The Microfoundations of School Segregation

An Agent-based Computational Approach

Edvard Nergård Larsen

Master’s thesis, Sociology
Department of Sociology and Human Geography
Faculty of Social Sciences

UNIVERSITY OF OSLO

June 2015
© Edvard Nergård Larsen

2015

The Microfoundations of School Segregation: An Agent-based Computational Approach

Edvard Nergård Larsen

http://www.duo.uio.no/

Print: Reprosentralen, UiO.
Summary

The aim of this thesis is to address potential causes of ethnic segregation in schools through developing and analyzing an agent-based model. An agent-based model is a computer program where individual agents interact with each other and an environment according to a set of behavioral rules. I argue that agent-based modeling is particularly suitable for theoretical exploration of complex macro-level patterns such as segregation. The thesis includes an overview of the theoretical foundations of agent-based modeling, and discussions of its use in the social sciences. I also discuss analytical sociology, middle-range theory and mechanism-based explanation, in order to provide a theoretical foundation for my methodological approach.

The model I have developed for this thesis represents households enrolling their children into schools, based on simple rules of preference regarding the ethnic composition of and the geographical distance to schools. Building on an existing model created by Victor Stoica and Andreas Flache, as well as Thomas Schelling’s model of residential segregation, analysis of the model seeks to investigate three central questions regarding the growth and persistence of school segregation. The first question deals with how mild preferences for avoiding being in a small minority among parents affect segregation patterns: Can segregation arise among tolerant parents, as Schelling’s model shows regarding neighborhoods, in schools? Second, the effect of including a preference for minimizing home-school distance among parents is examined. Finally, the development of segregation patterns over time is discussed. The results show that Schelling-type mechanisms can operate in schools as well as in neighborhoods. Including a preference for minimizing distance has a curbing effect on segregation, but the effect is dependent on how this preference is modeled. Finally, the model is empirically calibrated to represent a stylized map of Oslo, and the results compared with the purely theoretical model. This comparison shows that the overall behavior of the model is similar when initialized under empirically calibrated conditions, but that the effect of including a preference for minimizing home-school distance among households’ has a greater effect. Finally, I conclude with discussing the weaknesses of the model, and suggest further extensions for future work.
Acknowledgments

This thesis was written at the Department of Sociology and Human geography, University of Oslo. I would first of all like to express my gratitude to my two supervisors, Gunn Elisabeth Birkelund and Torkild Hovde Lyngstad.

Torkild, in the position of my primary supervisor, has contributed greatly with a steady flow of support, commentary, constructive criticism and ideas throughout the process. Special thanks are also in order for providing me with the dataset used for the thesis. Gunn, in the position of secondary supervisor, has far exceeded her obligations by always being available for advice, reading of drafts and brainstorming. In addition, I would like to thank them both for introducing me to the field of agent-based modeling and the use simulation in sociology, and for allowing me the opportunity to attend courses and workshops abroad – which provided knowledge absolutely critical for the development of this thesis. In this regard, I would also like to thank the talented and kind teachers, course organizers and fellow students at the Max Planck Institute for Demographic Research in Rostock.

I would also like to offer my thanks to friends enrolled in the master’s program for support, needed distraction and feedback throughout the last year, and to Elisabeth Ugreninov, for helpful comments and organization of the seminars. Karl Erik Holter from the Department of Mathematics also has my gratitude for invaluable technical and mathematical aid. Additionally, I am deeply indebted to Victor I. Stoica and Andreas Flache for developing the model my thesis is based on. Finally, I personally take credit for any and all mistakes in this thesis.

Edvard N. Larsen
Oslo, June 2015.
# Table of contents

1 Introduction ...................................................................................................................... 11  
   1.1 What is Segregation? ................................................................................................. 12  
   1.2 An Agent-based Model of Ethnic Segregation in Schools ........................................ 14  
   1.3 The Contribution of This Thesis ................................................................................ 15  
   1.4 Structure of the Thesis ............................................................................................... 16  
2 Previous Research ............................................................................................................ 18  
   2.1 Enrollment Policy and Parental Choice ..................................................................... 18  
      2.1.1 School Segregation and Residential Segregation .............................................. 19  
   2.2 The Causes of School Segregation ............................................................................ 22  
      2.2.1 Explanations of the Persistence of Ethnic Residential Segregation ................... 22  
      2.2.2 Explanations of the Persistence of School Segregation ..................................... 24  
   2.3 The Potential Effects of School Segregation ............................................................. 27  
   2.4 Summary .................................................................................................................... 28  
3 Theoretical Perspectives and Methodology ..................................................................... 30  
   3.1 The Analytical Approach........................................................................................... 31  
      3.1.1 Methodological Individualism and Mechanism-based Explanations ................. 32  
      3.1.2 Summary ............................................................................................................ 36  
   3.2 Agent-based Modeling in the Social Sciences .......................................................... 36  
      3.2.1 What is an Agent-based Model? ........................................................................ 37  
      3.2.2 Why Use Agent-based Models? ......................................................................... 39  
      3.2.3 Agent-based Modeling Within the Social Sciences: Three Examples ............. 42  
      3.2.4 Points of Contention Regarding Agent-based Modeling ................................. 50  
      3.2.5 Concluding Remarks .......................................................................................... 52  
   3.3 A Computational Approach to the Question of Persisting School Segregation....... 52  
4 Proposal for an Agent-based Model of School Segregation ............................................ 55  
   4.1 Overview of the Model.............................................................................................. 55  
      4.1.1 Household’s Preferences and Utility ................................................................. 57  
      4.1.2 The Searching Process: Satisficing vs. Optimizing ............................................ 60  
      4.1.3 Schools as Bounded Neighborhoods.................................................................. 60  
      4.1.4 Model Output: The Index of Dissimilarity......................................................... 61  
   4.2 An illustration: Following a Household and a School Through the Model............. 62
4.3 Changes and Additions Compared to the Original Model by Stoica & Flache (2014) 64

5 Analysis of Simulation Results ................................................................. 66

5.1 Theoretical Simulation Experiments .......................................................... 67

5.1.1 Simulation Experiment 1: Segregation Among Tolerant Parents .......... 67

5.1.2 Simulation Experiment 2: Can a Preference for Minimizing Distance Curb Segregation? ................................................................. 69

5.1.3 Simulation Experiment 3: Segregation Patterns Develop Over Time .......... 71

5.1.4 Segregation Dynamics Under Increased Mobility ................................... 74

5.1.5 Robustness and Internal Validity ............................................................ 76

5.2 Empirical Calibration of the Model ........................................................... 78

5.2.1 Empirical Simulation Experiment: Segregation Under Realistic Initial Conditions ................................................................. 79

5.3 Summary and Discussion of Results ......................................................... 85

6 Discussion and Conclusions ...................................................................... 88

6.1 Points of Weakness of the Approach ....................................................... 88

6.2 Possible Extensions of the Model ............................................................... 89

6.3 Conclusion ................................................................................................ 91

References ..................................................................................................... 93

Appendices .................................................................................................... 98

Appendix A: NetLogo code of the model .................................................... 98
List of figures

Figure 2.1: Share of minority-language students in lower secondary schools of Oslo, by share of minority-language students in the surrounding neighborhood, calculated using $k$-nearest-neighbors for two values of $k$ ................................................................. 20

Figure 3.1: Coleman's Bathtub with mechanism typology (Hedström & Ylikoski, 2010, p. 59). ................................................................................................................. 34

Figure 3.3: Schelling’s model of residential segregation implemented in NetLogo. On the left-hand picture, agents prefer at least 25% in-group neighbors, versus 26% on the right-hand side. ........................................................................................................................................................................... 47

Figure 3.4: Visualization of the Anasazi society as implemented in NetLogo. .................. 48

Figure 4.1: The visual interface of the model. ................................................................. 55

Figure 4.2: Graph of households’ utility function ......................................................... 58

Figure 4.3: Distribution of households’ threshold values, $T = 0.4, SD = 0.05$. ............ 59

Figure 5.1: D-index with tolerant agents, by majority share, after ten time-steps ......... 68

Figure 5.2: Boxplot of D-index by households’ search radius, one year of simulation.... 70

Figure 5.3: D-index over 25 years according to households’ search radius and population distribution. ......................................................................................................................... 72

Figure 5.4: D-index over 25 years according to search radius and percentage share of extra capacity in schools. .......................................................................................... 75

Figure 5.5: D-index over 25 years according to search radius and percentage share of extra capacity in schools, empirically calibrated population distribution. .......... 80

Figure 5.6: Majority share of schools: Simulated shares compared with real-life shares.... 83
List of tables

Table 3.2: Squazzoni’s classification of agent-based computational models (Squazzoni, 2012, p. 21-22) ................................................................................................................................... 44
Table 5.1: OLS regression, D-index as dependent variable. N = 3234 simulation runs. ........ 77
1 Introduction

Norway, along with several other European societies, has throughout the second half of the 20th century faced large-scale immigration from non-Western countries (Dustmann & Frattini, 2013). There is uncertainty regarding the long-run incorporation of minority members, and issues surrounding ethnic segregation are central in addressing this uncertainty. A central arena for successful integration of immigrants and their descendants is the educational systems, and ethnic segregation within schools is perceived as a threat to long-term integration: A persistent lack of exposure to majority-group members is considered a potential source of disadvantage for immigrant-background students, and as a breeding-ground for stereotypes among majority-background students. As of 2014, 59 out of 135 elementary schools in Oslo had a larger share of pupils with non-western immigrant backgrounds than pupils with a majority background. In six of these, the share was higher than 90% (Statistics Norway, 2014, 8.11).

Norwegian politicians have suggested introducing lower limit of majority-background students attending elementary schools in Oslo, to ensure a non-trivial presence of individuals of the majority culture and users of the majority language - this with the intention of securing successful, long-term social integration. In order to evaluate potential policy responses to various forms of segregation, one needs a solid understanding of the mechanisms that bring about patterns of segregation. Usually, segregation is understood as a macro-level phenomenon brought about as an often unintended consequence of micro-level behaviors (Schelling, 1971). Segregation patterns emerge from local choices made by individuals and households. This basic premise begs two questions: What mechanisms are involved in affecting the micro-level choices that are made, and how do these choices produce macro-level patterns of segregation? First of all, patterns of school segregation are naturally affected by correlated patterns of residential segregation. This can be assumed to vary with the degree of parental choice of schools for their children; the lesser the leeway for parental choice of schools, the larger the correlation between patterns of residential segregation and those of school segregation. Second, patterns of segregation are assumed to be influenced by the actors’ preferences for closeness or distance to members of another group. Like Thomas Schelling showed with regards to residential segregation, even weak preferences for a certain ethnic composition could potentially cause segregation patterns to emerge across schools: «Where there is parental choice of school, ethnic majority parents may opt for a ‘whiter’ school for their children, a self-perpetuating dynamic that exacerbates ethnic school segregation and has contributed to make the phenomenon a frequent source of public concern in
European countries» (Paasche & Fangen, 2011).

In order to examine the connection between the micro-level behavior of parents on one hand, and the macro-level patterns of segregation on the other, I will build an agent-based model of parents' school choice. Using this model, I will then simulate under what conditions different levels of school segregation emerge. Agent-based modeling is gaining prominence in sociology and allied disciplines. Such models can be defined as “computer programs in which artificial agents interact based on a set of rules and within an environment specified by the researcher» (Bruch & Atwell, 2013). They are systems of recursive functions, and each realization of an agent-based model is thereby a deductive process (Epstein, 2006, p. 56). Agent-based models are not to serve as replacements for other statistically oriented analysis; they are tools which can be used to better our understanding of complex macro-level phenomena. Under conditions not conducive to actual laboratory and field experiments, agent-based models provide a toolkit for experimentation in an artificial computer-based environment. Theoretical ideas and their consequences can be simulated, and provide an explicit link between individual behavior and societal outcomes. Agent-based modeling has surfaced as an interesting tool for use in sociological research (Tufte et al., 2010). As Bruch and Atwell (2013, p. 2) argues: “Because agent-based models explicitly link individuals’ characteristics and behavior with their collective consequences, they provide a powerful tool for exploring the social consequences of individual behavior.” Agent-based models may thus contribute to opening up explanatory “black boxes” between individual choice and levels of segregation.

1.1 What is Segregation?

In all spheres of society, individuals with similar characteristics tend to cluster either physically or socially. The term segregation refers to this clustering, and segregation can arise along countless such characteristics: Age, socioeconomic status, ethnicity or gender, to name a few. A suitable definition can be found in Bruch & Mare (2009, p. 270): “Segregation is the nonrandom allocation of people who belong to different groups into social positions and the associated social and physical distances between groups.” Segregation by gender can be found in the educational system and in the labor market, marriage may be segregated by age or social class, neighborhoods by income and/or ethnicity, and schools by both class and ethnicity. Segregation patterns may be harmless, such as an uneven distribution of dog-owners or the elderly in parks, but they may also lead to social structures that may be considered harmful. Segregated residential neighborhoods in major cities is an example of the latter,
where certain neighborhoods become associated with less available recreational activities, higher rates of unemployment, and poorer health (Sampson et al., 2002).

There is an extensive sociological literature on residential segregation, compared to the more limited amount of literature on school segregation, but the two phenomena are closely related. Massey & Denton (1988) argues that residential segregation can be understood as a macro phenomenon varying along five main axes of measurement: evenness, exposure, concentration, centralization and clustering. Evenness describes the distribution of minority members; they may be overrepresented in some neighborhoods, and underrepresented in others. This in turn affects exposure, or how much contact there is between majority and minority members due to spatial constraints, e.g. geographical distance. The degree of concentration increases when minority members are located within a small area compared to the majority, and this area may for instance be located around the urban core, leading to a higher degree of centralization. Finally, clustering describes the degree to which a minority is concentrated on few areas close to each other, or spread more widely among several neighborhoods. Although these measurements usually overlap empirically, it is argued that they are conceptually different (Massey & Denton, 1988, p. 282).

How do these general measurements of segregation apply when discussing school segregation? Minority members may be distributed so that they are over-represented in some schools and underrepresented in others, varying on the characteristic of evenness. This is perhaps the most central measurement of school segregation. This also affects the degree of exposure: Minority members may be distributed so that they rarely share schools with majority members, limiting exposure. The dimension of concentration measures the amount of physical space a minority occupies compared to the majority, and may be translated into the size of the schools where the ethnic minorities constitute the majority of students. In addition, these schools may also be centralized, for instance located generally closer to the urban core of a city. Finally, the measurement of clustering tells us to which degree the minority schools are located close to one another, compared to being scattered around various parts of a city.

As mentioned, evenness is perhaps the most central characteristic when measuring school segregation, i.e. the «differential distribution of two social groups among areal units in a city», where in the case of school segregation, areal units refer to schools. An uneven distribution constitutes segregation: «Evenness is maximized and segregation minimized when all units have the same relative number of minority and majority members as the city as a whole» (Massey & Denton, 1988, p. 284). In the case of schools, segregation is minimized when all schools have the same relative number of
minority and majority students as the age cohort currently enrolled in the school system as a whole (i.e. when the proportion of minority students is equal in all schools). In this thesis, the concept of segregation will refer to the lack of evenness in the distribution of majority- and minority-background students in schools.

1.2 An Agent-based Model of Ethnic Segregation in Schools

In 2014, Victor Stoica & Andreas Flache (2014) approached the question of how ethnic segregation in schools is generated by developing an abstract agent-based model representing households and schools. This model will be discussed more thoroughly in section 3.3, but its importance to the thesis implies that a short summary is required already at this point. The model represents a given number of schools located on a spatial grid, similar to a large chessboard, surrounded by a number of households belonging to one of two groups, representing households with either an ethnic majority or an ethnic minority status. Each household is assumed to have exactly one child of the age of enrollment in elementary school. Initially, the households are assigned a school in accordance with the catchment area (or school district) they inhabit, which corresponds to their closest school. Thus, the level of ethnic segregation initially matches the residential segregation of their neighborhoods. As the simulation is run, each household makes a decision regarding whether or not to change the school they have been assigned. This decision is determined by a utility function, where their preference for the ethnic composition influences their choice, along with their geographical distance to a given school. If the utility calculated through these preferences drops below a certain threshold value, the household will attempt to select a different school which would yield a higher utility.

This kind of model will allow us to closer investigate two mechanisms, and interplay between these mechanisms, in the generation of segregation patterns. These mechanisms are the two preferences among households in the model: The preference for a certain ethnic composition in schools, and the preference for minimizing home-school distance. The latter preference constitutes the effect of residential segregation on the patterns of school segregation. In addition, the model can be used as an experimental lab: Suggested policy changes, such as a removal of the “district school” policy, can be implemented, and the results examined. Initializing the model based on empirical data may also lead to interesting new results. This allows us to monitor in detail how changes in theoretical assumptions affect levels of segregation and, in turn, gives potentially valuable insight into the relative strength of
effects of the mechanisms involved in the simulation. As Stoica & Flache (2014) argues: «[...] integrating empirical information with theoretical models of segregation is a promising direction to assess competing explanations of school segregation (e.g. a preference based vs. a residential segregation based account) and to explore theoretically what the effects of policies could be that have been proposed to counter school segregation.» In addition, the empirical initialization allows for comparing simulation results with the actual level of ethnic segregation found in empirical data from Oslo more directly. If the model fails to generate qualitatively similar results, we may conclude that either: a) The mechanisms and assumptions made in the construction of the model are wrong, or b) that implementing additional mechanisms is required to capture the dynamics generating segregation patterns. For this thesis, the main purpose of the empirical initialization is to investigate whether or not the central dynamics operate in a similar way compared to the purely abstract model; however, I argue that more fruitful comparison with, and actual generation of real-world segregation patterns might be possible with further extension of the model.

What can we expect to find in such a model? The preference-based segregation dynamics emerge in Schelling’s model can be expected to appear in this model as well. Additionally, as shown by Stoica & Flache, we may expect to find that preferences for minimizing distance to schools curbs segregation patterns that might occur without these preferences (by discouraging "white flight"- the tendency of ethnic majority-group members to avoid schools with a high share of minority-group members). An agent-based model offers potential for insight into the interplay between different mechanisms producing macro-patterns of segregation, and provides a tool for “bottom-up” explanations of such complex phenomena. In addition, the agent-based model may guide further empirical research, by showing which mechanisms are and are not relevant in generating outcomes similar to those we observe in the actual population.

