# A data-driven agent-based model of primary school segregation in Amsterdam

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

**DOI**
10.1080/0022250X.2024.2340136
**Publication date**
2024
**Document Version**
Final published version
**Published in**
Journal of Mathematical Sociology

THE JOURNAL OF MATHEMATICAL SOCIOLOGY
2024, VOL. 48, NO. 3, 362–392
https://doi.org/10.1080/0022250X.2024.2340136

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

# A data-driven agent-based model of primary school segregation in Amsterdam

Eric Dignum [a], Willem Boterman [b], Andreas Flache [c], and Mike Lees [a]

[a]Computational Science Lab, University of Amsterdam; [b]Urban Geographies, University of Amsterdam; [c]Faculty of Behavioural and Social Sciences, University of Groningen

**ABSTRACT**

Theoretical agent-based models of residential and school choice have shown that substantial segregation can emerge as an (unintended) consequence of interactions between individual households and feedback mechanisms, despite households being relatively tolerant. However, for school choice, existing models have mostly been highly stylized, leaving open whether they are relevant for understanding school segregation in concrete empirical settings. To bridge this gap, this study develops an empirically calibrated agent-based model focusing on primary school choice in Amsterdam. Consistent with existing models, results show that substantial school segregation emerges when schools are chosen based on a trade-off between composition and distance, and also when households are relatively tolerant. Additionally, findings of (hypothetical) policy simulations suggest that it is important to understand which preferences for school composition and distance households have and how these interact. We find that the effects of policies aiming to reduce school segregation through geographical restricting mechanisms are highly dependent on those interacting preferences. Also, we assessed the contribution of residential segregation to school segregation. Our findings may have implications for methodologies aiming to estimate school choice preferences, such as discrete choice models, as these methodologies do not explicitly control for implications of these interactions and feedback mechanisms, which might lead to incorrect inference.

**ARTICLE HISTORY**
Received 2 November 2022
Revised 7 April 2023
Accepted 15 December 2023

**KEYWORDS**
Agent-based modeling; complex systems; complexity; empirical calibration; school choice; school segregation

## 1. Introduction

Numerous studies find substantial levels of school segregation in a wide range of educational systems, meaning that children with different characteristics, such as ethnicity, parental income/education or gender, do not attend the same schools. School segregation is generally considered an important societal issue as it is associated with unequal opportunities, outcomes, and eroding social cohesion (Boterman et al., 2019b; Wilson & Bridge, 2019). Despite policies aimed to counteract segregation, macro-level

**CONTACT** Eric Dignum e.p.n.dignum@uva.nl Computational Science Lab, University of Amsterdam, Science Park 900, Amsterdam 1098 XH, The Netherlands

analysis often finds stable, substantial, and sometimes even increasing levels of school segregation across most contexts (Thurner et al., 2018; Logan et al., 2008; Reardon & Owens, 2014). Primary school segregation in the Dutch city of Amsterdam is no exception: despite desegregation policies, segregation levels by parental income are stable, high, and even more severe than along ethnic lines (Onderwijsinspectie, 2022).

These ineffective policies could result from a lack of understanding of the mechanisms underlying school segregation. Commonly used methodologies in empirical school choice and segregation studies, such as discrete choice analysis, often treat households as isolated components that do not interact when choosing schools. Paradoxically, research finds that households use school compositions (e.g., socioeconomic and ethnic) to inform their school choices. Hence, households at least interact indirectly, observing school compositions (composed of other households) and current school choosers interact with previous ones constructing a feedback loop. It has already been shown that explicitly accounting for such interactions can lead to non-linear effects and tipping points (Boterman, 2021b, p. 0). These effects are hard to capture with techniques such as discrete choice analysis, which focus on individual decision-making in isolation and potentially miss these effects that result from interacting components (Maes, et al., Page, 2015). Moreover, such non-linear effects have also been found in empirically, in the form of White/Black flight, for example, where a certain group moves out of schools to avoid an unfavorable composition/quality (Cordini et al., 2019; Renzulli & Evans, 2005), where the initial movers trigger others to move out as well because of changing compositions, and so on, leading to a self-reinforcing process (feedback). Failing to capture such complexities in the analysis of individual school choices or changes in school compositions could lead to incorrect inferences about the underlying causes. Hence, several studies have argued that methods accounting for the interdependent actions of individual households can improve our understanding of school segregation (Boterman et al., 2022b; Perry et al., 2022).

