# GeneCity: A Multi Agent Simulation Environment for Hereditary Diseases

Demetrios G. Eliades[1], Andreas L. Symeonidis[1,2] and Pericles A. Mitkas[1,2]
[1]Electrical and Computer Engineering Dept.,
Aristotle University of Thessaloniki, GR541 24, Thessaloniki, Greece
[2]Intelligent Systems and Software Engineering Laboratory,
Informatics and Telematics Institute – CERTH, GR570 01, Thessaloniki, Greece
eldemet@ucy.ac.cy, asymeon@iti.gr, mitkas@eng.auth.gr

## Abstract

*Simulating the psycho-societal aspects of a human community is an issue always intriguing and challenging, aspiring us to help better understand, macroscopically, the way(s) humans behave. The mathematical models that have extensively been used for the analytical study of the various related phenomena prove inefficient, since they cannot conceive the notion of population heterogeneity, a parameter highly critical when it comes to community interactions. Following the more successful paradigm of artificial societies, coupled with multi-agent systems and other Artificial Intelligence primitives, and extending previous epidemiological research work, we have developed GeneCity: an extended agent community, where agents live and interact under the veil of a hereditary epidemic. The members of the community, which can be either healthy, carriers of a trait, or patients, exhibit a number of human-like social (and medical) characteristics: wealth, acceptance and influence, fear and knowledge, phenotype and reproduction ability. GeneCity provides a highly-configurable interface for simulating social environments and the way they are affected with the appearance of a hereditary disease, either Autosome or X-linked. This paper presents an analytical overview of the work conducted and examines a test-hypothesis based on the spreading of Thalassaemia major.*

## 1 Introduction

Analytical modeling problems are by nature complex and heterogeneous and the mathematical models usually used to solve them oversimplify the laws that govern systems, by presuming that populations are homogeneous and that they cooperate in harmony. This approach is also followed in the case of epidemiological studies and population genetics in Biology and Medicine. In order to address the problem of handling heterogeneous populations and introduce own will and individuality into the system an alternative approach has to be taken, which shall allow behavior differentiation between community members [16]. This approach is the exploitation of software agent technology combined with Cellular Automata, Genetic Algorithms and Social Sciences primitives. Such a combination allows the comprehension of heterogeneous populations and the stochastic processes on the dispersion of a hereditary disease they invoke, based on real time user intervention on system parameters. All this is, of course, not feasible with traditional mathematical models.

Humans in their social systems have language capacity and are reflexive and self-aware; for the software agents, this is another layer of complexity, that has not yet been resolved, theoretically or methodologically [13]. In general, interpreting a society and its mentality, remains a difficult task, since it is almost impossible to completely cover in full range all its aspects, be they visible, or hidden under a chaotic way of behavior.

Keeping these in mind, we have developed *GeneCity*, a multi-agent system (MAS) that simulates a human society in order to study three types of hereditary diseases and how these affect a population. Agents are provided with common social characteristics, such as education on the disease and fear for the birth of non-healthy descendants, and are provided with the ability to select their mate with respect to certain criteria such as wealth, topological distance, and age difference.

GeneCity deploys an initial population of autonomous agents under user-specified demographical and sociological constraints and allows it to evolve throughout the course of time. The user can, at any time, monitor changes in biological and social processes like mate selection, marriages,

deaths, births, genetic cross-over and mutation.

This paper aims to: a) introduce GeneCity and its core functionalities, b) explore the use of MAS in epidemiological simulations, within an easy-to-use and highly parameterized environment, and c) conduct a series of indicative experiments, leading to interesting conclusions.

The paper is structured as follows: Section 2 provides a survey on the related work on MAS simulation and epidemiology; Section 3 outlines the necessary background information needed in biology and sociology; Section 4 gives an overview of the specifications and design details of the project; Section 5 briefly describes the implementation of the system; Section 6 evaluates the system with a test-hypothesis, while Sections 7 discusses future work and concludes the paper.

## 2 Literature Survey

Traditional simulating techniques rely mostly on differential equations to describe the dynamics of a system. This is also the case for traditional epidemiological research, where the differential equations of Kermack and Mckendrick have been used. Their SIR model considers three classes: the susceptibles (S), the Infected (I) and those who have immunity (R) [20]. Typical concessions in the SIR model include a fixed population in a closed society, zero incubation period, a duration of infectivity that last as long as the clinical disease lasts, and zero deaths during that period. Mathematical models like SIR suffer from an inherited limitation; they consider sets as homogeneous, instead of heterogeneous [16].