1.3 The Contribution of This Thesis

The goal of this thesis is to theoretically address the impact of parental choice on school segregation. The contribution of my thesis is twofold. I will replicate an existing agent-based model, created by Stoica & Flache (2014), and extend this model further. The extensions revolve around two main assumptions of the model: How households’ preference for minimizing distance is modeled, and how the temporal aspect figures into the simulation.

The first extension is motivated by the idea that parental preference for minimizing distance to schools can be understood not only as a calculation where the closest schools yield higher utility, but
also as a distance threshold. Instead of assuming that schools are simply more attractive the closer they are to home, it is possible to assume that households operate with a “threshold”; when a school is located outside of a certain distance (perhaps representing a need for public transportation to arrive in time, or the need to change bus lines), it is excluded from the perceived options when parents are selecting a school. This “search radius” is specified when initializing the model, thus allowing for assessing the impact of choosing different “search radius” settings. I assume that if a preference for minimizing distance has a curbing effect on segregation patterns, modeling preferences using either strategy should generate the same results.

The second extension allows for letting the simulation run over what represents multiple years, and not just one initial enrollment. A characteristic aspect of segregation-producing processes is that they are potentially self-reinforcing over time. I argue for the assumption that this applies for school segregation as it does for residential segregation; the choices of previous households affect the ethnic composition of schools, thus affecting the choices of future households when enrolling their children. This process of self-reinforcement is usually referred to as segregation dynamics. Building on the framework provided by Stoica & Flache’s model, I allow for simulating not just one, but several years of school enrollment. This makes it possible to simulate dynamic segregation-generating processes over time. Furthermore, the model will be initialized with empirical data from Oslo, which offers an opportunity for investigating how the model behaves under more realistic initial conditions.

### 1.4 Structure of the Thesis

The thesis is divided into six main sections, including this introductory chapter. Section two provides an overview of research in the field of segregation, broadly construed, and more specifically residential segregation and school segregation. I discuss both the potential causes and consequences of segregation. In chapter three, I present an outline of analytical sociology and the mechanism-based approach towards sociological explanation. This meta-theoretical perspective is included in order to motivate the agent-based approach, and to embed agent-based modeling in a sociological context. This is followed by an overview of agent-based modeling, and a discussion of some social science examples of agent-based models, including Thomas Schelling’s seminal model of residential segregation. In section four, I introduce my proposed agent-based model of school segregation and the technical implementation of this model. Section five presents the actual analysis in two steps: First, I analyze how the abstract, theoretical model performs. Second, I do an analysis of the model, empirically calibrated with data from Oslo. Finally, in section six I summarize and discuss the results, and offer a
conclusion.
2 Previous Research

In this chapter, I discuss research on the mechanisms that produce patterns of segregation. In section 2.2 I will outline some central explanations for the persistence of ethnic residential segregation. I will then discuss whether or not these explanations are applicable when discussing school segregation patterns, followed by an overview of research done on the causes of school segregation more specifically. Finally, I will discuss research on the potential effects of an ethnically segregated school system. Before proceeding with this discussion on previous research, it is important to describe both the Norwegian school enrollment system and the patterns of school segregation observed in Oslo. Thus, I will begin with an overview the structural conditions governing school segregation in Oslo: The rules for enrollment and switching schools, patterns of residential segregation, and how these patterns correlate with patterns of school segregation. The agent-based model I will present in chapter 4 builds on these rules of enrollment, and the empirical data will be used to inform the model in final analysis of chapter 5.

2.1 Enrollment Policy and Parental Choice

Investigations into parental choice of schools are often motivated by the enactment of new school choice policies, such as going from a restricted policy to a free-choice policy (see Andersen, 2014; Taylor & Gorard, 2001). If one is to consider the potential effects of parental preferences for school segregation, there is a natural demand for there to be at least some degree of freedom of choice when parents are enrolling their children into the educational system. The degree of parental freedom in choosing a school is critical for the study of how segregation patterns emerge: With no freedom of choice, ethnic segregation in schools would simply be a product of residential segregation. Free choice, on the other hand, allows for parental preference to play its part in generating school segregation beyond existing patterns of residential segregation.

How is the situation regarding parental freedom in Oslo? Norway's elementary school enrollment system is characterized by a varying degree of parental choice, and this is due to the fact that these policies are managed at the municipality level. The Norwegian Union of Education gives the following description: «Students may be transferred to a school other than their local school when need and reason for such a transfer is stated. In most municipalities it is common that these applications are
approved insofar there is available capacity at the target school of the transfer» (my own translation). In principle, this implies a high degree of parental choice, although with a certain deterrent invoked by the bureaucracy involved in the process of changing schools. In addition, parents are guaranteed an available spot for their child at their designated local school. In the case of Oslo, this means that parents are allowed to freely choose schools for their children, as long as they actively engage in the switching/application process, and as long as there is capacity at the target school.

Unfortunately, there are no available centrally gathered data on how many elementary school students are actually enrolled in a different school than their designated local school, but qualitative research indicates that the phenomenon is widespread; Morken (2008) conducted a qualitative survey of parents of the Grorud area of Oslo, and found, among other things, that a large share of those questioned had either considered or actually engaged in switching schools for their children. Even though the Grorud area of Oslo is not representative of the city as a whole, the numbers can still be interpreted as an indicator that school switching is a rather widespread practice. This survey will be discussed more thoroughly in section 2.2.

2.1.1 School Segregation and Residential Segregation

In a system for school enrollment structured as in the case of Oslo, patterns of school segregation will naturally be affected by already existing patterns of residential segregation. Even under the assumption that a significant number of parents engage in enrolling their children in schools other than their local school, the available capacity of schools will still limit the degree of cross-district mobility. In this section, I will attempt to address how closely the patterns of school segregation correspond to residential patterns in Oslo. Even if there is no centrally gathered data available on the prevalence of school switching among parents, the relationship between patterns of residential segregation and school segregation can still provide some insight.
The above figure shows a scatter plot of all lower secondary schools in Oslo in 2014. The vertical axis shows the school’s share of minority-language students, while the horizontal axis shows the school’s share of minority-language residents (aged below 16) surrounding the school. This share is calculated using a $k$-nearest-neighbors approach based on the schools’ geographical locations, for two values of $k$: 500 and 1000. Linear regression lines are added for clarity. The data on the ethnic composition of schools is drawn from the municipality of Oslo (Statistics Norway, 2014, 8.11), while the data on the distribution of minority-language speakers in surrounding neighborhoods is drawn from GIS-coordinate census data. It is important to note that this is only an approximation of the patterns under...
scrutiny, as the $k$-nearest-neighbors calculation does not identically match the actual attendance boundaries of the schools. A more fine-grained approach would be to follow the lines of Saprito and Sohoni (2006), where mapping software is used to link school attendance boundaries with census data. Unfortunately, such an approach is beyond the scope of this thesis. In addition, I calculate numbers based on the entire population of children aged below 16 in the surrounding area: A more fine-grained approach would be based on the exact cohorts enrolled in the schools in the specified years. Still, as an aggregation meant for addressing the macro-level patterns, the above calculation should be able to provide some insight.

The purpose of this plot is to assess how closely residential segregation and school segregation correlate in Oslo. The correlation coefficients between the two variables (the share of minority-background individuals in the surrounding neighborhood and the share of minority-background students in the schools themselves) are 0.93 (for $k = 1000$) and 0.87 (for $k = 500$). This tells us that the level of school segregation is strongly correlated with patterns of residential segregation, yet there is still a portion of unexplained variance. The unexplained component is what this thesis addresses; how households’ choices of schools, given fixed residential locations, might affect school segregation patterns. If all parents were satisfied with the school they were allocated by default, the points on the above graph would be located very close to the regression line – a schools share of minority-language students would closely match the share of minority-language students of its surrounding neighborhood. Conversely, by observing the above graph, we can conclude that, while residential patterns go far in explaining school segregation, parental choice must be included in order to fully explain the variance between the ethnic compositions of schools. It is also worth noting that even if the share of minority-background students of each school would closely match the share in its surrounding residential area, we still would not be able to dismiss the potential significance of parental choice of schools: aggregate data does not tell us whether or not individual-level mobility between schools still occurs, even if patterns appear stable on the aggregate level.

Finally, I assess the overall levels of school segregation in Oslo using the Index of Dissimilarity (a measurement of distributional evenness that will be discussed in greater detail in chapter 4) for all of the secondary elementary schools shown in the above plot. The value of this index is 0.25, which can be interpreted as the share of either majority or minority-background students that would have to relocate for the distribution to be perfectly even across schools.
2.2 The Causes of School Segregation

Compared to the smaller literature on the causes of school segregation, residential segregation has historically been a more frequent target of sociological investigation. I argue that even though school and residential segregation are distinct phenomena, they obviously also have many features in common. Thus, an overview on research on residential segregation is required even when primarily discussing school segregation. In addition to the two phenomena being similar, residential segregation also constitutes a large component in the shaping of school segregation – especially in enrollment systems based on the idea of a local or district school with a corresponding catchment area.

2.2.1 Explanations of the Persistence of Ethnic Residential Segregation

As discussed in the introductory chapter, segregation patterns can be viewed as complex macro-level phenomena generated by several possible mechanisms and the interplay between them. Most research on the persistence of ethnic segregation stems from the U.S., which might be attributed to the overall higher degrees of residential segregation compared to European countries (Zorlu & Mulder, 2008, p. 246). Quillian (2002) argues that there are three main explanatory mechanisms of ethnic residential segregation commonly found in the literature on segregation in American cities. These can be summarized as follows:

1. Whites prefer predominantly white neighborhoods, and blacks prefer to live in neighborhoods with at least a certain amount of other blacks. These preferences together generate patterns of residential segregation.
2. White neighborhoods are preserved through the exclusion of blacks through various barriers to entry: discrimination among real-estate agents, violence against black newcomers etc.
3. Racial disparities in income or wealth, coupled with neighborhood differences in housing prices, causes segregated neighborhoods.

By analyzing migration patterns in U.S. cities, Quillian finds evidence that mainly supports the first explanation, in particular a preference among whites for predominantly white neighborhoods: “Whites appear to strongly desire to avoid more than a few black neighbors and move in patterns so as to realize this preference” (Quillian, 2002, p. 220). This suggests that white avoidance of neighborhoods that are
predominantly black is an important explanatory mechanism for the persistence of ethnic residential segregation. This conclusion is also supported by Crowder (2000), who finds that that:

The mobility decisions of White metropolitan householders are directly influenced by the racial conditions of the neighborhoods in which they live. The likelihood that individual White householders will leave the neighborhood during an annual interval increases with the size of the total minority population in the neighborhood, and the effect is strongest in the types of neighborhoods in which the vast majority of Whites are located. (Crowder, 2000, p. 251)

These studies both underline the fact that, while barriers to entry in certain neighborhoods and racial disparities in income or wealth most likely influence the persistence of residential segregation, the preferences for certain ethnic compositions in neighborhoods are of vital importance. In addition to white avoidance of predominantly black neighborhoods, often characterized by the term “white flight”, there is also evidence that supports a complementary tendency of blacks avoiding neighborhoods that are predominantly white. Crowder and South (1998, p. 25) identify significant movement of blacks into already predominantly black neighborhoods as an important factor in explaining residential segregation. Overall, there is reason to assume that a complex interplay of heterogeneous preferences needs to be included in an explanation of residential segregation.

Another important aspect contributing to residential segregation, in addition to mobility between neighborhoods, is the behavior of newcomer immigrants arriving in a country. This perspective is perhaps particularly relevant when discussing western European countries, due to the fact that large-scale immigration is a more recent phenomenon compared to the situation in the U.S. Zorlu and Mulder (2008) conducted a study on the determinants of initial and subsequent location choices among immigrants arriving in the Netherlands, and their results largely coincide with the observations made by researchers on residential segregation in the US. The main determinant of the choices made by arriving immigrants in the Netherlands is the ethnic composition of the neighborhoods: “Non-Western immigrants not only settle initially in more ethnically segregated neighborhoods, but also tend to migrate to segregated neighborhoods” (Zorlu & Mulder, 2008, p. 263). Thus, migration behavior, in the form of immigrants concentrating in areas where other immigrants already constitute a significant presence, can be a large contributor to the growth and persistence of residential segregation.
2.2.2 Explanations of the Persistence of School Segregation

As the studies of migration and residential patterns discussed above point to, there is reason to believe that individuals’ preferences for, or aversions against, certain ethnic compositions of neighborhoods is central to explaining segregation patterns. Explanations of school segregation usually also focus on these preferences, as the phenomenon can be understood as a product of the same mechanisms that produce residential segregation. When discussing how preferences among actors shape school segregation, however, an additional central preference needs to be taken into account: That of minimizing distance between home and school. The importance can be assumed to vary along the age of the children being enrolled in school; the younger the children, the more important distance becomes, for practical reasons. Teenagers, however, are more “mobile”, in the sense that they are able to commute greater distances independently. As noted in chapter two, there is a strong correlation between the composition of the population surrounding schools and the schools’ student composition. This can be attributable to the fact that minimizing geographical distance is important when parents are choosing schools for their children: If otherwise equally attractive alternative schools are available, it is reasonable to assume that parents’ opt for the school that is closest to home. Several studies of the deciding factors in parents’ choice of elementary schools also confirm this assumption (see Hunter, 1991; Hughes et al., 1993; Morgan et al., 1993).

In addition to the importance of geographical distance, one needs to examine which other preferences matter when attempting to explain school segregation. How does one approach the question of uncovering the motivation of parents when choosing schools? One approach is conducting either qualitative or quantitative surveys. An example of such an approach from Oslo can be found in Morken (2008), where parents in the Grorud area of Oslo were interviewed regarding their motivations for switching schools for their children. Although the research question is conceptually different from the motivations for initial enrollment into a certain school, one can assume that the preferences governing the decision are the same as those involved when initially choosing a school. Morken finds that the most common reasons stated for switching schools are not directly related to the ethnic composition of schools; difficulties in communication between the previous school and the household, bullying, the local reputation of the school, and issues regarding specific teachers. In fact, nearly all of those interviewed were explicit supporters of a multicultural school. On the other hand, some of these motivations are specific for the actual switching of schools, and thus inapplicable when discussing the initial choice of a school; relationships with specific teachers, feelings of social exclusion et cetera. It is possible to argue that a preference for ethnic composition would manifest itself before choosing a
school, and not be of consequence when this choice has already been made – i.e. if one later wishes to move ones child to a different school.

Kristen (2008) argues for understanding the choice of schools as a three-step sequential process consisting of the following stages: The perception of different school alternatives, the evaluation of the perceived alternatives, and finally access to the selected school (Kristen, 2008, p. 498). The perhaps most decisive part of the school choice process is the evaluation of perceived alternatives. This is where parental preferences come in, and one such preference is the preference for a certain ethnic composition of students in a given school. After analyzing survey data from the Essen region in northern Germany, Kristen points to several possible mechanisms that contribute towards school segregation: “[...] ethnic differences in school choice may stem from individual preferences for segregation or from school representatives’ ‘tastes for discrimination’” (Kristen, 2008, p. 507). The latter would not affect ethnic segregation in the schools of Norway, due to the nature of the enrollment system (school representatives are assumingly not in the position to act on any such tastes for discrimination) but the former, individual preferences for segregation, is recognizable from the literature discussed above. It is worth noting that Kirsten’s study mainly details the segregation of students of various ethnic origins into either public or denominational schools, the latter being somewhat particular to northern Germany. Nevertheless, the general mechanisms described might still apply to school systems elsewhere.

Karsten and colleagues (2003) argues that there are fundamental methodological problems when attempting to determine to which extent the ethnic composition of schools is considered when parents are deciding on schools for their children. In surveys investigating parent’s motivation for choosing a certain school for their children in general, as in Morken’s study of school switching motivation, ethnic composition is rarely mentioned as deciding factor. There are reasons to believe that such survey results are potentially biased, in that parents might wish to avoid socially unacceptable answers. A study conducted by Schneider & Buckley (2002) underlines this line of reasoning, by pointing out that there is a larger effect of race and class on school choice behavior than what parents admit in survey data. By studying a website where parents find information regarding their choice of school, they conclude with the following evidence:

In Schneider et al.’s (1998) study of expressed preferences, for example, less than five percent of the parents who were surveyed said that the race and economic background of the students in a school were among the most important characteristics of schools. Yet almost 30% of parents looked at student
demographic information early in their visit to DCSchoolSearch.com, making it the modal "response" category. (Schneider & Buckley, 2002, p. 138)

This leads them to concluding that using survey data as an approach to studying the motivation of parents choosing schools may have certain flaws. Few might want to admit that the ethnic composition of schools plays a part in the choice, while stating reasons such as test scores or the reputation of teachers instead:

Thus, when we move our research technique away from surveys, in which social desirability clearly affects response patterns, to more anonymous search behavior, the results are not as optimistic as those based on survey data. Moreover, this search behavior is congruent with preferences revealed by the studies of actual choice behavior-parents do care about academics but they also care very much about school demographics - something they will not admit to verbally. (Schneider & Buckley, 2002, p. 141)

Letting preferences for the ethnic composition of students matter in the choice of school may possibly be perceived as racism. This, in turn, might lead to parents stating other reasons for their choice of schools in surveys, and give biased results. In other words, what might seem as a plethora of deciding factors stated among parents for their choice of schools might mask the potential effect that ethnic composition actually has on the decision. Unfortunately, studies of searching patterns like the one conducted by Schneider & Buckley are of limited prevalence due to the difficulty of obtaining such data – in part due to privacy reasons. Alternatively, another approach differing from that of surveying motivations is to study the actual choices made by parents: Glazerman’s (1998) study of parents in Minneapolis revealed that instead of making choices based on the academic performance of the schools, as is often cited as a motivation in survey data, other factors served as better predictors: “Traditional measures of academic quality, test score levels and value added indicators, had little or no predictive power. Nor were choices influenced much by school safety or neighborhood quality, two factors frequently cited in the author's informal interviews. Race, distance, and SES on the other hand, had strong effects” (Glazerman, 1998, p. 20-21).

