaviour calculated. Then, this mean behaviour is compared with the behaviour observed in the real world. There is a chance that the two will not match. If the real world happens to be an outlier, the discrepancy could be very large. On the other hand, if the researcher starts with initial conditions that are taken from data, even if the real world is an outlier, the data will to some degree move the model in the direction of the real world, and such researcher will much more likely find a match between the model and the observed data.

One specific case where uniform random initial conditions are well known to be inadequate is the distribution of links between agents to form a network. Real social networks invariably display a higher degree of clustering than random graphs [24] and a number of algorithms are available to generate networks that more closely match empirical degree distributions [190].

An example of how an agent-based model can be improved by introducing data, in contrast with the random approach to initialisation, is a study of the Eurovision song contest [97]. This analyses voting patterns in this music contest, and begins with the hypothesis, ‘over a sufficiently long period of time the results of the Eurovision contest would approximate to random’. If the hypothesis were true, a simulation with random initial conditions and random voting schema should approach the real situation. But actually it does not. It is shown that introducing empirical data, such as the distance between countries (if a country is closer, people are more likely to vote for it) or a measure of the similarity of their cultures, improve the results of the simulations.

Thus, it is clear that data initialisation is a key step of this data-driven approach. It is specific for data-driven models, as it would not make sense in a theoretical model with a very high level of abstraction. Some examples have been already commented in 2.2.2, such as Galán’s water demand model [95] that loads databases into the agent-based model, or the Hedström unemployment model, which uses standard surveys [134]. Another interesting example is [164], that chose to analyse several empirical studies on German General Elections to extract a data-based distribution for initialising the model citizens.

However, this is not such a simple process, as it should be very much taken into account in the design. Is there enough collected data to feed the ABM? Can the agents population characteristics be loaded from external sources, instead of randomly generated? The amount of data that is possible to be loaded covers from agent attributes (e.g. loaded from surveys) to their relationships (in the cases in which it is available, like in online social networks), from their micro-actions (e.g. loaded from qualitative formalised data) to their distribution in space (through Geographic Information Systems). A special type of data, statistical surveys, are an important tool for quantitative sociologists. Their usefulness is out of doubt, and although they are intensively used in other simulation techniques (such as microsimulation), they are still far from being exploited by the ABM community. The following section will focus on the different data types and their usefulness, specially concerning the initialisation process.

2.6 A Step Forward: Introducing Surveys into ABM

2.6.1 From Microsimulation to Agent-Based Simulation

This section tackles how the ABM could be enriched using several concepts and methods of an essentially data-driven simulation type called Microsimulation.

The data-driven examples described in 2.2.2, together with the justification of a Data Initialisation step of 2.5.3, show how the issue of random initialisation can be addressed successfully by gathering data and feeding the model with it. This import of data for initialisation has some similarities with a technique called Microsimulation (also known as micro-analytic simulation) [112]. Microsimulation is a type of simulation that focuses on the behaviour of individuals over time. It consists on a process that begins with the initialisation of the individuals with empirical data (usually derived from a large sample survey). The simulation consists of repeatedly changing the simulated individuals according to a set of transition probabilities and transition rules (ideally, both extracted from empirical data). Afterwards, the aggregated effects of the treatment on the sample are observed, to infer the effects on the whole population. However, microsimulation does not model interactions between individuals, each of whom is considered in isolation.

Microsimulation has traditionally been used in areas where it is easy to obtain quantitative data, in the form of surveys and censuses (for the initialisation of individual units) and equations or rules (for defining agent behaviour). Although microsimulation has been successful in some problem domains such as traffic modelling and econometrics [36], it has been quite difficult to apply in social domains that are not so well structured or where there are important dependencies between agents [248]. Microsimulation is unable to model interactions between agents, an area where agent-based modelling is pre-eminent. Nevertheless, some aspects of microsimulation, such as basing the simulation on representative survey samples and using probability transition matrices to determine changes in the values of agent parameters, can usefully be applied to the design of agent-based models. Agent-based models usually follow an event-based rules approach rather than using transition probabilities. However, the limitations of modelling or the lack of sufficient data frequently make difficult to implement explicit rules and therefore they have to turn to other solutions, one of which is to use transition probabilities, which represent implicit rules. Qualitative information, although rarely used in agent-based modelling, can also be introduced [264, 265].

2.6.2 Simulation & Surveys: a Tortuous Relationship

Statistical surveys have been widely used in quantitative Sociology, being nowadays its main tool. On the other hand, microsimulation uses them intensively, as explained in the previous subsection. However, their application in agent-based social simulation is rare, even for the initialisation stage. The following literature review tries to clarify its spread and the most relevant examples of survey use in general social simulation.

Microsimulation provides several examples of the use of surveys in simulation. For instance, Brown and Harding [36] show how the data of Australian Bureau of Statistics surveys or the National Health Survey (NHS), both within the Australian context, can be imported to models that may guide government policies. The authors underline though, that surveys lack of geographical information or links among individuals. In [1] the authors develop the NHS model where they try to overcome the methodological limitations of surveys in the context of microsimulation. Their interesting method involves additional complementary data sets, in order to fulfil the gaps on family information or health conditions. Other microsimulation examples using surveys can be found in [112] and [117].

Tucker and Fletcher [249] present an example on mathematical modelling, using intensive use of surveys in order to build the needed equations, both for the trends and the parameter values. After building the model on domestic waste management, the output is validated against survey data, showing a good approximation to the general trends.

There are a few works in ABM that are interesting regarding this subject. For instance, Stroud et al. [236] deal with large quantities of data coming from the National Household Transportation Survey (USA) to study disease spread, and therefore needing transport patterns of each individual in each simulated household. However, instead of loading the whole data collection into the model, they just realise a random sample of households over the survey, and only using a small part of the variables that the NHTS provides.

Vindigni, Janssen and Jager [254] introduce surveys in social simulation by means of data mining (DM) techniques. Thus, survey data-sets on organic food consumption are studied and analysed with DM in order to induct behavioural rules for consumer agents.

There are several cases in which surveys are used to validate or calibrate the agent-based models, as it is done in the Mentat model (though this model uses them for other purposes as well). Thus, Moss [186], realises a review of possibilities of empirical validation in the field, including survey validation. Hoffmann et al. [141] make an effort formalising stock market simulation through qualitative and quantitative data, using some survey data from Dutch investors for calibration. López-Paredes and colleagues, in the FIRMABAR project [173], make intensive use of Valladolid urban survey data in order to both calibrate and validate the simulation.

Other works focus on analysing the surveys to extract statistical knowledge that is used in the simulation. For instance, Tucker and Smith [250] infer the forms of the probability distributions describing attitudes from questionnaire surveys of recyclers. A similar way can be found in [204], where the authors used survey data on locational preferences of residents within South-eastern Michigan to characterise the distributional parameters that were used to initialise the agents. Deffuant et al. used multiple data sources, mainly statistical, in the IMAGES project [62] in order to initialise the farmer populations.

Finally, Hedstrom [134] uses another method for importing survey data into social simulation, which is closest to the approach carried out in Mentat. This model about Stockholm youth unemployment imports directly demographic information from surveys to initialise the population.

2.6.3 Handling Data Sources: some Guidelines

Once it has been decided that data will be used to drive the simulation, the next questions are, what type of data, and where could the data be obtained?

It is desirable to have data from some representative sample of the target population. In practice, this usually means survey data from a large random sample of individuals, although it needs to be recognised that large representative samples, while statistically advantageous, also have some disadvantages:

1. If the sample is large, it is likely that the researcher will not be the person who designs or carries out the survey. More likely, the data will come from a government or market research source. This means that the survey will probably not include exactly the right questions phrased in the right way for the researcher's interests, and compromises will have to be made.

2. If the sample is random, it is unlikely that it will include much or any data about interconnections and interactions between sample members, so studying networks of any kind is likely to be impossible. This can be a serious problem when the topic for investigation concerns matters such as the diffusion of innovation, information, or social relationships.

3. Some data are inherently qualitative and not easily gathered by means of social surveys. For example, if one is interested in workplace socialisation (e.g. [265]), a survey of employees is a very crude and ineffective method as compared with

focused interviews, focus groups or participant observation (for more details on these standard methods of social research, see [106]).

Despite these issues, survey data can be valuable. It is particularly valuable when it is collected from panels, i.e. if the same individuals are interviewed at several times at intervals, such as every year. Panel studies are more or less the only way of collecting reliable data about changes at the individual level. Such data are valuable because they can be used to calculate transition matrices, referring to the probability that an individual in one state changes to another state (e.g. the probability of unemployment). With a sufficient amount of data, one can calculate such transition matrices for many different types of individual (i.e. for many different combinations of attributes). So for example, it becomes possible to calculate the rates of unemployment for young men, old men, young women and old women. However, if one tries to take this too far - differentiating according to too many attributes - the reliability of the computed probabilities will drop too far, because there will be too few cases for each combination of attributes. These probabilities provide the raw material for constructing probability distributions that may be used to simulate the effect of the passage of time on individuals.

The importance of obtaining data repeatedly over periods of time have been stressed here. This is because generally agent-based models are concerned with dynamical processes, and snapshots of the situation at one moment in time are of limited value and can sometimes even be misleading as the data basis for such models. While panel survey data are relatively rare compared with cross-sectional data, other forms of data collection about social phenomena are often more attuned to measuring processes. This is particularly the case with ethnography where the researcher observes a social setting or group continuously over periods of days, weeks or months. A third form of data collection is to use official documents, internet records and other forms of unobtrusive data that are generated by participants as a by-product of their normal activities, but that can later be gathered by researchers. Some examples are newspaper reports, web pages, and government reports. In these cases, it is often possible to collect a time series of data (e.g. using the Internet Archive `http://www.archive.org/` to recover the history of changes to a web site) and thus to examine processes of change.

Regardless of whether the data are quantitative or qualitative, it is often the case that they do not have to be collected afresh, but rather that data previously collected by another organisation, possibly for another purpose, can be used. Enormous quantities of survey and administrative data are stored in national Data Archives (European archives are listed at `http://www.nsd.uib.no/cessda/archives.html`) and increasing Archives are extending their scope to include qualitative data (e.g. in-depth interviews) as well (see for example, `http://www.esds.ac.uk/qualidata/`).

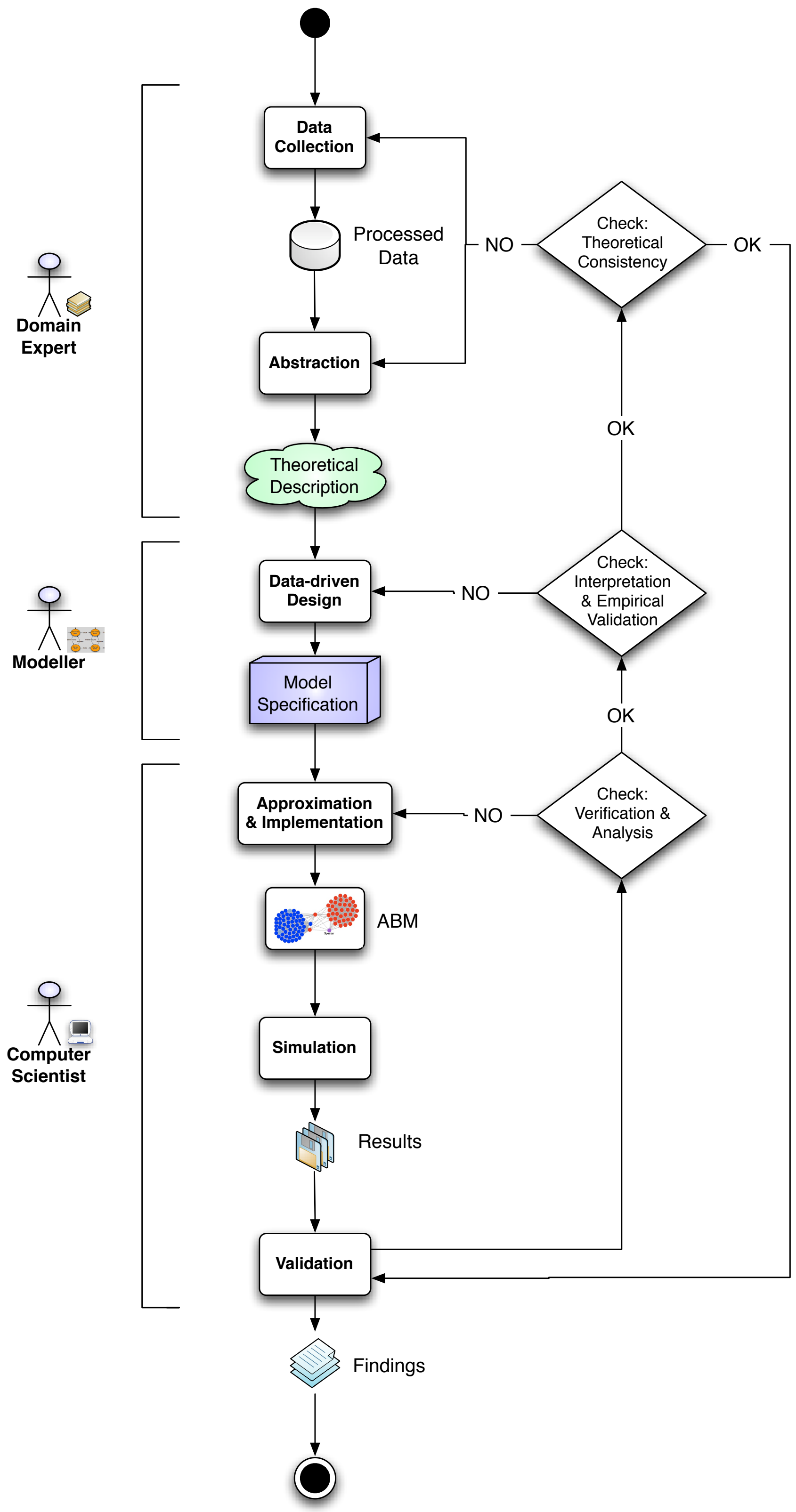

Figure 2.6: Proposed flow diagram for the data-driven agent-based modelling.

2.7 A Data-driven Agent-based Modelling Approach

2.7.1 The Data-driven Cycle

This methodological chapter has carefully extended the logic of simulation, defined a modelling approach appropriate for data-driven modelling and identified ways and methods for injecting data into simulation. However, a method integrating the whole perspective must be defined still, with specific guidelines to follow when applying the methodology to other models. This is the purpose of this section, that draws a collection of diagrams describing the data-driven cycle (partially inspired by the generic approaches of the works [74, 76, 94]), including the different tasks of the Domain Expert (or thematician), the Modeller and the Computer Scientist, together with the role of the Theory and auxiliary techniques that can be used.

The proposed agent-based modelling process is therefore driven by data. Thus, data play a key role that influence all the activities. Some of them are defined specifically for collecting and preparing data, and others make use of these data, for instance for building and initialising the model, or for its validation. In the Figure 2.6 the iterative cycle of interrelated activities is shown, together with the researcher roles implied in each of them. Thus, the Domain Expert should guide the Data Collection, and based on the observed data build a theoretical description through an Abstraction process. Afterwards, the Modeller builds the Model Specification based on such abstract description, also taking into account the empirical data available. Then, the Computer Scientist may program the actual agent-based model and simulate it. The Validation process implies tasks of the three roles, and depending on its success it might involve new cycles until obtaining the Findings which will be incorporated to the theory.

Next, each of these activities is detailed in another diagram. Thus, the main activities are described as follows:

Data collection from different sources. Note that there are different kinds of data, that can be classified in qualitative data (such as interviews or participant observation), or quantitative data (statistical equation-like or from samples such as surveys or panels). It includes primary data directly extracted from the real world (the target), or secondary data extracted from other researches empirically grounded. The task of selection of which data to collect, and how to do so, is naturally and unavoidably driven by the researcher actual perspective, theories and hypotheses, as explained in section 2.3.3.

Figure 2.7 details the whole process of data collections. First, the domain expert must select in which way the empirical data will be collected. This is dependent on the

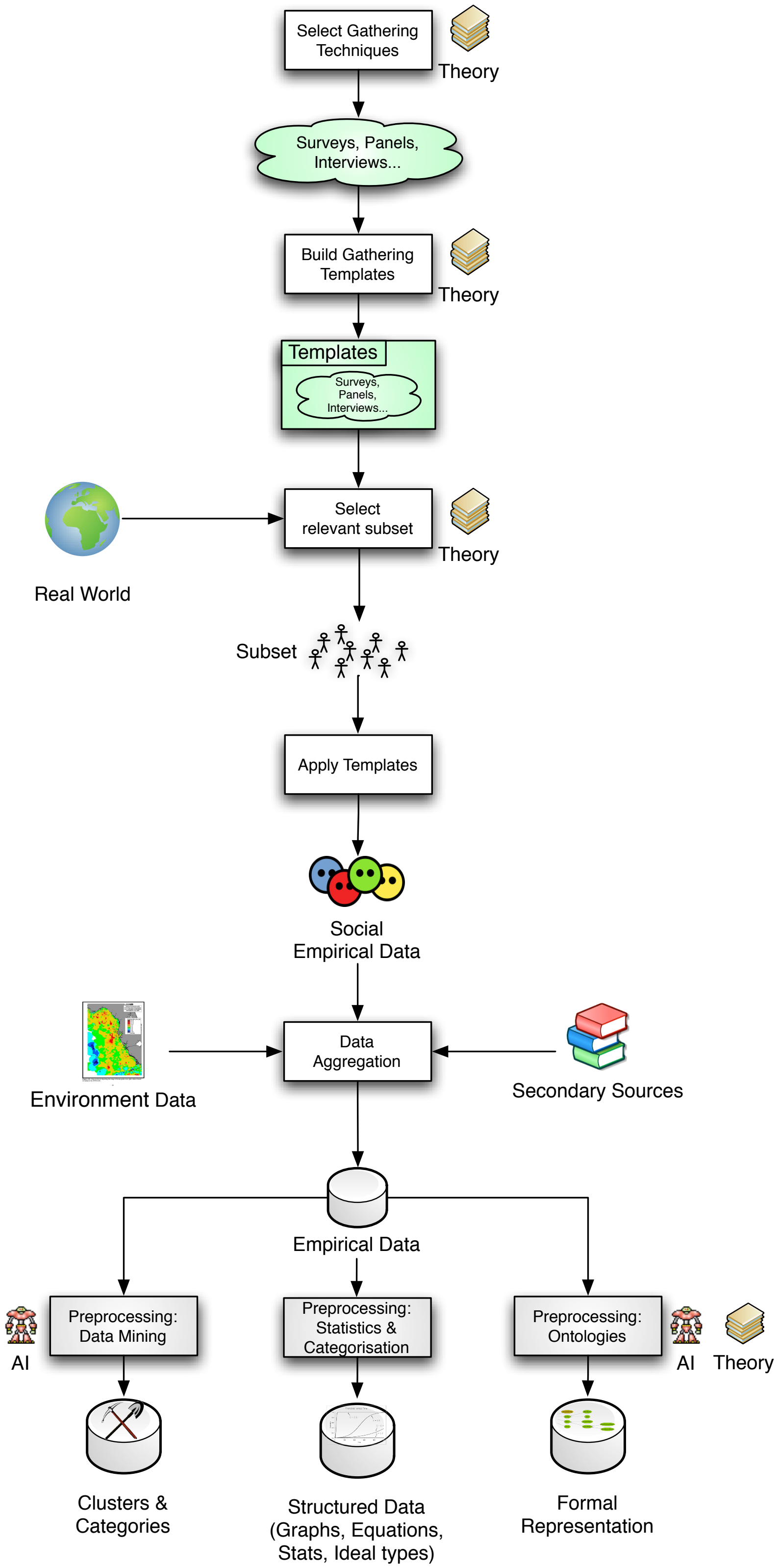

FIGURE 2.7: Sub-diagram concerning the Data Collection stage.

chosen approach (quantitative, qualitative) and the view of the world of the researcher[4] (positivist, ethnocentric, interpretativist, etc). Thus, a *gathering technique* or *method* will be chosen, such as a survey, a group of unstructured interviews or participatory observation. Then, the *templates* must be defined, that is, in the case of the survey, to write the survey itself, or in the case of an interview to specify which issues should be tackled. Afterwards, the representative sample, a subset of the population under study, is chosen in order to apply the chosen method with the templates. In the case of a survey, this would mean performing the survey over the random sample of individuals, while in a group of interviews would be to choose the representative people and carry it out. Therefore, the raw *empirical data* coming from the society studied is obtained.

Other sources may be also considered. *Secondary sources* such as research studies, documents or records generated by others (see section 2.6.3 for more information), could end up being as useful as primary sources [224]. Besides, additional data describing the environment, such as GIS, may be considered depending on the problem [95].

Several operations can be executed over the empirical data, that is, carrying out a *pre-processing* of the data. This is needed in order to structure the data to facilitate conceptualisation, abstraction, and correspondence with the model. Statistics or qualitative categorisation would provide such structured data in order to be able to handle it later. Besides, several AI technologies could be helpful in this step, such as Data Mining (extracting clusters from large collections of data) or Ontologies (which can help to organise the knowledge extracted in a hierarchical manner), as explained in section 2.7.2. Note than even though the whole Data Collection is assigned to the Domain Expert, dealing with AI technologies is a task of the Computer Scientist, who would act as an assistant.

Abstraction. After extracting knowledge from the empirical data, the domain expert must produce the first conceptualisation of the social phenomena under study, as shown in Figure 2.8. Following Galán et al. [94], 'this job involves defining the objectives and the purpose of the modelling exercise, identifying the critical system's components and the linkages between them, and also describing the most prominent causal relations'. This process of abstraction based on the considered theories and hypotheses, and taking into account the pre-processed data from the previous stage, would lead to an exhaustive *Theoretical Description* of the model.

Data-driven Design. A *Model Specification* must be derived from the *Theoretical Description* by the Modeller in the way defined by Figure 2.9. Thus, the Modeller is the responsible of formalising the description from the Domain Expert, but this is not

[4]This is an issue that appears in every discipline which requires a data collection process, not only in Social Sciences.

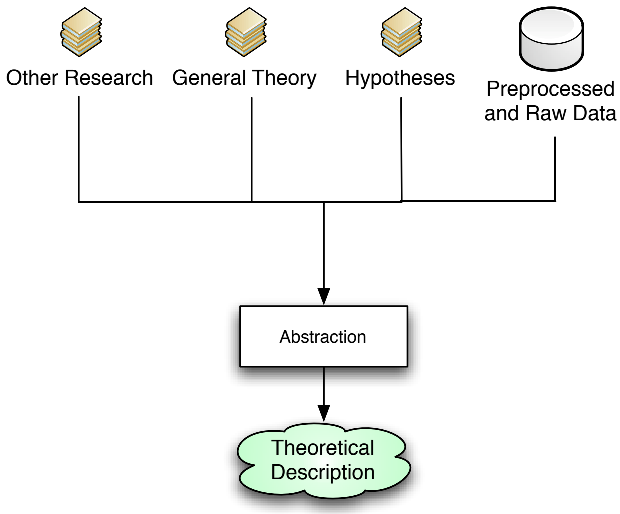

FIGURE 2.8: Sub-diagram concerning the Abstraction stage.

a clear and direct process. There will be multiple decisions to be made: due to the ambiguity and abstraction (and even inconsistency) inherent to natural language, there will be more than one formal model specifications which can be derived from a single theoretical description [94]. Thus, sometimes this task would involve (if it is possible) communication with the Domain Expert in order to clarify or extend the provided description.

In order to build the model specification (ideally with pseudo-code), considering that it is a 'data-driven design', this process must be grounded on empirical data. Thus, pre-processed data can be useful here, especially in an aggregated and structured form. The hierarchical representation given by the ontologies (potentially used in the Data Collection stage) can be of significant help in this stage [47], together with any clusters found in the Data Collection. On the other hand, a theoretical ingredient that could be considered is the set of templates defined by the Domain Expert in the Data Collection. Those templates concentrate the relevant information that must be taken into account, such as the variables considered in the survey. That is, the structure shell of the raw data, which may give an insight for the particularisation of the agents.

As it is described in-depth in section 2.7.2, again several AI technologies may prove useful depending on the problem. During the design of the agent-based model, when dealing with qualitative concepts typical from Social Sciences, fuzzy logic may be an option for introducing them in a formal manner. On the other hand, if the main issue is the adaptive behaviour that the agents should show, and if it is possible to increase the complexity of the agents, neural networks can be considered due to their potential in adapting behaviour [110].

The model specification which appears in Figure 2.9 is a complete specification, but it is possible to extend it for further improvement. Genetic algorithms might be used for optimising the agent behaviour, if there are several alternative strategies to choose from and there is a function or criteria to optimise [213]. Integrating a Natural Language Processing module for a representation of the output for informative purposes could be interesting in certain contexts [170, 188]. Finally, to prepare the simulation output so it can be processed by Data Mining techniques would be certainly useful in many cases, due to the large quantity of micro output that can be generated in a complex data-driven ABM [73, 126].

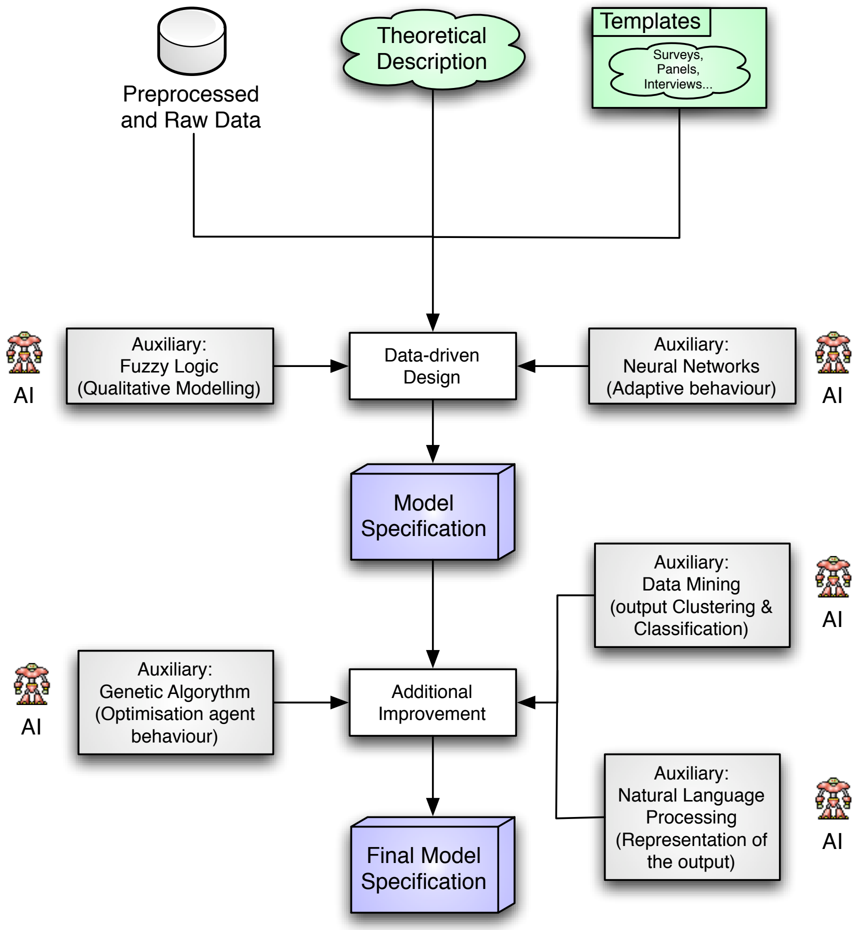

Figure 2.9: Sub-diagram concerning the Data-driven Design stage.

Implementation. Once the model specification is ready, the Modeller's tasks have finished and the Computer Scientist role should take over. First, the Implementation stage of Figure 2.10 is carried out. The Computer Scientist should not make any modelling or theoretical decisions, as the model specification should be completely clear and formalised (ideally with pseudo-code), with no inconsistencies or ambiguity. However, there are technical decisions to be taken, represented by the '*Approximation*' step in the figure. That is, to move the specification from the theoretical domain to the computational one. For instance, there is a subtle difference between a real number (potentially infinite... but computers cannot deal with infinite directly) and a floating point number, which is an accurate approximation. Galán et al. [94] shows another more elaborated example: 'The Navier-Stokes equations of fluid dynamics are a paradigmatic case in point. They are a set of non-linear differential equations that describe the motion of a fluid. Although these equations are considered a very good (formal and fully specified) model, their complexity is such that analytical closed-form solutions are available only for the simplest cases. For more complex situations, solutions of the Navier-Stokes equations must be estimated using approximations and numerical computation [220]. Deriving such approximations would be the task of the computer scientist's role.' Afterwards, the implementation will finally produce the complete agent-based model, ready to be executed.

Note this process of implementation can be automated using Model-Driven Development techniques, in which the tool would derive the code from the exhaustive model specification [198].

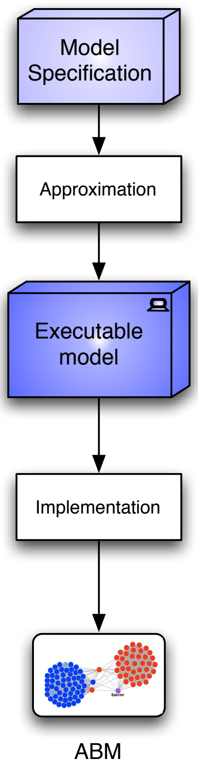

FIGURE 2.10: Sub-diagram concerning the Implementation stage.

Simulation. This is the attempt to explore the behaviour of an abstract model through the means of the execution of a computer program. This activity. though, also involves the exploration in the parameter space throughout the multiple executions from different initial conditions, as shown in Figure 2.11.

Concerning the model initialisation, although data structures can be used for the design of the model, it is actual data what should be used to define the initial population of agents and help in the initial parametrisation of the environment.

The process of '*exploration of the parameter space*' or '*introducing variation*' [5] consists on executing the model testing different scenarios and conditions in both individual and collective mechanisms and measures. If possible, to perform a sensitivity analysis (see section 2.3.4) would be recommended, but this is frequently not feasible in complex data-driven models. Thus, a selection of the relevant scenarios to simulate it is advised.

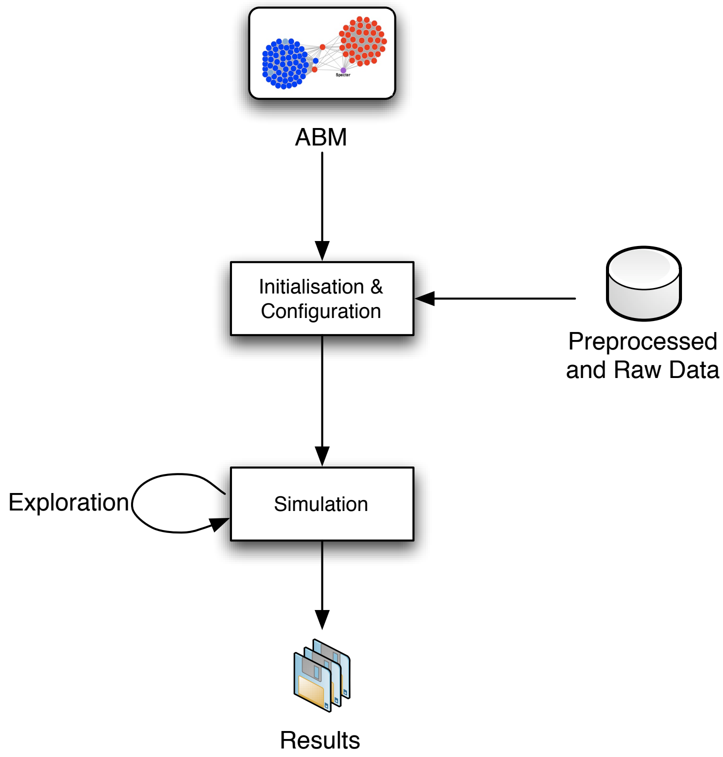

FIGURE 2.11: Sub-diagram concerning the Simulation stage.

Validation. The validation stage is considered fundamental, as explained in section 2.3.5. Figure 2.12 shows the steps that this stage can involve. Besides, the general Figure 2.6 shows those same steps but specifying which role takes care of each one, as all of them are implied in the process. The same diagram specifies the loops needed if any of the checks fail.

Thus, first a *Verification and Analysis of the Results* is carried out. An error here would lead the Computer Scientist to revise the implementation in order to find the bugs or structural mistakes that caused the problem. It is essential to verify the robustness and consistency of the output (not just the final state but also the evolving behaviour

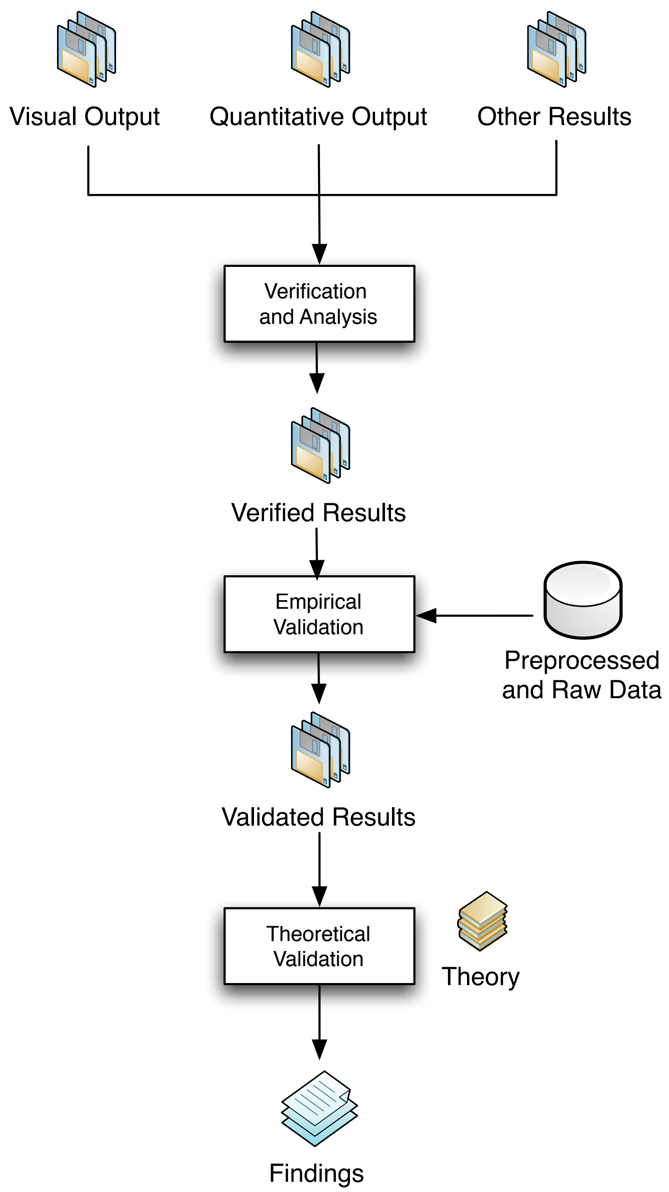

FIGURE 2.12: Sub-diagram concerning the Validation stage.

of the simulation) with the model specification and expected results, especially in limit cases. Afterwards, the Modeller realises an *Empirical Validation*, interpreting the results of the ABM, verifying that the program does what it was planned to do and comparing them to the empirical data. If there are problems in this step, it is the responsibility of the Modeller to re-design the model specification in order to improve the results. Finally, the Domain Expert carries out a *Theoretical Validation*, checking the consistency of the results with the theory and hypotheses through an interpretation process. Problems in this level, after overcoming the other checks, may lead to changes in the design or even re-definitions of the theoretical description of the problem. If important lacks in the model are found, it might be necessary to go back to the Data Collection stage and search for additional data, sources or research studies.

After all the validation steps have finished, the validated and interpreted results turn into *Findings*, which may be end up as a contribution to the current theory available.

2.7.2 Merging Artificial Intelligence Technologies

As it was shown in the previous section, collected data usually need some kind of treatment before it can be directly useful in the design stage. Moreover, there are multiple issues concerning ABM that should be addressed carefully, specially in the case of data-driven simulation. Social phenomena usually follow a smooth behaviour, which is hard to represent with standard programming algorithms. *Soft Computing* techniques [217] including neural or bayesian networks, fuzzy systems, and evolutionary computation can be helpful. Depending on the case, one or another can be used with promising results. For example, neural networks are useful for adaptive learning behaviours; fuzzy logic is helpful in modelling social processes; evolutionary algorithms usually substitute agents as another way of doing simulation, but they can also be used to optimise agent behaviour.

Other issues that can be addressed are the search for patterns and characterised groups (clusters) in the input data or in the simulation output. The larger the amount of data are, or the more complex the phenomena represented are, the more difficult is to find patterns and clusters. However, there are several AI tools, such as classifiers and data mining, which can make it rather easy [260]. Representation of the concepts is another complication. Ontologies represent an easy-to-handle interface with experts, and a formal view that can be inserted into the agent-based model. Natural Language Processing can be proposed for a better representation of the simulation output, prepared for non-experts. In the following paragraphs the utility of some of these techniques is further explained.

Fuzzy logic is oriented at modelling the imprecise modes of reasoning in environment of uncertainty and vagueness [267]. Thus, because vagueness is such a common thing in the social realm, it provides a useful way to handle this vagueness systematically and constructively [233]. There is an increasing interest among social scientists for adding fuzzy logic to the Social Science toolbox [157]. Even though it is still incipient, there are some examples of linking fuzzy logic with social simulation [79] (this is further discussed in section 5.4). Unlike traditional multi-agent models, where the agents are an over-simplification of real individuals, fuzzy agents take into account the stochastic component of the human behaviour.

Due to the ability of fuzzy logic for handling natural language (turning sentences like ‘A is quite stronger than B’ into functions) it has an important value when the modeller wants to import qualitative information into the model. Fuzzy logic is helpful for representing the narrative explanations of the expert or qualitative researches [79].

Data Mining is the process of extracting the patterns and relevant information from large amounts of data (such as databases and surveys). This technology can be used in several stages of the simulation process, as stated in [126]. First, it can manage those amounts of quantitative data and define clusters based on specified conditions. This extraction of macro patterns represent a rather complete analysis itself, and its results can be inserted into the model.

Moreover, the same process can be realised with the micro-output of the simulation, in case the output is adapted to provide a log file that records the state of each agent (or at least their final state) throughout the simulation period. This log file can be analysed to determine if the agents' states evolved in a consistent way. Data mining techniques can be applied to capture hidden patterns that are non-visible graphically. Furthermore, the clusters found before and after the simulation can be compared with each other to extract new conclusions. Those representative clusters have an analogy with the ideal types of Qualitative Sociology, and can be compared with this kind of research studies. Besides, data mining results may be helpful for the redesign and optimisation of the model, dropping redundant or spare aspects [126]. See section 8.2.3 for further information.

Ontologies are formal representations of sets of concepts in a specific domain, together with the formal relationships between those concepts [217]. Their immediate use is to classify the theoretical concepts and data. There are several ways of inserting them into the model [47], and through the use of an inference engine it is possible to deduct new knowledge during the execution. Besides, ontologies are closely associated with similarity functions, which may prove useful to model homophily dynamics [183]. Moreover, this classification can be mixed and contrasted with the clusters from data mining.

Natural Language Processing (NLP) [217] is an important sub-field of AI. It studies the problems related with automated generation and understanding of human languages. NLP can transform text samples into a structure suitable for the computer, allowing the direct import of qualitative information. However, the deep limitations of this technique nowadays have forced us to look to other part of the NLP: the natural language generation.

It is possible, after recording all the events and actions of every agent (as for the data mining case), to generate biography-like texts of a subset of them, enriching the output of the system [129]. These narrated life-stories would expose in natural language the micro behaviour of the agents. This friendly readable content would be useful for the non-technical audience, such as stake-holders, as shown already by Moss in [187]. If the agents described are selected among the most representative, this output would allow qualitative comparisons with other research studies, specially in the case of qualitative ideal types. A method for choosing the most significant and representative agents could be the use (again) of data mining clustering techniques, choosing the agents more similar to the ideal types found (see section 8.2.1 for an in-depth analysis of this issue).

2.8 Discussion and Difficulties

The introduction of empirical data in ABM has multiple advantages, as it has been mentioned. For instance, it is expected to improve the fit to the target behaviour, it allows more expressive and descriptive models, a more systematic validation and even the possibility of using several AI technologies to improve several aspects. However, it also implies some costs:

- In some cases, (such as very abstract and ambiguous models or when there is not available appropriate data to be used) complicating the model with empirical data will not benefit the results. Then, a KISS model would be the ideal approach. It is advisable to use empirical data but the way it can be applied in the different activities of the ABM process depends on several factors such as the model aim, type of available data, its structure, its reliability, its quantity, its ability to be processed, etc.

- The ease of understanding and communication associated with KISS is partially lost with this kind of modelling. However, a well-defined modular specification should be helpful in this sense. Moreover, in the 'Deepening KISS' approach (see section 2.4.3), the structured gradual process of increasing complexity facilitates understanding together with extensibility.

- Data-driven modelling demands a special effort in gathering data. Although this process is frequently required for validation, it may not have the intensity required here. The additional costs may not be worthwhile in certain cases (such as small models not needing a high level of descriptiveness). Moreover, validation of very abstract models can be structural (theoretically-based) or through a process of sensitivity analysis, not requiring a deep comparison with collected data. In those cases, a data-driven approach implies a high cost that may be difficult to justify, in spite of the expected improvement of results.

- In subsection 2.6.3 several specific difficulties related to the procedures have been addressed: surveys not providing exactly the required data; lack of information; qualitative or subjective data not easily gathered. Besides, if the data are extracted from several sources, it can be quite difficult to match it: different indicators, data not complementary or even contradictory. And handling large amounts of data makes still more complicated the process of deciding what is relevant. In all those cases representativeness and hypotheses should play an important role.

- Related to the previous problem, there are other fundamental issues concerning collecting data. For some models, especially those at a high level of abstraction, appropriate data may be impossible to obtain. Another problem, common also to microsimulation, is the requirement for large volumes of detailed data about individuals. Sometimes, the lack of data stems merely from the absence of suitable surveys and other data sources. But other times, the problem is more fundamental. For example, agent characteristics such as their emotional states are unobservable. In some models, the agents' current state depends on their previous circumstances (this is the case, for example, in models that incorporate path dependencies, or where agents have memory). However, it is rare for such histories to be recorded systematically in representative surveys. Panel studies may be a solution, but they are not common.

- A related issue is the need for dynamic data that measures changes over time in addition to the more usual 'snapshot' data sets typically available from surveys. It is also often hard to obtain information regarding networks and micro-interaction processes, unless one is dealing with very particular domains such as virtual communities where data are recorded as a side effect of electronic interactions [216]. Some of these problems can be overcome or worked around. For example, if modellers want to simulate a married couple, they can find a wife in a survey based on a random sample of individuals, but they also need an agent to represent her husband. Since the data are taken from a random sample, it is unlikely that the husband will also be in the survey. Strategies for dealing with this include creating an artificial 'husband', not based on anyone in the sample; or 'marrying' the woman to a different married man in the sample.

- About the technologies mentioned in 2.7.2, each one can be useful only in a limited range of cases. For instance, data mining needs large amounts of data to be effective. Fuzzy logic requires blur properties or concepts to deal with. Ontologies may be useless in cases where the classification is too simple. The output in natural language can be considered non-crucial for the implementation effort that it requires, although there are already several tools for NLP that could be useful depending on the context.

Apart from the recommendations presented to avoid these issues, some other general suggestions can be made, attending to the experience accumulated following this data-driven methodology:

- It is valuable to explore the problem background, focusing not only on the theoretical literature, but also on the availability of data.
- It is worthwhile to compare different collections of data and conclusions from diverse sources to give a stronger foundation to the model.
- In case it is needed to merge data from several sources, it is important to avoid inconsistencies. In all cases, representativeness and relevance are important criteria in selecting data manipulation procedures.
- The most valuable data are those that provide repeated measurements, preferably taken from the same respondents (as in a panel survey).
- The agent-based model should be designed so that it generates output that can be compared directly with empirical data.
- If the data are available, it is recommended to simulate the past and validate with the present, as it has been done in the case study (as explained in chapters 3 and 7).

2.9 Concluding Remarks

This chapter began with the classical debate around the abstraction level needed depending on the modelling aim. Along the different sections, the data-driven approach has been described and justified, making special emphasis in its methodological implications. Thus, the generic logic of simulation has been analysed in order to extend it for the data-driven case. The two opposite perspectives that emerged from the debate in the social simulation field were exposed: KISS, starting from simplicity, and KIDS, from full descriptiveness. A middle way approach has been defined, based on the iterative ‘deepening’ on model design details.

The injection of data [130], and especially survey data, into ABM, has been tackled, taking into account improvements from a special kind of simulation (Microsimulation), strongly data-driven. Therefore, a complete data-driven cycle has been specified and discussed, in which empirical data are not only used for validation purposes, but transversally in all the activities. Besides, the approach is complemented considering the different uses that Artificial Intelligence technologies can offer for these activities.

The effect of applying these guidelines would be to connect these agent-based models more closely to the social world that they intend to simulate, at the cost of the extra effort and complication involved in injecting empirical data into the simulation.

Chapter 3

Sociological Case Study: Mentat, the Evolution of Social Values in the Postmodern Spain

I finally decided that I'm a creature of emotion as well as of reason. Emotionally, I am an atheist. I don't have the evidence to prove that God doesn't exist, but I so strongly suspect he doesn't that I don't want to waste my time.

Isaac Asimov, 1982

3.1 Introduction

3.1.1 Introducing Simulation

Although this work is obviously framed in the Artificial Intelligence field, it works in a multi-disciplinary area where it could be approached in several ways. If it is considered as a 'Computational Sociology' Thesis, it is a must to consider and analyse its sociological implications and potentials. This chapter attempts to reflect those, focused on the Social Sciences audience.

Social phenomena are extremely complicated and unpredictable, since they involve complex interaction and mutual interdependence networks. Quantitative sociological explanations deal with large complex models, involving many dynamic factors, not subject to laws but to trends, which can affect individuals in a probabilistic way. According to Parsons [196], a social system is an interrelated and hierarchical set of components which interact to produce certain behaviours. Therefore, the target social system can be conceived as a collection of individuals that interact between them, evolving autonomously and motivated by their own beliefs and personal goals, and the circumstances of their social environment.

Many sociological problems are difficult to be addressed properly with traditional analytical and statistical techniques due to the diversity and great number of factors involved (e.g. evolution of culture), complicated dynamics (e.g. social networks), non-measurable social processes (e.g. psychological processes, world-size phenomena). Those problems are likely to be handled under the scope of the Complex Systems theory [23]. In this scope, agent-based systems have proved to be a proper framework to model and simulate these social processes [107].

The idea beneath Agent-Based Social Simulation (ABSS) is that the researcher may be able to understand this huge complexity not by trying to model it at the global level but as emergent properties of local interaction between adaptive autonomous agents who influence one another in response to the influences they receive [178] (an example can be seen in Figure 3.1). Because of that, the specification of characteristics and behaviour of each agent is critical, in what it can affect the dimensions of the problem under study.

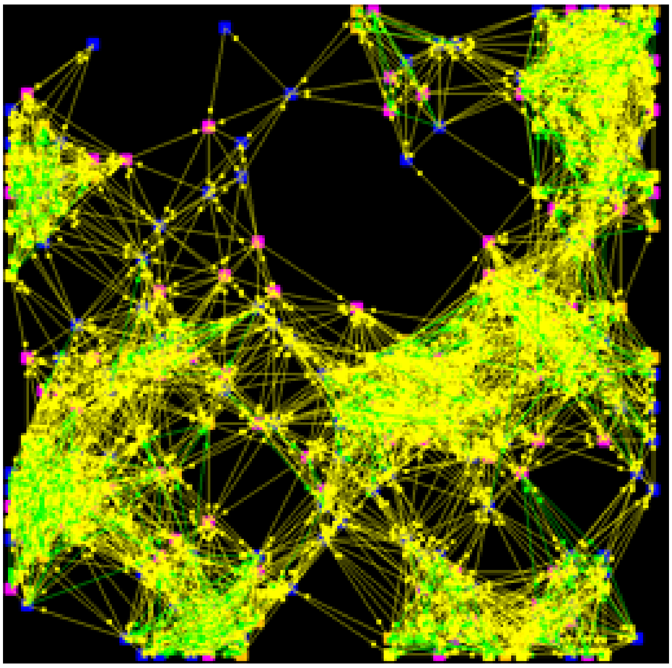

FIGURE 3.1: Snapshot showing emergent clusters in an agent-based simulation.

3.1.2 Some Contributions of ABM to the Quantitative Research

As it has been mentioned, the Social Simulation field could enrich the quantitative sociological research in several ways. Although here some of them will be outlined, a proper insight of the subject can be found in Gilbert & Troitzsch's work [107].

