# Socio-Spatial segregation in a caste based system: Application of Schelling's model of segregation to regional caste structures in India.

Shamil Khedgikar (ssk274@cornell.edu)

## Abstract:

Given the theory of dynamic segregation developed by Thomas Schelling to understand the underlying mechanisms of why segregation of any type exists in the society and is emergent in nature, this paper tries to apply the theory to the caste structures in a typical Indian community by creating an agent-based model. The highlight of the paper is the creation of 'conscious' agents, whose preferences change over time as a consequence of exposure to education, electronic media, migration and mutual interaction. Besides exposure, the agents also undergo aging, marriage and reproduction. These features allow us to understand the spatial dynamics of evolution of a closed, rural, caste based community into an urban, homogenous community over a time frame of 100 years. The model has been developed as a means to pursue the three questions; first, given an initial historic spatial distribution of segregation, how can Schelling's segregation model be applied to understand the change in this distribution over a period of time? Further, what are the most significant factors that cause change in the spatial pattern of segregation and how can these be captured quantitatively through an agent-based model? Finally, how can this agent based model be to be applied to peri-urban regions that are bound to undergo a redistribution due to urbanization and change of land ownership.

The region selected for analyzing the model is a part of the Coimbatore District region in the state of Tamil Nadu, India. There are two reasons for selecting the region for analysis. Firstly, a primary survey has been undertaken in this area well before urbanization spread across the region (Beck, 1976). Secondly, this area has a distinct geography which makes it accessible only through a fixed number of entry points within the area.

The methodology we undertake for the analysis follows from Thomas Schelling's model of segregation and its implementation undertaken as a part of the course. Firstly, the concept of preferences is modified to include selective preferences of caste systems based on inferences of (Hoff & Pandey, 2004) These selective preferences are then made to evolve over a period of time using simple mathematical functions. We record the observations of each of the functions and allow the model user to modify the parameters as per their choice. The environmental interaction of agents with the environment helps us to capture the effects of various phenomena such as accessibility to education, electronic media etc. Lastly, we find the density of emergent segregation patterns and try to find the overall probability of such patterns being generated across various regions in India.

# Introduction

## The Indian Caste System

Social segregation is consequence of power relationships and hierarchical structure of any given society. Segregation usually occurs on the basis of some intrinsic identities of individuals such as race, sex, religion, caste etc. However, spatial segregation is an emergent phenomenon which is a consequence of slight preferences at the individuals in associating with members from a similar or different social identity (Schelling, 1971).

The system of social segregation is highly dependent upon the nature of a society. The system of segregation is usually established and governed by those in power by virtue of a unique attributes that associates each individual within a ‘superior’ population and distinguishes them from the ‘inferior’ population. The power relationships that govern social segregation may be established through a religious doctrine or emerge as a consequence of normative practices. While normative practices evolve as societies progress and this may lead to homogenous society, it is segregation established through a religious doctrine that is difficult to eliminate. One such form of segregation is the *Varna* system. Established through rigorous rules as described in the Hindu Dharma-Shastras, this division of segregation has been an intrinsic part of the Indian society since the Vedic times. *Varna* is a Sanskrit term derived from the root word ‘*vṛ*’, meaning "to cover, to envelop, count, classify consider, describe or choose." According to the *Varna* system, the society is classified into four sections on the basis of division of labor. Varna contextually means "color, race, tribe, species, kind, sort, nature, character, quality, property" of an object or people in some Vedic and medieval texts. Each section of the society is assigned certain duties that they are entitled to perform over their life. The illustration below shows the heirarchial nature of the four *Varnas*; Brahmins, Kshatriyas, Vaishyas and Shudras.

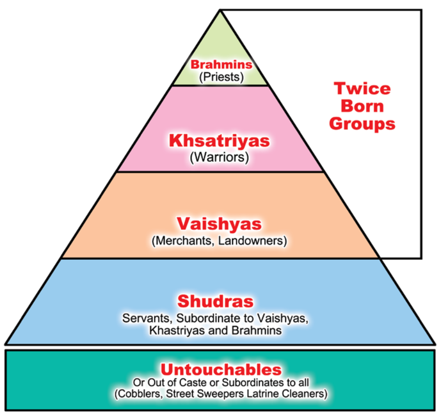



*Figure 1: The hierarchical Varna system*

Although the *Varna* system was established to ensure equal social division of labor, the Twice Born Groups i.e the priests, warriors, merchants and landowners were endowed with a greater social status through the doctrine. As the traditional Indian Hindu society grew, the *Varna system* lost its significance, to be replaced by a caste system or *Jati* system, that was hierarchical and dependent on occupation and lineage. This *jati* system evolved out of the *Varna* system and became a much stronger social hierarchy, percolating through religion and influence every aspect of an individual’s life. However, the hierarchical nature of the society persisted; it led to the severe oppression of the low castes by members of the higher castes. This caste system was abolished constitutionally by the government of India upon independence of the country in 1947; however, the relevance of caste was not affected due to a legislation passed by the Indian state. Being a predominantly agrarian society, the repercussions of the caste system manifested in the ownership of land, cattle and mutual interaction. The wealthier castes usually owned large amounts of land and the lower castes were voluntarily or forcefully made to work on these lands as laborers.

The Indian constitution re-classified the social hierarchy of the caste system by creating the Scheduled Castes (SC), Scheduled Tribes(ST) and Other Backward Classes(OBC) in order to implement affirmative action legislations and curb the oppression on lower caste communities in India. The effect of this rescheduling also enabled the Indian government to keep track of the demographic proportions of caste in India.

As per the 2011 Socio Economic Census of the Union of India, the distribution of castes across various religions is as follows:

*Table 1:Distribution of Population of each Religion by Caste Categories; (Source: http://secc.gov.in/)*

| **Religion/Caste** | **SCs** | **STs** | **OBCs** | **Forward Caste/Others** |
|---|---|---|---|---|
| Hinduism | 22.2% | 9% | 42.8% | 26% |
| Islam | 0.8% | 0.5% | 39.2% | 59.5% |
| Christianity | 9.0% | 32.8% | 24.8% | 33.3% |
| Sikhism | 30.7% | 0.9% | 22.4% | 46.1% |
| Jainism | 0.0% | 2.6% | 3.0% | 94.3% |
| Buddhism | 89.5% | 7.4% | 0.4% | 2.7% |
| Zoroastrianism | 0.0% | 15.9% | 13.7% | 70.4% |
| Others | 2.6% | 82.5% | 6.25 | 8.7% |
| **Total** | **19.7%** | **8.5%** | **41.1%** | **30.8%** |

The figures show the broader percentages of population by caste in the Indian society. However, this is the broad distribution of caste across rural and urban areas. The caste system has been deeply rooted in the Indian society; caste discrimination is still prevalent and practiced as a custom across many rural areas within the country. The untouchables and the lower castes (castes emerging from non-twice born groups) were restricted access from the earliest settlements of the Indian subcontinent (Mandelbaum, 1991). These restrictions extended from the allocation of a separate section of a settlement for housing of the lower caste individuals to not allowing the lower castes entry into temples and certain parts of a settlements. As a result, most lower caste communities settled at the periphery of rural settlements. These patterns of growth have been studied in (Beck, 1976) pertaining to the region surrounding the city of Coimbatore in India.

It is necessary to understand that the caste system is a deeply rooted normative system of social segregation in India. Individuals of higher castes attempt to maintain the 'purity' of their caste by marrying within their own castes. The ownership of land and practices of food exchange are dictated by the caste system.