Important methodological progress has been made in recent years. This is done, for example, by allowing for non-linear dynamics using Bayesian dynamical systems modeling (Spaiser et al., 2018) and Agent-Based Models (ABMs). ABMs are algorithmic models that consist of many, typically heterogeneous agents that interact (spatially, temporally, in social networks) and can respond to their time-varying environment all the while they pursue a certain objective (Bonabeau, 2002). They can systematically link preferences/constraints on the micro-level to their consequences for school segregation on the macro-level and can analyze policy implications (Bruch & Atwell, 2015). However, only a limited number of ABM studies have been conducted in the context of school choice research. Theoretical studies, with choice processes inspired by empirical work, have found that interactions play a crucial role in the observed levels of school segregation. These interactions can create feedback mechanisms that push even relatively tolerant households to attend segregated schools. Although this work has led to valuable insights, they have limited applicability to reality.

While empirically calibrated Agent-Based Models (ABM) have been utilized with success in modeling other social systems (Page, 2015), for school choice there have only been few. Mutgan (2021) uses discrete choice models to estimate parental preferences for eight different ethnic household groups of parents. These estimates are then incorporated into an ABM, concluding that Stockholm's ethnic school segregation is mostly due to ethnic residential segregation because most schools that are within reasonable distance are indistinguishable from each other from a composition perspective. However, this estimation procedure still assumes independent households and hence circumvents estimating preferences accounting for interactions, with possible biases. In another study, Ukanwa et al. (2022) use United States survey data, based on hypothetical school choice scenarios, to estimate school preferences of parents. They show that even if racial school compositions are made irrelevant, schools will still segregate because different racial groups have different school preferences for other school attributes besides composition. As these are stated preferences, feeding them into the ABM allows them to study how interactions between stated preferences unfold. However, the survey uses only 32 fictional schools with static characteristics, such as income/racial composition and average commuting time, that purposely differ substantially.

In this paper, using precise primary school locations and approximated residential locations of Amsterdam households, we study the effect of a wide range of distance/composition preferences and constraints. Allowing us to study how various scenarios and possible policies, including interactions, feedback loops and non-linear effects within the system, lead to empirically realistic patterns of school segregation by income. Hence, this is a different strategy than Mutgan (2021) as we assess which preferences/constraints and their interactions lead to empirically observed levels of school segregation, while still allowing for dynamic school attributes (e.g. composition) in the full system, contrary to the fictional, static and relatively small scenario of Ukanwa et al. (2022). There are several reasons for choosing Amsterdam as a case study. First, there is no strict geographic assignment mechanism (only a priority scheme) and households are free to apply to any primary school, loosening the link with potential residential processes, which are assumed to be fixed. Second, households have many schools available within a small radius, meaning residential segregation and distance preferences can theoretically even lead to less segregated schools relative to neighborhoods. However, the necessary conditions to achieve this are still an open question. Lastly, there are few cities that have the required data for these kind of models from which Amsterdam is one of the smallest, leaving it computationally tractable. Furthermore, as school segregation by income is more severe than along ethnic lines in Amsterdam (Onderwijsinspectie, 2022), income groups will be the focus.

Besides offering one of the first empirically calibrated ABMs for school choice and modeling the full city-scale, we show how our simulations provide a way to study a wide range of plausible and what-if scenarios. Even if we do not know parents' preferences precisely, we show that our empirically calibrated model implies that very strong tolerance needs to be assumed for increased choice to translate into lower levels of segregation, a result previously only derived by highly stylized models (Boterman, 2021b). Such mechanisms can only be understood if one considers the interactions within the system (Ladyman & Wiesner, 2020). Moreover, current empirical studies cannot start a school choice process from scratch, but only observe a snapshot, which might lead them to miss important non-linear effects of the interactions of current school choosers with the school compositions of previous years. Additionally, we conduct hypothetical experiments that are hard to observe in real-life: assessing the contribution of residential segregation to school segregation and analyzing effects of a policy that limits the choice set of parents for example. These experiments are relevant to policy makers because they can shed light on the effects of potential policy measures on processes that otherwise might take years to observe.