In summary, I argue that, based on the literature discussed above, school segregation may be understood as a potential product of two central mechanisms: First, school segregation is partly a product of residential patterns, mediated by a preference among parents for minimizing home-school distance. Second, a preference regarding the ethnic composition of schools seems to be of central importance when explaining the growth and persistence of school segregation patterns. Finally, the
literature underlines some methodological difficulties in investigating the motivations of parents, and the role of these motivations in the formation of school segregation.

2.3 The Potential Effects of School Segregation

While this thesis is concerned with identifying the causes of school segregation, a short discussion on its potential effects is included as well. The literature on the effects of school segregation is sparse, and most of it deals with segregation in the USA. The main reason for this is that the massive waves of immigration into western European countries constitute a relatively recent phenomenon, and the long-term effects of segregated schools may still not be fully developed. As an example, the descendants of immigrants constituting the large waves of immigration into the Nordic countries have only recently begun their entry into higher education and the labor market. What follows is an overview of research done to map the possible effects of school segregation.

As the empirical literature on the effects of school segregation grows, there is still little research done on the long-term effects, due to the problems mentioned above. Hermansen & Birkeland (2015) investigate these peer-effects among Norwegian students. Using register data, they find a modest positive effect of having immigrant classmates, controlling for school fixed-effects. This effect is stronger among students with immigrant backgrounds themselves:

The results show that students in cohorts with higher immigrant shares have a slightly higher likelihood of completing upper secondary education relative to students in cohorts with lower immigrant shares within the same school. This positive impact of immigrant classmates is substantially stronger among immigrant students than native students. (Hermansen & Birkeland, 2015, p. 24)

Their findings thus points towards the fact that the common concerns regarding immigrant-majority schools primarily revolve around socioeconomic status instead of ethnicity. These results also converge with a study of the educational trajectories of Turkish and Moroccan-origin youth in several European countries, conducted by Baysu and de Valk (2012). The researchers find that “In more open educational systems such as Sweden and Belgium, the segregation experience is less negative, absent or even positive for the academic school careers of the Turkish and Moroccan second generation» (Baysu & de Valk, 2012, p. 793-794).

Why are schools perceived as such important arenas for integration, and consequently, as sources of concern when segregated? One important factor is that social relations are essential
components of the education process: Students form lasting friendships with each other in schools, and if these schools are ethnically segregated, the friendship patterns may be ethnically segregated as well – leading to weakened integration overall. While Hermansen & Birkeland focus on the long-term effects on academic enrollment and achievement, Moody (2001) investigates how school segregation affects friendship segregation. He finds a complex relationship between ethnic heterogeneity and the degree of friendship segregation: There is a strong, positive relation between ethnic heterogeneity and friendship segregation, but this relation does not follow a linear pattern. Up to a certain point, heterogeneity increases friendship segregation, but when the levels of heterogeneity reach a certain threshold, the effect is reversed (Moody, 2001, p. 707-708). This relationship can be attributed to two interconnected mechanisms:

On the one hand, majority members may start to see minorities as a potential status threat once their numbers increase significantly. On the other hand, increasing numbers allow minorities to identify same-race friends that match on other attributes, leading to an increase in same-race friendship choice within the minority group. Once started, this process likely snowballs as increased same-race preference within one race solidifies the group structure within that race, making them appear more unified (and thus a greater status threat) to the other group. (Moody, 2001, p. 708)

As a minority group grows in size, it is thus both easier for students to find similar students within their ethnic group (i.e. potential friends), and easier for students to perceive members of the “other” group as just that; “others”, and a potential threat to status. This effect is weakened when heterogeneity reaches very high levels, such as in schools where there are several distinct ethnic groups of notable size: Race-specific preferences for friendships might have a “unifying” effect on the whole school, compared to schools where there are only two large groups with more explicit social structures and hierarchies.

2.4 Summary

In this chapter, I have discussed previous research on the causes and effects of ethnic segregation, both in general and in schools. Research indicates that in order to understand patterns of school segregation, one needs to understand the motivations and choices among parents’ and their children. Uncovering these motivations and preferences through surveys or interviews is a challenge, as noted by Glazerman (1998) and Schneider & Buckley (1998). In the next chapter, I will present a theoretical foundation for understanding segregation patterns, along with an overview of my methodological approach for this
thesis, agent-based modeling. Such a model represents an alternative approach to the explanation of school segregation, which can be employed in addition to surveys of motivation among parents and empirical studies of actual school choice behavior.
3 Theoretical Perspectives and Methodology

How do we understand segregation patterns? There are several features of segregation-generating processes that appear in a variety of contexts, and therefore seem to have some level of generality. One of these is that the segregation-generating processes are complex: Segregation patterns are macro-level phenomena that cannot be understood as a simple aggregation of individual action. To understand segregation processes, one has to take into account the interplay between heterogeneous actors and their beliefs, actions and preferences. A second common characteristic of segregation processes is that they tend to be self-reinforcing. The macro-level patterns of segregation may affect the process that generates them: For instance, a neighborhood becomes more attractive as affluent people move in, leading to an increase in housing prices. This, in turn, makes it difficult for the less affluent to afford to move to the neighborhood, thus fueling further segregation by income. Some patterns of segregation are coordinated and intentionally created: For instance, US schools in the 1950s were racially segregated by law. More common though, is that there is no central organizing entity influencing the generation of these patterns: In neighborhoods as well as in schools, segregation arises as a product of the behavior and preferences of individual actors.

When discussing macro-level phenomena that arise through micro-level interaction with no central coordination, the concept of emergence is central. The term was used by Durkheim ([1901] 1982), and describes macrosocial phenomena that consist of more than just their parts. Emergent properties are properties of a given system that cannot simply be explained as an aggregate of individual behavior. Durkheim uses the hardness of bronze to illustrate the concept: It “[…] lies neither in the copper, nor the tin, nor in the lead which have been used to form it, which are all soft or malleable bodies. The hardness arises from the mixing of them” (Durkheim, [1901] 1982, p. 39-40). In the same way, the mixing of individuals and their behavior can result in a social system with unforeseen emergent properties such as high levels of segregation. The question of emergence can be summarized as follows, inspired by Joshua Epstein: How do ensembles achieve the functionalities (or properties) their constituents lack? (Epstein, 2006, p. 2). In order to understand complex macro-phenomena such as segregation patterns, one thus needs to consider the interaction between individuals, not just these individuals’ preferences and motivations.

In this chapter, I will first present the strategy for sociological explanation referred to as analytical sociology. I will discuss the concept of methodological individualism and mechanism-based
explanation, focusing on the explanation of emergent macro-level phenomena such as segregation patterns. The inclusion of this meta-theoretical perspective is motivated by the nature of my methodological approach: Agent-based modeling builds on certain core theoretical principles, which I argue coincide well with the theoretical foundation found in analytical sociology. In the second part of the chapter, I provide an outline of agent-based modeling as a tool for sociological research, including a discussion of three influential examples of agent-based models employed within the social sciences. Finally, I will present an agent-based model of school segregation developed by Stoica & Flache (2014), which will serve as the foundation of my own model presented in chapter 4.

3.1 The Analytical Approach

In order to approach these complex, emergent phenomena, a general strategy for developing theory is useful. The approach found in analytical sociology is one such strategy.

Analytical sociology is concerned precisely with examining, and explaining, complex systems and their emerging patterns, such as those of segregation. The call for an analytical approach to sociological theory, based on explanatory social mechanisms, stems from the notion that much of social theory primarily deals with conceptualizing, labeling and relabeling the social world, instead of actual explanation (Hedström & Swedberg, 1998, p. 1). Proponents of analytical sociology argue that social theory should move beyond descriptive accounts, and shift focus towards explanation. These explanations center on the «cogs and wheels» of society, or the mechanisms that bring about certain macro-level phenomena: “A sustained focus on explanatory social mechanisms would allow sociological theory to re-connect with what we consider to be its most promising and productive era – namely, middle-range sociology of the kind that Robert K. Merton and Paul Lazarsfeld tried to develop at Columbia University after World War II” (Hedström & Swedberg, 1998, p. 1). Instead of basing explanations on the assumptions that the macro-level is simply the aggregate of micro-level preferences, or on the contrary, that macro-level patterns can be understood as products of other macro-level properties without considering micro-level interaction, one needs to explicate the social mechanisms that produces a given macro-level phenomenon.

I will now provide a brief overview of the analytical approach, mechanism-based sociological explanation, and Robert K. Merton’s notion of middle-range theory. This overview is included in order to establish a theoretical foundation for the agent-based approach, which will be discussed in section 3.2.
3.1.1 Methodological Individualism and Mechanism-based Explanations

Analytical sociology is a strategy for understanding the social world based on some theoretical and methodological principles. One of these principles is the principle of methodological individualism. Individualism is defined as “a methodological doctrine according to which all social facts, their structure and change, are in principle explicable in terms of individuals, their properties, actions, and relations to one another” (Hedström & Bearman, 2009, p. 8). In contrast, a structuralist approach to explanation sees social structures as the theoretical starting point. Methodological individualism entails that explanations should be founded on individuals, their behavior, and the interaction that goes on between them. The principle can further be divided into two versions: A “strong” version, where all reference to structural phenomena is excluded from the explanans, and a “weak” version where such explanations are deemed unrealistic. The latter version forms the foundation of analytical sociology.

Social structures are included in explanations, because a pure rock-bottom explanation is considered unrealistic: “...the realism and the precision of the proposed explanation will be greatly improved if we take certain macro-level properties such as relational structures as given and incorporate them into the explanation” (Hedström & Bearman, 2009, p. 10). Thus, (weak) methodological individualism does not exclude macro-level properties from explanation, as one would from a purely (strong) individualistic view. As noted in Hedström & Swedberg (1998): as the current social world is made up of “[...] causal histories of nearly infinite length, we can only hope to provide information on their most recent history” (p. 13). Hence, in order to provide reasonable explanations based on methodological individualism, certain macro-level facts or properties must be taken into account. Note that the principle of weak methodological individualism is sometimes referred to as structural individualism by some advocates of analytical sociology, for instance in Hedström & Bearman (2009): I will make no distinction between these terms for the purposes of this thesis, and consider the terms equivalent.

The explanations called for from this perspective are based on social mechanisms. Hedström and Bearman reviews several alternative definitions of a social mechanism (Hedström & Bearman, 2009, p. 6), and find they share common characteristics. A mechanism is a causal chain leading to the social fact that is to be explained. It describes actors, their actions, and the relations between them. A mechanism “...refers to a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome, and we explain an observed outcome by referring to the mechanism by which such outcomes are regularly brought about” (Hedström & Bearman, 2009, p. 5). This distinguishes mechanisms from societal laws or other general theoretical constructs; they act with a certain regularity, but can bring about various outcomes under different conditions. At the same
time, they are distinguishable from simple description by pointing out the causal chain that brings about a social phenomenon.

Elster (1998) distinguishes between two main types of mechanisms, which he coins type A mechanisms and type B mechanisms. The latter “arise when we can predict the triggering of two causal chains that affect an independent variable in opposite directions, leaving the net effect indeterminate” (Elster, 1998, p. 46). An example of such a mechanism, which Elster borrows from Cartwright (1983), would be the case of planting a certain type of flower in both rich and warm soil. The fact that the soil is rich would cause the plants to thrive, but the heat might cause them to perish. We do not know the net effect on the plants, but after observing the outcome, we can point to the causal chain as an explanation of what happened. Type A mechanisms, on the other hand, “arise when the indeterminacy concerns which (if any) of several causal chains will be triggered” (Elster, 1998, p. 46). This is exemplified by the fight-or-flight response found in many animals when exposed to shock or pain: The same condition triggers one of the two distinct responses.

As briefly mentioned above, analytical sociology aims to continue a development based on the principle of middle-range theory, as known from Robert K. Merton, and social mechanisms can be seen as the foundational building blocks of middle-range theory. Merton describes middle-range theories as follows: “Theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization and social change” (Merton, 1968, p. 39). Middle-range theory is thus more general than empirical explanations of specific social facts, but less general than theories aimed at describing the whole of society. Merton thereby distances his view from grand theory as found in structuralism on the one hand and from thick descriptions on the other hand. Sociology, in middle-range theory terms, can be viewed as an arsenal of semi-general theories, or mechanisms, which collectively help us explain various social facts. These mechanisms are not simply ad-hoc explanations of a specific social phenomenon: Mechanisms gain their explanatory power from their generality (Hedström & Swedberg, 1998, p. 10).

Social mechanisms can be classified into different categories, depending on which level they operate at. Hedström and Swedberg (1998) have developed a typology of mechanisms, illustrating this by employing the familiar image of Coleman’s “Bathtub”:
Coleman’s image underlines the primacy of methodological individualism by showing how the micro-level needs to be taken into account when explaining macro-level associations. Any social system or continuous social action can be conceptualized as a chain of these macro-to-micro-to-macro transformations, and a satisfactory explanation rests on understanding each step of each such chain. As Coleman himself argues, explanations should account for how the actions and motivations of individuals give rise to social organization, instead of assuming that macro-level phenomena are simple aggregates of individual characteristics and behavior (Coleman, 1994, p. 197). The three broad categories of mechanisms are termed situational mechanisms, functioning from the macro to the micro level, action-formation mechanisms, functioning at the micro-level, and finally transformational mechanisms, constituting the link back from the micro to the macro level.

A situational mechanism operates from the macro-level to the micro-level, more specifically, from any social structure onto the individual. “The individual actor is exposed to a specific social situation, and this situation will affect him or her in a particular way” (Hedström & Swedberg, 1998, p. 23). Erwing Goffman’s (1963) work on social interaction in public spaces is cited as an example of theory with situational mechanisms as its core. Action-formation mechanisms are mechanisms that
function entirely at the micro-level, detailing how “[…] a specific combination of individual desires, beliefs, and action opportunities generate a specific action” (Hedström & Swedberg, 1998, p. 23).

Finally, transformational mechanisms detail how individual action and interaction at the micro-level generate change or persistence of structures at the macro-level. These mechanisms can be constituents in explanations of how individual behavior lead to unexpected macro-level consequences: “Here a number of individuals interact with one another, and the specific mechanism (which differs depending on the nature of the interaction) shows how these individual actions are transformed into some kind of collective outcome, be it intended or unintended” (Hedström & Swedberg, 1998, p. 23). In our case, transformational mechanisms are of particular interest, because they help us understand how complex macro-level phenomena such as segregation can be brought about. In order to properly explain a macro-level social pattern, all categories of mechanisms need to be taken into account: a macro-level association alone can be of value as a description, but true explanation needs to explicate the micro-level mechanisms bringing the macro-level explanandum about.

In order to illustrate how the three types of mechanisms discussed above coexist and influence one another, one can consider Max Weber’s explanation of how the protestant work ethic contributed to the rise of capitalism (Weber, [1904] 2001). From the macro to micro-level, a certain religious doctrine instills a distinct set of values through situational mechanisms. These values, in turn, stimulate to a certain economic behavior through action-formation mechanisms. Finally, the new economic behavior generates a new macro-level outcome, capitalism, through transformational mechanisms. Relating this typology of mechanisms to the generation of school segregation patterns, we can also identify the three types of mechanisms at work. At the macro level, there are several factors which can be assumed to influence individual’s choices of schools: The ethnic composition of schools, residential patterns (for instance ethnically segregated neighborhoods), institutional rules detailing the degree of freedom when choosing schools, and several others. These factors all affect individuals - their beliefs, values and perceived alternatives - through situational mechanisms. Action-formation mechanisms turn these beliefs into a specific behavior of school choice, for example avoiding schools with a high share of students belonging to a different social group. Finally, this behavior leads to a new macro-pattern of school segregation through transformational mechanisms. This final link is often not a simple aggregate of individual behavior, but can, as Schelling’s model of residential segregation shows us (further discussed in section 3.2.3), be an unintentional bi-product of individual behavior.
3.1.2 Summary

The main advantage of a middle-range, mechanism-based approach to sociological theory is the explicit linking between *explanans* and *explanandum*: Identifying generative mechanisms “[…] helps us distinguish between genuine causality and coincidental association, and it increases the understanding of why we observe what we observe” (Hedström & Swedberg, 1998, p. 8-9). Without a focus on mechanisms, we are potentially left with what Hedström and Swedberg refer to as “black-box” explanations – where the explanatory mechanism detailing the relationship between two social entities is omitted. An example is research on the statistical relationship between social class and variables such as income or health. It is difficult to imagine social class as a causal agent directly influencing either of these variables; the statistical association between social background and income may tell us the overall “effect” of class on income or health, but it does not specify the causal chain producing this effect. In order to properly explain the relationship, the general mechanisms operating between class and other variables need to be specified:

A statistical “effect” of a class variable in contexts like these is essentially an indicator of our inability to specify properly the underlying explanatory mechanisms. The worse we do in specifying and incorporating the actual generative mechanisms into the statistical model, the stronger the “effect” of the class variable will appear to be. (Hedström & Swedberg, 1998, p. 11)

According to the analytical approach to sociology, the main purpose of sociologists is thus to identify and explicate mechanisms that can explain a social fact from the bottom-up. There is a call to move from variable-based to mechanism-based explanations in sociology. In the next section, I will discuss an approach to theoretical analysis that I will argue embodies the same principles found in analytical sociology, namely agent-based modeling.

3.2 Agent-based Modeling in the Social Sciences

Models offer us simplified means for analyzing empirical phenomena, by bridging the gap between theoretical assumptions and empirical reality: They are simplified representations of the real world (Lave & March, 1993). In the social sciences, statistical models are in widespread use. One common type of approach is that of various forms of regression analysis: Researchers formalize relationships between variables through mathematical functions, mapping regularities and correlations in empirical data. Models can also be used for attempting prediction of future states, but more commonly perhaps, *explanation* of present macro-level phenomena is often the main purpose of modeling. With the
massive changes to computing power and availability seen through the last thirty or so years, new approaches to modeling have grown more popular. One of the types of modeling that has increased in popularity and attention is agent-based modeling. The history of agent-based modeling can be traced back to the work of mathematicians John von Neumann and Stanislaw Marcin Ulam in the 1940s, later simplified by John Conway in his “Game of Life”-model (Tufte et al., 2010). In addition, the method shares roots with other computational forerunners such as system dynamics, microsimulation methods and cellular automata (Squazzoni, 2013, p. 19-20). In Norway, As early as in 1967, Norwegian sociologists Ørjar Øyen and Per Otnes (1967) expressed enthusiasm over the potential power of using what they called “machine simulation” in the social sciences, exemplified by a simple agent-based model representing the diffusion of rumors in communities.

