Mechanisms for increased school segregation relative to residential segregation: a model-based analysis

Dignum, E.; Athieniti, E.; Boterman, W.; Flache, A.; Lees, M.

DOI
10.1016/j.compenvurbsys.2022.101772

Publication date
2022

Document Version
Final published version

Published in
Computers, Environment and Urban Systems

License
CC BY

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Mechanisms for increased school segregation relative to residential segregation: a model-based analysis

Eric Dignum a,⁎, Efi Athieniti a,⁎⁎, Willem Boterman a,⁎⁎⁎, Andreas Flache d, Michael Lees a

a Computational Science Lab, University of Amsterdam, 1090 GH Amsterdam
b The Cyprus Institute of Neurology & Genetics, 1683 Nicola - Cyprus
c Urban Geographies, University of Amsterdam, 1018 WV Amsterdam
d Faculty of Behavioural and Social Sciences, University of Groningen, 9712 TS Groningen

ARTICLE INFO

Keywords:
Agent-based modelling
School choice
School segregation
Residential segregation
Complex systems
Complexity

ABSTRACT

Excess school segregation is a phenomena observed across many countries and one common explanation from the literature is the hypothesis that parents might want to live in a diverse neighbourhood, but when it comes to their children, they are less tolerant with respect to school compositions. This study uses an agent-based model where households face residential decisions depending on neighbourhood compositions and make school choices based on distance and school compositions. Results indicate that increased school segregation relative to residential segregation can be observed in large parts of the parameter space, even when the tolerance for households belonging to the other group is equal for neighbourhood and school compositions. Our results demonstrate that asymmetric preferences are not a requirement for excess school segregation and show that residential segregation combined with distance preferences play a key role in this increase.

1. Introduction

Residential and school segregation are often associated with the reproduction of inequalities, while at the same time, the educational system is considered to be at the heart of policies for ensuring equal opportunity, decreasing inequalities and promote integration (Butler & Hamnett, 2007). Hence, school segregation obstructs this and is therefore considered a major societal problem (Boterman, 2019).

Studies have identified two main factors that influence school segregation and that connect it to residential segregation: distance and school composition (Boterman, Musterd, Pacchi, & Ranci, 2019; Hastings, Kane, & Staiger, 2005; Oosterbeek, Sövágó, & van der Klaauw, 2021). Parents have the tendency to choose a school close to home and hence residential segregation patterns will be—partially—reflected in the population of schools. Households also prefer a school with a higher proportion of students from socio-economic/ethnic backgrounds similar to their own (i.e., homophily). This homophily is not only found for school choice, but also for residential choice (McPherson, Smith-Lovin, & Cook, 2001; Musterd, Van Gent, Das, & Latten, 2016).

When parents prefer a school or neighbourhood with a larger share of their own group or tend to avoid one with a small share, this obviously impacts segregation and might induce a self-reinforcing process. By choosing a neighbourhood to live in or a school to attend, a household directly influences the composition. As other households might base their decision on that changed composition, they interact and their choice is affected by that first household. Their decision could, in turn, influence the subsequent choices of all other households. Empirically this might be reflected in phenomena such as white flight, where households of a particular group opt-out of schools and/or neighbourhoods to avoid an undesirable school composition (Cordini, Parma, & Ranci, 2019; Renzulli & Evans, 2005). Hence, these factors can contribute to the sorting of children with particular characteristics (e.g., ethnicity, religion, gender, wealth, age, social class) between schools/neighbourhoods and hence result in segregation (Harris, 2016) as a collective outcome of individual school choices.

Strikingly, in various educational systems, the level of school segregation is reported to be consistently higher than that of residential segregation (Bellei, Contreras, Canales, & Orellana, 2018; Boterman, Musterd, Pacchi, & Ranci, 2019; Johnston, Burgess, Wilson, & Harris, 2006; Wilson & Bridge, 2019). Within the literature, one hypothesis is that parents might want to live in a diverse neighbourhood, but when it comes to their children, they are less tolerant with respect to school compositions and opt-out of neighbourhood schools (Cordini et al., 2019; Logan, Oakley, & Stowell, 2008; Renzulli & Evans, 2005). Dutch
schools are also found to be more segregated than their neighbourhoods (Boterman, 2018). However, there is also evidence suggesting Dutch parents find it desirable for their children to attend diverse schools (Boterman, 2013), making substantial school segregation undesirable from a parental as well as societal perspective.

Self-reinforcing processes or feedback loops as explanation of segregation are shown to emerge in theoretical school choice models (Stoica & Flache, 2014) and have been identified in both the empirical (Card, Mas, & Rothstein, 2008) and theoretical residential choice literature (Schelling, 1971). Moreover, drawing on Schelling’s famous argument, Zhang (2011) and Sage and Flache (2021) reason that the substantial segregation patterns do not emerge because people really want it (i.e., they are relatively tolerant for the other group), but segregation emerges in the first place because integration is an unstable state of the system and segregation remains because it is a stable state. Hence, both residential and school segregation are argued to be distinct emergent processes, but they are also intertwined as school choices are partially dependent on residential segregation (i.e., via distance preferences).

If this is the case, it is not immediately clear if we necessarily need to assume that parents are less tolerant in school choices to produce excess school segregation. Schools may segregate more (less) without the influence of residential patterns, if composition preferences are strong (weak) enough, because residential segregation is defined at a different spatial scale. Or, in other scenarios, residential segregation could push schools to an already segregated state if only distance preferences are at play, generating uneven distributions in schools that would drive even further segregation if school composition is introduced into parents’ decisions.

In a recent study, Dignum, Boterman, Flache, and Lees (2021) argue that these and other dynamics of school choice constitute a complex system and that earlier research neglected this complexity by primarily focusing on explaining the individual school choice behaviour of parents, implicitly assuming that explanations of collective level outcomes such as school segregation could be straightforwardly deduced as the sum of the isolated individual components. However, this might be insufficient to develop a correct understanding of the problem (Maes, 2022) as they often involve feedback loops, non-linearity and adaptation of the parts, resulting from the interactions of the individual components and their environment (Miller & Page, 2009). Therefore, a complexity perspective employing Agent-Based Models (ABM), as argued for by (An et al., 2021; Bruch & Atwell, 2015; Conte et al., 2012) and used by Zhang (2011) for modelling residential segregation and Stoica and Flache (2014); Sage and Flache (2021) for school segregation, might provide other explanations as to why the often observed gap between residential and school segregation can emerge.

This study examines an alternative hypothesis to explain differences in residential and school segregation. Specifically, we explore whether self-reinforcing segregation dynamics can induce levels of school segregation that exceed residential segregation and what conditions (or parameters) of these dynamics are likely to generate maximal excess segregation. Thus, we ask to what extent and under which conditions can school segregation levels exceed those of residential segregation and what is the role of these residential patterns, while keeping the tolerance level of parents/households for neighbourhood and school composition equal? This paper will examine the hypothesis by combining a residential and a school choice model into one ABM. Additionally, it extends them by using a local and global neighbourhood perception simultaneously, removing satisfaction thresholds, as well as a general preference function for both choice processes and a systematic analysis of the entire parameter space.