## 2. Data

Before presenting the Amsterdam primary school context in the next section, the data on which the simulations are based is first described in detail. Household data is gathered from the municipality of Amsterdam and is publicly available on several different, aggregated levels. The lowest aggregate level, "buurten" and referred to as neighborhoods from now on, is the used resolution in absence of individual-level data. The neighborhood boundaries are drawn by the municipality and consist of rivers, parks, and infrastructure (e.g., roads and railways). Specifically, the year 2019 (last complete one) from the Dataset Basisbestand Gebieden Amsterdam 2022 (BBGA) is used (Municpality of Amsterdam, 2022a). In 2019, Amsterdam had a total of 456,593 households and 72,565 children of primary school age (4–12 years old). On average (standard deviation), an Amsterdam neighborhood contained 1195 (713) households, of which 190 (163) primary school-aged children. The mean number of schools per neighborhood is 0.55 (0.78) schools, which are attended by 162 pupils (249) on average.

Per the neighborhood, the total number of primary school-aged children (4–12 years) and the percentage of people that belong to five income groups are available. These income groups are based on the after-tax income distribution of the entire population of the Netherlands. In this study, they are referred to as Quintile 1 (Q1) with an income less than €21,854, to Q5, more than €62,356. Q1 consists of the households that earn in the 0–20% quintile of the income distribution, Q2 households within the 20–40%, continuing up until

Q5. Reasons for choosing income over other household attributes is that segregation by income for Amsterdam primary schools is more severe than, for example, ethnicity (Onderwijsinspectie, 2022). Admittedly, it might not be straightforward to identify how much of a certain group (in this case Q1 and Q5) attend a certain school. Estimates of school compositions are not publicly available in the Netherlands, hence households have to use proxies for this in real life. We assume that people of lower and higher incomes (Q1 and Q5) are likely to be good proxies for differences that are instrumental in the school choice processes (Butler, 2003; Lareau & Weininger, 2003) and are better able to distinguish the "other" than, for example, Q3 versus Q4. Therefore, the extremes are selected, while Q2, Q3, and Q4 are left out. Moreover, excluding Q2, Q3, and Q4 also has computational reasons, reducing the number of households by more than 50%, decreases the runtime of the experiments drastically. However, as a robustness check, all five income groups are included in one experiment. In order to approximate the number of primary school-aged children belonging to Q1 and Q5 households, the number of 4–12-year-old children is multiplied by the percentage of households in these quintiles, and rounded to the nearest integer. Note that this assumes households of the different income percentiles have the same fertility rate and have only one child per household, which is an idealization due to lacking data. Although there is evidence that income levels and fertility rates are linked, it differs per context exactly how Doepke et al. (2022) and without proper data on this, we leave this for future work. However, for our modeling purposes this would change the relative compositions of the groups in our model, which could alter the segregation dynamics. Hence, this results in 47% (33,465) of Amsterdam's primary school-aged children in the model, 62% (20,748) of them Q1 households and 38% (12,717) of them Q5 households.

Combining the administrative boundaries of the neighborhoods and the number of households with primary school-aged children for the Q1 and Q5 group per neighborhood, the residential locations of households can be determined. Due to the openly available data, actual locations are not available and have to be approximated. Using building locations within the neighborhoods (Municpality of Amsterdam, 2022b), households are placed randomly in known built areas. Note that these buildings could very well be office spaces instead of actual houses, but parks, water, roads, and other infrastructure are discarded as household locations. As the simulation is executed thousands of times, a new random allocation is generated every time the model is started. This is to induce some variation in residential locations and avoids dependence on initial conditions.[1] Random within-neighborhood placement could miss potentially important within-neighborhood residential segregation

[1]More specifically, 120% of Q1 and Q5 households are generated per neighborhood. Every time the model is started, 100% out of the initial 120% is randomly sampled without replacement. This results in the simulation being executed with the actual known number of households (i.e. 100%), but with some residential variation.