In this section I will give an overview of agent-based modeling, focusing on its use within sociology. The first part of the chapter provides a perspective on the main goals behind agent-based modeling, how agent-based modeling differs from other kinds of models, and how these models are used in practice. The second part includes a brief overview of agent-based modeling within the social sciences, followed by a discussion on three examples of such models: Thomas Schelling’s model of residential segregation, Dean and colleagues’ model of the Anasazi tribal society, and Billari and colleagues’ “Wedding Ring” model of age-at-marriage patterns.

### 3.2.1 What is an Agent-based Model?

An agent-based model is a computer program where autonomous agents and their interactions are simulated following predetermined rules of the agents’ behavior. The theoretical starting point of any agent-based model is the agents themselves (often, but not always, representing individuals), their behavior, the interaction between agents, and the structures they are embedded in. Joshua Epstein considers agent-based models the starting point of a new approach towards social science in general, which he terms *generative* social science: Proper explanation of a macro-level phenomenon here rests on the ability to generate said phenomenon computationally. According to Epstein, agent-based models deal with answering the following question: “How could the decentralized local interactions of heterogeneous autonomous agents generate the given regularity?” (Epstein, 2006, p. 5) Regularities, in this case, describe any macro-level social pattern of interest. Although agent-based modeling is not necessarily to be understood as a new, separate paradigm for social science, the idea of *generating* to develop theory and explanation is central to the approach.
Agent-based modeling thus revolves around “generating” some macro-level regularity or system through modeling individual agents and their behavior. Using agent-based models in the social science is referred to by several equivalent terms: Generative social science (Epstein, 2006), computational sociology (Hummon & Fararo, 1995; Macy & Willer, 2002) or simply agent-based modeling. I argue that agent-based modeling should not necessarily be seen as a separate branch of sociological research, rather, agent-based modeling should be seen as one (of several) tools available to sociologists in researching emergent macro-level phenomena. Squazzoni (2012, p. 9-18) identifies six central ideas of sociology involving agent-based models. These ideas are the following:

1.) The primacy of models over grand theories and descriptive accounts: This aligns agent-based modeling with the mertonian notion of middle-range theory as discussed above. The main objective is to develop functional, analytical models; neither grand theoretical perspectives, nor rich descriptive accounts.

2.) The generative approach to explanation: Agent-based modeling gains its explanatory power through generating, from the bottom-up, the social fact that is to be explained.

3.) A pragmatic approach to the micro-macro link. This relates to the previous point: Agent-based modeling explicates the micro-macro link through bottom-up generation of macro-level phenomena.

4.) The pursuit of an unexcluded middle ground between deduction and induction, theory and data: Artificial societies, represented by agent-based models, allow sociologists to investigate reality without the chance of gathering data. Agent-based models are deductive, in that they begin with a set of theoretical assumptions regarding the generation of some social fact, and they are inductive, in that empirical (“grown”) data is analyzed.

5.) The focus on dynamics, process and change: Agent-based models allow researchers to investigate the dynamics of a social system by allowing for “[…] reproducing, synthesizing and visualizing space-time dynamics through the computer […].” (Squazzoni, 2012, p. 15)

6.) A tendency towards a trans-disciplinary/issue-oriented style of research: Artificial societies can potentially include systems and processes that are of interest to several disciplines, and knowledge from several disciplines can be combined when developing an agent-based model of a social system.
These core ideas are usually implicit in the motivation behind use of agent-based models. In addition, they underline the similar motivations of analytical sociology and the agent-based approach: The focus on bottom-up explanations in accordance with methodological individualism, and the primacy of analytical, middle-range models as over grand theory in particular.

An agent-based model can take various forms and be created in various programming languages, but Gilbert (2008, p. 2) proposes the following general definition: “[…] a computational method that enables a researcher to create, analyze, and experiment with models composed of agents that interact within an environment”. Agent-based models often, but not necessarily, involve visual representation of the social system being investigated. In addition, agent-based modeling can be viewed as a form of, or an extension of, microsimulation. Microsimulation in general is a tool for developing and testing hypotheses derived from empirical data. By letting an artificial population be subjected to transition rates and parameters estimated from empirical data, researchers can examine hypothetical scenarios and explore implications of what they have discovered in their statistical models. An example of this approach can be found in Rindfuss et al. (2010), where the effect of available child-care slots on fertility is the object of inquiry. After estimating the impact of the availability of child-care on fertility rates using statistical models, the researchers use microsimulation methods to create hypothetical scenarios with varying degrees of child-care coverage. In this case, the microsimulation provides insights into questions with predictive relevance: How would fertility rates look had there been full child-care coverage? What about the opposite case, with no child-care coverage?

The primary difference between traditional microsimulation and agent-based modeling is that agent-based models allow for modeling interaction in a more dynamic way. Microsimulation treats agents as isolated individuals, and does not allow for spatial interaction or heterogeneous behavioral rules (Gilbert, 2008, p. 18). In an agent-based model, on the other hand, agent heterogeneity, spatial relations and variations on behavioral rules are often central elements of the simulation, and allow the analyst to shed light on these issues and ultimately the link between the micro- and macro-level.

3.2.2 Why Use Agent-based Models?

Agent-based modeling may be seen as a manifestation of the central concepts of analytical sociology. The weak version of methodological individualism advocated in this approach fits well with the core concepts of agent-based modeling; bottom-up explanations, the focus on individuals and the interaction between them, and the explication of the micro-macro link. As discussed previously in this chapter, there is a call for an analytical approach where the “black boxes” of statistical association and variable-
based research is explicated and traced to explanatory mechanisms. Agent-based modeling can be seen as a potential answer to this call. As Manzo (2010) argues, there is a “structural homology […] between thinking in terms of mechanisms and building an agent-based model” (Manzo, 2010, p. 147): When creating an agent-based model of a social system, one needs to explicate all three links found in Coleman’s Boat – the situational mechanisms, the action-formation mechanisms, and the transformational mechanisms. There are several reasons to advocate for more frequent use of agent-based models within the social sciences.

First, agent-based models can function as tools for developing theory, by forcing researchers to formalize their underlying theoretical assumptions. When building an agent-based model, one has to transform theoretical and behavioral assumptions into logical or mathematical expressions and functions, which introduces a requirement for clarity and explicitness. For instance, a set of vague assumptions on what governs individual behavior is not sufficient to formulate an agent-based model of interaction; one has to code these assumptions into actual behavioral rules for agents to follow. It is impossible to specify an agent-based model using mere statistical information: one needs to have a clear perception about which explanatory mechanisms are at work in bringing about the pattern that is to be explained. Thus, developing an agent-based model can provide clarity about which theoretical assumptions, or mechanisms, one assumes to be generating a given macro-level phenomenon.

Second, agent-based models are useful for mapping complex processes. Agent-based models are typically employed in the study of processes that rely on complex, dynamic interactions between individuals, and not simply aggregation of many individuals’ behavior. These processes are often too complex to be easily understood using statistical or other mathematical models: For instance, if modeled as an equation based mathematical model, they tend to be “unsolvable”, in that they do not necessarily have stable equilibria to calculate. In this case, agent-based modeling can provide clarity both by showing how patterns emerge, and also by visualizing these patterns to improve our intuitive understanding of the processes generating them.

Third, agent-based models allow for assessing the relative importance of explanatory mechanisms. In some cases, competing mechanisms are assumed to be possible explanations for a given phenomenon. With an agent-based model, assessing the relative importance of these mechanisms is possible. One can ask the following question: Does mechanism A, B or both mechanisms have generative sufficiency - that is, the power to sufficiently generate the patterns we wish to understand? In Stovel & Fountains (2009) study of matching in the labor market, the authors show that several factors (employers with a preference for discrimination, segregated networks among those seeking
employment, or skill-level differences between groups) can generate different levels of segregation. An agent-based model can thus aid us in assessing the relative importance of these factors, and the interplay between them.

In addition, simulation techniques can be used to discover possible explanatory mechanisms by themselves. One can imagine asking the following question: What kind of mechanism would make the simulated outcome similar to the observed empirical outcomes? This approach is exemplified by Bearman, Moody and Stovel in their paper on adolescent sexual and romantic networks (Bearman et al., 2004). The object of interest in this case is the peculiar spanning tree-structure of adolescent sexual networks in a U.S. high school. In order to discover the potential mechanisms generating this structure, artificial networks were simulated according to a varied set of rules built on plausible assumptions regarding the sexual behavior of the students. This process eventually leads to the discovery of a potential social norm restricting the length of chains in the network: adding this norm to the rules of behavior of the agents in the simulations leads to networks similar to those found in observed data. The norm, which assumes that students seek to avoid status loss by being sexually involved with “(from a female perspective) one’s old boyfriend’s current girlfriend’s old boyfriend” (Bearman et al., 2004, p. 75), can thus be assumed to account for the peculiar network structure to be explained. Such a norm would most likely never be discovered through interviews of the students themselves: “Most adolescents would probably stare blankly at the researcher who asked boys: Is there a prohibition in your school against being in a relationship with your old girlfriend’s current boyfriend’s old girlfriend?” (Bearman et al., 2004, p. 75) Still, the researchers argue that such a norm makes sense intuitively, by referring to the potential for status loss in engaging in such a relationship. Ultimately, their approach illustrates how simulation methods can lead to the discovery of otherwise unnoticed potential explanatory mechanisms, along with assessing their relative importance.

Finally, agent-based modeling provides a tool for experimental approaches. Agent-based models are suitable for use as “artificial laboratories” used for experiments: By allowing the researcher to vary parameters to match hypothetical scenarios, simulation models allow for conducting various experiments that, in the real world, would be impossible to implement. This makes an agent-based model with a sufficient degree of realism a potential tool for forecasting, by varying parameters to match future scenarios and societies. It is important to note that, while using agent-based models for experiments is without doubt useful for building theoretical knowledge, ambitions of forecasting should come with a great deal of reservation. Developing an agent-based model with enough realism to capture the central dynamics of a system is one thing; attaining the realism needed to make anything
close to precise forecast, at least within the social sciences, will be very difficult, and for some problems likely unattainable.

It is worth noting that not all sociological explanations of phenomena or processes benefit equally from applications of agent-based modeling to scientific problems. As Macy and Willer (2002) notes, “ABMs are most appropriate for studying processes that lack central coordination, including the emergence of organizations that, once established, impose order from the top down” (Macy & Willer, 2002, p. 148) Agent-based modeling helps us understand macro phenomena, or global patterns, that cannot be reduced to aggregates of individual characteristics or behavior. Where interaction is key to understanding the formation of macro-level patterns, where there is no central organization that imposes order or structure from the top, agent-based modeling can be used to shed light on how micro-level interaction brings these patterns about.

3.2.3 Agent-based Modeling Within the Social Sciences: Three Examples

I have discussed the central ideas and concepts behind agent-based modeling, and it is worth noting that these do not necessarily address how agent-based models are used in practice. Bruch & Atwell (2013) suggests characterizing agent-based models along a continuous spectrum between “low-dimension realism” and “high-dimension realism”, according to their connection to empirical data. On the low-dimension end of the spectrum, we find stylized models where agents may have few attributes, simple rules of behavior, and may operate within an abstract world (e.g. a chess board-like grid). In models with high-dimension realism, on the other hand, agents usually have many attributes and the models include parameters based on empirical data, sometimes approaching what we could call an artificial society.

Even though tempting, aiming for as high a degree of realism as possible is usually not the best solution: Instead, the degree of empirical realism should ideally be governed by the research question the model seeks to answer: “A model’s success is determined not by how realistic it is but by how useful it is for helping understand the problem at hand.” (Bruch & Atwell, 2013, p. 8). Different types of models answer different types of questions. A stylized, theoretical model may for instance help “[...] develop new ways of thinking about a problem and provide a great deal of theoretical stimulation for existing empirical research [...]” (Bruch & Atwell, 2013, p. 7). On the other hand, some questions call for models with a higher degree of realism. For instance, if one wants to investigate a hypothetical scenario of a social system for predictive purposes, a high degree of realism may be required to provide
informative results. In this section, I present some examples of agent-based models within the social sciences, ranging from stylized theoretical models, to empirically based models with aims for “higher-dimension realism”. In addition to viewing agent-based models as ranging on a spectrum between low- and high-dimensional realism, a more formal classification scheme may be of use. The following map of agent-based models within sociology can be found in Squazzoni (2012, p. 21-22):
<table>
<thead>
<tr>
<th>Synthetic models</th>
<th>Analytical models</th>
<th>Applied models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial societies</td>
<td>Abstract models</td>
<td>Middle-range models</td>
</tr>
<tr>
<td><strong>Definition</strong></td>
<td><em>In-silico</em> social system surrogates</td>
<td>Theoretical models on general social phenomena</td>
</tr>
<tr>
<td><strong>Purposes</strong></td>
<td>Synthesizing components of social life realistically so as to explore intuitions on important aspects of the evolution of social behavior and structures that cannot be studied empirically or experimentally</td>
<td>Theory building and development through models that do not reflect any concrete and specific empirical instance</td>
</tr>
<tr>
<td><strong>Positive Consequences</strong></td>
<td>New insights might be incorporated/tested in more precise studies</td>
<td>Findings’ generalization</td>
</tr>
<tr>
<td></td>
<td>Favoring new connections between specialized knowledge might help a big picture view on social puzzles</td>
<td>Revealing non-obvious properties of social systems Exploring explanatory hypotheses Providing theoretical frameworks for empirical studies</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Critical Points</strong></td>
<td>Difficulty to transform results into empirically testable findings</td>
<td>Easy to lose sight of real empirical <em>explanandi</em></td>
</tr>
</tbody>
</table>

Table 3.2: Squazzoni’s classification of agent-based computational models (Squazzoni, 2012, p. 21-22)
Squazzoni’s map is a more fine-grained attempt at describing agent-based models, their purposes, their potential positive consequences, and critical points. The classification is useful when discussing agent-based model, to gain a better sense of their goals and limitations. The scheme, from left to right, classifies agent-based models into categories based on their generality and target *explanandum*. It is worth noting that these categories are not meant to be exclusive and rigid, but should be seen as a continuum ranging from the general to the more specific. Leftmost, we find artificial societies - simulations that seek to capture all elements of a social system. These models can represent a nesting of more specialized knowledge from various disciplines into a “big picture” view of a social system. They are not prevalent in the social sciences, but Epstein and Axtell’s (1997) “Sugarscape” model can be seen as a prototype for this kind of modeling. This model has also been expanded to examine, among other things, more realistic wealth distributions (Rahman et al., 2009) and historical demographic change (Dean et al., 1998).

Abstract models, on the other hand, are perhaps the most prevalent within the field of sociology. These models attempt to generate some general social structure or phenomenon that isn’t necessarily linked to a specific empirical setting, but rather focusing on building theory by providing insight on general mechanisms. The phenomena in question can for instance be the evolution of ethnocentrism based on arbitrary group differences (Hammond & Axelrod, 2006), the diffusion of trends (Strang & Macy, 2001), or “bandwagon” effects on collective action (Chwe, 1999). Middle-range models are similar to abstract models, differing in that they usually target a more specific class of empirical phenomena. With Case-based models one moves even further towards specific empirical phenomena by attempting to match some real-world, well-circumscribed social system. By infusing more empirical data into the specification of a model, a middle-range or abstract model can be transformed into a case-based model.

Finally, applied simulation models represent the end of the spectrum, in that they attempt replicate a highly specific empirical phenomenon with great detail and realism. When agent-based models are used for policy recommendation or forecasting purposes, they usually take the form of applied simulation models. Like artificial societies, applied simulation models are not very prevalent in sociology, but are used as tools for policy testing and forecasting in health (for an overview, see Nealon & Moreno, 2003). In the next section, I will discuss three examples of agent-based models employed in the social sciences, and relate these two both Squazzoni’s typology and to Bruch & Atwell’s notion of high- and low-dimensional realism.
Thomas Schelling’s model of residential segregation

Perhaps the most famous and the most frequently mentioned example of agent-based modeling in the social science is Thomas Schelling’s tipping-point model of residential segregation (Schelling, 1971). Perhaps ironically, the most celebrated model mentioned in the context of computational social science was not developed computationally at all – Schelling used a checker-like board on paper and tokens to represent the models agents. Schelling’s objective was to investigate how individual preference for having similar neighbors might give rise to segregated neighborhoods and cities. The agents of the model are located on a square grid representing an urban area, and they each have one of two available colors. Initially, the agents are randomly distributed, and they all follow a simple rule of action: if the share of same-colored neighbors drops below a certain threshold (specified by the modeler), the agents move to a different location on the grid. This is repeated for all agents until everyone is satisfied with their current position in the grid. The results show neighborhood formation and the emergence of segregation patterns, even among agents with only a weak preference for being surrounded by same-group neighbors.

On the axis between high- and low-dimensional realism models, Schelling’s model is located firmly on the low-dimensional side. In Squazzoni’s classification map, this would be an abstract model that aims to provide theoretical insight into the social mechanisms producing a general phenomenon. No parameter of the model is based on empirical data, the world represented is an abstract grid, and agents follow very simple rules of behavior. In spite of this, the model provides valuable insight into segregation-generating processes. When run, the model shows that a tipping mechanism occurs: As members of one group enter a neighborhood predominantly inhabited by the same group as themselves, members of the other group in this neighborhood, who otherwise would be satisfied, move out. The result of this is that even though agents express tolerance in their micro-level behavior, the interaction between them causes a cascade towards unintended patterns of segregation.
Schelling’s model shows another important theoretical point regarding the micro-macro link, through the fact that trivial changes in the agents’ tolerance for out-group neighbors can lead to massive changes in the degree of segregation observed at the macro-level. The difference between a preference for at least 25% in-group neighbors versus a preference for at least 26% produces a difference between 55% and 75% overall segregation (measured by the average number of in-group neighbors). If we were to only observe the macro-level patterns, one would likely conclude that the population of the city with the higher level of segregation of the two would have to be substantially more xenophobic. In fact, the difference between the two cities (a 1% decrease in tolerance) is trivial. This again demonstrates the difficulty of making assumptions of macro-to-macro-level associations, without taking account of the complex interaction processes that constitute the micro-macro link. What makes the Schelling model interesting is not only that it shows that micro-to-macro mechanisms can have tipping or cascading tendencies of self-reinforcement, but also that assumptions on macro-to-macro associations are highly error-prone.