## Effects of Urbanization

The effects of urbanization on rural settlements causes a change in the social, economic and political structure of settlements. The social change is characterized by creation of a 'homogenous' population; as a consequence of urbanization, the economic class becomes a more dominant normative basis for segregation rather than social caste. Throughout urban settlements, we start seeing pockets of population that have similar economic status as opposed to similar castes. The underlying mechanisms of this change are yet to be investigated in greater detail. A statistical analysis of caste based socio-spatial segregation, however, is presented in (Dupont, 2004). The second effect of urbanization is migration. The effect of migration due to urbanization has been studied in great detail throughout literature. The first study that captured this nuanced consequence of urbanization was the Harris and Todaro model, developed in 1970 (Harris & Todaro, 1970). This effect is characterized by migration of population from rural areas to an urban node. The last significant effect of urbanization is the exposure to new technology. As new digital telecommunication technologies make their way into rural markets, the connectivity of individuals to the rest of the society is greatly enhanced. This results in the creation of a 'network' link with the existing networks of society which is characterized by the increased awareness of agents as a result of exposure to the external world. This is a characteristic observation of most empirical research that has been undertaken by Manuel Castells in his most prolific work; The Information Age Trilogy.

The next section of the paper provides an introduction to the study area and a summary of key features and data that is relevant to the model.

## Study Area

The study area for this model is based in the region surrounding the city of Coimbatore in India. Coimbatore, is a major city in the Indian state of Tamil Nadu. Located on the banks of the Noyyal River surrounded by the Western Ghats, it is the second largest city in the state after Chennai and the sixteenth largest urban agglomeration in India. It is administered by the Coimbatore Municipal Corporation and is the administrative capital of Coimbatore district. It is one of the fastest growing tier-II cities in India and a major hub for textiles, industries, commerce, education, information technology, healthcare and manufacturing in Tamil Nadu. It is often referred to as the "Manchester of South India" due to its cotton production and textile industries. Coimbatore is also referred to as the "Pump City" and it supplies nearly half of India's requirements of motors and pumps. The city is one of the largest exporters of jewelry, wet grinders, poultry and auto components.

The population of Coimbatore as per the 201 Indian census was 1,601,438. The urban boundaries of the area represent a total area of 247.92 square miles. The city and surrounding region of Coimbatore has undergone rapid urbanization in the recent years. Urbanization in Coimbatore has been a consequence of increased connectivity through National and State Highways to the surrounding rural areas as well as increased investments in the infrastructure, textile and automobile sectors. The creation of TIDEL Special Economic Zone in the industrial estates surrounding the city has led to an increased migration of skilled and unskilled workers from the surrounding region. This in turn has led to the growth of a vibrant and healthy retail market within the city. Young professionals migrating to the city enjoy the career opportunities and the mild climate that the city has to offer. The city of Tamil Nadu is governed by the Coimbatore Municipal Corporation.

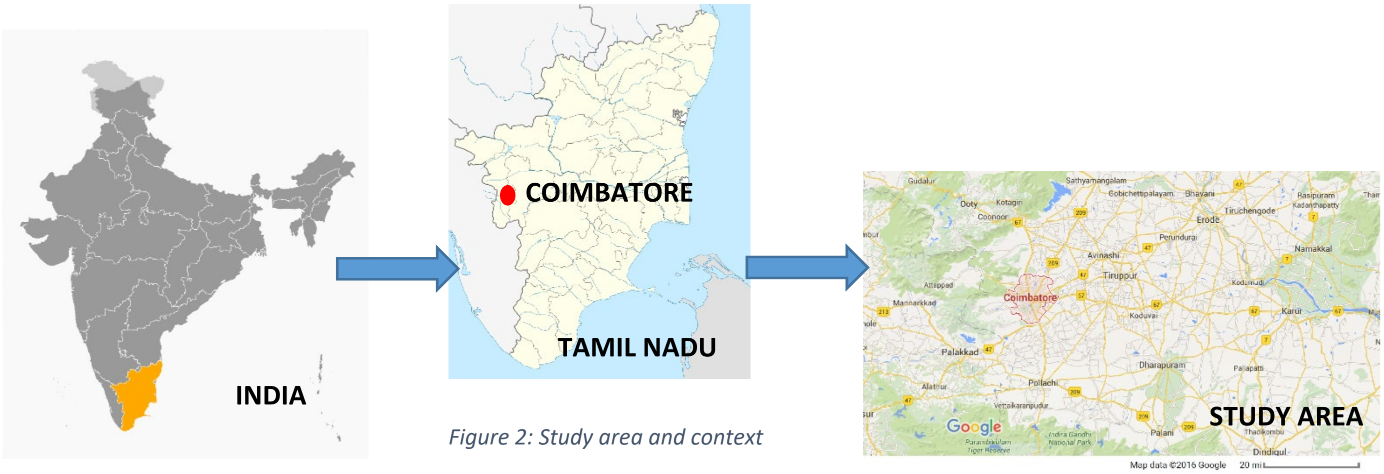



*Figure 2: Study area and context*

The above pictogram represents the study area for the model. The region surrounding Coimbatore was first studied by anthropologist Brenda E.F Beck (Beck, 1976) in the mid-70s. The salient features of this study area are as follows –

- Isolated settlement surrounded by high hills on all sides
- Existence of a dominant caste in the major settlement
- Peripheral settlements of lower castes around the major settlement in the region

The reason for choosing the study area is as follows: The initial historic spatial distribution comes from (Beck, 1976). Although regional caste systems in India have been studied extensively by Anthropologists from the 1960s and 1970s, only a few cases have tried to formulate theoretical models through empirical research on the spatial distribution of a community based on caste. (Beck, 1976) not only analyzes spatial distribution of caste based systems but also defines the classifications of spatial distribution theoretically. This implementation of this model requires an area that has undergone urbanization in a single node to understand the causal nature of changing preferences in patterns of segregation. The study area satisfies all the conditions and also provides us with a relative benchmark to assess the questions that have been considered. The delineated study area that we would be using for our model, adapted from the general spatial model of (Beck, 1976) is shown on the following page.

A, B, C and D represent the areas of dominant populations of the castes surrounding the region of Coimbatore. The numbers on the map denote the sampling points that have been used to delineate the study area which we shall adopt in our model. The red dot approximately represents the urban extent of the City of Coimbatore when the study was first undertaken. The consequent rings show the growth around the urban core due to industrialization, suburbanization and migration. The green ring surrounding the city show the approximate urban boundaries in the early 1990s (Census 1991). Finally, the black ring surrounding the green city represents the present extents urban boundaries of the area.

