Using Survey Data to develop agent-based models of spatial segregation

Abstract. Segregation processes depend on the distribution of neighbourhood preferences. These preferences are not constant among either the native or the immigrant population. This was already shown by Farley and colleagues (Farley et al. 1978) and Emerson and colleagues (Emerson et al. 2001). The developed show cards by Farley and colleagues (1978) were adapted and used for the German General Social Survey (ALLBUS) 2006.
Most studies refer to the segregation model developed by Schelling (2006). Schelling (2006: 150pp.) designed a dynamic, self-regulating system to describe the emergence of spatial segregation patterns. The model implies that actors are not aware of the consequences of their individual decisions. The fact that the traits of individual actors influence each other can lead to an interlinked process of spatial segregation. In this short paper the model uses results of the ALLBUS to investigate how different values effect the segregation of the social groups.
It can be shown that with the implementation of different levels of similar wanted the agents move more often and partly remain within different clusters. This makes it possible to examine how different values derived from the empirical analysis of the ALLBUS affect Schelling's model, which was created in Netlogo.

Keywords: *segregation, agent-based modeling, simulation study*

1. Introduction

With the increased immigration of refugees to Germany since 2015, there has been a strong increase in the proportion of people with a migration background as well as a strong diversification of immigrants' references to origin. Immigration and diversification are changing the ethnic composition of the population, especially in cities, and thus driving processes of ethnic segregation. The phenomenon of segregation is a well-known topic but in its theoretical foundations very limited. In this short paper, I will point out a theoretical approach for an agent-based simulation of segregation pattern and investigate how different levels of tolerance influence the outcome of the model. To develop the different level of tolerance data of the German General Survey (ALLBUS) 2006 will be analysed and implemented in an agent-based model.

2. Theoretical Background

The theoretical foundation of segregation is usually seen in the model of Schelling. Schelling (2006: 138) acknowledges that he leaves out two processes in his model. One is the organised actions, whether these are legal/illegal, forced or merely exclusionary, or subtle or blatant. The other is the effect of socio-economic aspects. By this, Schelling means above all processes that separate the rich from the poor or the better educated from the less educated (Schelling 2006: 139). In our contemporary understanding, we would call this process social segregation. He acknowledges that race correlates closely with income and thus influences residential choice. That means that the residential segregation is a result of these residential choices and can be seen as manifestation of social distance. He assumes that even if race is ignored, segregation occurs. Schelling (2006: 139) writes that the choice of neighbourhood is a choice of neighbours. This means for people who want a specific social group in their neighbourhood force to move in these areas because the majority is like a hint for specific characteristics. A explanation he uses is that other people have seen skin colour as a signal and the people now reflect the signal. In sociology, this is called the Matthew effect (Merton 1995). Schelling (2006: 141) acknowledges that there are always segregated neighbourhoods, but the characteristics can differ. In the US usually ethnic segregation is researched by skin colour and that it is difficult to find neighbourhoods that are neither 75% populated with people of white or black skin colour. In a comparison, Schelling writes that it is also difficult to find areas that have a balance that lasts long enough. Most segregation studies focus on ethnic segregation but there are other forms of segregation like social or demographic. In this short paper I will focus on ethnic segregation.

Schelling (2006: 141) admits that his model does not explain how segregation occurred. In this context, Schelling assumes that an analysis of the majority situation can only happen locally and that each group strives for this numerical superiority. When this is achieved, they try to segregate themselves from the smaller group. To understand segregation, the researcher needs to analyse the incentives that motivate or sustain behaviour. However, no equilibrium is sought by both groups, since the disappearance of a minority leads to complete segregation. At the same time, complete segregation is a

stable condition, while in other conditions a shift in the mix is always possible (Schelling 2006: 142p). The question that arises is how these mixtures are influenced by individual decisions. According to Schelling, some rules are used to obtain legitimacy for certain behaviours (Schelling 2006: 144).