2. Methods

This section describes the technical details of the ABM developed for simulating residential and school choices. The model is inspired by the ABMs used in Stoica and Flache (2014); Sage and Flache (2021), Schelling (1971) along with some extensions. For a clear and concise description of the model the Overview, Design concepts, Details (ODD) protocol (Grimm et al., 2010) is used. Note that the terms households, agents, parents and children are used interchangeably throughout this section as well as city and environment.

2.1. Basic principles and purpose

The model is based on the evidence that distance and composition are important factors in school choice (Böhlmark, Holmlund, & Lindahl, 2016; Burgess, Greaves, Vignoles, & Wilson, 2015; Chumacero, Gómez, & Paredes, 2011; Dignum et al., 2021; Glazer & Dotter, 2017). Moreover, substantial residential segregation is present in numerous cities around the world (Mustard, 2020; Taeuber & Taeuber, 2008), hence, one is inclined to find similar people living close to each other. If households prefer schools close to their homes, residential patterns will be reflected in schools to a certain extent, making residential segregation an important factor in determining school segregation. Furthermore, homophily, the tendency to be attracted to similar people (McPherson et al., 2001) is not only found in school choice but also for place of residence (Clark, 2019; Musterd et al., 2016; Smith, McPherson, & Smith-Lovin, 2014). Therefore, households have a specific level of tolerance for out-group people (i.e., the converse of homophily) when choosing a residential location, while the school choices implement a trade-off between tolerance and distance. As mentioned in the introduction, it is often assumed that the discrepancy between school and residential segregation emerges because parents are less tolerant with respect to school composition for their children than they are to compositions of their neighbourhood. The purpose of this study is to investigate whether having the same tolerance for dissimilar people in the residential and school choice, can produce higher levels of school segregation compared to residential segregation. If the model can demonstrate this, it would show that it is possible to produce a higher level of segregation in schools through alternative mechanisms, and that this does not necessitate a higher level of intolerance for school choice. Hence, one is able to observe if the discrepancy between both types of segregation still emerges and under what conditions we would expect maximal difference.

2.2. Entities, state variables and scales

2.2.1. Environment

The environment (or city) is a two-dimensional $N \times N$ lattice containing two types of household agents, reds and blues. In the US for example, this is often represented by Black/White (Reardon & Owens, 2014), but this indicator can be thought of as any characteristic defining households, such as ethnicity (Boterman, 2019), income (Oosterbeek et al., 2021), social status (Candipan, 2020), and so on. The population ratio is 50-50 and the number of occupied cells is controlled by the density $d \in (0,1)$. Note that a 50-50 ratio can resemble every two-group combination with equal shares in the population, so 30-30 is also possible for example. Moreover, the environment can resemble a city, but also smaller spatial scales where 50-50 might be a more realistic scenario.

There are $m$ neighbourhoods on the lattice and an equal number of schools, with each school placed in the centre of the neighbourhoods as

---

1 The full model code and resources used within this paper are available online at https://gitlab.computationalscience.nl/edignum/school-choice-understanding-segregation.
shown in Fig. 1a. In this way a one to one mapping of neighbourhoods to schools is established. Note, that the grid is assumed to be a regular grid and not a torus. Although this seems rather artificial, in reality, various cities have neighbourhood schools and/or schools are assigned catchment areas (Wilson & Bridge, 2019).

2.2.2. Households

Following the basic principles, households (agents) make residential and school choices. School choice is based on distance to, and composition of the school, whereas the residential decision is based on the neighbourhood composition only. One might argue that neighbourhood composition is not the most important nor only factor in residential choice. However, as it is deemed very important in school choice and to study the effect of equal tolerance, it is necessary to incorporate it in both processes. Nevertheless, neighbourhood composition is found to be of influence in diverse contexts, such as Singapore (Wong, 2013), the Netherlands (Van Gent, Das, & Musterd, 2019), the US (Clark, 2002) and is correlated to many other factors in residential choice. Hence, it could serve as an approximation of why one finds similar people clustering together in neighbourhoods (McPherson et al., 2001). Additionally, although the process generating the residential patterns might be far from realistic, the patterns themselves can be very diverse as composition preferences can vary according to four parameters (Eq. 4, 5). In the end this is what determines the distances to the different schools and what drives (part) of the choice.

To distinguish different choices for agents, a utility-based framework is used. Both empirical and theoretical work in residential (Musterd et al., 2016; Zhang, 2004) and school choice (Candipan, 2020; Hastings, Kane, & Staiger, 2009; Stoica & Flache, 2014) employ this methodology to study why households make the choices they do. In this model, agents continually make decisions to maximise their utility at each time step. The utility calculation and the decision process are described below.

2.2.2.1. Tolerance. All households have an optimal fraction (i.e., tolerance/homophily) of similar neighbours ($t_i \in [0,1]$), which is fixed throughout the simulation. For the analysis, the optimal fraction is equal for all agents and also the same for the school as for the neighbourhood composition ($t_{si} = t_{si} = t$).

2.2.2.2. Residential and school utility. Residential utility (3) depends only on the neighbourhood composition ($C_{ni}$), whereas utility for a school (1) depends on the school composition ($C_s$) and normalised distance ($D_a$) to a school (Stoica & Flache, 2014). More specifically, the utility of agent $i$ for a school $s$ is defined as follows:

$$U_i = C_{ni}D_a^{\alpha}$$

(1)

Where $0 \leq \alpha \leq 1$ is the parameter which controls the relative weight of composition over distance, hence, $\alpha=0$ means only distance is considered important in school choice. The distance of agent $i$ to school $s$ ($D_{ai}$) is normalised and calculated as:

$$D_{ai} = \begin{cases} \frac{\text{dist}_{\text{max}} - \text{dist}_{ai}}{\text{dist}_{\text{max}} - \text{dist}_{\text{min}}} & \text{if dist}_{ai} \leq \text{dist}_{\text{max}} \\ 0 & \text{otherwise} \end{cases}$$

(2)

Here $\text{dist}_{ai}$ is the Euclidean distance between a school $s$ and an agent $i$. It is normalised with respect to $\text{dist}_{\text{max}}$, the distance to the furthest school and $\text{dist}_{\text{min}}$, the closest school. The difference between the maximum distance and actual distance to the school for the agent is divided by the maximum distance that can be travelled (Eq. 2). Note that these minimum and maximum distances can differ per household. Residential utility depends on composition only, hence $\alpha=1$ and distance is irrelevant:

$$U_i = C_{ni}$$

(3)

For both residential and school choice, $C_i$ follows from the asymmetric utility function also used in Stoica and Flache (2014), with $x_i$ the proportion of similar agents in the neighbourhood/school for household $i$.

$$C(x_i, t_i, M) = \begin{cases} \frac{x_i}{t_i} & x_i \leq t_i \\ \frac{M(1-x_i)(1-M)}{1-t_i} & x_i > t_i \end{cases}$$

(4)

This class of preference functions is more commonly known as single peaked preference profiles (Ballester & Haeringer, 2011). It is maximum when the number of similar agents is equal to the optimal fraction ($t_i$) and there is a penalty for a homogeneous neighbourhood/school ($M$). Estimates of residential preferences from empirical data in the US (Clark & Fossett, 2008), Singapore (Wong, 2013) and for schools in the

Fig. 1. Grid size of $20 \times 20$, 16 neighbourhoods and 16 schools for visualisation purposes. There are two types of agents, reds and blues, black indicates an empty cell. The yellow circles indicate the school positions and the white lines indicate the neighbourhood boundaries. For the orange agent at (14,14), the green (dashed) neighbourhood represents its current administrative unit (bounded neighbourhood), while its local neighbourhood ($r=1$) is portrayed by the yellow (dotted) square.