*Understanding cultural change through agent-based modeling*

While Schelling’s model of residential segregation is highly abstract and stylized, there are many examples of agent-based models aiming for a higher degree of realism. One such
example is Dean and colleagues’ model of the Anasazi tribal society (Dean et al., 1998). The researchers sought to understand the processes that led to the sudden abandonment of the settlements in the Long House Valley around the year 1300, and also to explore in general the possibilities of using an agent-based model to simulate a population of human agents reacting to environmental change. One central question addressed is whether environmental factors alone can account for the sudden abandonment of the settlements, or if social or institutional factors need to be taken into account in order for the model to generate results that match the empirical data.

Figure 3.4: Visualization of the Anasazi society as implemented in NetLogo.

Such an undertaking calls for using rich empirical data when specifying and developing the agent-based model. Using archeological data on soil fertility for maize production, data on the area’s topography, dendroclimatology and geomorphology (to name some, but not all), the researchers attempted to reconstruct in great detail the Long House Valley area and its implication for human society in the given period. The model simulates a population of humans deciding where to farm and settle while consuming food, reproducing, aging and eventually dying. The simulation proves informative by allowing the researchers to test hypothesis about past events, for which there is only indirect evidence. The researchers
underline a central difficulty of historical sciences like archaeology; the inability to “rewind
and rerun” the tape of history. An agent-based model such as this one is seen as a potential
tool to allow something similar to “rewinding the tape”:

[…] with agent-based modeling we can execute numerous simulations to investigate
alternative outcomes of sociocultural processes under different initial conditions and
operational procedures. […] Thus, in terms of the Artificial Anasazi model, we could change
agent attributes, such as fecundity or food consumption, or introduce new elements, such as
mobile raiders, environmental catastrophes, or epidemics. (Dean et al., 2006)

A central result of the simulation is that environmental factors included in the model can
account for important demographic features in the Anasazi society, and also a steep decline in
population around 1300 A.D. – when the real-world Anasazi society vanished. The actual
disappearance is, however, not accounted for. This can lead to the conclusion that other
factors are needed to fully explain the Anasazi’s departure from Long House Valley, be they
cultural, institutional, or something else entirely. This illustrates the potential value of agent-
based modeling as a tool for exploring historical processes and assessing the relative
importance of explanatory mechanisms.

The Wedding Ring: An agent-based model of marriage and partnership formation

A more recent example of agent-based modeling in the social sciences is Billari and
colleagues’ “Wedding Ring” model of age-at-marriage patterns in societies (Billari et al.,
2008). The objective of the research by Billari and colleagues is to investigate whether current
assumptions in research on individual behavior and social interaction are sufficient in
explaining the patterns of partnership formation typical for Western societies. This is done by
creating an agent-based model that represents a population of males and females who form
couples and eventually get married. An agent-based model is helpful in this case, to assess
whether or not the current assumptions on the formation of marriage-patterns are sufficient to
generate the age-at-marriage patterns observed in empirical data.

The central assumptions underlying the model can be described as follows, based on
the work of Gudmund Hernes (1972) on the entry into first marriage. On one hand, an
individual’s search for a partner is affected by its social network: The larger the network, the
more potential relevant partners – and this network size varies with age. On the other hand,
individuals experience “social pressure”: People are increasingly likely to consider marriage as the share of married people within their network increases. These two effects also affect each other. If the share of married people in an individual’s network increases, the social pressure to get married grows. At the same time, the number of available (unmarried) partners decreases.

The simulation results provide insight into the possible mechanisms producing the age-at-marriage curve observed in empirical data. Running the simulation on a population of 800 individuals, they arrive at several conclusions: The first is that if the model only takes “social pressure” into account, that is, the peer-effect of married people in an agent’s network on their propensity to get married themselves, the age-at-marriage curve observed is not similar to the one observed in empirical data. However, when the age-dependent size of networks is taken into consideration, a different picture emerges: “Including the age dependency of the network size to determine (through social pressure) the acceptable range of potential partners, results in the emergence of a hazard that is qualitatively similar to the one that is well–known” (Billari et al., 2008, p. 71). This leads to the conclusion that the two mechanisms, social pressure and social influence (through network size), by themselves are insufficient for generating the age-at-marriage curve characteristic for western societies. Including both mechanisms, however, constitutes a candidate explanation for the phenomenon.

Compared to Schelling’s model of residential segregation, the Wedding Ring model moves us closer to the high-dimensional realism end of the spectrum: Population distributions, fertility and mortality rates are estimated from empirical data. At the same time, the world in which the agents interact and the interaction itself is stylized and abstract. The aim of the researchers is not to create realistic surrogate for the real marriage market, but to capture and replicate certain central dynamics and processes. In Squazzoni’s terms, the Wedding Ring model would be described as a case-based model: The aim is to map the complexity of a specific social phenomenon, using empirical data for estimation of the models parameters.

**3.2.4 Points of Contention Regarding Agent-based Modeling**

While these examples all show various applications of computational modeling currently being used within the social sciences, there are limitations to this approach that needs to be addressed. When advocating an agent-based computational approach, one often refers to the
The explanatory strength of being able to “generate” macro-level social patterns. This strength can be called into question: Let us assume an agent-based model manages to accurately reproduce the qualitative nature of some complex macro-level pattern using a theoretically grounded set of assumptions. Now, let us imagine a different model using a completely different set of assumptions being able to generate identical patterns. In this case, it is difficult to assess which causal mechanisms are more likely to be at work in the real-world process being modeled. This is a problem, but not without solutions. First of all, the problem can be addressed theoretically, by discussing which set of underlying assumptions are more coherent with current theory, and thus more plausible. Second, the problem might be addressed empirically: By comparing the initialization of both models with empirical data, one can assess which set of behavioral assumptions are the most closely aligned with real-world observations. As Epstein (2006) argues, agent-based models provide us with candidate explanations, and, as in any other science, there is work to be done on figuring out which micro-level specification is most tenable empirically (Epstein, 2006, p. 9).

As mentioned, a balance needs to be struck when creating an agent-based model between high- and low-dimensional realism. As discussed earlier, it might sometimes be tempting to include a multitude of parameters to achieve a more realistic representation of the process being modeled. This can be a double-edged sword; on one hand, one can achieve higher external validity by being able to represent the empirical world more closely. On the other, the clarity and ability to trace causal mechanisms might gradually be weakened. Macy and Flache (2009) illustrates this problem by pointing to what they call the “bumper sticker” of agent-based modeling, coined by Joshua Epstein; “If you didn’t grow it, you didn’t explain it”. When exploring a complex agent-based model, however, simply growing a macro-level pattern is not necessarily enough to provide an explanation. Assume one has an agent-based model capable of “growing” the macro-phenomena to be explained. A certain combination of parameter settings gives a close match to empirical reality, and the model shows that the underlying assumptions are capable of generating the observed patterns. However, if the model reaches a high level of complexity, understanding how these parameter combinations generate various macro-patterns can prove just as difficult as explaining the mechanisms by analyzing empirical data. Macy & Flache thereby suggests adding a second “bumper sticker”: “If you don’t know how you grew it, you didn’t explain it.” (Macy & Flache, 2009, p. 263). This illustrates the importance of keeping the model dynamics clear-cut, to preserve the strength of causal clarity provided by agent-based modeling. Macy & Willer (2002) underlines the point that keeping models simple enough for transparency is vital, and adding
unnecessary complexity “[...] undermines their usefulness as tools for theoretical research if we can no longer figure out how the model produces a given result” (Macy & Willer, 2002, p. 147).

3.2.5 Concluding Remarks

I have now provided an overview the core ideas behind agent-based modeling, the main purposes of such models, and three examples of agent-based modeling within the social sciences. These are just some of many examples that can potentially illustrate the fruitfulness of using computational methods within the social sciences. Agent-based modeling lets researchers model certain aspects of social life that other modeling strategies might not be able to handle with the same ease, and can function as a tool for developing insight into the complex micro-macro link, where transformational mechanisms are in play. As noted by Hedström and Manzo (2015):

[…] we have developed a range of methodological tools for describing how contextual and individual-level properties affect actor-level outcomes. What we largely are lacking, however, are tools for addressing the micro-to-macro link, that is, tools for assessing the macro outcomes individuals are likely to bring about when they act and interact over extended periods of time. Agent-based modeling is destined to become a crucial tool for addressing such questions. (Hedström & Manzo, 2015, p. 182)

Of course, this is not to suggest that agent-based modeling should replace traditional sociological research using statistical data and methods, or the analysis of qualitative data: Instead, agent-based modeling should be seen as complementary to these methods. Agent-based models can (and perhaps should) be informed by assumptions derived from empirical research, and can again inform further empirical research by providing further insight into the processes being investigated.

3.3 A Computational Approach to the Question of Persisting School Segregation

As I have discussed in chapter 2, the problem of bias leads to some potential difficulties when investigating the preferences motivating parents’ choice of schools. Even though both survey data and qualitative data can yield important knowledge on the most important mechanisms
affecting school enrollment, the question may also be approached from a computational perspective. In order to evaluate the possible mechanisms behind the emergence of segregated schools, Stoica & Flache (2014) employ an agent-based model representing an abstract world of parents choosing schools. As Schelling’s model of residential segregation shows us, segregation patterns can arise as a product of relatively weak preferences for being surrounded by same-group members – i.e. people belonging to the same ethnic group. Comparing the findings in Schelling’s model to a similar model of school enrollment can provide important insight into how segregation-producing dynamics function in schools.

Their first research question was as follows: Are Schelling-type preference dynamics sufficient to generate segregation in schools, as they do in neighborhoods? Their model can be classified as what Squazzoni terms an abstract model; the model is theoretical, and describes a general social phenomenon, namely segregation in schools. The intent is not to explain how segregation emerges within a specific city, but to theoretically address whether or not the postulated explanatory hypotheses are sufficient to generate segregation in a general system. The two assumptions under scrutiny in this case are: a) parental preference for a certain ethnic composition in schools, and b) parental preference for minimizing home-to-school distance. What follows is a brief overview of the model and their results.

The model is composed of two kinds of entities: schools and households. The schools are randomly distributed on a grid, and each school has a catchment-area which is made up of every location closest to the given school than to any other schools. Households are then distributed randomly on the grid, and each household initially has their child enrolled into the nearest school. Households are either majority-background or minority-background, represented by their color – as in Schelling’s model of residential segregation. When the simulation runs, every household evaluates whether or not they are satisfied with their currently allocated school. Their satisfaction depends on two parameters: the ethnic composition of the school, and the geographical distance from the household to the school. In addition, a weighting parameter is included, so the researcher can specify how much weight the households put on ethnic composition and distance. The unsatisfied households then have the option to select a new school; all schools with empty slots are evaluated, and the first one found that would make the household satisfied is chosen.

After running the simulation through various experiments, their findings indicate that unintended segregation can, as it does in Schelling’s model, arise in schools as well as in residential neighborhoods. In addition, when varying the weighting parameter governing how important distance is for households choosing schools, they find that preferences for
minimizing distance can suppress the tendency towards unintended segregation: “We found that this curbed segregation in an initially integrated population that would have segregated when parents' choices had only been driven by ethnic preferences.” (Stoica & Flache, 2014, 4.2) Thus, segregation in a population of tolerant agents can arise as an unintended consequence of individual behavior, but the level of segregation may again be suppressed by a competing preference for minimizing distance to schools. As mentioned earlier, this agent-based model will be the foundation of the approach in this thesis. The general framework of the model will be modified and expanded, and these modifications will be discussed in the next chapter.
4 Proposal for an Agent-based Model of School Segregation

This chapter will explain the agent-based model of school segregation. In the abstract version of the model, the simulation represents a map of a fictional city or an urban area with households and schools randomly distributed. This is similar to the Schelling model of residential segregation, and allows us to investigate if similar segregation-producing dynamics operate in the case of schools. In the empirically calibrated version, the geographical locations of the schools and the initial population distribution are based on census data from Oslo, more specifically, the region encompassing the twelve southernmost schools of the Oslo school district. This selection was made due to the differences in the ethnic composition the neighborhoods surrounding schools in the area, as a contrast to the homogeneous distribution provided in the abstract model. In this chapter I will first provide an overview of the model: Its purpose and its technical implementation, along with a discussion on the Index of Dissimilarity, which I use as a measurement of segregation. Second, I will provide an illustration of how the simulation proceeds for documentation purposes.

4.1 Overview of the Model

Figure 4.1: The visual interface of the model.
The main purpose of the model is to capture the dynamics that generate patterns of ethnic segregation in schools, and to assess how two preferences among parents, the preference of a certain ethnic composition and the preference for minimizing home-school distance, affects the growth and persistence of segregation patterns. The following section describes the basic structure of the model.

The model is based on two classes of agents; households and schools. These agents are located on a spatial non-toroidal grid, representing a city (or a specified area of a city). All geographical locations that are closer to a given school than to any other school constitute that school’s catchment area. All households located within this catchment area are allocated the given school as their default school, which I refer to as their local school. Schools in this model have a maximum capacity for students, which equals the number of households in their catchment area plus a variable percentage of vacant slots. This aspect is carried over from Stoica & Flache’s model reviewed in the previous chapter, and seeks to capture the variable nature of the maximum capacity of schools: Each school’s capacity is based on the size of its catchment area, and the variable percentage of vacant slots is added to ensure the possibility of actual mobility between schools. If there were no extra capacity, no household would be allowed to select a school other than that household’s local school. In addition, this mimics the enrollment system in Norwegian schools; as a parent, you are guaranteed a slot at your local school, but beyond this, schools have a limited capacity for enrolling students from outside of the catchment area. Households represent parents with exactly one child to be enrolled in the school system, and are divided into two populations based on minority or majority status.

When the simulation is initiated, the schools (the number of which is specified beforehand) are generated and randomly distributed on the spatial grid. For each step of the simulation, the following happens: Households, representing the population of students enrolling in schools one year, are generated and randomly placed on the grid. Their minority/majority status is then allocated according to a parameter specifying the percentage share of minority students. The households enroll their child into the closest school, their local school. Finally, each household checks whether their utility is below a threshold value, indicating that they are unsatisfied with their current school. Those unsatisfied will evaluate a given percentage of alternative schools, identifying the one that would yield the highest utility. The number of schools evaluated is based on a “search radius” parameter ranging between 0 and 1: When set to 1, all schools are evaluated. When set to 0.5, only the closest 50% are evaluated, and so on. After identifying the most desirable school
among those evaluated, the households enroll their children into the newly chosen target school. Finally, the households leave the simulation, and the schools store their population of students so that the next population of households will base their choice upon the previous year’s ethnic composition. This means that each time step of the simulation represents one year, thus allowing for modeling the evolution of segregation patterns over time.

The model is implemented using the software package NetLogo (Wilensky, 1999), which is a programmable modeling environment for building agent-based models and exploring multilevel systems. Originally created for educational purposes, NetLogo has grown into a solid modeling platform on its own, and is now one of the most widely used software packages for agent-based modeling in the social sciences. Ease of use and lower requirements for programming knowledge are often cited as reasons for its widespread use, while some criticize it for being less powerful and non-intuitive relative to other programming languages. For a comparison on various platforms and languages for developing agent-based models, see Nikolai and Madey (2009). The NetLogo code of this model is included in Appendix A.

4.1.1 Household’s Preferences and Utility

As mentioned above, the households evaluate schools based on the ethnic composition of the school’s student mass. In the original Schelling model, the equivalent calculation is done by simply dividing the number of same-color neighbors with the total number of neighbors – if this share drops below a certain threshold, the agent in question is unsatisfied with its current location. In this model, I have opted for a more nuanced method of evaluating the desirability of schools, as in Stoica & Flache’s model. The evaluation of schools consists of calculating the “utility” derived from the school in question, and is derived from the following equation:

\[
P(x) = \begin{cases} 
\frac{x}{fp}, & x \leq fp \\
M + \frac{(p-x)(1-M)}{p(1-f)}, & x > fp 
\end{cases}
\]

Where \(p\) is the total number of students in the school, \(x\) the number of students of the own group, \(M\) the specified utility when the same-color share is 100%, and \(f\) the specified share of same-colored students that would yield maximal utility (Stoica & Flache, 2014, 2.7). If a household is in a minority at its current school, the utility is simply the number of same-group students.
students divided by the ideal number of same-group students. If a household is in a majority at
its current school, the function is elaborated to reduce utility gained based on the parameter
$M$. This function builds on the work of Zhang (2004) and makes an important addition to
Schelling’s original model possible: A household’s utility at a certain school can decrease if
the share of same-colored students exceeds the ideal share. By lowering the value of $M$, it is
possible to investigate the patterns that emerge from agents with preferences against
segregation; that is, agents who would be unsatisfied with their current school if they were in
a too large majority. While not a central topic of this thesis, such an investigation provides an
interesting venue for future work. For comparison: In Schelling’s original model, an agent
with a preference against being in a minimum 30% minority would only care if his share of
same-colored neighbors dropped below this threshold. There would be no difference in
“utility” gained between a share of 80 percent or 100 percent – as long as the values are both
above the threshold of 30 percent. Meanwhile, according to this utility function, we instead
operate on an ideal share of same-colored neighbors (or co-students, in our case); this means
that the expected utility is lower if the share of same-colored neighbors is 100 percent,
compared to the ideal share of, say, 80 percent. The following figure shows a given
households utility:

Figure 4.2: Graph of households’ utility function.

Note that when the same-group share is 100% ($p = 1$), utility equals $M$, which is a parameter
that can be specified before running the simulation. If the share of same-group students equals
the ideal share $f$, the utility equals 1, its maximum. In addition to the parameters described
above, a threshold $T$ is specified, governing the point at which a household is unsatisfied. The baseline setting of $T$ for this thesis is 0.4, in order to mimic the “tolerant” behavior examined in Schelling’s model of residential segregation. This would mean that households would be unsatisfied once the utility of their current school drops below 0.4. The threshold value can be varied to represent various degrees of tolerance among household. In addition, some heterogeneity is added to this process by assigning a slightly altered threshold to each household. Instead of letting all households being unsatisfied at the exact same utility level, I have modified the threshold to instead be a random value, which follows a normal distribution with a mean equal to the value of $T$, and a standard deviation of 0.05. The standard deviation was chosen so that most households’ threshold will fall within at least 0.1 from the mean value $T$. The following plot illustrates this distribution when $T = 0.4$:

![Plot of distribution of households' threshold values, $T = 0.4$, SD = 0.05.](image)

Figure 4.3: Distribution of households’ threshold values, $T = 0.4$, SD = 0.05.