The question is how segregation arises and what theoretical foundation can be used for Schelling's model. In the following, I would therefore like to discuss the established-outsider configuration according to Elias and Scotson (1993).

The negotiation of norms is carried out through the process of communicating the action taken, i.e. the articulated advocacy. However, the exercise of the new norm leads to a demarcation between the behaviour of the opponents of the new norm and reality. Thus, a conscious demarcation is focused on. Elias and Scotson (1993) refer to such a figuration as an established-outsider configuration.

Established people, according to Elias and Scotson, are characterised by the conviction of a group charism in which all members of the group participate. With the group charism comes a strong "we" ideal and the belief that one's self belongs to a group of people who have a higher value (Elias and Scotson, 1993: 9). In this context, the self-image is shaped by the most positive subgroup of the establishment group. The counterpart to this is group shame. Group shame is characterised by a group being ascribed the "worst" characteristics of its "worst" subgroup (Elias and Scotson, 1993: 13). This results in a social devaluation, of the outsiders (Elias and Scotson, 1993: 9). The stigmatisation of groups is usually linked to collective ideas about certain groups. Through the ideas, the behaviour of the stigmatising group becomes excusable, since it is not the stigmatising group that has endowed the stigmatised group with the characteristic, but this characteristic comes from higher powers. Thus, the characteristic is used as a sign of inferiority or badness. This symbol not only has the function of relieving the stigmatising group of the burden of guilt, but also of defending the existing balance of power (Elias and Scotson 1993: 32p.). Segregation is a mixture of such structures. On the one hand side person would avoid moving into districts where they are in the position of the minority and on the other hand side leave the district if the neighbourhood changes. Segregation thus arises primarily out of the need to distinguish oneself from others and to maintain a positive group perception. The theories of Elias and Scotson (1993) can be used as a theoretical foundation of the processes described by the segregation model of Schelling.

The problem Schelling (2006) mentioned that this model is missing a theoretical basis can be solved when using this theory. Even if individuals do not want to separate themselves, the desire for at least 50% of the same social group in the environment leads to the emergence of segregation patterns.

In research on segregation, ABM is also used regularly. In an international comparison, many studies on segregation have already been conducted. Therefore, only a few selected studies will be discussed here. Hatna and Beneson (2012) were able to use ABM to generate a real-world picture of the distribution of religious communities in two Israeli cities by using census data from 1995 and taking into account residential preferences in relation to one's own religion and a similar prestige of the residential environment. Their models were able to replicate real-world mixed neighbourhoods. It turned out that not only one characteristic is sufficient to replicate segregation. Liu and colleagues (2019) considered actors who could influence neighbours in their ABM and showed that this changed relocation behaviour. Zuccotti and colleagues (2021) showed that segregation is influenced by socio-economic status and ethnicity. These models show that ABMs can be used to replicate real-world segregation patterns. There are no segregation models developed for the German context. Moreover, there is another aspect that is neglected. These models say nothing about the processes that lead to segregation. These models say nothing about the processes that lead to segregation. Building on the theory of Elias and Schelling, this model will look at how the model changes when real world data is used for the model. Here, the tolerance values function as an indicator of whether people accept other group members in their neighbourhood.

3. Hypothesis and Research question

Agent-based simulation studies (ABM) provide a better understanding of the course of segregation processes. Through modelling, it is possible to observe how individual characteristics generate different macro phenomena (Flache & de Matos Fernandes 2021: 453). One way to make the spatial segregation model more realistic is to use survey data. In this case the neighbourhood can be understood as a social network. This network consists of relationships with different strengths. These can range from friendships to more casual contact. To simplify the model, it is assumed in the simulation

study that the agents perceive their neighbours and adapt their actions accordingly. What preferences exist with regard to neighbourhoods and how do these affect segregation patterns? This is the research question that is the focus of this presentation.
With this theoretical foundation of Schelling (2006) and Elias and Scotson (1993) the hypothesis arises that persons reject neighbourhoods in which they are in the minority position and would avoid them. This hypothesis is supported by the model assumptions of Schelling and results from the theory of Elias and Scotson (1993). The specific tolerance values should arise from empirical data. The research question is how these values effect the model outcome.