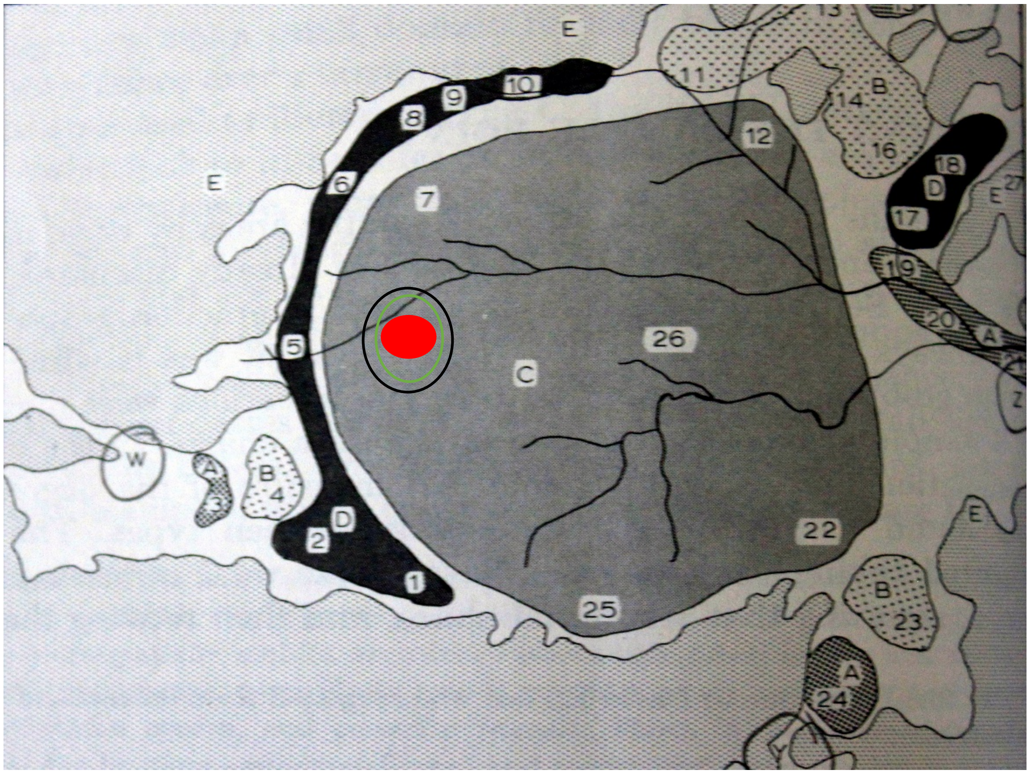



*Figure 3: Study area as appears in (Beck 1976) and the urban growth of Coimbatore*
*Red dot – Settlement area surrounding Coimbatore(1976); Green ring - Urban boundaries in the early 1990s;*
*Black Ring – Current city extents based on Google Maps.*

There are three major questions that this paper tries to address in the context of the entire region shown above study area. These are discussed in the next section along with the hypothesis of the paper.

## Questions

There are three major questions that this paper tries to address. These are as follows:

- **Given an initial historic spatial distribution of segregation, how can Schelling's model be successful in applied to understand the change in this distribution?**

Schelling's model is successful at explaining how the underlying choices pertaining to the tolerance of different individuals towards any common social, economic or political trait or feature (race or caste, income level, political affiliation etc.) leads to an emergent segregation amongst like demographic or economic activity groups. The simplest and the most dynamic means of visualizing this process is through the creation of an agent-based model.

The motivation of using an Agent Based model is a basic dissatisfaction with rational agents; there are several limitations of top-down models and the lack of a mathematical framework for solving real life social, economic and political problems has driven the research fraternity to seek out solutions through the paradigm of Complexity Theory which can be exploited through the creation of Agent Based models.

Coming back to the relevant question, we try to explore the emergence of such social patterns of segregation by applying the concept to the caste system. Any change in distribution of spatial patterns at the regional level would be characterized by the process of individuals adjusting their location in the context of space. However, we can see that the given region has a unique pattern of segregation that is a consequence of land ownership, food exchange hierarchies and the governing rules which have mostly been dictated through a religious doctrine. Thus, we need to first generate spatially relevant 'model space' in which we can reimagine the given pattern of segregation. We then need to recognize that the preferences of individuals are rather different from the Schelling model and is dictated by a social hierarchy existing in the area. Once we understand the relationships between each section of the caste in the surrounding region, we would be able to predict and simulate the changes through the use of an agent based model which is integrated with a GIS database for accurate spatial analysis

- **What are the most predominant factors which cause change in the spatial patterns of segregation and how can these be captured quantitatively through an agent-based model?**

One of the most distinct feature of agent based models is that they are very useful in situations where the state of equilibrium of a system is not defined (Axtell, 2000). We have already identified the effects of urbanization in the earlier section of the paper. We choose to focus on three major effects of urbanization that would cause a change in the spatial patterns of segregation. These are:

1. **Homogenization of the urban core**: As a consequence of urbanization, the core area of an urban settlement is likely to become homogenous. The urban core of a single settlement in a region represents the largest node of economic activity. However, as we are dealing with an economy that is based in the developing world, we need to be wary of the fact that manufacturing and services are the two primary types of economic activities that are the characteristic features of an urban area. The secondary nodes are thus the industrial estates, special economic zone and technology parks that usually get developed in the outskirts of an urban settlement. These primary and secondary nodes can be fairly considered as areas of homogenous population groups, where the relevance of social status dependent on caste is replaced by the economic class of the masses. In the case of our study area, Coimbatore is the urban settlement where we expect homogenization to occur and spatial patterns of segregation to change.

2. **Migration in the region:** Migration of population into a region is facilitated by accessibility. Given the geography of the study area, we know that travel into the urban region is possible only through road and railways. Further, (Beck, 1976) identifies four distinct points of entry that can be considered as the sources of incoming population into the region. Using these sources to analyze the impact of incoming population on existing segregation patterns within the region.

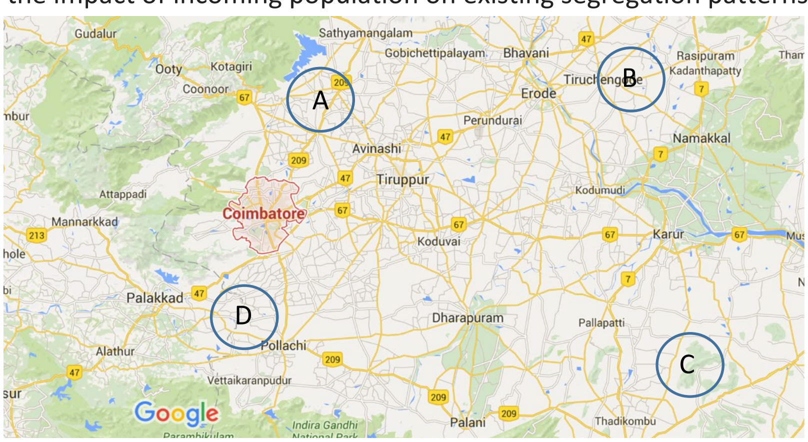



*Figure 4: Nodes of entry into the study*

Further, if we try to study the road and rail network of the region, we see that the four points (A, B, C and D) are the main nodes through which migrating population enters the study area.

3. **Creation of Technology Networks:** We borrow the 'networking logic' relationship underlying Manuel Castells' research (Castells, 1996) to quantitatively analyze the impact of exposure to digital technologies. Exposure to technology creates new networks that are manifested in space. Urban settlements are the nodes of such new networks and are responsible for strengthening the relationships between settlements and the rest of the world. These urban settlements become what Castells calls as the 'Space of Flows.' We can thus expect two changes in the segregation pattern that exists in our study area. Firstly, we can expect stronger links being formed between smaller settlements that have the potential to urbanize. Secondly, we can expect the tolerance level of individuals to increase with respect to populations who are of a different caste, race etc. as more and more people are connected to the entire world as a consequence of digital technologies.