As research has shown, multi-agent systems can efficiently implement distributed, by their nature, models [9]. According to Epstein and Axtell [1996], the agent-based simulation can be quite useful in the studying of macroscopic phenomena and behaviors, the spread of a disease and the dynamic evolution of populations. The use of genetic algorithms [17] in these models has given the agents the ability to be exposed to natural selection, widening their abilities from solving engineering problems [12], to strategic behavior in games (e.g., the "Prisoner's dilemma") [1]. In an agent-based model, each agent has its own genetic and cultural characteristics, which can be modified during its lifecycle through its interaction with the rest of the agents and the environment.

Cellular Automata (CA) [15] is another technique widely used in simulation systems, as a promising approach for understating social dynamics. CA have been exploited for epidemic infections and transmission, were each point in the two-dimensional lattice world represents a heterogeneous individual [19]. The use of CA provides topological information to the agent, that being another factor which affects disease transmission [8].

Various platforms have been developed for simulating various models, from simple ones to more complex, i.e., SWARM [23], Sugarscape [8], Ascape [24] and Echo [18]. Despite of the plethora of simulation tools, most epidemic models simulated are not related to hereditary transmission; rather they relate to physical contact between persons. From these models, we mention the Smallpox Bioterror simulation [7], based on the Ascape platform, which applies multilevel movement of the agents (work, school and hospital), where they spread the disease. In a similar manner, a simulation on the transmission of the SARS [19] searches for the best policy for dealing with the disease. Although there are many approaches for pathological epidemiology, they cannot be used as is for simulating and analyzing a hereditary disease in social-agent environments. This was the primary reason for developing GeneCity.

## 3 Background Information

### 3.1 Hereditary Diseases

A human has 46 chromosomes, 44 of which are autosome and 2 are sex-related (X and Y; Males have XY, while Females XX). The mutation of a gene can sometimes affect the normal functions of the body, leading to a hereditary disease. A person is thought to be a *homozygote* when he/she has a pair of identical mutated genes, and *heterozygote* when there is a pair of non-identical genes, one of which is mutated. A characteristic that may appear on a homozygote (patient) but not on a heterozygote (carrier of the trait) is called *recessive*, while a characteristic that appears to both is denoted as *dominant*. Transmission of a mutation occurs mainly as:

1. Autosomal Recessive
2. Autosomal Dominant
3. X-linked (or Sex-related)

Both the Autosomal and the X-linked transmission obey the Mendelian laws (e.g. if Parent1 has [ab] genes and Parent2 has [AB] genes, then children will have [aA, aB, bA, bB], with a 25% probability for each of these to occur). The difference of Autosomal with X-linked mutation is that in the latter case mutation can appear only on the X-chromosome.

### 3.2 Mate Selection

Many parameters affect mate selection in human beings, namely random meeting, beauty, wealth and capital, reproduction ability, even age difference [26, 4]. If choosing a partner is essential for the optimization and diversity of life on the planet, we can expect that choosing a partner can be

equally important in designing an artificial system also [2]. Health may also be one of the parameters affecting selection.

It is therefore obvious that mate selection would play a pivotal role in our simulation. In order to best match males with females based on their preferences (a problem also known as the "Stable Marriage Problem"), we have followed the approach introduced by D. Gale and L. Shapley [22].

# 4 Overview of GeneCity

## 4.1 System Model

The Genecity model abides by the specifications dictated by sociologists (concerning family-related matters) and biologists (concerning hereditary disease spreading matters). Their domain knowledge has been incorporated into the model by defining the environment and its governing rules, as well as the personal characteristics, beliefs and wills that each agent posses.

The personality of each agent instantiated in the system is defined by modifying certain parameters, as well as the initial scenario of interaction with the other agents. Key hypotheses that the model takes into account include: the artificial society is comprised of married and single agents; all agents want to have children; all agents are capable of child birth; each agent can get married only once; relatives are not allowed to get "married"; a male (or female) may propose (to mate) to a number of females (or males) and decide on the best candidate; social characteristics in each agent affect the mating decision, as well as the number of children the agent will bring to life.

## 4.2 Specifications

GeneCity is an agent society that provides a graphical user-interface for users to monitor evolution and modify parameters. It comprises five types of agents, each responsible for its own field of "expertise". These types are:

1. The *SuperAgent*, which is the coordinator of the system and the GUI agent. The SuperAgent has full control over every agent and is capable of changing all system parameters in real time.

