Simulating Social Systems with Matlab - Residential Segregation in Zurich

by

Brechbuehl Sonia and Buechi Jonathan

Semester Project

Fall Semester 2013

Chair of Sociology, in Particular of Modeling and Simulation, ETH Zürich, 2012

Supervising Assistant: Woolley Olivia
Further Assistant: Kuhn Tobias

Simulating Social Systems with Matlab - Residential Segregation in Zurich

Brechbuehl Sonia, Buechi Jonathan

Chair of Sociology, in Particular of Modeling and Simulation, SOMS, ETHZ
Zurich, Switzerland
2013

Abstract In this report the segregational behavior of Zurich's population was studied using a more dimensional model similar to the one Schelling came up with in the sixties. An agent based simulation model was implemented in Matlab and information about origin, age and family state was provided by the statistical department of the city of Zurich. In the model, each agent was given a random site in one of the twelve districts of Zurich and then according to her/his own characteristics she/he had to choose whether she/he wanted to stay or leave the place. Tolerances towards the four Von-Neumann neighbors of each agent for all the three characteristics and towards the average of the district of the characteristics were changed in the simulations. An other parameter was the "threshold happiness" an agent must have at a certain place to be willing to stay there under this particular neighborhood circumstances.
The results show, as expected, that if the agents care much about a certain characteristic of their neighbors there is much segregation concerning this characteristic. Furthermore, it was found that the Von-Neumann neighbors cause much more segregation than the average of the whole district, which damps the phenomena. It was achieved to define certain tolerance parameters for which the end state of the segregation is comparable to reality and is reasonable. In a second part the influence of situation-independent moving was studied. It turned out that the average happiness of the agents can not be increased by letting 5% of the agents move no matter whether they like the given conditions or not at a time step.

Acknowledgement We would hereby like to thank Michael Bröniger from the Presidential Department of the City of Zurich for his help and for providing the data used in this work. Furthermore, we want to thank the Assistants Olivia Wooley and Tobias Kuhn for their kind help.

Contents

1 Individual Contribution

Table 1 shows the individual contributions to the project of the two group members.

Table 1: Contribution of the group members

Brechbuehl Sonia	Buechi Jonathan
Data research	Data research
Idea and Construction of model	Idea and Construction of model
	Implementation Code
Writing report	
Analysis and Evaluation	Analysis and Evaluation
Presentation	Presentation

2 Introduction and Motivation

The segregation of people in living areas of cities is a well known phenomena which was investigated strongly in the past. Because we are living in Zurich we would like to find out more about the social situation here. Zurich is a multicultural city, many people from different origins and social backgrounds are living here. A combination of circumstances and characteristics of people makes their segregational behavior a complex problem. In order to investigate this interactions, in this paper a general model based on the data of Zurich is implemented and tries to simulate their segregational behavior due to differences in age, family state and origin.

2.1 Theory

2.1.1 Residential Segregation

Residential segregation can be a result of different origins. Firstly there is the non-tolerance of people towards others of different religion, origin, color, sex, age, income, language etc. It is clear that this segregation aspect is due to individual choices that discriminate. A model which describes this segregational aspect was brought up by Schelling [1]. Segregation can also be economically determined, e.g. by housing prices

in certain areas, such that certain people can not afford to life in expensive areas which are mostly also more attractive i.e. "better". As a result of this, one can often observe ghetto formation in bigger cities [2].
This paper only treats the segregation that results from forms of discrimination. The people in the model have different origins, age and family state and a certain tolerance to accept people which are different from themselves in their direct neighborhood. It can be understood as an extension of the work by Schelling [1] to more than one dimension and more than one tolerance parameter.