4. Methods

To investigate this hypothesis, I proceeded in two steps. In the first step, the data from the ALLBUS 2006 and 2016 were analysed. There, in accordance with Farley et al. (1978), the consent to the different composition of neighbourhoods was asked. The first step is a descriptive evaluation of this item. In the second step, a model was programmed in Netlogo (Wilensky 1999). The model that takes up the data from the ALLBUS.

5. How the model works

The basic structure does not differ much from the segregation model of Wilensky (1997), only the specifications made make it significantly more complex than the original model. These changes will be discussed in the following. First, the model assumes two distinguishable groups. For convenience, the agents have been coloured blue and red. This colouring can be perceived by the individual agents. The two groups are randomly distributed, just as is the case with Schelling. Due to the assumption of neighbourhoods, there are no further social ties between the agents. The neighbourhood in this case is therefore based on pure perception by the agents. The social identity is based on one's belonging to a social group. In the following, we will discuss which settings can be made in the model.

It is possible to vary the density of the agents. The density of the agents can be varied between 0 and 100% via a controller. In this way, the assumptions of Schelling can be used. In addition, there are two controllers for the neighbourhood preferences. In the model the share of similar and unsimilar neighbours can be changed through sliders. Therefore, in this text the term sliders is used. The %-similar-wanted slider specifies how many agents of the same social group are desired in the neighbourhood. A second slider (%-unsimilar-wanted) specifies how many different neighbours are desired in the neighbourhood. Both sliders can be set to a value between 0 and 100.

There is the possibility to expand and limit perception. On the one hand, there is the possibility to activate the Neumann neighbourhood. The Neumann neighbourhood is a specific implementation in Netlogo and effects the number of neighbours the agents perceive. If the Neumann neighbourhood is deactivated, the agent perceives its eight neighbours completely. If the Neumann neighbourhood is activated, the agent only perceives one neighbour above him, one below him and one to the right or left of him. This means that the perception of the neighbourhood is reduced by half. With the slider radius-neigh, the neighbourhood can be extended to the perception of up to ten neighbours. There is also the possibility that there are not only two, but four groups in the model. This is done with the switch four-groups? Slider Noise adds a disturbance term to the model that specifies a random value and thus makes the model more realistic. In addition, monitors can be used to observe the concrete number of members in each group, how many agents are happy and how similar the agent groups are. To begin with, there are the variables of how high the average value of equal neighbours desired by all agents is and what percentage of all agents are unhappy.

Table 1 shows the properties of the agents at the beginning.

property	value	concept
Happy?	true/false	If happy is false the agent moves and look for a better place nearby.
similar-nearby	numeric	how many neighbouring patches have a household with my group?
other-nearby	numeric	how many have a household of another social group?
total-nearby	numeric	sum of other and similar variables
similar-nearby-fraction	numeric	fraction of neighbours who have the same social group than me

attractiveness-now	numeric	to compute and store current attractiveness once moving is considered
patch-now		to remember current patch before household starts moving
similar-nearby-opt		optional new patch: how many similar are nearby
other-nearby-opt		optional new patch: how many household of another social group are nearby
total-nearby-opt		optional new patch: how many household are there nearby in the potential new spot
similar-nearby-opt-fract		optional new patch: fraction of similar nearby
ethnicity	numeric	belonging to a specific group
my-%-similar-wanted	numeric	the threshold for this particular agent
tolerance-level-neighbours	numeric	influence of tolerance to the neighborhood composition

Table 1: properties of the agents at the beginning

The values of the belonging to one of the social groups, the tolerance-level, similar-wanted and unsimilar-wanted are drawn randomly.