2. The *MatchAgent*, which is responsible for the proper functioning of the mating processes.

3. The *MediaAgent*, which simulates all forms of mass media within a society, providing information on a disease.

4. The *HumanAgent*, representing the single.

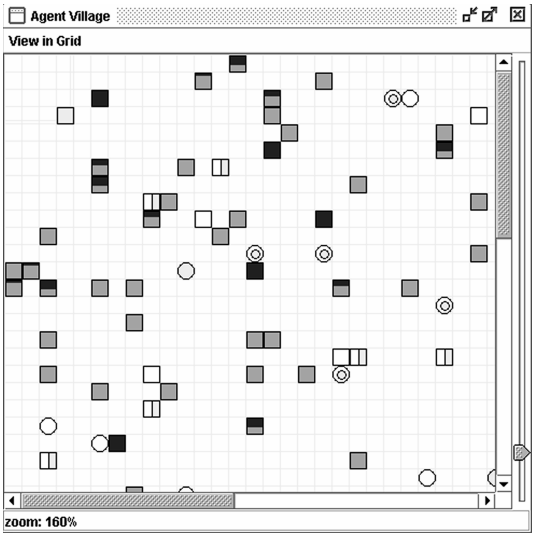


**Figure 1. The GeneCity grid**

5. The *FamilyAgent*, representing the family.

GeneCity also provides a number of monitoring tools, including real-time graphical and genealogical tree representations, for the exploring of mutation evolution to become feasible.

## 4.3 The Environment

The environment is defined as a two-dimensional ($x \times y$) grid specified by the user. Each cell in the grid may be either a vacant space, a HumanAgent (female, male; healthy, heterozygote, homozygote) or a FamilyAgent (Figure 1). In order for agents to be represented upon the grid, medical genealogical symbols [5] have been adopted (Figure 2). Symbols 1 and 2 represent healthy Males and Females, 3 and 4 Carrier Males and Females, and symbols 5 and 6 represent Patient Males and Females, respectively.

## 4.4 Human Agents

The genetic code of HumanAgents comprises 31 genes (bits): 9 genes imprint the biological characteristics of the agent, while the rest are related to its social characteristics. The biological characteristics include Sex (1-bit), Type of

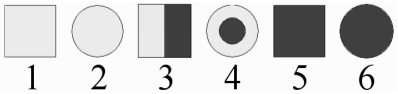


**Figure 2. Medical symbolisms for HumanAgents**

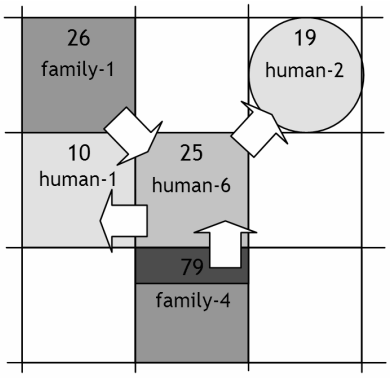


**Figure 3. Information Exchange between agents**

Disease (2-bits), Mutation Gene (2-bits), Health Status (1-bit), Reproduction Ability (1-bit), and Phenotype Visibility (2-bits), while the social characteristics include Wealth-Beauty (4-bits), Degree of acceptance on new information (4-bits), Power to Influence others (4-bits), Fear to diseases in general (5-bits), and, finally, Extroversion (5-bits).

Instead of an one-point crossover, we have implemented a multiple-point one, on the Mutation Gene, Wealth, Acceptance, Influence, Fear, and Extroversion characteristics. The others are directly specified by the user and remain intact.

### 4.5 Information Degree

In real life, information on diseases comes from versatile sources, namely health education, mass media (newspapers, television, radio and internet), as well as human interaction. In order to incorporate this parameter into the system, we have specified a *Degree of Information* for the HumanAgents. Three factors affect this parameter:

1. MediaAgent influence. The MediaAgent influences part of the HumanAgents group.

2. Family Education. During childhood, agents grasp the family mentality on the disease. The degree of grasping is analogous to their degree of acceptance of new information.

3. Moore Neighborhood Learning. Within one's (societal) neighborhood, a number of agents may exist, exchanging information on a disease. Figure 3 illustrates a case where Human-6 provides information to Humans 1 and 2, and receives information from Families 1 and 4.