This methodology is naturally adapted to explore theories concerned with social processes, representing their dynamic aspects of change. And as Epstein states, it facilitates explanation and understanding of such social phenomena and can be used to challenge the robustness of prevailing theory through perturbations. Furthermore, it may allow in certain cases to suggest improvements or policy implications through the simulation of different scenarios, discover new questions and reveal the apparently simple (complex) to be complex (simple) [81].

Social Simulation provides the possibility of establishing a reasonable prospective analysis, starting with a single survey wave. One more advantage over other prospective methods is that, for a chosen future moment, an equivalent database is available (the simulation output after evolving till that specific moment), with the analytic benefits associated to it (being able to correlate variables, complicated with aggregated data). In the case of retrospective research studies based on incomplete temporal series, it is possible to get deeper into a specific instant, for a better understanding of data evolution.

Another fundamental point is the comparison between the micro and macro perspectives. It is possible the study and track individual agents, comparing them to the ideal typologies [124]. Also, it can be very interesting to analyse the repercussions of the agents behaviour micro-rules on the macro level along time, that is, it is possible to investigate emergence processes. Or vice-versa, to analyse the effects of macro emergent dynamics on the individual decisions.

The ability of agent-based social simulation to handle complex social networks increments its potential. Social networks empower the system's non-linearity, turning some micro behaviours, together with the nodes/agents interactions (through their links), in a global self-organised emergent behaviour.

The Network Theory, from Physics, provides several ways to study them, and even classifications depending on their topology. Each topology has some associated properties and phenomena, which can result in exciting implications for Social Sciences. A popular example is the property of Small-World networks called 'Six degrees of separation' [256], or the importance of 'hubs' in scale-free networks.

This trend constituted the field of Social Network Analysis (SNA) [255], and formalised innovative methods for such networks, with specialised software built specifically for it. SNA studies not only the implicated actors (nodes), but above all their mutual relations. Or, a step forward, the interactions among the agent internal state and the topologic structure of the network (adaptive networks) change. Thus, due to its development, the study of certain phenomena has been expanded, such as the 'weak ties' and their discovery as essential for the structural cohesion.

In the same line, in the Agent-Based Models (ABM) several dynamics can be studied, which are difficult to handle with other methodologies. For instance, the homophily or friendship dynamics, building up complex interrelated networks. The matchmaking processes, or the temporal evolution of the links strength (together with their emerging and breaking) are feasible to study in this field.

Thus, ABM not only focuses on the prediction of results. Its true potential is discovered in the process of understanding the phenomena, and analysing the multiple (in a way, collateral) effects that emerge in the self-organising social network. In an ABM, there are multiple aspects that could be studied: from the micro behaviour to the social dynamics that the interaction produce; from the space restrictions to the emergent aggregated effects, or even the influence of the macro level into the micro behaviour. Such an insight provides a richer context for further exploration than the classical equation models, even if both can be used as a black-box for plain prediction purposes.

As an added value, the fact of having a computing system which encloses a relatively complex model of the problem under study has another clear benefit: the possibility of integration with other Artificial Intelligence (AI) technologies. The agent system can be expanded depending on its particular needs: neural networks for adaptive micro-behaviour [107]; fuzzy logic to model social processes [125]; genetic algorithms to optimise agents strategies [213]; data-mining and classifiers to extract clusters and new non-visible patterns [126]; ontologies for the knowledge formal representation and communication with experts [177]; natural language processing for an adequate human-readable output in the micro level [124].

To sum up, it is relevant to cite the '16 reasons other than prediction' to build an agent-based model, proposed by Joshua Epstein in the 2nd World Conference on Social Simulation (available in [81]):

1. Explain (very distinct from predict)
2. Guide data collection
3. Illuminate core dynamics
4. Suggest dynamical analogies
5. Discover new questions
6. Promote a scientific habit of mind
7. Bound (bracket) outcomes to plausible ranges
8. Illuminate core uncertainties
9. Offer crisis options in near-real time
10. Demonstrate trade-offs / suggest efficiencies
11. Challenge the robustness of prevailing theory through perturbations
12. Expose prevailing wisdom as incompatible with available data
13. Train practitioners

14. Discipline the policy dialogue
15. Educate the general public
16. Reveal the apparently simple (complex) to be complex (simple)

3.1.3 Objectives of Mentat from the Sociological Perspective

Mentat[1] began as part of the middle-term objective of increasing the usability of ABSS tools for sociologists [221, 222], who are usually not skilled in computer programming. Thus, a logical aim emerged: to analyse and model a complex real sociological issue, with an agent-based model that could lead to a generalisation of the process.

The problem addressed by Mentat is the simulation of the evolution of social values in the Spanish society, in the period 1981-1999. This phenomenon is deeply interrelated with other social processes: the cultural modernisation process, religiosity patterns, political behaviours and demographical dynamics [9, 115, 147]. Thus, the scope of Mentat was quickly broadened, covering the simulation of social values together with other interrelated factors. The computational model should represent the evolution of these social phenomena, as it is described in [201].

Once the model has gone through a formal validation, taking into account the theoretical framework and the available empirical data, the social simulation would allow further explorations of the insight of these processes: approaching non-observable phenomena; dynamic emergent processes; implications of the micro level on the macro level and vice-versa; and pseudo-experimental situations, 'what if...' scenarios, based on the modification of parameters and rules.

Besides, the implemented model could establish the basis for a future application on projections of expected evolution of social values and other parameters in the Spanish society. This would be possible by introducing the most recent data instead of 'simulating the past', as it is done here. However, to make accurate predictions is always a difficult objective, and this model does not try to focus on that. Nevertheless, the modelling has taken care of handling only empirical data that could be available in the current period, and using 'future' data only for validation. In this case, 'simulating the past' is a need: with a historical series showing the evolution of the phenomena simulated, the model

[1] 'Mentat' is a discipline (and profession) in Frank Herbert's science-fiction book *Dune* (1965). Mentats are humans trained to mimic computers: human minds developed to staggering heights of cognitive and analytical ability. Unlike computers, however, Mentats are not simply calculators. Instead, the exceptional cognitive abilities of memory and perception are the foundations for supra-logical hypothesising. Mentats are able to sift large volumes of data and devise concise analyses in a process that goes far beyond logical deduction. They can extract the essential patterns of data, and deliver useful conclusions with varying degrees of certainty. Their main task in the Dune universe is to be socio-political advisors due to their capacities for understanding social processes.

should show, with the same initial data, a behaviour matching the series. Consequent modifications are guided by the coherence degree with this data.

The chosen subject can be considered just a particular phenomenon of a wider problematic: the systemic study of human mentalities. This systemic study of a contemporary open and complex society covers a broad spectrum of socio-cultural phenomena such as values, beliefs and attitudes. This type of research has been traditionally faced through public opinion surveys, but it is rare in the Social Simulation area.The most similar subfield is opinion dynamics (OD) [137], where many abstract models try to study the evolution of opinions and how people interact and convince each other.

However, the approach followed in this work is essentially different for several reasons: OD models rarely make use of data, while here the approach is completely data-driven; OD models are very abstract, usually following a KISS perspective [111], while Mentat does not (as explained in section 2.4.3); opinions are volatile and can be quite easily mutated, while values are rather stable over time; usually, OD models do not use time scales, so it is not provided the amount of time for the shown convergence, but this work is framed in a time and a space.

From this perspective, this work could be used as a base and reference for further sociological studies that may start from previous survey research (together with theoretical frames and diverse secondary sources), and wish to expand their research and analysis conclusions giving it another turn, from the new approach of social simulation. There are several advantages of this step forward, defined in section 3.1.2, such as the increase of formalisation and systematisation of knowledge and conclusions (demanded by the model programming), the proper validation, and the definition of its limitations.

Most of the challenges addressed along this work are common to a wide range of themes, specially those driven by quantitative Social Sciences methods. Thus, the procedures and solutions described could be applied without great effort to several sociological researches, but as well to political science research or public opinion research.

3.2 Theoretical Framework

3.2.1 Data Input: The European Values Study and other sources

In the case of this work (and potentially in other similar ones), the main empirical source is the data file of a survey, while other works were used as secondary sources: demographic information from official census, other surveys, qualitative researches, and theoretical bibliography regarding the subject.

The European Values Study (EVS) [84] has been used to provide the values of the agent characteristics[2]. The EVS is, as its promoters say, 'the most comprehensive research project on human values in Europe' (a question sample in Figure 3.2). It is a survey repeated every 10 years approximately (1981, 1990 and 1999, with the 2008-2009 in progress) in most European countries, making possible to compare several countries or periods among them. An example of how its comparative analysis can be carried out, using the aggregated data for each country, can be seen in Figure 3.3 which shows the distribution of the importance of God all over Europe. In particular, this work uses the Spanish sample of the EVS 1981[3]. This sample[4] is proportionally representative of the whole Spanish society in that moment (not counting underage individuals). This 'old' survey (1981) was used as initial reference to be able to validate the model with the later ones (1990, 1999).

Q1 Please say, for each of the following, how important it is in your life.

		very important	quite important	not important	not at all important	DK	NA
v1	Work	1	2	3	4	8	9
v2	Family	1	2	3	4	8	9
v3	Friends and acquaintances	1	2	3	4	8	9
v4	Leisure time	1	2	3	4	8	9
v5	Politics	1	2	3	4	8	9
v6	Religion	1	2	3	4	8	9

Q2 When you get together with your friends, would you say you discuss political matters frequently, occasionally or never?

1 – frequently
2 – occasionally
3 – never

8 – don't know (spontaneous)
9 – no answer (spontaneous)

☐ (v7)

FIGURE 3.2: EVS question sample. Source: European Values Study Master Questionnaire 2008 Final Version [84].

This way, the agents in the artificial society assume real values of their characteristics, representative of the population, in those theoretically relevant variables: religiosity, values, moral permissiveness (towards several sensitive issues such as abortion), political ideology, and others relevant issues for the social network or context, such as gender, age, civil state, children number, educational level and socio-economic status.

[2] The World Values Survey (with a graphical representation in Figure 3.4 [148]) emerged from the European Values Study.

[3] The EVS survey realisation usually takes more than a year to be completed. Thus, it is not accurate to say either EVS-1980 or EVS-1981, as it was carried out in both years. Therefore, along this work those two years are used indistinctly, but it should be noted that they refer to the same EVS on 1980/1981. The same can be said about the EVS 1999/2000.

[4] It has as maximum limit of random error a ±2.08%, with a confidence level of 95.5% and p=q=50%.

On the other hand, this is complemented with available secondary information, mainly introduced to model the demographic evolution of the Spanish society: life expectancy, maternity age, children per woman, etc.

During the design of Mentat, special emphasis has been made in not introducing any input not available at the moment of the fieldwork (1981). The aim is to contrast the data with 'future' measurements to check if the simulation is 'predicting the past' correctly. If this objective is accomplished, the used procedures could be adequate for inserting more recent starting points.

The only exception to the previous rule was the introduction of some demographic evolution equations (built from real data taken from census and dynamic projections). The reason is that this 'future' information is usually available in the form of demographic projections (in the Spanish case, by the official National Institute of Statistics [152]). These projects usually work very well in short and mid-terms, and thus the procedure followed here still can be replicated with more recent data.

3.2.2 Theoretical Background of the Study on Religiosity Evolution

Western Europe is one of the world regions where the secularisation process has advanced the most, according to the WVS data [148]. In the European context, Spain is one of the countries where this secularisation advance has appeared and accelerated in the last three decades. This was due to the loss of social influence of the Catholic Church in a traditionally Catholic country, and till recently one of the most religious nations in Europe (the current distribution of the importance of God all over Europe according to EVS can be seen in Figure 3.3). It is clear the growing distance among the population and that institution, its dogmas of faith, rules or commands, together with the associated religious practice and thus its loss of influence in other life circles.

However, the classical assumption of considering secularisation as a need in the modernisation process and the modern world as secularised, is losing ground in the theoretical-sociological debate. Nowadays, another alternative has a growing weight: the possibility of the unique character of this process in the current period and constrained to Europe, instead of having an universal scope [60]. This idea is reinforced by the evidence of how, during the last decade, religion importance is growing in the international board: fundamentalism at its very peak; the growing ability of Islamism of generating social identities; the demographical evidence that shows how the religion believers are growing and non-believers are shrinking (as the Third World grows faster than the First) [193].

Thus, as the paradigm of 'secularisation' should be seriously revised, many researchers of religion in Europe prefer to focus on the modernisation process, closely

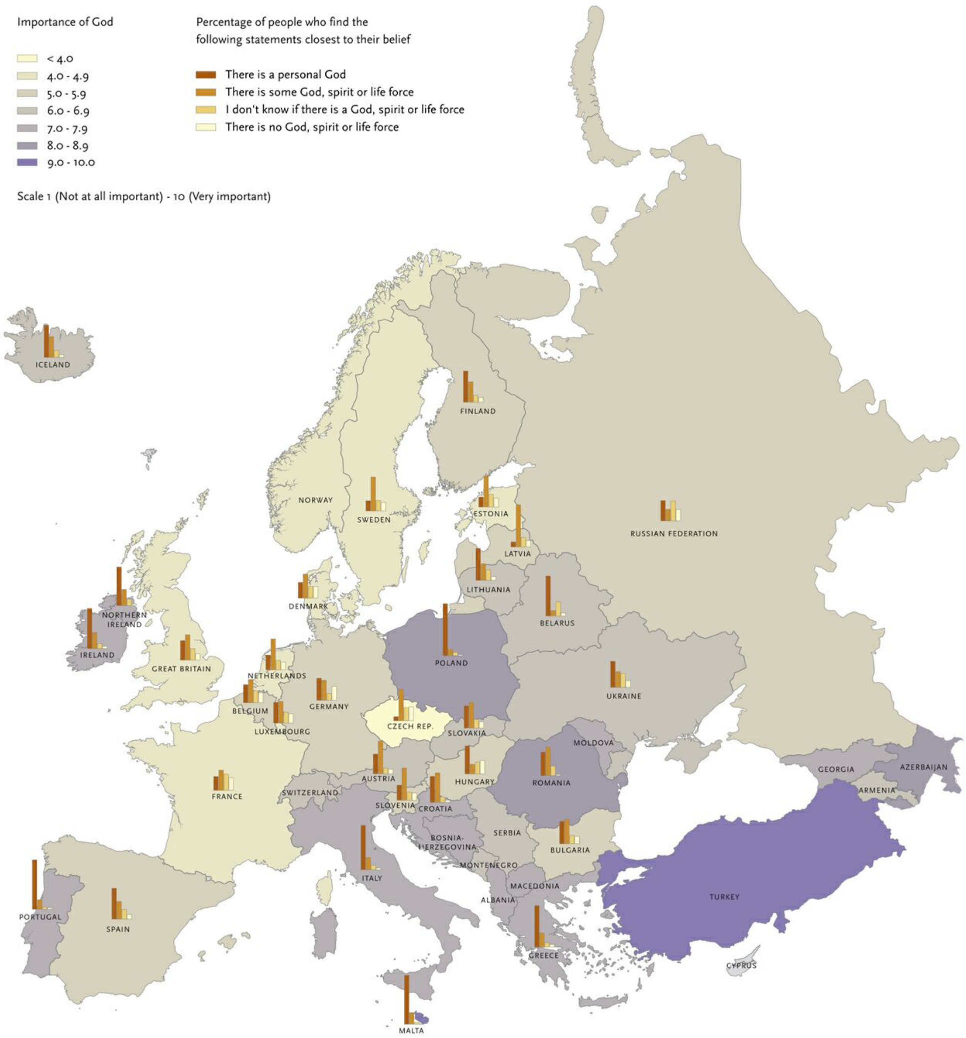

Figure 3.3: An example of the EVS comparative potential, according to a social values question: the importance of God. Source: Halman, L., Luijkx, R., van Zundert, M. 'Atlas of European values'. Brill Academic Pub (2005) [114].

related to secularisation [27, 59]. Following this trend, this work considers that the European modernisation processes are the basis of the religious transformations, specially the strong weakening of traditional religiosity.

The term 'modernisation' is used here without opposing it to other concepts as 'post-modernisation' or 'globalisation' that could be understood as manifestations of an advanced modernity. From this theoretical view, the problematic and under debate secularisation concept moves aside, leaving the focus to the modernisation: the existence of an opposing tension between the Institutional Catholicism and cultural modernisation. Thus, the Church reveals itself as a historically and current social force against the modernisation process [71]. This is clear in the Spanish case, with a high impact on the mentalities all along the second half of 20th Century, due to the building of National-Catholicism[5]. This ideology empowered Catholicism as the legitimating ideology of Franco regime and linked religion with the political, social and moral conservatism. From this perspective, the ‘religious feeling’ in Spain is intimately associated to this conservatism, while the modern mentalities (political, social and moral progressivism) reacted naturally opposing to those religious mentalities. This is the main reason why the study of religiosity in Spain must be associated with the study of political and socio-cultural manifestations.

Modernisation is a transforming force of the ‘religious fact’, regardless of its promoting unbelieving. It affects religiosity, transforming it and producing a pro-conservative reaction [72, 140]. Thus, it is interesting not only to consider the weakening of Catholicism but also to differentiate among the ones that reject modernisation and others that, keeping their religious identity, assume it by restructuring their ‘religious feeling’. This leads to the need of a classification of religious forms, which is fundamental in the modelling of the system.

Thus, four main patterns of religiosity are defined: ‘Ecclesiastical’ (EC), ‘Low-intensity’ (LI), ‘Alternatives’ (AL) and ‘Non-religious’ (NR) [11, 12]. ‘Ecclesiastical’ are individuals relatively close to the hierarchy (Vatican and bishops), catholic practitioners who believe in the Church and attend Catholic mass weekly. The ‘Low-intensity’ ones have a moderated religiosity, with notably less practice. However, they still believe in the Church and go sometimes to religious celebrations. ‘Alternatives’ have a strong religious identity but do not trust the Church, and reject to attend religious services, although they usually conserve their catholic identity. Last, the ‘Non-religious’ are people who do not trust Church neither see themselves as religious people. The majority are atheists and agnostics, but not all of them.

Thanks to the surveys that were done during the last 20 years (EVS, WVS [84, 150]), it is clear that the ecclesiastical pattern has decreased considerably, while the non-religious and alternatives have been increased. Low-intensity increased till 1990 and decreased after till 2000, where they reached the 1980 level [15].

Besides, through the same surveys, other implications can be discovered: EC tend to be traditional and conservative, while the NR are the most modern. However, AL are moderately more modern than the average, and LI are intermediate[15].

Those results concerning the religious patterns confirm the theory that states that cultural modernisation is a fundamental driving force for religious change [193]. It supports the relation between the conservation of religiosity (associated to traditional

[5]In clear symmetry with National-Socialism, allied of the Franco Regime since the very beginning.

mentalities) and the move away from religion (associated to modern mentalities). Moreover, it covers the transformation of such individuals who are open to the modern culture without sacrificing religion.

Thus, due to the importance of modernisation to explain religious and values change, in order to continue, it is compulsory to explain how the concept of cultural modernisation is understood and operated.

3.2.3 Theoretical Background of the Cultural Change

From the point of view of the European Values Study (EVS) and World Values Survey (WVS), the Spanish society is in the middle of a long process of change in values. Such values interact with the religious transformations, promoting the secularisation of society, while the religious traditions partially determine the value patterns in societies [115, 193].

Thus, it can be observed that the European Catholic countries were economically developed later and less than the others, and one single Church has monopolised the religious offer. Protestant countries have been more secularised after being exposed for centuries to the modernisation processes [59]. Therefore their religiosity has been more spread in Europe. An important factor to consider in this difference is the proven conservatism and rejection of modernisation processes by the Catholic Church in comparison with other churches. However, this is not the only factor, as others may be important such as the thesis of secularisation based on the increment of existential safety or the theory of the religious markets [193]. However, the last EVS (1999/2000) shows a possible convergence and approximation among both country blocks, because the Catholic countries are moving away from the traditional paradigm.

The 'Cultural Modernisation' (CM) can be defined as a macro-trend, a constellation of interconnected phenomena in the values sphere, that is, under the scope of values, beliefs and attitudes and closely related with behaviours. Supported by the pioneering works of Abraham Maslow [180], and some other important works by Inglehart [147, 148] and Halman [83], the process is outlined in table 3.1. It is important to consider two dimensions of this process: individualisation and post-materialism.

First, the concept of social individualisation of Halman, which can be considered a macro-trend inside CM. In this context, individualisation can be understood as the historical and social process in which values, beliefs, attitudes and behaviours are oriented towards personal choices, promoting the autonomy, self-expression and freedom of the individual, and reducing the dependence from traditional principles, authorities and institutions [9, 14].

Those strongly affected by this macro-trend are sensible to independence, development of their own criteria, controlling their attitudes and behaviours from their own

Table 3.1: The cultural modernisation process: a comparison of traditional, modern and postmodern societies, according to Inglehart.

	Traditional	*Modern*	*Postmodern*
Central societal project	Survival in a stationary economy	Maximise economic growth	Maximise subjective welfare
Individual values	Religious norms & community traditions	Motivation for success	Postmodern & post--materialist values
Authority system	Traditional authority	Rational-legal authority	Loss of importance of both religious & legal authorities

Source: Inglehart R., 'Modernization and Postmodernization', Princeton University Press (1997) [148].

point of view and not imposed by external religions, ideologies or social authority. Thus, personal happiness as driving principle, plurality of subjective world views, democratisation and self-determination are closely connected to it.

A second dimension can be found in Inglehart's post-materialism, based on Maslow motivations theory [180]. According to Maslow, the more basic needs are covered, the more complex needs and motivations arise. Thus, a motivational pyramid can be drawn, where individuals tend to prioritise the superior needs while covering the inferior ones: physiological, safety, social belonging, esteem, self-actualisation needs. Inglehart, inspired by this theory, developed the idea that economic welfare after World War II brought to advanced industrial societies a progressive change of values: from valuing material welfare (mainly economical issues) towards other post-materialistic priorities based on self-expression and personal realisation.

These two dimensions are a developed version of the bidimensional classification of Flanagan [88], who hardly criticised Inglehart's one-dimensional view of values as a strong simplification of the values evolution social process. In fact, Inglehart ended up rectifying and assuming a bidimensional view in his last works, as shown in Figure 3.4, which condenses the results of the WVS considering both dimensions [148].

However, the EVS series do not provide consistent indicators for realising a comparative analysis of any of the two dimensions. This is the reason why a selection of other significant indicators from the EVS was selected, taking into account their close association with the two dimensions considered. Moreover, those indicators have a close relation with beliefs, attitudes, religiosity and ideology trends. For the methodological details of the selection check [9, 14, 15].

Thus, the most relevant variables in the model for the study of cultural change include characteristics concerning the 'mental state' of individuals:

- The religious typology previously explained in section 3.2.2.

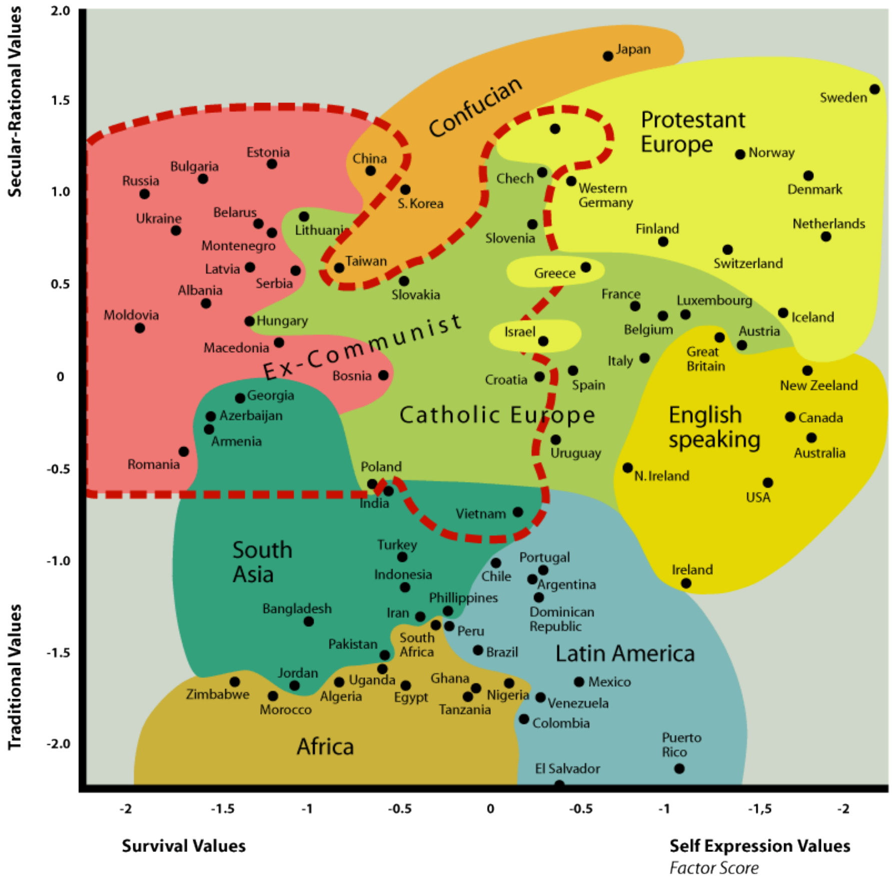

FIGURE 3.4: The Inglehart-Welzer Cultural Map of the World, where countries are located depending on their aggregated values of the two dimensions of Cultural Modernisation. Spanish values are located close to the centre. Source: Ronald Inglehart and Christian Welzel, 'Modernization, Cultural Change and Democracy'. New York, Cambridge University Press, 2005: p. 64 [151] based on the World Values Surveys [150].

- Political ideology, measured with the classical one-dimensional spectrum[6] 1-10 (left-right), significantly relevant to measure mentalities, and interrelated with religiosity.

- Several structural characteristics, such as educational and economic level, gender, age or civil state.

- Tolerance levels towards controversial topics: prostitution, abortion, euthanasia, divorce, homosexuality and suicide. These tolerance levels are good indicators as they show a high correlation with other indicators related to cultural modernisation, especially with the individualisation. Thus, a high tolerance would

[6]The political ideology is imported from the EVS in this 1-10 form. However, due to technical reasons associated to the use of fuzzy logic, this spectrum was re-defined in the real interval [0,1], as explained in Chapter 5.

be associated to liberal and non-religious minds. Figure 3.5 shows several statements (several of them expressing 'tolerances') projected over the two dimensions of cultural modernisation: post-materialism (with the two extremes of traditional and rational) and individualism (with survival and well-being/self-expression as opposites).

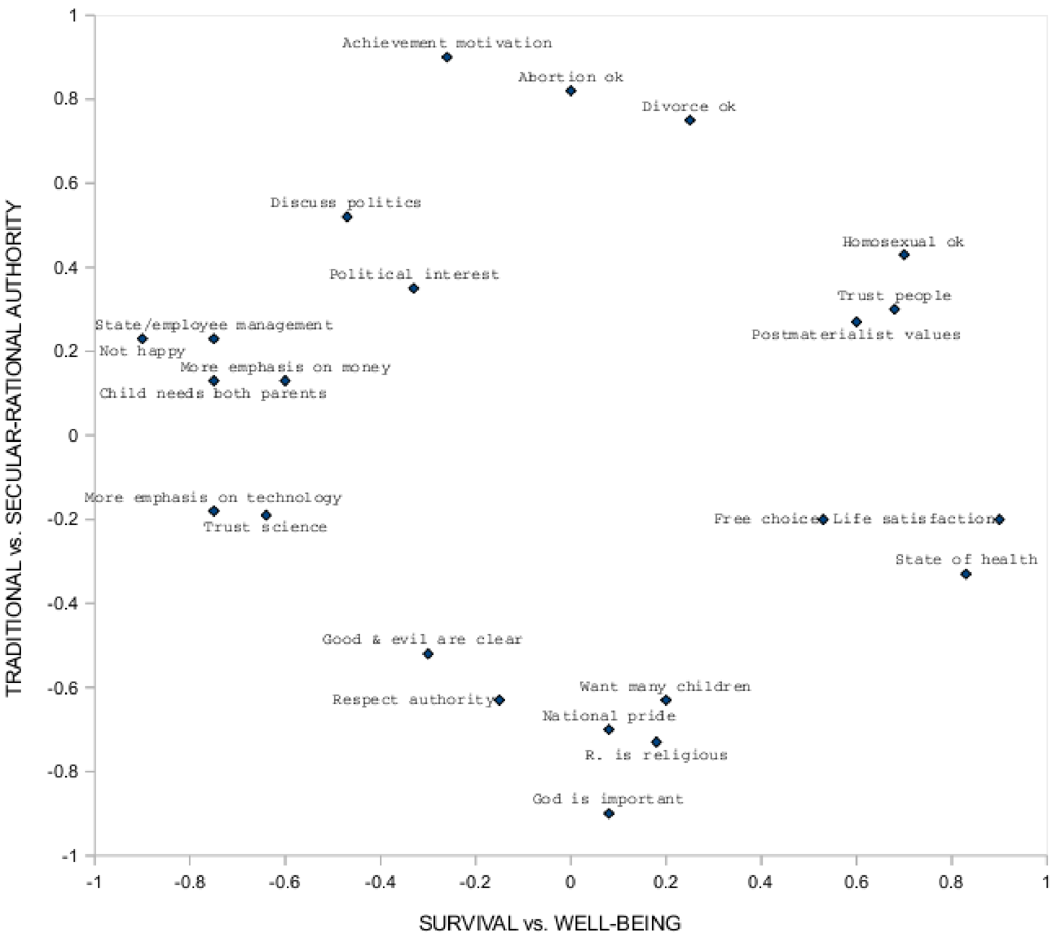

FIGURE 3.5: Scatter chart of Authority and Survival or Well Being. Source: R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148].

3.2.4 Social Influence vs. Demography

According to the values surveys and the literature, the process of evolution of social values has two sides: the intergenerational perspective (widely theorised by Inglehart) and the intragenerational perspective. Both perspectives can be named vertical and horizontal changes, respectively. Intergenerational dynamics take into account the changes across generations, which are socialised in different values. Intragenerational changes consider the internal changes in a life course, that is, the evolution inside each generational group.

On one hand, the individuals of different generations are socialised in clearly differentiated values and attitudes. Such socialisation happens in the early ages of life, during adolescence and early youth, the moment where values, beliefs and basic attitudes are interiorised. Afterwards, these values tend to prevail over all the person's life. Even though there are chances for change, this pattern will have a great influence for a lifetime [147].

On the other hand, there is an intragenerational change, that is, evolution inside each generational group. Thus, subjects modify their values and attitudes along the adult's life as well, and not only during youth. However, these changes have a minor relevance in comparison with any of the intergenerational changes. Thus, cultural and value changes happen mainly in 'generational jumps', and only secondarily by changes in individual's lives (influenced by the socio-cultural environment), as shown in Figure 3.6.

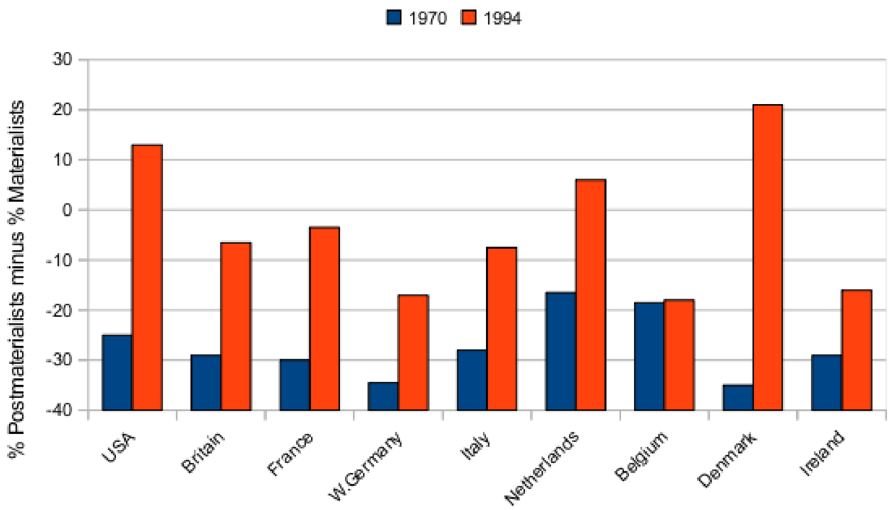

FIGURE 3.6: The shift toward Postmaterialist Values among nine Western Societies, 1970-1994. Source: European Community Surveys (February 1970, Fall 1994), and U.S. National Election Surveys from 1972 and 1992. Appeared in R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148].

This model aims to determine to which extent the demographic dynamics explains the magnitude of mentality change in Spain. The demographic factor neither causes nor determines the change in values, but it does exert an important influence on the velocity and intensity in which it manifests itself, so (and for this reason) it possesses an important predictive ability for its evolution. This is mostly due to the fact that the changes in values have been chiefly (but not exclusively) generational changes, hence the generational replacements (say, the death of elders, carriers of the most traditional and conservative values, and the arrival of youngsters, bearers of emerging values, as shown in Figure 3.7) constitute a significant sociological inertia.

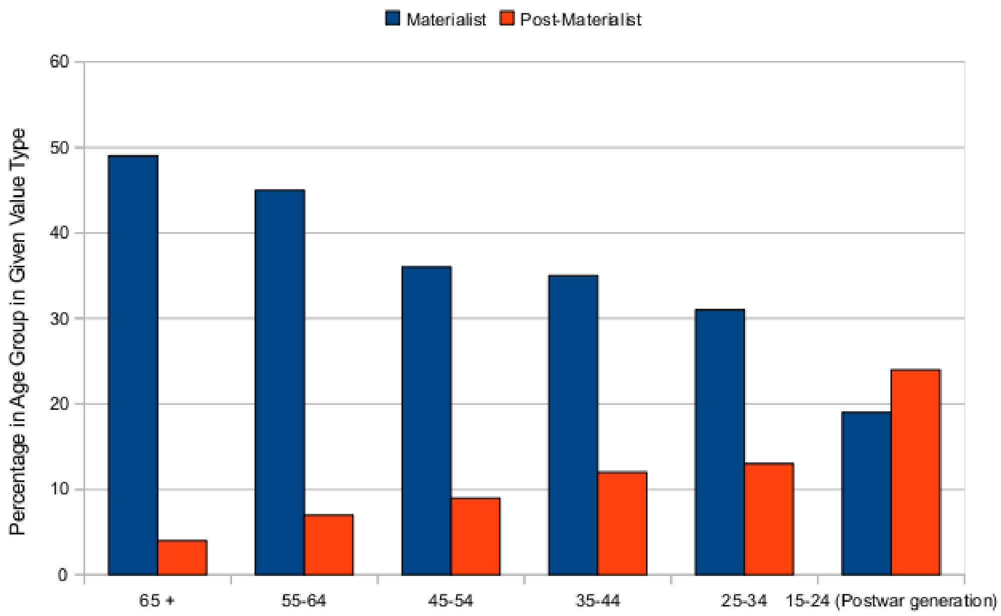

FIGURE 3.7: Value type by age group, among the Publics of Britain, France, West Germany, Italy, Belgium, and The Netherlands in 1970. Source: European Economic Community Survey Feb. 1970, appeared in R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148].

Therefore, the values of each simulated individual must remain constant, but not their aggregation in the whole society, as its demography is changing over time. This will reflect the intergenerational changes but not the intragenerational ones. However, this isolation is an optimal way to analyse and appreciate the predicting effect of the demographic dynamics[7].

3.2.5 The Key Role of Modelling Demography

If the intergenerational differences constitute the essential factor for the explanation and prediction of the evolution of values, beliefs and religiosity forms, the demographical factors acquire a fundamental role in the model. Therefore, the model can explore to which extent the demographic dynamics (and the intergenerational changes associated) are responsible of the magnitude of the evolution of the value survey data.

Spain appears in the value surveys as one of the countries where the variable *Age* is more discriminant, that is, it helps to differentiate mentalities better than in other countries (at least in the period 1980-2000), as shown in Figure 3.8 from [148]. This is because in the studied period, the country suffered multiple social, political and cultural changes, both deep and fast. Thus, the early socialisation models, according to

[7]This should be complementary to the temporal series of the values evolution of the different generations and its associated statistical analysis.

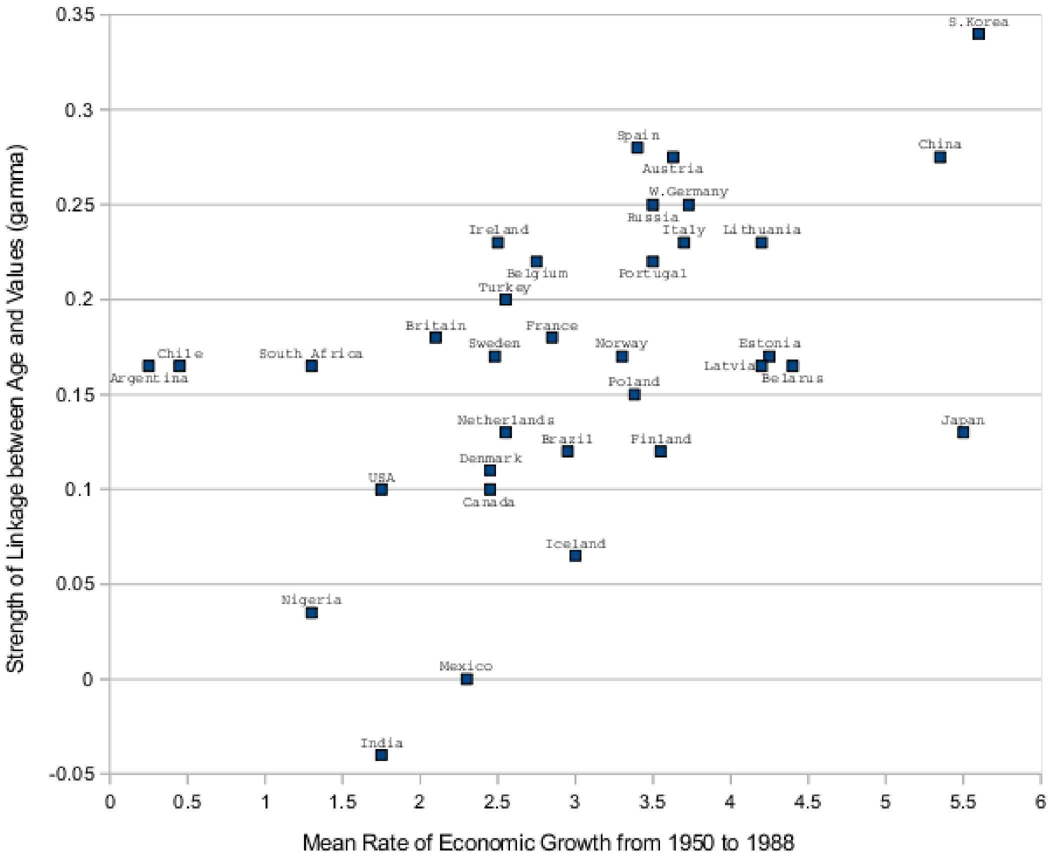

Figure 3.8: Societies with fast-growing economies have relatively large differences between the values of young and old. Note the high position of Spanish society. Source: World Values Survey 1993, appeared in R. Inglehart, 'Modernization and Postmodernization'. Princeton University Press (1997) [148].

Inglehart's works, changed frequently, causing that individuals with different ages have significantly different mentalities [12, 13].

Thus, from the previous theory it can be inferred that the demographic dynamics in Spain should reach a high predictive value of the prediction of such mentalities for several decades, since the generational gaps will tend to keep along time. A way of validating this hypothesis is to study the effect of demographic dynamics without taking into account the intergenerational changes. If the theory is right, it should be observed that the simulation data from 1980 should adjust relatively well (even when time is taken into account) to those obtained empirically in the later curves of 1991 and 1999 (the wavs from the EVS that will allow the validation).

Therefore, the modelling of such demographic dynamics is essential. This process is described in [118] and in sections 6.3.3, 7.2.3, 7.2.4 and 7.2.5.

3.2.6 Size Matters: Simulating Survey Individuals

The model attempts to simulate the demographic dynamics of the studied period from 1980 to 2000, checking to which extent such dynamics explain the magnitude of mentality change in Spain, as shown in the survey data. In order to do so, the agent-based model is fed with the survey data of 1980, using it as initial configuration. However, this means that the agent-based model will have just as many individuals as the survey can provide (in this case, EVS-1980 is a sample of 2303 individuals), and this rises several issues.

First, it is important to insist in the sample's statistical representation, that is, the sample is a good approximation of the behaviour of the whole population (with a small margin of error), assured by the statistical sampling methods. Thus, those 2303 individuals[8] are a valid representation (reproducing in small scale the same proportions of the reference universe) of the whole Spanish population in 1980 (around 40 million people), in every variable considered by the EVS. Thus, it will be used to establish the initial conditions of a model initialised with them.

Another important issue appears when considering the interaction of those agents. Even though the attributes of those individuals represent the Spanish population, will they behave dynamically as they should to represent the general behaviour?

The size of a town/village/city is important in the structure of interactions. In small habitats such as towns below 5000 inhabitants, relationships are usually closed and it is common for all people to know each other. This is the opposite of what happens in large metropolis, where interaction networks are open and fluid, less conditioned by physical distance than in a town.

The issue appears because a group of a few thousands of individuals could end up showing the global behaviour of a town. This is the case especially if the interaction rules are defined considering the general population, which is determined mainly by urban spaces. The main difference of a town, a small closed society, compared with a city is this structure of interactions: as each individual knows everybody, the chances of interacting and influencing others is significantly larger than in a city.

This 'town effect' can only be reduced increasing the size of the artificial society or limiting the number of interactions among the subjects. This last option was chosen in the model, as it is theoretically composed by urban citizens (majority in Spain, and most of the EVS sample). The limitation was implemented through the restriction of people that an agent can know along its life, as explained in 5.5.1 and 6.2. Instead of considering the average of around 100 contacts that an individual usually has, the number of friends was chosen, as it reaches maximums of 40-50.

[8] As section 7.2.3 explains in-depth, this number was increased in the simulation, with the introduction of several hundreds of children not represented in the survey.

3.3 Conclusions: the Sociological Model Principles

Throughout this chapter the main supporting theoretical background of the problem to model, the evolution of social values in Spain from 1980 to 2000, has been addressed. The reasons for using social simulation and its potential contributions to Sociology are summarised in the introductory sections. Afterwards, the EVS was analysed, together with the main concepts underlying the sociological model: for instance, cultural change, religiosity evolution, or intra/inter-generational dynamics.

Next, the sociological principles that have been addressed in-depth will be integrated to summarise here the theoretical model. The main decisions taken in the following chapters are consistent with this model, and are partially derived from it.

1. This work has followed a perspective using both qualitative and quantitative sources, 'triangulating' its findings [22, 55, 155]. However, its main empirical source is the European Values Study (EVS) survey [84].

2. The Cultural Modernisation macro-trend (section 3.2.3), is supported by Inglehart [147, 148] and Halman [83]. This hypothesis defends the process of modernisation of societies from a Traditional perspective to a Modern and later Postmodern one. This process is deeply related to the economic development and the change in values. Its main two dimensions, as considered here, are individualisation and post-materialism. Individualisation [83] promotes individual autonomy, self-expression and personal freedom, and discourage the dependence from traditional principles, authorities and institutions (including the political parties, the state and the Church). Post-materialism [148], based on the works of Maslow [180], assumes the progressive preference from material welfare towards post-materialistic priorities based on self-expression and personal realisation.

3. As the Individualisation process implies a progressive secularisation of the society, religiosity tends to decrease in the advanced industrial societies, promoting an increasing distance from the Church institution. Moreover, the values encouraged from the Church represent an obstacle in the development of the cultural modernisation. This progressive weakening is especially intense in the case of orthodox religiosity (coined Ecclesiastical in the religious typology of section 3.2.2). On the other hand, moral permissiveness and tolerance levels towards traditional controversial topics increase significantly (that is the case of abortion or homosexuality).

4. The process of values evolution has two sides: intergenerational and intragenerational changes (section 3.2.4). Intergenerational dynamics take into account the

changes across generations, which are socialised in different values. Intragenerational ones are considered the internal changes in a life course, that is, evolution inside each generational group.

5. Inglehart studied in-depth the intergenerational changes [147], arguing that values are acquired during the early life stage, estimated around adolescence and first youth. He states that it is in this important period when values are shaped, and then they will keep rather stable for the rest of the person's life. Later changes are possible, but sensibly more moderated. Therefore, Inglehart gives intergenerational changes the main responsibility for values evolution over time, while considering intragenerational changes of minor relevance in comparison.

6. In this context defined by Inglehart, it is expected an essential influence of parents' values in their children's. That is, the liberal non-religious families will tend to educate their children in those liberal values (for instance, with equivalent or even less religiosity), while conservative and religious parents would do the opposite.

7. Assuming Inglehart's theory, intergenerational changes drive social values evolution. However, those changes are completely dependent on demographic dynamics. Demographic inertia pushes towards the modernisation of society, due to multiple factors:

    (a) The elder generation is the most socialised in traditional and religious values. When these values begin disappearing, they are replaced by others socialised in more modern sensitivities.

    (b) On the other hand, new generations are more and more influenced by the values of individualisation and post-modernisation.

    (c) The influence of religion in the new socialisation processes is more and more reduced.

8. Thus, this demographic inertia has an expected important implication in the strength of the following effects, as shown in the temporal series of the values surveys [150]:

    (a) Decrease of ecclesiastical religion and traditional values (religious or non-religious, such as other conservative values).

    (b) Increase of non-religious identities.

    (c) Expected stabilisation or probable increase of religious forms reconciled with individualisation values.

    (d) Important increase of social permissiveness (tolerance) of aspects related to the existing tensions among religious and secular views of the world (abortion, divorce, euthanasia, suicide, homosexuality, etc.).

9. On the other hand, it should be considered the essential importance of parents' values and the stability of those acquired during childhood. Then, the socialisation process, which plays an important role, can be understood as a simple 'social inheritance' from the parents. Note that this is a simplification, as the socialisation process is influenced as well by the environment and other socialising agents.

10. Based on the previous points, the model should provide rather accurate results on modernisation indicators just considering the demographic evolution (intergenerational changes) and the socialisation process (shaping of values). Thus, the values from the parents can be transmitted to the children without loosing consistency with the theory. Moreover, values can be considered stable on the course of a person's life once they are 'inherited' from the parents. Societal values evolution will then be driven by demographic dynamics.

11. Due to the relative importance of demography (section 3.2.5), an effort on its empirical support must be done. Then, the population pyramids will be taken into account, together with, for instance, the reproduction ages or the children rate per couple. The agent model must have a demographic micro behaviour in order to find a partner (changing their civil state) and have children (who will inherit their values).

12. Then, even though the intragenerational changes are not considered in this stage (although they may be tackled in other models as explained in section 8.2.4), the model needs the development of relationships among individuals. Those can be based on friendship dynamics which can lead to finding a proper partner and may lead to offspring. The friendship dynamics in the model are based on sociological research, which has been included in Chapter 5 due to the requirement of contextualisation. Note that individuals tend to have stronger friendship ties to those with similar characteristics: similar social class, close age, proximity in space, similar values. Besides, the model can define different levels of friendship, from acquaintances to friends and spouse. Thus, a complex social network is built, with different link kinds (family and friendship) and strengths (a continuous degree).

13. The modernisation indicators that are considered in the model have been selected taking into account their relevance to represent the described dynamics of individualisation and the viability of their empirical validation. They are listed in section 3.2.3.

14. The period and society chosen for the case study have been the last twenty years of 20th Century in Spain. Throughout this period, this society is affected by a fast modernisation force which clearly changes the general shape of the social values

snapshot from 1980 to 2000. The main dynamics explained here (secularisation, individualisation increase, religiosity fall, tolerance increase) can be found in the Spanish society, being transformed during those 20 years. Besides, the variable age is significantly discriminant in the Spanish case, increasing the dependence of the demographic dynamics. Therefore, with the sociological model principles stated, the evolution of social values in Spain during this period should be a viable task to achieve.

Chapter 4

The Micro View: the Mentat Agent

With the myth of the State out of the way, the real mutuality and reciprocity of society and individual became clear. Sacrifice might be demanded of the individual, but never compromise: for though only the society could give security and stability, only the individual, the person, had the power of moral choice—the power of change, the essential function of life. The Odonian society was conceived as a permanent revolution, and revolution begins in the thinking mind.

Ursula K. Le Guin, *The Dispossessed*, 1974

4.1 Introduction

Human behaviour is a complex system by itself [107]. Therefore, the simulation of such behaviour must be constrained to specific contexts where only some aspects of this complexity are modelled. And thus, the agent used in the model will be just a partial representation of some specific human actions. The choice of which subset of the wide human behaviour is more relevant (and thus should be modelled) is an open problem that has a different answer depending on the chosen perspective: it is an extension of the classical issue of the 'human nature'. Traditionally, economists tend to give importance to the selfish and rational part ('homo economicus'), while sociologists focus on the social capabilities (Aristotle's 'zoon politikon') and psychologists tend to see humans as mainly irrational and emotional.