6. Results

In the ALLBUS, people were asked in which residential area they live or in which residential area they would not like to live. Based on the vignettes of Farley and colleagues (1978), a typical residential area was drawn. This consisted of 49 units in order to adapt it to the German context (GESIS 2011, 2017, Wasmer et al. 2007). The results are shown in the table. It can be seen that homogeneous residential areas are also rejected by a small proportion. It is therefore more realistic for purely homogeneous neighbourhoods to be rejected and thus for a certain number of the other social group to live there. It turns out that many people prefer a proportion of foreigners between 8 and 50%, which are clearly different values than those assumed by Schelling. It is also evident that neighbourhoods are only rejected above 66%. - rework

	Want to live in the district	**Do not want to live in this district**

	Proportion of foreigners (48 other units)	2006	2006
1	0	0,81	0,05
2	8,3	0,81	0,02
3	16,6	0,73	0,03
4	25	0,56	0,07
5	33,3	0,4	0,12
6	41,6	0,29	0,19
7	50	0,19	0,27
8	58,3	0,11	0,36
9	66,6	0,06	0,46
10	75	0,03	0,58
11	83	0,01	0,68
12	91,6	0,01	0,76
13	100	0,01	0,99

Table 2: Calculations from the ALLBUS 2006

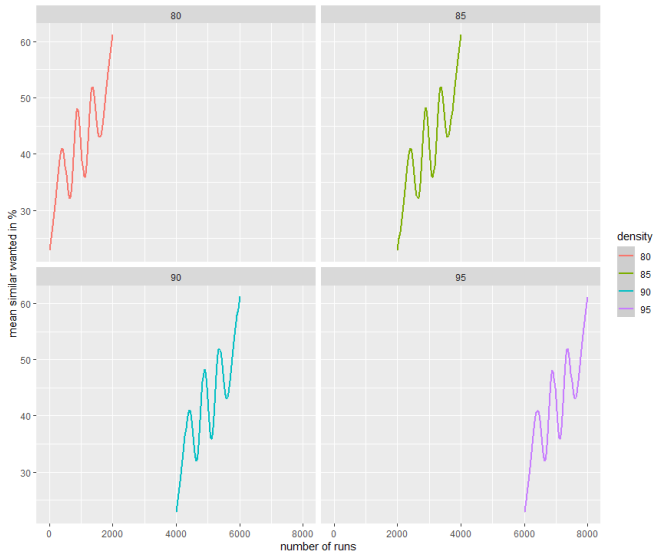
In the model this implications were implemented by the variables %-similar-wanted and %-unsimilar-wanted. In order to better adapt the values to reality, they were randomly selected from the range between 0 and the set limit in the sliders. To answer the research question of how the different values affect the model, an experiment was created in the Behaviorspace in Netlogo. The different values for "density", "%-similar-wanted" and "%-unsimilar-wanted" were varied. The parameter variations can be found in Table 3. These variations are orientated on the results of the descriptive analysis of the data. The maximum of the slider "%-unsimilar-wanted" is set at 20% because in the next step under 50% of the people want to live in this area. The variation of the slider "%-similar-wanted" is oriented on the empirical data, because if around 75% people from a other live in the area only 1% want to live there. The density varies between 80 and 95 because 80 is the value Schelling (2006) choose and between 90 and 95 because most cities in Germany have a vacancy rate of 5 to 10%. 100 runs were carried out per parameter combination with a termination after 500 steps, in total there were 8000 runs of the model.

concept	name of the parameter	values
random effect	noise	0.03
density	density	80, 85, 90, 95