(which is not captured now). If this would be the case, this reflects itself especially in case of strong distance preferences, increasing school segregation compared to "random within-neighborhood placement" scheme that is utilized.

For schools, the actual addresses are publicly available on the "Dienst Uitvoering Onderwijs" website (Dienst Uitvoering Onderwijs, 2022). These addresses are translated into coordinates, after which they can be placed alongside the households on the Amsterdam map. For the generated households, the spatial allocation can be seen in Figure 1. Euclidean distances from every household to every school serve as approximations to the actual travel time to the respective schools.

## 3. Primary school choice in Amsterdam

In the Netherlands children attend primary school when they turn 4 years old. This takes around 8 years to complete, with an equal number of consecutive grades. Currently, Amsterdam has 202 primary schools which are, Euclidean distance wise, fairly equally distributed with respect to the Q1 and Q5 income groups used in this study. Figure A1 shows that both groups have a similar average and substantial absolute number of schools available within a certain radius.

There are various factors influencing primary school choice and resulting segregation in Amsterdam, of which the socio-economic/ethnic composition of schools and home-to-school distance are the most important components (Manzo, 2013, pp. 0,0; Oosterbeek et al., 2021; Ruijs & Oosterbeek, 2019). The influence of these two factors on school choice and their effect on school segregation are described in more detail in the following sections. Note that distance and composition have separate sections, but they are inherently linked if households choose school further away because of composition preferences, for example.

### 3.1. Distance

Households prefer to send their children to a school close to home. However, relatively more socio-economically advantaged households are found to be more willing/able to travel than their counterparts (Manzo, 2013; Oosterbeek et al., 2021). There are multiple explanations for why households might choose schools further away, but one is that parents might deem the composition of schools closer to home unfavorable and hence attend one further away (Jongejan & Tijs, 2010; Oosterbeek et al., 2021). This immediately establishes an interaction between distance and composition preferences, which is discussed in more detail in the Composition section. Another explanation is that they travel further for specific school profiles, as the Dutch constitution allows schools to profile themselves along religious or pedagogical lines (Boterman et al., 2019b; Dronkers, 2016).

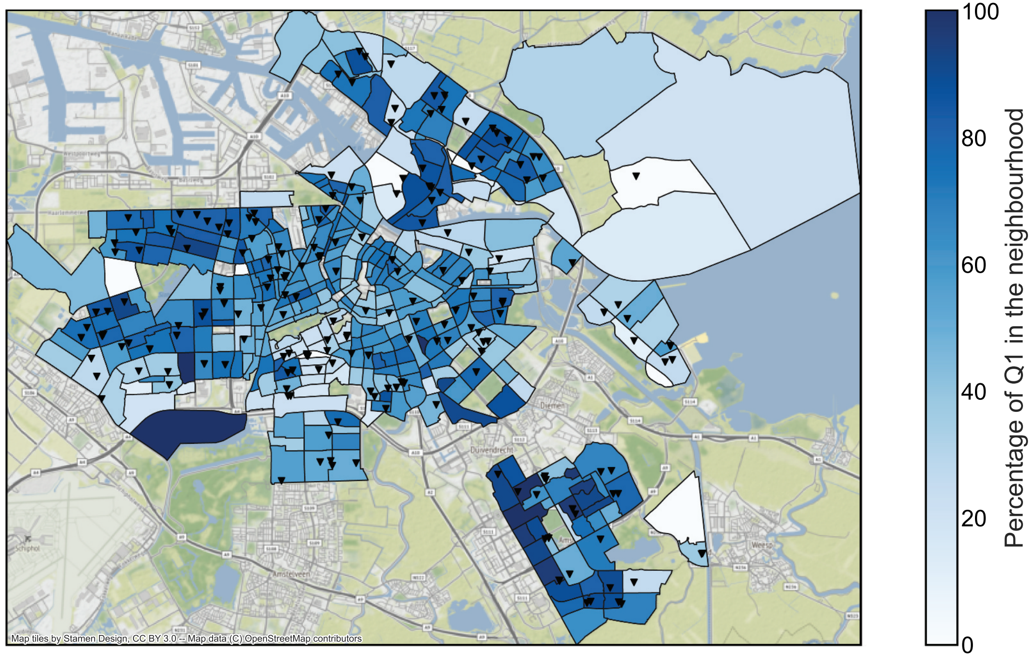


**Figure 1.** School locations (black triangles) and the percentage of Q1 households. Note that neighborhood size or color intensity do not say anything about the number of households within the neighborhood. No data is used for the neighborhoods between the South-East and the rest of Amsterdam because these are officially in other municipalities.