(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Utility

Netherlands (Boterman, 2013) reflects people actually preferring some level of diversity over homogeneity. Note that utilities are on the unit interval.

The preference function portrayed in Fig. 2 provides more flexibility than the ones used in traditional Schelling models (Schelling, 1971), which are step functions combined with a satisfaction threshold. Agents are unsatisfied below the optimal fraction ($t_i$), while satisfied and indifferent for everything above. However, this behaviour can be approximated (red-dotted line) by setting $M=1$, resulting in a utility of 1 for $t_i \leq x_i \leq 1$. The green line represents an agent that ideally wants a 50–50 mix and equally penalises being the only agent of its group in its neighbourhood/school or having none of the other group ($M=0$). The blue and red line each have a larger value of $M$, although households with the blue preference function still receive more utility for their optimal fraction over homogeneity. This allows for the modelling of different preferences. For example, in countries where one might view at diversity differently than in Europe and/or ethnic preferences might differ from those for income/education in the same context.

2.2.2.3. Neighbourhood perception. Findings from a residential choice study in the Netherlands (Van Gent, Boterman, & van Grondelle, 2016) indicate that agents have variable perceptions for their neighbourhood. Within a larger neighbourhood, households sometimes associate themselves with a smaller portion of the larger spatial unit. In order to incorporate this in the model, the choice of residential location in the model is based on the bounded neighbourhood (e.g., census tract, administrative unit) and on more specific location characteristics within the local neighbourhood (based on their immediate neighbours). Following this, the parameter $b \in [0,1]$ is used to weigh the proportion of similar agents of the bounded neighbourhood ($x_{\text{bounded}}$) versus that of the local neighbourhood ($x_{\text{local}}$). Where the bounded neighbourhoods are the fixed, white squares as in Fig. 1a and the local neighbourhood is the Moore neighbourhood with radius $r$ centred around the agent. Moreover, the size of the former is controlled by the number of neighbourhoods ($m$) and size of the grid ($N$), the latter by the radius, $r$. See Fig. 1a for a visual representation of both neighbourhoods for a specific agent. Hence, the households consider when selecting a residential site is a weighting of its local neighbourhood and its larger, bounded neighbourhood. This is an important—more realistic—addition over both the Zhang (2004); Stoica and Flache (2014) models, as they look at local neighbourhoods only.

Combining the preference function with the neighbourhood perception, one can model considerable diversity of preferences by varying the optimal fraction ($t_i$), penalty for homogeneity ($M$), the radius ($r$) and neighbourhood perception ($b$) or estimating them from data. Estimates from interviews (Boterman, 2013), questionnaires and/or discrete models (Oosterbeek et al., 2021) could provide input here. Hence, the model is theoretically able to incorporate—some—of the heterogeneity (i.e., diversity between households) that is often found in empirical studies of residential and school choice (Hastings et al., 2009; Musterd et al., 2016; Oosterbeek et al., 2021).

$$x_i = b x_{\text{bounded}} + (1 - b) x_{\text{local}}$$

2.2.2.4. Moving. Households move with the objective of maximising their individual utility. In reality, agents might move greedily to improve their state, but with some noise because of insufficient information or mistakes. To model this, a multinomial choice process is implemented as in Sage and Flache (2021). That is, agents consider $K$ random empty residential locations or $m$ schools, including their current one and calculate the probability of moving to a new location. Large increases in utility are more likely to be chosen, but locations that give an agent a decrease in utility also have a non-zero probability. Note that agents move without incurring any costs.

$$\Delta U_i(\text{location } r, \text{current location } q) = U_{\text{new}}(x_i, t_i, M) - U_{\text{old}}(x_i, t_i, M)$$

$$P(\text{location } r | \text{current location } q) = \frac{e^{\beta \Delta U_i(r,q)}}{\sum_{r=1}^{K} e^{\beta \Delta U_i(r,q)}}$$

Varying the temperature ($\beta$) allows control over the randomness in the decision process, where agents become perfect utility maximisers as $\beta$ tends to infinity and the choice is fully random if $\beta=0$. For the school process, all schools are considered at every step. Hence, $U_{\text{school}}$ can be substituted for $U_{\text{loc}}$, $K$ for $m$ and distance is included in the utility calculation.

2.2.2.5. School capacity. Lastly, the number of students that can attend a school is controlled by the capacity parameter $c$.

Fig. 2. Single-peaked utility function. For the blue line, the agent obtains the maximum utility at $t_i$ and only $M$ if the neighbourhood or school is homogeneous with respect to their own group. The green (dashed) and red line (dotted) show alternative values for $t_i$ and $M$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
capacity $= 1 + \frac{\text{number of households}}{m}$

$m$ is the number of neighbourhood/schools and one free spot is added (i.e., plus 1) to ensure there is place for every household in the system if $c=1$. When $c=1$, a school can house a bounded neighbourhood in number of households (roughly 291). If $c$ increases, schools can house more students, which allows for larger school to emerge.

### 2.3. Process overview and scheduling

Although people are reported to move neighbourhoods for schools (Boterman, 2013; Butler & Hamnett, 2010) and hence evidence for a bidirectional relationship between residential and school choice exists, this model is simplified by simulating the residential process first. Note that this could be more realistic for educational systems where there are no catchment areas such as the Netherlands. Moreover, one can argue that once residential choices have been made, although possibly informed by the location of schools, the actual initial school choice is made after the residential patterns have emerged and can be considered sequential. Note that this reasoning excludes parents that move neighbourhoods after an initial school has been chosen, but the percentage of households leaving both neighbourhood and school is estimated to be less than 5% for the US (Swanson & Schneider, 1999).

Initially, all agents are assigned a random residential location and an optimal fraction, $T_i$. In each successive iteration, a random fraction ($f$) of the agents are chosen and they move in random order to maximise their utility. In the following step either all agents are considered again (replacement) or from the unmoved agents (without replacement), which is controlled by the scheduling parameter. This process is repeated until convergence or until the maximum amount of steps is reached. The processes are considered converged when the average utility, standard deviation of utility and segregation measure all have a Mean Absolute Deviation (MAD) lower than the convergence threshold parameter ($T_c$). Using the mean of the last window size ($T$) steps, all of the values need to be within the convergence threshold distance of the mean. Once the residential process has finished, the school choice process starts and households start choosing schools. The same $T_c$ and maximum steps hold for the school process, but school segregation is calculated using schools as administrative units, rather than the bounded neighbourhoods for the residential process. All parameter values remain fixed during each model run.