Thus, explicitly or not, every agent-based model follows one or another of these perspectives. Several agent types are discussed in the next section in order to frame the decisions regarding micro aspects taken in Mentat. This framework chooses to build a specific sociological agent architecture, the 'Mentat Agent' or 'M-Agent', useful for a collection of problems that can be modelled with instances of this general agent model (this family of problems is listed in the next section 4.3). The actual problem of the evolution of the Spanish values will be used as an example of the contextualisation of such architecture.

As other framework-related decisions, the use of a specific agent architecture significantly facilitates the application of the proposed data-driven methodology. Other problems from the same family will be modelled in a similar way following the structural guidelines of this architecture.

After the state of the art of agent models (section 4.2), a specification of the M-agent architecture is presented (section 4.3). Afterwards, its contextualisation in the studied problem is carried out in the next sections, specially focusing on the attribute definition and micro-behaviour.

4.2 Agent Models: A Review

There are different approaches to undertake the issue of the agent model. An agent can be as basic as a reactive cell in cellular automata [261] or even reach the level of a complex entity using a cognitive model. Cognitive models arose in the last years, especially from the cross of cognitive science with artificial intelligence advances. They try to provide a frame for psychological mechanisms, processes and knowledge structures, mostly supported by process-based theories [239] and from a bounded rationality approach [230]. A cognitive model is frequently interrelated with a specific cognitive architecture. A cognitive architecture provides a specific framework for more detailed modelling of cognitive phenomena, through specifying essential structures, divisions of modules and relations among modules, always embodying fundamental theoretical assumptions [237].

The debate concerning the agent model choice is an open issue which still drives many discussions. Is it possible to create a general modelling principle for its application in every model, or at least the most of them? Should social psychology models be considered and integrated into the models, or should they be ignored for the sake of simplicity? When does social simulation need cognitive models? Is the BDI model a valid general-purpose architecture? [65]

Most of these questions can only be answered partially or with ambiguous responses. Besides, depending on the chosen approach, the answer may be biased towards the 'human nature' view of the modeller, as explained in section 4.1. Next, a general review will be tackled before exposing the architecture of the Mentat agent in the next section.

Gilbert has argued that social simulations do not always need to be coupled to cognitive models, as in some cases using them would only complicate the research. Besides, taking into account the dual nature of social processes, working on societal and individual levels requires the consideration of both levels and the interaction dynamics among them [103]. Deffuant et al. discussed in [65], together with other epistemological issues, the need of a cognitive model and the difficulty of finding a global unifying principle for agent modelling.

Computer scientists and MAS designers are familiar with the Beliefs-Desires-Intentions architecture (BDI). It is a model proposed by Bratman [34], in which agents take decisions based on their knowledge (Beliefs), individual objectives (Desires), leading to plans for further actions (Intentions). This model is based on the Rational Choice Theory [49] widely used in the classical Economics or Game Theory models (that is, it is an extension of the rational agent used there). Its main advantage is to provide a robust ordered frame structure, which allows researchers to clearly model the agents.

However, it implies important constrains: agents should be typically homogeneous, completely rational and selfish [209]. As collected by [185], multiple researchers argue against this reductionism: ‘human beings act, primarily if not solely, on the basis of self-interest’ [67], ‘Egoism is not an assumption but the assumption underlying Neoclassical Economics, which is, in turn, the dominant approach to the discipline in this country’, or the observation of Hazlitt that a society comprised entirely of either altruistic or egoistic agents would not be ‘workable’ [133]. Besides, there are important critics concerning the lack of experimental grounding of the architecture [65]. Moreover, Gilbert states ‘[the BDI architecture] is not based on experimental evidence or on theoretical analyses of human cognition, but rather on what we might call ‘folk psychology’ ’.

Therefore, the BDI framework is ideal to model the roles of the workers in a factory, or vehicles traffic dynamics [68, 171, 208] but in some other environments (such as the sociological context of this work) another approach is needed.

Thus, multiple cognitive models, using bounded rationality [230], have emerged. They add complexity to the classical rational agent or begin with different premises (for instance, suppressing the infinite computational capacity and global information premises, characteristic from the rational agent). They usually take into account social psychology theories to build what it has been coined agent cognitive models [237, 239]. Depending on the cognitive model, its approach will try to focus on different issues that were ignored in the rational agent, stressing their importance and including them in the model. Frequently, these issues are related to the social capabilities of the agents. For instance, [53] empower the social learning capabilities as fundamental for social simulation agents, or [238] focuses on organisational theories and the agent roles inside them. On the other hand, Antunes proposes an alternative to utilitarian rational models stressing the importance of values from a cognitive point of view [4, 6].

Plenty of cognitive models provide their own cognitive architecture, as they are usually closely interrelated. In fact, that is the case of several cited works, such as [238] or [6].

However, the debate around cognitive architectures and which structures they should have, depends on the social theories that are taken into account. Even though integrating attempts have been made for generalising such global cognitive architecture [40], the debate is still open. Efforts like the one of Carley et al. [41] throw some light on the issue. This work tests several cognitive architectures (including humans) for their performance in the frame of organisation theory, in both micro and macro levels. The results suggest how in the micro level more cognition is needed, while in the macro level simpler models perform better. There are some other interesting examples of cognitive architectures, such as the one proposed by Sabater et al. [218], which develops a consistent architecture taking into account a cognitive theory of reputation, tested through a evaluation using fuzzy logic (appropriate for modelling qualitative concepts, common in cognition). In [31] the authors propose a cognitive architecture for an economic simulation of firms. Although in Economics, simulations frequently make use of the rational agent, this work uses bounded rationality agents, with limitations in attention, time and memory. Others can be found in reference books on the issue such as [237, 239].

In the case under study, a social agent architecture was built, but not attempting to design a new general cognitive model but an ad hoc architecture for a specific family of problems (which is specified in the next section). The model designed for Mentat is supported by sociological theory, and strongly driven by empirical data.

4.3 A General Architecture for the M-Agent

The Mentat Agent (M-Agent) does not follow a complex cognitive model but falls into the sociological definition of human behaviour. The aim of building an unified general-purpose valid-for-all agent model is considered too ambitious and even utopian [65], even though some attempts have been made [184]. Thus, this architecture has been built as general as possible, not framed in a specific behavioural theory, but constrained to a family of problems where it can be adapted to. Next, a specification of those problems is detailed:

- From a bird's-eye view, this agent framework facilitates the exploration [5] of complex data-driven models[1], as the framework allows the isolation of certain layers and modules. Thus, it is feasible to analyse the weight of different factors in the resulting aggregated effect. E.g. opinion dynamics vs. demographical evolution,

[1]The ease of communication of complex data-driven models is partially lost, if they are compared with the theoretical KISS models. Moreover, in such large and complex models, frequently it is difficult to realise a sensitivity analysis to determine the influence of each factor in the final result. A flexible framework like the proposed one, with modules optionally disabled, can facilitate this task.

media effect, second order emergence, norms emergence, cognitive capabilities, learning mechanisms and others.

- More specifically, this framework is optimised for studying the evolution of multiple individual characteristics in a given society and period, especially in contexts with abundant quantitative sociological data. Depending on the model, the focus can move from the micro behaviours (such as values or relationships) to the macro indicators (e.g. unemployment rates [134]).

- Besides, the framework is also prepared to study any problem from the whole agent-based computational demography spectrum, such as population pyramids, migration patterns, marriage and family dynamics, mortality crises effects, fertility decline and ageing consequences, etc [29].

Collateral to the case study chosen, it is possible to explore the possibilities of different techniques in social simulation. Thus, some specific problems that may require specific technologies can be easily integrated as new modules inside the system. For instance: GIS, Artificial Intelligence (as explored in this work), statistical, analytical or visual tools, etc.

And thus, there are other sets of problems which this approach is not very well adapted to:

- KISS theoretical non-data-driven models: prisoner dilemma simulations and others from game theory, for example [159].

- Specific contexts were the agents follow very particular behavioural patterns, and thus a significant effort to adapt the framework would be needed: electronic discussions, workers in a factory, emergence in an airport, etc.

- Highly complex psychology-based cognitive agents that require an elaborated ad hoc agent architecture for specific reasons, even though the M-Agent architecture attempts to integrate them.

The M-Agent architecture is structured following the Layers architectural pattern[2]. This pattern helps to structure applications that can be decomposed into groups of subtasks in which each group of subtasks is at a particular level of abstraction. Layers can be organised through composition (an association of classes in which one owns the

[2]Architectural patterns are software patterns that offer well-established solutions to architectural problems in software engineering. Besides, they help to document the architectural design decisions, facilitate communication between stake-holders through a common vocabulary, and describe the quality attributes of a software system as forces. An architectural pattern expresses a fundamental structural organisation schema for a software system, which consists of subsystems, their responsibilities and interrelations. In comparison to design patterns, architectural patterns are larger in scale. [19]

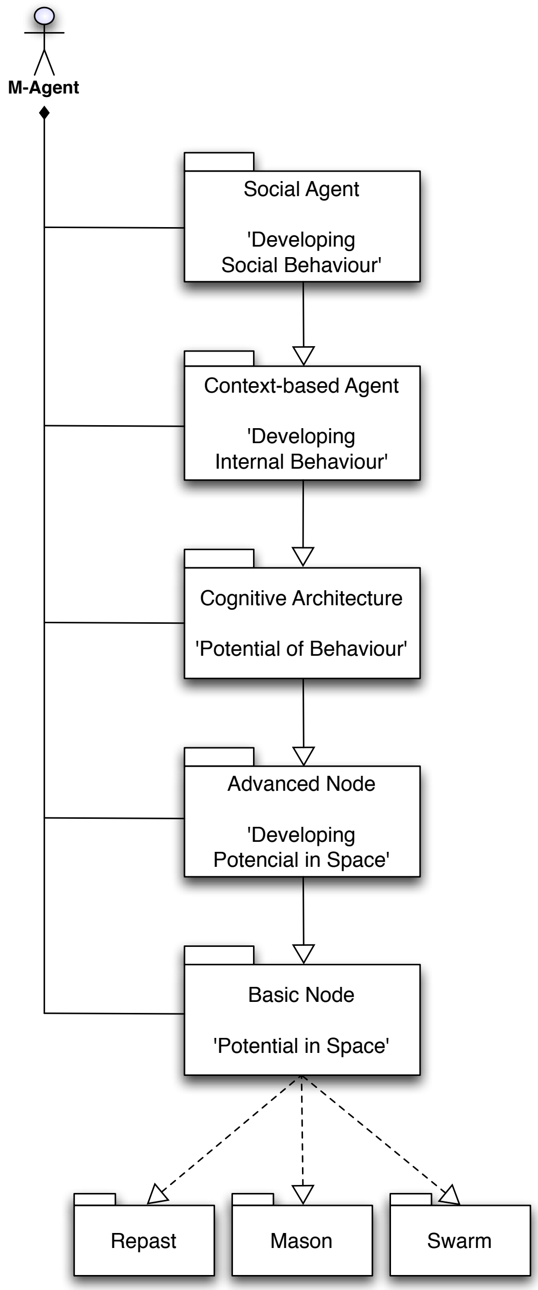

FIGURE 4.1: Diagram showing the global M-agent architecture. Note that each layer inherits the behaviour of the previous one. *Basic Node* may inherit the behaviour of the basic node class of the agent library used, such as Repast, Mason, or others.

others), inheritance (defining specialisation of classes, as generalised in Figure 4.1 and specified in the Mentat UML diagram 4.7) or aggregation (an ordinary composition which does not imply ownership) [19].

Thus, each layer holds a different set of knowledge and micro-actions. The whole M-Agent is composed by the set of all the layers, as shown in Figure 4.1, and each model can give more importance to one or another depending on the specific problem. This approach should guide the implementation but it is neither necessary nor recommended that each layer is implemented just in a single class: design patterns may be applied for the structure of each layer. Thus, the five layers defined are:

1. **Basic node properties**. Figure 4.2 shows the properties for the 'Potential in Space', that is, the capacity to be a draw-able node, and part of a network of nodes, without any internal states or behaviour. It is prepared to follow the 'Adapter' (or wrapper) design pattern, which provides an interface so other classes can work together [96]. Thus, it will be particularised depending on the libraries used: the simple *DefaultDrawableNode* from the Repast framework [194], or the basic node from other frameworks (Mason, Swarm, Netlogo [206]).

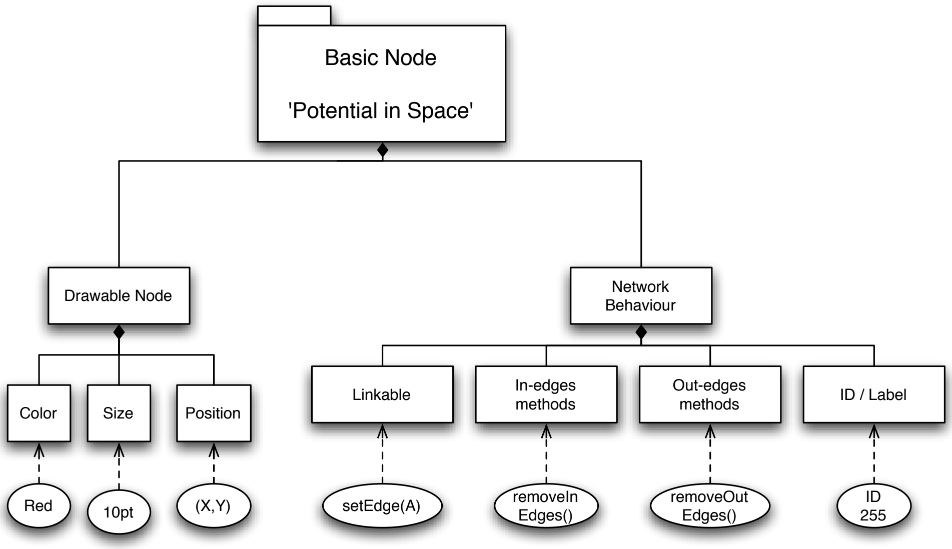

FIGURE 4.2: Sub-diagram showing the architecture of the Basic Node layer.

2. **Advanced node properties**. This layer, shown in Figure 4.3, develops the potential of the previous layer placing the node in a specific space (or several spaces), with spatial actions to gather information from the environment (e.g. neighbourhood knowledge) or move inside it. Classical software agent sensors/perceptors [217] are implemented here, together with any basic spatial or communication action that should be carried out. E.g. if agents need a specific communication language (such as KQML [86] or FIPA-ACL [166]), its related capabilities should be implemented here.

3. **Cognitive Architecture**. Figure 4.4 shows the cognitive layer, the 'potential of behaviour', as it provides the capabilities for acting in a context before placing the agent in a specific context, the same way the basic node provided basic capabilities before existing in a space. Obviously, if the agent uses a particular cognitive model, it should be implemented here. Abilities such as learning or planning, together with

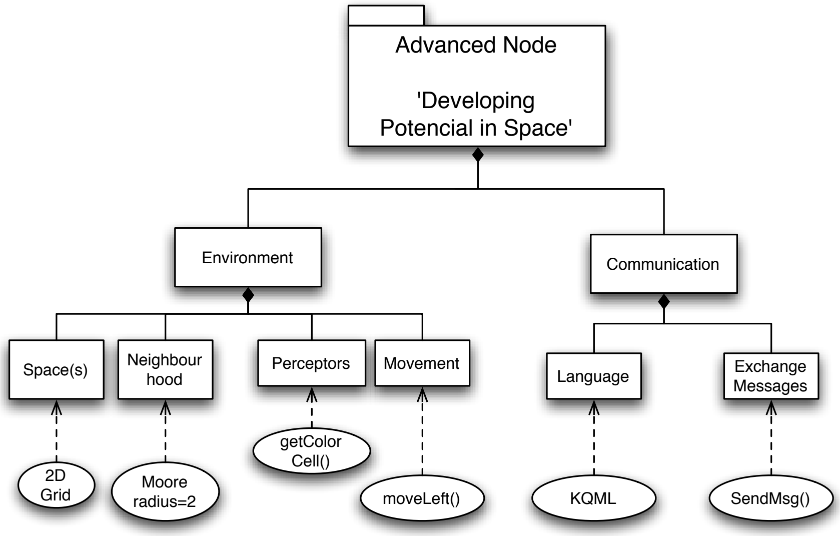

FIGURE 4.3: Sub-diagram showing the architecture of the Advanced Node layer.

the (empty) knowledge base should be included, as it can be seen in the diagram. However, in the Context-based layer these abilities may be further specified.

4. **Context-based agent**. Even in very abstract models, agents do need particular characteristics depending on the problem under study. This is defined by the context of the problem, and thus Figure 4.5 shows the layer that includes the contextualisation of the agent's internal state. Therefore, specific model attributes, knowledge base or mental state should be implemented in this layer, adding the particular structures needed, together with their associated behaviour (such as similarity operators or transition functions). If the agent follows a specific life cycle or automata-based changing state sequence, it must be defined in this layer. Besides, the initialisation of the agent state should be carried out here. Therefore, this layer would hold any data-driven initialisation process, if any.

5. **Interaction dynamics**. Agents do not only receive information from their environment. They perform operations with this information, plan responses and initiate reactions depending on their actual behaviour. Thus, complex interaction actions are modelled in this layer as shown in Figure 4.6, using the communication protocol implemented before. This should be the layer most affected by any data-driven design carried out in the ABM, as it holds the contextualisation of the social behaviour. E.g. opinion dynamics [137], matchmaking or demographic dynamics (as presented in the forthcoming chapters).

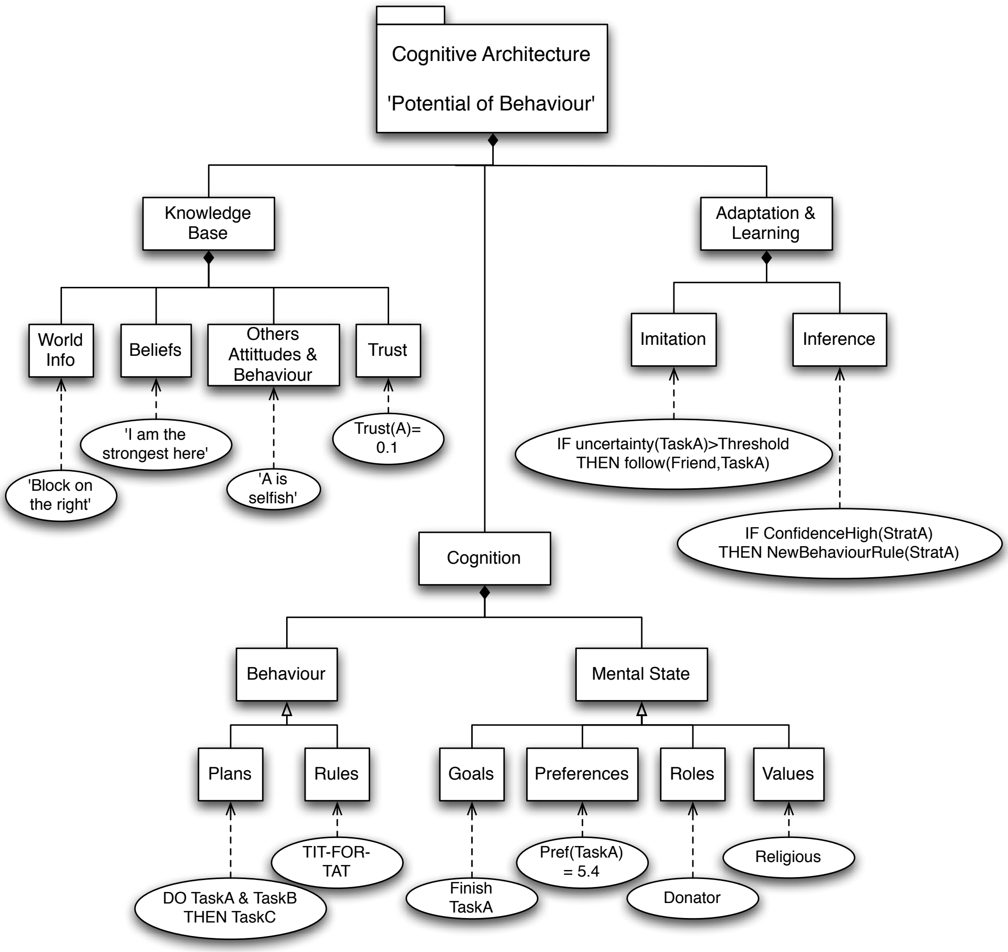

FIGURE 4.4: Sub-diagram showing the architecture of the Cognition layer.

4.4 The Agent Architecture Applied to this Case Study

4.4.1 Contextual Agent Architecture

This section will tackle the application of the M-Agent architecture to the specific problem explained in chapter 3. Besides, the methodology of chapter 2 is applied through every step of the model. Thus, the agents are designed following a data-driven approach, through the 'Deepening KISS' method.

The 'basic node' layer is, as suggested in the previous section, an 'adapter' that can be particularised as the basic node of the Repast framework, *DefaultDrawableNode* [194]. The second layer, 'advanced node', is implemented on top of it: a simple agent is defined, which represents an individual in the defined space: a grid, as described in 6 (although, as tackled there, there could be various spaces). It has 'node' properties, so it can belong to a connected graph of agents, forming a basic network, and have a specific position in the space. An UML representation of it, with its basic node methods, can be seen in the class *SimpleIndiv* diagram of Figure 4.7.

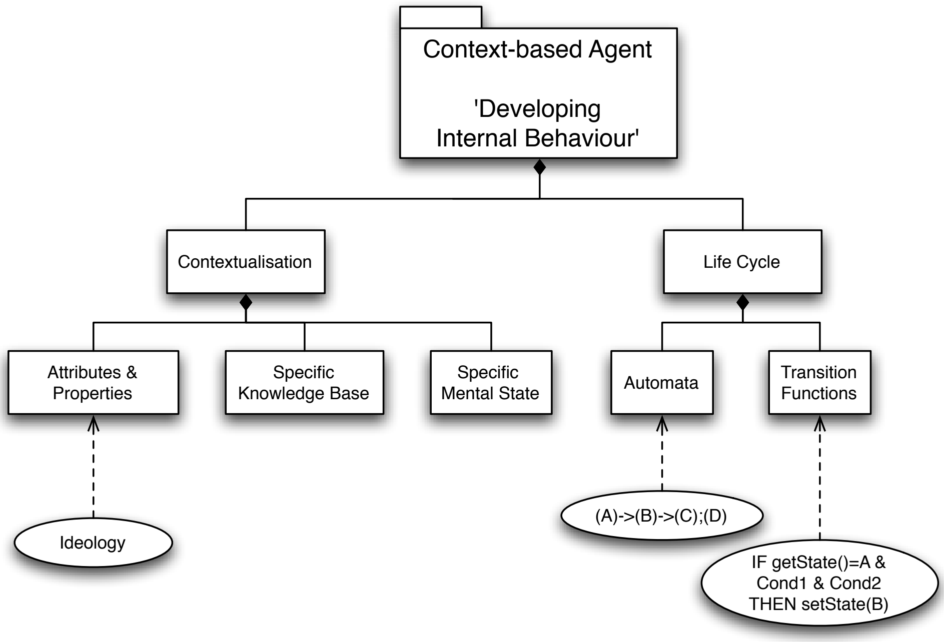

FIGURE 4.5: Sub-diagram showing the architecture of the Context layer.

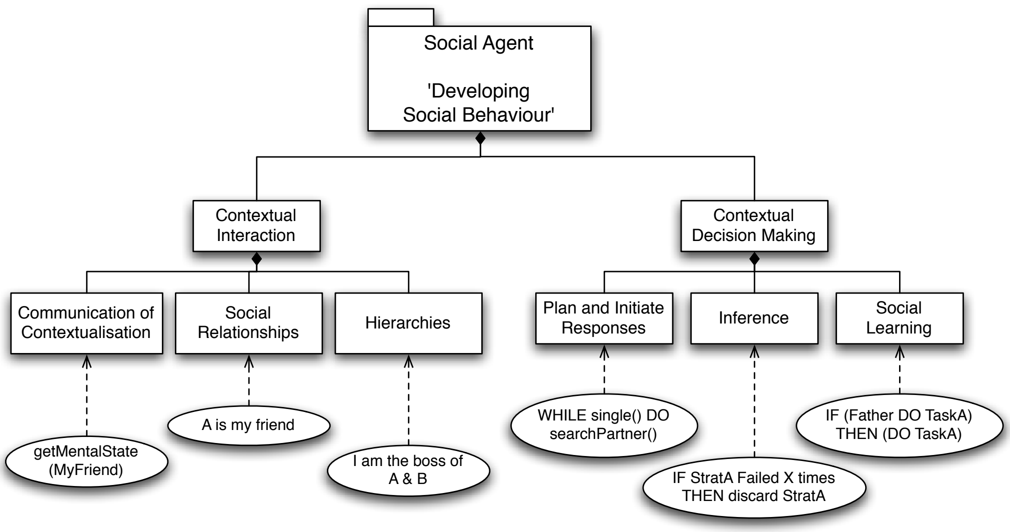

FIGURE 4.6: Sub-diagram showing the architecture of the Social layer.

Once this basic KISS agent is implemented, its design complexity is increased gradually ('deepening'). These agents do not present cognitive capabilities, so the needed structures of the cognitive layer are merged into the contextual one.

The agent represents an individual in the problem under study, so it needs the attributes chosen relevant for this context: the architecture of the 'context-based agent' layer. According to the data-driven approach, those attributes are loaded from empirical data: in this case, from the EVS, described in the subsection 3.2.1. However, only the

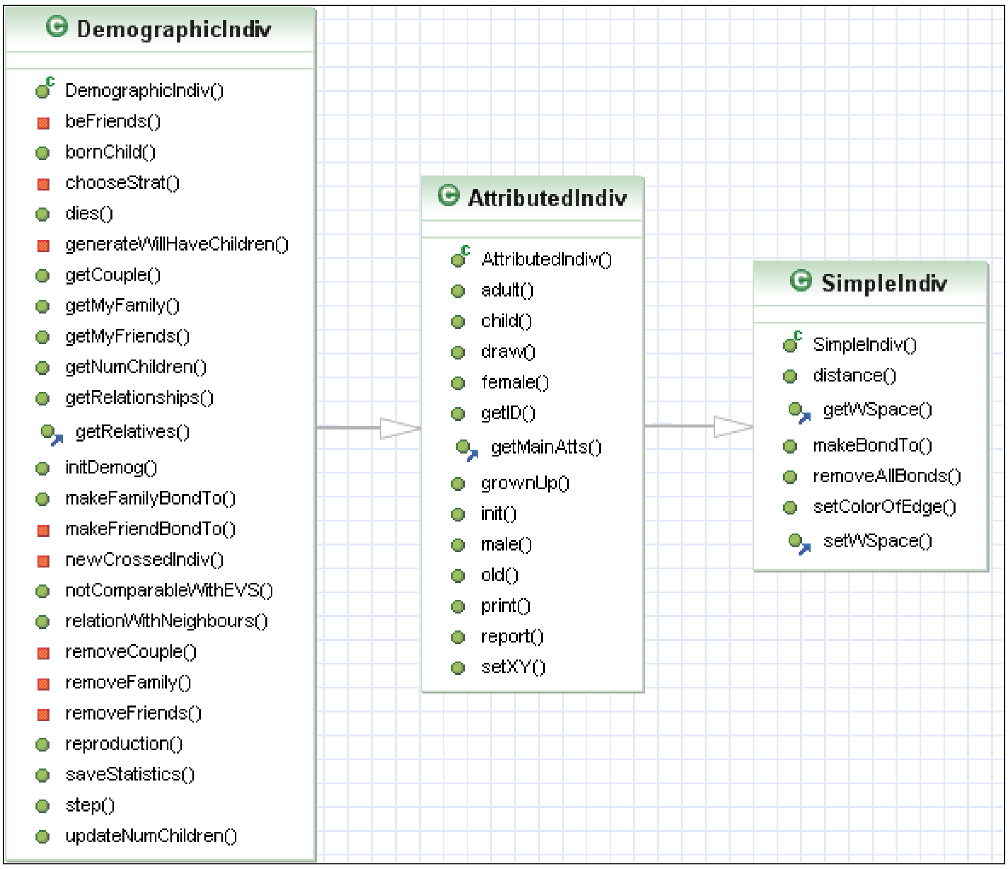

FIGURE 4.7: UML diagram of the agent hierarchy classes.

significant subset of variables were selected. This agent is represented in the UML diagram of *AttributedIndiv* of Figure 4.7, together with some of its basic methods (and associated subclasses such as Figure 4.8).

Finally, the complex agent dynamics are modelled according to the 'social agent' layer, providing micro-actions for the agent to be able to take decisions. Furthermore, the agents are designed depending on a demographic model (as presented in section 3.2.5 and 6.3.3), so they must grow older and follow the life cycle limitations (such as the child impossibility for having children). This agent would be the third class of Figure 4.7, *DemographicIndiv*, where its most important methods are shown. Each one of the three agent classes inherit the behaviour of the previous ones.

As this agent model obeys the needs of the actual problem, there are several constrains to their behaviour. Thus, even though the agents interact with each other, they do not change others' internal states. This is because Mentat is testing the weight of the intra-generational values evolution, not the inter-generational evolution, as it is explained in 3.2.4. Besides, the agents are motionless: they do not move in the space. This is because the space does not only represent the geographical space, as it is explained in 6.2. However, the framework can easily overcome both limitations if it is required by the chosen problems.

4.4.2 Agent Attributes

The agent attributes can be divided in several categories. Depending on their nature they are classified in:

- *Auxiliary programming attributes* (e.g. internal counters or statistical parameters).
- *Social characteristics attributes* (such as the education level).
- *Demographical attributes* (e.g. life stage).

The most relevant ones according to the sociological model are the social characteristics, which are loaded from the EVS, importing the survey indicators used there. Thus, they can be further classified, as detailed:

TABLE 4.1: Classification of agent attributes.

General characts.		**Tolerance levels**		**Others**	
Name	*Type*	*Name*	*Type*	*Name*	*Type*
Gender	*Boolean*	Prostitution	*Integer*	Church Att.	*Integer*
Age	*Integer*	Abortion	*Integer*	Confid. Church	*Integer*
Education	*Integer*	Homosexuality	*Integer*	Is-religious	*Integer*
Economy	*BoundedReal*	Euthanasia	*Integer*		
Religion	*Category*	Suicide	*Integer*		
Ideology	*BoundedReal*	Divorce	*Integer*		

Apart from the self-explained gender and age, the other variables may need some explanation. For example, the socio-economic status, an indicator of education level and the ideological positioning. The indicator of educational level is defined by the age of giving up school as a main activity[3]. The indicator of social status is a standardised factor (average zero and standard deviation one), and it should stay close to constant in time. The indicator for political ideology was obtained from a scale of ideological self-positioning in the spectrum of left-wing/right-wing positions: 1 means extreme left-wing and 10 means extreme right-wing[4]. The religion parameter follows the 4-groups typology defined in the literature described in the subsection 3.2.2. All the tolerances follow a similar pattern: they represent in a degree from 1 to 10 the tolerance of the individual regarding the specific sensible subject. *Church attendance* is also represented the same way.

The programmed internal structure of the attributes can be observed in the UML diagrams 4.9 and 4.10. All the attributes (and the *AcceptanceGrade* class, which is

[3] The reason why the educational level was not quantified using the well-known classification (Primary, Secondary, BSc, MSc, etc) is due to the lack of unification of the education systems in the studied countries.

[4] This is the classical indicator for political ideology used in both Sociology and Political Science. The surveyed individual decides with which part of the spectrum is identified.

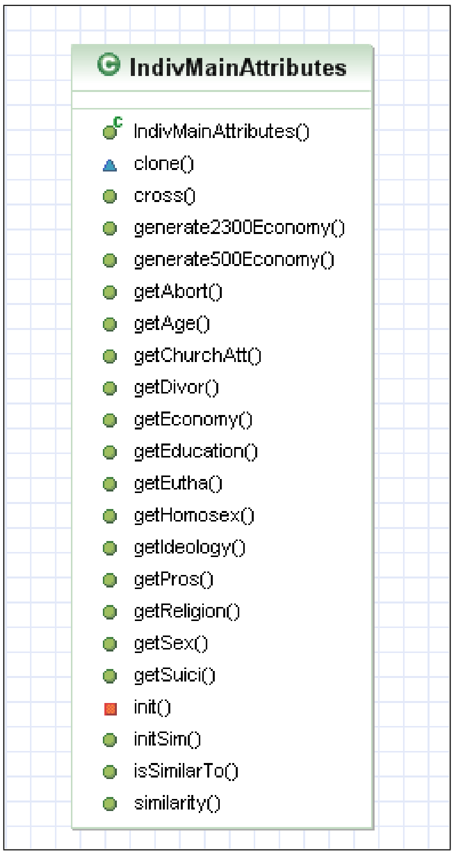

Figure 4.8: UML diagram of the main attributes aggregation class. This belongs to the *AttributedIndiv* class, and thus it is part of the context-based layer.

as well an *Attribute*) implement the *Attribute* interface, which forces the inclusion of methods for the random value generation, value by default and normalised return. Thus, the attributes follow a similar parallel structure. However, each of them handles its associated specific functions: e.g. the age provides the life stage (child, adult, elder) depending on certain thresholds, or the religion provides a clusterisation following a theoretical typology explained in Chapter 3. There are two special cases: first, the *Economy* supports the import of just the Economic Class (low, middle-low, middle, middle-high, high classes) or the specific Economic Status found in the EVS, with a higher level of complexity; second, the tolerance levels shown in 4.10 are all represented in the same way: a degree in the real scale [1,10], and equivalent operations over it.

4.4.3 Life Cycle

Real individuals follow life cycle patterns: they can be children, adults or elder people, and move from one to the next stage, changing their particular behaviour. Besides, they follow the classical scheme for every animal: they are born, grow up, relate to others, reproduce and die.

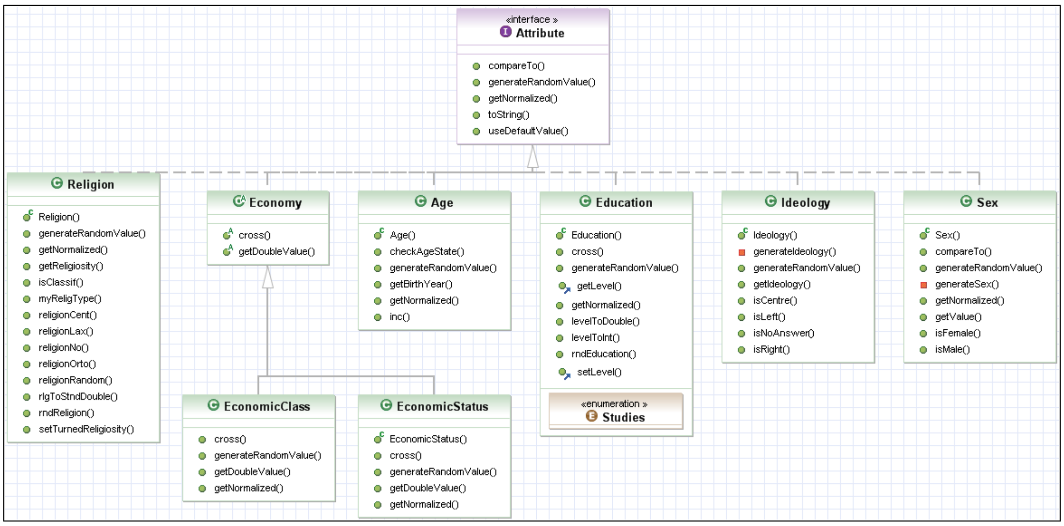

FIGURE 4.9: UML diagram of the main attributes.

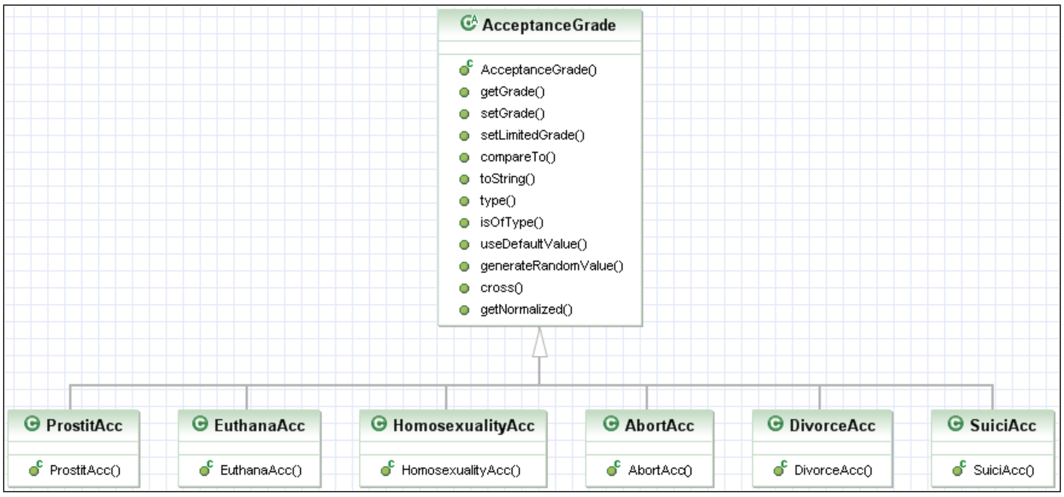

FIGURE 4.10: UML diagram of the tolerance attributes.

The changing state sequence of the agent life cycle is shown in Figure 4.11. Thus, in Mentat the agents are born inheriting the characteristics of their parents. Afterwards, they usually relate to other people, and some of them may end up being their close friends. When the agent reaches maturity, it can search for a spouse. In case of finding one they will marry, and will be able to have children (if the partner does not die before). The couple may have children and give them their inheritance in term of cross of attributes, including values representing the socialisation process, as supported in section 3.3. Note that agents can neither divorce nor die before elderly, because the percentage of both phenomena are not significant enough to be considered in the model.

Therefore, every agent can be child, adult or elder, with their particular behaviour associated, as it is shown in the life cycle figure and summarised in table 4.2. For instance, a child cannot have a spouse, only adults can reproduce, and only the elderly

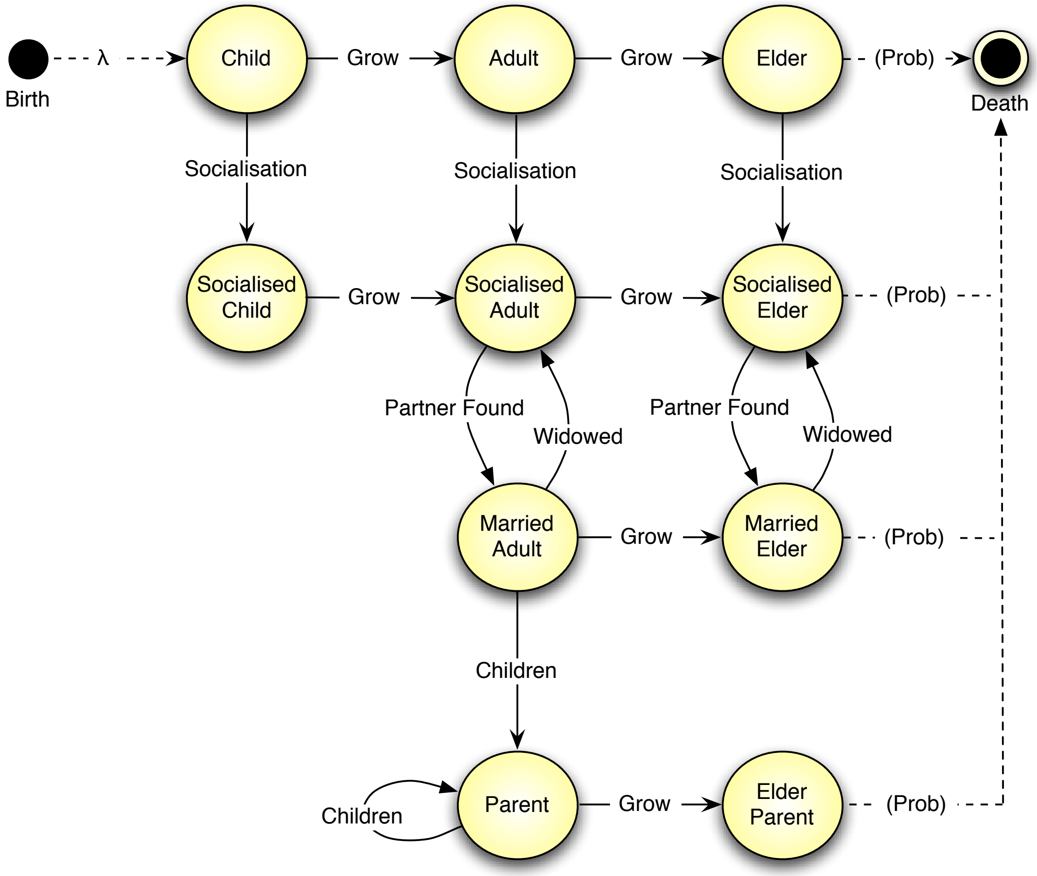

FIGURE 4.11: Finite state automaton showing the agent life cycle.

have chances to die at some point. The operations associated to the life cycle are clear in the *AttributedIndiv* class of the UML diagram 4.7. Those behaviours will be further explained in the next chapter of social dynamics.

TABLE 4.2: Different behaviour according to the life cycle.

	Children	**Adult**	**Elderly**
Grow up	x	x	x
Establish friendships	x	x	x
Find spouses		x	x
Reproduce		x	
Die			x

4.5 Concluding Remarks

Agent-based modelling gives the micro aspect an essential role in the simulation, opposing to other simulation techniques such as system dynamics, focused on modelling the macro [156]. Thus, in order to apply the proposed data-driven methodology, it is essential to define an agent architecture that is able to capture representative data of the problem domain, together with its contextual behaviour.

In this chapter, multiple agent models have been discussed, in order to clarify the background in which one approach is selected. Only afterwards it is possible to undertake the task of defining a flexible layered agent architecture. This architecture does not attempt to cover the whole spectrum of problems possible in ABM, but a family of them which has been specified.

The selected case study, the evolution of social values, constitutes an example of ABM that follows this architecture. Its application has been described, including some implementation descriptions, formal UML diagrams and data sources. Besides, the general life cycle of the agents is summarised, although the socialisation and matchmaking processes are detailed in the next chapter.

The potentials underlying this architecture should be explored in further implementations of other problems, preferably sensibly different to the one carried out by Mentat, to test its robustness and flexibility. It is intended to be useful for researchers trying to approach problems that belong to the defined family.

However, this framework includes not only micro aspects, but also complex social interactions and macro processes, which are developed in the following chapters.

Chapter 5

Social Dynamics

We know that there is no help for us but from one another, that no hand will save us if we do not reach out our hand. And the hand that you reach out is empty, as mine is. You have nothing. You possess nothing. You own nothing. You are free. All you have is what you are, and what you give.

Ursula K. Le Guin, *The Dispossessed*, 1974

5.1 Introduction

This chapter tackles two main objectives. First, it describes in detail the social dynamics carried out in Mentat. Second, it proposes Fuzzy Logic as a helpful technique to deal with the kind of approximate or uncertain knowledge associated with Social Sciences problem modelling. In order to do so, it analyses the application of fuzzy logic to the Mentat model (as it was described in [125]).

Social dynamics, in this context, is referred to the collection of agent interactions, their conditions and associated mechanisms, together with the emergent behaviour that they cause. Regarding of the difficulty of studying such dynamic processes and the frequent lack of data associated to them, social dynamics have attracted attention from several fields. Thus, the available literature concerning these dynamics is broad and diverse. For instance, [135] states that 'using cellular automata is a promising modelling approach to understand social dynamics', and even classical system dynamics [156] attempt to do so. A more theoretical and generalist approach would be Kluver et al. works [162, 163], where the authors provide a purely mathematical framework for the study of dynamics in social systems.

However, the exploration of social dynamics is the strongest point of agent-based models, as it is the simulation technique that better handles such dynamics and studies the emergent self-organised processes that they imply. Therefore, multiple models have approached them from different perspectives and in different case studies. Under the frame of social dynamics, there are studies concerning opinion dynamics, cultural dynamics, language dynamics, crowd behaviour, hierarchy formation, information diffusion, etc. In this section, some relevant agent-based models tackling these issues will be reviewed.

Social dynamics are notably explored in opinion dynamics models, such as the works of Deffuant et al. [63, 66], which are usually KISS models which attempt to explore the diffusion of opinions through the agent social network, till stable (clustered) states are reached. However, the work [64] is an interesting exception that explores opinion dynamics from a data-driven approach. This complete model, result of a large project [62], incorporates economic rational anticipation, decision function over several criteria, dynamics of information transmission. It was partly derived from interviews and experimental data, and provides an insight on political extremism, including economic interests, expectations, and past history. Its data-driven and multiple-source approach is rather similar to the one carried out in Mentat.

The case of Petter & Andreas' work [203] provides an interesting example which deals with some concepts that appear in Mentat. This model studies youth subcultures dynamics from a rather simple model, dealing with Boolean friendship and even considering age as a factor (and thus, demographical issues).

Besides, interesting attempts of integration and formalisation of social dynamics have been made, although the field is rather fragmented in multiple problems related to social dynamics (as mentioned, from opinion dynamics to information diffusion or crowd behaviour). An impressive work from the Physics field that attempts to integrate multiple models, comparing them from the same approach, is Castellano et al. work [43]. The works of Carles Sierra et al. constitute an interesting attempt to systematise the social dynamics involving negotiation [85, 195], from the basic ones of game theory to the complex argumentation-based systems, proposing frameworks to integrate the different approaches [154]. More general works from the software engineering field try to make a contribution for providing an architecture to the agent social actions. Part of them tackle the classical question of social order: 'How is it possible to maintain social order in the face of intelligent agents that are capable of autonomously taking self-interested decisions?' [153]. This is the case of Castelfranchi's [42] which formalises the implications and differences of the dynamic social order and social control architectures, reviewing different ways of approaching it from engineering. This is one of the main issues of the newly term coined 'socionics', which is an attempt to imitate 'bionics' but with Sociology and Computer Science in mind [179]. Some other works merge social theories into software engineering in order to approach the same issue. This is the case of the works of Rubén Fuentes et al. [92, 93] who use activity theory in order to handle conflict resolution in agent simulations, from a software engineering approach [91].

Some of the attempts made to model social dynamics have used fuzzy logic as an auxiliary tool for ABM. FLAME [78] introduces it in order to build an adaptive model for emotional behaviour, with important learning capabilities. On the other hand, first Cioffi-Revilla [48] and afterwards Jens G. Epstein [79] make a call for the use of fuzzy logic in social simulation. J.G. Epstein shows some of its potentials through the extension with fuzzy logic of Joshua M. Epstein and Robert Axtell's popular Sugarscape model [82].

The following section will provide a summarised overview of Mentat social dynamics without taking into account the use of fuzzy logic, which is introduced in sections 5.3 and 5.4. The communication among the neighbours is tackled in section 5.5, while the whole friendship dynamics are described in-depth in 5.6. The last sections discuss the results of applying fuzzy logic and some concluding remarks.

5.2 Overview of Mentat Social Dynamics

The social dynamics of Mentat have been fuzzified in order to improve their behaviour. However, such social dynamics can be outlined without taking into consideration if they are crisp or fuzzy. This is the purpose of this section: to provide a summarised general insight of Mentat interactions in order to achieve a comprehension of them without considering fuzzy logic. The following sections will explain how these mechanisms are improved through a fuzzification process.

Similar to many examples in the literature, Mentat uses as world space a 2D squared grid NxN, whose characteristics can be found in section 6.2. Among the different types of grid neighbourhoods explained later on in 5.5.1, the 'Extended Moore Neighbourhood' of radius 6 was selected, in order to reach the average number of personal friends [116]. Thus, each agent can have local interactions with an average of 52 agents (based on the mean density), who may end up being acquaintances or friends. The insight of these calculations can be found in section 5.5.1.

Each agent compares itself with all the other agents that is able to establish communication with. Thus, a similarity function is defined to be able to quantify such comparison: the more similar two agents are, the greater value its similarity function will provide (both similarity functions are further specified in section 5.5.2). Such similarity function (for in-depth descriptions of the similarity functions tested, see 5.5.2) depends on a subset of the agents' attributes, which are compared one per one to compose the total similarity. This subset was selected following the expert sociologist advice and the sociological model, as explained in 3.2.1.

Then, neighbours can be rated and ordered in terms of similarity. On the other hand, the 'proximity principle' states that the more similar two individuals are, the stronger their chances of becoming friends are (for a throughout sociological analysis, see 5.6.1).