However, these kinds of schools might attract only certain groups of parents and thus increase school segregation (Borghans et al., 2015; Boterman, 2020). For example, highly educated parents are found to be more attracted to schools of a specific pedagogical denomination (e.g., Montessori). As these profiles and other identities are argued to be important in choice of school, this could induce these parents to travel further. Conversely, their counterparts might stick to closer schools, increasing segregation (Boterman, 2019). Similar reasoning can be applied to school quality. Although quality is a highly ambiguous concept, parents have a strong preference for academic performance of schools and even more so if the family is advantaged (e.g., education, income) (Manzo, 2013; Ruijs & Oosterbeek, 2019).

Taking this one step further, households might even move neighborhoods for school characteristics, making their preferred schools more accessible from a distance perspective (Boterman et al., 2019b; Burgess et al., 2015). This could strengthen the link with residential segregation and would result in these households living close to their schools. On the other hand, parents might move into more diverse neighborhoods (i.e., gentrification), but avoid the neighborhood schools and travel further to attend schools more to their preference (Candipan, 2020; Renzulli, 2005).

More specific to Amsterdam is the primary school allocation system that gives you priority at the eight closest primary schools (Gemeente Amsterdam, 2020). However, as shown in Figure A1, around 50% of Q1 and more than 50% Q5 households have more than 8 schools within a 2 km radius. Moreover, 86% of Amsterdam primary school choosers in 2020 attended a priority school (Breed Bestuurlijk Overleg, 2022) and the average distance to the attended school per ethnicity/educational group is less than 1600 m (Muncipality of Amsterdam, 2017). Note that the latter information is not available for the income groups that are used here, but as these parents also belong to the income groups, it is assumed that Q1/Q5 households also attend schools at least this close on average. However, it is less clear how this priority scheme and the spatial distribution of schools affect the level of school segregation. Will offering more choice decrease the effect of residential segregation in schools and allow for households to desegregate compared to what otherwise would be expected if everyone is forced to attend their closest school (i.e. fully determined by residential segregation)? Earlier theoretical studies have shown that strong distance preferences could restrict school segregation in an already residentially segregated setting, but only if preferences of parents are sufficiently tolerant (Boterman, 2021b, p. 0).

### 3.2. Composition

Households tend to attend schools that have a larger proportion of children belonging to their own group (e.g., income, ethnicity, socio-economic) (Jongejan & Tijs, 2010; Oosterbeek et al., 2021). Similarly, some groups are found to avoid schools with a large share of other groups (Karsten et al., 2003) and there is also an expressed interest for diverse schools in Amsterdam specifically (Manzo, 2013). One underlying mechanism for this is the choice of homophily, the preference to attend schools that have similar others (McPherson et al., 2001). However, it is important to distinguish between observed homophily (i.e. school segregation) and the choice of homophily (i.e. preference), as other possible mechanisms for could be underlying the observed homophily besides the choice of homophily (Wimmer & Lewis, 2010). For example, residential segregation projecting itself in schools, as parents have a preference for close schools but not for composition necessarily. In this case (substantial) residential patterns and distance preferences are part of the explanation. In the extreme case, when everyone attends or is forced to attend their closest school, residential patterns fully determine school segregation. In real life, this is often not the case and people might (want to) live in diverse neighborhoods, but distance preferences or the ability to travel for schools are found to be heterogeneous and group specific. For example, affluent households might – want to – live in a diverse neighborhood (i.e. gentrification), but when it comes to schools affluent households might avoid the "neighbourhood" school and could be more willing or able to travel further to – perceived – more favorable schools (Manzo, 2013; Candipan, 2020). Note that a specific profile or school quality/academic performance could also reflect itself in school compositions. For example, if parents with a certain religious affiliation cluster together in schools with a similar profile (Dronkers, 2016) or if children from highly educated parents are the majority in a school, you can expect academic performance to be better (Boterman et al., 2019b).