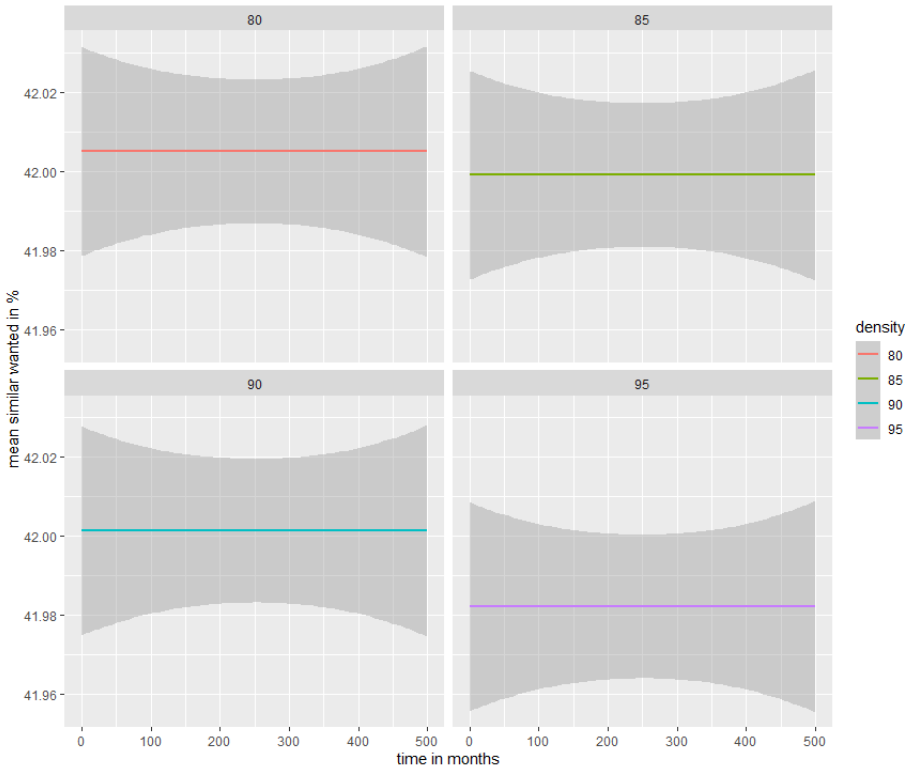
surrounded by agents of the same social group	%-similar-wanted	50, 60, 70, 80
surrounded by agents of the other social group	%-unsimilar-wanted	0, 5, 10, 15, 20

Table 3: Variation of the parameters in the experiment

Graphic 1 shows the variation of the means of the tolerance of all agents and the number of the runs. On the one hand, the new parameters no longer result in a state of equilibrium in the model, but the runs have to be aborted. On the other hand, the tolerance values increase plateau-wise, which is related to the selected population density. On the other hand, the tolerance values rises up to 60%, which means that society seems to become more tolerant simply because a small amount of mixing is desired by the people.



Graphic 1: Mean of the similar wanted depending on the number of runs, separated plots by density



Graphic 2: Distribution of mean of tolerance depending on the steps and variation of the parameters and corresponding confidence interval.

Graphic 2 shows the distribution of the mean of tolerance depending on the steps and variation of the parameters and the corresponding confidence intervals. The confidence intervals are overlapping, which means that there is no clear difference between the experiments is possible. The mean is differentiating between the experiments and is lower with a density of 95 than with a density of 80.

7. Summary and Limitations

In this short paper, the theoretical basis for segregation has been presented first. For this purpose, the theories of Schelling and Elias and Scotson were linked. This provided