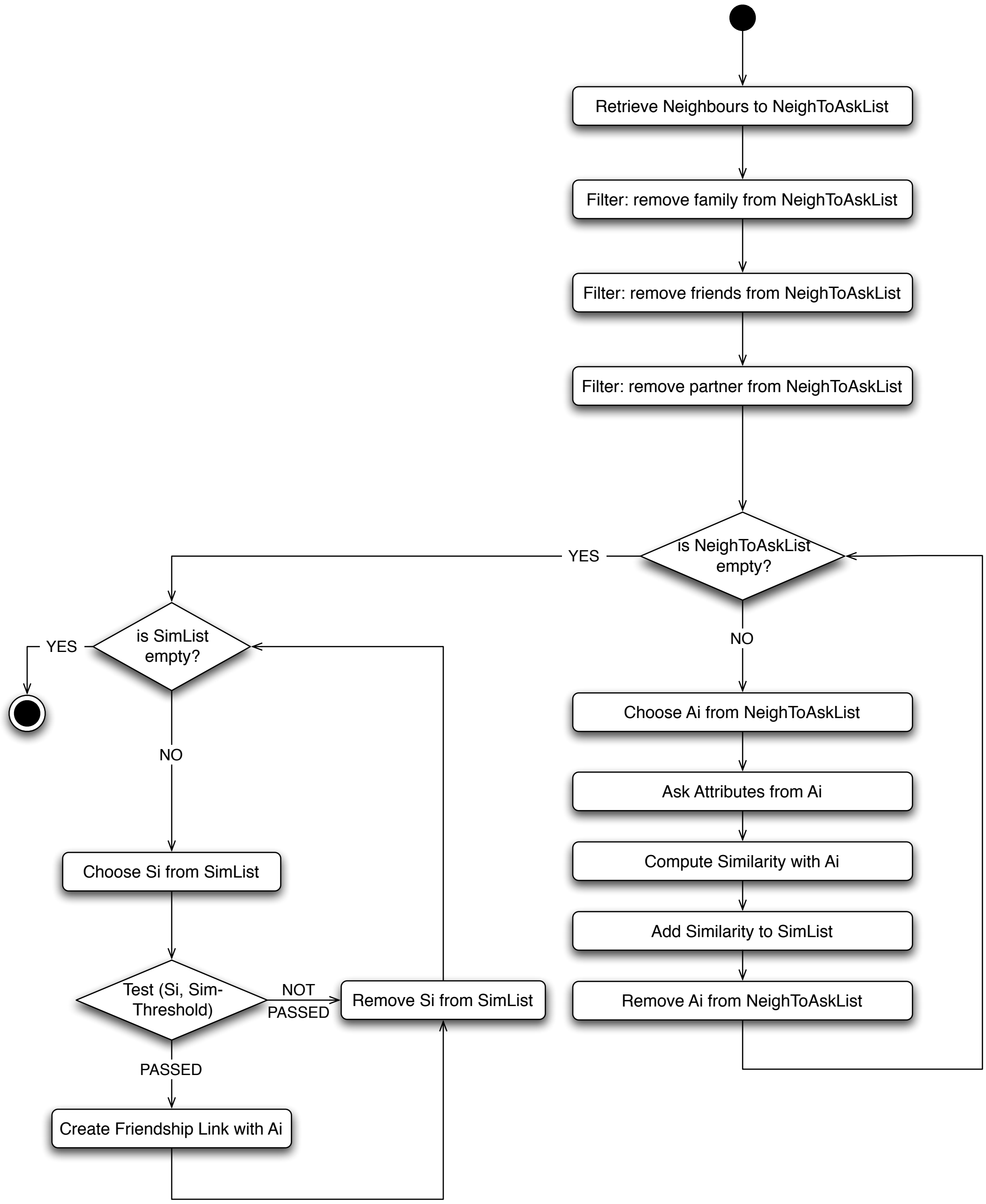

FIGURE 5.1: Flow Diagram summarising the process of emergence of a friendship link.

Thus, the chances of two neighbours of becoming friends is directly proportional to their similarity. Moreover, if friendship is considered a gradual relationship, the more similar two agents are, the better friends they will become over time. This friendship process is summarised in Figure 5.1. There are several methods for quantifying this emergence and evolution of friendship, and the one selected for Mentat is described and supported in 5.6.2.

The same as neighbours, friends can be also rated and ordered in terms of similarity, which is useful for the matchmaking process. First, friends are filtered taking into account Boolean questions such as 'is this agent single?', 'is it an adult?', 'does it have

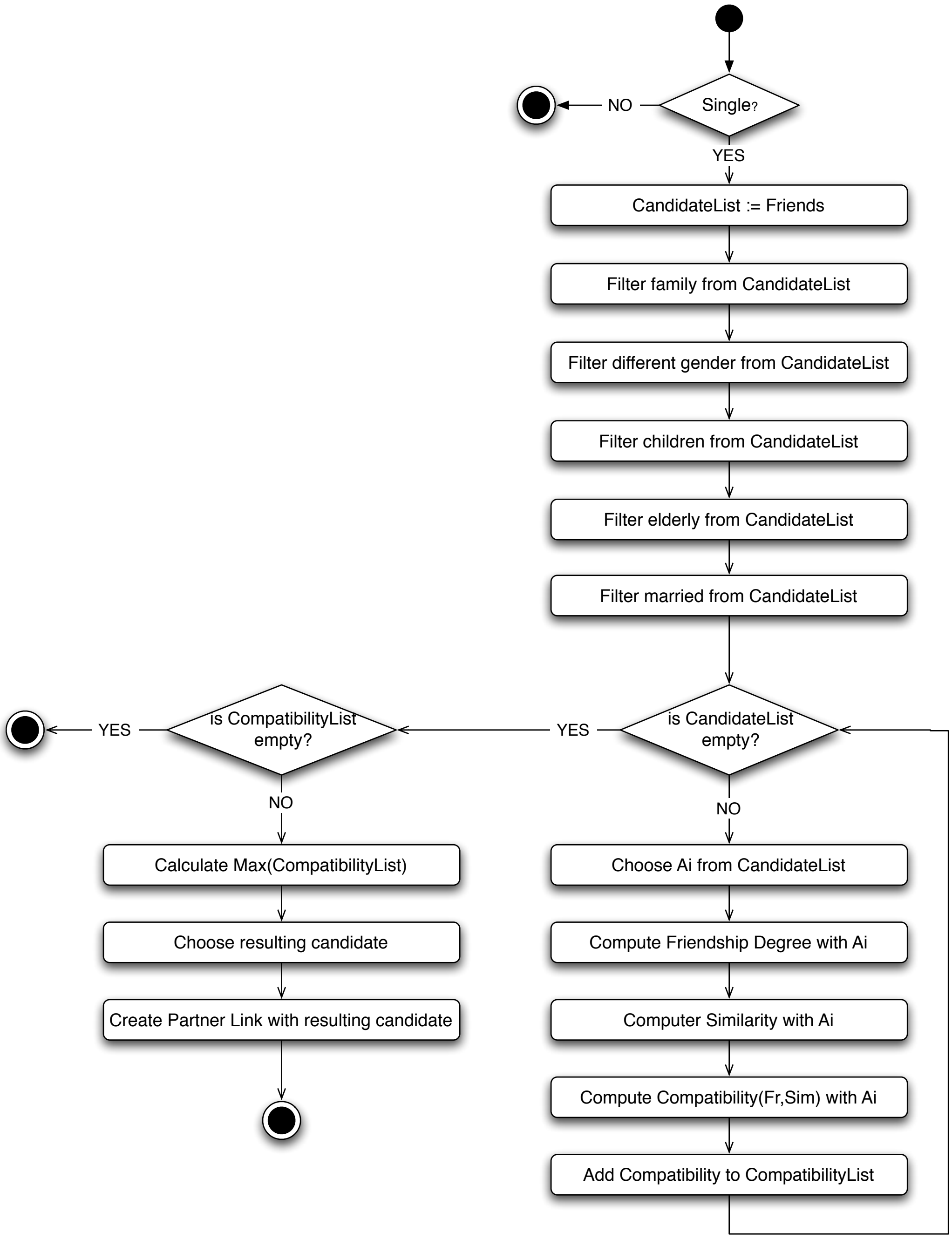

Figure 5.2: Flow Diagram summarising the process of partner finding (matchmaking).

opposite gender?'. Afterwards, the selected subset are the 'candidates' for becoming a spouse. The selection among the candidates is followed in the order defined by similarity: the first one of the queue who passed the filter will be chosen (check 5.6.3). This matchmaking process is summarised in Figure 5.2.

As far as couples appear, so do families. Each couple will have chances of having some children following the demographic model criteria, specified by a set of probabilistic equations defined in 6.3.3.

The whole collection of social interactions and associated behaviours have been outlined here without the need of fuzzy logic. However, in the following sections, its introduction will be tackled and its utility for the modelling of the explained social dynamics, proved.

5.3 Introducing the Fuzzy Mentat

The following sections analyse how fuzzy logic can be useful when building models for agent-based social simulation (as it was described in the works [48, 125]). This analysis comprises different aspects that can be *fuzzificated* from an ABM, in order to improve and refine the social processes modelled. Following this approach, it is applied to a specific social problem: the current model Mentat.

As it was exposed in the previous section, the social dynamics in Mentat are based on the social relationships appearance and evolution, taking into account other relevant issues such as communication or similarity.

Even though it can be found many literature regarding friendship networks, weak links / acquaintances, relationship evolution and so on, researchers are still far from understanding all the processes involved. After studying the available theory, it has been decided to use the 'proximity principle' in order to model friendship dynamics. This principle affirms that the more similar two individuals are, the stronger their chances of becoming friends are. Thus, this chapter attempts to model the processes in which strangers turn to be acquaintances, those turn into friends, and some friends into couples (as it was explained in Hassan et al.'s [131]).

In order to do that, this approach will start with an already implemented ABM, a crisp (non-fuzzy) Mentat, in a mature stage of development [118]. And from that point, the application of the friendship modelling is accomplished using fuzzy logic. Therefore, it is exposed how the theory has been guiding the fuzzification process step by step, resulting in a fuzzy ABM, the final version of Mentat. The whole process consisted in the fuzzification of the agent characteristics, the similarity process, the fuzzification of the friendship relationship together with the introduction an evolution function, and a new partner matchmaking calculation. Comparing the results of the different ABM, it can be assessed that the fuzzy version deals with the problem in a more accurate way.

5.4 Preliminary: Fuzzy logic in a fuzzy environment

5.4.1 Framing the Issue

Certain complex problems tackled by ABM cannot be properly modelled with the typical simple agent models. There are several ways of tackling such complexity, and depending on the studied context, one or another may be used.

An example of a system that requires further considerations in agent modelling could be some sociological analysis derived from the European Value Study and the World Value Survey by Inglehart [193]. In these surveys there are many questions about the degree of happiness, satisfaction in different aspects of life, or trust in several institutions. Although there is some kind of categorisation for the possible answers, such as 'Very much' or 'Partially', there is always some degree of imprecision, which is difficult to model with discrete/crisp categories. Even more, when the individual is evolving to different positions, some of these values get even more undefined. This issue arises also when modelling agent relationships such as friendship: is it possible to measure a degree of friendship between two persons?

Similarity and friendship degree are blurry concepts, and this uncertainty must be modelled rigourously. In this context, an appropriate way of increasing the model complexity in order to improve its refinement can be Fuzzy Logic. Fuzzy logic is oriented at modelling the imprecise modes of reasoning in environment of uncertainty and vagueness, an usual feature in the Social Sciences realm [233].

5.4.2 Reviewing the Importance of Fuzzy Logic in ABM

Individuals are often vague about their beliefs, desires and intentions. They use linguistic categories with blurred edges and gradations of membership, for instance: 'acquainted or friend'. Fuzzy logic is oriented at modelling the imprecise modes of reasoning in environment of uncertainty and vagueness [267, 269], as it has been discussed above, for some aspects in the study of human societies. Thus, because vagueness is such a common thing in the social realm, fuzzy logic provides us with a useful way to handle this vagueness systematically and constructively [233].

There is an increasing interest among social scientists for adding fuzzy logic to the Social Sciences toolbox [157]. Likewise, even though it is still incipient, there are several examples of researches linking fuzzy logic with social simulation. Since the early proposal of Cioffi-Revilla [48], in which the author gave preliminary arguments for the use of fuzzy logic in Social Sciences and even computer simulations, some models followed this path. For instance, in some models, agents decide according to fuzzy logic rules; 'fuzzy controls' or 'fuzzy agents' are expert systems based on 'If $\rightarrow$ Then' rules where the premises and conclusions are unclear. Unlike traditional multi-agent models, where these completely determined agents are an over-simplification of real individuals, fuzzy agents take into account the stochastic component of the human behaviour.

Some authors have proposed to improve the agents' strategy choices within the iterated prisoner's dilemma using fuzzy logic decision-rules [172]. Others researchers have claimed that simulation based on two-player games can use fuzzy strategies when analytic solutions do not exist or they are computationally complicated to obtain (because

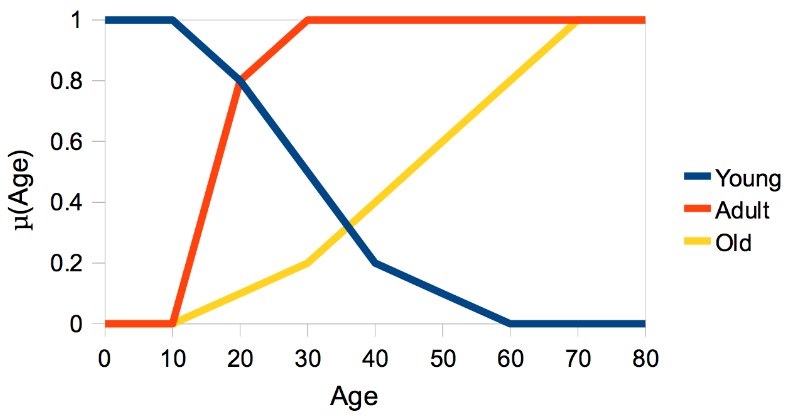

FIGURE 5.3: A graphical example of concept definitions in fuzzy terms.

agents use fuzzy strategies; i.e., 'If I think my opponent will choose action x, I will choose action y') [258]. Agent fuzzification has been already applied for social network analysis, as in the work of Carbó, Molina and Dávila [39], but considering mainly relationships among agents. Relevant examples of fuzzy multi-agent based models are the Fort and Pérez work on spatial dilemmas [89], Jens Epstein's extension of the sugar-space model with fuzzy agents [79] and interesting computational modelling of 'fuzzy love and romance' [231]. The most of these are abstract KISS ABM where fuzzification helped to model the social process studied.

5.4.3 Introducing Fuzzy Sets

It is common in ABM to take into account Social Sciences works or experts advice, who should be consulted repeatedly along the process. However, frequently they use linguistic variables and terms to express their knowledge. This linguistic concepts will probably be better represented with fuzzy sets. In this section, a few basic mathematical definitions of general fuzzy set theory are introduced, in order to support the comprehension of the fuzzification process carried out in the next sections.

Given a universe of discourse U, a fuzzy set μ U $\rightarrow$ [0,1] on U is a mapping that gives a membership degree in the interval [0, 1] to every element of U [267]. Note that classical sets are particular cases of fuzzy sets. There are many human characteristics that have not a clear boundary or depend on the interpretation or context, such as *young* or *beauty*. They can be represented by a fuzzy set on the set of human people, giving them a membership degree of the characteristic, such as for instance, 0.8 or 0.2. An example of such membership degree μ is represented in Figure 5.3, where μ(Age) can be observed. There, three fuzzy sets (functions) corresponding to three concepts are plotted, with a different membership degree depending on the age of the individual.

Given some fuzzy sets representing characteristics, sometimes it is needed to model algebraical operations on them, as 'young' and 'beauty', that must also be modelled, for example to trigger a rule in approximate reasoning systems. In order to use different operators, it is needed to define those algebras, generalising the classical sets.

Therefore, a fuzzy relation R, generalising the classical logic relations, can be defined as R: $U \times U \rightarrow [0, 1]$ [266]. Fuzzy relations have multiple applications to represent degrees of relations between objects in an universe. Those relations can also be characterised without a clear border, such as the *Friendship* relationship. Classical relations can not express some types of information as 'we are more or less friends', or 'a little bit friends'. Fuzzy relations are mostly used in Artificial Intelligence applications to represent degrees of similarity (that define unclear groups or clusters) or to model implications rules to make inference with uncertainty, imprecision or lack of knowledge.

There are other definitions of fuzzy concepts related to the mathematical insight of the fuzzification process carried out here. See [125] for a formal clarification of basic operators (t-norms and co-norms), properties as t-transitivity or t-indistinguishability, and fuzzy similarity (generalising the classical equivalence relations [268]).

5.5 Communication

5.5.1 Neighbourhood

Similarly to cellular automata [261, 262], ABM usually works on a 2D square grid NxN as the world space where the agents 'live'. Although it is not a requirement, it is a popular simplification of the continuous and complex world space that has proven very powerful regarding of its simplicity. The use of the grid, its characteristics and typology can be found in the subsection 6.2.

In spite of the chosen grid, agents will have an 'interaction neighbourhood' [136] formed by the agents close enough to establish communication. That 'close enough' is defined by two decisions: the radius parameter and the choice of using Moore or Von Neumann neighbourhoods[1] (figure 5.4). In the case of Radius = 1, Von Neumann's is defined as the four cells above and below, left and right, while Moore's is the group of all eight surrounding cells. However, radius can be increased to cover more cells, in a distance equal to the radius. Thus, a cell with Moore neighbourhood of radius 2 will have a maximum (as the cells can be occupied by agents or not) of 8+16 = 24 neighbours. Note that radius could be increased till it covers the whole grid. However, this is not common as the agent paradigm tries to promote local knowledge of every agent (increasing decentralisation). Nevertheless, it can be done adapting the influence of each agent with an inverse power law depending on the distance [167]. In the case of the Mentat model, a Moore neighbourhood of radius = 6 was selected (a maximum of

[1]There are, however, other regular neighbourhood types, such as hexagonal or triangular [135, 262].

168 agents around, and an average of 52 based on the agents density of 3.2 calculated in section 7.2.2), in order to reach the typical number of personal friends [116][2].

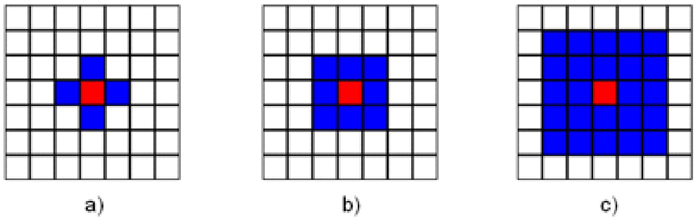

FIGURE 5.4: Different neighbourhoods: a) Von Neumann; b) Moore; c) Extended Moore with radius=2.

As long as agent A is a neighbour of agent B, they will be able to communicate (figure 5.5). This communication can be translated in terms of exchange of messages (with an agent communication language such as KQML [86]), direct influence (affecting the agent internal state), and in general transmission of information in any way. In some systems the agents need to compare themselves with their neighbours, to take decisions based on that new information (maybe selecting only some agents to establish further communication). This comparison is usually carried out through similarity measures. The basic crisp version of Mentat uses a discrete gratification similarity, while the fuzzy Mentat defines a new fuzzy similarity.

5.5.2 Similarity Functions

5.5.2.1 Crisp Gratification Similarity

The basic behaviour of a similarity function is that, the greater the closeness among the two compared objects, the greater the value it must return. In the case of Mentat, agents with a collection of characteristics are compared with each others. This function should determine how similar two individuals are, but in a sociological manner, not purely mathematical. Thus, it is obvious that some characteristics are sociologically more relevant than others to determine such similarity: e.g. age is more important than hair colour (obviously not included in this model) or civil state (included). With the help of a Sociology expert, five main characteristics were chosen to determine the similarity: age, political ideology, religiosity, educational level and economic status. This similarity will be used for the friendship emergence and for the spouse selection.

---

[2]As it is explained in section 3.2.6, the number chosen should be 100, as the average number of contacts. However, due to theoretical discrepancies justified in that section, the number is limited to the number of personal friends

In the basic crisp Mentat, every function was defined as a crisp (non-fuzzy) function. Thus, similarity was not different, and it was modelled and implemented through a not-normalised gratification algorithm. For each characteristic (from the chosen ones) it was defined a range of 'closeness' (with three categories, 'close', 'standard' and 'far' distances). Depending on such 'closeness' of the two characteristic values compared (e.g. the ideology of A and the ideology of B), the similarity counter was gratified (in the 'close' case), left equal ('standard') or penalised ('far'). The sum of all the results of the five attributes comparisons was the output of the function. Thus, this amount of points, could be compared with other results from other agents, and determined which agent was more similar in a crude and superficial way.

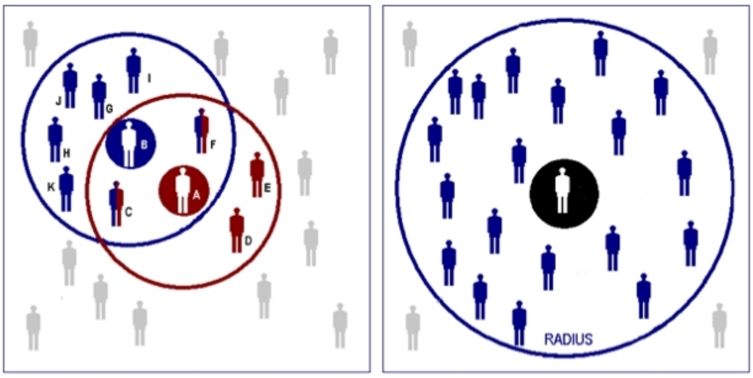

FIGURE 5.5: Graphical representation of the agent communication radius.

5.5.2.2 Building a Fuzzy Similarity

Mentat's gratification technique is not very sophisticated and could be improved. The use of fuzzy logic would significantly increase its accuracy. But the use of fuzzy operators/functions needs the fuzzification of the attributes where they are applied. Therefore, fuzzy sets over these variables are formally defined (in a generalised manner).

Thus, the agent attributes, rather different from each other, were normalised in the real interval [0, 1] (each one depending on its original range). For example, let $\mu_{economy}$:U→[0,1] be the fuzzy set that gives an economic grade based on the economy variable of the individual. This fuzzy set can be defined by segments with different growth (high class, middle class, working class...) or by a linear function. This way, an individual with a $\mu_{economy}(\text{ind}) = 0.7$ would represent a person quite wealthy. Each fuzzy set would be defined by a similar process.

Afterwards, the fuzzy similarity can be defined using a T-indistinguishability, which generalises the classical equivalence relations. It can be obtained from the negation of a T*-distance, where T* is the dual t-conorm of the t-norm T. A complete mathematical explanation beneath this can be found in Valverde's [252], but roughly the distance between the attributes of the two agents compared is 'how far are they', so its negation will point out 'how similar are they'. This way the aggregation of the similarities of each couple of fuzzy sets (by default, normalised) will return the total similarity rate among two individuals. The negation used is a fuzzy strong negation N [228], the distance operator 'd' was defined as the difference of the fuzzy values, and the aggregation chosen is an OWA operator [263]. Thus, the fuzzy relation is defined as follows:

$$R_{similarity}(ind, ind2) = OWA(\forall \mu_i \in ind, N(d(\mu_i(ind), \mu_i(ind2)))) \tag{5.1}$$

An OWA (Ordered Weighted Averaging) [263] is a family of multi-criteria combination (aggregation) procedures. By specifying suitable order weights (whose sum will result always 1) it is possible to change the form of aggregation: for example, the arithmetic average in the example OWA would need a value of 0.5 to both weights. The weights of the OWAs chosen in Mentat configuration will have, by default, standard average weights[3].

Both similarities, the crisp and fuzzy ones, have been implemented using the Strategy design pattern, as it can be observed in the UML diagram 5.6.

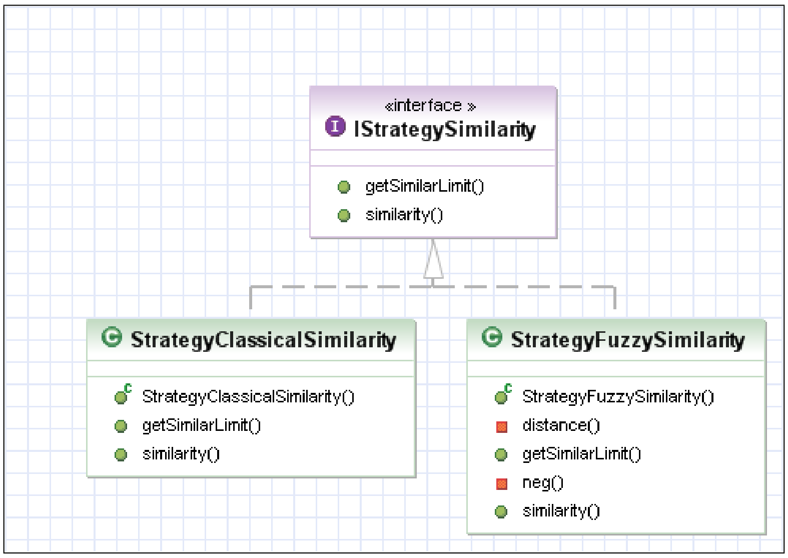

FIGURE 5.6: UML diagram of the different similarity strategies.

[3]There are two reasons to explain why OWAs were used for the formalisation instead of standard average functions. First, the generalisation aspirations of this fuzzification process, so it could be applied in other ABMs, encourage the maximum level of abstraction. The possibility of choosing the most appropriate weights for each component could be useful to adapt to different problems (for example, allowing different weights for each main attribute in the similarity definition). Second, this weight selection is allowed inside Mentat configuration: even if the main testing has been used using a single T (Lukasiewicz) and standard weights, both can be tweaked to obtain different results or to focus on different phenomena.

5.6 Friendship Dynamics

5.6.1 Understanding Friendship

Selecting a friend is among the most personal of human choices, and thus it is not surprising that friendship groups tend toward social homogeneity. Members of the working class usually associate with other workers, and middle-class individuals generally choose friends who are middle class.

A preliminary step to constructing a friendship modelling is an examination of the way that the social context structures friendship choice. Contextual explanations for individual behaviour argue that (i) individual preferences and actions are influenced through social interaction, and (ii) social interaction is structured by the individual's social characteristics [144].

This is consistent with the important homophily principle in social networks of [183]. Principles of 'meeting' and 'mating' by which strangers are converted to acquaintances, acquaintances to friends, and even maybe friends into partner, follow the same rules. Meeting depends on opportunities alone (that is, to be in the same place at the same time, more in subsection 6.2); instead, mating depends on both opportunities and attraction. How readily an acquaintance is converted to close friendship depends on how attractive two people find each other and how easily they can get together.

The 'proximity principle' indicates that the more similar people are, the more likely they will meet and become friends [253]. Therefore, features like social status, attitudes, beliefs and demographic characteristics (that is, degree of 'mutual similarity') channel individual preferences and they tend to show more bias toward homogeneous friendship choices.

5.6.2 Strangers → Acquaintances → Friends: a Fuzzy Logistic Function

Similarity, proximity or friendship are vague or blurry categories, because they do not have clear edges. For this reason, a model of friendship dyads was developed, using the general framework presented above, but considering similarity and friendship as continuous variables. Thus, the fuzzy similarity will be used, and friendship will be redefined as a fuzzy relationship. Besides, because friendship occurs through time, Mentat was considered in dynamic terms, letting the friendship to evolve over time. Let R_{friend}:UxU → [0,1] be the fuzzy relation on the set of individuals that give a degree of 'friendship'. This fuzzy relation gives a degree of friendship in the interval [0, 1] for every couple on individuals. Let Ind be an individual in U. The crisp set Friends (Ind) is defined as the set of all the individuals x in U whose R_{friend}(Ind, x) is greater than 0.

Therefore, every individual will have a range from true close friends to just 'known' people with the rest of individuals. Note that some restrictions to this definition could be introduced in order to suit context needs.

The friendship process is conceived as a search for compatible associates, in terms of the proximity principle, and where strangers are transformed to acquaintances and acquaintances to friends as a continuous process over time.

It is proposed the hypothesis that a logistic function[4] [30] can describe formally the 'friendship relationship' or degree of friendship for every couple of individuals:

$$\frac{dF}{dt} = rF(t)\left(1 - \frac{F(t)}{K}\right) \tag{5.2}$$

The equation expresses the hypothesis that friendship increases over time[5]; thus, at each point of time, *F(t)* defines the minimum degree of friendship that is given as an initial condition ($0 <$ *F(t)* $< K$); K is the maximum degree of friendship that agents can reach (K can be understood as the level of 'close friends'), and finally r value defines the growth rate of friendship. However, this equation does not include the 'proximity principle' described above. This principle can be included in equation 5.2 by modifying the growth rate r and stating it as follows: the more similar in social characteristics two individuals are, the higher the growth rate of their friendship is (r needs to be made sensitive to the similarity value). Thus, the following equation can be expressed:

$$r = S \times J \tag{5.3}$$

Where S is a measure of similarity and J defines a multiplicative factor that increases the magnitude of S within r. The objective of J is turning r more sensitive to S values, and specially sensitive to high S values. For this reason, J describes an exponential growth depending on S values. J can be formalised as follows:

$$J(s) = J_0 \times e^{ps} \tag{5.4}$$

Where J_0 is the initial value of *J*, *P* defines the constant of proportionality and S is the similarity value between the individuals. The following graphs show how the friendship

---

[4]The equation 5.2 used is a standard logistic function, and it was derived mathematically by Peter Hedstrom in 'Dissecting the Social: On the Principles of Analytical Sociology' [134], chapter 'Social interaction and social change'. This equation shows a behaviour consistent with the sociological supporting theory of section 5.6.1 and the advise of a domain expert: a sociologist.

[5]In Mentat, friendship ties cannot be weakening nor broken. If two individuals are not similar enough, their tie will never grow strong, but once appeared it will always exist. This simplification was acceptable according to the domain expert advice.

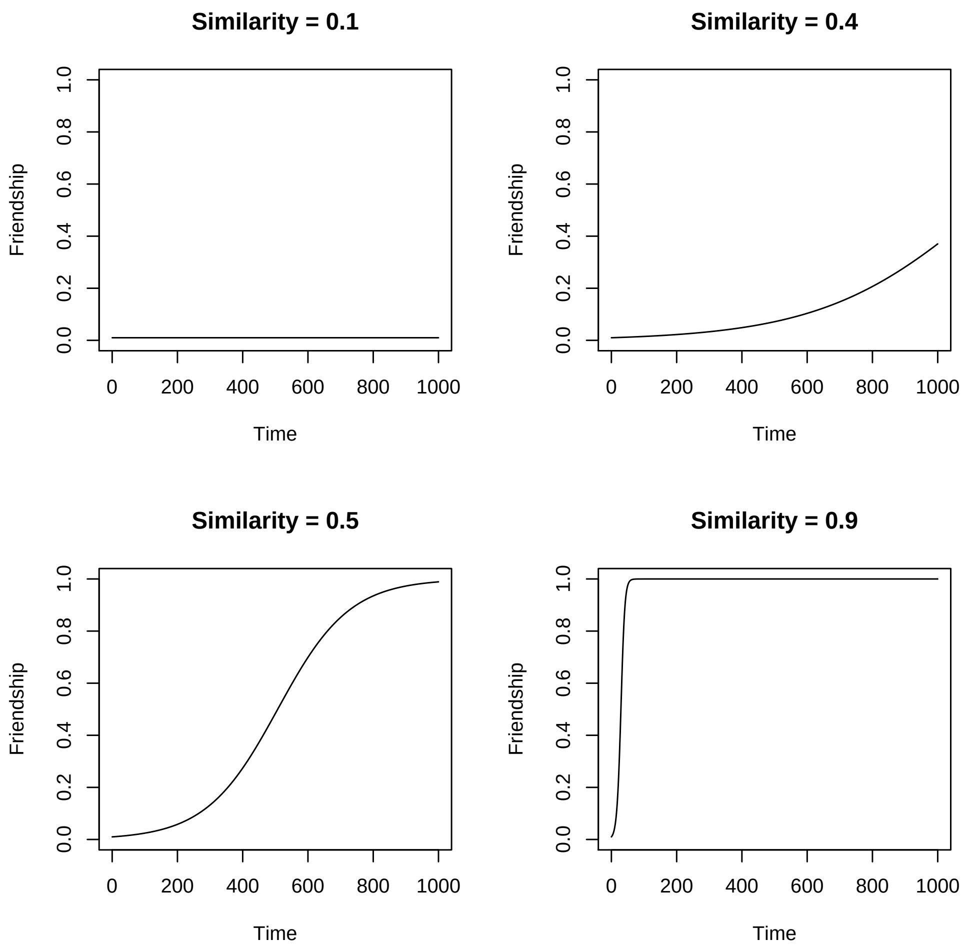

FIGURE 5.7: The logistic function chosen for the evolution of friendship, for several one-to-one similarity values.

will develop over time given different initial conditions[6].

5.6.3 Friends → Partner

The social system also simulates matchmaking. Once a couple is formed (for example, via a marriage), a stable couple can be defined as a crisp relation: two persons are a couple or they are not. It is proposed to learn this classical relation R_{couple}:UxU $\rightarrow$ $\{0,1\}$ by using approximate reasoning and fuzzy inference techniques.

[6]In the Figure 5.7 of the figure it is assumed that P is equal to 5.8 and J_0 is equal to 0.001. In Figure graph-J it is assumed that K is equal to 1 and F_0 is equal to 0.01; r value is equal to $S{\times}J(s)$.

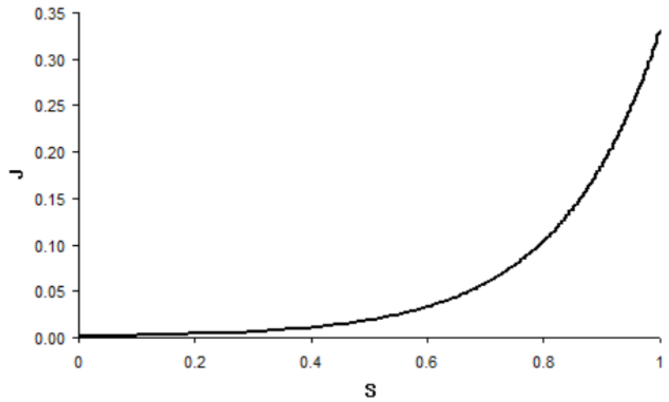

FIGURE 5.8: Exponential growth of parameter J, in relation with the similarity S.

Note that if a R_{couple} relation is known, a 'single' and 'married' crisp set on U is also know, defined as:

$$married(ind) = \begin{cases} 1, & \exists ind_2/R_{couple}(ind, ind_2) = 1 \\ 0, & otherwise \end{cases} \tag{5.5}$$

$$single(ind) = NOT\ `married'(ind) \tag{5.6}$$

It does not exist a specific process for finding a partner in previous social simulation models. After asking a Sociology domain expert, a general definition for it was built up. An agent will find a partner (always with a random possibility of failure not defined in the mathematical definitions) between the friends that have different gender, are adults, and do not have a partner yet. The chosen one will be the 'most compatible' one of its friends, where compatibility is defined as the aggregation of how strong their friendship is and how similar they are (this is an oversimplification with respect to real life, and it is future work to allow some randomness in this process).

This important information of how couples can be formed can be obtained by inferring a fuzzy relation $R_{compatible}$:UxU $\rightarrow$ [0,1], which can be obtained using a fuzzy aggregation operator, [263], and operations on the classical set 'has not partner':U $\rightarrow$ {0,1}, the fuzzy set 'adult':U $\rightarrow$ [0,1], the classical relation 'has different gender':UxU $\rightarrow$ {0,1}, and one fuzzy rule of inference, where the premise is the conjunction of the classical sets.

An OWA is again applied to formally define the $R_{compatible}$:UxU $\rightarrow$ [0, 1] fuzzy relation using the $R_{similarity}$:UxU $\rightarrow$ [0, 1] and the R_{friend}:UxU $\rightarrow$ [0, 1] fuzzy relation, as the following mapping:

$$R_{compatible}(Ind, Ind2) := OWA(R_{friend}(Ind, Ind2), R_{similarity}(Ind, Ind2))$$
$$= w_1 * R_{friend}(Ind, Ind2) + w_2 * R_{similarity}(Ind, Ind2)$$
$$\text{for all Ind, Ind2 in U, where } w_1 + w_2 = 1. \quad (5.7)$$

After computing the $R_{compatible}$ fuzzy relation , a $R_{`couplecandidate'}$:UxU → [0, 1] fuzzy relation can be computed by using the 'single' set and the 'Adult' fuzzy set on U as follows:

$$R_{couple-candidate}(Ind, Ind2) := Adult(Ind) \text{ AND } Adult(Ind2) \text{ AND}$$
$$`single'(Ind) \text{ AND } `single'(Ind2) \text{ AND } R_{compatible}(Ind, Ind2). \quad (5.8)$$

Note that the AND conjunction is implemented by a t-norm. It can be done many times because of the associative property of t-norms. Thus, T(x, y, z) = T(T(x, y), z).

Then, the method to marry an individual Ind is to find Ind2 such that it maximises the fuzzy relation $R_{`couple-candidate'}$(Ind, Ind2). Then it is possible to compute R_{couple}(Ind, Ind2) := 1.

5.6.4 Family

As a result of the friendship dynamics exposed in the previous subsections, agents can find spouses to match with. Those couples may have children, following the restrictions and equations explained in 6.3.3. Thus, it produces the emergence of families with parents, children and brotherhood. The family is represented as a different group type, with a different link colour. No third-level relationships are considered (grandparents, uncles, cousins) in order to focus on the family nuclei. Those nuclei are localised and aggregated in space, as children tend to appear close to their parents. This process promotes the appearance of clusters in the social network, where the new-born agents will relate with the friends of their family, relating to the friends of the brothers or parents and even to their children. This link has not been fuzzified, as it was not found necessary.

Both 'Family' and 'Friends' individual groups have been implemented as 'Group' classes, with common general operations, as shown in 5.9. However, the associated operations of the agent concerning these groups have been externalised in separate classes. Thus, 'IndivFamily' is an extension to the agent class which wraps its 'Friends', while 'IndivFamily' does the same with 'Family', as shown in 5.10.

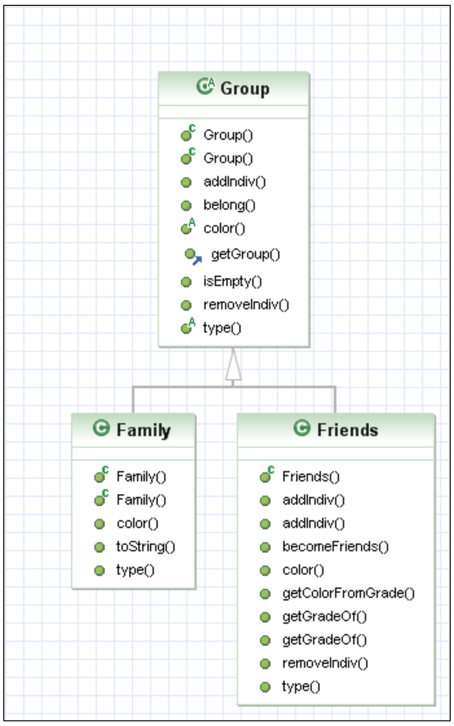

FIGURE 5.9: UML diagram of the possible agent groups.

5.7 Results and Discussion

Here a comparison between several versions of Mentat is presented, although it is better framed and exposed in the experimentation chapter, in section 7.3.3. Four different implementations, each one with two configurations have been analysed, focusing on three measures. The fuzzy modifications have been grouped in two main ones: '*Fuzz-Sim*', when the attributes are normalised and the similarity operator fuzzified, as stands the subsection 5.5.2.2; and '*Fuzz-Fri*', when the friendship turns to be fuzzy, evolving over time and affecting the partner choice (subsection 5.6.2). The four ABM represent all the possible combinations between these, represented in the table 5.1 in the pair (*Fuzz-Sim*, *Fuzz-Fri*). Thus, there is a crisp version of Mentat, $Mentat_{Crisp}$ (with its crisp similarity explained in section 5.5.2.1), with no fuzzy properties; the $Mentat_{FuzzSim}$, simply the same model but with the '*Fuzz-Sim*'; the $Mentat_{FuzzFri}$ with '*Fuzz-Fri*' but not '*Fuzz-Sim*'; and last $Mentat_{FuzzAll}$, with all the fuzzy modifications (that is, both *Fuzz-Sim* and *Fuzz-Fri*).

The two configurations deal with two possible ways of friendship emerging: one promoting random friends (and therefore an agent can be linked to a non-similar neighbour) and the other promoting similarity-based friends (and therefore an agent will rarely be linked to a non-similar neighbour, as it will give priority to the most similar ones). This is not a trivial decision, because the friendship evolution function already deals with

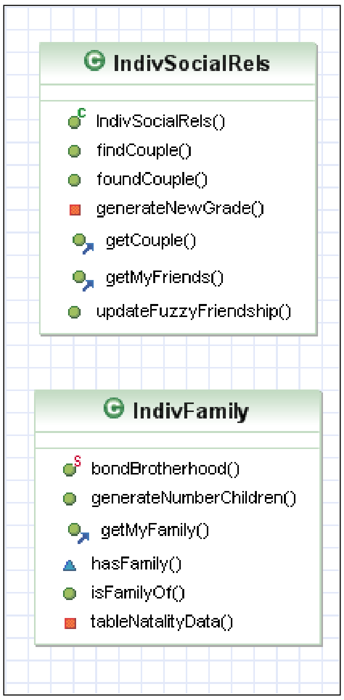

Figure 5.10: UML diagram of the agent class extensions with the social relationships related methods.

similarity, and if a neighbour is not similar at all, it will never be more than an acquaintance. It is not evident if the closer way to real-world is giving double strength to similarity (in the second option) or letting randomness to decide who will be the friend (and thus maybe ignoring similar people). It has to be mentioned that none of the two configurations is so deterministic and both are based on probabilities.

The parameters analyse the couples and how they are affected by the changes in the configuration and fuzzification. The $R_{Similarity}$ shows the proximity taking into account all the characteristics of each partner in a couple. The $R_{Friendship}$ focuses on the friendship link between them, which in a way (according to the logistic function) depends on their similarity too, but also in the time spent together. The $R_{Compatibility}$ is taken as an average of the other two. The values have been obtained after averaging the output of several executions of each version. Thus, in every execution, it is the mean of the property in every couple.

As the first two models have a Boolean friendship, their compatibility is always the same as the similarity. In the first configuration, when the friendship is rarely involved in the neighbours linking, the similarity rates are noticeably similar in all the versions. However, in the second one it is clear that the ones with fuzzy similarity slightly increase their success. But the bigger changes can be observed in the friendship: $Mentat_{FuzzAll}$ beats the other versions with a greater $R_{Friendship}$ and $R_{Compatibility}$, specially in the second configuration. These results approach the theoretical qualitative assessments made by the sociological friendship theory considered [144, 253] and the domain expert.

Intuitively, the couples of $Mentat_{FuzzAll}$ are more similar to each other (following the proximity principle), and take into account the degree of friendship in the process of choosing partner (which implies historical evolution of the relationship).

TABLE 5.1: Comparison among the different ABM, in increasing order of fuzzification.

	$Mentat_{Crisp}$	$Mentat_{FuzSim}$	$Mentat_{FuzFri}$	$Mentat_{FuzAll}$
	(0,0)	(1,0)	(0,1)	(1,1)
Config. Random-friendship				
Mean couple $R_{Similarity}$	0.76*	0.77	0.76*	0.77
Mean couple $R_{Friendship}$	(**)	(**)	0.72*	0.80
Mean couple $R_{Compatibility}$	0.76*	0.77	0.54*	0.62
Config. Similar-friendship				
Mean couple $R_{Similarity}$	0.73*	0.77	0.73*	0.78
Mean couple $R_{Friendship}$	(**)	(**)	0.54*	0.76
Mean couple $R_{Compatibility}$	0.73*	0.77	0.39*	0.59

*: The crisp Mentat similarity has other range, but here they have been normalised in the interval [0,1] in order to be compared.
**: When the friendship is not fuzzified, all the couples are friends (as this is a Boolean property).

5.8 Concluding Remarks

This chapter has exposed Mentat's social dynamics in-depth, detailing the social relationship mechanisms and how the social network emerges. Besides, throughout the different sections a fuzzification process was carried out, redefining several crisp model concepts as fuzzy concepts and relations. Fuzzy logic improved the behaviour of the agent-based model in those aspects with some degree of uncertain knowledge.

Thus, some concepts of social relationship dynamics were explained, including an evolution function that was applied for the changing of friendship over time. And fuzzy logic was introduced step by step. Therefore, fuzzy sets over each agent attribute were defined, together with a new fuzzy similarity operator that would influence friendship emergence and partner choice. The friendship relationship nature and importance in the model was significantly modified, fuzzifying it, making it evolve using the defined logistic function, and letting it influence in the partner choice as much as the similarity rate. The results of these changes were found positive, as long as they improve the proximity to the qualitative assessments of the theory.

To sum up the fuzzy application, a sociological friendship theory has been exposed. It was searched where it can be applied, found the useful tools to do that, implemented the application and extracted a collection of results that are used to validate the model against the theory.

Future research lines that could be followed could take into account other interesting friendship theories. There are deep studies in homophily in social networks [183] that could be implemented. An aspect that the Mentat model ignores but it is important enough to be considered is the stability of friendship [257]. Besides, the fuzzy Mentat could be extended to analyse the importance of weak links along one's life, a new possibility that the crisp Mentat did not allow [56]. Other extensions to the fuzzy Mentat are exposed in section 8.2.2.

Chapter 6

The Macro View: the Global Conditions

It was curious to think that the sky was the same for everybody, in Eurasia or Eastasia as well as here. And the people under the sky were also very much the same–everywhere, all over the world, hundreds or thousands of millions of people just like this, people ignorant of one another's existence, held apart by walls of hatred and lies, and yet almost exactly the same–people who had never learned to think but were storing up in their hearts and bellies and muscles the power that would one day overturn the world.

George Orwell, *1984*, 1948

6.1 Introduction

The micro level of an agent-based model focuses on the behaviour of each individual agent, while the macro level is centred in the agent society as a whole, together with its environment. The previous chapter provides the connection among the two views, linking the micro agent decisions to the social dynamics [80, 270]. Thus, it has described part of the macro level already: the emergent social network. There are, however, other aspects to be considered from the macro perspective.

The agent environment is the virtual world in which the agents interact. It constitutes the space were agents are located, the medium for their communication and commonly their geographical reference when considering local interaction [105]. Every agent-based model has an environment, although there are some examples where this aspect has a reduced importance (such as the global interaction examples of stock market simulations [176]). An environment can be *spatially explicit* if it represents a specific geographical space, such as the mapping of an existing city [95] or a region of states [44]. However, other models can distribute agents in an abstract environment based on other features instead of geography: opinion space [251], knowledge space [104], etc. Section 6.2 will tackle Mentat's environment in-depth.

Other macro aspects must be taken into account as well. That is the case of demography, which can play a fundamental role in certain agent-based models [29]. Thus, there are interesting examples such as the data-driven model on German migration patterns of [138], which recalls to Schelling's pioneer works [225]. More related to Mentat's model, there is [247], that studies marriage and mate through a combination of top-down demographic approach and a bottom-up psychological approach (with a bounded rationality cognitive model). This combination is somehow similar to the one presented in Mentat, with bottom-up social dynamics and top-down demographic equations, which will be presented in 6.3.3.

Finally, the model macro behaviour depends on other specific design decisions: the activated modules and the configuration parameters. Those aspects constrain the agent society evolution and must be taken into consideration: this is explained in section 6.3, together with the description of its UML design.

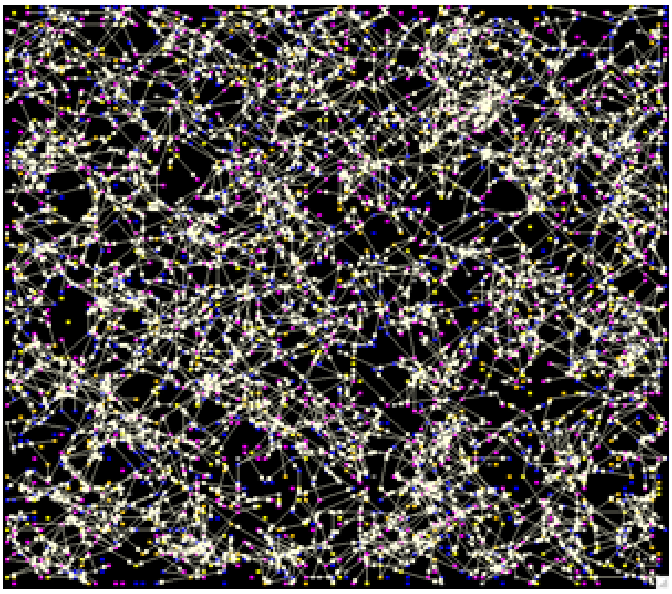

FIGURE 6.1: A global view of Mentat in action.