## 4. Modeling primary school choice in Amsterdam

To model primary school segregation in Amsterdam, the school choice process in the COMPASS[2] model of Merry & Boterman (2022c) is used and adapted. As mentioned in the Data section, households' residential locations are approximated and fixed throughout the school process. Note that this residential approximation is different every time a simulation starts, to avoid dependency on one specific initial residential allocation. To start the school process, households are randomly allocated to schools initially (i.e. at time step 0). In subsequent time steps, households base their school choice on the home-to-school distances and current school compositions. This process runs until the whole system reaches a stable state and hence household preferences/constraints result in a certain level of school segregation.

Every household assigns a utility to a school that consists of two components: distance and composition. For the distance to a school, closer to home means

[2]https://gitlab.computationalscience.nl/edignum/compassproject.

a higher utility. for a school's composition, the closer to the household's optimal fraction ($t$) of similar people in the school, the more utility they obtain. This means Q1 households consider the fraction of Q1 children attending the schools and vice versa for Q5. Both utilities are bounded by 0 and 1 and are weighted according to a parameter $\alpha \in [0, 1]$. This combined utility is calculated for every school and results in a ranking of all the schools. If a household is allowed to switch, it tries to attend its highest-ranked school, if there is insufficient capacity it tries its second, and so on. This process is repeated for every household. Below the utility calculations and simulation are described in more detail.

### 4.1. Distance utility

School choice research indicates that households prefer schools close to home; however, it is less clear how home-to-school distance affects the choice of parents. Is there are a hard cutoff, where households filter schools on distance and only consider schools within a certain radius or is there a more gradual decrease? Existing studies have used distance as a filter to approximate "feasible choice sets" of parents and reduce computational expense (Zwier et al., 2022) and distance decaying functions (Oosterbeek et al., 2021; Ruijs & Oosterbeek, 2019). As empirical research for Amsterdam is scarce on this, a sigmoidal function is used that can accommodate both choice processes (i.e. filtering and more gradual decrease), parameterized by $p$ and $q$ (Eq. 1). This allows us to study how both types of distance preferences influence school segregation. The distance to a certain school is transformed in a utility ($U_d is$) between 0 and 1 using a sigmoidal function. Note that $d$ stands for distance, $i$ and $s$ are the specific household and school, respectively. The parameter $p$ determines at which distance a household receives a utility of 0.5, with 1 being the maximum, while $q$ controls the slope (larger indicates a less gradual decrease). Figure 2 presents three possible realizations of this. When $q$ is large it approximates filtering, whereas if it is smaller the decrease is more gradual. For small $p$ and $q$ households clearly differentiate between schools that are very close to home. For example, they would assign noticeably different utilities to a school 500 m away, compared to one 600 m away.

$$U_{dis} = \frac{1}{1 + \left(\frac{d_{is}}{p}\right)^q} \tag{1}$$

### 4.2. Composition utility

The composition utility follows from the asymmetric utility function also used in Merry & Boterman (2022c). The fraction $x_{is}$ represents the proportion of similar households to household $i$ attending school $s$ (Eq. 2). As parents are