### 4.6 Fear of Transmission

Fear plays a pivotal role in families, when confronted with the case of giving birth to a child that suffers from a hereditary disease. They feel guilty and pointed-out and they adopt either the "No more pregnancies" or the "Neglect or accept the danger in future pregnancies" attitude [5]. The dilemma of having a child or not, when knowing beforehand that it will be diseased after birth, creates new ethical decisions to the society [11]. Research has shown that negative experiences in past pregnancies affect this decision [21].

This fear is incorporated into the HumanAgents and FamilyAgents through the *Degree of Fear*. Increased Degree of Fear implies increased cautiousness in decisions related to mate selection and child bearing. The degree changes in a reinforcement learning manner; for example, when homozygote patients are born, the fear is increased, always with respect to the patients' personal general fear value within their genetic code.

In general, the fear of transmission and the degree of information parameters are relevant to each other. A person with high degree of fear but with no information about the disease, and a person with high information values but not fear about it, will act in the same manner: the first in ignorance and the second in recklessness.

### 4.7 Mating Preference Value

When agents reach their mating age, they create, based on their preferences, a list of all the candidate agents of the opposite sex. Consequently, the list is sorted on the *value of preference* for each agent, given by a function that considers physical proximity, age and age difference, state of health, ability to reproduce, wealth or beauty, and the probability to give birth to a child suffering from a disease. It should be noted that the system does not allow mating between relatives up to the 2nd degree. As already mentioned, mating is conducted under the supervision of the MatchAgent.

## 5 Implementation

GeneCity provides a multi-functional user interface to facilitate researchers in their experiments. It has been implemented in Java v.1.4.1 and all the agents are developed over the Java Agent Development Framework (JADE) framework [3], which conforms to the FIPA specifications [10]. The system allows full control over the agents, from the moment of their creation until they are terminated. GeneCity agents communicate via a common ontology (the SimulationOntology), designed and built with Protégé 2000 [14]. A snapshot of GeneCity "in action", with all its internal windows expanded, is illustrated in Figure 4.

Through the "Settings" menu, the user can configure the biological, as well as the demographical and sociological parameters of the simulation at hand. The "Agent Remote Control" interface provides the user with the ability to modify the weights of the Preference Value function, to introduce a new agent and, finally, to activate or deactivate Neighborhood and Media Learning throughout the course of a simulation.

In addition, GeneCity provides a number of system indicators, as well as information on the agents who live and die in the system. The medical/genealogical tree of an agent can be viewed by selecting it, either on the agent table, or on its position in the GeneCity grid. Data containing information on healthy, carriers, and patients are available during the execution of the simulation. The system can also provide a real-time graph of the total population, as well as of the populations of the carriers and the patients. Finally, all data generated during the simulation are recorded and can be stored in text files for posterior analysis.

## 6 Evaluation

In order to demonstrate the proper functioning of GeneCity, we have set up a real-life scenario, based on the Thalassaemia disease.

### 6.1 The Thalassaemia Example and Population Screening

Thalassaemia major, also known as Mediterranean Anaemia or Cooleys Anaemia, is a hereditary disease transmitted in an Autosome Recessive manner. If not treated adequately, Thalassaemia major can exhibit severe phenotype symptoms with poor growth, facial and other bone deformities, fragile bones and bone fractures, enlarged liver and spleen, as well as impairment of normal physical activities. Recent studies have shown that regular transfusion therapy with safe and appropriately processed blood, combined with regular and effective iron chelation, can improve patients' quality of life [6]. Thalassaemia is usually met across southern Europe, in some central European countries, in Middle East, southern China, as well as in northern Africa and South America. It is estimated that about 1.5% of the global population are carriers of the disease, although it is certain that this number is a gross underestimate, as Eleftheriou points out.