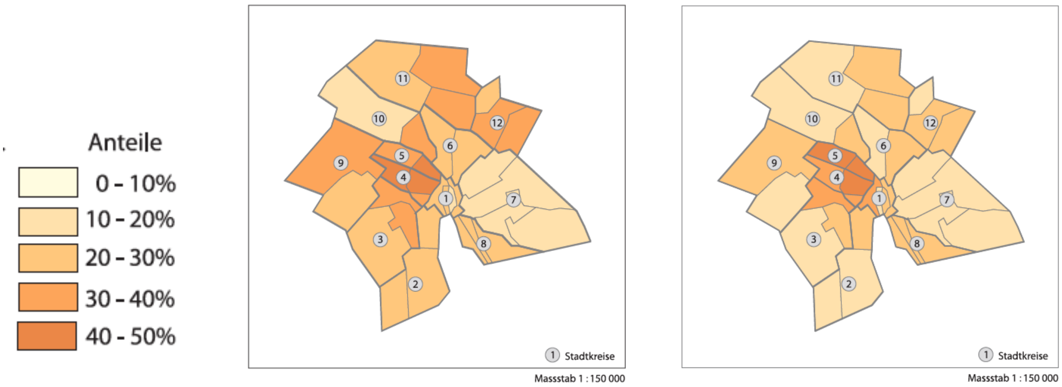
2.1.2 Agent Based Modeling

Residential segregation is a dynamic and interactive process. Therefore one has to make use of a model which considers each of the actors in it as an autonomous individual. Each is able to make decisions and move accordingly in the next time step. Agent based Modeling (ABM) was firstly used in the late 1940s. It is a way to simulate autonomous agents and their actions and interactions with each other and the environment. ABM tries, so to say, to see the results of this in macro-level phenomena [2]. In this way, ABM is a good choice to simulate dynamically the discrimination based segregation in an area due to choices of the agents [3].

2.1.3 Segregation in the city of Zurich

Using the results of studies done by the statistical department of the government of the city of Zurich (see [4] and [5]), one can easily see that the actual situation in Zurich is quite much segregated.
In figure 1 part b) one can see as an example the foreign resident population in the different districts of Zurich in year 2000. There are districts with very little foreign population like the districts number 7 or 10 and in others like 4 live up to 50 per cent foreigners. Looking at statistics about family situation and age distribution, there is also a segregation visible. Furthermore, comparing the numbers from year 2000 to earlier statistics one can observe trends and movement among the inhabitants. In figure 1 the foreign resident population in year 1990 is shown in image c).

An even bigger difference one can see in the data on the income, wealth and social state of the inhabitants. This aspect, however, is not treated in this work.



(a) Legend (b) year 2000 (c) year 1990

Figure 1: Foreign resident population in the twelve districts of Zurich [5]

2.2 Research Questions

One of the main goals of this project is to find out whether there can be seen different segregational behaviors due to the choice of the tolerance parameter of the agents. In the model will be six tolerances, three for the Von-Neumann neighbors, one for each criterium, and three for the differences from the agents characteristics and the respective average of all agents in the district. Can segregation be seen in the same extent if one looks only on the district average and not considers the Von-Neumann neighbors? (This latter case is what Schelling was been doing in [1] for one dimension.)
An other point is to find out, how threshold values have to be chosen to have "reasonable" segregational behavior. Furthermore, it was investigated whether a higher "overall happiness" can be reached when a certain amount of people leave their places in each step independent from their current happiness with their neighborhood, i.e. if people just move away sometimes. The idea behind this is, to make agents leave their site at which they are currently pleased (but not maximally happy yet) and give them the opportunity to find a better matching place. An other aspect is then, that the site becomes available for an other agent which might fit in better there but would not have had the chance to move there otherwise, since the first agent would never have left the site. The idea is to reach an equilibrium state with higher average happiness due to this newly induced movement.

3 Methods

3.1 The Data

The simulations are run with data provided from the statistical department of the city of Zurich. The data is given as can be seen in table 2. The first column represents the statistical zone, the second column the family state, the third the origin, the fourth the age and the last column the number of persons for which these criteria are true. Overall, a set of $333,105$ agents was used.

Table 2: Example of data used in the simulation

Statistical Zone	Family state	Origin	Age	Number of people
11104	Einzelperson	SchweizerInnen	45.45695	1661

The age (column 4) is always averaged over the number of people for which the other three criteria (stat. zone, family state and origin) are true.
Unfortunately, it was not possible to get more detailed data on the age nor data on the income and wealth of the people.

3.2 Model and Implementation

All codes for the different steps as well as the frame code which runs the simulation on several cores and codes for the evaluation are uploaded on *GitHub*.
Comment: For simplicity only male pronouns are used for the agents in the following description.

3.2.1 Part I: Simulation without additional moving

The model used is an agent based model with the agents representing each an inhabitant of the city of Zurich. As explained in the previous section, they have three characteristics: An origin, i.e. Swiss or foreign, a family state and an age. This was implemented by using a structural array for each agent. A vector with all the $333,105$ structural arrays was created.