- **How can one modify the primary agent-based model to be applied to peri-urban regions that are bound to undergo a redistribution due to urbanization and change of land ownership?**

The primary agent based model here refers to the simplest form of model that creates segregation patterns between agents of similar color depending upon the threshold of each individual. In order to apply this agent-based model to peri-urban regions such as our study area, we need to work across two layers of the model. The first step is creating a GIS dataset that captures the minimum necessary features of the built environment. We utilize existing data on built up, sprawl and road networks to create the dataset. Then, we overlay relevant parameters of the three predominant factors discussed above to impose restrictions on the interaction of agents. This is followed by understanding the nature of land ownership, which informs the division of land parcels within the study area. We then utilize the demographic data on caste population to generate agents in proportions to match the ownership levels of each agent is associated with a parcel. After observing the simulation, we are able to generalize a framework that helps us apply this model to another area. We accomplish this by creating a general methodology for implementation of the model.

The rationale for the creation of such a model has been established and we have a broad understanding of how each of these questions would be answered. However, before we proceed into the details of the model, it is important to step out of the modelling system to justify the creation of such a model and understand the policy relevance of this model. This would also help us in narrowing down on the specific research articles and models that need to be reviewed to enhance the creation of the proposed model in the most effective manner.

## Justification

The creation of this model is justified for the following reasons:

- Numerous agent based models have been developed to understand segregation that occurs in urban areas; residential areas in particular. However, analysis of segregation patterns at the regional scale have not been extensively studied. Regional agent based models try to focus mostly on issues of economic relevance. Thus, the relevant questions that we are seeking answers to are justified from the point of view of enriching the literature on agent-based models.
- There does not exist any study that tries to apply Schelling's methodology to the segregation that exists as a consequence of a caste system. Although numerous qualitative and quantitative research has been undertaken on studying the implications of caste based segregation (Hoff & Pandey, 2004) (Dupont, 2004), agent based model is a promising well tested methodology that can help inform our observations and recommendations to curb caste based segregation in urbanizing communities.

## Policy Relevance

India is currently the fastest growing major economy in the world. Towards the path to becoming one of the developed countries in the next twenty years, one of the biggest roadblock that the Indian society faces is the overlapping social and economic disparity. This overlap has been a consequence of the long standing hierarchical structure of the society which allowed the socially affluent members of the society to become wealthier while those belonging to the lower castes were left isolated from access to infrastructure and the influence of urbanization. This overlap was further strengthened as a consequence of the wave of urbanization during the early days of post-liberalization of the Indian economy (1993-1999). The young wealthy members of high castes in communities from the rural areas migrated into the urban areas, while the older members of the higher castes maintained their stronghold on the land ownership. As a result, the poor, lower caste community members of the society were left out from contributing to the formal economy and this has led to a low workforce participation rate in most of the Indian states.

(Todaro & Smith, 2015) first laid down the three factors that contribute to economic growth in developing countries. These are;

1. Capital Accumulation
2. Workforce Growth
3. Technological advances

Urban areas are the nodes within a country where capital is accumulated. Workforce growth in urban areas is a consequence of prospects of better economic opportunities that urban areas have to offer. It is not possible to sustain the supply migrating populations with jobs consistently. As a result, informality grows in urban economies in developing countries. (Dupont, 2004) shows that most informal settlement typologies in urban agglomerations have a higher proportions of people who belong to the lower sections of the caste hierarchy. Further, since most technological advances and workforce growth is manifested in urban areas, the population in rural agrarian parts of the country cannot effectively contribute the growth of the economy. While urban areas grow and prosper, diversity in these areas reduces segregation as people have more exposure and consequently more tolerance. On the other hand, rural societies remain influenced by traditional doctrines and rules and thus fail to utilize the complete potential of the workforce available across all sections of the society.

The proposed model is only a first step towards gaining insights into the dynamics of caste based segregations in Indian societies. In order to successfully develop a methodology for the proposed model, it is important to identify broad themes of research that is encompassed in the paper. In the next section, we review three peer-reviewed papers that help in the clear framing of the methodology for the model.

## Previous Studies

The three papers that have been reviewed and their broad themes of research are as follows:

1. **Constructing and implementing an agent-based model of residential segregation through vector GIS – A.T.Crooks**
   ***International Journal of Geographical Information Science*. Vol. 24, No. 5, May 2010, 661–675**
2. **Modelling the impacts of urban upgrading on population dynamics – Schwarz et.al**
   ***Environmental Modelling & Software* Vol. No. 78, August 2016 150-162**
3. **Combining Segregation and Integration: Schelling Model Dynamics for Heterogeneous Population – Erez Hatna and Itzhak Benenson; *Journal of Artificial Societies and Social Simulation* Vol 18. October 2015**

### GIS and Agent Based Modelling

1. **Constructing and implementing an agent-based model of residential segregation through vector GIS – A.T. Crooks**

This article summarizes the methodology for creating a geographically explicit agent-based model that is coupled with vector GIS datasets, which allows capturing of the geometric features and the socio-economic characteristic of geographies. Further, Schelling's theory of dynamic segregation (Schelling, 1971) is applied to the geographic dataset through the use of an agent based model. The extensions suggested in the model are of relevance for our study.

The critical aspect that creates the rationale for the paper is that most agent based models(ABM) that have been created utilizing GIS datasets are rooted in grid-based structures. Thus, while such ABMs provide valuable insights into geographic details, the geometric details of the environment are disregarded. The ability to represent an urban system as a series of spatial objects – spatially referenced relevant points, lines and polygons help in overcoming a very significant limitation of grid based ABMS; the unidimensional nature of analysis that disregards the built form of residential neighborhoods in most countries. Further, using explicit geometries allow the author to gain further insights into mechanisms of segregation as spatial analysis queries such as unions, intersections, buffer analysis of these geometries can be undertaken. Lastly, the 'tips' of genesis and exodus as defined by Schelling in his seminal paper are clearly understood not only at the block level, but also with respect to the transitions that occur at the borders of two distinct neighborhoods. The basic model structure can be understood through the following diagram which is part of the original publication.

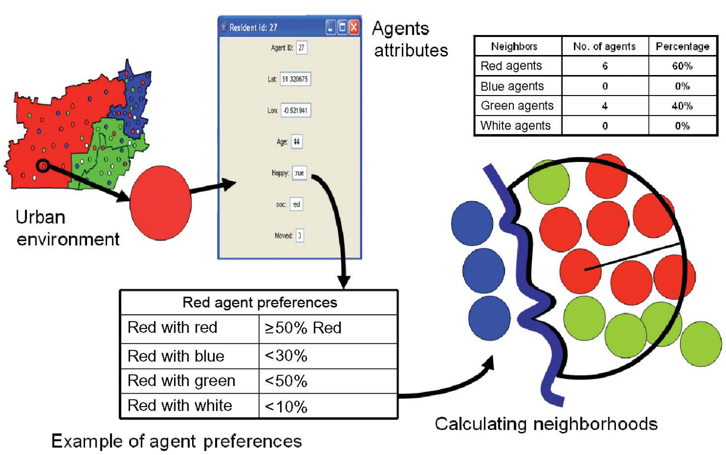



*Figure 5: Model Structure: Integration of Geometric datasets and agent based models*

The urban environment of the model is comprised of delineated neighborhoods which are initially comprised of agents with varied preferences. The color of the enclosed geometry of a neighborhood (could be a borough, block etc.) helps us visualize how notions of 'predominantly white' or 'predominantly black' neighborhoods are generated over the course of migration of individuals or household units from and to different neighborhoods.

The next stage of analysis undertaken in the paper is understanding the quantitative and temporal dimension of the phenomenon of 'emergence' simultaneously. Emergence is a process whereby larger entities, patterns, and regularities arise through interactions among smaller or simpler entities that themselves do not exhibit such properties. The segregation patterns that are being studied are emergent in different resolutions as per the author. This means that the level of segregation patterns observed is dependent on the scale at which the system is being viewed or the scale at which the model is operating. At lower scales, the pattern of segregation is more granular; however, with the increase of geographical extent, we see a larger segregation emerging.