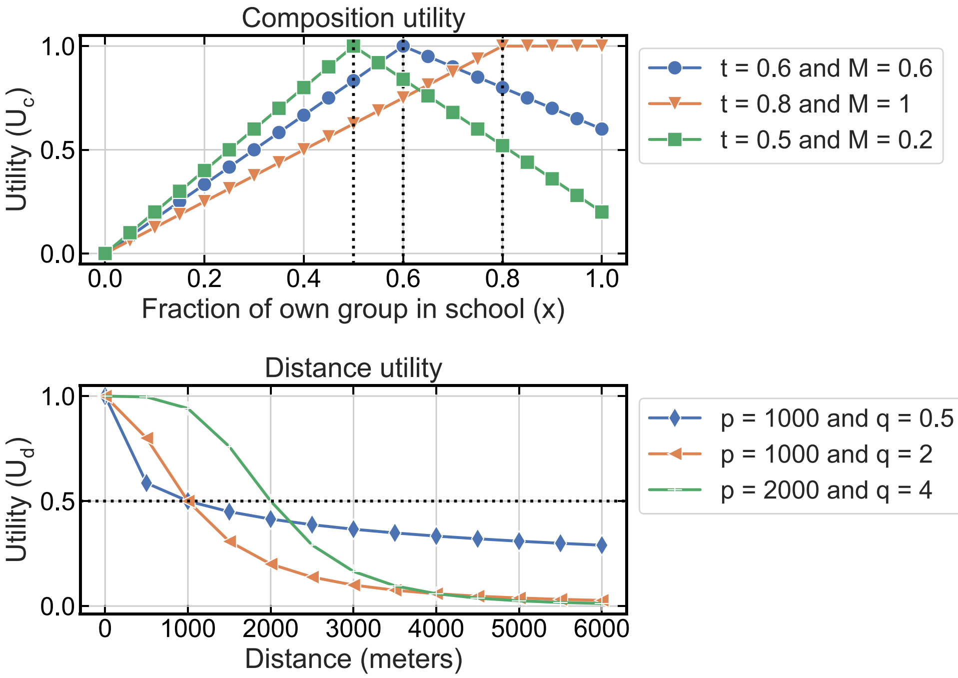


**Figure 2.** Distance and composition utility functions for several values of the parameters.

found to prefer schools with more of their own group attending, this is modeled as a linearly increasing function. However, hypotheses for primary schools in the Netherlands also suggest households actually prefer some level of diversity over homogeneity (Manzo, 2013). Hence, assumption is that households have an "optimal fraction" of similar households in a school $t \in [0, 1]$ for which they get the maximum composition utility of 1, but there is also a penalty $M \in [0, 1]$ when a school is homogeneous from the perspective of that household. Meaning that if household $i$ only has similar households attending that school, it receives a composition utility of $M$. Figure 2 shows three possible realizations when $t$ and $M$ are varied. If the optimal fraction is low, households are relatively tolerant, if $M$ is low, households strongly prefer some diversity.

$$U_{cis} = \begin{cases} \frac{x_{is}}{t} & x_{is} \leq t \\ M_i + \frac{(1-x_{is})(1-M_i)}{1-t} & x_{is} > t \end{cases} \tag{2}$$

These utilities are then combined using a weight $\alpha$:

$$U_{is} = \alpha U_{cis} + (1 - \alpha) U_{dis}. \tag{3}$$

When $\alpha = 0$, only distance to school governs the choice and if $\alpha = 1$ only school compositions. Note that this linear combination is chosen for ease of interpretation instead of the Cobb-Douglas utility function used in Merry & Boterman (2022c).

### 4.3. Simulation

To simulate primary school choice in the entire city of Amsterdam some additional parameters are needed. Firstly, schools are given a minimum and maximum capacity. This capacity depends on the multiplier $c \geq 1$, which makes sure schools cannot grow too large and there are at least enough places for every child (Eq. 4). For example, when the multiplication factor is two ($c = 2$), schools are allowed to grow two times as large as the average school size in the city. Additionally, a minimum capacity of 50 pupils per school is used, as empty schools are not realistic. Although not further investigated in this study, the maximum capacity can be a possible policy instrument, as Boterman (2021b) find that asymmetric preferences and limited school capacity can impact school segregation, particularly in residentially segregated settings.

$$\text{capacity} \in \left[50, \left\lceil c \cdot \frac{\text{total students}}{\text{total schools}} \right\rceil\right] \tag{4}$$

Every household ranks schools according to the utility it would give them (in descending order). After a random initial allocation (to calculate utilities), they try to attend their highest-ranked school in the subsequent time steps. If that school is full, the household tries to attend the second ranked school, and so on. Additionally, children roughly spend 8 years in Dutch primary schools, but for simplicity and convergence of the model, the specific in- and outflow of these children is not modeled nor do they incur any switching costs.