The city of Zurich can be imagined as a vector of length $333,105$ divided into twelve parts, each with length according to the relative size of districts one to twelve. To avoid places which are not taken by an agent the respective fraction of $333,105$ is always rounded down. This implicates that six agents are not participating in the events of one time step. However, since these are always randomly chosen it does not affect the results. Due to reasons explained in the previous section, the sites don't have any characteristics themselves. Furthermore, a vector with six entries is introduced, each defining the importance or weight of a given criteria for the Von-Neumann neighbors and for the district average. This can be interpreted as the tolerance of the agents (all have the same) towards the criteria (the bigger the number the smaller the actual tolerance). To be able to compare the results for different tolerances this vector is normalized to six (since there are six entries). Furthermore, a "happiness-threshold value" (HTV) is introduced, which defines the threshold where the agents stay or leave a given site. This value is chosen for the different simulations between zero and six, zero meaning the agents do not care about anything in their neighborhood and six meaning they are not to be pleased by any than the absolute perfect situation.

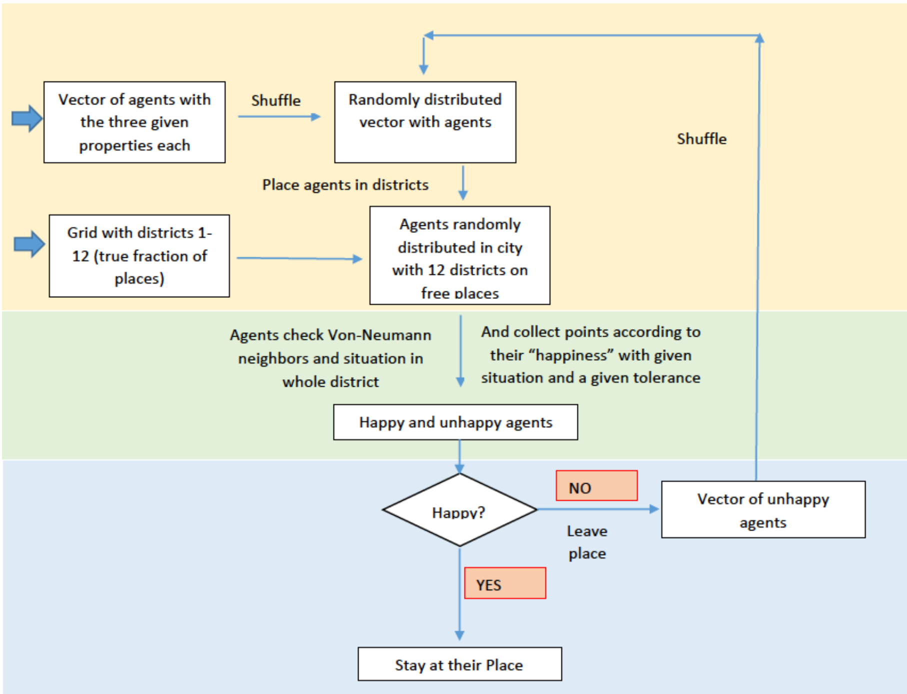


Figure 2: Schematic overview of the model

With the randomization of the agents and the known lengths of the district this can be understood as if all agents were thrown randomly onto the sites of the city. This is illustrated in the yellow part of figure 2. The agents only see within their district and can not look across the borders, i.e. periodic boundary conditions for each district are used.

The percentage of Swiss and families and the average age of each district are calculated. In the blue part of figure 2 it is illustrated that each agent checked its four nearest neighbors and the averages of the respective district he is currently in. For each of the six criteria the agent collects a number between zero and one, zero meaning he is the only one of his kind (origin and family state) or his age is very different from the neighborhood, i.e. he is very unhappy with the situation for a given criteria, and one means he is perfectly fitting in. These numbers are then weighted with the according tolerance and a "current-state-happiness value" (CSHV) is generated as the sum of all the weighted numbers. Due to all the normalizations done, this number is in the interval $[0,6]$. Comparing each agent's CSHV to the HTV defined previously, it can be decided for each agent whether he stays where he is or leaves the site ($CSHV > HTV = happy = stay$). In the blue part of figure 2 it can be seen that the agents which are happy stay at their location in the city (i.e. keep their position in the vector), while all unhappy agents were shuffled and given a new place and then checked again.