#### 2.3.1. Emergence

The households in the model have a certain optimal fraction of similar households in their search for a neighbourhood and school. From these choices, segregation in both processes is expected to emerge, in line with the modelling of the separate processes (Sage & Flache, 2021; Stoica & Flache, 2014; Zhang, 2011). However, as residential patterns have an influence via the preference for distance, the difference between the two levels of segregation and their sensitivity to the input parameters are harder to grasp, and of primary interest in this study.

#### 2.3.2. Adaptation and learning

Agents are allowed to adapt by moving to another school or place of residence. They only have memory going back one time step and are aware of all compositions of every empty location/school they consider. Hence, they are “myopic” (optimise given the current state of the system).

#### 2.3.3. Objectives

Agents have no other objectives than to maximise their utility by moving residential location and schools until the system reaches a stable state. However, due to stochasticity (2.3.6), the convergence threshold and/or maximum steps of both processes, it might be that some of the agents could have done better. Note that agents do not incur a cost of moving and can move indefinitely if the convergence threshold allows.

#### 2.3.4. Prediction and sensing

At the start of a time step $t$, all households have full information on all bounded/local neighbourhoods, empty locations and schools in the environment. However, it is only after all agents that were chosen to move in time step $t$ actually have made their decisions that the new compositions are updated. This means the last agent to move still uses the compositions that were calculated at the end of step $t - 1$, while potentially the environment may have substantially changed due to decisions of agents that were moved before. One exception is the empty locations. To facilitate moving more households than the density allows, the empty locations are updated continuously. Otherwise, one can only move a fraction of $1 - d$ agents, where the last agent has no choice anymore. Note that $r$ and $m$ control the radius of the local and bounded neighbourhoods respectively. For schools this is controlled by the capacity parameter ($c$).

#### 2.3.5. Interaction

There are no direct interactions defined between agents, yet the agents affect each other indirectly, by changing the neighbourhood/school compositions and distances through their choices.

#### 2.3.6. Stochasticity

Various parts of the model are stochastic in nature. Firstly, the initial residential locations and schools attended are randomly generated. Moreover, the considered empty locations are randomly chosen for each household and the population shares are not exactly 50–50 but drawn from a Binomial distribution with $n$ equal to the population size and $p=0.5$, this introduces some extra randomness in the initial conditions besides location. Lastly, the temperature parameter ($\beta$) controls the amount of randomness allowed in the choice process.

#### 2.3.7. Collectives

Households belong to a bounded neighbourhood and multiple local neighbourhoods of other agents. Also, in the school process, households are always enrolled in a school. Hence, the agents form collectives within these neighbourhoods and schools.

#### 2.3.8. Observation

At every time step and for each group type, the number of agents in the bounded neighbourhoods and schools are calculated, as well as their utilities, local neighbourhoods and distances to the schools they are attending. For each simulation run, all parameter values are saved, as well as the number of steps it takes to reach convergence, which will be equal to the maximum steps if any of the runs do not converge. From these values the residential and school segregation measures are calculated at each time step.

#### 2.3.9. Initialisation

Both models start from randomised initial conditions. In the residential choice model, the agents are randomly placed on the grid, excluding cells where schools are placed. In the school model the agents are assigned to a random school at first. The reasons for this is because the purpose is to obtain the stochastic stable state that the models will achieve in the long run, without biasing their choice with an initial configuration already affected by other model settings, for example residential segregation patterns.

#### 2.5. Input data

No empirical data is used explicitly in the model, however, in order to constrain the parameter space, computational constraints and common sense are used to limit the range of particular parameters (see Table 1). The model comprises 15 parameters that can all be varied,
Table 1 Parameters, their description, ranges in the sensitivity analysis and nominal values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
<th>Substantiation</th>
<th>Nominal value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size (N)</td>
<td></td>
<td></td>
<td>90</td>
</tr>
<tr>
<td>Composition</td>
<td>50–50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density (d)</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of schools/neighbourhoods (m)</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Households</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal fraction (τ)</td>
<td>[0.40, 0.80]</td>
<td>&lt;0.4 convergence issues</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;0.8 full segregation</td>
<td></td>
</tr>
<tr>
<td>Composition/distance trade-off (α)</td>
<td>[0, 1]</td>
<td>Full range</td>
<td>0.2</td>
</tr>
<tr>
<td>Utility at homogeneity (M)</td>
<td>[0, 1]</td>
<td>Full range</td>
<td>0.6</td>
</tr>
<tr>
<td>Neighbourhood perception (b)</td>
<td>[0, 1]</td>
<td>Full range</td>
<td>0.2</td>
</tr>
<tr>
<td>Radius (r)</td>
<td>(1, 2, ..., 10)</td>
<td>Smaller than bounded</td>
<td>3</td>
</tr>
<tr>
<td>Amount of randomness in moving decisions (β)</td>
<td>[1, 10, 50, 100]</td>
<td>Computational</td>
<td>100</td>
</tr>
<tr>
<td>Considered residential locations (K)</td>
<td>(1, 2, ..., 10)</td>
<td>Computational</td>
<td>5</td>
</tr>
<tr>
<td>School capacity (c)</td>
<td>[1, 5]</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Simulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum number of steps in both processes</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convergence threshold (τ_s)</td>
<td>(0.001, 0.005, 0.01, 0.02)</td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>Window size (l)</td>
<td>(10, 20, ..., 50)</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Fraction of agents moved (f)</td>
<td>(0.25, 0.5, 1)</td>
<td>Computational</td>
<td>0.25</td>
</tr>
<tr>
<td>Temperature (T)</td>
<td>(1, 10, 50, 100)</td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

However, segregation measures are sensitive to absolute and relative composition changes, as well as the spatial definition of the neighbourhoods and schools (Mazza & Punzo, 2015). For example, it is possible to observe the same spatial ordering of agents, yet a different level of segregation when the number of neighbourhoods or schools change. As the study examines the effect of behavioural parameters on the model output (i.e., segregation), rather than changes in levels of segregation due to different spatial units, these parameters are fixed. However, it is important to note that in more realistic scenarios both could affect the level of segregation.