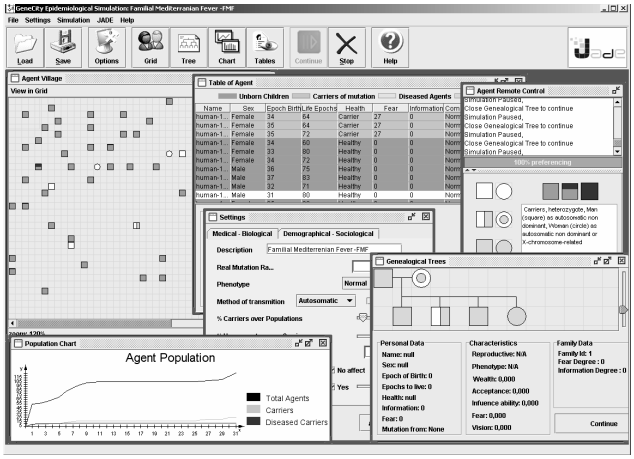


**Figure 4. GeneCity during simulation**

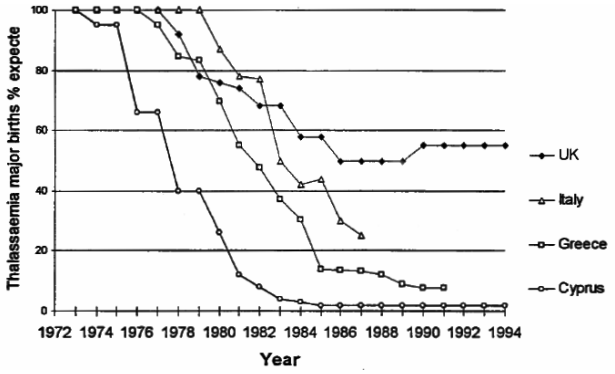


**Figure 5. The efficiency of Genetic Screening**

When an inherited condition is common and serious with detectable carriers, it may be appropriate to offer genetic screening to the whole population [25]. *Genetic Screening* is a basic test that is systematically run on a defined population, in order to identify a group with increased genetic risk. In Cyprus, for example, where the carrier frequency was 1 out of 7 people in the 70's, a policy of community education and premarital screening for Thalassaemia has reduced affected births to less than 5% of expectation. In the UK, on the other hand, only 50% of couples at risk are detected on time. Figure 5 shows that countries like Cyprus, Greece, and Italy, that promote premarital screening, have succeeded significantly in reducing Thalassaemia major births.

### 6.2 Simulation Data and a Test Hypothesis

The evaluation conduced is based upon the example of Thalassaemia. We have set up a population of 100 people, 15% of which are carriers of the trait and we consider that 25% of the carriers are patients. The fact that the disease is transmitted in an Autosome Recessive manner implies that the affected agents lifespan is not influenced. We have accepted that there is no prior information on the disease. Each epoch, the MediaAgent broadcasts information on the disease to all agents, while Neighborhood learning

abilities are enabled. Only health affects the mating function. The percentage of successful screening, i.e. not allowing marriages between two carriers or not allowing the birth of children with disease, is defined by *Pre-birth Control* (PbC) variable. By altering the *Pre-birth Control* (PbC) parameter, we have run a number of experiments, in order to validate our hypothesis: *Pre-birth control can indeed affect the number of patients and carriers in a population and decrease in the birth rate of children with the hereditary disease.*

The base case of the experiment is when no pre-birth control is conducted.

## 6.3 Results

Five experiments were conducted with five different values of PbC, namely 0, 20, 50, 80, and 100%. The societies were allowed to reach an equilibrium (328 Epochs) and the resulting data were used for further processing.

### 6.3.1 Frequency of "diseased" Births

The histogram in Figure 6 illustrates the frequency of "diseased" births per epoch, for the experiments with PbC values 0% and 80% respectively. For the base case, where PbC=0%, the births of diseased children increase exponentially, whereas for PbC=80%, they are reduced dramatically. Therefore by increasing the PbC, there was decrease in the frequency of "diseased" births.

### 6.3.2 Patients through the Epochs

Figure 7 illustrates the number of patients over the total population, for all the experiments. For the base case of PcB=0% there is a linear increase, from epochs 100 to 300, with patients reaching 9% of the total population. For the rest of the experiments, the curves follow an exponential decrease. For PbC=20%, the number of patients decreases and remains below 1% of the population, whereas for PbC=50% and 80%, patients are almost diminished to 0%. Especially for the case where PbC=100%, patients extinct in less than 100 epochs.