Table 3: Tolerances for the different simulations

first three average of district, last three van Neumann neighbors $[origin, age, family, origin, age, familystate]$
$[0,0,0,1,0,0]$, $[0,0,0,0,1,0]$, $[0,0,0,0,0,1]$, $[1,0,0,1,0,0]$, $[0,0,0,2,1,1]$ $[0,1,0,0,1,0]$, $[0,0,1,0,0,1]$, $[0,0,1,1,0,0]$ $[0,0,0,4,1,1]$, $[0,0,0,10,1,1]$ $[10,0,0,1,0,1]$

The simulation is run for threshold values ranging from 3 up to 4.5 (4, 4.1, 4.3, 4.5, 3.0, 3.4, 3.6, 3.8, 3.9). For all the threshold values the tolerances are changed as listed in table 3. All these simulations are run over 500 cycles. A few of them are redone with 1000 cycles.

3.2.2 Part II: Simulation of situation-independent moving included

In a second part of the project, always $15'000$, i.e. approximately 5% of the agents are moving away from their place no matter whether they like the conditions at a given time or not. This agents are chosen randomly. The ones who don't like their place leave as before. In this part the simulation is run for the conditions given in table 4. These are chosen based on the results of the first part of the project.

Table 4: Tolerances and HTVs chosen for the second part of the project

HTV	$[origin, age, family, origin, age, familystate]$
3.6	$[0,0,0,4,1,1]$
3.6	$[0,0,0,2,1,1]$
3.6	$[0,0,0,10,1,1]$

4 Results and Discussion

4.1 Segregation due to different tolerances and HTV

Firstly, the results of the simulations with only one non-zero entry in the tolerance vector are discussed. Therefore, one has to keep in mind that a high entry means high favor of the own kind. As expected, the segregation is very clearly visible for this particular characteristic as can be seen in figure 3 for the origin. The average happiness of the agents in the end state in 3 is almost at 6, which is the highest value possible. Also looking at the evolution of the CSHV one can see that it approaches this value quite quickly. The evolution of the CSHV of the same simulation as in figure 3 is plotted in 4. In addition, it can be observed that the origin and the family state cause a stronger segregation than age. This is also not very surprising since, as mentioned above, the ages are averaged. However, it strikes that there seems to be a correlation between the age and the origin as well as between the age and the family state while the family state and the origin do not show this correlation behavior. This is probably due to the fact that the age is always averaged with respect the origin and family state in each statistical zone.

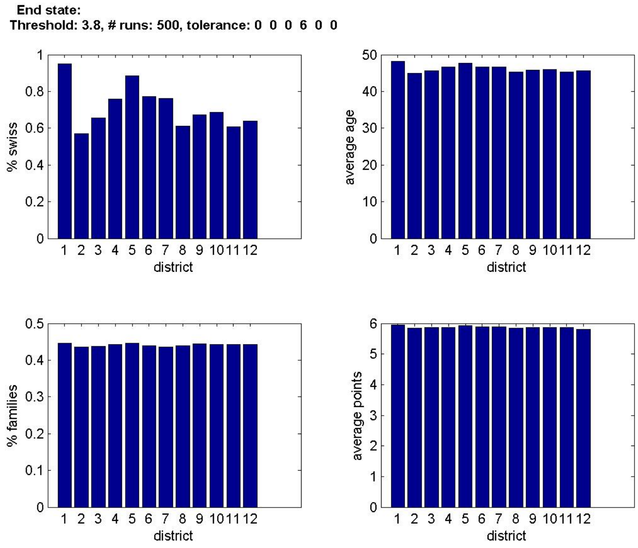


Figure 3: The results for threshold value 3.8. The only non-tolerance is against different origin among the Von-Neumann neighbors.