This represents a slightly more realistic approach, by making each households behavior less predictable, and acknowledging the fact that real-world parents do not necessarily have the exact same criteria when evaluating their satisfaction. Thus, instead of $T$ being an absolute value, it describes the mean of the distribution of thresholds among all households.
4.1.2 The Searching Process: Satisficing vs. Optimizing

In an agent-based model where agents engage in a searching process and choosing between various options, the way in which they search and settle for a choice can be important for the outcome of the simulation. This has been a subject of debate in various recreations and extensions of the Schelling model. Should agents choose the first option that is better than their current location? Should they survey all available options, and then select the one that would satisfy them the most? Or should they choose the first one that would put them above the threshold for satisfaction?

These questions relate to the terms *satisficing* vs. *optimizing*. Satisficing, a portmanteau of *satisfy* and *suffice*, was introduced by Herbert Simon (1956) as a suggestion on a more realistic representation of human decision making. In practice, satisficing agents do not operate on perfect information; they survey options until they find one that would make them satisfied. This leads to agents not necessarily making the optimal choice, and in many cases provides a more realistic picture of how human decision making actually works. On the other hand, an optimizing agent would survey all possible alternatives, and then maximize his utility by choosing the best one. In my model, the unsatisfied households searching for an alternative school are optimizing; I argue that this more realistically represents the real-life school choice process. Surveying alternative schools is a well thought-out process, and information regarding the ethnic composition (and distance) of all alternative schools are usually readily available for parents making the decision. Thus, satisficing behavior might be more realistic in a model of residential segregation (where the available choices for habitation are vast and information not necessarily perfect), but optimizing behavior might more accurately reflect the process of choosing a school.

4.1.3 Schools as Bounded Neighborhoods

As noted earlier, this model seeks to capture the same segregation-generating dynamics that Schelling found in his model of residential segregation. In this regard, it is important to discuss how schools in this model differ from neighborhoods in the original Schelling model. In Schelling’s model, a neighborhood consists of the closest locations surrounding an agent’s position on the grid. This means that each neighborhood is unique to each agent; naturally, they overlap, but each agent has its neighborhood defined according to his own unique residential location.
In this model, schools replace neighborhoods. They differ in a significant way from Schelling’s original neighborhoods; they are static and defined generally, as opposed to centering on a particular household. Instead, they are more similar to what Schelling in a later model elaborated as “bounded neighborhoods” (Schelling, 1978). Bounded neighborhoods are fixed blocks of residential locations, and considerably larger than the local neighborhoods found in the original Schelling model. This has important implications for how the model generates segregation patterns. From a random population distribution, larger fixed neighborhoods (schools, in our case) make it more difficult for the local ethnic composition to differ widely from the distribution of the population as a whole. This means that, in a population of households that would be satisfied by an ethnic composition that matches the population in general, there is less of a chance that a given household will be unsatisfied with its allocated school when the model is initialized. The implications of this will be further explored in the analysis of simulation results.

4.1.4 Model Output: The Index of Dissimilarity

The most widely used measurement of segregation, or, more specific, evenness, as described by Massey & Denton (1998), is the index of dissimilarity (D-index henceforth). The D-index is obtained from the following formula, originally formulated by Duncan & Duncan (1955):

\[
D = 0.5 \sum \left| \frac{N_{ia}}{N_a} - \frac{N_{ib}}{N_b} \right|
\]

Where \(N_{ia}\) is the population of group \(a\) in the \(i\)th tract, \(N_{ib}\) is the population of group \(b\) in the \(i\)th tract, \(N_a\) is the population of group \(a\) in the city as a whole, and \(N_b\) is the population of group \(b\) in the city as a whole (White, 1983, p. 1009). In the case of measuring school segregation, tracts would here refer to schools. One of the main advantages of using the D-index is that it is relatively simple to interpret: The D-index is the share of either group \(a\) or \(b\) that would have to relocate in order for there to be an even distribution that matches the city as a whole. If the D-index is at its maximum, 1, the distribution is maximally uneven. If it is at its minimum, 0, the distribution of the groups matches the population as a whole, i.e. an
absence of any segregation. The fact that the values of the D-index ranges between 0 and 1 is also an advantage for analysis and comparison. It is worth noting that the expected value of the D-index in a randomly distributed population is slightly above 0, as there will always be some degree of unevenness: It is extremely unlikely that the local distribution of every tract is identical to that of the population as a whole (see Winship, 1977; Corteze et al., 1976).

One central criticism of the D-index is that it does not take into account the spatial relationship between tracts, or in our case, schools. White refers to this as the “checkerboard” problem:

Allow the squares on a checkerboard to represent parcels (neighborhoods, tracts, blocks). Once the composition of each parcel (square) is given, any spatial rearrangement of them will still result in the same calculation for $D$. […] A city in which all the nonwhite parcels were concentrated into one single ghetto would have the same level of calculated segregation as a city with dispersed pockets of minority residents. (White, 1998, p. 1010-1011)

This problem is not specific to discussions residential segregation. All the minority-dominated schools in a city may be either evenly dispersed, or closely clustered together in the urban core: In either case, the D-index would show the same level of segregation. In spite of this objection, the D-Index is still the most widely used measurement of spatial segregation, and will be used throughout this thesis. Its potential weaknesses will be discussed as they become relevant. An agent-based model has one advantage in overcoming this problem. In addition to providing the D-index at every stage of the simulation, one can also visually observe the patterns of segregation not captured by the index itself. This allows for analyzing exactly which schools end up with a skewed distribution of minority students, and whether or not these schools are clustered together or evenly distributed on the grid.

4.2 An illustration: Following a Household and a School Through the Model

To further illustrate the workings of the model, I will describe the exact course of a household from the beginning of the simulation, until the household exits the simulation. First, the household is created and assigned to a random location on the grid. The household is then allocated a color; either green or red, representing minority or majority background. In this
case, we will let the color of the household be green. The household then enrolls its child into the closest school, its allocated local school.

Next, the household checks whether or not it is satisfied with its currently allocated school. This is done by calculating the utility using the function showed in figure 1. In this case, if the share of green students enrolled in the school last year was less than the ideal share of similar agents specified by the researcher, i.e. \( x < f_p \), the utility is calculated as \( x \) divided by \( f_p \). If the share of green students were larger than the ideal share, the utility is instead calculated as \( M + \frac{(p-x)(1-M)}{p(1-f)} \), where \( M \) is the utility derived if there were no members of the other group (here, red students) in the school. In our case, let us assume that the number of green students in the current school is \( x = 45 \) the total number of students \( p = 256 \), and the households’ ideal share of same-group members is \( f = 0.8 \). This gives the following utility: \( \frac{45}{256 \times 0.8} = 0.219 \). After calculating the utility, the household checks whether or not this falls below the utility threshold specified by the researcher; in this case, we let this threshold value be 0.4, to represent fairly tolerant households. The household is randomly assigned a threshold value from a distribution with a mean of 0.4 and a standard deviation of 0.05. In this case, the value assigned is 0.34. As 0.219 is less than 0.34, our green household is unsatisfied with its current school: There are too few students from the same ethnic group at the school assigned to it by default.

The household now begins searching for an alternative school, preferably one with a larger share of other green students. It chooses the closest schools, the number of which is specified by the researcher. In this case, let us assume that distance does matter, so the households only evaluate the closest half of all the schools: the search radius parameter is set to 0.5. Say our simulation contains 25 schools; our household now begins to evaluate the 13 (the number is rounded up) closest schools. This is done in the same way as the utility calculation described above. After calculating the expected utility for each of these 13 schools, the household switches to the school which will yield the highest utility, i.e. the one with the most favorable ethnic composition. Schools with no vacant slots (the amount of which is specified by the researcher before initializing the model) are excluded from consideration. This entails that there is a certain “first come, first served”-principle operating: If any given school has been filled up with students from unsatisfied households originally belonging to other schools, it is unavailable for selection for the rest of the households. The order of which households are selected to evaluate their satisfaction and potentially switch schools is random, and this matches the assumptions that schools accept all applications for enrollment until their maximum capacity is reached. Finally, after subtracting one green
student at its original school and adding one at its newly chosen school, the household exits the simulation: All households are removed to make room for the next cohort of households, representing the following year’s first-year students and their parents. This is repeated for each time-step, or year, of simulation.

Meanwhile, the schools in the model follow a simpler procedure. They are first generated and randomly distributed, before the households. After the households enroll their children into their closest school (before evaluating their satisfaction), the schools register their current student composition. Their variables are the total number of students, the total number of green students, the total number of red students, and their total capacity. The total capacity is set to the number of households enrolled, plus an extra percentage specified by the researcher. As an example, let us imagine the local school of the household described previously; the total number of students could for instance be 256, with 45 green students and 211 red students. If the parameter governing the extra capacity in schools was set to 25 percent, the school would have a maximum capacity of 320 students. When the unsatisfied households evaluate their options and start switching schools, this composition might change; depending on the threshold for satisfaction, the green students (being in a small minority) might apply to different schools. In addition, unsatisfied red students from nearby schools might perceive this school as a good candidate for switching, due to its high share of red students. The net result would then be a drop in the number of green students, and an increase in the number of red students. These numbers are then registered and stored after the current households leave the simulation, to be used by next year’s households in their evaluation and so on.

4.3 Changes and Additions Compared to the Original Model by Stoica & Flache (2014)

Comparing my model to Stoica & Flache’s model, there are two key differences. First of all, parental preference for minimizing distance to a certain school is not modeled as a linear function. Instead, I have opted to specify a “search radius”, given as a percentage share of the total schools by distance, in which parents are willing to send their children. This captures the fact that the “utility” given from the distance between the household and a given school might not follow a linear curve, but instead work as a threshold. If a school falls outside of a common bus line and requires a switch of public transportation means, the perceived utility might rapidly drop towards zero. On the other hand, if two schools are within walking
distance, the perceived difference in utility between the two might be negligible. Thus, in this model, the evaluation and choice of a new school is a two-step process: First, the households’ exclude all schools that fall outside of their acceptable distance (their “search radius”). Afterwards, they evaluate the remaining schools and make an optimal decision regarding ethnic composition.

Second, and perhaps more importantly, this model allows us to investigate the segregation-generating processes over time. In Stoica & Flache’s model, one batch of households are simultaneously (technically, the choices happen sequentially, but the choice of one household does not affect the choices of other households) selecting schools for their children. Thus, the simulation ends when all households have either chosen a new school or decided to stay at their allocated local school. In my model, new populations of households enter the simulation for each year, and their choices of schools are affected by the choices of previous cohorts. This allows us to investigate the tipping dynamics that make Schelling’s model so interesting; segregation-generating processes are self-perpetuating, and may reinforce themselves over time. The level of segregation in a group of schools may fluctuate, grow or stabilize over time, and either of these developments should be possible within this agent-based model.
5 Analysis of Simulation Results

In this chapter, I will present a series of experiments using the agent-based model in attempting to answer some central questions regarding segregation-generating dynamics in schools. The first of these questions stems from the findings of Thomas Schelling in his original model, which shows us that relatively strong patterns of residential segregation can appear even among fairly tolerant individuals. Can similar dynamics appear in the case of school segregation as well? This is also the first question addressed by Stoica & Flache, in their comparison between a Schelling-type model of residential segregation and their own model of school segregation.

The second question my model seeks to address details the relationship between parental preference for minimizing geographical home-school distance, and their preference for certain ethnic compositions in schools. Perhaps the most central difference between traditional Schelling-type models of residential segregation and Stoica & Flache’s model of school segregation is that the individuals (or households, in the school model) are immobile: their geographical location is fixed, and the choice being modeled revolves around choosing a school in contrast to moving to a new neighborhood. This, in combination with the assumption that geographical distance matters in the choice of schools, motivates investigating how a preference for minimizing distance impacts segregation-generating processes. Stoica & Flache finds that including a preference for minimizing distance among households has a curbing effect on segregation patterns; I will attempt to replicate these findings, in addition to offering an alternative implementation of such a preference.

Finally, perhaps the most interesting extension of my model compared to Stoica & Flache’s model is the inclusion of a temporal aspect. This allows for investigating the relationships discussed above over time, potentially providing insight into the self-reinforcing nature of segregation-generating processes in schools. On the basis of these questions, I formulate the following three hypotheses:

1. Segregation can arise among tolerant agents in schools, as it can in residential neighborhoods.
2. Including a preference among households for minimizing home-school distance can curb these tendencies towards self-organized segregation.
3. School segregation-generating tendencies are self-reinforcing and increase over time.
The first two hypotheses are built on the findings of Stoica & Flache, and the third hypothesis is derived from the findings in Schelling’s original model of residential segregation. When the simulation is run over an extended time span, the behavior of one set of households is allowed to affect the behavior of future households, thus gradually reinforcing patterns of segregation.

5.1 Theoretical Simulation Experiments

5.1.1 Simulation Experiment 1: Segregation Among Tolerant Parents

In the first experiment of the agent-based model, the objective is to replicate the findings of Stoica & Flache. We seek to answer the following question: Can patterns of segregation among tolerant agents arise in schools, as it does in Schelling’s neighborhoods? I define “tolerant agents” as those who would accept being a part of a minority in given school, but who would engage in switching schools if the share of same-group members would drop below 30 percent. This mirrors the basic assumption behind Schelling’s model of residential segregation. In addition, the households have a certain integrative preference, in that their perceived optimal share of same-group members is 70%, while \( M = 0.8 \). This means that their perceived utility drops if the share of same-group members is larger than 70%, down to a minimum utility of 0.8. This captures the fact that parents have a certain preference for a multiethnic environment, as long as their children are not in too small of a minority. The initial distribution of ethnicities is varied between three population scenarios: 60%-40%, 70%-30%, and 80%-20%. We set the number of households and schools to match the basic scenario found in Stoica & Flache, with 6000 households and 25 schools. The household search radius when evaluating schools is set to 100% (of all schools), to represent no preference relating to distance at all. Each school has an extra capacity of 25%, to allow for some movement between schools. Each parameter combination is replicated 100 times to achieve consistent results. As in Stoica & Flache’s model, each simulation represents one year of households enrolling their child into school, and all switches happen quasi-simultaneously: Each household evaluates and potentially switches schools in a random order, but the choices are treated as happening simultaneously, thus the choices of one household does not influence the choices of other households, except for potentially limiting their available schools due to occupying otherwise vacant slots.
Figure 3 shows the D-index after one time-step varied across the three different initial distributions of ethnicities. The results indicate that segregation can, as it does in Schelling’s model of residential segregation, arise among tolerant agents – under some initial population distributions. We observe mild segregation (with a mean D-Index of 0.26) in the 70-30 population distribution, and stronger segregation in the 80-20 distribution (mean D-Index of 0.43). These differences can in part be attributed to the measurement itself; the D-index is sensitive to changes in the population distribution. The differences are also consistent with expectations based on how the model is structured: Under these parameter settings, households are unsatisfied if they are in a minority smaller than 30%. In a randomly distributed population where the minority group constitutes 40%, there is a very small probability that any household will be in a minority small enough to actually switch schools. Thus, the mean D-index is less than 0.1, which is the expected value in a randomly distributed population. On the other hand, in the 70-30 and 80-20 population distributions, the probability of any given school having an ethnic distribution where the minority is small enough to become unsatisfied is larger. Among these three population distributions, the 70-30 mix is
perhaps the most interesting: In this case, our households are satisfied by an ethnic
distribution in their school that matches the overall population distribution. This leads to the
possibility of a state of equilibrium where all schools have a population distribution matching
the overall population, and all households are satisfied. The model shows that even though
such a state of equilibrium is possible, substantial levels of segregation are generated.

5.1.2 Simulation Experiment 2: Can a Preference for Minimizing
Distance Curb Segregation?

In the second experiment, we seek to further replicate the findings of Stoica & Flache, namely
that including a preference among households for minimizing distance limits the amount of
segregation. The underlying assumption is that parents not only want to ensure a substantial
number of same-group students surrounding their children, they also seek to minimize travel
distance between the home and the school. Including such a preference for minimizing
distance should limit the options for “fleeing” their allocated school in favor of one with a
more suitable ethnic composition, thus limiting overall levels of segregation.

The initial parameters are specified as experiment 1, with 6000 households and 25
schools. We vary the parameter governing the agents’ search radius between 20% (the
households only consider the five closest schools), 50% (the closest-half of all schools are
considered) and 100% (the households consider all schools). To mimic Stoica & Flache’s
model, we run the simulation for only one time-step; no new households enter the simulation,
and the simulation ends after each unsatisfied household has gone through the process of
evaluating alternative schools (with the potential result of enrolling their child in a different
school). The simulation is run for 100 repetitions for each parameter combination, resulting in
300 simulation runs. In this experiment, I initialize the model with a 70-30 population
distribution, where we can expect substantial levels of segregation among tolerant households
(M=0.8, F=0.7 and T = 0.4).

It is here important to note the difference between how the preference for minimizing
distance is modeled in this model versus Stoica & Flache’s model. In Stoica & Flache’s
model, the utility derived from the ethnic composition of a school and the utility derived from
the distance to a school is modeled as a Cobb-Douglas utility function: A weighting parameter
$\alpha$ is used to govern the relative importance of the two factors. This entails that as the
importance of minimizing distance to schools rises, the importance of ethnic composition
lessens. In this model, the importance of minimizing distance is instead modeled as a search
radius governing the percentage of all schools to be evaluated. I assume that if a consideration for distance has an impact on the processes generating segregation in schools, it should be evident in both models. The following figure shows a boxplot of all simulation runs organized by the search radius parameter:

Figure 5.2: Boxplot of D-index by households’ search radius, one year of simulation.