As areas grow and decline, the effect of migrating population becomes significant. This effect of migration translates as the addition or removal of agents from the system. These impacts are the last set of observations incorporated into the paper. Adding such dynamic processes help us make the model close enough to recreate the patterns of segregation observed in real life.

The paper is an attempt to 'disaggregate' dynamics of nuanced characteristics of society that cause residential neighborhoods and settlements to exhibit patterns of segregation. In the authors words, *'the model highlights how theories and concepts pertaining to urban phenomena can easily be abstracted within geographically explicit ABMs, helping further our understanding of how processes within cities operate and thus raising the importance of incorporating space and geometry when modeling urban systems.'*

## Cellular Automata and Agent Based Modelling

**2. Modelling the impacts of urban upgrading on population dynamics – Nina Schwarz, Johannes Flacke, Richard Sliuzas**

The purpose of reviewing this paper has been to get a comparative understanding of the other most popular technique utilized in developing agent based models of spatial relevance; Cellular Automata. Further, since the model is based on understanding the dynamics of urban upgradation within cities of the 'Global South', the paper becomes even more relevant to the proposed model in the paper.

In strictly mathematical terms, cellular automata refer to one of a set of units in a mathematical model that have simple rules governing their replication and destruction. They are used to model complex systems composed of simple units such as living things or parallel processors (Batty, 2005).

The papers objective is to study the effects of upgraded infrastructure in urban areas, the scope of upgraded infrastructure and the temporal nature of investment on the quality of infrastructure, housing development and income segregation in a typical urban settlement. The most striking feature of this paper is the depth of decision making capacities that have been programmed into the agents to understand the emergent perception of the urban infrastructure and the effects of the upgradation on the agents themselves. Such a process of treating agents as evolving decision makers utilizes the concept of backpropagation within a network that represents the decision making process.

The name of the system that has been developed in the paper is known InformalCity. It is a city that contains 49 districts that are spatially located in a grid structure. The housing typology of neighborhoods defined in the model is the simplest form of one plot, one house neighborhood. The center of the city, the Centroid of the cellular automatons is the central business district of the city.

At the beginning of the simulation, the city is initialized with agents representing households, being either tenants or owners. These households settle on a plot in one of the districts and can move to other plots in other districts in consecutive time steps but do not move within a district. Furthermore, owners can add rooms to their house, and tenants can become owners provided enough they have enough savings. When starting the simulation, users can set a number of parameters. These parameters include the calibration: the initial number of households, the number of plots per district, the cost to build three rooms of a house, and the population density threshold for the decrease of infrastructure quality. In addition, users can choose if they want to include urban upgrading in the simulation and can select between various options: the timing of the upgrading, the selection of target districts, the scope of upgrading efforts, cost coverage and maintenance.

The decision making process of the agents is elaborate in nature and can be understood through the following diagram:

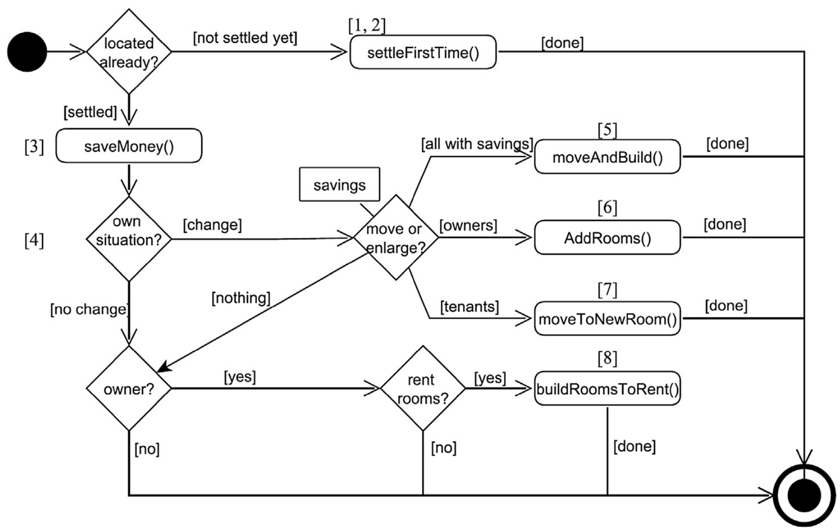



*Figure 6: Decision making of agents in InformalCity*

The key outcome of the study is that it highlights the limitations of urban upgradation schemes and the sheds light on why such schemes fail in developing countries. A real life example of such a scheme failing both fiscally as well in terms of enhancing the mobility of populations from India is the Jawaharlal Nehru Urban Renewal Mission. The central (federal) program was initiated in India in the year 2007 and its aim was based upon the idea of upgrading urban infrastructure for urban renewal in an attempt to achieve inclusive urban development. The model proves theoretically that such a policy is bound as the equilibrium condition for the 'ideal' urban upgradation program does not exist. Further, the model proves that policies that try to implement a single set of initiatives without internalizing the context of space, character of neighborhood and the opinions of individuals is bound to fail.

The last paper reviewed looks at analyzing the real life observations of patterns of segregation and integration through the help of spatial statistics.

## ABM and Spatial Statistics

4. **Combining Segregation and Integration: Schelling Model Dynamics for Heterogeneous Population – Erez Hatna and Itzhak Benenson; Journal *of Artificial Societies and Social Simulation* Vol 18. October 2015**

This paper tries to demonstrate that the variety of the Schelling model's steady segregation patterns is richer than the segregation–integration dichotomy and contains patterns that consist of segregated patches of each of the two groups, alongside areas where both groups are spatially integrated. The analysis of such patterns for the cities of New York and Chicago are undertaken by considering a general version of the model in which the mechanisms of the agents' interactions remain the same, but the tolerance threshold varies between the agents of both groups.

In order to grasp the point for which this model has been made, it is first necessary to understand the equilibrium conditions as defined in the Schelling's paper (Schelling, 1971). Through the use of a conceptual framework, Schelling was able to show that segregation occurs only if the tolerance level is above 1/3. On the other hand, the model remains integrated or in its initialization state if the tolerance level is below 1/3. This paper tries to compute threshold levels of patterns that are spatially visible through the use of remotely senses data. The pockets of relevance are chosen on the basis of a parameter known as Shannon Entropy Index. The Shannon Entropy Index is a quantitative measure that reflects how many different types of data (which refers to ethnic groups of populations) there are in a given spatial or non-spatial dataset.

The following diagram shows the comparative distributions of populations and the pockets in New York City and Chicago chosen by authors for their analysis.