6.2 Environment: A World. A Space. A Grid.

Mentat attempts to simulate the Spanish society in the last decades of 20th Century. The model is fed by the Spanish section of the EVS-1981, as this data collection is a statistical representation of the Spanish society. However, the agents are constrained in a common space where they can know their local neighbourhood and have different kinds of relationships. This reductionist step has its underneath logic in statistics and Quantitative Sociology, as explained in section 3.2. Thus, the aggregation of these thousands of agents in the same space will represent the whole society behaviour along time. But this has an obvious dilemma associated: although the original EVS individuals are distributed all across Spain, it does not make sense to map the model agents in the Spanish territory. If that was the case, how would they be able to contact each other? The constrain of these agents in the same space is a must.

Therefore, the agents 'soup' must float in some kind of liquid, with certain 'physical' constrains and rules. This world geography must be rather abstract and generic, in order to be able to generalise the real world: not the Spanish map, but an abstract environment where they could relate to each other. In ABM, inherited from the cellular automata, it is common to represent the abstract space with a 2D squared grid NxN [136]. The main issue that this abstraction can imply is the 'town effect' explained in section 3.2.6.

Although its obvious limitations, grids are a popular and acceptable simplification of the continuous and complex world space that has proven remarkably powerful regarding of their simplicity. The use of a grid is independent from other system characteristics, such as mobility (although usually only in two dimensions), agent complexity (as they can have an advance cognitive model in spite of their simple graphical representation) or communication (the agents are not isolated, but in contact with others agents).

There are several grid world types: they can be a torus (where North is connected to South and East to West) or allow walls (with a sphere where E-W are connected, or a square where no walls are connected). They can allow several agents in the same 'cell' or not. They can allow blanks (spaces with no agents) or not. They can be squared or with other shapes, such as hexagonal (with each cell having 6 neighbour cells). Even they can have more than two dimensions.

In Mentat, the grid used is a standard 2D squared Grid, though a 2D squared torus can be selected. However, as the differences among them are minimal, and the interaction in the results is not significant[1], along the text the world is considered a 2D grid. Such grid allows several types of neighbourhood, and as it was described in 5.5.1, an extended Moore neighbourhood of radius 6 was chosen.

It could be argued that, as the agents do not move nor have different environments such as work place, studies centre or home, this 2D space is rather limited. Two agents who are far in the grid will never communicate nor be able to be friends. Thus, agents local interaction may not be logical in this context. However, the counter-argument is

[1]The torus allow a few more interactions that were not possible before, minimising the possibilities of isolated agents. However, in Mentat such agents are rare or inexistent. Therefore, the influence of the torus in the output is anecdotic: a small rise in the average number of friends is the most noticeable consequence.

that this 2D grid does not represent only the geographical space. It represents the random component of friendship: the 'meeting', different than the 'mating'. As explained in 5.6.1, meeting depends just on opportunities (that is, to be in the same place at the same time); instead, mating depends on both opportunities and attraction. Both processes together are the basis for the conversion of strangers into acquaintances, friends or even spouse. It is obvious that the 'mating' is decided by the agents themselves, as explained in chapter 5. But the 'meeting' depends on living in the same area (common geographical space, such living in the same city), but as well on studying/working in the same place (the other layers of different local environments). The 2D grid represents the whole process and represents it spatially. Therefore, if agent A is close to agent B it is because they can find each other in the same space and time. It does not matter if B lives in another city and just goes every weekend to the same NGO as A does. However, if B is located far from A in the grid, that means that they do not share a space-time, even if they are neighbours in reality, and thus they do not have chances of becoming friends. It is a similar idea to the opinion space found in several opinion dynamics models [137]. An alternative solution would consist in defining several spaces (grids) where agents are located. Thus, the same agent would have a location, for instance, in the 'housing' space and in the 'work' space. In each of those environments, the agent would develop relationships (with neighbours and work colleagues). However, the opinion space option has been preferred, as it fits better with the friendship social network concept (the transitive property is well-defined in a single network but not in a collection of networks).

The world space in Mentat has been encapsulated in the class *WorldSpace*, which makes use of the Repast *Object2DGrid* an equivalent to handle the grid properly. Its most important methods, together with the auxiliary *Bond* class (an extension of Repast *DefaultEdge*), can be seen in 6.2. An equivalent to both edges and grids classes can be found in other simulation frameworks such as Mason, Swarm or NetLogo [206].

6.3 Insight of the Model

6.3.1 The Model Design

The agent architecture has been structured in a modular way, using a layer architectural pattern, as described in section 4.3. The global model is built with similar aims by using the architectural pattern Model-View-Controller (MVC). In MVC, the system is divided into three different parts: a Model that encapsulates application data and the logic that manipulates that data, independently of the user interfaces; one or multiple Views that display a specific portion of the data to the user; and a Controller associated with each View that receives user input and translates it into a request to the Model. Views

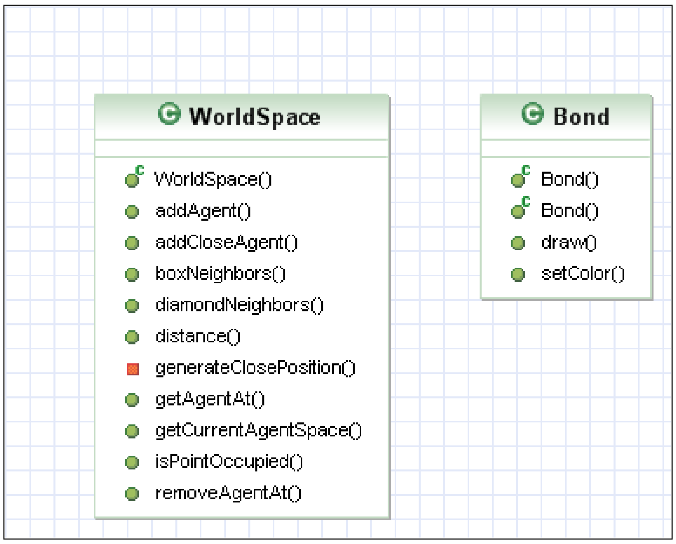

FIGURE 6.2: UML diagram of the world classes.

and Controllers constitute the user interface. The users interact strictly through the Views and their Controllers, independently of the Model, which in turn notifies all user interfaces about updates. This way the model logic can be completely isolated from the graphical interface [19].

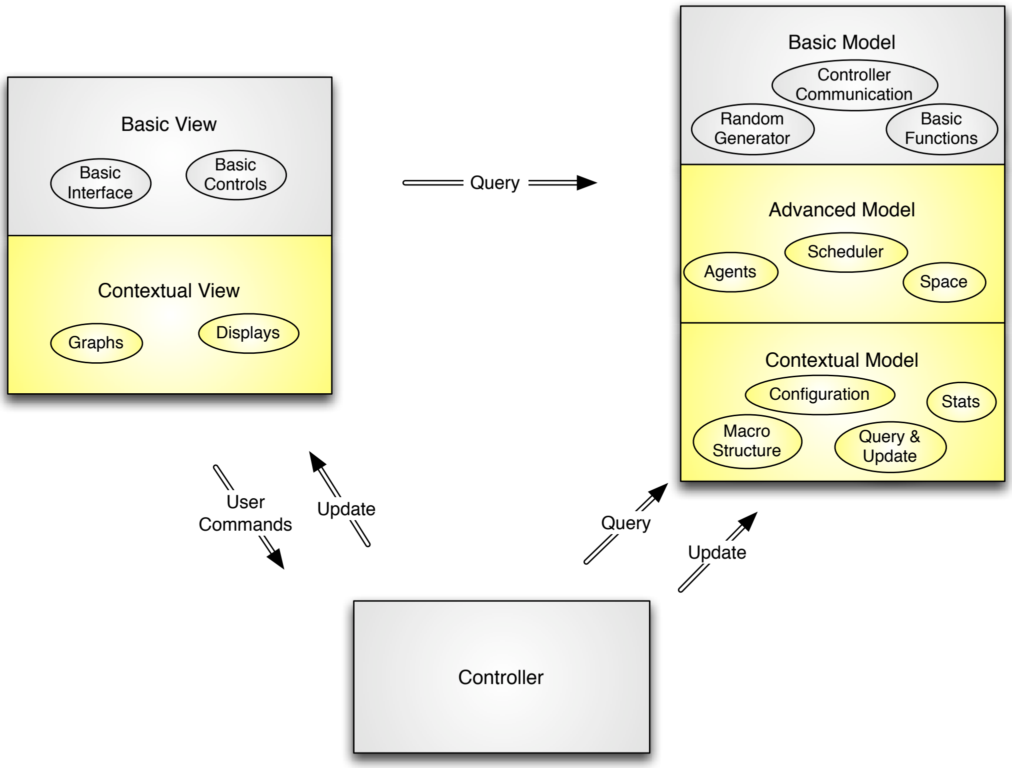

FIGURE 6.3: The Model Architecture from the Model-View-Controller (MVC) pattern perspective.

Figure 6.3 shows this MVC architecture of the Mentat agent-based model, where the Model and the View are subdivided in several modules. The three darker modules (Basic View, Basic Model and Controller) are frequently implemented in the simulation framework used (Repast, Mason, Swarm) while lighter modules (Contextual View, Advanced Model and Contextual Model) are extension layers for the current architecture:

1. **Basic Node:** low-level functions for a basic agent model, such as initialising random generators, needed programming functions (event listeners), play/stop/pause simulation capabilities, etc. It usually corresponds to the basic model class from the chosen simulation framework (such as *SimModelImpl* in Repast [194], in the case of Mentat).

2. **Advanced Node:** basic and minimalist but executable model, with the agent collection, a space for them and some needed functions (e.g. schedule, step, shuffle, print). In Mentat, it corresponds to the class *BasicModel* shown in Figure 6.4.

3. **Contextual Model:** context-based model, where the specific characteristics and operators that depend on the studied problem are implemented. Depending on the ABM structure, it can end up holding big loads of computation. As an example, in Mentat the class *ComplexModel* of Figure 6.4 contains the statistics calculation (*Stats* class in Figure 6.5), the social-structure class that loads the initial population (*SocialStructure* class in the same figure), and the collection of configuration parameters (*Config* class).

4. **Basic View:** this is the main graphical support that typically every simulation framework would provide. A basic interface and collection of controls to interact with the user are implemented here.

5. **Contextual View:** this layer is an extension to the Basic View which should provide the additional graphical part that the particular model needs, but being isolated from the rest of the model functions. It can deal with the whole graphical output of the model, supported by the functions given by the Basic View, including the visual representation of the agent simulation, and the real-time graph generation. In Mentat, this layer is composed by two model classes: *SpaceModel* and *GraphicModel*, from Figure 6.4. The first one was needed just for programming purposes: as a container for the auxiliary classes needed for the data source sequences in Repast graphs.

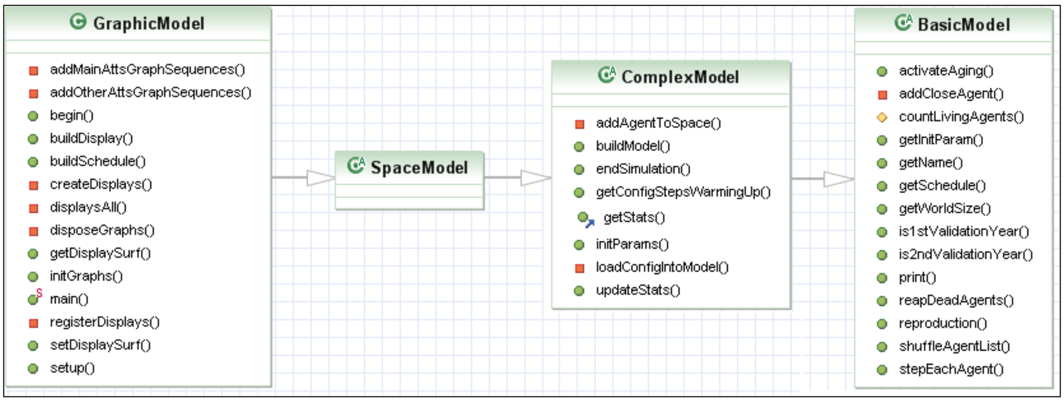

Figure 6.4: UML diagram of the model hierarchy classes.

6.3.2 Configuration: Main Parameters

Mentat has been implemented with an emphasis on modularity and configurability. In order to configure the different social processes taking place (adaptive friendship network, matchmaking, demography) a collection of parameters is needed. As one of the aims of this work is to build a flexible social agent framework, additional activation parameters (shown in Table 6.2) for the enabling or disabling of modules have been implemented. Thus, Mentat can be understood as 'several models in one': the exploration of the model space is facilitated through the activation of some of the mechanisms, comparing the results with other combinations. Thus, a crisp Mentat can be compared with a fuzzy Mentat (or even just partially fuzzy), an empirical initialisation with a random distribution, several levels of demographic dynamics, and different combinations of all of them. Besides, each of the models (each of the different active modules combinations) can be configured with different parameters (shown in Table 6.1), testing different number of agents, thresholds, neighbourhood, density, etc.

In the table 6.1, the main parameters are listed. A few parameters were not included in the system as they were not considered relevant enough, such as the choice of 2D Grid or Torus, or the choice of Von Neumann or Moore neighbourhoods. Other parameters were not included, as they are a result of the included ones, such as the *Number of agents*, which depends on the *Base Population* and the *Introduce Initial Children*. Several parameters have been useful while testing and 'deepening' (subsection 2.4.3) while exploring the possible models space.

6.3.3 The Demographic Model: Equations

One of the key elements of the model is the demography. In the beginning it was implemented through a set of statistical distributions: Normal (av,sd) curves where 'av' (average) and 'sd' (standard deviation) were fixed taking them from empirical research studies. However, these demographical variables, such as the number of children per

Table 6.1: Main parameters of Mentat.

Name	Range	Default	Comments
General Config.			
Base Population	500, 2303	2303	First versions of Mentat used a random sample of 500 agents over the possible 2303 full sample provided by the EVS.
Warming-up	Integer	100	Tests were done with a warming-up from 0 to 1,000 steps, and the best results were found around this value (7.2.4).
Similarity limit	Real	0.95 / 2.5	Threshold for considering two agents similar (fuzzy & discrete strategies) (5.5.2), found experimentally.
World Config.			
Density agents	Float	3.2	Defines the grid size depending on the number of agents. The equivalence is 3000 agents $\sim$ 100x100 grid. If a density significantly below or over this one is selected, the agent dynamics vary rather a lot. The best results according to the friendship sociological literature seem reached around this value.
Neighbour radius	Integer	6	An extended Moore neighbourhood with radius 6, to communicate with $\{168$ cells / density of 3.2 $\} \sim 52$ agents, justified in 5.5.1.
Friendship Config.			
Prob. Random Fr.	Float	0.005	Chance of two known people to become friends (in the beginning, with the minimum link strength). This is the first option for choosing friends, as explained in 5.6. This value has been tested to fit the sociological theory.
Prob. Chosen Fr.	Float	0.4	Chance of two similar people to become friends (in the beginning, with the minimum link strength). This is the second option for choosing friends, as explained in 5.6. This value has been tested to fit the sociological theory.

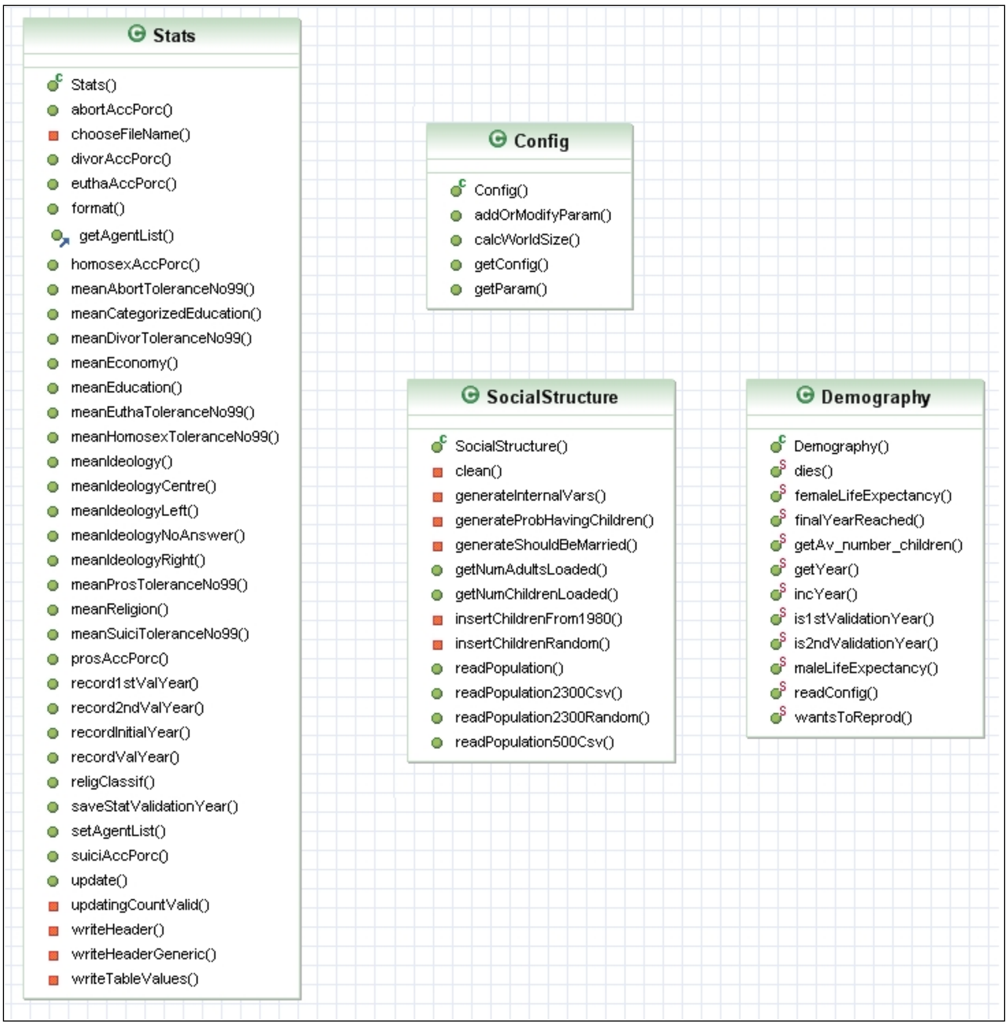

FIGURE 6.5: UML diagram of the auxiliary model classes, for the statistical measures, parameter configuration, initial population and demographic dynamics.

couple, are not static over time. They evolve slowly but noticeable in 20 years. This is why these static distributions were substituted by demographic equations which represent the evolution of these parameters. The use of this data does not break the potentials of projection and prediction of the framework, as it is explained in subsection 3.2.1. The process and reasons for introducing these equations in Mentat is tackled step by step in subsection 7.2.5. For every equation, it was followed the advice of a domain expert (a sociologist) who chose the optimal data sources for each needed mechanism, and helped to build a proper equation for it. Such data sources are mainly the actual EVS and the statistical data, temporal series and projections from the Spanish 'National Institute of Statistics' (INE [152]), together with a subpart of INE: the Spanish census [234]. For each case, first an equation was directly derived from the empirical data (such as Figure

Table 6.2: Activable mechanisms of Mentat.

Name	Range	Default	Comments
Module Activation			
Empirical Init.	Boolean	True	Enabled: population is initialised loading the EVS selected variables. Disabled: population randomly initialised.
Initial children	Boolean	True	Enabled: 716 new agents are introduced (explained in 7.2.3).
Fuzzy similarity	Boolean	True	Enabled: the similarity follows the fuzzy strategy (5.5.2). Disabled: the gratification strategy is used (5.5.2).
Fuzzy friendship	Boolean	True	Enabled: friendship is fuzzy and can evolve over time following a logistic function. (5.6.2). Disabled: friendship relationship is Boolean.
Eq # Children	Boolean	True	Enables the equation for the evolving number of children per couple (6.3.3).
Eq Marital status	Boolean	True	Enables the equation to follow the EVS marital status attribute (6.3.3).
Eq Prob children	Boolean	True	Enables the equation for the probability of having children (6.3.3).
Eq Life Expect.	Boolean	True	Enables the two equations for the life expectancy evolution (6.3.3).
Eq Age 1st Child	Boolean	True	Enables the evolution equation for the age of mothers in having the first child (6.3.3).

6.6), and afterwards, if possible, a linearisation[2] of such equation was calculated, with a minimum error (shown in Figure 6.7).

Thus, these equations are defined as follows:

- *Eq. Number of Children*: the number of children has evolved rather fast in the period under study (demographic data from INE [152] and EVS). In 1980, the average of number of children per couple was 2.2. In 1999, however, it was 1.19. This equation maps this evolution, assigning to every year its respective average. The actual function was decomposed for every year instead of building a simplified equation, for a better match with the data.

- *Eq. for 'marital status'*: the EVS provides specific information about the marital status of each individual. However, this are several issues related to this variable.

[2]The linearisation method refers to the calculation of a linear approximation of a given function.

First, the EVS specifies if the person is married or single but also divorced, separated or widowed, which are not considered in the initial conditions of the model. Besides, this parameter cannot be compared with other personal characteristics such as ideology, as it is essentially a link with another individual. Therefore, it depends on the social network structure: every married agent should have an assigned spouse within the agent-based model limits. Moreover, couples are a result of the friendship dynamics: the individuals always choose a similar person as a spouse. As the EVS does not say to whom each agent marries, each one should be matched with another married one, and thus always resulting in forced non-similar couples. Even if that is ignored, two individuals who are married should be close enough to communicate, which implies breaking the random distribution. Because all of these reasons, this parameter was not loaded, but transformed into an equation which is taken into account in the agent micro-decisions. If the individual should be married, it will encourage the behaviour of searching for a spouse.

$$Prob_{single}{}^{3} = -0.0014 * age^3 + 0.2921 * age^2 - 17.965 * age + 343.85 \tag{6.1}$$

- *Eq. of probability of having offspring*: depending on the agent's age, it will have more or less probability of having offspring. It is very unlikely that a 19 years old person would have (or even want) children, at least in the Spanish society. Thus, this equation gives the probability of a certain agent to have children, depending on its age. It is empirically grounded, based on the data of the EVS-1981.

$$Prob_{children} = -0.0018 * age^3 + 0.319 * age^2 - 18.3 * age + 349.9 \tag{6.2}$$

- *Eq. Life Expectancy*: this is a set of two equations, one for each gender, used to represent the evolution of life expectancy in the studied period (demographic data from INE [152]). The average age of death has evolved from 72.1 to 74.4 (male) and from 78.2 to 81.6 (female), and these equations will give such age given the current year during the simulation.

$$MaxAge_{male} = 0.1183 * year_{current} - 162.12 \tag{6.3}$$

$$MaxAge_{female} = 0.177 * year_{current} - 272.21 \tag{6.4}$$

[3] As it was explained previously, the constant values and form of each equation was directly derived from the data sources cited, with the aid of the domain expert, and followed by a linearisation of the function obtained.

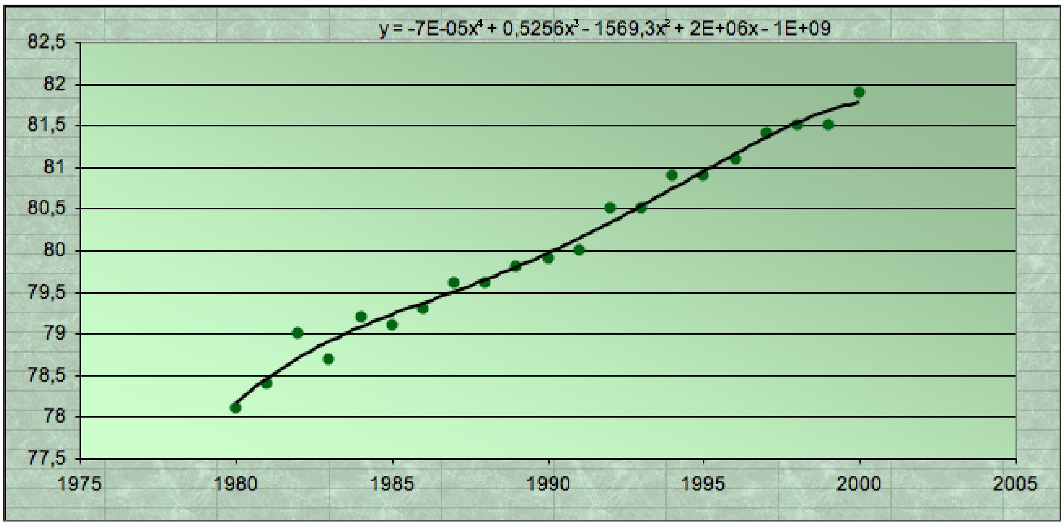

FIGURE 6.6: Representation of the female life expectancy, according to the disperse empirical data from the INE [152] (before the linearisation of Figure 6.7).

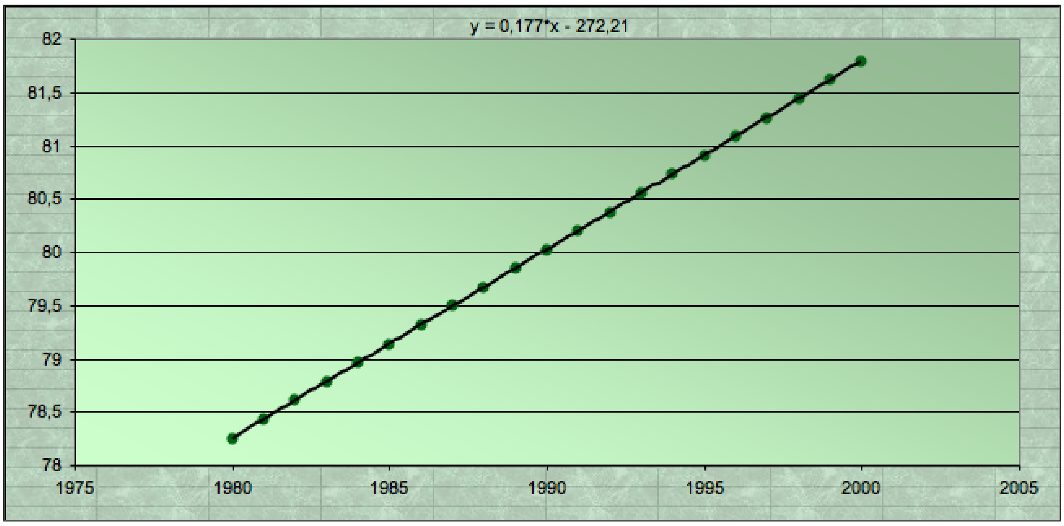

FIGURE 6.7: Linear approximation to the female life expectancy, with a mean error of 2.8% compared to Figure 6.6.

- *Eq. for the age of the first child*: this is a piecewise equation which provides the average age of mothers for having the first child, as this age has evolved along time (from 24.9 in 1980, to 29.0 in 1999). These data are extracted from the EVS.

$$MeanAge_{Reprod} = \begin{cases} round(0.1722 * year - 316), & year \leq 1989 \\ round(0.2353 * year - 441.31), & otherwise \end{cases} \tag{6.5}$$

6.4 Concluding Remarks

This chapter has completed the macro description of the Mentat model. Here, the environment where social dynamics take place has been discussed. Besides, the agent architecture drawn in chapter 4 is complemented with a parallel description of the layers that the general model should have. And similar to then, Mentat is used as a case study to show its actual implementation.

Continuing with Mentat's description, the main free parameters of the agent-based model are classified and described, such as the neighbourhood radius or the warming-up length. Moreover, Mentat's framework shows its modularity specifying the list of mechanisms that can be enabled or disabled, depending on the desired exploration to carry out. The description of the parameters and modules reveal a data-driven design which is supported by the strongly data-driven demographic model. Such model is compound by a set of optional empirical equations explained in-depth in the previous section.

Mentat's approach and framework have been deeply explained along the last chapters. However, there are still several tasks to carry out. Mentat's construction followed a 'Deepening KISS' (section 2.4.3) approach, so this evolution should be carefully detailed in order to justify design decisions. Besides, this framework should facilitate the exploration of the model space, but only a few tests were done already. It is the time for the experimentation, which will be exposed in the next chapter.

Chapter 7

Experimentation & Evolution of Mentat

A process cannot be understood by stopping it. Understanding must move with the flow of the process, must join it and flow with it.

The First Law of Mentat
Frank Herbert, *Dune*, 1965

7.1 Introduction

This chapter will describe the whole process of Mentat development, including the justification of every major decision, its incremental evolution from scratch to a complete framework, and the experimentation for validating its general behaviour, together with its theoretical foundations. This task is carried out in an exhaustive way, detailing in each modelling step all the features of the ABM version under study. The progressive modelling is facilitated due to the modular and flexible architecture used in the framework. This in-depth explanation is thought to be useful for other modellers who must face similar challenges.

Throughout the modelling process the data-driven methodology of chapter 2 was followed, solving the different problems by means of injecting additional empirical data into the simulation, when possible. As part of such methodology, the exploratory method 'Deepening KISS' was defined. This method focuses on the gradual increase of complexity in different aspects of the model, exploring different branches of the model space, and comparing the resulting models in order to find the optimal combinations (see section 2.4.3). Such increase of complexity is performed with gradual modifications of, for instance, the procedure for representing the concept C. If C is represented by a constant, it could be changed into a variable. Following this idea, a gradual sequence of complexity increase could be drawn: constant $\rightarrow$ variable $\rightarrow$ random distribution $\rightarrow$ empirical distribution → complex mechanism, etc. This approach was used along the development of the Mentat ABM, as it is exposed in the following section.

Concerning the general implementation decisions, it should be mentioned that Java Development Kit (JDK) 5.0 was used along the development, within the Eclipse IDE and supported by the Repast J 3.1 framework (Figure 7.1). Repast [50, 194, 212] is nowadays considered one of the leading social simulation toolkits [206, 246], together with Mason [174, 175, 181] and Netlogo [189, 244, 245].

FIGURE 7.1: For the implementation of Mentat, Java 5.0, the Eclipse IDE and Repast J 3.1 were used.

Other used languages and formats have been XML for the communication with external applications, CSV as a preferred spreadsheet/database format for importing large amounts of data into the model, and UML for the definition and specification of the different class diagrams. Concerning the external programs, it is relevant to mention the use of SPSS for the management of sociological surveys [192]. Besides, for the exploration of other research lines (described in the next chapter), Herodoto [129] was used as a tool for Natural Language Processing, Weka [143] for Data Mining over large data collections and OntoBridge [70, 210] for the OWL ontology reasoning.

7.2 The Model Evolution

7.2.1 The First Prototype

The Mentat project began with an early first simple prototype where some simple testing were carried out. This KISS model, as defined in section 2.4.1, has been the base for further development. Thus, according to the 'Deepening KISS' approach (section 2.4.3), this model will increase its complexity step by step, exploring the model space. Those steps will be explained in the following sections, together with its associated experimentation.

Only 500 agents were used in the first prototype, in order to deal with a small number of interactions. Besides, each simulation was executed for only 200 steps, with an equivalence of 1 year = 10 steps, as it was already trying to focus on the period 1980-2000. Even though this prototype was notably simple, it can already be framed inside the data-driven approach. Thus, the 500 individuals that were included had attributes imported from a simplified version of the EVS-1981[1] data (section 3.2.1): a random sample of 500 individuals taken from the 2303 of the whole EVS-1981 Spanish sample (an early test with around a 100 agents can be seen in Fig. 7.2).

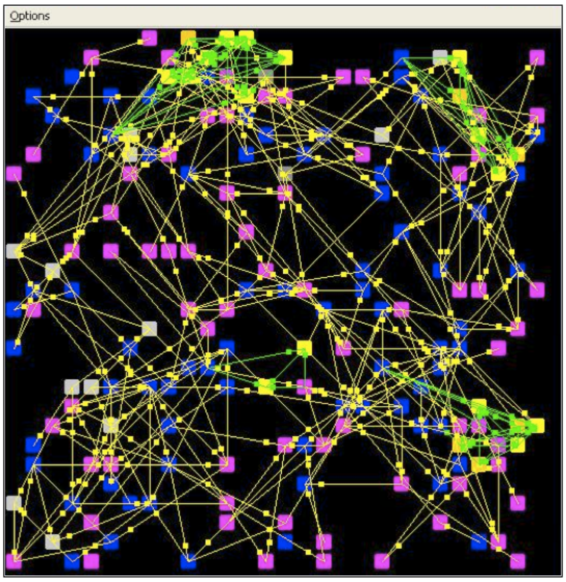

FIGURE 7.2: A view of the preliminary prototype with a few agents. Agent colours depend on their age and gender.

The help of the sociologist domain expert was needed since the beginning, in the preliminary discussions concerning the model. The model conceptualisation, the establishment of limits to the problem and the approach were some of the fundamental issues tackled. Afterwards, during the implementation, his help was essential, for instance, for the selection of a small subset of critical attributes over the hundreds available in the EVS, or the definition of the first simple behaviour of the agents, always with the sociological theory in mind [9].

Only six attributes were chosen, treated in a simplified way: gender, age, educational level, economical level, political ideology and religiosity. The agents behaviour was rather basic, as their demographic rules were static. Following the KISS principle, only Normal distributions were used, without allowing collected data to play a key role in the early design. Another example of its simplicity is that agents had a fixed life-length (for example, males lived till they were 72 years old). The agent relationships were rather basic, with Boolean friendship occurring randomly and matchmaking occurring when possible.

[1]The EVS survey realisation usually takes more than a year to be completed. Thus, it is not accurate to say neither EVS-1980 nor EVS-1981, as it was carried out in both years. Therefore, along this work those two years are used indistinctly, but it should be noted that they refer to the same EVS on 1980/1981. The same can be said about the EVS 1999/2000.

Table 7.1: Summary of the first prototype characteristics.

Parameter	**Value**	**Comments**
Number of agents	500	
Time scale	1 year = 10 steps	Limited for testing
Space size	60x60	Limited for testing
Warming-up length	N/A	
Similarity method	Trivial	
Demography mechanism	Trivial	
Social dynamics mechanism	Trivial	
Supporting theory	Values (partial)	Partially using social values research [9]
Supporting data	EVS (partial)	Small random sample over EVS
Validation	Unsuccessful	

It is not surprising that this model, although able to extract some graphs and statistics, did not provide an output close to real data. For example, the demographic dynamics were inconsistent, with periods without growth (as there were no agents reproducing) and periods of rapid increase. Table 7.1 summarises the main characteristics of this model.

7.2.2 Mentat 0.1 - First Results

After that first basic attempt, and following the defined data-driven methodology (and thus its 'Deepening KISS'), the system grows in two ways: structural modularity and behaviour complexity. First, the ABM is modified in order to build not only a model but also an agent framework. Such framework must be easy to extend, and able to isolate modules, enabling, disabling or even substituting them, in order to consider the impact of different aspects. Only after that, the exploration of the model space is significantly facilitated, and the model complexity increased.

Thus, relationships began taking into account the similarity among agents. Such similarity was built with a gratification method based on the main agent attributes, as explained in section 5.5.2.1. Friendships emerged based on the statement 'the more similar, the more chances to become friends', linking the probability of friendship to the similarity degree. Matchmaking took also that statement into account in order to choose the spouse (both last mechanisms explained in section 5.2). Besides, family links were created, avoiding the matchmaking among the close family (brotherhood, children), and allowing the emergence of family nuclei. That is, the families tend to cluster as the children are born close to the parents, and thus are related to their environment. Figure 7.3 shows those friendship and family links.

The system still uses the sample of 500 individuals, as its speed is practical for rapid development. The length of the simulation is extended, reaching an equivalence of 1 year

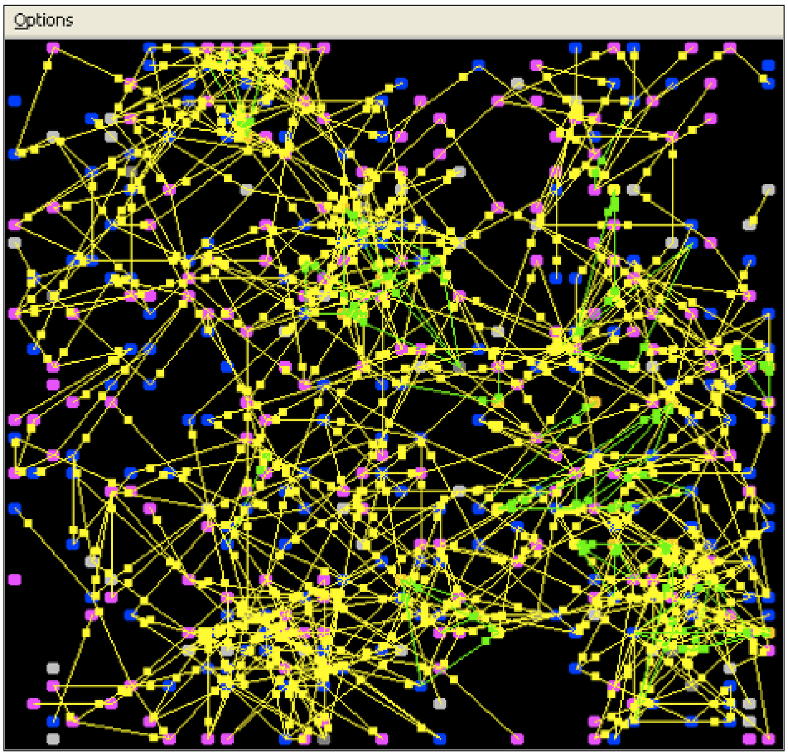

Figure 7.3: Mentat with 500 agents showing friendships (yellow links) and family nuclei (green links).

Table 7.2: Evolution of the religious patterns of the Spanish society, according to the European Values Study (1980, 1990, 1999) [84].

	1980	1990	1999
Ecclesiastical	33	25	22
Low-intensity	22	26	23
Alternatives	14	17	19
Non-religious	31	32	35

= 50 steps. This version is the first one to provide some acceptable results, here just concerning the matching of the religiosity typology evolution (tackled in section 3.2.2). Supported by these first results and according to sociological research (as explained more in-depth in section 5.5.1), testing was carried out in order to find optimal values for several parameters. Thus, a Moore neighbourhood of diameter 6 was established, together with an optimal density for the agent distribution in space of Cells/Agents= 3.2 (see section 5.5.1 for justification of these values). Thus, for 500 agents, a 40x40 grid was chosen. An example of the associated improvement by changing space size *ceteris paribus*[2] of results can be seen in Figures 7.4 and 7.5, in comparison with Table 7.2.

[2]'Ceteris paribus' is a Latin phrase meaning 'all other things being equal'. This is a common assumption in scientific inquiry: a causal connection between two states is studied by modifying the studied

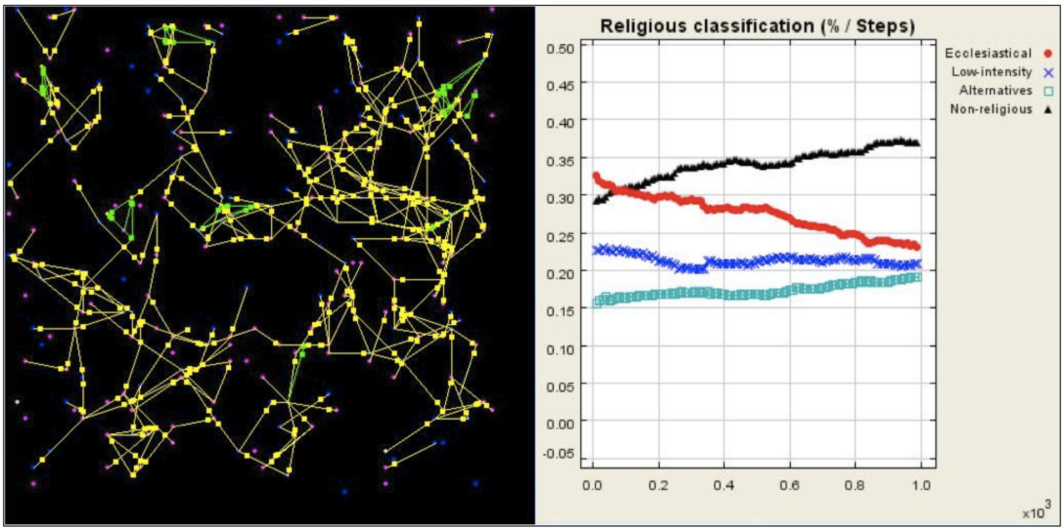

FIGURE 7.4: Density testing: changing space size *ceteris paribus*. 500 agents in a 60x60 space, showing the evolution of the religiosity patterns (1980-1999). Compare with Table 7.2.

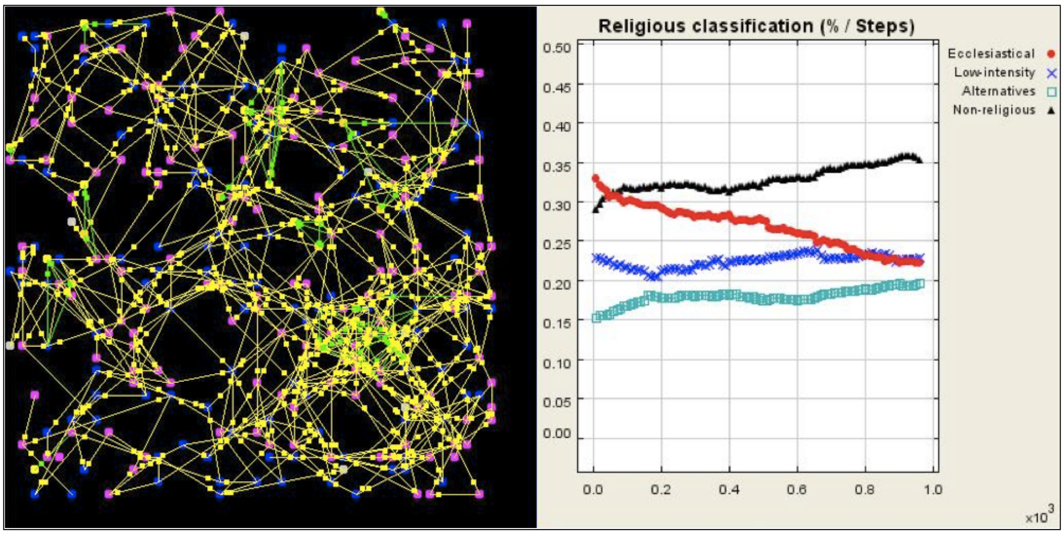

FIGURE 7.5: Density testing: changing space size *ceteris paribus*. 500 agents in a 40x40 space, showing the evolution of the religiosity patterns (1980-1999). Compare with Table 7.2.

This model could not get through the demographic validation successfully, as its population evolution does not match realistic evolution. However, some improvements were carried out, providing the agent with different behaviour depending on the agent life cycle (applying section 4.4.3's automata and constrains), and not just a graphical representation as in the previous prototype (elder are grey, children are yellow or orange, etc.). Besides, demographic rules based on Normal distributions with empirically

variables but holding all the other relevant factors constant, in order to rule out the possibility of other factors which could override the relationship between the antecedent and the consequent [226] grounded parameters [152] were introduced (such as the number of children per couple). Table 7.3 describes its main characteristics.

Table 7.3: Summary of the characteristics of Mentat 0.1.

Parameter	Value	Comments
Number of agents	500	
Time scale	1 year = 50 steps	
Space size	40x40	Fixed density of Cells/Agents= 3.2
Warming-up length	N/A	
Similarity method	Gratification	
Demography mechanism	Basic	Based on Normal distributions with average empirically grounded
Social dynamics mechanism	Functional	Boolean Friendship and Matchmaking based on similarity
Supporting theory	Values, Friendship	Social values research [9]. Friendship sociological research [144, 253]
Supporting data	EVS, Demographic data	Small random sample over EVS. Demographic stats [152]
Validation	Religious patterns evolution	Demographic validation unsuccessful

7.2.3 Mentat 0.3 - Deepening: Need of Children

In order to increase the ABM's accuracy, the next step carried out is the substitution of the random sample of 500 agents. Thus, Mentat is fed instead with the full sample of 2303 individuals from the Spanish section of the EVS-1981, so that the model must handle this population of agents.

Thus, the initial conditions of the ABM are given by this sample, whose structure is convenient to analyse carefully. This initial demographic pattern is a probabilistic one, with a representative distribution given by the EVS-1981, where the maximum random error limit is $\approx 2.08\%$, the trust level is 95.5% and $p = q = 50\%$ [84]. This distribution represents the Spanish population, with an adequate distribution in many different variables: education, economic level, political ideology, values, etc. However, there is a methodological problem difficult to face: as it is based on a survey, no data from underaged people (where underage is under 18 years old) can be found. Children do not participate in surveys, regarding of the age limit chosen. This problem would be minor if just the short-term was considered, such as some months or a year. However, dealing with a 20 years evolution, the problem is deep and important: in 1999, each agent that in 1981 was a child is now an adult (even the ones of 0 years, that in the end are 18) capable of reproducing and altering the global patterns. Furthermore, those 'children' (some of them initially are 17 years old, and they could even have new children in the 80's) represented the 23.33% of the total Spanish population at that time... a too large amount for being ignored, especially considering that lots of them will reproduce during those 20 years. As a side effect of these 'missing children' (besides, those children would have expected to grow up and have families, that now cannot appear in the simulation) the population of the simulation drops more than a 20% during those 20 years (a whole old generation dies, but only a few children are born). In order to approach a realistic population evolution, this demographic issue should be fixed somehow.

Therefore, the immediate task is to introduce that large amount of individuals, which does not have associated data available. The characteristics of the underaged born after 1981 are based on the EVS-1981. Although the initial idea was to use the information of the EVS-1990 to generate them, it was discarded because of several reasons. A big part of them (the closest to 18 yeas) were much more similar to the 1981 group than to the 1990 generation. Besides, the 1990 data would give individuals already changed by the influence of other circumstances, and the aim is to study the influence of demography in these variables. This way the other two EVS will still be used just for validation, and the work will continue being general enough (considering that, in many cases when trying to make predictions, there are not available 'future data' compared with the initial time).

As the population pyramids for these periods are available (in the official statistics [152]), it is easy to know how many individuals of each range of ages are needed: in total, 716 new agents must be included, taking into account the amount for each cohort[3] of individuals. Their characteristics[4] have been assigned from other existing agents of the EVS-1981, chosen randomly from the ones under 30 years. This procedure is justified in that the children should be similar to the youngest ones available. Note that in the real society this new generation keeps values slightly more modern than the simulated ones. However, with the options available, it turns to be an acceptable approximation for solving the problem without an exponential growth of complexity.

But solving this demographic issue, including the underaged individuals with a group of characteristics equivalent to an adult, raises two new inconveniences. First, it disrupts reality as long as children are not mature enough for having stable values. Second, it makes it difficult to compare with the available empirical data (basically, surveys and studies made after 1981), always referred to people over 18 years old. Both problems can be solved by filtering the output of the simulation and not considering the underaged in each time step (as if a survey was done in the actual population of the simulation). Note that, as the simulation is dynamic and agents grow older, more people will be included

[3]That is, the new 716 individuals must be divided into the corresponding age-group population according to the official statistics [152]. For instance, there are 40 individuals of 17 years old but 35 of 1 year old.

[4]Not all the characteristics have been copied to the new individuals: the adult age and the civil state do not make sense in an underage, so they were reset.

in the filter. Anyway, some demographic statistics will remain unfiltered, such as the total number of agents or the percentage of children and adults. A summary of the main features of this Mentat version are listed in Table 7.4.

TABLE 7.4: Summary of the characteristics of Mentat 0.3.

Parameter	**Value**	**Comments**
Number of agents	2303 + 716 = 3019	EVS full sample + introduced children
Time scale	1 year = 50 steps	
Space size	98x98	Fixed density of Cells/Agents= 3.2
Warming-up length	N/A	
Similarity method	Gratification	
Demography mechanism	Basic	Based on Normal distributions with average empirically grounded
Social dynamics mechanism	Functional	Boolean Friendship and Matchmaking based on similarity
Supporting theory	Values, Friendship	Social values research [9]. Friendship sociological research [144, 253]
Supporting data	EVS, Demographic data	EVS. Demographic stats [152] and population pyramids [152]
Validation	Values evolution (partial), Demographic evolution (partial)	Values validation: fit in religious patterns and approximation of others. Demographic validation partially successful.

7.2.4 Mentat 0.4 - Deepening: Warming-Up

After the 'jump' from the Mentat with 500 individuals to the Mentat with 3019, there are still structural deficiencies concerning the demography management in the system. A particular observation is that in the first years of simulation no agents are born. Obviously, this does not obey the empirical demographic evolution, and the reason for this misbehaviour is that the agents begin isolated from each other (spatially close, but with no links between them). They invest in finding friends and maybe spouse their first years in the simulated world. When the population as a whole has already established a robust linked network, they begin showing the expected macro output.

Thus, in order to deal with this issue the needed initial marriages must be introduced, as it was done with the needed initial children. However, even though the percentage of people that should be married is available (and even the agents that are married or not, as it is given in the EVS), there is not a way to find out with whom: it completely depends on the network distribution, as an agent only has communication with its local environment. This is one of the issues that was discussed in section 2.8, the difficulties of obtaining information concerning networks and micro-processes.