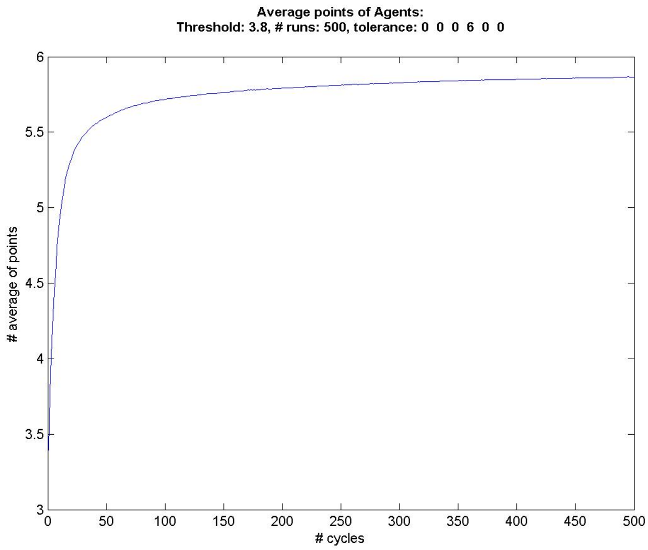
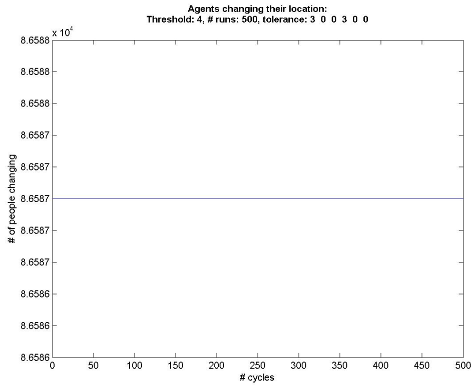
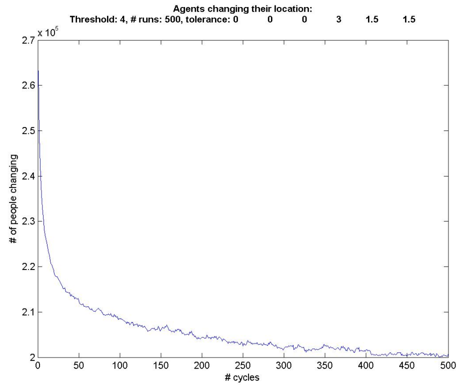


Figure 4: Evolution of the CSHV for threshold value 3.8 and the tolerance vector $[0,0,0,6,0,0,]$.

Furthermore, the results show clearly, that the Von Neumann neighbors cause strong segregation while the average of the whole district tends to preserve the given random



(a) Case where the agents care about the origin of their direct neighbors as well as about the origin of all others living in the same district.



(b) Case where the agents only care about characteristics of their Von-Neumann neighbors.

Figure 5: The amount of agents changing in each cycle for two different tolerance vectors.

state. Inserting a medium to high non-tolerance for a characteristic concerning the whole district, this has a damping effect. After a few cycles almost no change in the number of agents who leave their site can be observed. In figure 5 this dumping effect is visible in a) compared to the evolution of the CSHV in b) which results from the simulation with the same HTV$=4$ but tolerances only concerning the Von-Neumann neighbors. Also simulations where e.g. the origin of the whole district was a big issue for the agents (tolerance 5) and the origin and family state were just weighted with 0.5 each for the nearest neighbors show this trend. One can see almost no segregation for the origin. Surprisingly, some segregation is visible for the criterium of the family state. Only when one of the criteria for the whole district is weighted with 6, there is a bit of segregation visible. However, this only happens for rather small HTV. For thresholds > 4, there is nothing happening even for this extreme case. This can be explained by the randomness the agents are placed onto the sites of the city. The agents seem not be able to see the characteristics (represented by other agents), against which they have a possible non-tolerance, in the crowd of the district and therefore, the segregation is evened out by this random distribution.

The most reasonable distribution of agents seem to result out of tolerances only depending on the nearest neighbors. The cases $[0,0,0,4,1,1]$ and $[0,0,0,3,1.5,1.5]$ show