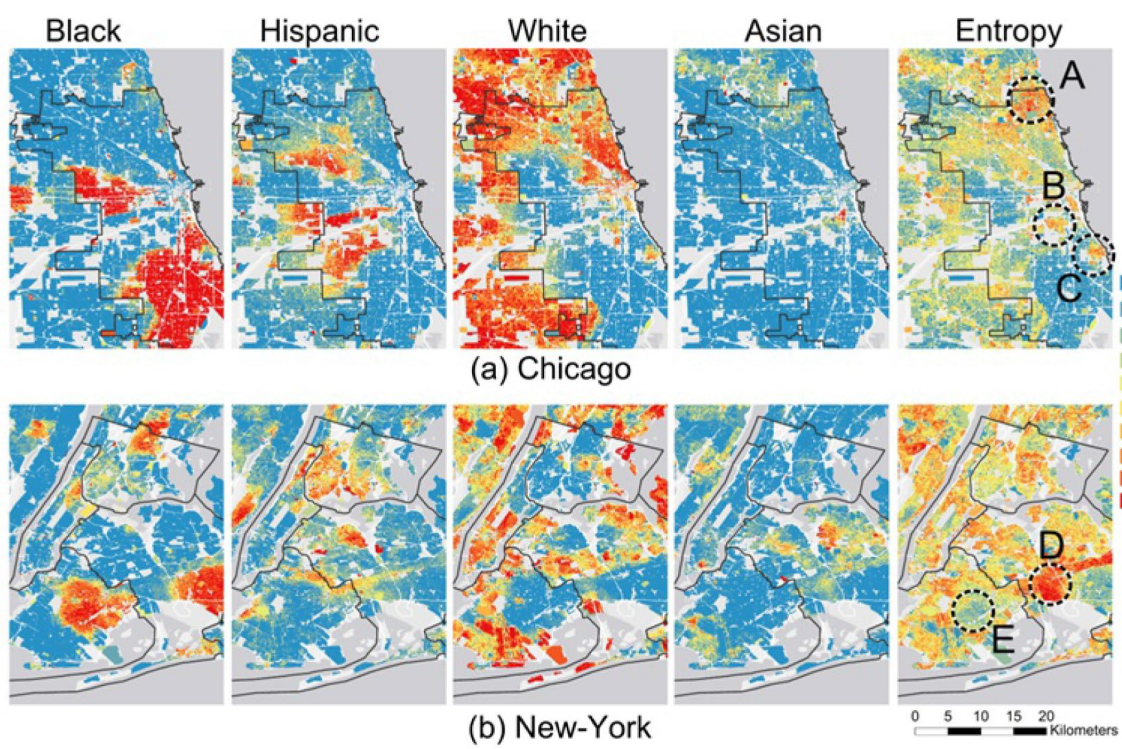



*Figure 7: Selection of Case Study pockets; Ethnic integration and segregation thresholds*

Using a well revised methodology of analysis similar to the one used by the authors in their publication titled **Schelling Model and Ethnic Residential Dynamics: Beyond the integrated – segregated dichotomy of patterns – Erez Hatna & Itzhak Benenson; Journal *of Artificial Societies and Social Simulation* Vol 15. January 2012,** the authors undertake the analysis of extracting temporal segregation patterns that spatially auto correlate with the patterns chosen for analysis. Across the five pockets (A, B, C, D and E) the authors compute the correlations for tolerance, color and the percentages of individuals of a particular ethnic group. This is followed by a sensitivity analysis that allows the authors to check the robustness of the developed model. Having successfully applied a similar model for understanding the dynamic process of segregation in the within the central cities of Yaffo and Ramle in Tel Aviv Israel for exploring spatial mechanisms of segregation within groups of A*rab Muslims, Arab Christians and Jews,* the sensitivity analysis of the paper proves that the model is robust. Finally, the authors are able to show that in the case of four ethnic groups studied across two cities, patterns of mixed integration and segregation occurs at tolerance levels below 1/5 and above 2/3.

Through the case studies, we have been able to gain an insight into nuances of spatial modelling that would help us in the formulation of the model. After exploring three main themes, we are in a position to detail out and specify the characteristics of the model using a pseudo-code an ODD framework.

# Algorithm

The following pseudocode will help in the implementation of the program that will be created in the NETLogo integrated programming development platform.

**Note on NETLogo:**
NetLogo is an integrated development platform that was created by Uri Wilensky. It was created in the spirit of "low threshold and no ceiling" to facilitate learning and creation of Agent Based Models.

## Pseudo Code

**Procedure: GIS Extension**

Import shapefile1 – Brahmin and Kshatriya Study Area
Import shapefile2 – Vaishya Study Area
Import shapefile3 – Shudra Study Area
Import shapefile4 – Untouchables Study Area

**Procedure: Setup**

GIS Dataset: Use Shapefiles
Create n turtles
Set Shape circles
Set size default
[Brahmin Study Area : Envelope
Kshatriya Study Area : Envelope]
Assign n/2 Turtles **(50%)**
Set Color Blue
Vaishyas Study Area : Envelope
Assign n/5 Turtles **(20%)**

Shudra Study Area: Envelope
Assign n/10 Turtles **(10%)**
Untouchables Study Area: Envelope
Assign n/10 Turtles **(10%)**

Set Minimum Radius – Input Variable MaxR
Set Maximum Radius – Input Variable MinR
Set Minimum Threshold – Input Variable MaxT
Set Maximum Threshold – Input Variable Min T

**Procedure: Assessing neighborhood**

Check %turtles in MaxR - Totalnbr
Check %turtles same color in MaxR – Variable Maxnbr
Check %turtle same color in MinR – Variable Minnbr
If Maxnbr - Minnbr >= (MaxT – MinT)*Totalnbr
Execute procedure Relocate

**Procedure: Relocate**
Choose empty point in study area

**Procedure: Migrate**
Create new turtles at migration (x,y) nodes
Relocate

Following this logic generally, we program the features in NETlogo.

## ODD Framework

**Purpose:**

The purpose of the model is to understand the underlying mechanisms of caste based segregation that occur due to the interaction of individuals in a typical Indian society. Based on Schelling's theory Dynamic Segregation  (Shelling, 1971) occurs, the model tries to simulates the effects of urbanization, education, electronic media, migration and on an initially caste based segregated rural community in the region of Coimbatore District in India. By creating 'conscious' agents who change their tolerance level to other agents based on their exposure to different social factors, the model simulates the growth of a population through aging, marriage and death of agents. This helps us understand how the power of decision making can affect segregation patterns as a geographical region is urbanized.

**Entities, State Variables and Scales:**

There are a total of 700 entities at initialization in the model. These entities represent populations of various castes that are part of the are defined based on the following state variables. These are:

- **Age** – Randomly distributed; The lowest age is 1, highest being 70.
- **Caste factor** – Four castes based on the social hierarchy; Brahmins, Kshatriyas, Vaishyas and Shudras.

- **Tolerance** – Each individual has a tolerance level that is determined within a range defined by the user, the number of surrounding entities of a similar caste and the changes in the environment.

***Purpose***

The purpose of the model is to understand how spatial segregation across a particular scale is not necessarily a consequence of individual preferences but rather an emergent phenomenon. The model was first hypothesized by Nobel-prize winning game theorist Thomas Schelling and has since been used to describe how segregation across social and economic and physical classifications. This model is an adaptation of Schelling's theory and has been created to develop a better understanding of the scientific principles underlying segregation across two different members of a population.

***Entities, State Variables and Scale***

Four types of agents, differentiated on the basis of their color, representing caste are bound in a spatial environment delineated by the use of a GIS dataset (point data) that represents the study area within the city of Coimbatore in the model. The properties which are inherent to agents and defined uniformly across all agents include variable sensing radius or in-radius set with a minimum and maximum limit as well as a threshold level that determines a percentage of agents surrounding an agent of different color will force the agent to change its position incrementally. The state variable governing the movement of the agent is uniform for all agents and can be varied for generating different typologies of segregation.

Other common state variables include the shape of the agents (square) and their size relative to their boundary as well as their respective location which is assigned randomly to the agents in this particular model.