The figure shows a weak positive correlation between an increase in the households' search radius and the segregation outcomes: a larger search radius leads to higher levels of segregation. This is consistent with what Stoica & Flache finds running their simulation; introducing a preference for minimizing distance does in this case, as in theirs, curb the formation of segregation patterns, although the impact is less substantial. With a search radius of 1, i.e. all schools are evaluated when an unsatisfied household decides to switch schools, all unsatisfied minority-group members will cluster and fill up the schools that initially had the most favorable ethnic composition for minority-group members. When the search radius is lowered, i.e. to 0.2, each unsatisfied household only considers the five closest schools. This causes the “centralized” clustering tendency described above to break down and overall segregation levels are lowered. Note that the percentage of vacant slots in each school during these simulations is held at 25%: It is possible that the effect of limiting the search
radius is stronger if the share of vacant slots is increased, and this possibility will be addressed in section 5.1.4.

5.1.3 Simulation Experiment 3: Segregation Patterns Develop Over Time

So far, the model shows results consistent with those presented by Stoica & Flache: Patterns of school segregation can emerge as an unintended consequence of the behavior of tolerant parents, and these tendencies are curbed by including a preference for minimizing home-school distance. Next, I want to examine how these patterns evolve over time. This extension of the model seeks to capture the central aspect of Schelling’s original model, namely that segregation can arise through “tipping” processes: The choices of one individual can trigger the choices of other individuals, causing a cascade towards segregation. Do patterns of school segregation behave in a similar way? Assuming they do, one would expect that segregation solidifies and increases for each year of simulation.

The simulation is now run following the same initialization as in the previous two experiments: households are tolerant (M=0.8, F = 0.7 and T = 0.4), the population distribution is varied between 80-20, 70-30 and 60-40, and the households’ search radius is varied between 0.2, 0.5 and 1. I now let the simulations run for 25 time-steps, simulating 25 years of school enrollment, compared to a single time-step in the previous experiments. Every year, 6000 new households enter the simulation, and base their choice of school on the ethnic composition from the previous cohort of households. Each parameter combination is simulated for 100 repetitions to achieve greater consistency, resulting in a total of 900 simulation runs. The mean D-index for each year is then plotted.
Figure 5.3: D-index over 25 years according to households’ search radius and population distribution.
The first thing to notice is that, unsurprisingly, not much happens in the 60-40 population distribution setup, for reasons previously discussed: The probability of any given household to ever have their utility fall below the threshold of satisfaction is very low (approximately 0.06%, under the assumption that each school on average houses 250 students). Meanwhile, in the two other population distribution setups, there is a notable level of segregation after the first year of simulation. These levels proceed to grow over the course of the first ten years, before beginning to stabilize. This growth is in turn influenced by the search-radius parameter: When all schools are evaluated in the searching process among households, the D-index stabilizes at a higher level, compared to when only the closest 50% or 20% are considered. In addition, the difference is higher when observed after multiple years of simulation, compared to a single year. Note that the initial curve is steeper in the 80-20 population distribution setting: This is due to the fact that in such an initialization, the probability of all minority-group households being unsatisfied in the first year of enrollment is very high. This, in turn, leads to a massive movement of students during the first year, causing a rapid rise in segregation up to the maximum allowed by the capacity of schools and the search radius of the households.

What can be concluded from the above results? First of all, including a temporal aspect seems to be crucial: Levels of segregation measured after the first year differ greatly from levels observed after ten years. Thus, drawing conclusions solely based on observing the simulation after one year can result in an understatement of the potential segregation-generating dynamics of the model. Secondly, assessing the importance of the households’ search radius after one year of simulation might lead to the conclusion that its role in curbing the formation of segregation patterns is trivial, while the impact seems to be substantial if the simulation is run for an extended period of time. Still, compared to the findings of Stoica & Flache, the inclusion of such a preference is not enough to eliminate segregation patterns altogether, only to decrease overall segregation levels. This difference I attribute to the difference in modeling between my approach and that of Stoica & Flache: Due to the relationship between the preference for a certain ethnic composition and the preference for minimizing distance as represented in their model, maximizing the importance of distance implies eliminating the effect of a preference regarding ethnicity. This, in turn, might contribute to exaggerating their findings. On the other hand, in my model, the two preferences are independent of each other, which reveals a more modest effect of the distance-preference on segregation outcomes.
5.1.4 Segregation Dynamics Under Increased Mobility

As discussed above, levels of segregation seem to grow slightly over time in the 70-30 and 80-20 population distribution regimes. This observation is in line with the central findings of Schelling in his original model: Segregation among tolerant agents increases gradually due to the “tipping” mechanism found in their interaction at the micro level. In order to investigate this potential for self-reinforcement in the school model, I run the previously described experiment again, this time varying the percentage of extra capacity available in schools. The underlying assumption motivating this additional experiment is that the default amount of extra capacity specified until now, 25%, can serve to limit the growth of segregation patterns over time: Increasing this extra capacity in schools serves to increase the potential mobility, allowing households more freedom to maximize their utility when selecting schools. The extra capacity represents how flexible the schools are with regards to changes in the size of the student mass: One can assume that in a system where schools adjust their capacity based on their popularity, segregation-generating processes are operating with less restrictions.

Thus, in the next experiment, I run the simulation described above with a 70-30 population distribution. This time, the parameter governing the percentage of vacant slots in schools is varied between 5%, 25% (as in the above experiments), and 50%. As an example, when the parameter is set to 50%, a school with an initial population of 200 students has a maximum capacity of 300 students. On the other hand, if the parameter is set to 5%, the school is only allowed to accept ten students applying from outside of the schools catchment area. If the self-reinforcing tendencies described above work as we assume, one can expect that overall levels, and the growth, of segregation will increase as the share of vacant slots is increased. Conversely, I expect that segregation is limited when the parameter is set to 5%.
Figure 5.4: D-index over 25 years according to search radius and percentage share of extra capacity in schools.
The above figure shows the levels of segregation over 25 years in the 70-30 population distribution under the three different scenarios regarding the percentage of extra vacant slots in each school. The center plot is a repetition of the previous experiments: The percentage of extra vacant slots of each school is set to 25. This means that each school’s capacity each year is set to the number of enrolled students in the previous year, with additional vacant slots equal to 25% of this number. This allows for some mobility, and reflects that the most popular schools increase their size over time to stimulate demand. The left- and right-hand plots show the same simulations with the percentage of extra capacity decreased to 5% and increased to 50% respectively.

These results confirm the suspicion that the amount of vacant slots of schools serve to limit the overall levels of segregation. Increasing the amount of vacant slots, i.e. allowing for more mobility between schools, serves to accentuate the patterns observed in the previous experiment. In addition to overall higher levels of segregation, these results confirm to a greater degree the findings of Stoica and Flache: Incorporating a preference for minimizing home-school distance among households does curb overall levels of segregation, and this effect is more clearly visible when the amount of vacant slots is increased. When there is no consideration of distance involved in the households’ choices of schools, the unsatisfied minority-member households are to a greater extent free to cluster in those few schools with an initial distribution most favorable towards the minority. When, in addition, the schools are allowed to adjust their size to a greater extent (by having more vacant slots each year), the number of minority-member students allowed into each of those schools increases. On the other hand, these few minority-dominated schools will over time be vacated by now unsatisfied majority-member households, further reinforcing segregation patterns.

5.1.5 Robustness and Internal Validity

Most agent-based models describe a system or dynamic involving one or more stochastic features and, this model is no exception. The initial distribution of households and their background (majority or minority), and the distribution of threshold values among households constitute such features. To investigate whether or not the results of my simulation experiments are robust, i.e. to assess the internal validity of the model, I will run a large number of replications varying all parameters. These results will then be analyzed using a regression model, and compared to the results from the previously executed experiments. In
addition, such an analysis allows us to assess the relative effect of each parameter, independent of other variables, on the overall segregation levels.

Table 5.1: OLS regression, D-index as dependent variable. N = 3234 simulation runs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.115***</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.061***</td>
<td></td>
</tr>
<tr>
<td>Search radius</td>
<td>0.006***</td>
<td></td>
</tr>
<tr>
<td>Percentage extra capacity</td>
<td>0.039***</td>
<td></td>
</tr>
<tr>
<td>Majority Share</td>
<td>0.174***</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.981***</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3.240</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.604</td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.603</td>
<td></td>
</tr>
<tr>
<td>Residual Std. Error</td>
<td>0.184 (df = 3234)</td>
<td></td>
</tr>
<tr>
<td>F Statistic</td>
<td>986.831***  (df = 5; 3234)</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01; ***p < .001

Table 5.1: OLS regression, D-index as dependent variable. N = 3234 simulation runs.

The above table shows the results of a linear regression model with the D-index as the dependent variable, using the previously discussed parameters as covariates: The threshold value \( T \), the ideal share of similar households \( F \), the households’ search radius when evaluating schools, the percentage of extra vacant slots in schools, and the overall majority share in the population of households. The variables are normalized on a scale between 1 and 10.

The coefficients for each independent variable are significant to the 0.05 level. The strongest effects are found in the parameter governing the households’ threshold, \( T \), and the overall majority share of the population. This is to be expected: Increasing the threshold for when households are satisfied equals less tolerant households, which in turn implies higher levels of segregation. In this case, increasing the value of \( T \) by 0.1 leads to an overall increase of the D-index by 0.115. As for the overall majority share, the strong effect is in line with what we observed in the first simulation experiment: The smaller the average minority-share of schools, the higher the probability that the minority-group households’ utility will fall below the satisfaction threshold, thus causing them to switch schools. Finally, we see that
both increasing the households’ search radius and increasing the amount of extra vacant slots in schools are associated with higher overall levels of segregation, which supports my findings in experiment 2 and 3. In conclusion, this simple statistical analysis of the model output coincides with the results from my previous experiments, which implies that the previous results are robust. In addition, the regression reveals that the parameter with the weakest effect on overall segregation levels is the search radius parameter, indicating again that the effect of including a distance preference only has modest impact on overall segregation. This can be contrasted with the findings of Stoica & Flache, where their implementation of the distance-minimizing preference was sufficient to curb self-organized segregation entirely.

5.2 Empirical Calibration of the Model

In this second part of the chapter, we move on to empirical calibration of the model. The objective of this empirical calibration is to assess how the model dynamics explored in the first part of this chapter operate under more realistic initialization settings. In order to investigate this, the model is altered to represent an urban region in Oslo, compared to the abstract area represented until now. I use empirical data described in chapter 2, the k-nearest-neighbors-calculations centered on the schools, to specify the initial distributions of minority-majority-members surrounding each school.

In this section, I will revisit some of the simulation experiments conducted above, with the only difference being the initial population distribution and the geographical positioning of the schools. This will allow comparison between the purely abstract results and the empirically calibrated results: The central question to be investigated in this case is whether or not the model behaves differently when initialized with a more realistic population distribution. In practice, the theoretical simulation represents a geographical area with no residential segregation: This allows us to examine the potential effect of parental choice on school segregation, controlling for the effect of residential segregation. The empirical calibration, on the other hand, lets us explore the effect of parental choice in a system where there are preexisting patterns of residential segregation. In addition, I will compare the simulated student compositions of the schools with data on student compositions in their real-life equivalent schools of Oslo. The objective of this calibration is not necessarily to be able
to reproduce the segregation patterns found in Oslo, but rather to investigate the behavior of the model under more realistic initial conditions. In order to realistically “grow” the actual patterns of segregation in Oslo, the initial population would have to be mapped to the catchment areas (or school districts), and these residential patterns would have to be updated over time and specified in accordance with a real-life time span. Although an interesting direction for further work on the model, such a process is beyond the scope of this thesis. Hopefully, a comparison between simulated student compositions and real-life student compositions will still provide insight into the potential weaknesses of the modeling strategy.

5.2.1 Empirical Simulation Experiment: Segregation Under Realistic Initial Conditions

In this experiment, I will re-run the simulations examining the first and second hypothesis discussed above, namely, whether or not segregation can arise in schools among tolerant agents, and how the inclusion of a preference for minimizing distance affects this process. All parameter settings are maintained from the experiments conducted in the previous run, the only difference being the initial population distributions. The following figure shows the D-index over 25 years of simulation, and the search radius parameter, representing the households’ preference for minimizing distance, is varied between 0.2, 0.5 and 1. In addition, the amount of extra capacity is varied between two settings: 5%, which represents limited mobility between schools, and 25%, which represents more flexibility regarding the size of schools. Increasing the extra capacity of schools, thus increasing mobility, was shown to accentuate and increase segregation patterns. How similar is this effect under a more realistic population initialization?
Figure 5.5: D-index over 25 years according to search radius and percentage share of extra capacity in schools, empirically calibrated population distribution.
How do these results compare to those in the previous part of this chapter, where the population was randomly distributed according to three population distribution regimes? The first thing to note is that in this empirically calibrated population distribution, the baseline level of segregation is higher than in the randomly distributed populations used previously. In a randomly distributed population where every household enrolls their child in the local school, the baseline segregation is close (but, due to stochasticity, not equal) to zero. In this experiment, on the other hand, the expected baseline segregation is $D = 0.46$, i.e. if no household switches schools, the D-index would remain at 0.46.

Compared to the results from the random population distributions, these patterns are generally similar: When the extra capacity is limited (5%), the growth and persistence of segregation is curbed. When schools are allowed to adjust their capacity to a greater degree (25% extra capacity), the D-index is higher, to the point of reaching 1, i.e. maximum segregation, when households consider all schools (search radius = 1). The effect of introducing a limit to the households search radius is also similar: The larger the search radius (i.e. households consider a greater number of schools when making their selection), the higher the overall levels of segregation. Compared to the previous results, this effect is even greater in the empirically calibrated model, especially between the cases where households consider all schools (search radius = 1), and when they consider half or less of all schools (search radius = 0.5 or 0.2). This difference can be attributed to the fact that in the empirically calibrated initial distribution, some schools are initialized with a significantly larger share of minority member students than others. Whether or not these minority-dominated schools fall within or outside the search radius of a minority-member household located elsewhere thus matter greatly for the overall segregation levels.

Finally, I will compare the simulated distribution of students from the empirical model with the distributions found in empirical data from Oslo. The majority shares of students, as of 2014, are compared with two different simulation settings: First, the baseline setting, where no school switches are allowed. In this setting, every household enrolls their child into the local school, thus making school segregation equal to residential segregation. The second comparison is made between the simulated shares after having run the simulation for 5 time steps, equaling 5 years. For this purpose, I have chosen a modest set of parameter settings to approximate a more realistic setting; the extra capacity allowed in each school is set to 5% (to emulate the fact that real-life schools are most likely rigid in terms of adjusting their overall
capacity for each year), the households’ are tolerant \((F=0.7, M=0.8, T=0.4)\), and they evaluate the closest-half of all schools (search radius = 0.5).
Figure 5.6: Majority share of schools: Simulated shares compared with real-life shares.
The above bar graph shows the two comparisons: The topmost graph shows the majority share of each school in the baseline simulation setting, where the model is initialized, but no switches are allowed, compared to the empirical majority shares of each school. In the lower graph, the simulated shares are measured after ten years of enrollment, and the model behaves as described in the experiments conducted previously.

The first thing to note is the similarity between the baseline simulated shares and the empirical shares. There are some deviations, but overall patterns are very similar: This can be interpreted that residential patterns to a very high extent contribute to the school segregation patterns found in Oslo. What happens when the simulation is run, under certain conditions, for ten years? The graph above indicates that, in accordance with the experiments conducted above, the following process takes place: School 6, in this case, has a majority share small enough to cause majority-group members to be unsatisfied with their current school. These households then search for alternative schools, settling on schools where the majority-share is satisfying. This leads to overall higher majority shares in the rest of the schools, which in turn causes minority-group members to relocate to other schools. Schools numbered 2, 3, 7 and 8, that had high majority-shares in the baseline simulation, are now completely homogeneous, with a majority share of 1.

The results after simulating enrollment for 10 years deviate further from the real-life student compositions compared to our baseline experiment, instead of providing a closer match. This leads to one of two possible conclusions: Either the a), the mechanisms and assumptions made in the construction of the model are wrong, or b) that implementing additional mechanisms is required to capture the dynamics generating segregation patterns. In this case, I argue that both conclusions are possible, although it is reasonable to expect that the abstractions made in the design of this model do not come close to a realistic simulation of school enrollment in a real-life system. The dynamics revealed in the theoretical experiments, namely that the growth and persistence of segregation patterns is possible even in a population of tolerant agents, could still be at work in real-life settings - but additional mechanisms, and modifications of the mechanisms currently included, are required to capture these dynamics in a realistic way.
5.3 Summary and Discussion of Results

In the first part of this chapter, I have conducted three separate simulation experiments in order to investigate different aspects of the model’s dynamics. The first experiment concerned whether or not segregation can arise among tolerant agents, as it does in Schelling’s famous model of residential segregation. The second experiment seeks to answer the question of how a preference among households for minimizing home-school distance factors into the process: Does such a preference, as it does in Stoica and Flache’s model, serve to curb the formation of segregation patterns? Finally, the third experiment investigates how the simulation behaves if allowed to run over several years, as opposed to just one. In addition, the final experiment is investigated under different levels regarding mobility between schools, by increasing the amount of vacant slots in each school. In the introduction to this chapter, I formulated a set of three hypotheses:

1. Segregation can arise among tolerant agents in schools, as it can in residential neighborhoods.
2. Including a preference among households for minimizing home-school distance can curb these tendencies towards self-organized segregation.
3. School segregation-generating tendencies are self-reinforcing and increase over time.

Regarding the first hypothesis, the results of the first experiment indicate that segregation can arise among tolerant agents in schools, as it can in neighborhoods. This confirms that the model works as intended, in that it captures the central tendencies of Schelling’s model of residential segregation. Segregation levels are higher when the overall ratio between majority- and minority-member households is uneven, but this can in part be attributed to the D-index being sensitive to changes in the overall population.

The second hypothesis is supported by the second simulation experiment, and more insight is gained when the temporal aspect of the model is included. When observed after only one time step, the model shows that limiting the households’ search radius has a weak, but positive effect on overall levels of segregation. This effect is increasingly visible when the model is run for 25 years. When running the simulations under increased mobility (by
increasing the amount of vacant slots in schools), the effect of the distance preference is has an even stronger impact on overall segregation levels.