This missing social information cannot be loaded from empirical data. It is possible to force those links by inventing them, but it is preferable to let the agents decide with their own criteria. First, the information available is loaded: which agents are married/single in 1981. Second, the simulation can begin... but freezing the years counter. In this special period, coined 'Warming-up', the agents neither get older, nor have children, nor die. But they do communicate with each other, building new friendship and marriage links (taking into account if they should be married in 1981). After a certain period of steps have passed, the 'Warming-up' finishes and the year counter begins. This way, the real simulation begins with the agents already linked, and from 1981 new agents are born, achieving a more accurate global behaviour in comparison with empirical data.

However, when is the network considered robust enough? In other words, how many steps should the warming-up stage have? Several period lengths have been explored, and the results of section 7.3 will show the improvements that different periods give. Table 7.5 shows a summary of this system version.

TABLE 7.5: Summary of the characteristics of Mentat 0.4.

Parameter	**Value**	**Comments**
Number of agents	2303 + 716 = 3019	EVS full sample + introduced children
Time scale	1 year = 50 steps	
Space size	98x98	Fixed density of Cells/Agents= 3.2
Warming-up length	100→1000	Multiple periods have been explored. The lengths shown in the results are 100, 500 and 1000
Similarity method	Gratification	
Demography mechanism	Basic	Based on Normal distributions with average empirically grounded
Social dynamics mechanism	Functional	Boolean Friendship and Matchmaking based on similarity
Supporting theory	Values, Friendship	Social values research [9]. Friendship sociological research [144, 253]
Supporting data	EVS, Demographic data	EVS. Demographic stats [152] and population pyramids [152]
Validation	Values evolution (partial), Demographic evolution (partial)	Values validation: fit in religious patterns and approximation of others. Demographic validation partially successful.

7.2.5 Mentat 0.5 - Deepening: Complex Demographic Equations

Last changes arise new dilemmas. In the previous Mentat of 500 agents (version 0.1), the agents only searched for a spouse whenever they wanted to reproduce, so the number of total couples was the same as the number of couples with children. But after the last modifications which take into account the agents already married before 1981, this does not make sense. The most of those old couples probably had already all the children they wanted to have, and should not have more in the 80's and 90's. There is a new need of finding out which agents desire to have children during the simulation.

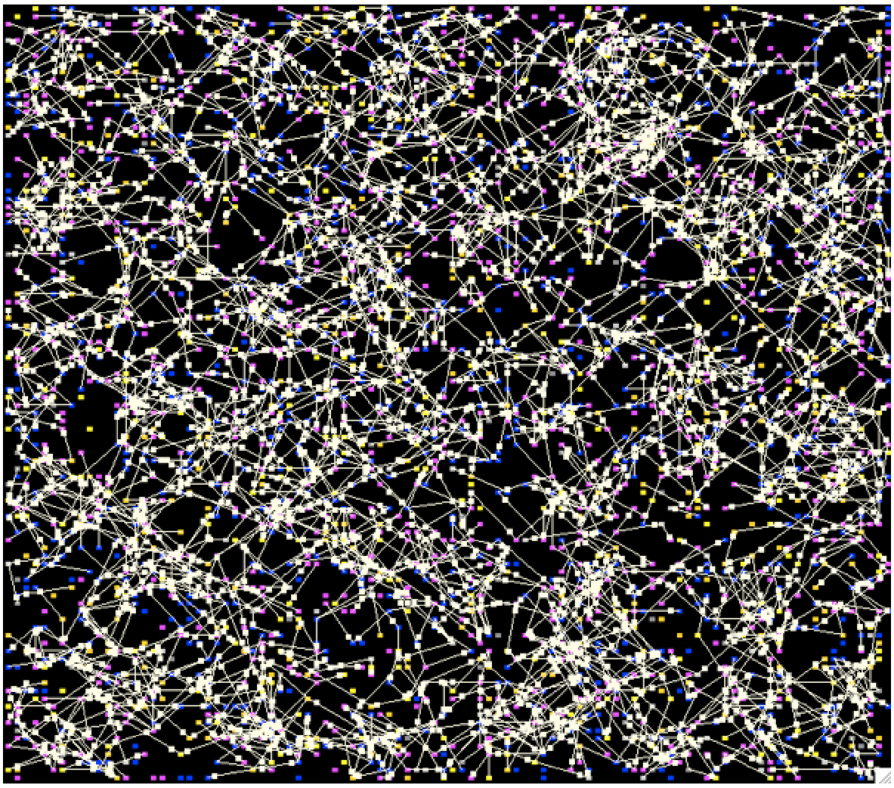

Figure 7.6: Mentat 0.5 space view, with three thousand agents and Boolean friendship.

This fact is determined by multiple factors (for instance, economic class or religiosity) but the one with the biggest weight in the election/possibility of having children, is the age. An equation can be calculated with the available data and with the statistics technique of non-linear regression. Such equation will allow to determine the probability of having children of a given individual of a certain age. For example, a 35 years old agent has only a 23% of probability.

The same mechanism can be used for other choices, following the 'deepening' approach. Instead of loading the information of being single from the EVS (as mentioned in the last section), it can be calculated through another regression equation, so it is generalised and usable for every agent (and not only for the initial loaded ones).

This effort of achieving demographic convergence with the target by adding new complexity layers can be expanded. Thus, it is decided to substitute the rather simple method of deciding the number of children (a Normal distribution with mean on the Spanish mean in 1981) with the regression equation of birth rate taking into account the actual year. The change on the ABM behaviour is immediate, as the birth rate has decreased rapidly in those years: from a 2.2 in 1981, falling to 1.19 in 1999. In fact, from an infra-populated model it twists to an over-populated one. This behaviour can be understood easily, as a data-driven birth rate has been introduced without compensating it with a proper data-driven life expectancy. Thus, the previous Normal distribution is substituted by two equations for indicating the completely different results for men and women. In this case, the empirical data show a clear increase in those years, due to the improvements in health and quality of living. Another significant change in this period is the age of women when having the first child, which has been increased considerably. Therefore, the last normal distribution is substituted for another appropriate regression equation based on the data available.

After these modifications, the agent life cycle has changed and grown in complexity, but it is still easily understandable. It is richer in information and empirical data, as the equations have been generalised from such data [45, 84, 152]. Moreover, the results have been improved significantly, as it is shown in section 7.3. For the construction of every equation, it was followed the advise of a domain expert (a sociologist in this case) who chose the optimal data sources for each mechanism needed, and helped to build a proper equation for it. More information concerning the equations, including their actual formula, can be found in section 6.3.3.

However, this solution has its own limitations. The first two equations, based on age, try to solve the 'time issue', as an agent has different behaviours depending on its age. But the equations themselves evolve in time too. Maybe that 35 years old agent has a 23% of probability in 1981, but a 28% in 1995. Anyway, the calculations would turn to be thorny and unnecessary complicated for the small real difference that it would entail. And the other equations, that take into account the year instead of the age, do not imply this difficulty.

A snapshot of the agent space of this development state of Mentat can be found in 7.6. The 3019 agents are represented using a colour code, depending on their state (child, adult, elder) and gender (male, female). Besides, there are two types of relationships: friendships (yellow links) and families (green links). For instance, a female child (orange square) may have several friends (with yellow links) and two parents (with green links). On the other hand, a summary of the main features of this version is presented in Table 7.6.

7.2.6 Mentat 0.9 - Deepening: Complex Relationships

As it was explained in-depth in chapter 5, friendship and social similarity are blurry and continuous social processes which should not be handled in a Boolean or constant way. Therefore, along that chapter a fuzzification process is carried out in order to improve the social dynamics taking place in the agent-based model. The result includes dynamic evolution of a continuous friendship, an improved similarity measure and matchmaking more in tune with theoretical assessments, everything supported by the sociological friendship theory. These improvements have a subtle influence in the overall quantitative macro results, but do have an importance in the micro qualitative perspective, as it was

TABLE 7.6: Summary of the characteristics of Mentat 0.5.

Parameter	**Value**	**Comments**
Number of agents	2303 + 716 = 3019	EVS full sample + introduced children
Time scale	1 year = 50 steps	
Space size	98x98	Fixed density of Cells/Agents= 3.2
Warming-up length	100→1000	Multiple periods have been explored. The lengths shown in the results are 100, 500 and 1000
Similarity method	Gratification	
Demography mechanism	Equation-based	Dynamic equations empirically grounded
Social dynamics mechanism	Functional	Boolean Friendship and Matchmaking based on similarity
Supporting theory	Values, Friendship	Social values research [9]. Friendship sociological research [144, 253]
Supporting data	EVS, Demographic data	EVS. Demographic stats and equations [45, 84, 152] and population pyramids [152]
Validation	Values evolution, Demographic evolution	Values validation: good approximation of multiple variables. Demographic validation successful.

exposed in the chapter. The final model after these new modifications is synthesised in Table 7.7.

7.2.7 Mentat 1.0: The Complete ABM

The final version of Mentat is achieved after subtle improvements of the last version. As the validation in different aspects already provides a good approximation to the empirical research, new secondary characteristics were added to the ABM. Thus, new attributes were loaded from the EVS in order to map their evolution, such as the tolerance levels. Besides, the module for extracting statistics was expanded, and multiple tests were carried out. The model shows consistent results, showing structural similarity among executions and being robust against noise. Moreover, the graphical interface was slightly improved. All these minor changes did not cause any modification in the structure of the model nor in its results, as shown in its Table 7.8. This final version of Mentat was further explored inserting new modules with other artificial intelligence techniques, as it is shown in the future work of chapter 8, but it is not considered here.

7.3 Analysis of the Results

In section 7.3, several collections of results from the different versions of Mentat are shown and analysed. Next section 7.3.4 will carry out a throughout synthesis of the conclusions that can be extracted from these results.

TABLE 7.7: Summary of the characteristics of Mentat 0.9.

Parameter	**Value**	**Comments**
Number of agents	2303 + 716 = 3019	EVS full sample + introduced children
Time scale	1 year = 50 steps	
Space size	98x98	Fixed density of Cells/Agents= 3.2
Warming-up length	100→1000	Multiple periods have been explored. The lengths shown in the results are 100, 500 and 1000
Similarity method	Fuzzy	
Demography mechanism	Equation-based	Dynamic equations empirically grounded
Social dynamics mechanism	Fuzzy & Dynamic	Fuzzy Friendship which evolves over time. Matchmaking based on similarity and friendship degree.
Supporting theory	Values, Friendship	Social values research [9]. Friendship sociological research [144, 253]
Supporting data	EVS, Demographic data	EVS. Demographic stats and equations [45, 84, 152] and population pyramids [152]
Validation	Values evolution, Demographic evolution, Micro social processes	Values validation: good approximation of multiple variables. Demographic validation successful. Theoretical validation of micro social dynamics

TABLE 7.8: Summary of the characteristics of Mentat 1.0.

Parameter	**Value**	**Comments**
Number of agents	2303 + 716 = 3019	EVS full sample + introduced children
Time scale	1 year = 50 steps	
Space size	98x98	Fixed density of Cells/Agents= 3.2
Warming-up length	100→1000	Multiple periods have been explored. The lengths shown in the results are 100, 500 and 1000
Similarity method	Fuzzy	
Demography mechanism	Equation-based	Dynamic equations empirically grounded
Social dynamics mechanism	Fuzzy & Dynamic	Fuzzy Friendship which evolves over time. Matchmaking based on similarity and friendship degree.
Supporting theory	Values, Friendship	Social values research [9]. Friendship sociological research [144, 253]
Supporting data	EVS, Demographic data	EVS. Demographic stats and equations [45, 84, 152] and population pyramids [152]
Validation	Values evolution, Demographic evolution, Micro social processes, Robustness	Values validation: good approximation of multiple variables. Demographic validation successful. Theoretical validation of micro social dynamics

7.3.1 Random Initialisation vs. Empirical Initialisation

First of all and before any other results, a test for justifying the data-driven essence of Mentat is presented. As it was explained in chapter 2, the data-driven approach has different aims than the theoretical approach. As it is focused on expressiveness, it frequently provides a more accurate output for the given case study. For instance, if an ABM is initialised with data, its output should match an empirical validation better than an ABM initialised with a random distribution. This initial test was executed in Mentat, taking benefit from its ability to enable and disable its different modules.

Thus, Mentat 0.5 was executed with random initialisation (Mentat-RND) and with the EVS empirical initialisation. But in order not to let other variables to influence (*ceteris paribus*), only the Age was taken into account and thus imported from the EVS. Therefore, both versions have exactly the same structure except for the source of the ages of the initial agent population. In Mentat-RND this attribute has been assigned using an uniform random distribution in the range [0, 75].

The indicators for the comparison of both versions have been chosen as they are directly affected by the demographic model and the population pyramid (not more than the age distribution). The indicators are: the percentage of old people, the ratio of single to married agents, and the overall population growth rate (determined by the change in the number of couples and their age).

As Mentat-RND's output is unstable, with noticeable changes between executions, an aggregation measure was needed. Its results are averaged over 15 executions to allow stochastic variations in its output. Mentat 0.5 is almost stable between executions because of its fixed initialisation and so the means shown are based on just 10 runs. Their results are compared with the empirical data from EVS in Table 7.9.

Table 7.9: Comparison between EVS, the random initialised version and the data-driven version of Mentat 0.5.

	EVS/Census*			Mentat-RND			Mentat 0.5		
	1980	1990	1999	1980	1990	1999	1980	1990	1999
% 65+ years	16*	18*	21*	19	24	29	15	19	24
% Single	28	29	29	-	45	37	-	42	35
% Pop. Growth	-	-	+8%*	-	-	+10.1%	-	-	+7.2%

Source: European Values Study (1980, 1990, 1999) [84]
* Source: Spanish Population Census (1981, 1991 and 2001) [234].

Considering the proportion of older people first, the Spanish Census [234] shows that it has been growing, starting at 18% in 1980 and reaching 21% by 1999. Mentat-RND begins with almost the correct figure in the (simulated) year 1980, but the rate of growth is much faster than it should be. On the other hand, Mentat 0.5 shows a better fit to the empirical data, as shown in Table 7.9.

The observed proportion of single people is steady over time. The number of couples in the ABM is directly proportional to the number of friendship links, so the ratio of single to married agents is an appropriate measure of the network cohesion. In Mentat 0.5, the attributes of the individual agents are initialised from the 1980 EVS data, but not the couples, as there is no information about links between members of the sample in the EVS. Therefore, the simulation must start by creating such links to build the network structure. Only after some execution steps, the proportion of couples does converge to a steady state. It can be observed that Mentat 0.5 is again closer to the survey data than the randomly initialised version. Continuing both simulations beyond 20 shows a convergence to a proportion of single people in the range [28,30], but this is reached more slowly by Mentat-RND.

For the case of population growth, the randomised version generates a rate of 10.1%, higher than the Census (8%), while the data-driven version has a growth rate slightly lower (7.2%) than the Census. Overall, the data-driven Mentat 0.5 provides a closer fit to the empirical data than the randomly initialised Mentat-RND for all three of these parameters.

7.3.2 Validation: Discussing the Structural Modifications

In this subsection, a validation of the structural modifications carried out in the model from the version 0.1 towards the 0.5 is considered. This chain of modifications always followed the 'Deepening KISS' philosophy (section 2.4.3), increasing gradually the complexity of the model. As the three changes involved in 0.3, 0.4 and 0.5 are mainly demographical, and most of them are closely interrelated, they are all analysed at the same time. The version Mentat 0.5 is compared with Mentat 0.1 in order to consider the real improvement of the steps taken. However, it is not appropriate to compare an ABM of 500 agents using a partial sample with another using the full sample of 2303 agents provided by the EVS (plus the agents generated because of the underage issue). Thus, all the comparisons were done with a slightly improved version of Mentat 0.1, using the full sample of 2303 agents instead of the 500. Thus, the contrast is fair, with the demographic changes as only cause of the differences.

Table 7.10 compares the EVS and the two Mentat versions, along the different evolutions of a sample of chosen variables, in the period studied (1981 $\rightarrow$ 1990 $\rightarrow$ 1999). The measurements have been made by extracting the information a) from the three different EVS of those years, b) from statistical calculations (over several executions) in the previous version of the system, Mentat 0.1, in the specified years, and c) equivalent calculations made in Mentat 0.5. All variables are calculated considering only the individuals over 18 years. A wide range of different statistical measures were extracted to test the consequences of every change. Even so, mainly means and percentages are

TABLE 7.10: Validation: comparison between EVS, a slightly improved Mentat 0.1 (as it has 2303 agents instead of 500) and Mentat 0.5 (with its 3019 agents).

	EVS			Mentat 0.1 (2303)			Mentat 0.5 (3019)		
	1981	1990	1999	1981	1990	1999	1981	1990	1999
% Male	49	47	49	49	48	48	49	48	49
% Female	51	53	51	51	52	52	51	52	51
Age (mean)	45	43	46	45	51	57	45	47	49
% 65+ years	16*	18*	21*	15	23	31	15	19	24
% Single	28	29	29	100	82	79	100①	42②	35③
% Single ②							100①	34②	30②
% Single ③							100①	29③	28③
Studies (ending age)	16	N/A	17	16	16	16	16	18	18
Economy	0	N/A	N/A	0.028	0.062	0.095	0	0.09	0.105
Ideology									
% Left	29	33	31	29	30	32	29	33	36
% Centre	18	19	23	18	17	18	18	18	17
% N/A	30	25	24	30	30	29	30	29	27
% Right	22	23	21	23	22	22	23	22	20
Ideology (mean)	4.85	4.65	4.75	4.85	4.82	4.73	4.85	4.74	4.59
Tolerances (scale 1 to 10)									
Abortion	2.89	4.13	4.34	2.89	2.96	3.06	2.89	3.08	3.3
Divorce	4.79	5.51	6.10	4.79	4.92	5.09	4.79	5.13	5.4
Euthanasia	3.18	3.97	4.73	3.18	3.24	3.34	3.18	3.43	3.6
Suicide	2.26	2.25	2.77	2.26	2.29	2.35	2.26	2.36	2.5
Religious Typology	1981	1990	1999	1981	1990	1999	1981	1990	1999
Ecclesiastical	33	25	22	33	31	29	33	29	25
Low-Intensity	22	26	23	22	23	22	22	23	22
Alternatives	14	17	19	14	14	15	14	16	16
Non-religious	31	32	35	31	31	33	31	34	37
Pop. growth			8%*			1%			7.2%①
Pop. growth ②									8.6%②
Pop. growth ③									10%③

*: The source is the Spanish population census of the years 1981, 1991 and 2001 [234]. EVS does not show accurately these data.

①, ②, ③: With a 'Warming-up' of 100 steps, 500 steps or 1000 steps, respectively. When not specified, 100 steps are used.

shown here, as they can easily reflect the whole population behaviour. Note that the structural similarity of both versions of Mentat has simplified the analysis, and it can be assessed that these values have a minimum error between executions.

The analysis of the results can begin with the most simple measure: the gender. Its evolution is steady, but it has a small change in 1990. This change is not shown at all in the old system, but can be appreciated quite well in the newer one. On the other hand, the age mean increases just slightly (+2%), but in both Mentat versions an incorrect significant rise is observed. Anyway, it is clear that the old version has a much worse output (+22%) than the newer one (+8%).

The analysis of the percentage of old people has more importance than the previous variables. The evolution of this indicator reflects an important part of the population pyramid structure. Moreover, it is more sensible than the others, and it suffers bigger changes. The EVS data are not accurate here, because they take into account other factors, but in the table the data from the census [234] has been included as empirical source. The increase, as previously observed, is minor in the old version, but moderately good in the newer one.

The percentage of single individuals in the population is a factor related to the agents social network. An agent can be single if a) it does not have single adult opposite-gender friends to have as a partner, or b) it does not seek for a partner (for example, because it is a child). In the beginning, the agents find the problem (a), but with time the social network should grow in complexity and cohesion, so the predominant problem is (b). It can be seen that the percentage of singles should remain quite stable near 30. However, in both Mentat versions a surprising '100' appears in the beginning. This can seem weird, as in all the other measurements the 1981 value matches the EVS. The logic beneath this is that the measurement of 1981 has been done in the very beginning of the simulation, even before than the 'Warming-up' (the introduction of friendships and marriages before the years counting). Therefore, in that moment, all the agents are isolated, so all of them find themselves with the problem (a). Obviously, the size of the 'Warming-up' is crucial for this variable: the more time you leave them to interact, the more couples there will be. Leaving apart the Mentat 0.1, with unreliable output in this matter, it is preferable to concentrate in the new one. Three different sizes have been tested in-depth: 100, 500 and 1,000 steps (note that one year, out of the warming-up, is 50 steps), and the different outputs are shown in the table. After the 'Warming-up', the social network acquires consistence and approaches a lot to the ideal value (that is just measuring the connectivity between nodes). Note how a wider size gives more cohesion: with 500, the ideal is achieved after 1500 steps (500 of 'Warming-up' plus around a thousand of the nearly 20 years of simulation), and logically it can be achieved before with 1000, after the same 1500 steps (1000 of 'Warming-up' plus 500 of 10 years). The stability after that objective is reached (in 1999 it is still nearly 30, instead of continuing the falling), matches accurately the observed reality, and can lead to the situation where the problem (b) is the widely dominant. From the Network Theory point of view, this is a property of small-world networks [3].

On the other hand, the size of 'Warming-up' has another logical effect: the more couples there are, the more children will be born. Anyway, as most of the young couples always find someone, it does not have a big influence in the population growth. As it can be checked in the end of the table, the real growth in that period is around +8%. Mentat 0.1 fails again, but the other one, with different sizes, achieves much better results. The side effect of the extra 1000 steps that gave great results with the percentage of singles yields a bigger error here. Note that the warming-up length does not have other important side effects, so it was not shown in the most of the other variables (where the usual length of 100 steps was considered).

The level of studies is slightly increased. But the education is categorised (in five ranges), so it cannot be appreciated properly (only with the 'jump' from 16 to 18). On the other hand, the 'economic status' is not calculated properly in the EVS, but it is known that it should remain constant or with a minimum increase, and this is exactly what happens in the simulation.

The political ideology follows a noticeably similar trend in both the EVS and Mentat 0.5. However, Mentat shows a higher slope, in the means (2.0% vs. 5.3% of decrease) and in the different percentages. The tolerance measures show a slower increase than in reality, and in both measures the differences between the two Mentat versions are not particularly big. This is due to several facts. First, the intra-generational changes were not modelled, so the agents main attributes remain static over time (and these variables are extremely sensible to those influences). Second, the new generation introduced (700 agents) has rather similar characteristics to the ones from 18 to 30 (as it has been explained), but they should be more modern than that. Finally, the simulation is not able to display the slight move to the right that occurred in the Spanish society during the governments of José María Aznar (1996-2004). However, this can be understood if considered that the simulation is drawn from 1981 data, while Spain was still in the period of democratic transition. In the case of the tolerance measures, it is clear that they are very much affected by the intra-generational changes, empowered by the media and especially opinion dynamics (for instance, homosexuality tolerance should not be considered 'inherited' from the parents, while religious beliefs could).

One of the best indicators for the evolution of values here is the religious typology, strongly based on them. This indicator is predicted with significant accuracy, regardless of the different curves that each type follows (rapid fall, hill, smooth rising and smooth growing, respectively), and with better results in Mentat 0.5.

Overall, it can be assessed that the objective of improving the results of Mentat through the deepening process was accomplished. The Mentat 0.1 provides acceptable trends in some important aspects of the simulation, like the evolution of the religious typology or the political ideology. However, as it has been commented, it has important issues dealing with the demography of the agent population. With Mentat 0.5 these problems are addressed, keeping the acceptable results that Mentat 0.1 achieved, and improving a collection of other indicators. This methodology has been followed, isolating every part that should increase its complexity, re-implementing it, analysing the result of the step, and comparing it to the previous situation and the reality. The result is a model that still can be easily handled, especially considering the problem under study. But it has been growing in complexity gradually from the previous version, so it can deal better with several issues with a higher level of expressiveness.

7.3.3 Validation: Discussing the Fuzzification Process

The modifications of the next versions of Mentat after the 0.5 did not modify substantially the results presented in the previous Table 7.10. However, they had influence in other features: they modified the fundaments of the ABM's social dynamics, with a key importance in the micro qualitative perspective, as exposed in chapter 5. Moreover, these changes essentially improve the theoretical structural validation of the model, according to the sociological theory in use.

A synthesis of those changes, even though they were extensively explained in section 5.7, is tackled here.

Mentat 0.5's complexity was gradually increased, exploring the different fuzzification alternatives, and combining them. Even though multiple features were altered by this process, the modifications can be grouped in two groups: '*Fuzz-Sim*', grouping the normalisation of attributes and fuzzification of the similarity operator; and '*Fuzz-Fri*', grouping the fuzzification of friendship, its dynamic evolution with a logistic function based on similarity, and its influence in partner choice.

Table 7.11 shows all the possible combinations of these groups, beginning in the non-fuzzy Mentat 0.5 and ending in the fuzzified Mentat 0.9.

Besides, the tests were carried out taking into account that there are two possible ways of emerging friendship: one promoting random friends (and therefore an agent can be linked to a non-similar neighbour) and the other one promoting similarity-based friends (and therefore an agent will rarely be linked to a non-similar neighbour, as it will give priority to the most similar ones). This is not a trivial decision, because the friendship evolution function already deals with similarity, and if a neighbour is not similar at all, it will never be more than an acquaintance. It is not evident if the closer way to real-world is giving double strength to similarity (in the second option) or letting randomness to decide who will be the friend (and thus maybe ignoring similar people).

TABLE 7.11: Comparison among different Mentat versions, in increasing order of fuzzification.

	Mentat 0.5	$Mentat_{FuzzSim}$	$Mentat_{FuzzFri}$	*Mentat 0.9*
	(0,0)	(1,0)	(0,1)	(1,1)
Config. Random-friendship				
Mean couple $R_{Similarity}$	0.76*	0.77	0.76*	0.77
Mean couple $R_{Friendship}$	(**)	(**)	0.72*	0.80
Mean couple $R_{Compatibility}$	0.76*	0.77	0.54*	0.62
Config. Similar-friendship				
Mean couple $R_{Similarity}$	0.73*	0.77	0.73*	0.78
Mean couple $R_{Friendship}$	(**)	(**)	0.54*	0.76
Mean couple $R_{Compatibility}$	0.73*	0.77	0.39*	0.59

*: Mentat 0.5's gratification similarity has other range, but here they have been normalised in the interval [0,1] in order to be compared.
**: When the friendship is not fuzzified, all the couples are friends (as this is a Boolean property).

It has to be mentioned that none of the two configurations is deterministic and both are based on probabilities.

The three chosen parameters analyse the couples and how they are affected by the changes in the configuration and fuzzification. The $R_{Similarity}$ shows the proximity taking into account all the characteristics of each partner in a couple. The $R_{Friendship}$ focuses on the friendship link between them, which in a way (according to the logistic function) depends on their similarity, but also in the time spent together. The $R_{Compatibility}$ is taken as an average of the other relations. The values have been obtained after averaging the output of several executions of each version. And in every execution, it is the mean of the property in every couple.

As the first two Mentat versions have a Boolean friendship, their compatibility is always the same as the similarity. In the first configuration, when the friendship is rarely involved in the neighbours linking, the similarity rates are notably similar in all the versions. However, it is clear in the second one that the ones with fuzzy similarity slightly increase their success (specifically, a 5.48% of increase). But the bigger changes can be observed in the friendship: $Mentat$ 0.9 exceeds the other versions with a greater $R_{Friendship}$ and $R_{Compatibility}$, specially in the second configuration (where the difference reaches 51.28% in the best case). These results approach the theoretical qualitative assessments made by the sociological friendship theory considered [144, 253] and the domain expert. Intuitively, the couples of $Mentat_{FuzzAll}$ are more similar to each other (following the proximity principle), and take into account the degree of friendship in the process of choosing partner (which implies historical evolution of the relationship).

Finally, the four different orders of fuzzification described above for both RSimilarity and RFriend of couples, were compared by using the statistical test "One-Way Analysis of Variance" (ANOVA), in order to detect evidence of difference among the population means. The Fisher's statistical significance test equals to 7.281, with a P-value P $\leq$ 0.0001. This small P-value provides strong evidence against null hypothesis, namely, that the difference in the means among the four orders of fuzzification are by chance, both for $R_{Similarity}$ and $R_{Friendship}$ of couples. Therefore, the differences among the means analysed can be attributed to model's fuzzification.

7.3.4 Theoretical Validation: Discussing its Sociological Consistency

Model validation is the process of determining that the model behaviour represents the real system to satisfactory levels of confidence and accuracy, which are determined by the intended model application and its application domain. When dealing with complex systems, as it is frequent on ABM, the traditional methods used in model validation are not widely accepted [37]. In such cases, a suitable option for the validation of the conceptual model is to check whether the theoretical foundations and assumptions are reasonable within the context of the objectives of the simulation. This structural validation is sometimes performed on the basis of participatory methods with domain experts and stake-holders [173].

In Mentat, an empirical validation was performed comparing the macro quantitative output of the model to the empirical data available, as shown in the previous section 7.3. However, it is interesting to carry out as well a structural validation based on the model theoretical foundations, according to the domain expert.

The results are satisfactory, even though the model could be further developed. The achievements in the approximation to the religiosity patterns evolution are significant, together with other important population indicators, especially after highlighting that the considered data come from 1981. Note that the construction of the ABM has followed the hypothesis and statements shown in the sociological model of section 3.3, and the results confirm the assumptions made then.

Such evolution just affects to a part of the socio-cultural changes: the ones concerning the inter-generational dynamics. Through the modelling process it has been understood the key importance of the appropriate modelling of the demographic dynamics in order to reach the proposed aims concerning the evolution of social values. Even though demography does not directly affect social changes, their oscillations over time are, to a certain extent, the consequence of demographic evolution. That is, as the elder and more conservative and religious are dying, the society turns towards more modernity (with all its implications) and less religiosity. These results support Inglehart's theory which assesses that social values are rather stable and they stay without deep changes in the course of the whole life.

The results achieved in this work are determined by the actual society under study. In the Spanish society, the variable 'age' has a high discriminating power, as it was explained in chapter 3. However, this factor does not necessary have to keep its weight throughout the following years. In fact, it is feasible to consider that it is already declining with the disappearing of the generation who lived the Spanish Civil War (1936-1939). In those hypothetic circumstances, and in others where the simulated society does not have a high discriminating 'age', a model equivalent to this one would not provide this accuracy in the output. For these reasons, it would be interesting to complement this work with a research on the opinion dynamics and horizontal influence that produce the intra-generational changes, even if those changes have a minor weight in the socio-cultural changes. Such model could be integrated in Mentat case study, built within the same framework but in another separated ABM, or built completely apart from this work.

On the other hand, this model has been useful because of other reasons apart from the *'prediction'* of the results of the EVS, as it is acknowledged in section 3.1.2. As Epstein states [81], agent-based modelling can have many other aims and reasons. Thus, Mentat has been useful for the explanation and better understanding of the social processes under study, including friendship emergence, social values evolution, partner matching and demographic dynamics. It can be used to guide future research on the field, especially other models that seek to study any of the social processes simulated. It suggested dynamic analogies for the friendship evolution, that obey theoretical foundations and could be validated against new empirical work collected ad hoc. It challenged the robustness of Inglehart's theoretical and empirical works, and provided new support to them. The modelling process helped to identify the lack of data sources for certain aspects, as found in this chapter. And, last but not least, Mentat worked as a case study to prove the feasibility of the agent framework proposed in this work.

7.4 Concluding Remarks

Throughout this chapter the evolution of the ABM Mentat has been exposed, step by step. Beginning in its first rough implementation, and according to the Deepening KISS method proposed in chapter 2, successive complexity layers have been added. Thus, after identifying problems in a mechanism, it has been substituted with a more sophisticated one, in an effort of combining simplicity (as new complexity is only added by exploring the state space and in case it improves the general behaviour) and expressiveness (as the focus is on behaviour and validation, not in keeping it simple). The result has been a successful implementation of a case study that exemplifies how the suggested data-driven methodology and the proposed agent framework tackle a complex problem.

The agent framework modularity allows the extension of the model without a complete re-structuring of the implementation. The model could be further extended in several ways to explore other issues related with the case study.

For instance, Mentat could be modified in order to study a different period, such as 2000–2020 instead of 1980–2000. The model should be fed with different data: instead of the EVS-1980, the EVS-2000 should be used for the initialisation of the agents. Besides, the demographic equations should be adapted to that period, using the projections available from the Spanish official statistics [152]. As the population pyramids changed as well, the number of missing initial children generated will be modified, and they should be based on the new data. That is, only the data sources must be changed, while the rest of the ABM remains unaltered. Although it may be interesting, carrying out such a test is not recommended due to the impossibility of empirically validating a future time period. However, another modification that could be tested and that can be validated is the study of social values in the same period but in a different country, such as France. In this case, the same modifications of the model should be performed (that is, the data sources), but it could be validated with the French section of the EVS-1999.

Another interesting issue would be the introduction of some migration phenomena into the model. During the second part of the period under study (1990-2000), the immigrant population in Spain increased following a growing pattern. Although still small by then, in later periods (e.g. 2000-2010) or other countries (e.g. France), their effect into the social values evolution cannot be ignored. If the appropriate data are collected (such as characteristics of the immigrants and growing patterns), it is feasible to introduce them into the model gradually, with minor modifications in the ABM structure.

Chapter 8

Conclusions

"Alright," said Deep Thought. "The Answer to the Great Question..."
"Yes...!"
"Of Life, the Universe and Everything..." said Deep Thought.
"Yes...!"
"Is." said Deep Thought, and paused.
"Yes...!"
"Is."
"Yes...!!!...?"
"Forty-two," said Deep Thought, with infinite majesty and calm.

Douglas Adams, *The Hitchhiker's Guide to the Galaxy*, 1979

8.1 Summing-up

8.1.1 Summary of Achievements

This work has covered multiple dimensions: epistemological, methodological, technical and sociological:

- *Epistemologically*, it has joined the debate on the basis of social simulation from a experimental point of view [65]. Thus, this work does not follow purely theoretical and abstract simulations that cannot be empirically validated (that is, through an experiment). On the other hand, this work neither follows a positivist approach, in which social processes are reducible to quantitative observations and effects, with sociological theory playing a weak role. Then, this work encourages a move towards empirical experimentation guided by theory, dealing with both theory and quantitative/qualitative data.

- To achieve such aim, it proposes a *methodology* following the data-driven approach, which guides the injection of empirical data into the simulation, bringing them closer to reality, while acknowledging the important role of theory in the whole process. Therefore, the approach is complemented with a systematic method for the exploration of the model space in order to achieve comprehensible but descriptive models.

- This methodology is supported *technically* by the specification and implementation of a modular agent framework in order to facilitate the exploration of the model space and the incremental construction of those models, focusing on the data-driven trend.

- Following the exposed perspective, a case study was developed in-depth to test and validate the application of the proposed methodology and framework. The case study addresses the *sociological* issue of social values evolution, together with friendship emergence and demographic dynamics involved. This model supported the importance of the demographic dynamics in the explanation of the social values evolution in the post-modern Spain.

The construction of Mentat, the case study, can be summarised in a series of key milestones. A data-driven methodology is applied intensively through the course of its development. The modelling process has been realised (a) bottom-up and (b) top-down. (a) is represented by the social network arising from the micro behaviours and friendship dynamics. (b) is hold in the elaborated demographic model. Both (a) and (b) have been theoretically supported, and implemented within a modular agent framework, designed in incremental layers. The modules of Mentat can be enabled or disabled in order to explore the model space following the 'Deepening KISS' of chapter 2. The model is validated from a quantitative macro perspective (empirical validation), from a qualitative micro perspective (social dynamics matching the theoretical assumptions) and from a macro theoretical perspective (discussing its sociological consistency).

Different techniques of Artificial Intelligence are applied and combined in the model, testing the adaptability of the framework and their use for social simulation. Together with the software agents, Fuzzy Logic has been intensively used as shown in Chapter 5. Other technologies such as Natural Language Processing or Data Mining have been further explored, as detailed in Chapter 8. Furthermore, several guidelines have been provided on how to apply others (ontologies, genetic algorithms) that may end up being useful in the modelling process, as exposed in Chapter 2.

Mentat serves as a case study of the methodology and framework, but it provides as well some sociological insight of the studied problem by giving new support to specific theories. Specifically, the ABM stresses the key significance of demographic dynamics in the studied problem: the evolution of social values in Spain during the last decades of 20th Century. This implies that intergenerational changes are considerably more important than intragenerational ones in this Spanish context, and supports Inglehart's theories of values evolution [147].

8.1.2 Contributions to the Field

This work has contributed to the social simulation field in several ways, which are listed here. The list of related publications, grouped by relevance, are exhaustively detailed in section 8.1.3. In this section, the contributions are listed organised by theme, where the published articles are referenced.

- There has not yet been developed a methodology focused on the data-driven trend, and that is the main aim of this work. Several modifications of the logic of simulation were tackled, more appropriate to the data-driven models [128, 130]. Besides, an iterative approach for the modelling process, coined 'Deepening KISS' is suggested [118, 119, 122]. Finally, the method is completed with a full specification of the data-driven modelling cycle, taking into account the logic of simulation proposed, the deepening approach and the possible artificial intelligence techniques that can be useful [122, 123, 126].

- In order to provide tools for the application of this methodology, a modular agent framework is specified and implemented. Such framework would provide a flexible environment to develop a family of sociological problems, and it is specially designed to follow the data-driven guidelines [120, 121, 129, 200, 201].

- Following this perspective, a complete case study was developed, providing with this example an insight of the methodological steps to carry out and the application of the framework. The chosen problem is the sociological issue of the evolution of social values during the 20 years of the beginning of the Spanish post-modernisation process. A theoretical frame was constructed in order to support every modelling decision in sociological theories and hypotheses [8]. Mentat, the resulting ABM, is empirically grounded and theoretically supported, and provides a deep analysis on the social processes tackled [7, 15, 131].

- Testing the robustness of the framework, multiple Artificial Intelligence (AI) techniques were integrated in different modules, combining them with the software agents of Mentat, with different results. The most significant results were found with the insertion of Fuzzy Logic into the model, and thus the ABM was completely integrated with fuzzy agents and relationships [124, 125, 131, 132]. Other AI technologies gave interesting results for social simulation, such as the introduction of natural language processing for the automatic generation of a biography-like output for the agents [127, 169–171], as it is described in the future work. Moreover, the potentials of data mining for its application in the field has been further explored, with promising results [126].

8.1.3 Published Works

The contributions of this work have led to a number of articles in both international conferences and journals. In this section, they are classified exhaustively attending to their relevance. For a thematic classification of the publications, see 8.1.2.

Articles published in international journals:

1. Pavón, J., Arroyo, M., Hassan, S., Sansores, C. Agent-based modelling and simulation for the analysis of social patterns. *Pattern Recognition Letters*, 29(8):1039–1048 (2008) [201] (extended version of [200]) explores the possibility of using agent-based meta-modelling languages on the ABM building process, using Mentat's prototype as a case study.

2. Hassan, S., Garmendia, L., Pavón, J. Introducing uncertainty into social simulation: Using fuzzy logic for Agent-Based modelling. *International Journal: Reasoning-based Intelligent Systems* (2009) (To appear) [125] provides arguments for the use of fuzzy logic in agent-based modelling, based on the experience in the fuzzification of Mentat.

3. Arroyo-Menéndez, M., Hassan-Collado, S. Simulación de procesos sociales basada en agentes software. *Empiria - Revista de metodología de ciencias sociales*, (14):139–161 (2007) [15] (extended version of [7]) reviews the advantages of using agent-based modelling from the sociological perspective, using as a case study the problem of social values evolution.

Articles published as book chapters or post-proceedings in the subseries Lecture Notes in Artificial Intelligence, part of the Lecture Notes in Computer Science series from Springer:

4. Hassan, S., Pavón, J., Antunes, L., Gilbert, N. Injecting data into Agent-Based simulation. In K. Takadama, G. Deffuant, C. Cioffi-Revilla (Eds.), *The Second World Congress on Social Simulation (tentative book title)*, Springer Series on Agent Based Social Systems. Springer, Washington, D.C. (2009) (To appear) [128] (extended version of [130]) proposes a modification of the classical logic of simulation in order to adapt it to the data-driven approach, and suggest guidelines to handle empirical data in ABM.

5. Hassan, S., Salgado, M., Pavón, J. Friends forever: Social relationships with a fuzzy Agent-Based model. *Hybrid Artificial Intelligence Systems, Third International*

Workshop, HAIS 2008, Springer-Verlag, (5271):523–532 (2008) [131] tackles Mentat's social dynamics, including the supporting sociological theory on friendship, the dynamic evolution of friendship links and the application of fuzzy elements.

6. Hassan, S., Antunes, L., Arroyo, M. Deepening the demographic mechanisms in a Data-Driven social simulation of moral values evolution. In N. David, J.S. Sichman (Eds.), *Multi-Agent-Based Simulation IX, Revised selected papers*, vol. 5269 of *Lecture Notes in Artificial Intelligence (from the Lecture Notes in Computer Science)*, Springer-Verlag, 167–182 (2008) [118] (extended version of [119]) proposed the 'Deepening KISS' methodology for the exploration in the model space, as a mid-way between KISS and KIDS.

7. Hassan, S., Garmendia, L., Pavón, J. Agent-Based social modeling and simulation with fuzzy sets. *Advances in Soft Computing*, Springer-Verlag, (44):40–47 (2007). From the conference 2nd International Workshop of Hybrid Artificial Intelligence Systems (HAIS07) [124] formally describes the fuzzification process of Mentat.

8. León, C., Hassan, S., Gervás, P., Pavón, J. Mixed narrative and dialog content planning based on BDI agents. *Lecture Notes in Artificial Intelligence, from the Lecture Notes in Computer Science Series*, Springer-Verlag, (4788):150–159 (2007). Selected papers from 12th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2007 [170] (extended version of [171]) modifies an early version of Mentat in order to connect it to a Natural Language Processing external module, in a Fantasy context for the generation of automatic agent biographies.

Articles published in the main international conferences in the field of social simulation, including the World Congress of Social Simulation and the Conference of the European Social Simulation Association. Other important indexed conferences are listed here, such as AAMAS, MABS, CAEPIA or IJCAI.

9. Hassan, S., Gutiérrez, C., Arroyo, J. Re-thinking modelling: a call for the use of data mining in data-driven social simulation. In *Proceedings of the 1st Workshop on Social Simulation at the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*. Pasadena, CA (2009). (To appear) [126] is a call for the use of Data Mining methods in order to improve the data-driven modelling process, providing methodological steps to do so and a case study based on Mentat.

10. Hassan, S., Fuentes-Fernández, R., Galán, J.M., López-Paredes, A., Pavón, J. Reducing the modeling gap: On the use of metamodels in Agent-Based simulation. In *Proceedings of the 6th Conference of the European Social Simulation Association (ESSA09)*. Guildford, UK (2009). (To appear) [123] attempts to facilitate the communication among the several roles that can take place in the modelling process, by means of using meta-models as middle-range language.

11. Hassan, S., Antunes, L., Pavón, J. A Data-Driven simulation of social values evolution. In Decker, Sichman, Sierra, Castelfranchi (Eds.), *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2009)*, 1337–1338. IFAAMAS, Budapest (2009) [120] gives a short description of the Mentat framework characteristics, with special focus on the technical issues.

12. Hassan, S., Antunes, L., Pavón, J. Mentat: A Data-Driven Agent-Based simulation of social values evolution. In *Proceedings of the Multi-Agent-Based Simulation 2009*. Budapest (2009). (To Appear in Springer Lecture Notes in Artificial Intelligence) [121] summarises the work realised on Mentat from a bird's eye view, including a general overview of the main achievements.

13. Hassan, S., Pavón, J., Gilbert, N. Injecting data into simulation: Can Agent-Based modelling learn from microsimulation? In *Proceedings of the World Congress of Social Simulation 2008*. Washington, D.C. (2008) [130] proposes a modification of the classical logic of simulation in order to adapt it to the data-driven approach. It uses as case study two versions of Mentat.

14. Hassan, S., Antunes, L., Pavón, J., Gilbert, N. Stepping on earth: A roadmap for data-driven Agent-Based modelling. In *Proceedings of the 5th Conference of the European Social Simulation Association (ESSA08)*. Brescia, Italy (2008) [122] is a call for the use of data-driven agent-based modelling, providing some guidelines to do so.

15. Hassan, S., Antunes, L., Arroyo, M. Deepening the demographic mechanisms in a Data-Driven social simulation of moral values evolution. In N. David, J.S. Sichman (Eds.), *Proceedings of the MABS 2008*, 189–203. Springer, Estoril, Portugal (2008) [119] proposes the 'Deepening KISS' methodology for the exploration in the model space, as a mid-way between KISS and KIDS.

16. Hassan, S., Pavón, J., Arroyo, M., León, C. Agent based simulation framework for quantitative and qualitative social research: Statistics and natural language generation. In F. Amblard (Ed.), *Proceedings of the Fourth Conference of the European Social Simulation Association (ESSA07)*, 697–707. Toulouse, France (2007) [129] presents Mentat as a modular framework for both quantitative and qualitative social research, using statistics and natural language processing techniques.

17. León, C., Hassan, S., Gervás, P., Pavón, J. Mixed narrative and dialog content planning based on BDI agents. In *Proceedings of the CAEPIA-TTIA'07*, vol. I, 27–36. Salamanca (2007) [171] modifies an early version of Mentat in order to

connect it to a Natural Language Processing external module, in a Fantasy context for the generation of automatic agent biographies.

Articles published in other international congresses:

18. Arroyo, M., Hassan, S. Marco teórico-sociológico y operativización para modelar un sistema multi-agente sobre la evolución de la religiosidad española. In F.J. Miguel (Ed.), *Proceedings of the 2nd Workshop on Social Simulation and Artificial Societies Analysis (SSASA'08)*, vol. 442, 8. CEUR Workshop Proceedings, Barcelona (2009) [8] describes the sociological frame of the Mentat project, identifying the theoretical supports and the justification for several modelling decisions concerning the evolution of social values.

19. Hassan, S., León, C., Gervás, P., Hervás, R. A computer model that generates biography-like narratives. In A. Cardoso, G.A. Wiggins (Eds.), *Proceedings of the 4th International Joint Workshop on Computational Creativity*, 5–12. London (2007) [127] defines a whole framework for the automatic generation of stories, trying to evaluate and measure the creativity of the process.

20. Arroyo, M., Hassan, S. Simulación de procesos sociales basada en agentes software. In *Actas del IX Congreso Español de Sociología*. Barcelona (2007). Grupo de trabajo I: Metodología [7] reviews the advantages of using agent-based modelling from the sociological perspective, using Mentat as a case study.

21. León, C., Hassan, S., Gervás, P. From the event log of a social simulation to narrative discourse: Content planning in story generation. In *Proceedings of the AISB Annual Convention*, 402–409. Newcastle, UK (2007) [169] tackles the automatic generation of a fairy tale based on an agent-based simulation (a modification of Mentat).

22. Hassan-Collado, S., Mata-Garcia, M.G., Salvador, L.G. Aplicaciones de lógica borrosa a sistemas sociales con agentes software. In *Actas del Primer Congreso Internacional de Matemáticas en Ingeniería y Arquitectura*, 256–265. Madrid (2007) [132] describes a first fuzzification of Mentat, including some dead-ends and some successful attempts.

23. Pavón, J., Arroyo, M., Hassan, S., Sansores, C. Simulación de sistemas sociales con agentes software. In Actas del Campus Multidisciplinar en Percepción e Inteligencia, vol. I, 389–400. Albacete (2006) [200] explores the possibility of using agent-based meta-modelling languages on the ABM building process, using the example of religiosity evolution.

8.2 Future Work

8.2.1 Natural Language Processing

There are multiple works that show how social simulation can be a useful tool for quantitative researches [178]. However, qualitative researches have been used in this field mainly in an assistant way, working from the quantitative perspective and letting it guide the analysis and results view, as it can be seen in [241]. There are some interesting works related to how to introduce this qualitative knowledge into an ABM [264]. On the other hand, here it is proposed the use of the results of agent-based social simulation also for qualitative social scientists, even knowing that the intensive use of computers is mostly linked to quantitative researchers. This is performed by recording 'stories' of agents, allowing to get knowledge about their particular evolution and build a narrative personal story, which may provide insight on their motivation along time.

Thus, each individual is taken as a whole, with the holistic perspective of qualitative researchers. From this perspective, the evolution of each individual has a great importance, and instead of considering the global emergent trend, the focus moves towards people lives. Then, in order to explore this research line, a Mentat version was modified to introduce some of this preliminary work.

Thus, it is natural to provide with a name and surname for each agent: now each one represents a person instead of a number or an ID (it is rather different 'i214 died' and 'Pablo Martínez died'). Besides, the agents have received the possibility of 'living' events across their lives, events that could change their future decisions. All these changes converge in the new system output: a narrated life-story in natural language. This is achieved by adding a new module to the system: a natural language processing (NLP) tool. This tool is a simple automatic narrator designed with most usual techniques on this area [211]. Its main purpose is to add more content to the analysis of the agent-based model, and in this way completing the graphical output.