***Process Overview and Scheduling***

The model generated creates a preliminary number of agents. Four populations of the two agent types are initially placed into random locations of a neighborhood represented by a geospatial map which is created as a 'sub-space' of the existing world with the geometry of a Torus. After placing all the agents on the map, specifically the centroids of the points in the study area that combine to create the sub-space, each cell is either occupied by an agent or is empty when the model is ready to be executed.

We now try to determine if each agent is satisfied with its current location. A satisfied agent is one that is surrounded by at least *t* percent of agents that are like itself, where *it* is the threshold specified by the observer of the model upon its setup. This threshold *t* is a predefined state variable which is constant throughout the execution of the model. This is an unrealistic assumption because in reality everyone might have a different threshold they are satisfied with. Note that the higher the threshold, the higher the likelihood the agents will not be satisfied with their current location.

For example, if for a particular blue agent, the threshold level is 0.3, t = 30%, agent X is satisfied if at least 30% of its neighbors are also Red. If fewer than 30% are Red, then the agent is not satisfied, and it will want to change its location on the map subject to a fixed size of movement. Further, the notion of neighbors here uses the concept of a sensing radius which is the distance specified by the observer originating from the center of each agent. Also, there exists an option of randomly adding more agents of a particular color during the execution of a simulation which allows us to gauge how segregation process in a society could be influenced by the influx of a particular type of population.

Now, as the model starts execution, each agent tries to maximize its happiness and it does so based upon its variables which are explicitly defined by the observer. As a consequence, depending upon the parameters defined, we might see agents of similar colors aggregating in specific locations within the boundary of a map.

***Design Concepts***

1. **Basic Principles –** The basic concept underlying this model is Schelling's theory of neighborhood segregation. The idea behind the model is that even though each individual does not try to segregate with agents of similar properties, the system as a whole end up segregating as a result of each agent's wish to maximize its utility. In this particular model, the design elements include 'setup' and 'go' buttons intended to start, pause and stop simulations. Also, there are separate buttons which allow for the addition of agents within the simulation. These buttons have been incorporated into the model through the use of standard features of the software package and hardcoding commands to the respective buttons themselves. Further, there are numerical inputs which allow for the specification of number of agents that are required to be created at the beginning of each simulation and then allow us to add more agents of a particular color while the simulation is running. In terms of variable and consistent parameters, there are minimum and maximum threshold levels and the sensing radius which are determined by the use of slider bars in the interface of the model. Finally, for being able to track the segregation process in real time, there are two graph boxes which have been setup; one that allows for the tracking of percentage of happy agents (agents which have achieved a state of localized

equilibrium) over time and one that displays the percentage of agents with same colored agents in its neighborhood defined by the sensing radius that has been specified by the observer for each agent in the model.

2. **Emergence –** The model has four important emergent results that are influenced by the parameters in consideration. The first is that in most cases, the agents will end up being segregated into separate pockets and then reach a state of equilibrium until the execution of the model is manually stopped or until more agents have not been added to the model by the observer. The second emergent consequence is that when the system is not able to attain segregation and as a result continues being in a state of inhomogeneity, with the percentage of happy agents fluctuating in a timely manner. Third type of emergent behavior involves a complete chaotic behavior of the system and in this case, there is no fixed pattern that emerges in the system. Finally, the last emergent scenario is that of no movement, in which case the generated system is in itself in an equilibrium position and none of the agents have an incentive to move. The parameters defined in the algorithm that influence these emergent behaviors are the sensing radius, the threshold maxima and minima and the number of agents generated in the system.
3. **Adaptation –** In terms of adaptation to the environment, the only property that helps determine the agents whether they should move in the next unit time is whether based on the sensing radius they have assumed, whether there are a threshold number of neighbors in its proximity at that instant or not. It is on the basis of this condition that a particular agent would be happy or unhappy. In case the agent is happy, it will continue occupying the same cell/patch. If not, it will randomly move by a specified unit within its spatial bounds only to reassess its condition. This is essentially a two-step algorithm that will allow for the continuous simulation of the model. This interaction is direct in nature as agents interact with other agents in order to determine their state in the system and proceed accordingly.
4. **Objectives –** The objective of each agent is to be happy. It is thus, out of this simplistic objective hardcoded into the behavior of each agent that segregation patterns emerge despite spatial constraints.
5. **Learning Prediction –** There aren't any predictive properties of agents that are a part of this model. It is assumed here that each agent or in real life terms each individual has the same amount of tolerance for people with different intrinsic properties such as color. However, this is only an assumption that allows for the simplistic execution of the model. The learning prediction

of the agents can at the most be considered to be *tacit,* that is it does not change over time but implicitly allows agents with similar properties to aggregate.

6. **Sensing –** In terms of sensing the environment, the agents use the concept of in – radius, which is the number of center-points of agents that lie in a range less than or equal to the radius defined by the observer. The agents are assumed to know their own color and identify other agents with different color in the specified in radius. Also, the agents are programmed to know whether they are happy or unhappy at a particular time and are assumed to know when they are in a position to move by the specified distance dependent upon their current situation.
7. **Interaction –** The interaction between agents is direct; this means that the agents interact with other agents directly and also allow for the accommodation of other agents that are introduced into the simulation while the model is actually running.
8. **Stochasticity –** The first stochastic element of the model is the creation of the model in its initial state. The location of agents is not deterministic and is generated randomly. The same logic applies to the addition of agents during the simulation of the model. In order to ensure that a newly generated agent does not have the same attributes as that of an existing agent, the concept of generating two random numbers and then adding them to the location coordinates of any existing agent is used. This ensures that each agent occupies his exclusive space in the model space and does not lead to any discrepancy when agents are determining whether they are happy or not.
9. **Collectives –** There is a collective aspect of the model that is not explicit in nature. We can see that aggregations in agents with similar colors occur as a consequence of each agent trying to maximize its happiness. This, as a feedback allows for more agents to aggregate around the same set of agents which triggers a segregation agglomeration pattern that persists until all the agents have been segregated.
10. **Observations –** We observe from the model that every system, dependent upon the number of agents, minimum and maximum sensing radius and threshold level will actually choose from one of the four well-defined states and persist in equilibrium in that state until an external intervention such as addition of more agents or restarting of the simulation are caused by the observer. This leads us to conclude that segregation is a highly dependent phenomenon which can be influenced by a multiple set of variables that govern the movement of agents in bounded space.

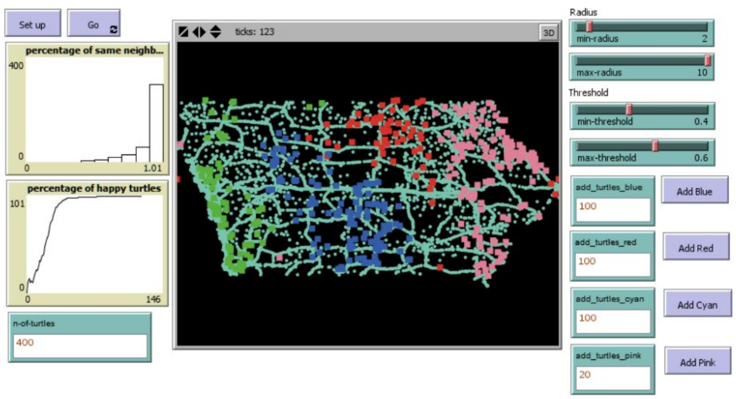
# Methodology

## Objective Presentation

In order to create a model that can help us understand the answers to the following questions, we add complexity to our model using a step by step approach.