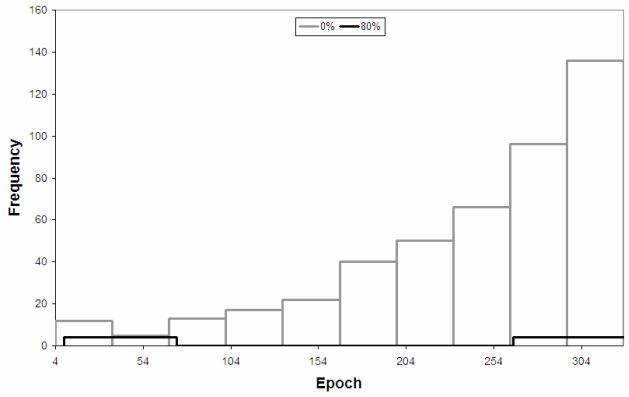


**Figure 6. Frequency Histogram of Births with Hereditary Disease**

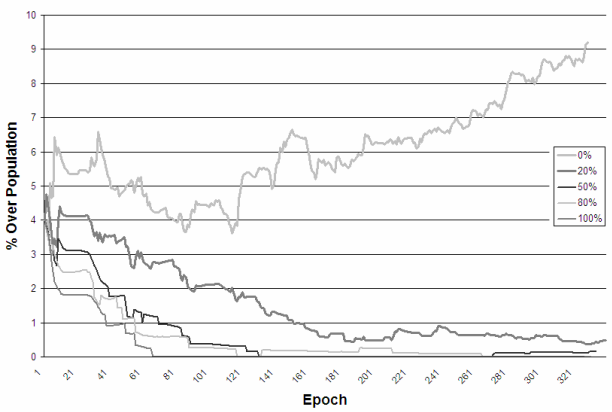


**Figure 7. Rate of Patients over Total Population**

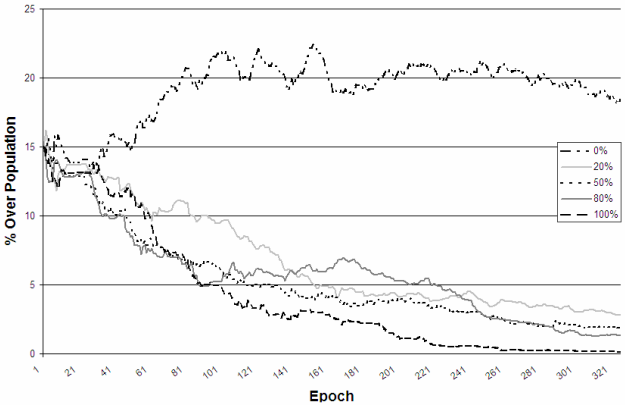


**Figure 8. Rate of Carriers and Patients over Total Population**

### 6.3.3 Trends of Trait

In order to better support our initial hypothesis, we have studied the epidemiological spread of the mutation in the genetic code of the population for different PbC values. We note that for the base case of PbC=0%, there is an exponential increase in the number of people carrying the trait, while in all the others cases the trend decreases exponentially (Figure 8).

### 6.4 Discussion on Results

Apart from the consequences hereditary diseases have from a medical perspective, an increased number of patients implies increased treatment costs for the state, as well as increased psychological pressure for the patients and their families. It is, thus, obvious that the need to efficiently treat hereditary diseases is imperative.

Statistical studies on hereditary diseases require efficient measuring of the number of the patients and the carriers for large time-spans, and this makes it very difficult to safely verify simulation results based on complex mathematical models. By the use of agent technology, however, a more straightforward depiction of reality is achieved, assisting researchers in understanding the dynamics of such multivariate system, when lacking real-world information.

The Genecity artificial societies grow exponentially, with a percentage of natural growth of 7.8 per 1000 people, according to real demographic data for some European countries. Based on the experiments conducted and the extracted results, pre-birth control can, indeed, reduce the number of people suffering from hereditary diseases. In fact, the decrease rate of the patients in the total population, based on the variation of the Pre-birth Control variable, appears to be in correlation with the actual decrease rate of real populations. The fact that our initial hypothesis is valid, provides a gateway for dealing with the problem.

Nevertheless, more parameters may be taken into account and change the scenery in pre-birth control; ethics is one of them. If abortion is an option for couples, then different societal behaviors may arise.

This brings up an interesting issue: To what degree should the screening policy and the pre-birth control be applied in a population? Results suggest that a good strategy for eliminating a hereditary disease with the minimum amount of families affected is full pre-birth control planning and screening, along with proper health education.

## 7 Conclusion and Further Work

GeneCity provides an integrated framework for modeling environments, simulating behaviors and monitoring society dynamics. It can be used as a Decision Support System for researchers working in related fields, to better understand the complex nature of the epidemiological concrescences of a hereditary disease. It combines primitives drawn from the fields of medicine, biology, and machine learning, in order to produce a reliable, well-built simulation tool. It provides a user-friendly console for tuning and re-tuning the various societies developed. Through this paper we have attempted an overall description of the tool. GeneCity functionalities were discussed, while a test-hypothesis based on Thalassaemia major was provided in order to prove the added-value of the framework developed.