a theoretical foundation for the agents' chains reaction. Subsequently, the model developed was presented in Netlogo and its mode of operation was described. The rules followed the theoretical foundation were translated in mathematical terms and implemented in the model. With the ALLBUS results, individual preferences about ethnic composition can be fed into the segregation model in a more complex but also empirically more realistic way. In this paper, an agent-based model is presented that works with the empirical preferences from the ALLBUS. The distributions are rudimental implemented in the model. It turned out that the different preferences ensured that the agents became more tolerant overall and accepted more agents of the other group in their environment over time. Looking at the models also shows that the respective segregated areas in the model do not remain stable as in Schelling's model, but move. These mechanisms need to be investigated further. In addition, the values from the ALLBUS should be implemented more precisely in further models. At present, only random numbers are implemented as lower limits to implement a level of uncertainty as well. In later models, the empirical distributions are to be implemented as well as lower and upper limits. At the moment the individual tolerance is set at the beginning and not changing anymore. The individual tolerance is changing during the life course this should be implemented that the individual tolerance can be shaped as well by the corresponding neighbourhood. It is important to mention that the behavioural rule from Schelling is implemented in this model, which is critical and should be investigated in future research, if this rule reflects correctly the real world. In addition, only a simple descriptive evaluation was carried out, with the extension of the agents' attributes, better cause-effect relationships can be established. It would be interesting to investigate how the models behave when there are four groups of agents with different distributions. There is still more possibilities to use empirical data for agent-based models. In the next step more attributes should be added to the agents and as well attributes should be added to the patches.

References

1. Emerson, Michael O.; Chai, Karen J.; Yancey, George (2001): Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans. In: *American Sociological Review* 66 (6), S. 922. DOI: 10.2307/3088879.
2. Farley, Reynolds; Schuman, Howard; Bianchi, Suzanne; Colasanto, Diane; Hatchett, Shirley (1978): "Chocolate city, vanilla suburbs:" Will the trend toward racially separate communities continue? In: Social Science Research 7 (4), S. 319–344. DOI: 10.1016/0049-089X(78)90017-0.
3. Farley, Reynolds, and William H.Frey. (1994): "Changes in the Segregation of Whites from Blacks during the1980s: Small Steps toward a More Integrated Society." *American Sociological Rev.* 59 (February):23–45.
4. Flache, Andreas & de Matos Fernandes, Carlos A. (2021): Agent-based computational models. In: Gianluca Manzo (Hg.): Research handbook on analytical sociology. Cheltenham, UK and Northhampton MA, USA: Edward Elgar Publishing (Research Handbooks in Sociology), S. 453–473.
5. Friedrichs, Jürgen; Triemer, Sascha (2009): Gespaltene Städte? Soziale und ethnische Segregation in deutschen Großstädten. 2. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden.
6. GESIS - Leibniz-Institut für Sozialwissenschaften (2011). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2006. ZA4500 Datenfile Version 2.0.0. Köln: GESIS Datenarchiv.
7. GESIS - Leibniz-Institut für Sozialwissenschaften (2017). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2016. ZA5250 Datenfile Version 2.1.0. Köln: GESIS Datenarchiv.
8. Hatna, Erez; Benenson, Itzhak (2012): The Schelling model of ethnic residential dynamics: Beyond the integrated - segregated dichotomy of patterns. JASSS 15 (1): DOI: 10.18564/jasss.1873.
9. Liu, Zeyu; Li, Xueping; Khojandi, Anahita; Lazarova-Molnar, Sanja (2019): On the extension of Schelling's segregation model. 2019 Winter Simulation Conference (WSC): 285-296. DOI: 10.1109/WSC40007.2019.9004848.
10. Merton, Robert K. (1995). The thomas theorem and the matthews effect. Social Forces, 74(2), 392-422.
11. Schelling, Thomas C. (2006): Micromotives and macrobehavior. [New ed.] with a new preface and the Nobel Lecture. New York, NY: Norton (Fels lectures on public policy analysis).
12. Wasmer, Martina; Scholz, Evi; Blohm, Michael (2007): ZUMA-Methodenbericht 2007/09. Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2006. Hg. v. GESIS - ZUMA.

13. Wilensky, U. (1997). NetLogo Segregation model. http://ccl.northwestern.edu/netlogo/models/Segregation. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
14. Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
15. Zuccotti, Carolina; Lorenz, Jan; Paolillo, Rocco; Sánchez, Alejandra Rodríguez; Serka, Selamavit (2021): Exploring the dynamics of neighborhood ethnic segregation with agent-based modelling. An empirical application to Bradford. DOI: 10.31235/osf.io/gmzdp.