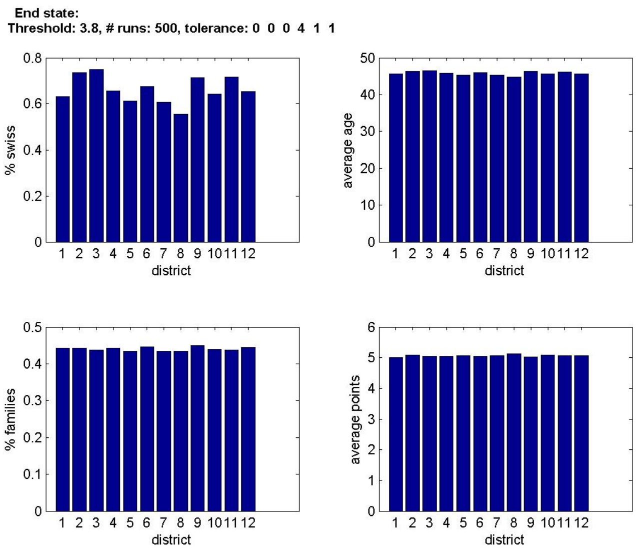
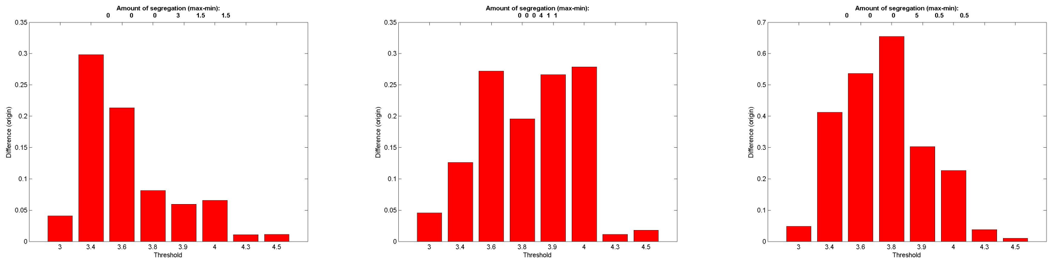


Figure 6: The results for HTV 3.8 and the indicated tolerance vecor

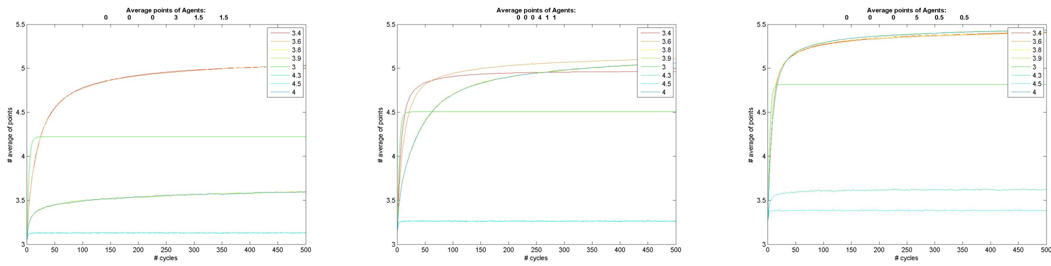
results which are comparable to reality in terms of the distribution of families and Swiss in the districts. Here one has to keep in mind that in the simulation the tolerances were normalized to 6. Therefore the case $[0,0,0,3,1.5,1.5]$ is where the initial values where chosen to be $[0,0,0,2,1,1]$. Figure 6 shows more or less how it could be in a real city. Comparing it to 3 it strikes that the average happiness is never as high for this case as it is in case of only taking into account one parameter (see also figure 8 for a similar case). Nevertheless, also for this latter case a stable equilibrium seems to establish, which, however, is still not reached after 500 cycles.

Thirdly, it can easily be seen that the simulation shows segregational behavior for HTVs between 3.4 and 4. This is shown in figures 7. In figures a) to c) the difference in percentage of Swiss in the district with most Swiss and with least Swiss is shown for all the different HTVs. It can be seen that above 4 and below 3.4 for none of the cases actual segregation is observable. (An example is the case of threshold 4.5 with tolerance vector $[0,0,0,4,1,1]$ where nothing is visible anymore.) Only for the case where only one criterium of the Von-Neumann neighbors was weighted with 6 HTV of 4.5 still show segregation. This can be explained as follows: For a threshold below 3.4 after the first few cycles the agents are not moving anymore because they are satisfied very quickly. Nevertheless, they are not yet very happy, just "happy enough" to stay.



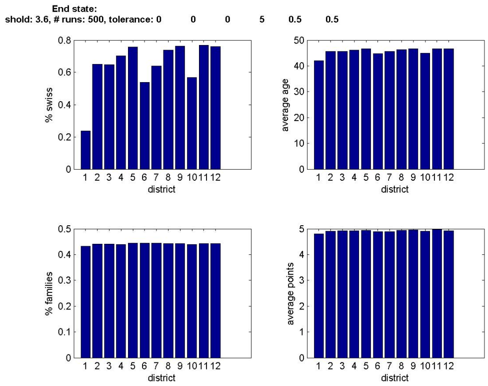
(a) (b) (c)

Figure 7: Plot of difference in percentage of Swiss in the district with most Swiss and with least Swiss the for all the different HTVs. The figures a), b) and c) show the situation for different tolerance vectors