Future research efforts include the development of data analysis tools to be implemented within the GeneCity framework, the creation of different cultural groups within the population, the concurrent existence of various mutations of the disease, and the treating of more complicated, contagious diseases.

## References


[1] R. M. Axelrod. *The evolution of cooperation*. Basic Books, New York, 1984.

[2] W. Banzhaf and F. H. Eeckman, editors. *Evolution and Biocomputation, Computational Models of Evolution: Evolution and Biocomputation*, volume 899 of *Lecture Notes in Computer Science*. Springer, 1995.

[3] F. Bellifemine, A. Poggi, and R. Rimassa. Developing multiagent systems with JADE. *Lecture Notes in Computer Science*, 1986:89–101, 2001.

[4] D. M. Buss. Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, 12:1–49, 1989.

[5] J. M. Connor and M. A. Ferguson-Smith. *Essential medical genetics*. 3rd ed., Oxford; Boston, Blackwell Scientific, c1991, 1991.

[6] A. Eleftheriou. *About Thalassaemia*, volume 4. Thalassaemia International Federation Publications, 2003.

[7] J. Epstein, D. A. T. Cummings, S. Chakravarty, R. M. Singa, and D. S. Burke. Toward a containment strategy for smallpox bioterror: An individual-based computational approach. CSED Working Paper No. 31, December 2002.

[8] J. M. Epstein and R. Axtell. *Growing Artificial Societies: Social Sciences from the Bottom Up*. MIT Press, 1996.

[9] J. Ferber. *An Introduction to Distributed Artificial Intelligence*. Addison-Wesley, 1999.

[10] T. F. Foundations. Foundation for intelligent physical agents specifications. Technical report, The FIPA Consortium, 2003.

[11] A. Giddens. *Sociology*. Polity Press/Blackwell Publisher, 4th edition, 2001.

[12] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, USA, 1989.

[13] C. Goldspink. Modelling social systems as complex: Towards a social simulation meta-model. *Journal of Artificial Societies and Social Simulation*, 3(2), 2000.

[14] W. Grosso, H. Eriksson, R. Fergerson, G. Gennari, S. Tu, and M. Musen. Knowledge modeling at the millennium (the design and evolution of Protégé-2000). In *Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management*, 1999.

[15] R. Hegselmann and A. Flache. Understanding complex social dynamics: A plea for cellular automata based modelling. *Journal of Artificial Societies and Social Simulation*, 1(3), 1998.

[16] E. Hillbrand and J. Stender, editors. *Many-Agent simulation and Artificial Life*. IOS Press, 1994.

[17] J. H. Holland. Genetic algorithms and classifier systems: Foundations and future directions. In J. J. Grefenstette, editor, *Proceedings of the second international conference on genetic algorithms and their applications*, pages 82–89. Ed. Hillsdale, New Jersey: Lawrence Erlbraum Associates, 1987.

[18] P. Hraber, T. Jones, and S. Forrest. The ecology of echo. *Artificial Life*, 3(3):165–190, 1997.

[19] C.-Y. Huang, C.-T. Sun, J.-L. Hsieh, and H. Lin. Simulating sars: Small-world epidemiological modeling and public health policy assessments. *Journal of Artificial Societies and Social Simulation*, 7(4), 2004.

[20] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society*, A(115):700–721, 1927.

[21] N. Kyriazis. *Population Reproduction: Theoretical Approaches and Empirical Studies (in Greek)*. Gutenberg, Athens, 1992.

[22] D. G. McVitie and L. B. Wilson. The stable marriage problem. *Management Science/OR*, 14(7):486–490, 1971.

[23] N. Minar, R. Burkhart, C. Langton, and M. Askenazi. The swarm simulation system, a toolkit for building multi-agent simulations, 1996.

[24] M. Parker. What is ascape and why should you care? *Journal of Artificial Societies and Social Simulation*, 4(1), January 2001.

[25] WHO. Primary health care approaches for prevention and control of congenital and genetic disorders. Report of a who meeting, cairo, egypt, 6-8 december 1999, World Health Organization, 2000.

[26] P. Zak and K. Park. Population genetics and economic growth. URL http://ssrn.com/abstract=239990, August 2000.