The main metric used to measure both residential and school segregation used in this study is Theil’s index (Theil & Finizza, 1971), because of its better mathematical properties than for example the commonly used Dissimilarity index (Duncan & Duncan, 1955). However, this does not mean Theil’s measure is without flaws. Although finite-size effects for Theil’s measure of segregation have not been thoroughly studied as far as the authors are aware, for the commonly used Dissimilarity index (D) Carrington and Troske (1997) show that substantial differences in D occur when the population size is low or group proportions are highly skewed. To avoid substantial finite-size effects these two parameters are fixed in the sampling scheme at a value of N=90 (i.e., height and width of the grid) and a group composition of 50–50 respectively. Moreover, the density (d) also controls the number of households and is fixed at d=0.9. Additionally, the number of neighbourhoods and schools are explicitly used in the segregation calculations. However, changing their number (m) can also change the level of segregation while the actual spatial ordering is the same in the grid. Therefore this parameter is fixed as well. As an indication the average primary school size in Amsterdam is used, which lies between 290 and 300 pupils per school (Gemeente Amsterdam, 2021). 25 schools—and hence neighbourhoods—, together with a density of 0.9 lead to an average school size of 291 pupils. Lastly, the one-to-one-mapping of neighbourhoods and schools is important, because varying these separately may introduce measurement artefacts. Note, that to maintain the one-to-one mapping in the segregation calculations, residential segregation is calculated using the bounded neighbourhoods only. This implies that the way residential segregation is measured is different than the neighbourhood perception of households (if b<1). However, such administrative units or census tracts are often used in reality to measure segregation levels. Moreover, the way in which local versus bounded neighbourhoods influence the decision processes of households is not fully understood. How these changes impact the choices of agents and the measures of segregation separately is not trivial and left for future studies.

3. Results

The interest of this study lies in how household behaviour affects the levels of residential and school segregation. Although these outcomes are completely determined by the input parameters, due to randomness and interactions, it is often hard to grasp exactly how these mechanisms result in the observed output. Moreover, as the behaviour of the agents depends on the specific parameter values, varying one parameter while fixing the others at a nominal value may cause different dynamics than changing them simultaneously. Therefore, a Global Sensitivity Analysis (GSA) is conducted where multiple parameters are varied simultaneously to evaluate the effect that the individual parameters and their possible interactions have on the output of interest. See section 3.2 for a more detailed explanation. However, this is computationally costly and the system needs to have reached a stable state in both processes. Otherwise, inference regarding the behaviour may change substantially if the simulation is allowed to run for a longer amount of time. Therefore, we first conduct a convergence analysis of parameters that are assumed to mostly affect the rate of convergence instead of the level of segregation. GSA methods often do not provide enough details on the workings of the mechanisms between input and output. Additional analyses, where only the most influential parameters are varied systematically (i.e., local sensitivity), are carried out to understand how these parameters influence the levels of segregation.

3.1. Analysis of convergence

The number of considered residential locations K ∈ (1,5,10), the fraction of agents moved f ∈ (0.25,0.5,1), the scheduling method (with or without replacement) and temperature β ∈ (1,10,50,100) are assumed to affect the rate of convergence. For example, if all households consider 10 locations every time step, more of the phase space can be explored in parallel which may lead to faster convergence, yet at a slower pace. Hence, these parameters are varied to assess which window and threshold are deemed appropriate.

Parameters that are varied later in the GSA probably affect convergence as well. If all households require homogeneous neighbourhoods/schools to obtain the full utility, a stable state might be reached faster than if they have an optimal fraction of 0.6. Therefore, in addition to varying the convergence parameters, a coarse Sobol sequence with sample size 256 is employed to assess convergence within the full parameter space. That is, k, M, b, r, a, c are varied simultaneously and the resulting 1024 model runs are then distributed evenly over all combinations of the convergence parameters (72 in total). 500 steps is deemed sufficient for processes to have reached a stable state. This is visually confirmed by looking at Fig. A.1, where all 25 randomly drawn model runs are shown to have reached an equilibrium. However, it also shows that all processes could have been terminated before step 500. For every time step in both processes, an appropriate termination criterion is tested with varying window sizes ∈(10,20,....,50) and thresholds ∈(0.001,0.005,0.01,0.02). The level of segregation at convergence is then compared with the level of segregation at step 500, to see if the two
values agree. A process is considered converged if segregation, the mean and standard deviation of all household utilities are all within a threshold distance \(T_d\) for the last window size time steps. Hence, only stable attractors are analysed, not cycles, where the system could possibly switch between states, although in a real world social system exogenous conditions could possibly disturb / limit cycles.

Inspecting Figs. A.2 A.3, A.4 and A.5, one can see that considering 5/10 empty locations and a temperature of 100 decrease the difference in residential segregation. The fraction of agents moved is fixed at 0.25, which decreases accuracy, but drastically decreases computational time. Therefore, \(T=100, K=5\) (less calculations per time step than 10) and \(f=0.25\). The analysis for the school process is similar, except for the with replacement option performing better A.8, hence is chosen. Figs. A.6, A.7 show the differences in residential segregation for these fixed values. Based on this, a 30–0.01 are all within 0.03 of the end level of segregation and runs are still terminated quite quickly. For the school process (Fig. A.9) all combinations with a window size larger than 10 lead to the difference being smaller than 0.015.

### 3.2. Global sensitivity analysis

To quantify the influence of the parameters, various global sensitivity analysis methods have been proposed in the literature (Pianosi et al., 2016). One of the common approaches is Sobol’s method with the Saltelli sampling scheme, which is based on the variance decomposition of the output (Pianosi & Wagener, 2015). The main benefit of the Sobol method is that it estimates the direct effect of parameters (i.e., first-order effects), and the higher-order and total-order effects. However, inspecting the output distributions of residential and school segregation (Fig. 3) it shows that large parts of the parameter space result in zero or fully segregated states and hence bi-modal output distributions. This could jeopardise the assumption that the variance is a meaningful metric for the output distribution. Therefore, two moment-independent global sensitivity approaches, PAWN (Pianosi & Wagener, 2015) and Delta (Borgonovo, 2007), are used to verify the ranking and mapping of the Sobol method.

Additionally, most global sensitivity methods assume that parameters are independent, which is guaranteed in this scenario as it is a fully theoretical model with no dependencies built in. Theoretically, parameters with no influence should have zero first- and total-order indices, yet as one is dealing with estimations, this might not be the case in practice. Therefore, a dummy variable with no relationship to the outcome is included in the simulations, parameters close to the estimated dummy indices can then be assumed to be of little influence. Note that \(\alpha\) also acts as a dummy in the residential process, as it is only affects school choices and hence school segregation.

The sample size for the Saltelli sampling scheme is set to 2048, which results in 32,768 model runs when second-order effects are included. Means and confidence intervals are estimated on all of the model runs and 100 bootstrapped samples. The same model runs—generated by the Saltelli sample—are used to estimate the PAWN/Delta indices for computational efficiency. The seven remaining parameters are all varied along their full range, except for the optimal fraction, which is drawn from \(\epsilon\{0.4, 0.8\}\). Optimal fractions below 0.4 face convergence issues, whereas above 0.8 one observes full segregation despite other parameter values, as well as convergence problems.

From 32,768 model runs, all residential processes converged according to the convergence criteria, while for the school process only 13 did not (0.04%). The output distributions (Fig. 3) show that in most cases full segregation is reached for both processes and also a substantial, but smaller peak at zero segregation. Often, the levels of school- and residential segregation are quite close, or the former is slightly higher, but there are parts of the parameter space where school segregation exceeds the level of residential by more than 0.2 (absolute difference). The global sensitivity analysis is used to try and discover what parameters are most influential in affecting these different levels of segregation.