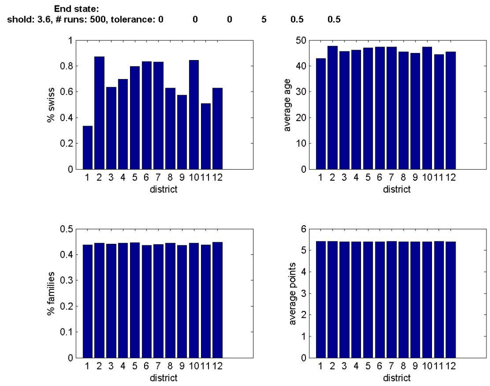


(a) (b) (c)

Figure 8: The difference in the segregation for each characteristic of the agents and their CSHV after 500 cycles



(a) in each timestep 15’000 agents leave their site independent of the situation



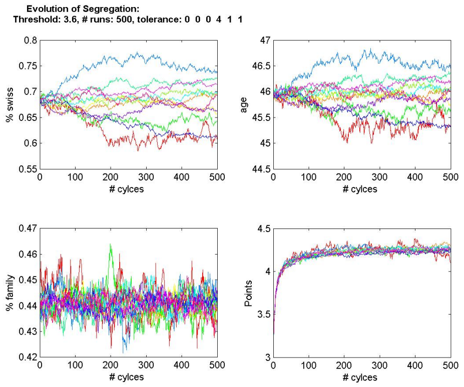
(b) only unhappy agents leave their site

Figure 9: The end states are plotted for the two models with HTV 3.6 and tolerance vector [0, 0, 0, 5, 0.5, 0.5]

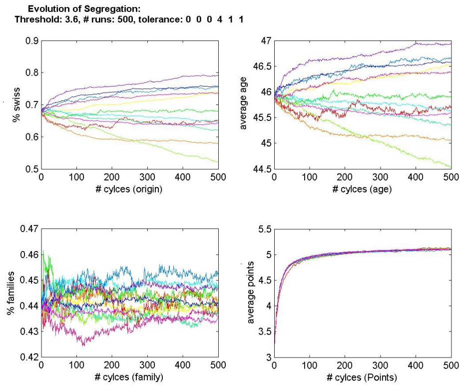
This is shown in figures 8. For values above 4 the agents do not come to rest because they are with almost no situation to be pleased. Even for the simulations with 1000 cycles still many agents are moving (see figures 8) and no real segregation is therefore to be observed. This of course is not reality since not a third of the population moves so often. It is not very surprising that the ones which are happiest (have most happiness points in the end of the simulation) are the ones which live in a rather segregated area according to their tolerances. This is clearly visible when comparing figures 8 and 7.

4.2 Segregation with additional situation-independent moving

For the second part the simulation were only performed on HTV and tolerance vectors which lead to good results in the first part. The goal was to see whether The simulations with the condition that in each cycle 15’000 i.e. 5% of the agents leave their sites no matter if they like their current neighborhood or not lead to very surprising results. In figure 9 the results for HTV 3.6 and the tolerances $[0,0,0,5,0.5,0.5]$ are shown. On the left hand side the new results with the extra agents moving are shown and on the right hand side the result from the original simulations are presented, where only the unhappy agents left their place. It is astonishing, that the extra movement of agents leads to lower segregation for all criteria and to a lower overall happiness. A possible



(a) in each timestep 15’000 agents leave their site independent of the situation



(b) only unhappy agents leave their site

Figure 10: The evolution of the segregation for the 12 districts as well as the district averages of the CSHV are presented in the four sub pictures for a HTV 3.6 and the tolerances [0, 0, 0, 4, 1, 1]

explanation would be, that the agents which are in the original case happy also have to leave and end up in a state where they are less happy and do not find so easily a better state again. This effect is even stronger for the simulations done for the tolerances $[0,0,0,4,1,1]$ and $[0,0,0,3,1.5,1.5]$. This can be explained that the situation in these cases is more complex and the agents can not only become happier due to one criterium i.e. it is worse to leave a site where they were happy and harder to find one where they feel good again. This is shown in the lower right pictures in figures 10.