The third hypothesis is also supported by my findings. When households make their choices of schools based on the ethnic composition found in the previous year, segregation is reinforced, and thus grows over time. Thus underlines the importance of including a temporal aspect when analyzing segregation dynamics. These findings are similar to the dynamics found in Schelling’s original model of residential segregation, although the mechanisms causing the stabilization of segregation over time operates in a different manner. In Schelling’s model, agents are allowed to relocate an infinite number of times, thus creating the tipping-effect over time. In my model, households only have one opportunity to switch schools. This is a modeling decision made for purposes of maintaining some realism: I argue that parents continuously switching between schools for their children does not accurately represent the real-life equivalent process. Thus, I argue that this difference creates a natural limit to how much segregation can “grow” over time, partly governed by the amount of extra capacity of each school. In Schelling’s model of residential segregation, this limit is achieved when all households are satisfied (or fail to find a location in which they would be).

Summarizing the results from exploring the theoretical model, I will conclude with the following, based on the three hypotheses: Segregation patterns can arise in schools as an unintended consequence of behavior among tolerant parents. The level of segregation is higher when schools are allowed to adjust their capacity to a greater degree, according to the number of applicants they receive each year. Finally, and particularly in the cases where there is such a large degree of vacant slots in schools, the growth of segregation is curbed when a parental preference for minimizing home-school distance is included. Thus, my model supports the findings of Stoica & Flache, in addition to showing the importance of observing the dynamics of segregation over time.

Finally, the empirically calibrated experiment shows that the dynamics investigated in the theoretical experiments also operate in a similar manner when initialized with a realistic population distribution. While the results from the theoretical model indicate that parental choice has the potential to cause substantial levels of school segregation by itself, without preexisting patterns of residential segregation, the empirically calibrated experiments indicate that the effect of parental choice, especially when no preference or regulations regarding distance is involved, is stronger when there are preexisting patterns of residential segregation. Regarding real-life school systems, it is evident that patterns of residential segregation
constitute an important cause of the growth and persistence of school segregation. However, parental choice has the potential to increase these levels, especially in cities where residential patterns are segregated as well.

Comparing the simulated results to actual student compositions in the area of Oslo represented, however, reveals that the model lacks either the central assumptions or the sophistication to realistically simulate real-world patterns of segregation. The comparison reveals that the baseline simulation patterns, i.e. when households simply enroll their children in their local schools, constitute the closest match to empirical reality. One possibility is that in the real-life case of Oslo, other mechanisms are counteracting the segregation-generating effects that my model simulates. Another possibility is that the behavior rules included in the model simply fail to capture the actual behavior of parents in the Oslo region. Data on the frequency and patterns of actual school switching in Oslo would go a long way in providing further insight into this problem. It is also possible that there is a mismatch between analytical strategy and empirical *explanandum*: In the case of Oslo, residential segregation might be the principal component in explaining school segregation. In other school systems, parental choice might have a more severe impact. This could be investigated by initializing the population distributions and geographical locations of schools based on empirical data from other real-world cities.
6 Discussion and Conclusions

The purpose of this thesis was to explore how different transformational mechanisms generate patterns of school segregation, by analyzing an agent-based model. The two mechanisms I have modeled are two preferences among parents assumed to have an effect on school segregation, derived from previous research on the causes of segregation patterns. These preferences revolve around the ethnic composition in certain schools, and minimizing the geographical distance between home and school. More specifically, I wanted to investigate whether the dynamics found in Schelling’s model of residential segregation also operate in a similar model of school segregation, and also how including a preference regarding geographical distance affects this dynamic. Ideally, such a model can create a framework for further theoretical exploration and the testing of hypotheses by extending and altering the parameters of the model. Additionally, this framework can potentially be modified to apply to segregation in other areas apart from schools, for instance organizations or labor-markets.

6.1 Points of Weakness of the Approach

The overall goal of this approach has been to explore theoretically the mechanisms causing the persistence and growth of ethnic segregation in schools. In many ways, the model is only a step in the direction of a full-fledged simulation of the school enrollment process, and the next section will detail possible extensions to be made in future incarnations. Still, some fundamental issues need to be summarized and addressed.

First of all, the model is lacking in realism to the point where comparison with real-life patterns is difficult. Serious investigation of the consequences of for instance policy changes has not been addressed in this thesis, due to the lack of both detail in the model itself, and the lack of realistic empirical approximation to data. Thus, the external validity of this model is severely limited – at its current stage it functions as a tool for exploring theoretical implications, but approximating a real-world scenario needs further extension and sophistication. Second, the sophistication of the actual behavioral rules operating in the model can be criticized for lacking detail. This is in part due to the fact that only two preferences are modeled: the preference for avoiding being in too small of a minority, and the preference for minimizing home-school distance. An example of the problems caused by the lack of realism
can be observed in how the segregation patterns generated in this model are usually the product of minority-group members choosing schools with a substantial share of other minority-group students, while in reality; research suggests that the “flight” of majority-group members away from such schools also plays an important role. This can potentially be addressed by the model in its current state, by experimenting with heterogeneous preferences between groups: For instance, assuming the majority-group has a lower tolerance for the share of minority-group members than the minority-group members themselves.

Finally, there is a central problem with theoretical agent-based models, which was discussed in chapter 3, section 3.2.4. While this model shows that the mechanisms I have included in the model are sufficient in generating substantial levels of segregation by themselves, this does not necessarily imply that real-world patterns of segregation come about due to these mechanisms. This is a problem which all agent-based models to some degree face: How can we know that the mechanisms we have postulated are in fact the mechanisms operating in the generation of real-world segregation patterns? As a potential answer to this, I will underline that the model shows a candidate explanation for the growth and persistence of school segregation, which derives its assumptions from previous empirical research: Yet it does not necessarily show the correct explanation of school segregation. By adding and examining other potential explanatory mechanisms to the model, it is possible to investigate whether there are other candidate explanations which could generate similar (or even more realistic) patterns. For this thesis, I have opted for a simple, abstract approach, in order to preserve explanatory clarity. There are several potential directions for adding sophistication to the model.

### 6.2 Possible Extensions of the Model

As mentioned above, this agent-based model, even in its empirically calibrated form, still does not take into account other possible explanations of the processes generating school segregation. There are several possible extensions to the model that might prove fruitful.

First of all, empirical studies of the considerations made when choosing schools provide a few concrete motivations stated by parents. One of these is the reputation of a given schools educational quality. If the share of minority students correlates with the schools achievements on national tests, what could be perceived as parents moving their children to “whiter” schools may instead be parents moving their children to schools with a better
academic reputation. In order to include a preference for academic quality in this model, some important considerations are to be made: Under which assumptions do we model the relationship between minority/majority status and academic performance? How, precisely, should academic performance be measured? Would this preference be modeled separately from the preference for a certain ethnic composition? These questions lead to challenges, but answering them adequately could provide a more fine-grained picture of school segregation dynamics.

Second, there is an argument to be made for including social networks in the model. Kirsten (2008) notes that Turkish-origin parents in Germany perceive fewer options when selecting schools due to a lack of information, thus choosing schools with a large minority share and reinforcing segregation. Let us say we assume that parents’ networks of friends, and the friendship network of their children themselves, play a role when determining which school to choose. If these networks are ethnically homophile, the resulting behavior might lead to higher levels of segregation – in spite the lack of a preference for certain ethnic compositions. An extension of the model allowing for networks with varying degrees of network homophily and their effects on behavior would be a fruitful direction for further work.

In addition, further extensions can be made to the empirical calibration of the model. In its current incarnation, the empirically calibrated model is still abstract - the population distribution is calibrated to match the actual distribution located around the schools, but a point of criticism can be made with regards to the realism of this process. For instance, the actual distribution to be matched is, more realistically, the share of minority/majority households with children at the given ages for entering schools. When such data is available, the actual school districts should be mapped with GIS coordinate data to more closely replicate the actual distributions of students enrolling in each school. The population in the actual school districts - not just the surrounding neighborhoods – can then be represented. At the same time, this process can be made even more realistic by mapping the population distribution over a specific time frame, simulating the changes in residential patterns over time. On the school-side of the model, the actual capacity of schools and the variance of capacity over time could also be calibrated to empirical data. This would produce a more fine-grained empirical calibration, leading to results with increased external validity.

Finally, the model in its current incarnation details the effect of parental choice of schools given fixed residential locations. A venue for further exploration is to actually merge
the model with a Shelling-type model of residential segregation: Households are both allowed
to change their residential location in addition to choosing their schools. With such a model, it
would be possible to explore the possible effects of schools on residential mobility: For
instance, parents moving out of certain neighborhoods when they have a child ready for
school enrollment in order to secure a slot at a school they consider attractive. This leads to
the need for a more complex utility calculation on behalf of households, but combining a
residential model with a school model is without doubt an interesting direction for further
work on segregation patterns.

In addition to extending the model further, my results can be interpreted as having
some implications for further empirical work. As discussed in chapter 3, agent-based models
can be of value in providing directions for further empirical investigation. My model it
underlines the point provided by Schneider and Buckley (2002): Even if survey data doesn’t
reveal a tendency for choices based on ethnic preferences, these preferences (even while mild)
can still generate heavily segregated school systems. Taking this into account, further
empirical work on the emergence of school segregation might benefit from focus on actual
behavior instead of surveying motivations and preferences. In other words, while survey data
is prone to be influenced by what is deemed socially acceptable reasons for choosing certain
schools, actual behavior is not.

6.3 Conclusion

Agent-based modeling has increasingly gained traction as a tool for investigating complex
macro-level phenomena within sociology. In this thesis, I have explored the application of
agent-based modeling on the question of school segregation. Building on the theoretical
framework found in analytical sociology, I argue that agent-based modeling is suitable as a
tool for exploring transformational mechanisms: The processes where interaction and
behavior of individuals at the micro-level shape and bring about complex patterns at the
macro-level. Building on the framework developed by Stoica & Flache (2014), I have
developed an agent-based model representing an urban area with households selecting schools
for their children. Analysis of the behavior of the model has shed light upon the processes
generating school segregation as a result of ethnic preferences among households, and a
preference for minimizing distance to schools. The simulation results support the original
findings of Schelling’s model of residential segregation, namely that segregation can
theoretically arise as an unintended consequence of interaction between tolerant agents – in schools as well as in neighborhoods. In addition, the model further develops knowledge on the interplay between preferences for minimizing distance and the previously mentioned Schelling-type preferences for ethnic composition in schools. I find support of the findings from Stoica & Flache’s model, although with some moderations. A preference for minimizing distance among households has the potential of limiting overall segregation levels, but not the ability to eliminate segregation patterns completely.

The results also underline the importance of including a temporal aspect when investigating segregation-generating processes, by revealing different segregation outcomes over time. In this regard, the most central finding of the model is that segregation patterns rapidly self-reinforce and grow, before stabilizing, over time. This leads to the conclusion that measuring segregation over time provides a different, and perhaps more nuanced perspective. Further extension of the model can potentially allow for more fruitful comparison with real-life segregation patterns, and allow for greater realism. Finally, there is the question of enrollment policy. The results of analyzing this model shows the potential effect of even a just a modest preference among parents for keeping their children away from being in a too small ethnic minority. Substantial segregation can arise under conditions where parents actually prefer an integrated, multiethnic environment. The implications of this for enrollment policy seeking to limit segregation are perhaps bleak, but it is again important to remember the high level of abstraction this model is based on. Perhaps, when this framework is expanded and empirically calibrated to a more realistic level, policy changes and their effects can be modeled and investigated.
References


All sources used in this thesis are stated.

Number of words in this thesis, including tables and figures: 26110.
Appendices

Appendix A: NetLogo code of the model

breed [households household]
breed [schools school]

globals [  
total-red  
total-green  
total-households  
]
schools-own [  
curr-n-total  
curr-n-green  
curr-n-red  
prev-n-total  
prev-n-green  
prev-n-red  
capacity  
dissim  
being-considered  
full?  
temp-ethnpref  
temp-distpref  
temp-utility  
]

households-own [  
happy?  
near-school  
target-school  
n-same  
n-total  
optimal-n-same  
ethnpref  
distpref  
utility  
moved?  
]

to setup  
  clear-all  
  reset-ticks  
  spawn-schools  
  edit-schools  
end

to go  
  spawn-households  
  set total-red count households with [color = red]
set total-green count households with [color = green]
set total-households n-of-households
enroll
ask schools [ set capacity (round (curr-n-total + (curr-n-total * perc-extra-capacity / 100))) ]
if ticks = 0 [ ask schools [set-current-to-previous] 
ask schools [calc-dissim] ]
ask households [check-happy]
move-unhappies
ask households [re-check-happy]
ask schools [update-numbers]
ask households [ die ]
ask schools [calc-dissim]
ask schools [ 
ifelse prev-n-green >= prev-n-red 
[ set color green ]
[ set color red ]
]
tick
end

;## Setup commands

to edit-schools 
ask schools [ 
set shape "house"
set size 5
set color gray
set curr-n-green 0
set curr-n-red 0
set curr-n-total 0
set full? false
]
end

to spawn-schools 
ask patch 16 25 [sprout-schools 1]
ask patch 7 45 [sprout-schools 1]
ask patch 11 55 [sprout-schools 1]
ask patch 32 5 [sprout-schools 1]
ask patch 56 12 [sprout-schools 1]
ask patch -16 1 [sprout-schools 1]
ask patch -31 -48 [sprout-schools 1]
ask patch 10 -23 [sprout-schools 1]
ask patch 49 -19 [sprout-schools 1]
ask patch 37 -46 [sprout-schools 1]
ask patch -5 17 [sprout-schools 1]
ask patch -68 -47 [sprout-schools 1]
ask patch -7 -64 [sprout-schools 1]
ask patch -59 -67 [sprout-schools 1]
ask patch -53 45 [sprout-schools 1]
ask patch -18 -29 [sprout-schools 1]
ask patch -39 -14 [sprout-schools 1]
ask patch 60 61 [sprout-schools 1]
ask patch 58 29 [sprout-schools 1]
ask patch 59 -48 [sprout-schools 1]
ask patch -24 64 [sprout-schools 1]
ask patch -26 36 [sprout-schools 1]
ask patch -35 7 [sprout-schools 1]
ask patch 58 -69 [sprout-schools 1]
ask patch -55 21 [sprout-schools 1]
end

;## Go commands

to spawn-households
  ask n-of n-of-households patches [
    sprout-households 1
  ask households [
    set shape "house"
    set size 2
    set color red
    set moved? false
  ]
  ask n-of (round (n-of-households * perc-green)) households [
    set color green
  ]
end
to enroll
  ask households [
    set near-school min-one-of schools [distance myself]
    ifelse color = green [
      ask near-school [
        set curr-n-green (curr-n-green + 1)
        set curr-n-total (curr-n-total + 1)
      ]
    ]
  ]
to set-current-to-previous
  set prev-n-green curr-n-green
  set prev-n-red curr-n-red
  set prev-n-total curr-n-total
end
to check-happy
  ifelse color = green [
    set n-same [prev-n-green] of near-school
    set n-total [prev-n-total] of near-school
    set optimal-n-same (n-total * ideal-similar-wanted)
    ifelse n-same <= (n-total * ideal-similar-wanted) [
      set ethnpref (n-same / optimal-n-same)
    ]
  ]

set distpref 1
set utility (ethnpref ^ weight * distpref ^ (1 - weight))
ifelse utility >= random-normal threshold 0.05
[ set happy? TRUE ]
[ set happy? FALSE ]
end

to re-check-happy
if moved? = TRUE [ 
ifelse color = green
[ set n-same [prev-n-green] of target-school ]
[ set n-same [prev-n-red] of target-school ]
set n-total [prev-n-total] of target-school
set optimal-n-same (n-total * ideal-similar-wanted)
ifelse n-same <= (n-total * ideal-similar-wanted) [ 
set ethnpref (( n-same + 1) / (optimal-n-same + 1) ) 
]
[ set ethnpref (homogen-pref + (((n-total - n-same) * (1 - homogen-pref)) /  (n-total * (1 - ideal-similar-wanted))))]
set utility ethnpref
ifelse utility >= random-normal threshold 0.05
[ set happy? TRUE ]
[ set happy? FALSE ]
]
end
to search-schools
let close-schools min-n-of (round (count schools * search-radius-perc)) schools [distance myself]
ask close-schools [ set being-considered true ]
ask schools with [being-considered = true] [ 
if full? [set being-considered false] ]
ask schools with [being-considered = true] [ 
ifelse [color] of myself = green
[ ifelse prev-n-green <= prev-n-total * ideal-similar-wanted [ 
set temp-ethnpref ((prev-n-green + 1) / (prev-n-total * ideal-similar-wanted + 1)) 
]
[ set temp-ethnpref (homogen-pref + (((prev-n-total - prev-n-green) * (1 - homogen-pref)) / (prev-n-total * (1 - prev-n-total * ideal-similar-wanted))))]
]
[ ifelse prev-n-red <= prev-n-total * ideal-similar-wanted [ 
set temp-ethnpref (prev-n-red / prev-n-total * ideal-similar-wanted) ]
]
set temp-ethnpref (homogen-pref + (((prev-n-total - prev-n-red) * (1 - homogen-pref)) / (prev-n-total * (1 - prev-n-total * ideal-similar-wanted))))
]
prev-n-total * ideal-similar-wanted)))
]
]

set temp-utility temp-ethnpref
set target-school max-one-of schools [temp-utility]
ask schools [
  set temp-utility 0
  set being-considered false
  set temp-ethnpref 0
  set temp-distpref 0
]
end

to move-unhappies
  ask households with [happy? = FALSE] [
    if moved? = false [
      search-schools
      move
    ]
  ]
end

to move
  ifelse color = green [
    ask near-school [set curr-n-green (curr-n-green - 1)]
    ask near-school [set curr-n-total (curr-n-total - 1)]
    ask target-school [set curr-n-green (curr-n-green + 1)]
    ask target-school [set curr-n-total (curr-n-total + 1)]
  ]
  [ ask near-school [set curr-n-red (curr-n-red - 1)]
    ask near-school [set curr-n-total (curr-n-total - 1)]
    ask target-school [set curr-n-red (curr-n-red + 1)]
    ask target-school [set curr-n-total (curr-n-total + 1)]
  ]
  ask target-school [ if curr-n-total = capacity
    [ set full? true]
  ]
  set moved? true
end

to calc-dissim
  set dissim sqrt(((prev-n-red / total-red) - (prev-n-green / total-green)) ^ 2)
end

to update-numbers
  set prev-n-green curr-n-green
  set prev-n-red curr-n-red
  set prev-n-total curr-n-total
  set curr-n-green 0
  set curr-n-red 0
  set curr-n-total 0
  set full? False
end