Fig. 4a indicates that the optimal fraction \(t_0\), homogeneity penalty \((M)\), neighbourhood perception \((b)\) and radius \((r)\) are the most important parameters according to all four sensitivity indices for residential segregation. The dummy, \(a\) and \(c\) are considered non-influential by the first-, total-order and PAWN index, as their confidence intervals include zero. Although the PAWN index does not agree with the entire ranking, there is a clear distinction between the top four and the rest (also verified by the Delta measure). Similarly, for school segregation the ranking of all metrics is the same for the top four. Note that in both processes the total-order index for the most influential parameters is considerably larger than the direct effect (i.e., Sobol first-order), which indicates that interactions between parameters are also affecting the variance of the output. The second-order Sobol indices (Fig. 4c, d) show that the interaction between the homogeneity utility \((M)\) and optimal fraction \(t_0\) is important (significantly different from zero) for both residential- and school segregation. For the remaining parameters, either higher-order interactions are more important or the sample size needs to be increased for more accurate estimates.

As a robustness check, both the first- and total-order indices are calculated for different sample sizes\(^2\) (Figs. A.10, A.11). Additionally, the sensitivity indices and average effects have been recalculated using another segregation metric called the Isolation index (Massey & Denton, 1988), these checks do not alter the qualitative results.

As all indices agree on the distinction between the top four parameters and their sensitivity to residential/school segregation, the focus will lie on the union of both subsets. This means the school capacity is treated as non-influential and excluded. For the subsequent analysis a systematic parameter sweep is conducted, however, whenever a parameter is not varied, it needs to be fixed at a reasonable value. The average effects observed in the GSA (Fig. A.12) represent the average level of segregation observed when “binning” a specific parameter. Binning the values on the x-axis (50 bins), while the y-axis represents the average of the observed levels of segregation when all other parameters are still allowed to vary (bands show the 2.5% and 97.5% percentile). Using these average effects for Theil’s measure of residential and school segregation, the following fixed/nominal values are chosen \(t_0=0.5, M=0.6, r=3, b=0.2, a=0.2\) and \(c=2\). Reasoning behind this is that for the optimal fraction \((t_0>0.5)\), utility at homogeneity \((M>0.6)\), radius \((r>3)\), neighbourhood perception \((b>0.2)\) and \(a=0.2\), neighbourhoods and/or schools are already substantially segregated \((>0.7)\). By choosing the parameters at or below these boundaries the expectation is to see more interesting dynamics at lower, more realistic levels of segregation. Moreover, if segregation emerges even for \(t_0=0.5\) it is only expected to worsen for larger values.

### 3.3. Mechanisms of segregation

Although most parameter configurations seem to lead to either fully or zero segregated states, by systematically varying the subset of parameters, one can try and understand the mechanisms of segregation and what values lead to specific levels. As the total-order indices are substantially larger than the first-order effects, there might be higher-order interactions important in explaining the variance of the output distributions. Therefore, not only the direct effects are studied, but also second-order interactions. Nominal values and parameter ranges used are presented in Table 1.

The direct effects of the parameters (Fig. 5) show that as the optimal fraction \((t_0)\) of similar households increases, both levels of segregation increase as well. This is expected, as households explicitly require a higher fraction of similar agents for more utility. However, the increase is steeper and starts earlier for school segregation. Because segregation is bounded by 1 from above, the difference between school and

\(2\) This is possible with Sobol as Saltelli sampling generates a sequence.
residential has a maximum around 0.55–0.60. Given an optimal fraction lower than 0.65–0.70, the level of school segregation seems to be more affected than the level of residential segregation. Simultaneously varying $t_i$ with $M$ reveals similar patterns. Increasing the value of $t_i$ shows an increase in residential and school segregation, which is steeper for small values of $t_i$ and flattens when approaching 1 (Fig. 6). Increasing

Fig. 3. Residential, school segregation and their difference (school - residential). One occurrence represents the level of segregation as measured in the last step (i.e., converged or maximum steps is reached) of the model run using Theil’s measure of segregation.

Fig. 4. Sobol’s first- (blue) and total-order (orange) indices for the 7 parameters that are varied, complemented by the moment-independent PAWN (green) and Delta (red) measures of sensitivity. Confidence intervals are based on bootstrapping 100 times and the parameters are sorted on their value for the first-order Sobol. Figure (a) shows the sensitivity for residential segregation, (b) for school and (c) and (d) show their second-order Sobol indices (minus or plus the value in brackets delivers the 2.5% and 97.5% percentile respectively). All of the outputs are based on Theil’s measure of segregation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
M (i.e., reducing the penalty of homogeneity) results in an earlier increase in segregation if the optimal fraction \( t_i \) is increased simultaneously. Again, school segregation seems to be more affected than residential, resulting in a maximum difference between the two for values between 0.5 and 0.7 for varying values of \( M \). Hence, relatively tolerant households (i.e., their optimal fraction is to be the slight majority) maximise the gap between the two levels of segregation.

The \( \alpha \) parameter controls the relative weight of composition over distance in school choices. Hence, \( \alpha \) should not influence residential segregation and, when only distance matters (\( \alpha = 0 \)), school segregation should be equal to residential. The general pattern is that as households place more emphasis on school composition over distance (i.e., \( \alpha \) increases), schools start to become more homogeneous compared to neighbourhoods and the difference between the two levels of segregation increases (right plot in Fig. 5). Also when the other parameters are varied, this pattern occurs (Figs. A.16, A.18, A.19). Note that there is a reduction in school segregation at \( \alpha = 1 \), which might be due to asymmetric scales of the composition and distance component.

Although, both are bounded between zero and one, most distance utilities are 0.7 or higher, while the composition utilities range from 0.5 to 1. This also makes the interpretation of \( \alpha \) dependent on \( t_i \) and \( M \) as these control the composition utility.

When homogeneous neighbourhoods/schools (\( M \)) are severely penalised (i.e., equals 0), households are pushed more to find their optimal fraction (nominal value equals 0.5) — for which they receive the highest utility of 1 — and hence segregation is low. As this parameter starts to increase, the penalisation becomes less severe and segregation is allowed to emerge (Fig. 5), however, it seems that residential and school segregation are affected differently as the gap is largest when \( M = 0.6 \). This effect is less pronounced, but still present if \( b \) or \( r \) are increased together with \( M \) (Fig. A.20, A.21). This is probably due to both \( r \) and \( b \) being only explicitly present in the residential choice process and because residential segregation seems most affected. Increasing \( r \) means local neighbourhoods approach bounded neighbourhoods in size,
making it harder to obtain your optimal fraction as others have the same objective and the grid is finite, increasing residential segregation. Larger values of \( b \) (i.e., bounded neighbourhoods have a larger weight), forces agents to form clusters aligned with the bounded neighbourhoods, which are used in the segregation calculations, hence worsen residential patterns. Increasing both simultaneously (Fig. A.22) increases residential segregation rapidly as it is harder to find your optimal fraction in two relatively large and different neighbourhoods (i.e., local and bounded). As residential segregation is already substantial, the differences between school and residential are low. Hence, if either households have a large local neighbourhood or place a lot of emphasis on their bounded one, substantial segregation is allowed to emerge very quick. Note that this is also due to the assumption that the bounded neighbourhoods being directly associated with the school catchment area, which in reality does not have to be the case.