This NLP module has been designed following the usual approach based on rules. The system is heavily oriented towards Content Determination (filtering the facts, providing only those facts that are considered to be significantly interesting for the reader), and Discourse Planning (ordering those filtered facts in such a way that the reader perceives a coherent story). Besides, rather simple Surface Realisation based on templates for creating the final text is addressed, to show a human readable form of the final content. This architecture merged in Mentat can be observed in the diagram of Figure 8.1. The most natural way of selecting one (or a few) individual stories is to find the qualitative 'ideal types'. That is, the life-stories of the most representative individuals, the ones that show what has happened during the simulation. These individuals reflect the 'macro' changes of the complex society. For instance, a hyper-inflation process can be shown through the fall of living quality of a representative person. This person

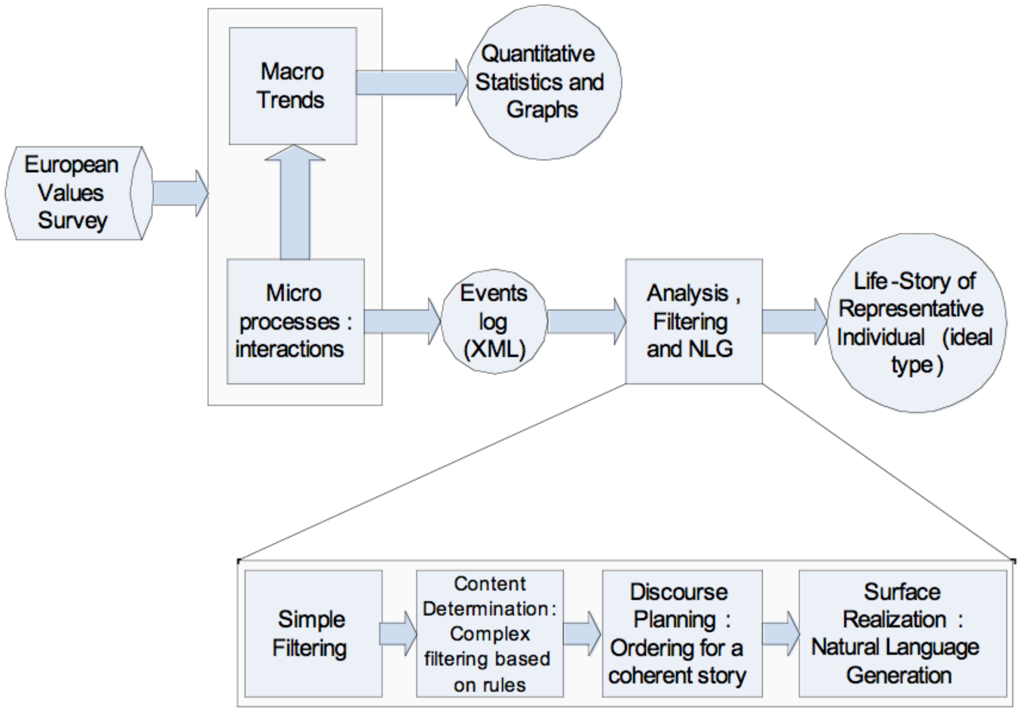

Figure 8.1: Diagram showing the architecture of Mentat double output: stats and biographies.

story would capture how he/she is forced to buy cheaper products and stop having some luxuries.

To achieve Content Determination, the NLP tool takes every individual as a whole being, and analyses the set of agents depending on its configured rules. Those rules, clearly context-dependent, have to be defined in order to choose correctly the representative individual(s). They should measure the interest (weight) of different life-events of the agents and their relationships with others. Thus, a domain expert should collaborate in the construction of those rules in order to be theoretically grounded. Therefore, only the most representative individual will be selected (the closest agent to the qualitative ideal type) from a set of many possible agents (all of the simulation). In this prototype version, the tool selects only one of them, the one who is considered to be the most relevant. More details about how this tool works can be found on [129, 169].

Afterwards, the execution of the simulation generated an XML file, together with the quantitative and visual output. This XML is the result of logging the trace of every event of each agent throughout their simulated lives. Agents that die and are born, find a partner or a friend, together with their life events: everything is recorded. Next, the NLP module will process this XML, as explained in [169] in order to provide the narrative output. Figure 8.2 shows a comparison of the XML and the preliminary generated text by the prototype, although the bigger effort is to select 'Rosa Pérez' among the hundreds of individuals. Biographies of agents that have been selected to build a story are not as

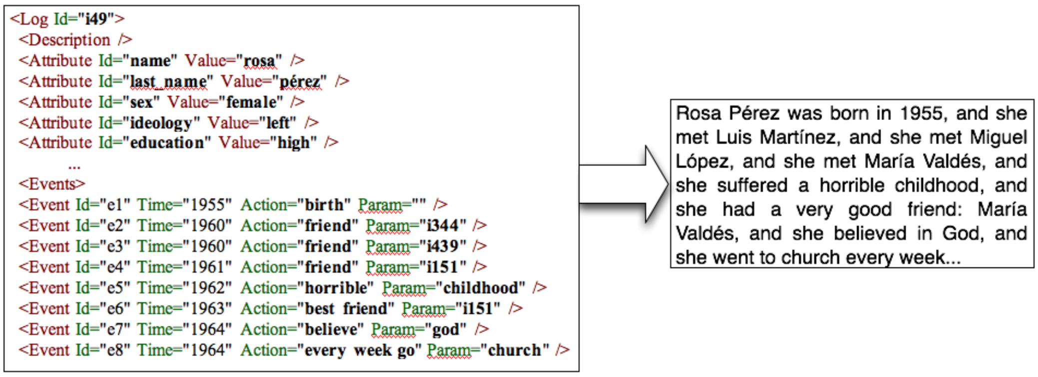

FIGURE 8.2: Sample of an XML log of an agent turned into a piece of life story of a representative individual.

expressive as real textual narrations, but can be directly compared with them (because of its natural language format). Due to the huge amount of textual and narrative material that qualitative research deals with, a narrative story is a natural way for offering an output. This could help to understand the real individuals, compared with other possible life-stories, and it could be rather useful due to the lack of individuals that usually qualitative researches have. Moreover, data mining clustering (as explained later in this chapter) can be used for building the qualitative ideal types of the simulation, and thus the real individuals of qualitative research can be better understood in comparison with the simulated stories of those ideal types.

An alternative application to this new output is to provide complementary information, simple to understand for any human. This is especially relevant for those who are not familiarised with the computational simulation, such as stake-holders. An stake-holder could read the text with an example a character life, and better understand the social process being simulated, as it is performed in [187].

There is significant future work in this research line, as there are plenty of issues to improve before those generated biographies can be considered trustworthy. Both, the rules and the agent events should be theoretically supported. Besides, the ideal types of the ABM should be found using clustering techniques in order to correctly select the appropriate individuals [126]. The text generation should be significantly improved

8.2.2 Other Fuzzy Extensions

The use of fuzzy logic in ABM has been tested successfully during the chapter 5, using it in several aspects of the model. However, those are not the only uses that have been explored. This section proposes further fuzzification that could be carried out in other ABM aspects, using Mentat as a case study.

For instance, in case of modelling opinion dynamics, such process could be easily fuzzified. In order to explore this possibility, an scenario in which Mentat has social influence (horizontal intragenerational dynamics) can be considered.

Each agent has a collection of attributes which can be influenced by other, and thus modified to some extent. Some attributes, such as gender or age cannot be affected by others' influence, but some others, such as ideology, can be influenced by the political trends of friends. This local influence is, by definition, a 'fuzzy concept': the influence on a person cannot be easily quantified. Thus, after a throughout analysis and with the help of the domain expert, the following formal proposal is built.

Let X be a fuzzy set on the universe U, expressing a human characteristic. The fuzzy set variation of the attribute 'X' on U, denoted Δ^X: U $\rightarrow$ [0,1] determining the influence of the characteristic X by the environment of each individual 'Ind' can be defined as the aggregation of the influence of all its relatives, friends and spouse. This influence is determined by the 'proximity' of the person, the distance 'd' between the attribute selected, and how young 'Ind' is (if the agent is younger, it will be more receptive to be influenced):

$$\Delta^X(Ind) = OWA_{i=1\ldots N}(R_{proximity}(Ind, Ind_i) \text{ AND } d(X(Ind) - X(Ind_i)) \text{ AND } \mu_{young}(Ind)). \tag{8.1}$$

Note that the Product t-norm is used to compute the 'AND' conjunction.

Let $\text{R}_{proximity}(\text{Ind}_n,\text{Ind}_i)$: UxU$\rightarrow$ [0,1] be the fuzzy relation on the set of individuals that give a degree of 'proximity'. This fuzzy relation is defined by the aggregation (OWA) of the classical relation 'couple' with the fuzzy relations R_{friend} and R_{family}:

$$\begin{aligned} R_{proximity}(Ind, Ind2) &:= \\ OWA(R_{couple}(Ind, Ind2), R_{friend}(Ind, Ind2), R_{family}(Ind, Ind2)) &= \\ w_1 * R_{couple}(Ind, Ind2) + w_2 * R_{friend}(Ind, Ind2) + w_3 * R_{family}(Ind, Ind2) & \end{aligned} \tag{8.2}$$

Finally, the evolution of an attribute is determined, for each individual, as

$$X(Ind) := OWA(X(Ind), \Delta^X(Ind)) \tag{8.3}$$

Other aspects of Mentat can be fuzzified, as it may happen in plenty of other models of social processes. For instance, another interesting process to be modelled is the inheritance of the parents' characteristics. In Mentat, it is solved by giving the children an averaging of each attribute of the parents. However, fuzzy operators can be defined to deal with this issue. Thus, the fuzzy connectives of composition 'o' can be used for obtaining the attributes of the new individuals (together with a recommended random mutation factor not included in the mathematical definition, which should be introduced in favour of diversity):

$$\forall X \; attribute \; of \; Ind, \; \mu_x(Ind) := \mu_x(Father(Ind)) \; o \; \mu_x(Mother(Ind)) \tag{8.4}$$

Another relevant aspect to consider is the agent life cycle states. The life cycle state of an agent is defined by its age, and determines its behaviour, as specified in section 4.4.3. Therefore, an agent in the state of 'child' cannot find a stable partner, while an 'old' agent will not have children and will have greater probability of dying. But, where are the limits among states? In the crisp systems, there are threshold limits that determine that change. For example, if an agent has an age over 18 it is in the 'adult' state, but if it is slightly under it, it is in 'child' state, with a radically different behaviour. This is very different to the real behaviour: observing the social process, the changes are gradual, as people get older. Therefore, this is another case for applying fuzzy logic, even though here it is rather difficult: it is easy to define how young an individual is, but it is difficult to change gradually its behaviour (anyway, it is an implementation problem, which will not be tackled here).

Another approach not referred here is to take into account the space and time dimensions. Even though space is implicitly covered when an agent is allowed to communicate only with its nearby ones, the fact of an agent is closer than other is ignored: these are geographical issues. This could be seen as another fuzzy relation, where 1 is the closest neighbour, and 0 not known at all. About timing, it must be said that all the definitions here should be taking time into account, because in the simulated system continuous time does not exist: time is discretised in time steps. This way, all the operations require a step of time. For instance, a right way would be:

$$X^{s+1}(Ind) := X^s(Ind) + \Delta^{X,s}(Ind) \tag{8.5}$$

where s is the number of time steps.

8.2.3 Data Mining

Data mining techniques (DM), by nature, depend on the availability of large amounts of data, which is processed and classified and/or clustered, extracting new knowledge such as hidden patterns. The main reason for this delay is the mentioned KISS approach: till the appearance and growing popularity of complex data-driven models, the ABM rarely used large amounts of empirical data.

Even though there are a few works that began to consider DM as useful for ABM, it is a trend far from being spread. In [161] it is proposed the use of a DM technique, association rules, to validate the simulation output and analyse the real-world data.

These rules would discover unexpected relationships among the categorical variables in the simulation and empirical data. The model is validated checking the existence of inconsistencies among the rules generated for each data set. However, no results of the approach are shown.

This path is beginning to be explored, using Mentat as experimental ground. In the work [126], a methodological approach to Data-driven ABM through the use of Data-Mining is presented. The proposed method is framed and deeply described, stage by stage, with a focus on the Analysis process. Besides, Mentat is used as a data-driven case study, explaining how each of the stages was followed in its context.

8.2.4 Mentat Extensions

The framework's modularity and flexibility make it an ideal candidate to explore the application of different techniques and approaches, together with extensions of the model exposed. Some of the possible model extensions, such as the introduction of immigration, were commented in section 7.4 of previous chapter. Besides, in the same section an interesting modification is discussed: the change of temporal or spatial context of Mentat. Note that the only ABM modifications needed for such changes would be the data sources, which specify the time and space characteristics[1].

Another interesting open issue is the modelling of the different space layers that can be defined. As it was commented in section 6.2, there is a dilemma associated to the distribution of the survey individuals in the space: 'the original EVS individuals are distributed all across Spain... but it does not make sense to map the ABM individuals in the Spanish territory. If that was the case, how would they be able to contact each other? The constrain of these agents in the same space is a must'. Thus, the agents must be constrained to an abstract space as the 2D grid defined in that section. But this is a complex and open issue that can have more elaborated solutions. There are multiple spaces where proximity could be defined in parallel, and the same agent could belong to several spaces at once. Thus, there can be a friendship space, family space, work-relations space, but also a division between physical space and online virtual space, for instance.

An issue Mentat is rather poor is at graphical visualisation. This fact is worsen with the high density of friendship links that emerge during the 20 years simulation. This is strongly linked to the framework used, as Repast J does not provide a good

[1]This is correct as far as the change of context is not too extreme. A temporal change of 20 years ahead or before in Spain can be simulated in the same way. But an attempt of modelling the ancient Hispania with the same ABM would not be feasible even if the data sources were accessible. Thus, the same can be said of spacial issues: moving the context to Germany or even Romania is acceptable. But trying to model Nigeria, with a very different social structure (and thus friendship network) or stability (and so many other factors involved), would be out of reach for Mentat.

collection of visualisation libraries. The high cost of using external ones and integrate them successfully pushed this issue to the end of the priority queue. This problem would be easily solved by porting Mentat to the recently appeared 'Repast Simphony', which did not exist when this work was began. Simphony offers an integration with the VisAD scientific visualisation package, and allows both 2D and 3D environments (as shown in Figure 8.3). Another artificial intelligence technique that was explored in Mentat

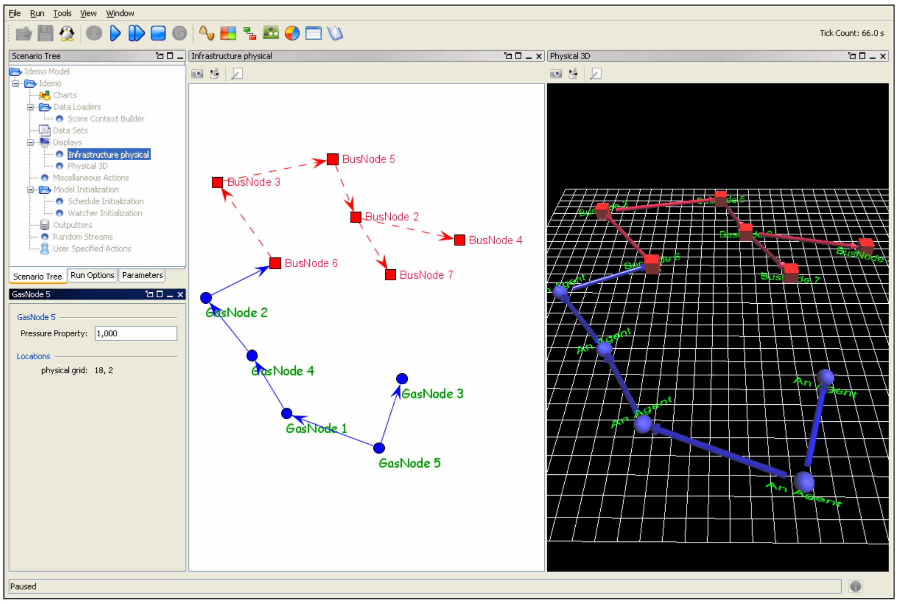

FIGURE 8.3: Screenshot of Repast Simphony. Source: Repast agent simulation toolkit web page: http://repast.sourceforge.net [212].

was the insertion of ontology-based reasoning in the agents. This way, the inference engine that ontologies allow would open more complex deductions when communicating agents. However, the expected efficiency difficulties that were found when trying to allow inference reasoning to 3000 agents discouraged this research line.

Mentat focuses on analysing the weight of demography in the mentalities change, in the case study. However, this aim has another face: which is the weight of the horizontal influence (intragenerational effects)? It would be significantly interesting to apply this approach to the same period and society, but focusing on the opinion dynamics that affect social values. It could be inserted in the existing Mentat model or, in order to isolate it better, build it in another ABM. In any case, its results could be crossed with current Mentat's and offer a significantly wider view of the problem.

Appendix A

Apéndice: Síntesis del trabajo

A.1 Introducción

A.1.1 Contexto

Un sistema social es un grupo de individuos que interactúan entre ellos, evolucionando autónomamente y motivados por sus propias creencias y objetivos personales, junto con las circunstancias de su entorno social. Los sistemas sociales son un ejemplo de sistemas complejos no lineales, ya que su comportamiento no puede ser expresado como la suma de comportamientos de sus partes. Así, normalmente exhiben un fenómeno emergente cuyas propiedades necesita de nuevas categorías a un nivel mayor de abstracción para poder ser estudiadas.

Estos sistemas complejos son difíciles de estudiar debido a que no pueden ser comprendidos de forma analítica: es muy frecuente que no se pueda hallar un conjunto de ecuaciones que permita describir todo el sistema [197]. No obstante, dicho comportamiento no lineal puede ser estudiado a través de la construcción de un modelo y su simulación [107].

La idea detrás de la simulación social basada en agentes es que el investigador pueda comprender la complejidad no tratando de modelar desde el nivel global, sino analizando las propiedades emergentes resultantes de la interacción local entre agentes autónomos que se influyen mutuamente como respuesta a las influencias que reciben [178]. Por ello, la especificación de características y comportamiento de cada agente es crítica en lo que respecta a las dimensiones del problema estudiado.

Este método facilita un enfoque bottom-up para el análisis del comportamiento macro en sociedades de entidades que interactúan [20]. Este comportamiento agregado se llama emergente, ya que el comportamiento colectivo e incluso individual no puede ser predicho de las condiciones iniciales de la simulación [54].

Sin embargo, existen algunas limitaciones al tratar de simular sistemas sociales. El principal problema es que el individuo, representado como agente software, es por sí sólo un sistema complejo, cuyo comportamiento es impredecible y menos determinado que el de un agente, el cual tiene comportamiento y percepciones que pueden ser diseñados con relativa simplicidad. En la práctica no es posible considerar la simulación de innumerables matices que pueden encontrarse en un sistema social real con respecto a la interacción social, caracterización del entorno, etc. Por ello, es inviable pretender la simulación de un sistema social en todas sus dimensiones[107]. Para poder tomar un enfoque adecuado, debe ser construido un modelo como abstracción del fenómeno a estudiar. Por otro lado, el investigador debería limitar su simulación a procesos sociales concretos en un contexto "sistémico" e interactivo (adecuado para la simulación social y su requisito de cientos de entidades). De ese modo, la simulación de sistemas sociales sería considerada en términos de centrarse en un proceso concreto (sin intentos de abarcar en exceso) [77].

A pesar de dichas limitaciones, el paradigma de agentes ofrece numerosas ventajas para expresar la naturaleza y peculiaridades del fenómeno social. No obstante, la mayoría de los modelos de simulación basados en agentes son relativamente simples, similares a autómatas celulares [262]. Esto puede ser válido para estudiar el comportamiento emergente que resulta del comportamiento pseudo-determinista de los agentes. Sin embargo, cuando los individuos son considerados como entidades mentales complejas con capacidades cognitivas como creencias y valores, este enfoque puede resultar demasiado limitado. Debido a ello este trabajo promueve un enfoque fomentando la complejidad frente a la simplicidad, en múltiples dimensiones: técnicamente (con la especificación de un framework modular que integra varias tecnologías de inteligencia artificial), epistemológicamente (moviéndose de una visión abstracta y teórica hacia una dirigida por datos) y metodológicamente (definiendo una metodología para adaptar el enfoque dirigido por datos al modelado basado en agentes).

A.1.2 Motivación

Los métodos de investigación normalmente siguen una de las dos vías tradicionales para hacer ciencia: inducción y deducción. Inducción es el descubrimiento de patrones en los datos empíricos, mientras que deducción implica la especificación de axiomas para probar consecuencias lógicas que son derivadas de ellos. No obstante, de acuerdo con varios autores como Axelrod [20] y Gilbert [101, 102], la simulación puede ser definida como la "tercera vía" de hacer ciencia.

> *"Como la deducción, [la simulación] comienza con un conjunto de supuestos explícitos. Pero al contrario que la deducción, no prueba teoremas. En su lugar, una simulación genera datos que provienen de un conjunto de reglas especificadas rigurosamente en lugar de la observación directa del mundo real. Mientras que la inducción puede ser usada para encontrar patrones en*

> *los datos, y la deducción puede ser usada para encontrar consecuencias de supuestos, el modelado para la simulación puede ser usado como ayuda para la intuición."* [20]

Así, el modelado computacional puede ser definido como la construcción asistida por ordenador de una abstracción de un sistema observado para un propósito concreto [235], y la simulación computacional es entendida como un tipo particular de modelado [107] con el propósito de "dirigir un modelo de un sistema con entradas adecuadas, y observar sus correspondientes salidas" [20, 33].

Hay múltiples tipos de simulación útiles para las ciencias sociales [107]. Los modelos basados en agentes (ABM[1]) presentan diversas características que los convierten en una opción muy atractiva para simular fenómenos sociales complejos. Como se expuso en la sección A.1.1, facilitan el estudio de la no-linearidad de los sistemas complejos (como las sociedades humanas). Es más, se pueden encontrar diversas ventajas para su uso en ciencias sociales. Los ABM pretenden establecer una correspondencia directa entre entidades en el fenómeno social y los agentes computacionales, junto con las relaciones entre ellos. Este hecho desarrolla un potencial para representaciones naturales pero formales de entornos y fenómenos sociales [80]. Además, los ABM están estructuralmente preparados para modelar los procesos de emergencia y auto-organización típicos de los sistemas complejos. Esto se debe a que los agentes son entidades autónomas cuya interacción local en redes sociales complejas desarrolla el citado comportamiento emergente.

Hay varias propuestas metodológicas sobre el "cómo" construir modelos basados en agentes. Una metodología recopila experiencia en procesos, métodos y herramientas, proporcionando una forma robusta y sistemática de aproximarse a nuevos problemas. La sección 2.1 revisa distintas metodologías, junto con sus herramientas asociadas, provenientes del campo de sistemas multi-agente (SMA) e ingeniería del software orientada a agentes (AOSE). Sin embargo, estas metodologías no se centran en la simulación de "sociedades artificiales de agentes autónomos" [52] y por tanto normalmente no tienen en cuenta importantes consideraciones. Por ejemplo, los procesos de emergencia y auto-organización, la concepción del individuo como más que el agente racional dirigido por objetivos, o el rol de la teoría en la metodología (para una revisión en profundidad, acudir a la citada sección 2.1).

---

[1] ABM puede referirse a dichos modelos basados en agentes, "Agent-Based Model", o al proceso de construirlos: "Agent-Based Modelling". Los ABM son esencialmente distintos que los Sistemas Multi-Agente (SMA) de Inteligencia Artificial. Un ABM se refiere a un modelo computacional en el marco de la simulación social, centrado en el modelado y simulación de sociedades. Estas sociedades están compuestas por individuos que son representados por agentes autónomos, y por tanto por sistemas multi-agente. Pero tanto el campo como las implicaciones de estos acrónimos divergen sustancialmente.

Otros intentos de construir una metodología para el proceso de modelado provienen del propio campo de ABM. La sección 2.1 revisa los principales trabajos al respecto. Por ejemplo, las tres vías de modelar fenómenos de McKelvey [182] (positivista, teórica y para sistemas complejos), junto con los esfuerzos de mejorarla de Goldspink [109] y Gilbert [207]. Estos esfuerzos se aproximan a las fases de la lógica clásica de la simulación [107] de distintas maneras, explorando sus sub-fases y las relaciones entre ellas.

Sin embargo, existe una distinción epistemológica importante entre los modelos basados en agentes que puede tener implicaciones metodológicas, y que todas estas metodologías no consideran. Los modelos pueden ser clasificados en función de su objetivo de investigación [75]. En ocasiones, el investigador busca un modelo genérico para explicar el fenómeno social, desde un alto grado de abstracción y de forma suficientemente simplificada para ilustrar fácilmente una teoría o hipótesis concreta. Por contra, el investigador puede preferir centrarse en la expresividad del modelo, en la que se haga una extensa descripción empírica de un caso de estudio concreto [75, 223]. El primer caso corresponde a la llamada "investigación dirigida por teoría", mientras que la segunda es la "investigación dirigida por datos", como se observa en el continuo de la Figura A.1.

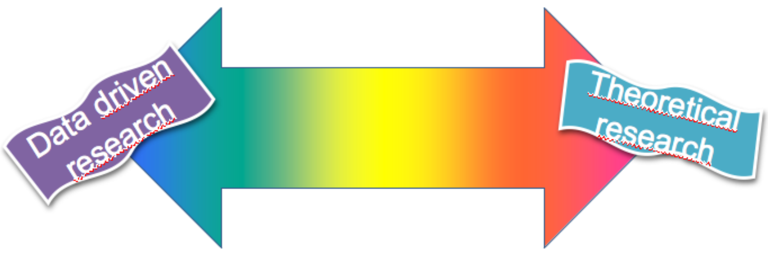

FIGURE A.1: La investigación dirigida por datos y dirigida por teoría pueden ser entendidas como dos tendencias opuestas en el modelado basado en agentes.

Hoy día, la mayor parte de los modelos son concebidos desde el punto de vista de la investigación dirigida por la teoría. Aunque pueden encontrarse en diferentes puntos de la parte derecha del espectro de la Figura A.1, normalmente ven el modelo principalmente como ilustración de una teoría.

Sin embargo, y sin ignorar el importante papel de la teoría, los modelos pueden buscar principalmente expresividad y descripción. Y para ello, pueden tener requisitos que no son abordados por dichas metodologías genéricas, a menudo centradas en modelos teóricos. Por ejemplo, temas como la inicialización empírica, las limitaciones de la recolección de datos, la validación empírica intensiva, o el papel de los datos en el diseño no son normalmente considerados en estas metodologías.

Así, existe una carencia de una metodología en ABM que, asumiendo que la investigación dirigida por datos tiene un enfoque y objetivos sensiblemente distintos, ofrezca un flujo estructurado para el desarrollo de modelos dirigidos por datos. Esta metodología debería considerar el papel clave de los datos empíricos a lo largo de las fases de modelado. Esta carencia ha provocado que los modelos dirigidos por datos existentes hayan sido construidos sin un marco común (como se observa en el estado del arte de modelos dirigidos por datos de la sección 2.2.2).

Así, puede afirmarse que cuando existen datos empíricos disponibles de la observación del caso de estudio, el proceso de modelado y simulación implica nuevas fases. Por ejemplo, los datos empíricos (o al menos distribuciones empíricamente probadas) deberían de ser considerados en lugar de utilizar las distribuciones aleatorias abstractas, en un esfuerzo por acercar el modelo al fenómeno real. Además, pueden contribuir en el diseño y la calibración del modelo (cf. [122] y sección 2.5).

Este trabajo pretende contemplar la relevancia de los datos empíricos en el proceso de modelado basado en agentes. Además, siguiendo el ejemplo de las metodologías AOSE, junto con esta metodología se ofrece un framework de agentes. Este framework se construye sobre el enfoque dirigido por datos que se expone, y facilita el desarrollo de modelos basados en agentes para una familia de problemas. Algunos problemas representativos de esta familia son:

- Estudiar la evolución de múltiples características individuales en una sociedad y periodo dados, especialmente en contexto de abundancia de datos sociológicos cuantitativos. Pueden ser internos como valores y comportamientos, o generales como las tasas de paro [134].

- Estudiar múltiples problemas del campo de la demografía computacional basada en agentes, como pirámides de población, patrones migratorios, dinámicas de amistad y emparejamiento, efectos de crisis de mortandad, consecuencias del descenso de la fertilidad y envejecimiento de la población, etc. [29].

- Este framework facilita la exploración [5] de los modelos complejos dirigidos por datos[2], ya que el framework permite el aislamiento de ciertas capas y módulos. De esta forma, es viable analizar el peso de los distintos factores en el efecto agregado resultante. E.g. dinámicas de opinión frente a la evolución demográfica, o el efecto de los medios, la emergencia de segundo orden, emergencia de normas, capacidades cognitivas, mecanismos de aprendizaje y otros.

Es posible explorar las posibilidad de distintas tecnologías en la simulación social simultáneamente al desarrollo del caso de estudio que se ha tomado. Algunos problemas

[2]La facilidad comunicativa de los modelos complejos dirigidos por datos es perdida en parte, si son comparados con los modelos teóricos KISS. Además, en modelos de tal dimensión y complejidad, a menudo es complicado realizar un análisis de sensibilidad que determine la influencia de cada factor en el resultado final. Un framework flexible como el propuesto, con módulos desactivables opcionalmente, puede facilitar enormemente esta tarea.

concretos pueden requerir tecnologías específicas que pueden ser integradas fácilmente como nuevos modelos en el sistema. Por ejemplo: sistemas de información geográfica (GIS), inteligencia artificial (como se explora en este trabajo), herramientas estadísticas, analíticas o visuales, etc.

Además, la elección de un problema complejo en el dominio de la sociología como caso de estudio, como la evolución de valores en la sociedad de la España postmoderna, es considerado esencial para apoyar la viabilidad del enfoque en ciencias sociales. Este problema es modelado siguiendo la metodología propuesta y utilizando el framework de agentes dado. Este ABM debería de proporcionar una mayor comprensión del fenómeno social bajo estudio.

Así, el caso de estudio se centra en el análisis de la evolución de diversos indicadores en España entre 1980 y 2000, centrándose en los valores sociales y predisposiciones mentales. Este periodo posee gran interés para la investigación, debido al gran cambio de valores y actitudes que la sociedad enfrentó entonces. Los casi 40 años de dictadura finalizaron en 1975, cuando el país estaba lejos de Europa en todos los indicadores de progreso, incluyendo los valores dominantes y el nivel de modernización (era una sociedad más tradicional, conservadora y religiosa). Sin embargo, en el año 1999, los datos cuantitativos (e.g. encuestas) muestran no sólo una fuerte convergencia con los valores europeos, sino llegando a posicionarse como una de las sociedades más modernas (en lo que a valores se refiere) [8, 148]. Así, el cambio en España ha sido desarrollado con una gran velocidad e intensidad durante el periodo bajo estudio. El problema a confrontar sería éste: el gran cambio de mentalidad en esta sociedad, en este periodo.

En las últimas décadas, las ciencias sociales estudian temáticas que, debido a poseer una naturaleza interna y subjetiva, han sido tradicionalmente difíciles de estudiar e incluso evitadas. Éste es el caso de los estudios en felicidad [168], emociones [57, 90], dinámicas de religiones [17, 18, 158], y valores [142, 214].

La investigación en ciencias sociales se ha aproximado al complejo tema de los valores (no valor económico, sino valor moral o social) desde distintas perspectivas. No obstante, el estudio del cambio de valores y su evolución se ha transformado sustancialmente desde los estudios de Ronald Inglehart al respecto [2, 147, 148].

Sus esfuerzos están enmarcados principalmente en el campo de la sociología cuantitativa. El estudio europeo de valores (EVS, de European Values Study), conducido en un principio por Jan Kerkhofs y Ruud de Moor [84], pero cuya extensión y profundización debe ser atribuida a Inglehart [147].

La EVS es una encuesta realizada desde 1981 en la mayor parte de Europa occidental, en distintas oleadas cada 9-10 años. Agrega los datos por países, haciendo un conjunto de preguntas relacionadas con valores e información complementaria a una muestra representativa de cada país [149] (para más detalles, ver sección 3.2.1). Este método proporciona posibilidades para la comparación entre distintos países en el mismo periodo, o en el mismo país en distintos periodos. Inglehart ha terminado extendiendo el procedimiento de la EVS a una encuesta a escala mundial: la World Values Survey (WVS) [150].

Los estudios de Inglehart se centran en la existencia de la macro-tendencia de la Modernización Cultural [147, 148]. Sus primeros trabajos fueron extendidos y apoyados por otros investigadores como Halman [83] y Flanagan [88]. Además, fueron estudiados en profundidad en el caso español por Arroyo [9], quien lo contrastó con otras fuentes tanto cualitativas como cuantitativas. Estos autores defienden la hipótesis del proceso de modernización de las sociedades desde una perspectiva tradicional hacia una moderna y luego postmoderna. Para un análisis de sus diferentes dimensiones y consecuencias, junto con otras investigaciones en el área, ver sección 3.2.3.

De acuerdo con los trabajos de Inglehart [148], este proceso de modernización cultural sería causado principalmente por las dinámicas inter-generacionales y sólo secundariamente por las intra-generacionales. Las dinámicas intergeneracionales tienen en cuenta los cambios entre generaciones, que son socializadas en diferentes valores. Las dinámicas intra-generacionales son consideradas cambios internos en el curso de la vida de una persona, es decir, la evolución dentro de un mismo grupo generacional. Esta perspectiva es justificada con la argumentación de que los valores son adquiridos y formados durante una etapa temprana de la vida (el llamado proceso de socialización), y se mantienen relativamente estables el resto de la vida de la persona. Es decir, cambios posteriores son posibles, pero de forma moderada. En dicho contexto, se espera una influencia determinante de los valores de los padres en los de los hijos.

Asumiendo la teoría de Inglehart y los estudios de Arroyo en la evolución de la religiosidad [10, 11, 14], el cambio de valores está principalmente dirigido por la inercia demográfica (responsable de las dinámicas intergeneracionales) y el proceso de socialización de los niños en sus familias. De ahí se puede deducir que un ABM que tuviera en cuenta principalmente estos procesos debería proporcionar resultados relativamente fiables en los indicadores de modernización (como religiosidad, ideología política o tolerancia en temas controvertidos). Si el comportamiento emergente observado en una simulación derivada de la teoría (y preferiblemente dirigida por datos empíricos) fuera consistente con los datos empíricos sobre el cambio de valores, significaría un apoyo a las hipótesis de Inglehart. Para un análisis más detallado del componente teórico-sociológico, junto con las razones para usar ABM en este problema ver el Capítulo 3, y ver la sección 7.3.4 para la validación teórica del modelo.

La evolución de valores es afectada por un gran número de factores dinámicos interrelacionados: género, edad, educación, economía, ideología política, religiosidad, familia, red social de amistades, patrones de emparejamiento y reproducción, ciclos de vida, tolerancia con respecto a determinados temas, etc. Así, este caso de estudio encaja muy bien el nivel de complejidad típico de los ABM dirigidos por datos.

De este modo, en el modelo son incluidas grandes cantidades de datos empíricos, junto con los supuestos teóricos esbozados aquí. Por ejemplo, la EVS se utilizará para inicializar la población de agentes; para diseñar el comportamiento de los agentes se usarán investigaciones cualitativas y la guía de un experto en el dominio; la literatura sociológica se utiliza para especificar la red social de amistades; ecuaciones empíricas reflejarán la dinámica demográfica española; y una validación empírica y otra teórica confirmarán la consistencia del modelo. Acúdase a la sección 2.6.2 para una revisión de otros ABM usando encuestas y a la sección 5.1 para otros trabajos de dinámicas sociales en simulación.

A.1.3 Objetivos

El principal objetivo de este trabajo es el desarrollo de una metodología y las herramientas apropiadas para facilitar el estudio y análisis de una familia de sistemas sociales. Su enfoque se centra en modelos expresivos y descriptivos, y por tanto fuertemente dirigidos por datos. Así, el subconjunto de problemas donde puede ser aplicado es determinado por su proximidad al fenómeno social[3]. La familia de problemas (explicada en profundidad en la sección 4.3) cubre todo el espectro de la demografía computacional basada en agentes [29], junto con el estudio de múltiples fenómenos sociales concretos dirigidos por datos. Este enfoque facilita la exploración [5] de estos modelos complejos, ya que el framework permite el aislamiento de ciertas capas y módulos. Así, es plausible analizar el peso de diferentes factores en el efecto agregado resultante.

La realización de esto conllevará dos contribuciones principales:

1. *Una metodología para simulación social basada en agentes dirigida por datos.* Para poder ser sistemático a la hora de aplicar el enfoque dirigido por datos a los modelos basados en agentes, se requiere una metodología. Este trabajo aborda la elaboración de dicha metodología para el proceso de construcción de modelos basados en agentes dirigidos por datos, utilizando datos empíricos no sólo para la validación sino también para el diseño e inicialización. Además, esta metodología propone guías para incorporar datos de encuestas (la principal herramienta de la sociología cuantitativa) y para integrar otras técnicas de Inteligencia Artificial en los modelos basados en agentes. Para una explicación detallada, acudir al Capítulo 2.

[3] Un problema teórico y abstracto sería modelar “los patrones de inmigración”, mientras que un problema cercano al fenómeno social sería “el flujo migratorio de españoles a Alemania en los años 60”

2. *Un framework de agentes modular y flexible* que dé soporte a la aplicación de la metodología especificada. Desde un enfoque de ingeniería del software, los investigadores del campo deben de ser promovidos a construir frameworks extensibles centrados en un contexto y una familia de problemas en lugar de construir una miríada de modelos aislados. Dicho framework de agentes consiste en una arquitectura modular de agente social (en el Capítulo 4) junto con la implementación de todo un sistema dirigido por datos flexible y extensible, compuesto por módulos activables de forma independiente. Este framework debe de ser capaz de tratar una familia de problemas que puedan ser validados empíricamente.

De este modo, los objetivos definidos implican las siguientes tareas:

1. *Estudiar como caso de estudio la evolución de valores sociales*, junto con otros parámetros sociales relacionados. Este problema tiene un alto grado de complejidad debido a la cantidad de factores interrelacionados implicados. El uso de un caso real concreto permite una valoración empíricamente testada de la metodología y el framework. Por tanto, el modelo basado en agentes denominado Mentat estudia el cambio de valores en la sociedad española durante el periodo 1980–2000 debido al proceso de postmodernización. Este modelo apoya la hipótesis de la importancia capital de la demografía en este complejo proceso. Para una perspectiva teórica profunda puede acudirse al Capítulo 3, mientras que para los detalles de implementación y experimentación puede verse el Capítulo 7.

2. *Especificar las dinámicas sociales* que proporcionan al modelo basado en agentes, con las capacidades de evolución demográfica, emparejamiento de individuos similares y dinámicas de amistad. Las decisiones tomadas deben estar apoyadas tanto teórica como empíricamente. Para una explicación detallada, ir a los Capítulos 5 y 6.

3. *Integración de múltiples tecnologías en modelado basado en agentes*: La modularidad de Mentat permite la exploración de modelos, con nuevos módulos integrando tecnologías de Inteligencia Artificial en un framework de agentes sociales. La metodología propuesta recomienda posibles usos de la IA en distintas partes del ciclo de modelado. De este modo, el modelo ha sido un campo de experimentación para la integración de la simulación social basada en agentes con lógica fuzzy, procesamiento de lenguaje natural o data mining:

   (a) *Promover el uso de agentes software fuzzy.* Este modelo promueve la exploración de los potenciales de la lógica fuzzy, especialmente en el contexto de la simulación social. Mentat utiliza agentes fuzzy y relaciones de homofilia fuzzy entre ellos. El proceso evolutivo de fuzzificación es explicado paso a paso, definiendo los procesos generales para fuzzificar un modelo nítido (no difuso o fuzzy). Debido a la adaptación natural de la lógica fuzzy a los conceptos de los seres humanos, encaja en las categorías de las ciencias sociales adecuadamente, permitiendo una mejora sustancial en la calidad del modelado. Para más detalle consultar el Capítulo 5.

(b) *Exploración de nuevos tipos de salida para los modelos.* Tradicionalmente, los modelos basados en agentes tienen una salida macro en forma de estadísticas (quizá incluyendo métricas de Análisis de Redes Sociales), gráficas y la visualización del espacio simulado. La modularidad de Mentat permite la exploración de salidas alternativas para la simulación. Una posibilidad es rastrear todas las acciones de los agentes y exportar su comportamiento micro en un fichero XML. Éste puede ser posteriormente procesado por un módulo externo que puede seleccionar el individuo más representativo del fenómeno estudiado. A continuación, y usando técnicas de procesamiento de lenguaje natural, es posible generar una biografía en lenguaje natural de la vida del agente seleccionado. Esto proporcionaría una comprensión a nivel micro del comportamiento general del modelo basado en agentes, presentándolo de una forma asimilable por cualquiera. Otra alternativa consistiría en la extracción de clusters tanto de la entrada del sistema (los datos empíricos de entrada) como de la salida (datos desagregados de la simulación) utilizando técnicas de data mining. Esto puede ser útil, entre otros usos, para una validación de segundo orden del modelo. Ambas tecnologías son estudiadas en el trabajo futuro, Capítulo 8.

A.1.4 Estructura del documento

En primer lugar, se debe hacer notar que no existe un capítulo específico de estado del arte. Ello es debido a que cada capítulo desarrolla su propia revisión de los principales trabajos concernientes al ámbito de cada tema. Así, cada capítulo está auto-contenido.

Este trabajo ha sido escrito con tres tipos de lectores en mente: informáticos, especialistas en simulación social y sociólogos puros. El vocabulario y contenidos están pensados para estos tres grupos, con un sesgo hacia conceptos formales y técnicos. No obstante, determinados capítulos están más enfocados a uno u otro perfil, permitiendo distintas lecturas de este volumen como se explica más adelante.

Los siguientes capítulos constituyen esta tesis doctoral:

1. **Introducción**. Este capítulo realiza una breve síntesis de los principales aspectos del campo de la simulación social, junto con una descripción de las motivaciones y objetivos del trabajo. Su propósito es servir como marco introductorio a cualquier lector.

2. **Metodología: introduciendo datos empíricos en simulación social**. Este capítulo representa la principal contribución de esta tesis: una metodología para el desarrollo de modelos basados en agentes dirigidos por datos. El proceso de modelado basado en agentes clásico es descrito, analizando cada una de sus fases. Las problemáticas que surgen al tener en cuenta el enfoque dirigido por datos son discutidas, realizando una revisión de las distintas metodologías existentes. Todo ello lleva a la exposición y discusión de la metodología propuesta.

3. **Caso de estudio sociológico: Mentat, la evolución de valores en la España postmoderna**. Éste es el caso de estudio elegido para la validación de la metodología y framework propuestos en este trabajo. Este capítulo está centrado en la teoría, y aunque puede tener poco interés para el informático, constituye la base para comprender las implicaciones sociológicas del caso de estudio.

4. **La visión micro: el agente de Mentat**. Describe el agente social construido siguiendo la metodología y framework propuestos, preparado para una familia de problemas concreta. Se discuten las alternativas existentes y se define su arquitectura, para concluir analizando el ciclo de vida del agente y sus principales características.

5. **Dinámicas sociales**. Este capítulo describe una de las mayores contribuciones de este trabajo. Agrupa una serie de conceptos clave: el proceso de fuzzificación de Mentat (y su justificación), el marco de comunicaciones de los agentes, las estrategias de similitud probadas, y las dinámicas sociales del sistema. Estos procesos son presentados formalmente, explicando el comportamiento de la red social que se genera.

6. **La visión macro: las condiciones globales**. Desde un enfoque a alto nivel se explica el sistema basado en agentes y su entorno. Se explican las principales decisiones tomadas desde el enfoque macro y los parámetros del sistema. Por último, el modelo demográfico, de importancia crítica en Mentat, es detallado.

7. **Experimentación y evolución de Mentat**. Este capítulo describe la evolución de la implementación del sistema, explicando cada estado de desarrollo de Mentat y cada decisión clave que lo ha hecho evolucionar. Siguiendo el enfoque dirigido por datos del "Deepening KISS" (explicado en el Capítulo 2), el modelo aumenta su complejidad en cada iteración. Por último, la salida es analizada desde el punto de vista técnico y sociológico.

8. **Conclusiones**. En el capítulo final, las principales contribuciones son explicadas en detalle. Además, las distintas líneas de trabajo futuro, centrándose en las

fusiones con otras técnicas de Inteligencia Artificial, son expuestas y discutidas en profundidad, con algunos resultados preliminares.

Dependiendo del perfil del lector de los tres definidos, se sugiere un flujo de lectura distinto. El especialista en simulación social será capaz de extraer el máximo beneficio de este trabajo, al constituir su principal audiencia. Así, el conjunto de la tesis puede resultar interesante, y deberían seguir el flujo presentado por el diagrama A.2. No obstante, dos capítulos de especial relevancia son el 2: Metodología y 5: Dinámicas sociales.

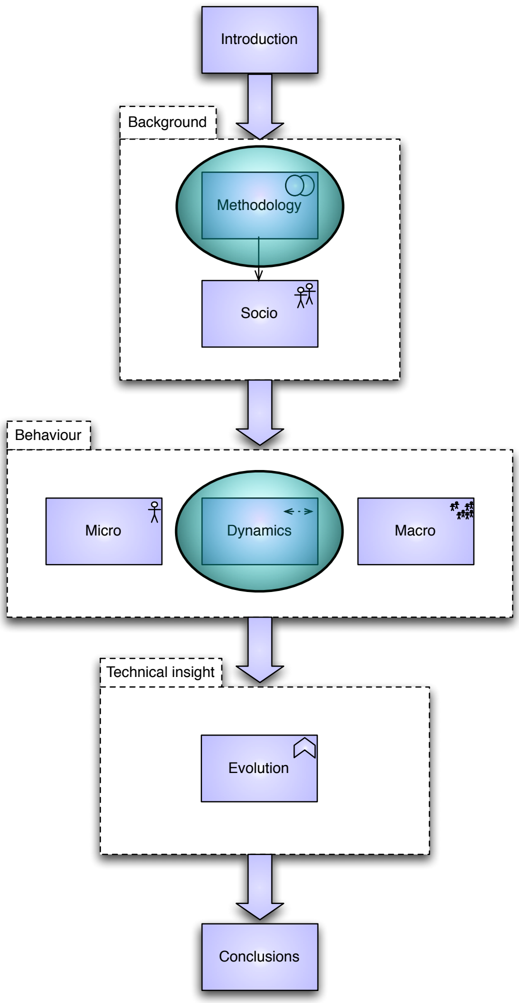

FIGURE A.2: El diagrama de flujo para la lectura del documento recomendado para los especialistas en simulación social.

Por otro lado, y como puede esperarse, sociólogos puros e informáticos estarán interesados sólo en partes de este trabajo. Los informáticos y expertos en Inteligencia Artificial (IA) deberían centrarse en el capítulo metodológico y especialmente en el de experimentación y evolución. Para aquellos interesados en aplicaciones de IA, los de

Dinámicas sociales y Conclusiones (la parte de trabajo futuro) pueden ser estudiados con detenimiento.

Por último, los sociólogos puros con pocas o ningunas nociones de informática aún tienen cabida en este trabajo. El capítulo introductorio debería dotarles del marco adecuado para abordar fácilmente el capítulo sociológico. Las últimas secciones del capítulo de experimentación y las primeras de las conclusiones cubren temas que pueden resultar relevantes. Finalmente, para el sociólogo que quiera profundizar en el campo, el capítulo de dinámicas sociales es recomendado encarecidamente.

A.2 Conclusiones

A.2.1 Resumen del trabajo

Este trabajo ha cubierto diversas dimensiones: la epistemológica, metodológica, técnica y sociológica:

- *Epistemológicamente*, se ha sumado al debate sobre los fundamentos de la simulación social desde un punto de vista decididamente experimental [65]. Así, este trabajo no sigue el enfoque de simulaciones abstractas y puramente teóricas que no pueden ser validadas empíricamente (es decir, a través de un experimento). Por otro lado, tampoco toma el enfoque positivista, en el cual los procesos sociales son reducibles a las observaciones y efectos cuantitativos, con la teoría sociológica jugando un papel secundario. Este trabajo apoya la perspectiva de la experimentación empírica guiada por la teoría, dando relevancia a dicha teoría y a los datos, tanto cuantitativos como cualitativos, por igual.

- Desde esa perspectiva, propone una *metodología* siguiendo el enfoque dirigido por datos, que guía la introducción de datos empíricos en la simulación, acercándola a la realidad pero reconociendo la importancia clave de la teoría en todo el proceso. Este enfoque es complementado con un método sistemático de exploración del espacio de modelos para alcanzar modelos comprensibles y descriptivos a la vez.