The continuous moving of the agents is also visible in 10, where the evolution over all time steps is shown. The left hand side again shows the curves for the simulation where 15’000 agents had to move and the right hand side the evolution in the original simulation. While the agents in b) quickly reach their comfortable state and the rather smooth curves show a more or less clear trend the situation in a) is much wilder. The agents do not seem to reach an equilibrium in the different districts and also the CSHV is fluctuating much more during the whole simulation in a). However, neglecting the fact of the fluctuations, these curves seem to follow the same trend in terms of their gradients. The curves in a) seem to be shifted down for more than half a point compared to those in b).

Finally, a film was created to being able to watch the dynamics of the segregational behavior. It is visible, as explained above, that for some choices of parameters the relaxation happens quite fast and for others it takes more cycles. A result which is visible only in this way of presenting the result is, that one is able to see, where the dynamics take place. Obviously the agents which are in an agglomeration of people of their own kind do very rarely change site. After a few cycles all the dynamics happen at the boarders of the agglomerations, i.e. where the different kinds of agents meet. The agents at the boarders of the domains also have smaller CSHV at the end of the simulation. This is very comparable to the results of Monte Carlo simulations applied on physical systems.

5 Conclusions and Outlook

A city with much segregation in this model seems to please the agents more, i.e. the average CSHV is very high if there is much segregation, this can be seen in figures 7 and 8. However, these situations are not very close to a real ones, since there is very extreme segregation (like it shows up if only one of the characteristics is weighted very strongly for the nearest neighbors). There is also this correlation between the characteristics which can not be explained in this model.

HTVs between 3.4 and 4 and tolerance vectors such as $[0, 0, 0, 3, 1.5, 1.5]$ lead to a moderate but still very well visible segregation. In this case the agents do also have reasonable CSHVs. Of course this results of the simulation have to be looked at carefully, since important characteristics of the sites in the city are not taken into account and the agents don't differ in their social status (e.g. education), income or wealth. These kind of characteristics are also as much, if not even more, important for studying the formation of ghettos, with all serious and grave consequences, in a city.

In this study we could show that the Von-Neumann neighbors have a bigger influence on the decision of the agents than the average of the whole district, in this sense it was shown the same behavior of the agents as already Schelling wrote about in [1] only in a more dimensional case.

The idea of raising the average amount of happiness points by kicking some agents out of their site no matter if they like the current situation or not, did not work out as expected. Perhaps the amount 15'000 is too high and therefore the lowering of the agents moving to only a number slightly higher than 15'000 can not be achieved. An other possibility might be to not kick out 5% in every cycle but only do this after 10 timesteps. More ideas for further studies on that topic are:

- More runs of the same simulation to get statistical relevance
- Simulate for more cycles to get equilibrium, i.e. to see for how long there is still not an equilibrium state reached
- Simulating with an initial state which is already segregated. This may lead to an influence of the first three entries in tolerance vector as well.
- Expand the model with data on social status, income or wealth.
- Simulate with a city in which the different districts have different characteristics.

References

[1] Thomas C. Schelling: Models of Segregation, *The American Economic Review Vol. 59, No 2, Papers and Proceedings of the Eighty-first Annual Meeting of the American Economic Association*, **1969**, pp. 488-493

[2] Li Yin: The Dynamics of Residential Segregation in Buffalo: An Agent-based Simulation, *Urban Studies Journal Limited*, **2009**, pp. 2749-2270

[3] Mark Fossett and Warren Waren: Overlooked Implications of Ethnic Preferences for Residential Segregation in Agent-based Models, *Urban Stud.*, 42: 1893-1917

[4] Stadt Zürich, Präsidialdepartement, Statistik Stadt Zürich: Essential Zurich 2013, **2013**

[5] Corinna Heye, Heiri Leuthold: Segregation und Umzüge in der Stadt und Agglomeration Zürich

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of Originality

This sheet must be signed and enclosed with every piece of written work submitted at ETH.

I hereby declare that the written work I have submitted entitled

____

is original work which I alone have authored and which is written in my own words.*

Author(s)

Last name	First name
BRECHBÜHL	SONJA
Büchi	Jonathan

Supervising lecturer

Last name	First name

With the signature I declare that I have been informed regarding normal academic citation rules and that I have read and understood the information on 'Citation etiquette' (http://www.ethz.ch/students/exams/plagiarism_s_en.pdf). The citation conventions usual to the discipline in question here have been respected.

The above written work may be tested electronically for plagiarism.

12/12/13
Place and date

Signature

*Co-authored work: The signatures of all authors are required. Each signature attests to the originality of the entire piece of written work in its final form.

Print form