In the Appendix (Figs. A.23, A.24) the OFAT analysis is repeated for a 70–30 and 90–10 composition. The main results while varying the optimal fraction and utility at homogeneity still stand for 70–30 (i.e., that school segregation rises faster than residential), only the increase starts at a larger value. For 90–10 the effect disappears for the optimal fraction. All the other effects disappear, is probably due to the minority having a harder time finding a high utility spot for values of the optimal fraction that are larger than 0.5. These are fractions are difficult as they are in the minority.

3.4. The role of residential patterns in school segregation

One benefit of simulations is that one can explore—hypothetical—scenarios that might be hard or even impossible to encounter in the real world. In empirical school segregation research it is often the question what role residential patterns play in the level of school segregation (Boterman, Musterd, Pacchi, & Ranci, 2019). However, in cities with existing residential segregation one cannot observe the level of school segregation in a scenario with no residential patterns. To circumvent this, simulations are conducted where the school process runs without the residential process, such that one can observe the level of school segregation without residential patterns. Fig. 8 shows that schools often segregate as much with, as well as without residential patterns. However, residential patterns seem to have a clear, but minor effect when \( \alpha, M, b \) are relatively small and for the entire range of \( r \). Fig. 7 shows that at least part of the extra school segregation is due to systemic effects, as the majority group within a school tends to travel further on average than the minority group for a more favourable composition (i.e., to obtain a higher utility). This effect decreases as \( \alpha, a \) and \( M \) approach 1, which all place more emphasis on composition. In these cases both travel almost equal distances on average and schools are approaching full segregation. It is worthwhile to note that this discrepancy between the minority and majority agents happens without the minority agents specifically having a smaller weighting for the distance compared to the composition. In each simulation everyone has the same \( \alpha \) and one group is forced to travel further due to systemic effects. Note that \( \alpha=1 \) is excluded, in this case distance does not matter and agents travel very far, which would otherwise distort the plot.

4. Conclusion

The purpose of this paper is to study alternative explanations as to why one might observe school segregation exceeding residential segregation, besides assigning households to be less tolerant for school compositions compared to neighbourhood compositions (Cordini et al., 2019; Logan et al., 2008; Renzulli & Evans, 2005). It extends the model of Stoica and Flache (2014) by basing residential choices on a mixture of a local and bounded neighbourhood, removing the satisfaction thresholds, and by employing a convergence- and global sensitivity analysis on the most important parameters. Moreover, it examines the role of residential patterns in school segregation.

The results show that in a large part of the parameter space one can observe increased school segregation relative to residential segregation, while retaining the assumption that preferences for composition are equal for both residential and school compositions. Especially when households prefer to be the slight majority (\( t_i \in [0.5,0.7) \), homogeneity is penalised sufficiently (\( M \in [0.6,0.8] \)), school composition is more important than distance preferences (\( \alpha>0.5 \)) or when the local neighbourhood is emphasised over the bounded (\( b<0.2 \)). Analyses suggest
that there are two mechanisms behind this difference.

Firstly, part of the increase is due to residential segregation (Fig. 8) and the school process being run after the residential. The residential segregation model tends to generate residential distributions that are more segregated than would be needed to satisfy composition preferences of agents, due to the self-reinforcing dynamics. This is consistent with Schelling models and their extensions (Clark & Fossett, 2008; Schelling, 1971; Zhang, 2011). Given distance preferences, households would also—all other things being equal—prefer to attend the nearest school. However, suppose they would choose the nearest school, then due to the residential segregation this school is unlikely to have a local neighbourhood exactly matching the preferred composition of parents. Thus, at least one group is underrepresented, starts to move out (i.e., they are willing to travel further for a more favourable composition as shown in Fig. 7) and hence triggers the whole cascade again, but now for school choices. Assuming this starts from a level of school segregation already roughly reflecting the current residential segregation (i.e., nearest school choice), the cascade can now further increase school segregation above this level. This suggests that excess school segregation cannot only be observed when households are less tolerant for school segregation relative to residential is also due to the different scales. This indicates that the increase in school segregation through residential segregation. In some circumstances, when the effect of residential on school segregation is marginal, the scope for intervention (to say, reduce school segregation) may be limited. Despite the fact that school segregation still being higher than residential in those cases, our model suggests that this is driven by school choice and not residential.

Using a theoretical ABM to model school segregation provides several advantages over traditional methodologies employing empirical data. ABMs allow the modelling of commonly found characteristics such as feedback loops, non-linearity and spatial effects all found in school choice (Dignum et al., 2021), with the observed levels of segregation emerging from the individual behaviour of the households (Heppenstall, Crooks, See, & Batty, 2011). Additionally, simulations can potentially cover and study the entire parameter space as well as explore hypothetical scenarios. The model also highlights that contexts matters; it is important to grasp how parents value distance and composition, perceive their neighbourhood/school compositions and come to a decision, to understand the mechanisms behind the collective outcome of segregation in realistic scenarios. Moreover, segregation itself and counteracting policies, are processes that possibly span years. Hence, simulations could be helpful in determining causes and effects over long periods of time more quickly.

4.1. Limitations and further research

However, there are also limitations and interesting directions for further research. Firstly, no explicit empirical data is used to calibrate the model, and partly because of that, simplifying assumptions have been made to end up with a tractable model. The fact that residential choices are based on composition only and school choices on distance and composition is unrealistic (Boterman, Musterd, Pacchi, & Ranci, 2019) and there is substantial evidence for heterogeneity of households patterns on school segregation. If households perceive their neighbourhood and schools at different scales in reality, this increase of school segregation relative to residential might be due to these varying perceptions rather than a decrease in tolerance. This finding also has important policy implications for those regulating and measuring school segregation through residential segregation. In some circumstances, when the effect of residential on school segregation is marginal, the scope for intervention (to say, reduce school segregation) may be limited. Despite the fact that school segregation still being higher than residential in those cases, our model suggests that this is driven by school choice and not residential.

Using a theoretical ABM to model school segregation provides several advantages over traditional methodologies employing empirical data. ABMs allow the modelling of commonly found characteristics such as feedback loops, non-linearity and spatial effects all found in school choice (Dignum et al., 2021), with the observed levels of segregation emerging from the individual behaviour of the households (Heppenstall, Crooks, See, & Batty, 2011). Additionally, simulations can potentially cover and study the entire parameter space as well as explore hypothetical scenarios. The model also highlights that contexts matters; it is important to grasp how parents value distance and composition, perceive their neighbourhood/school compositions and come to a decision, to understand the mechanisms behind the collective outcome of segregation in realistic scenarios. Moreover, segregation itself and counteracting policies, are processes that possibly span years. Hence, simulations could be helpful in determining causes and effects over long periods of time more quickly.