- Esta metodología es apoyada *técnicamente* por la especificación e implementación de un framework de agentes modular para facilitar la exploración del espacio de modelos y la construcción incremental de éstos, centrándose en la perspectiva dirigida por datos.

- Para validar la aplicación de la metodología y framework propuestos, se ha desarrollado en profundidad un caso de estudio de elevada complejidad. Éste estudia el problema de la evolución de valores sociales desde un punto de vista *sociológico*, junto con los procesos relacionados de emergencia de la amistad y dinámica demográfica. El modelo basado en agentes desarrollado apoya la importancia de dicha dinámica demográfica en la explicación de la evolución de valores en la España postmoderna.

La construcción de Mentat, como ha sido denominado el sistema del caso de estudio, puede ser resumido en una serie de aspectos. La propuesta metodología dirigida por datos es aplicada intensamente a lo largo de su desarrollo. El proceso de modelado ha sido realizado (a) bottom-up y (b) top-down. (a) es representado por la red social que surge del comportamiento micro y las dinámicas de amistad. (b) se sostiene en el elaborado modelo demográfico. Todo ello ha sido implementado en el marco del framework de agentes, diseñado de forma modular y en capas incrementales. Las capacidades de Mentat han sido estructuradas en módulos desactivables para poder explorar distintas combinaciones de modelos, siguiendo con la metodología definida en el Capítulo 2. El modelo es validado desde un enfoque macro cuantitativo (validación empírica), desde un enfoque micro cualitativo (correspondencia de la dinámica social con los supuestos teóricos) y desde un enfoque teórico (discutiendo su consistencia sociológica).

Además, distintas tecnologías de Inteligencia Artificial han sido incorporadas al modelo, probando la adaptabilidad del framework y la utilidad de éstas en simulación social. Junto con los agentes software, se ha utilizado intensamente la lógica fuzzy como se comprueba en el Capítulo 5. Otras tecnologías como el procesamiento de lenguaje natural o el data mining han sido exploradas, como se detalla en el Capítulo 8. Además, se han ofrecido guías sobre cómo aplicar otras que pueden resultar útiles (ontologías, algoritmos genéricos) durante el proceso de modelado en el Capítulo 2.

Mentat ha servido de caso de estudio para la metodología y el framework, pero a su vez ofrece un alto grado de comprensión sobre el problema, otorgando nuevo apoyo a determinadas teorías sociológicas. En concreto, este modelo enfatiza la importancia de la dinámica demográfica en el caso de estudio elegido. Esto implica que los cambios intergeneracionales son considerablemente más importantes que los intra-generacionales, al menos en el contexto español en ese periodo, reforzando así las teorías de Ronald Inglehart [147] sobre el cambio de valores.

A.2.2 Contribuciones al campo

Este trabajo ha contribuido al campo de la simulación social de diversas formas que son enumeradas aquí. La lista de publicaciones relacionadas, agrupadas por relevancia, están detalladas en la sección 8.1.3. En esta sección, las contribuciones se organizan por tema, con los artículos publicados referenciados en el tema respectivo.

- Aún no se ha desarrollado una metodología centrada en la tendencia dirigida por datos ("data-driven"), y ése es el principal objetivo de este trabajo. Se han realizado varias modificaciones a la lógica clásica de la simulación para hacerla más apropiada a los modelos dirigidos por datos [128, 130]. Además, se ha definido un método iterativo para el proceso de modelado denominado "Deepening KISS" (Profundizando en el KISS, donde KISS es el principio equivalente a la navaja de Occam) [118, 119, 122]. Finalmente, este método es completado con una especificación completa del ciclo de modelado dirigido por datos, teniendo en cuenta la nueva lógica de la simulación propuesta, el método "Deepening KISS" y las posibles tecnologías de inteligencia artificial que puedan resultar útiles [122, 123, 126].

- Para proporcionar herramientas para la aplicación de esta metodología, un framework de agentes modular ha sido especificado e implementado. Este framework dota al modelador de un entorno flexible para desarrollar una amplia familia de problemas (principalmente sociológicos), y está especialmente diseñado para seguir las directrices del enfoque dirigido por datos [120, 121, 129, 200, 201].

- Siguiendo esta perspectiva, se ha desarrollado un completo caso de estudio, proporcionado con este ejemplo un conocimiento más claro de las fases metodológicas necesarias para la aplicación del framework. El problema elegido ha sido la evolución de valores durante los últimos 20 años del siglo XX en la sociedad española, en los cuales esta sociedad va sufriendo el proceso de postmodernización. Un marco teórico para dicho problema ha sido construido para apoyar cada decisión de modelado en teorías e hipótesis sociológicas [8]. Mentat, el modelo basado en agentes resultante, está apoyado empírica y teóricamente, proporcionando un análisis profundo de los procesos sociales simulados [7, 15, 131].

- Probando cómo de robusto y flexible es el framework, distintas tecnologías de Inteligencia Artificial (IA) han sido integradas en diferentes módulos, combinándolas con los agentes software de Mentat, con distinto resultado. Los resultados más prometedores se han encontrado con la introducción de lógica fuzzy en el modelo. Es por ello que esta tecnología ha sido integrada por completo en los agentes y relaciones fuzzy [124, 125, 131, 132]. Otras tecnologías de IA que han resultado interesantes para la simulación social han sido el uso de procesamiento de lenguaje natural para la generación automática de biografías de agentes relevantes [127, 169–171], como es descrito en profundidad en el trabajo futuro del Capítulo 8. Además, los potenciales que puede desarrollar el data mining al aplicarlo en este campo han sido explorados, con resultados preliminares prometedores [126].

Bibliography

[1] Abello, A., Lymer, S., Brown, L., Harding, A., Phillips, B. Enhancing the australian national health survey data for use in a microsimulation model of pharmaceutical drug usage and cost. *Journal of Artificial Societies and Social Simulation*, 11(3):2 (2008).

[2] Abramson, P.R., Inglehart, R. *Value change in global perspective.* University of Michigan Press (1995).

[3] Albert, R., Barabási, A.L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97 (2002).

[4] Antunes, L. *Agentes com Decisao Baseada em Valores.* Ph.D. thesis, PhD thesis, Universidade de Lisboa, Lisboa, Portugal (2001).

[5] Antunes, L., Coelho, H., Balsa, J., Respicio, A. e*plore v.0: Principia for strategic exploration of social simulation experiments design space. In S. Takahashi, D. Sallach, J. Rouchier (Eds.), *Advancing Social Simulation: the First World Congress*, 295–306. Springer, Kyoto, Japan (2006).

[6] Antunes, L., Faria, J., Coelho, H. Improving choice mechanisms within the BVG architecture. In *Intelligent Agents VII, Proc. of ATAL 2000*, vol. 1986 of *Lecture Notes in Artificial Intelligence.* Springer–Verlag (2001).

[7] Arroyo, M., Hassan, S. Simulación de procesos sociales basada en agentes software. In *Actas del IX Congreso Español de Sociología.* Barcelona (2007). Grupo de trabajo I: Metodología.

[8] Arroyo, M., Hassan, S. Marco teórico-sociológico y operativización para modelar un sistema multi-agente sobre la evolución de la religiosidad española. In F.J. Miguel (Ed.), *Proceedings of the 2nd Workshop on Social Simulation and Artificial Societies Analysis (SSASA'08)*, vol. 442, 8. CEUR Workshop Proceedings, Barcelona (2009).

[9] Arroyo-Menéndez, M. *Cambio cultural y cambio religioso. Tendencias y formas de religiosidad en la España de fin de siglo.* Tesis Doctoral. Ed Complutense. Servicio de Publicaciones, Madrid (2004).

[10] Arroyo-Menéndez, M. Hacia una espiritualidad sin iglesia. In J.F. Tezanos (Ed.), *Tendencias en identidades, valores y creencias*, 409–436. Fundación Sistema (2004).

[11] Arroyo-Menéndez, M. La fuerza de la religión y la secularización en europa. *Iglesia Viva*, (224):99–106 (2005).

[12] Arroyo-Menéndez, M. Religiosidad centrífuga: ¿un catolicismo sin iglesia? *Iglesia Viva*, 111–119 (2005).

[13] Arroyo-Menéndez, M. Religiosidade e valores em portugal: comparação com a espanha ea europa católica. *Análise Social*, 42(184):757–787 (2007).

[14] Arroyo-Menéndez, M. Individualizacion y religion en la europa catolica. *Revista Espanola de Sociologıa (RES)*, 9:61–85 (2008).

[15] Arroyo-Menéndez, M., Hassan-Collado, S. Simulación de procesos sociales basada en agentes software. *Empiria - Revista de metodología de ciencias sociales*, (14):139–161 (2007).

[16] Atkinson, C., Kühne, T. Model-driven development: a metamodeling foundation. *IEEE software*, 36–41 (2003).

[17] Ausloos, M., Petroni, F. Statistical dynamics of religions and adherents. *Europhysics Letters*, 77:38002 (2007).

[18] Ausloos, M., Petroni, F. Statistical dynamics of religion evolutions. *Physica A: Statistical Mechanics and its Applications* (2009). In press.

[19] Avgeriou, P., Zdun, U. Architectural patterns revisited - a pattern language. In *Proc. 10th European Conference on Pattern Languages of Programs (EuroPlop 2005)*. Irsee, Germany (2005).

[20] Axelrod, R. Advancing the art of simulation in the social sciences. *Complexity*, 3(2):16—-22 (1997).

[21] Axtell, R.L., Epstein, J.M., Dean, J.S., Gumerman, G.J., Swedlund, A.C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., Parker, M. *Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley*, vol. 99. National Acad Sciences (2002).

[22] Babbie, E.R. *The Practice of Social Research.* International Thomson Publishing Services (1998).

[23] Bar-Yam, Y. *Dynamics of complex systems.* Westview Press (2003).

[24] Barabasi, A., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T. Evolution of the social network of scientific collaborations. *APS Meeting Abstracts*, 27012 (2002).

[25] Batty, M. Agent-based pedestrian modeling. *Environment and Planning B: Planning and Design*, 28(3):321–326 (2001).

[26] Beck, K., Andres, C. *Extreme programming explained: embrace change.* Addison-Wesley Professional (2004).

[27] Beckford, G., Giner, H.L. *La religione degli europei. Fede, cultura religiosa e modernitá in Francia, Italia, Spagna, Gran Bretagana, Germania e Ungheria.* Turim, Edizione della Fondacione Giovanni Agnelli (1992).

[28] Bernon, C., Gleizes, M.P., Peyruqueou, S., Picard, G. Adelfe: A methodology for adaptive multi-agent systems engineering. *Lecture Notes in Computer Science*, 156–169 (2003).

[29] Billari, F.C., Prskawetz, A. *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour.* 1 ed. Physica-Verlag Heidelberg (2003).

[30] Blanchard, P., Devaney, R.L., Hall, G.R. *Differential Equations.* 2 ed. Brooks Cole (2002).

[31] Boero, R., Castellani, M., Squazzoni, F. Micro behavioural attitudes and macro technological adaptation in industrial districts: an agent-based prototype. *Journal of Artificial Societies and Social Simulation*, 7(2) (2004).

[32] Boero, R., Squazzoni, F. Does empirical embeddedness matter? methodological issues on Agent-Based models for analytical social science. *Journal of Artificial Societies and Social Simulation*, 11 (2005).

[33] Bratley, P., Fox, B.L., Schrage, L.E. *A guide to simulation.* Springer Verlag (1987).

[34] Bratman, M.E., Intention, P. *Practical Reasoning.* Harvard University Press, Cambridge, MA (USA) (1987).

[35] Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., Mylopoulos, J. Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8(3):203–236 (2004).

[36] Brown, L., Harding, A. Social modelling and public policy: Application of microsimulation modelling in australia. *Journal of Artificial Societies and Social Simulation*, 5(4):6 (2002).

[37] Brown, T.N., Kulasiri, D. Validating models of complex, stochastic, biological systems. *Ecological Modelling*, 86(2-3):129–134 (1996).

[38] Bruch, E.E., Mare, R.D. Neighborhood choice and neighborhood change. *American Journal of Sociology*, 112(3):667–709 (2006).

[39] Carbó, J., Molina, J.M., Dávila, J. Fuzzy referral based cooperation in social networks of agents. *AI Commun.*, 18(1):1–13 (2005).

[40] Carley, K., Newell, A. The nature of the social agent. *Journal of mathematical sociology*, 19(4):221–262 (1994).

[41] Carley, K.M., Prietula, M.J., Lin, Z. Design versus cognition: The interaction of agent cognition and organizational design on organizational performance. *Journal of Artificial Societies and Social Simulation*, 1(3):1–19 (1998).

[42] Castelfranchi, C. Engineering social order. *Lecture Notes In Computer Science*, 1–18 (2000).

[43] Castellano, C., Fortunato, S., Loreto, V. Statistical physics of social dynamics. *0710.3256* (2007). Reviews of Modern Physics 81, 591-646 (2009).

[44] Cederman, L.E. *Emergent actors in world politics: how states and nations develop and dissolve*. Princeton University Press (1997).

[45] Centro de investigaciones sociológicas (CIS). (Spanish sociological research centre) (2008). Http://www.cis.es.

[46] Chalmers, A.F. *What is this thing called science?* Hackett Publishing Company (1999).

[47] Christley, S., Xiang, X., Madey, G. An ontology for agent-based modeling and simulation. In *Agent 2004 Conference* (2004).

[48] Cioffi-Revilla, C.A. Fuzzy sets and models of international relations. *American Journal of Political Science*, 25(1):129–159 (1981).

[49] Coleman, J.S., Fararo, T.J. *Rational choice theory: Advocacy and critique*. Sage Publications (CA) (1992).

[50] Collier, N. Repast: An extensible framework for agent simulation. *The University of Chicago's Social Science Research* (2003).

[51] Comte, A. *A General View Of Positivism.* Kessinger Publishing, LLC (2007).

[52] Conte, R., Gilbert, N., Sichman, J.S. MAS and social simulation: A suitable commitment. *Lecture Notes in Computer Science*, 1534:1–9 (1998).

[53] Conte, R., Paolucci, M. Intelligent social learning. *Journal of Artificial Societies and Social Simulation*, 4(1):U61–U82 (2001).

[54] Conte, R., Castelfranchi, C. *Cognitive and Social Action.* UCL Press, London (1995).

[55] Corbetta, P. *Social research: Theory, methods and techniques.* Sage (2003).

[56] Csermely, P. *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks.* 1 ed. Springer (2006).

[57] Damasio, A.R., Sutherland, S. *Descartes'error: Emotion, reason, and the human brain.* Quill New York: (2000).

[58] David, N., Sichman, J.S., Coelho, H. The logic of the method of agent-based simulation in the social sciences: Empirical and intentional adequacy of computer programs. *Journal of Artificial Societies and Social Simulation*, 8(4) (2005).

[59] Davie, G., Hervieu-Léger, D. *Identités religieuses en Europe.* La Découverte (1996).

[60] Davie, G. *Religion in modern Europe: A memory mutates.* Oxford University Press (2000).

[61] Dean, J.S., Gumerman, G.J., Epstein, J.M., Axtell, R.L., Swedlund, A.C., Parker, M.T., McCarroll, S. Understanding anasazi culture change through agent-based modeling. In *Dynamics in human and primate societies: agent-based modeling of social and spatial processes*, 179–205. Oxford University Press (2000).

[62] Deffuant, G. Improving agri-environmental policies: a simulation approach to the cognitive properties of farmers and institutions. *Final report of project FAIR*, 3 (2001).

[63] Deffuant, G., Amblard, F., Weisbuch, G., Faure, T. How can extremism prevail? a study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5(4) (2002).

[64] Deffuant, G., Huet, S., Bousset, J.P., Henriot, J., Amon, G., Weisbuch, G. Agent based simulation of organic farming conversion in allier département. *Complexity and Ecosystem Management*, 158–189 (2002).

[65] Deffuant, G., Moss, S., Jager, W. Dialogues concerning a (possibly) new science. *Journal of Artificial Societies and Social Simulation*, 9(1) (2006).

[66] Deffuant, G., Neau, D., Amblard, F., Weisbuch, G. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(4):87–98 (2000).

[67] DesJardins, J.R. *Environmental ethics: an introduction to environmental philosophy.* Wadsworth Belmont, CA (1993).

[68] Dia, H. An agent-based approach to modelling driver route choice behaviour under the influence of real-time information. *Transportation Research Part C*, 10(5-6):331–349 (2002).

[69] Diamond, J.M. Archaeology: Life with the artificial anasazi. *Nature*, 419(6907):567–569 (2002).

[70] Diaz-Agudo, B., Jiménez-Diaz, G., Recio-Garcia, J.A. How to teach semantic web?: a project-based approach. In *Proceedings of the 13th annual conference on Innovation and technology in computer science education*, 98–102. ACM New York, NY, USA (2008).

[71] Díaz-Salazar, R. *Iglesia, dictadura y democracia: Catolicismo y sociedad en España (1953-1979).* Hoac (1981).

[72] Díaz-Salazar, R. *España laica. Ciudadanía plural y convivencia.* Espasa, Madrid (2008).

[73] Domingos, P., Richardson, M. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 57–66. ACM, San Francisco, California (2001).

[74] Drogoul, A., Vanbergue, D., Meurisse, T. Multi-Agent based simulation: Where are the agents? *Lecture Notes in Computer Science*, 1–15 (2003).

[75] Edmonds, B. Simulation and complexity-how they can relate. *Virtual Worlds of Precision–Computer-Based Simulations in the Sciences and Social Sciences*, 5–32 (2005).

[76] Edmonds, B., Building, A., Street, A. The use of models-making MABS actually work. *Center for Poicy Modelling* (2000).

[77] Edmonds, B., Moss, S. From KISS to KIDS - an 'Anti-simplistic' modelling approach. In P. Davidsson, B. Logan, K. Takadama (Eds.), *MABS*, vol. 3415 of *Lecture Notes in Computer Science*, 130—144. Springer (2004).

[78] El-Nasr, M.S., Yen, J., Ioerger, T.R. FLAME—Fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3(3):219–257 (2000).

[79] Epstein, J.G., Möhring, M., Troitzsch, K.G. Fuzzy-Logical rules in a Multi-Agent system. In *SimSocVI Workshop, Groningen*, 25 (2003).

[80] Epstein, J.M. Agent-based computational models and generative social science. *Complexity*, 4(5):41–60 (1999).

[81] Epstein, J.M. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12 (2008).

[82] Epstein, J.M., Axtell, R. *Growing artificial societies: Social science from the bottom up*. The Brookings Institution and The MIT Press, Washington, D.C. and Cambridge, MA (resp.) (1996).

[83] Ester, P., Halman, L., Moor, R.D. *The individualizing society: Value change in Europe and North America*. Tilburg University Press (1994).

[84] European values study (EVS). Http://www.europeanvalues.nl.

[85] Faratin, P., Sierra, C., Jennings, N.R. Negotiation decision functions for autonomous agents. *Robotics and Autonomous Systems*, 24(3):159–182 (1998).

[86] Finin, T., Fritzson, R., McKay, D., McEntire, R. KQML as an agent communication language. In *Proceedings of the third international conference on Information and knowledge management*, 456–463. ACM New York, NY, USA (1994).

[87] Fishwick, P.A. Simulation model design. In *Proceedings of the 26th conference on Winter simulation*, 173–175. Society for Computer Simulation International San Diego, CA, USA (1994).

[88] Flanagan, S.C. Changing values in advanced industrial societies: Inglehart's silent revolution from the perspective of japanese findings. *Comparative Political Studies*, 14(4):403 (1982).

[89] Fort, H., Pérez, N. The fate of spatial dilemmas with different fuzzy measures of success. *Journal of Artificial Societies and Social Simulation*, 8(3) (2005).

[90] Frank, R.H. *Passions within reason: The strategic role of the emotions*. Norton New York (1988).

[91] Fuentes, R., Gómez-Sanz, J.J., Pavón, J. Activity theory for the analysis and design of multi-agent systems. *Lecture Notes in Computer Science*, 110–122 (2004).

[92] Fuentes, R., Gómez-Sanz, J.J., Pavón, J. Managing conflicts between individuals and societies in multi-agent systems. In *Engineering Societies in the Agents World V: Fifth International Workshop (ESAW-2004), Lecture Notes in Artificial Intelligence*, vol. 3451, 106–118. Springer (2004).

[93] Fuentes, R., Gómez-Sanz, J.J., Pavón, J. Social analysis of Multi-Agent systems with activity theory. *Lecture Notes in Computer Science*, 526–535 (2004).

[94] Galán, J.M., Izquierdo, L.R., Izquierdo, S.S., Santos, J.I., del Olmo, R., López-Paredes, A., Edmonds, B. Errors and artefacts in Agent-Based modelling. *Journal of Artificial Societies and Social Simulation*, 12(1):1 (2009).

[95] Galán, J.M., Lopez-Paredes, A., Olmo, R. An agent based model for domestic water management in valladolid metropolitan area. *Water Resources Research*, 45(W05401) (2009). (In press).

[96] Gamma, E., Helm, R., Johnson, R., Vlissides, J. *Design patterns: elements of reusable object-oriented software*. Addison-Wesley (1995).

[97] Gatherer, D. Comparison of eurovision song contest simulation with actual results reveals shifting patterns of collusive voting alliances. *Journal of Artificial Societies and Social Simulation*, 11 (2006).

[98] Geller, A. The emergence of individual welfare in afghanistan. In *Paper präsentiert am 20th International Political Science Association (IPSA) World Congress. Fukuoka*, vol. 9, 13 (2006).

[99] Geller, A., Moss, S. Growing QAWM: an Evidence-Driven declarative model of afghan power structures. *Advances in Complex Systems*, 11(2):321–336 (2008).

[100] Geller, A. Power, resources and violence in contemporary conflict: Artificial evidence. In *Proceedings of the World Congress of Social Simulation 2008* (2008).

[101] Gilbert, G.N. Simulation as a research strategy. In *Social Science Microsimulation [Dagstuhl Seminar, May, 1995]*, 448–454. Springer-Verlag London, UK (1996).

[102] Gilbert, N. Simulation: A new way of doing social science. *American Behavioral Scientist*, 42(10):1845–1847 (1999).

[103] Gilbert, N. When does social simulation need cognitive models? *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*, 428–432 (2006).

[104] Gilbert, N., Pyka, A., Ahrweiler, P. Innovation networks-a simulation approach. *Journal of Artificial Societies and Social Simulation*, 4(3):8–34 (2001).

[105] Gilbert, N. *Agent-Based Models.* annotated edition ed. SAGE Inc. (2007).

[106] Gilbert, N. *Researching Social Life.* Third edition ed. SAGE Ltd. (2008).

[107] Gilbert, N., Troitzsch, K.G. *Simulation for the Social Scientist.* 1 ed. Open University Press (1999).

[108] Glaser, N. Contribution to knowledge modelling in a multi-agent framework (the CoMoMAS approach). *PhDthesis, Universite Henry Poincare, Nancy I, France* (1996).

[109] Goldspink, C. Methodological implications of complex systems approaches to sociality: Simulation as a foundation for knowledge. *Journal of Artificial Societies and Social Simulation*, 5(1):1–19 (2002).

[110] Grossberg, S. Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1(1):17–61 (1988).

[111] Guillaume, D., Davis, N., Frederic, A., Gérard, W. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3:87–98 (2000).

[112] Gupta, A., Kapur, V. *Microsimulation in Government Policy and Forecasting.* Elsevier Science Publ. Co. (2000).

[113] Guyot, P., Drogoul, A. Designing Multi-Agent based participatory simulations. In H. Coelho, B. Epinasse (Eds.), *Proceedings 5th Workshop on Agent-Based Simulations*, 32–37. Publishing House, Lisbon (2004).

[114] Halman, L., Luijkx, R., van Zundert, M. *Atlas of European values.* Brill Academic Pub (2005).

[115] Halman, L., Pettersson, T. Differential patterns of secularization in europe: Exploring the impact of religion on social values. *Religion in Secularizing Society: The Europeans' Religion at the End of the 20th Century*, 48–75 (2003).

[116] Hamill, L. New communication technologies & travel - the story so far (2007).

[117] Harding, A. *Microsimulation and Public Policy*, vol. 232 of *Contributions to Economic Analysis.* Elsevier Science Publ. Co., Amsterdam (1996).

[118] Hassan, S., Antunes, L., Arroyo, M. Deepening the demographic mechanisms in a Data-Driven social simulation of moral values evolution. In N. David, J.S. Sichman (Eds.), *Multi-Agent-Based Simulation IX, Revised selected papers*, vol. 5269 of *Lecture Notes in Artificial Intelligence (from the Lecture Notes in Computer Science)*, 167–182. Springer-Verlag, Estoril, Portugal (2008).

[119] Hassan, S., Antunes, L., Arroyo, M. Deepening the demographic mechanisms in a Data-Driven social simulation of moral values evolution. In N. David, J.S. Sichman (Eds.), *Proceedings of the MABS 2008*, 189–203. Springer, Estoril, Portugal (2008).

[120] Hassan, S., Antunes, L., Pavón, J. A Data-Driven simulation of social values evolution. In Decker, Sichman, Sierra, Castelfranchi (Eds.), *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2009)*, 1337—1338. IFAAMAS, Budapest (2009).

[121] Hassan, S., Antunes, L., Pavón, J. Mentat: A Data-Driven Agent-Based simulation of social values evolution. In *MABS 2009 Proceedings*. Springer, Budapest (2009). (To Appear in Springer Lecture Notes in Artificial Intelligence).

[122] Hassan, S., Antunes, L., Pavón, J., Gilbert, N. Stepping on earth: A roadmap for data-driven Agent-Based modelling. In *Proceedings of the Fifth Conference of the European Social Simulation Association (ESSA08)*. Brescia, Italy (2008).

[123] Hassan, S., Fuentes-Fernández, R., Galán, J.M., López-Paredes, A., Pavón, J. Reducing the modeling gap: On the use of metamodels in Agent-Based simulation. In *Proceedings of the Sixth Conference of the European Social Simulation Association (ESSA09)*. Guildford, UK (2009). (To appear).

[124] Hassan, S., Garmendia, L., Pavón, J. Agent-Based social modeling and simulation with fuzzy sets. *Innovations in Hybrid Intelligent Systems (Springer-Verlag)*, (44):40–47 (2007). From the conference 2nd International Workshop of Hybrid Artificial Intelligence Systems (HAIS07).

[125] Hassan, S., Garmendia, L., Pavón, J. Introducing uncertainty into social simulation: Using fuzzy logic for Agent-Based modelling. *International Journal: Reasoning-based Intelligent Systems* (2009). (To appear).

[126] Hassan, S., Gutiérrez, C., Arroyo, J. Re-thinking modelling: a call for the use of data mining in data-driven social simulation. In *Proceedings of the 1st Workshop on Social Simulation at the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*. Pasadena, CA (2009). (To appear).

[127] Hassan, S., León, C., Gervás, P., Hervás, R. A computer model that generates biography-like narratives. In A. Cardoso, G.A. Wiggins (Eds.), *Proceedings of the 4th International Joint Workshop on Computational Creativity*, 5–12. London (2007).

[128] Hassan, S., Pavón, J., Antunes, L., Gilbert, N. Injecting data into Agent-Based simulation. In K. Takadama, G. Deffuant, C. Cioffi-Revilla (Eds.), *The Second*

World Congress on Social Simulation (tentative), Springer Series on Agent Based Social Systems. Springer, Washington, D.C. (2009). (To appear).

[129] Hassan, S., Pavón, J., Arroyo, M., León, C. Agent based simulation framework for quantitative and qualitative social research: Statistics and natural language generation. In F. Amblard (Ed.), *Proceedings of the Fourth Conference of the European Social Simulation Association (ESSA07)*, 697—-707. Toulouse, France (2007).

[130] Hassan, S., Pavón, J., Gilbert, N. Injecting data into simulation: Can Agent-Based modelling learn from microsimulation? In *Proceedings of the World Congress of Social Simulation 2008*. Washington, D.C. (2008).

[131] Hassan, S., Salgado, M., Pavón, J. Friends forever: Social relationships with a fuzzy Agent-Based model. *Hybrid Artificial Intelligence Systems, Third International Workshop, HAIS 2008*, (5271):523–532 (2008).

[132] Hassan-Collado, S., Mata-Garcia, M.G., Salvador, L.G. Aplicaciones de lógica borrosa a sistemas sociales con agentes software. In *Actas del Primer Congreso Internacional de Matemáticas en Ingeniería y Arquitectura*, 256–265. Madrid (2007).

[133] Hazlitt, H., Hazlitt, H. *The foundations of morality.* Van Nostrand (1964).

[134] Hedstrom, P. *Dissecting the Social: On the Principles of Analytical Sociology.* Cambridge University Press (2005).

[135] Hegselmann, R., Flache, A. Understanding complex social dynamics: A plea for cellular automata based modelling. *Journal of Artificial Societies and Social Simulation*, 1(3):1 (1998).

[136] Hegselmann, R. Understanding social dynamics: The cellular automata approach. In *Social Science Microsimulation [Dagstuhl Seminar, May, 1995]*, 282–306. Springer-Verlag (1996).

[137] Hegselmann, R.K. Opinion dynamics and bounded confidence models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3) (2002).

[138] Heiland, F. The collapse of the berlin wall: Simulating state-level east to west german migration patterns. *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, 73 (2003).

[139] Held, D. *Introduction to Critical Theory: Horkheimer to Habermas.* University of California Press (1980).

[140] Hervieu-Léger, D. *Religion as a Chain of Memory.* Rutgers University Press (2000).

[141] Hoffmann, A.O.I., Jager, W., Eije, J.H.V. Social simulation of stock markets: Taking it to the next level. *Journal of Artificial Societies and Social Simulation*, 10(2):7 (2007).

[142] Hofstede, G. *Culture's consequences: International differences in work-related values.* Sage (1984).

[143] Holmes, G., Donkin, A., Witten, I.H. WEKA: a machine learning workbench. In *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, 357–361 (1994).

[144] Huckfeldt, R.R. Social contexts, social networks, and urban neighborhoods: Environmental constraints on friendship choice. *American Journal of Sociology*, 89(3):651 (1983).

[145] Iglesias, C., Garrijo, M., Gonzalez, J. A survey of agent-oriented methodologies. *Lecture notes in computer science*, 317–330 (1999).

[146] Iglesias, C.A., Garijo, M., Gonzalez, J.C., Velasco, J.R. Analysis and design of multiagent systems using MAS-CommonKADS. In *Proceedings of the Intelligent Agents IV: Agent Theories, Architectures, and Languages: 4th International Workshop, ATAL'97*, 313. Springer, Providence, USA (1998).

[147] Inglehart, R. *The silent revolution: Changing values and political styles among Western publics.* Princeton University Press Princeton (1977).

[148] Inglehart, R. *Modernization and postmodernization: Cultural, economic, and political change in 43 societies.* Princeton University Press (1997).

[149] Inglehart, R. *Human Beliefs and Values: a cross-cultural sourcebook based on the 1999-2002 values surveys.* Siglo XXI (2004).

[150] Inglehart, R., university Consortium for Political, I., Research, S. *World Values Surveys and European Values Surveys 1981-1984, 1990-1993, and 1995-1997.* Inter-University Consortium for Political and Social Research (2000).

[151] Inglehart, R., Welzel, C. *Modernization, cultural change, and democracy: The human development sequence.* Cambridge University Press (2005).

[152] Instituto nacional de estadística (INE). (National statistics institute). spain (2008). Http://www.ine.es.

[153] Jennings, N.R. Commitments and conventions: The foundation of coordination in multi-agent systems. *Knowledge Engineering Review*, 8:223–223 (1993).

[154] Jennings, N.R., Faratin, P., Lomuscio, A.R., Parsons, S., Wooldridge, M.J., Sierra, C. Automated negotiation: prospects, methods and challenges. *Group Decision and Negotiation*, 10(2):199–215 (2001).

[155] Jick, T.D. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, 602–611 (1979).

[156] Karnopp, D., Rosenberg, R., Perelson, A.S. System dynamics: a unified approach. *IEEE Transactions on Systems, Man and Cybernetics*, 6(10):724–724 (1976).

[157] Katz, A., vom Hau, M., Mahoney, J. Explaining the great reversal in spanish america: Fuzzy-Set analysis versus regression analysis. *Sociological Methods Research*, 33(4):539–573 (2005).

[158] Kelley, D.M. *Why Conservative Churches are Growing: A Study in Sociology of Religion with a new Preface*. Mercer University Press (1996).

[159] Kendall, G., Yao, X., Chong, S.Y. *The Iterated Prisoners' Dilemma: 20 Years on*. World Scientific Publishing Company (2007).

[160] Kennedy, C., Theodoropoulos, G. Intelligent management of data driven simulations to support model building. In *Principles of Advanced and Distributed Simulation, 2006. PADS 2006. 20th Workshop on*, 132 (2006).

[161] Kennedy, C., Theodoropoulos, G., Sorge, V., Ferrari, E., Lee, P., Skelcher, C. AIMSS: an architecture for data driven simulations in the social sciences. *Simulation*, 1:2 (2007).

[162] Kluver, J., Schmidt, J. Control parameters in boolean networks and cellular automata revisited from a logical and a sociological point of view. *Complexity*, 5(1) (1999).

[163] Klüver, J., Stoica, C., Schmidt, J. Formal models, social theory and computer simulations: Some methodical reflections. *Journal of Artificial Societies and Social Simulation*, 6(2) (2003).

[164] Kottonau, J., Pahl-Wostl, C. Simulating political attitudes and voting behavior. *Journal of Artificial Societies and Social Simulation*, 7(4) (2004).

[165] Küppers, G., Lenhard, J. Validation of simulation: Patterns in the social and natural sciences. *Journal of Artificial Societies and Social Simulation*, 8(4) (2005).

[166] Labrou, Y., Finin, T., Peng, Y. Agent communication languages: The current landscape. *IEEE Intelligent systems* (1999).

[167] Latane, B. The psychology of social impact. *American Psychologist*, 36(4):343–356 (1981).

[168] Layard, R. *Happiness: Lessons from a New Science*. Penguin (2006).

[169] León, C., Hassan, S., Gervás, P. From the event log of a social simulation to narrative discourse: Content planning in story generation. In *Proceedings of the AISB Annual Convention*, 402–409. Newcastle, UK (2007).

[170] León, C., Hassan, S., Gervás, P., Pavón, J. Mixed narrative and dialog content planning based on BDI agents. *Lecture Notes in Artificial Intelligence, from the Lecture Notes in Computer Science Series (Springer-Verlag)*, (4788):150–159 (2007). Selected papers from 12th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2007.

[171] León, C., Hassan, S., Gervás, P., Pavón, J. Mixed narrative and dialog content planning based on BDI agents. In *Proceedings of the XII Conferencia de la Asociación Española Para la Inteligencia Artificial (CAEPIA-TTIA'07)*, vol. I, 27–36. Salamanca (2007).

[172] Lomborg, B. Nucleus and shield: The evolution of social structure in the iterated prisoner's dilemma. *American Sociological Review*, 61:278–307 (1996).

[173] Lopez-Paredes, A., Sauri, D., Galan, J.M. Urban water management with artificial societies of agents: The FIRMABAR simulator. *SIMULATION*, 81(3):189–199 (2005).

[174] Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K. Mason: A new multi-agent simulation toolkit. In *Proceedings of the 2004 SwarmFest Workshop*, vol. 8 (2004).

[175] Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., Balan, G. MASON: a multi-agent simulation environment. *Simulation*, 81(7):517 (2005).

[176] Lux, T., Marchesi, M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500 (1999).

[177] M., R. Ontology and epistemology for agent-based wayfinding simulation. *International Journal of Geographical Information Science*, 15:653–665 (2001).

[178] Macy, M.W., Willer, R. From factors to actors: computational sociology and agent–based modeling. *Annual Review of Sociology*, 28:143–166 (2002).

[179] Malsch, T., Schulz-Schaeffer, I. Socionics: Sociological concepts for social systems of artificial (and human) agents. *Journal of Artificial Societies and Social Simulation*, 10(1) (2007).

[180] Maslow, A.H. *Motivation and Personality*. 3rd revised edition ed. Longman Asia Ltd ,Hong Kong (1987).

[181] Mason multiagent simulation toolkit. Http://cs.gmu.edu/˜eclab/projects/mason/.

[182] McKelvey, B. Complexity theory in organization science: Seizing the promise or becoming a fad? *Emergence*, 1(1):5–32 (1999).

[183] McPherson, M., Smith-Lovin, L., Cook, J.M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444 (2003).

[184] Mentges, E. Concepts for an agent-based framework for interdisciplinary social science simulation. *Journal of Artificial Societies and social simulation*, 2(2):4 (1999).

[185] Metzler, T. Can Agent-Based simulation improve dialogue between science and theology? *Journal of Artificial Societies and Social Simulation*, 5(1) (2001).

[186] Moss, S. Alternative approaches to the empirical validation of agent-based models. *Journal of Artificial Societies and Social Simulation*, 11(1):5 (2008).

[187] Moss, S. Scott moss. simplicity, generality and truth in social modelling, part 2: Demonstration. In *Proceedings of the 2nd World Congress on Social Simulation (WCSS 2008)*. Washington, D.C. (2008).

[188] Moss, S. Simplicity, generality and truth in social modelling, part 2: Demonstration. In *Proceedings of the World Congress of Social Simulation 2008* (2008). (To appear).

[189] NetLogo. (Northwestern University). http://ccl.northwestern.edu/netlogo/.

[190] Newman, M.E.J., Watts, D.J., Strogatz, S.H. Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 1) (2002).

[191] Nicolaisen, J., Petrov, V., Tesfatsion, L. Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. *Evolutionary Computation, IEEE Transactions on*, 5(5):504–523 (2001).

[192] Nie, N.H., Bent, D.H., Hull, C.H. *SPSS statistical package for the social sciences*. McGraw-Hill, New York (1970).

[193] Norris, A.P., Inglehart, R. *Sacred and Secular.* Cambridge University Press, Cambridge (2004).

[194] North, M.J., Collier, N.T., Vos, J.R. Experiences creating three implementations of the repast agent modeling toolkit. *ACM Transactions on Modeling and Computer Simulation*, 16:1–25 (2006).

[195] Parsons, S., Sierra, C., Jennings, N. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292 (1998).

[196] Parsons, T. *The Social System.* The Free Press/Macmillan (1964).

[197] Pavard, B., Dugdale, J. An introduction to complexity in social science. *Tutorial on complexity in social science, COSI project* (2002).

[198] Pavon, J., Gomez-Sanz, J. Agent oriented software engineering with INGENIAS. *Lecture notes in computer science*, 394–403 (2003).

[199] Pavón, J., Gómez-Sanz, J., Fuentes, R. Model driven development of multi-agent systems. *Lecture Notes in Computer Science*, 4066:284 (2006).

[200] Pavón, J., Arroyo, M., Hassan, S., Sansores, C. Simulación de sistemas sociales con agentes software. In *Actas del Campus Multidisciplinar en Percepción e Inteligencia*, vol. I, 389–400. Albacete (2006).

[201] Pavón, J., Arroyo, M., Hassan, S., Sansores, C. Agent-based modelling and simulation for the analysis of social patterns. *Pattern Recognition Letters*, 29(8):1039–1048 (2008).

[202] Pavón, J., Gómez-Sanz, J. Agent oriented software engineering with INGENIAS. In *Multi-Agent Systems and Applications II*, vol. 2691 of *Lecture Notes in Artificial Intelligence*, 394–403. Springer-Verlag (2003).

[203] Petter, H., Andreas, G. Modelling the dynamics of youth subcultures. *Journal of Artificial Societies and Social Simulation*, 8(3):3 (2005).

[204] Polhill, J.G., Parker, D., Brown, D., Grimm, V. Using the ODD protocol for describing three Agent-Based social simulation models of Land-Use change. *Journal of Artificial Societies and Social Simulation*, 11(2):3 (2008).

[205] Popper, K. *The Logic of Scientific Discovery.* Reprint ed. Routledge (1992).

[206] Railsback, S.F., Lytinen, S.L., Jackson, S.K. Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9):609 (2006).

[207] Ramanath, A.M., Gilbert, N. Towards a methodology for agent-based social simulation research. *Agent-based Social simulation SIG, Barcelona* (2003).

[208] Rao, A.S., Georgeff, M.P. BDI agents: From theory to practice. In *Proceedings of the first international conference on multi-agent systems (ICMAS-95)*, 312–319. San Francisco (1995).

[209] Rao, A.S., Georgeff, M.P. Modeling agents within a BDI-Architecture. In R. Fikes, E. Sandewall (Eds.), *Proc. of the 2nd Intl. Conf. on Principles of Knowledge Representation and Reasoning (KR'91).* Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, Cambridge, MA (1991).

[210] Recio-García, J.A. OntoBridge.

[211] Reiter, E., Dale, R. Building applied natural language generation systems. *Natural Language Engineering*, 3(01):57–87 (1997).

[212] Repast agent simulation toolkit. Http://repast.sourceforge.net/.

[213] Roach, A. Con:Sims: simulating consumer behaviour (2007).

[214] Rokeach, M. *The Nature of Human Values.* LC control NO. 72092870 ed. Free Pr (1973).

[215] Rosenblueth, A., Wiener, N. The role of models in science. *Philosophy of Science*, 316–321 (1945).

[216] Roth, C., Taraborelli, D., Gilbert, N. Measuring wiki viability. *An empirical assessment of the social dynamics of a large sample of wikis. Submitted* (2008).

[217] Russell, S., Norvig, P. *Artificial Intelligence: A modern approach.* Prentice-Hall (1995).

[218] Sabater, J., Paolucci, M., Conte, R. Repage: Reputation and image among limited autonomous partners. *Journal of Artificial Societies and Social Simulation*, 9(2) (2006).

[219] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S. *Global Sensitivity Analysis: The Primer.* Wiley-Interscience (2008).

[220] Salvi, R. *Navier-Stokes equations: theory and numerical methods.* Chapman & Hall/CRC (1998).

[221] Sansores, C., Pavón, J. Agent-Based simulation replication: A model driven architecture approach. *Lecture Notes in Artificial Intelligence*, 3.789:244–253 (2005).

[222] Sansores, C., Pavón, J., Gómez-Sanz, J. Visual modeling for complex Agent-Based simulation systems. *Lecture Notes in Artificial Intelligence*, 3.891:174–189 (2006).

[223] Sansores-Pérez, C.E. *Metodología para el estudio de sociedades artificiales (PhD thesis)*. Universidad Complutense de Madrid (2007).

[224] Sapsford, R., Jupp, V. *Data collection and analysis*. Sage (1996).

[225] Schelling, T.C. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1:143—186 (1971).

[226] Schlicht, E. *Isolation and Aggregation in Economics*. Springer Verlag (1985).

[227] Schreiber, G. *Knowledge engineering and management: the CommonKADS methodology*. MIT press (2000).

[228] Schweizer, B., Sklar, A. *Probabilistic Metric Spaces*. Dover Publications (2005).

[229] Schwenk, G., Reimer, T. Simple heuristics in complex networks: Models of social influence. *Journal of Artificial Societies and Social Simulation*, 11(3):4 (2008).

[230] Simon, H.A. *Models of bounded rationality*. MIT Press Cambridge, MA (1982).

[231] Situngkir, H. The ribbon of love: Fuzzy-Ruled agents in artificial society. Tech. rep., BFI Working Paper Series (2007).

[232] Sloman, A. Explorations in design space. In *Proc. of the 11th European Conference on Artificial Intelligence* (1994).

[233] Smithson, M.J., Verkuilen, J. *Fuzzy Set Theory: Applications in the Social Sciences*. 1 ed. Sage Publications, Inc (2006).

[234] Spanish population census 1981, 1991, 2001. (Instituto nacional de estadística (INE). national statistics institute) (1981).

[235] Sterman, J.D. A skeptic's guide to computer models. *MIT System Dynamics Group.* (1991).

[236] Stroud, P., Valle, S.D., Sydoriak, S., Riese, J., Mniszewski, S. Spatial dynamics of pandemic influenza in a massive artificial society. *Journal of Artificial Societies and Social Simulation*, 10(4):9 (2007).

[237] Sun, R. *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. Cambridge University Press (2006).

[238] Sun, R., Naveh, I. Simulating organizational decision-making using a cognitively realistic agent model. *Journal of Artificial Societies and Social Simulation*, 7(3) (2004).

[239] Sun, R. *The Cambridge Handbook of Computational Psychology.* 1 ed. Cambridge University Press (2008).

[240] Taylor, A.J. The verification of dynamic simulation models. *Journal of the Operational Research Society*, 233–242 (1983).

[241] Taylor, R.I., University, M.M. *Agent-based Modelling Incorporating Qualitative and Quantitative Methods: A Case Study Investigating the Impact of E-commerce Upont the Value Chain.* Manchester Metropolitan University (2003).

[242] Thagard, P. *Computational philosophy of science.* MIT press (1993).

[243] Thayer, R.H., Christensen, M.J. *Software Engineering, The Development Process (Practitioners).* 3 ed. Wiley-IEEE Computer Society Pr (2005).

[244] Tisue, S., Wilensky, U. NetLogo: a simple environment for modeling complexity. In *International Conference on Complex Systems*, 16–21 (2004).

[245] Tisue, S., Wilensky, U. NetLogo: design and implementation of a multi-agent modeling environment. In *Proceedings of Agent 2004*, 7–9 (2004).

[246] Tobias, R., Hofmann, C. Evaluation of free java-libraries for social-scientific agent based simulation. *Journal of Artificial Societies and Social Simulation*, 7(1):6 (2004).

[247] Todd, P.M., Billari, F.C. Population-wide marriage patterns produced by individual mate-search heuristics. *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, 117 (2003).

[248] Troitzsch, K.G., Mueller, U., Gilbert, G.N., Doran, J.E. *Social Science Microsimulation.* Springer Verlag (1996).

[249] Tucker, P., Fletcher, I. Simulating household waste management behaviours part 2: Home composting. *Journal of Artificial Societies and Social Simulation*, 3(3) (2000).

[250] Tucker, P., Smith, D. Simulating household waste management behaviour. *Journal of Artificial Societies and Social Simulation*, 2(3):31 (1999).

[251] Urbig, D., Lorenz, J., Herzberg, H. Opinion dynamics: The effect of the number of peers met at once. *Journal of Artificial Societies and Social Simulation*, 11(2):4 (2008).

[252] Valverde, L. On the structure of F-Indistinguishability operators. Tech. rep., Computer Science Division, University of California at Berkeley, Berkeley, California (1984).

[253] Verbrugge, L.M. The structure of adult friendship choices. *Social Forces*, 56(2):576–597 (1977).

[254] Vindigni, G., Janssen, M.A., Jager, W. Organic food consumption. *British Food Journal*, 104(8):624–642 (2002).

[255] Wasserman, S., Faust, K. *Social Network Analysis: Methods and Applications*. 1 ed. Cambridge University Press (1994).

[256] Watts, D.J. *Six Degrees: The Science of a Connected Age*. 1st ed. W. W. Norton & Company (2003).

[257] Wellman, B., lin Wong, R.Y., Tindall, D., Nazer, N. A decade of network change: Turnover, persistence and stability in personal communities. *Social Networks*, 19(1):27–50 (1997).

[258] West, J.E., Linster, B. The evolution of fuzzy rules as strategies in Two-Player games. *Southern Economic Journal*, 69(3):705–717 (2003).

[259] Williams, M.B. Deducing the consequences of evolution: a mathematical model. *Journal of Theoretical Biology*, 29(3):343 (1970).

[260] Witten, I.H., Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. 2 ed. Morgan Kaufmann (2005).

[261] Wolfram, S. Theory and applications of cellular automata. *Advanced Series on Complex Systems, Singapore: World Scientific Publication, 1986* (1986).

[262] Wolfram, S. *A New Kind of Science*. 1 ed. Wolfram Media (2002).

[263] Yager, R.R. Families of OWA operators. *Fuzzy Sets Syst.*, 59(2):125–148 (1993).

[264] Yang, L., Gilbert, N. Getting away from numbers: Using qualitative observation for agent-based modeling. *Advances in Complex Systems*, 11(2):175–186 (2008).

[265] Yang, L., Gilbert, N. Getting away from numbers: Using qualitative observation for agent-based modelling. In F. Amblard (Ed.), *Proceedings of the ESSA'07: Fourth Conference of the European Social Simulation Association*, 205–214. Toulouse, France (2007).

[266] Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 100(Supplement 1):9–34 (1999).

[267] Zadeh, L.A. Fuzzy sets. *Information and Control*, 8:338–353 (1965).

[268] Zadeh, L.A. Similarity relations and fuzzy orderings. *Information Science*, 3:177–200 (1971).

[269] Zadeh, L.A. Fuzzy sets. In *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A. Zadeh*, 394–432. World Scientific Publishing Co., Inc. (1996).

[270] Zambonelli, F., Omicini, A. Challenges and research directions in agent-oriented software engineering. *Autonomous Agents and Multi-Agent Systems*, 9(3):253–283 (2004).