1. **Segregation of Four types of agents**

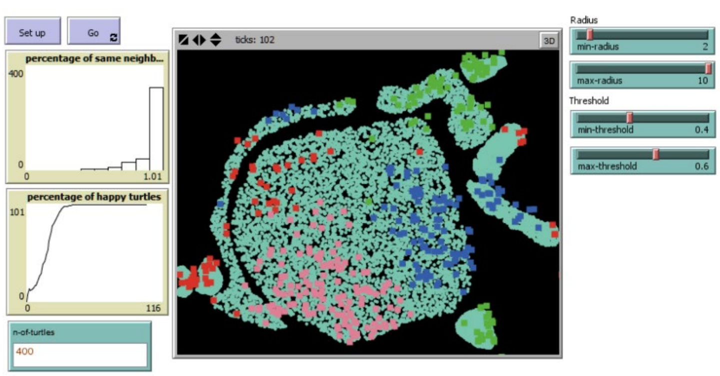
There are four castes of relevance that we are interested in studying. The typical Schelling's model uses two types of agents. We first introduce 2 additional agents into the default model. The proportions of all four populations are currently equal. Then, we use a sample GIS dataset (state map of Iowa) to simulate our results; the outcome is as follows:




2. **Incorporating dataset of Study Area**

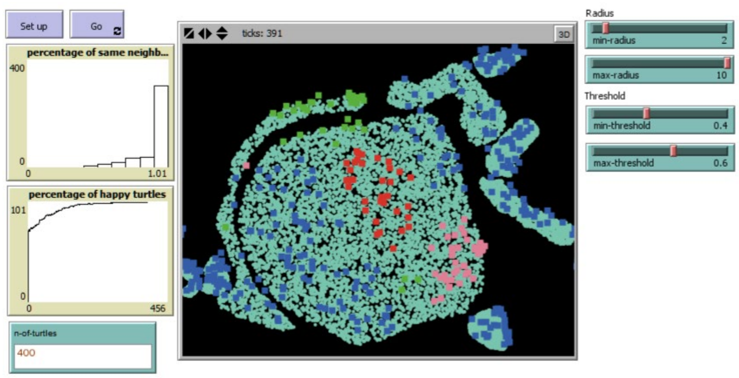
The GIS dataset that we are interested, surrounding the region of Coimbatore is imported into the NETLogo code. As we are interested in using the model-space which lies entirely within the study area, we import a dataset of points that have the shape of our study area. There are a total of eight spatial geometries, each with a uniform density of 1000 points per geometry.

Next, we run the same code to check for compatibility of our dataset. The outcome is as follows:

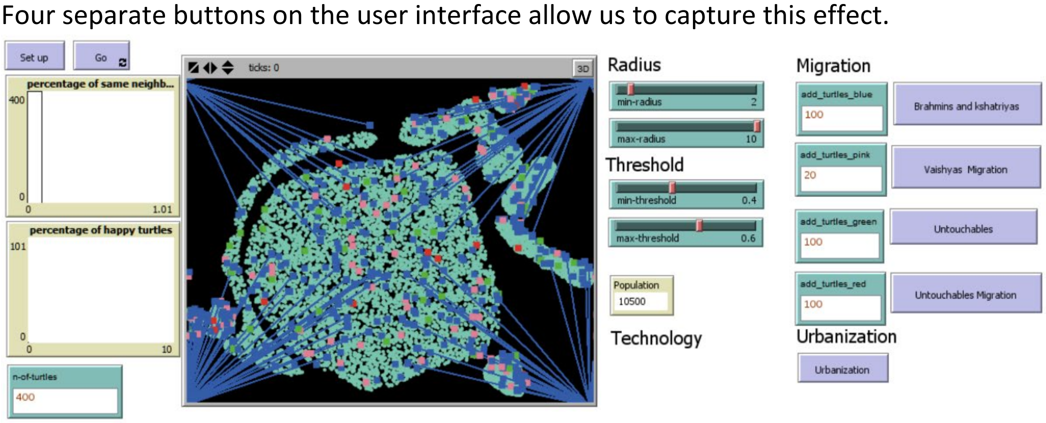



Note that the option for adding additional turtles has been removed as the feature is not of relevance to our model.

3. **Changing proportions of castes; defining representative colors.**
The color coding that we follow is as follows:
**Blue – Brahmin, Kshatriya population (50% of total population)**
**Pink – Vaishya population (20%)**
**Green – Shudra population (15%)**
**Red – Untouchables (5%)**
The proportions are adjusted and the total population of the region is set to 500. With the population changes in place, we run the simulation to get the following segregation pattern;
**Note – Because of the irregular geometry of the study area, we are not able to generate simulations with larger agents. The movement from one isolated geometry to another is entirely non deterministic as it involves generation of random residuals as x and y coordinates for the agents.**




4. **Birth, aging and death:** We now incorporate the process of birth, aging and death into the system. The birth of new agents occurs when the new agent is treated has been satisfied in his pursuits to find a neighborhood of choice. For counting the number of new turtles born, we use the hatch command.
5. **Migration:** In order to capture the effects of migration, spatially choose the reference points of entry as per the geography of the region and allow for the entry of agents into the study area. We track the patterns of migration in order to get a better understanding of their movement. Four separate buttons on the user interface allow us to capture this effect.



## Sub Models

There is a significant amount of coding that is required for the implementation of the model. The effects of segregation need to be studied carefully along with the implementation framework that is necessary for understanding how socio-spatial segregation occurs. The following section tries to convey the idea of how the effects of migration, urbanization and exposure to technology can be captured.

**Migration:** To an extent we have been able to capture the effects of migration. We see how migration of new agents can be stochastically programmed into the model. However, from a conceptual point of view, migration would lead to the new pockets of homogenous population groups to migrate to the area.

**Urbanization:** The effect of urbanization would be captured through dynamically adjusting the threshold levels in the urban core that represents homogenous space. Any movement of turtles in this space (the urban core) would stop their further movement; thus creating a homogenous urban core.

**Technology:** Effects of technology can be captured by restricting mobility of networks along specific geometries. These geometries would link the markets within the study area, represented through agents in such a manner that they can be visualized.

### Conclusion

So far we have only been partially successful in constructing an accurate model to account for the effects of phenomena on urban migration. However, we do see some interesting consequences that provide insight into the dynamics of segregation.

## References


Axtell, R., 2000. Sciences, Why Agents? On the Varied Motivations for Agent Computing in Social. *Center on Social and Economic Dynamics Working Paper No. 17.*

Batty, M., 2005. *Cities and Complexity.* 1st ed. Boston: The MIT Press.

Beck, B. E., 1976. Centers and Boundaries of Regional Caste Systems: Towards a General Model. In: C. A. Smith, ed. *Regional Analysis: Economic Systems, Volume 2.* New York: Academic Press, pp. 255-290.

Castells, M., 1996. *Rise of the Network Society.* 1st ed. Los Angeles: University of California at Berkeley Press.

Dupont, V., 2004. Socio-spatial differentiation and residential segregation in Delhi: A question of scale?. *Geoforum; Elsevier,* Volume 35, pp. 157-175.

Harris, J. & Todaro, M., 1970. Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review,* 60(1), pp. 126-142.

Hoff, K. & Pandey, P., 2004. Belief Systems and Durable Inequalities: An Experimental Investigation of Indian Caste. *World Bank Group Policy Research Paper Series.*

Mandelbaum, D. G., 1991. *Society in India.* 5th ed. Los Angeles: University of California Press .

Schelling, T., 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology,* Volume 1, pp. 143-186.

Todaro, M. & Smith, S., 2015. *Economic Development.* 6th ed. s.l.:Pearson .