4.1. Limitations and further research

However, there are also limitations and interesting directions for further research. Firstly, no explicit empirical data is used to calibrate the model, and partly because of that, simplifying assumptions have been made to end up with a tractable model. The fact that residential choices are based on composition only and school choices on distance and composition is unrealistic (Boterman, Musterd, Pacchi, & Ranci, 2019) and there is substantial evidence for heterogeneity of households
considering neighbourhoods and schools (Hastings et al., 2005; Chu-macero et al., 2011; Oosterbeek et al., 2021). The coupling of empirical data can reduce the need to make these unrealistic assumptions and for systematic sweeps of the entire parameter space.

Furthermore, residential and school choice are found to be intertwined and not sequential processes (Boterman, 2019; Burgess et al., 2015). Although moving schools might be less costly than moving neighbourhoods, school choices could feedback into residential choices if parents move for school characteristics (Butler & Hamnett, 2010; Holme, 2002), for example due to geographic assignment mechanisms (Boterman, Musterd, Pacchi, & Ranci, 2019). This potentially increases residential segregation and decreases the gap between the two. Moreover, not only can preferences differ between people, they can also change over time. Positive contacts with out-group people are likely to decrease prejudice and social distance (Semyonov & Glikman, 2009), which in our case would be resembled by a decrease in the optimal fraction and reduce segregation. Additionally, in reality there exist households without children that might be part of their neighbours’ neighbourhood perception, but not of the school compositions. Also, the model assumes no cost of moving, especially if the groups resemble income differences, this might be an interesting dynamic to take into account. High-income parents might have more means to move neighbourhoods (increasing residential segregation), yet their counterparts might take greater agency in choosing schools (without geographic constraints) as they cannot easily move.

Moreover, when loosening the one-to-one mapping of schools and neighbourhoods (i.e., unequal numbers or placement) and/or varying population structures (i.e., multiple groups, different ratios), one approaches a more realistic scenario, but measuring and comparing residential and school segregation potentially becomes more complicated as well. The experiment on the role of residential patterns in schools already shows the effect of the former on the latter is not trivial. Moreover, being a small minority in the current model will make it very hard to find optimal fractions larger than 50%. Hence, other population structures could also look at relative optimal fractions. Lastly, the time scale at which the two processes converge is arbitrary and does not necessarily reflect reality.

Hence, our future research intends to consider these limitations and develop suggestions to improve the modelling of the residential- and school choice processes. Lastly, the model could also be used to study the effect of exogenous shocks to the system, such as waves of migration, school choice policies and/or institutional rules. Will a segregated/integrated state be stable against a wave of more/less tolerant residents coming in for example?

Acknowledgments

This paper is part of the Computational Modeling of Primary School Segregation (COMPASS) project which is funded by the Dutch Inspectorate of Education and the City of Amsterdam. The fourth author acknowledges financial support by the Netherlands Organization for Scientific Research (NWO) under the 2018 ORA grant ToRealSim (464.18.112) and the research program Sustainable Cooperation – Roadmaps to Resilient Societies (SCOOP) funded by NWO and the Dutch Ministry of Education, Culture and Science (OCW) in its 2017 Gravitation Program (grant number 024.003.025).

Appendix A. Appendix
Fig. A.1. Mean utility, standard deviation utility of households and segregation (Theil) per process. Every line represents one random model run out of 25 in the residential or school process from the convergence analysis.

Fig. A.2. Difference in residential segregation compared to step 500 for varying window sizes and thresholds.
Fig. A.3. Difference in residential segregation compared to step 500 for varying window sizes and thresholds.

Fig. A.4. Difference in residential segregation compared to step 500 for varying window sizes and thresholds.

Fig. A.5. Difference in residential segregation compared to step 500 for varying window sizes and thresholds.

Fig. A.6. Difference in residential segregation compared to step 500 for varying window sizes, thresholds and fixed values of $\beta=100$, $K=5$, $f=0.25$ and scheduling with replacement.

Fig. A.7. Steps to convergence for varying window sizes, thresholds and fixed values of $\beta=100$, $K=5$, $f=0.25$ and scheduling with replacement.
Fig. A.8. Difference in residential segregation compared to step 500 for varying window sizes and thresholds.

Fig. A.9. Difference in school segregation compared to step 500 for varying window sizes, thresholds and fixed values of $\beta=100, K=5, f=0.25$ and scheduling with replacement.

Fig. A.10. Sobol first- (blue) and total-order (orange) indices for residential segregation and varying sample sizes. Bands indicate the 2.5% and 97.5% percentiles of 100 bootstrapped samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. A.11. Sobol first- (blue) and total-order (orange) indices for school segregation and varying sample sizes. Bands indicate the 2.5% and 97.5% percentiles of 100 bootstrapped samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. A.12. Average effects of varying parameters using the Saltelli sampling scheme. Residential (blue), school segregation (orange) and their difference (school - residential, green) are plotted. Observations are binned using 50 bins, where each of the bins represents the average level of—or difference in—segregation observed. The parameter of interest is “fixed” at the bin value, while the others were allowed to vary. The bands represent the 5th and 95th percentiles of the observations within a bin. The difference of school and residential segregation is based on Theil’s index. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. A.13. Sobol’s first- (blue) and total-order (orange) indices for the 7 parameters that are varied, complemented by the moment-independent PAWN (green) and Delta (red) measures of sensitivity. Confidence intervals are based on bootstrapping 100 times and the parameters are sorted on their value for the first-order Sobol. Figure (a) shows the sensitivity for residential segregation, (b) for school and (c) and (d) show their second-order Sobol indices (minus or plus the value in brackets delivers the 2.5% and 97.5% percentile respectively). All of the outputs are based on the Isolation index. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. A.14. Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.
Fig. A.15. Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

Fig. A.16. Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

Fig. A.17. Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.
Fig. A.18. Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

Fig. A.19. Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

Fig. A.20. Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.
Direct and interaction effects of Utility at homogeneity (M) and Radius (r)

![Figure A.21](image_url)  

Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

Direct and interaction effects of Radius (r) and Neighbourhood mix (b)

![Figure A.22](image_url)  

Varying two parameters at a time. Residential (left), school segregation (middle) and their difference (school - residential, right) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

Varying one factor at a time

![Figure A.23](image_url)  

Varying one parameter at a time for a 70–30 composition. Residential (blue), school segregation (orange, dashed) and their difference (school - residential, green, dotted) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Fig. A.24. Varying one parameter at a time for a 90–10 composition. Residential (blue), school segregation (orange), and their difference (school - residential, green, dotted) are plotted. The plotted level of segregation is the average over 30 model runs, the bands represent the 2.5% and 97.5% percentiles. Nominal values of the parameters that are not varied can be found in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References