Households are not given exactly the same optimal fraction ($t$), utility at maximum ($M$) and $\alpha$. They are sampled from a truncated normal distribution bounded between 0 and 1, with a mean and a standard deviation $\sigma$ that can be varied. Additionally, the fraction of households ($f$) allowed to change schools in every time step can be varied as well, which leads to a reduction in computation if it can be fixed at a low value. Lastly, one run of the model is considered converged if the average and standard deviation of all households' utility at their current schools and school segregation (measured by the Dissimilarity index) have not changed by more than 0.01 in the last 30 time steps. These numbers are determined in the convergence analysis with the COMPASS model (Merry & Boterman, 2022c). Algorithm 1 gives a brief overview of the simulation in pseudo-code.

## 5. Results

Although the expectation is that the household parameters controlling the school preferences of the households are affecting the level of school segregation the most, there might be other influential parameters or interactions between them that influence the model dynamics. Moreover, the simulation parameters summarized in Table 1 are mostly introduced for the convenience of the model or simulation, but lack a specific sociological meaning here. Hence, ideally they do not affect the level of school segregation and can be fixed at nominal values to speed up computation and/or focus on the other parameters.

Global Sensitivity Analysis (GSA) is conducted first, where all parameters are varied simultaneously. A GSA allows one to rank the parameters according to how sensitive the output is to changes and possibly distinguish influential from less-influential parameters and find important interactions. However, with this sort of analysis it can be hard to observe the mechanisms behind the effects of parameters and the level of school segregation. Hence, to do this, a Local Sensitivity Analysis (LSA) is carried out, where the parameters are varied individually and systematically. Hence, one can infer what happens when one parameter changes at a time or when two are varied simultaneously (i.e. interaction effects). Additional details of the methodology can be found in Merry & Boterman (2022c) and all indices are available in the appendix or upon request.

For two measures of school segregation as output, a clear distinction can be seen for the optimal fraction ($t$), utility at homogeneity ($M$) and $\alpha$ compared to the rest of the parameters (Figures A2-A4), meaning that these parameters are most influential with respect to the output. Additionally, the school capacity parameter, $c$, $p$, $q$ and the fraction of households allowed to switch schools (fraction moved, $f$) are considered somewhat influential based on their sensitivity indices.

**Table 1.** Parameters, their description, ranges in the sensitivity analysis and nominal values.

| | Range | Substantiation | Nominal value |
|---|---|---|---|
| **Household/school parameters** | | | |
| Optimal fraction ($t$) | [0.40, 0.80] | $<$ 0.4 convergence issues<br>$>$ 0.8 full segregation | 0.5 |
| Composition/distance trade-off ($\alpha$) | [0, 1] | Full range | 0.2 |
| Utility at homogeneity ($M$) | [0, 1] | Full range | 0.6 |
| $p$ | [0, 4000] | | 1000 |
| $q$ | [0, 6] | | 2 |
| School capacity ($c$) | [1, 5] | | 2 |
| Minimum school size | | Too small schools not allowed | 50 |
| **Simulation parameters** | | | |
| Standard deviation of $t, M, \alpha$ ($\sigma$) | [0, 0.05] | Computational | 0.01 |
| School steps | | | 300 |
| Convergence threshold | | | 0.01 |
| Window size | | | 30 |
| Fraction of agents moved ($f$) | [0, 1] | Full range | 0.125 |

Figure A5 shows the change in school segregation if the parameter is varied. Note that in the GSA all parameters are allowed to vary and hence these are "average" effects. All four school segregation measures show that small values for the fraction of households moved ($f$) and the school capacity ($c$) lead to lower levels of segregation. This is to be expected; if the former is very small, few households will be able to switch schools, limiting possible impacts on overall segregation levels, while if the latter, $c$, is small, this does not create much opportunity for moving between schools. Although $f$ has some effect, this seems to be the case only when it is very close to zero. Hence the effects of $c$ and $f$ can be explained and are fixed at nominal values for simulation reasons together with $\sigma$. The fraction of households moved, $f$, is fixed at 0.125 and $c = 2$ to decrease computations per time step but still allows for sufficient movement. The standard deviation $\sigma$ does not show any effect in the entire GSA (Figure A5) and is fixed at $\sigma = 0.01$. These simulations/computational parameters are expected to mainly impact the convergence rates and hence fixing them is not affecting the analysis of the mechanisms underlying school segregation. This also means that the most influential parameters are those with an actual sociological meaning, which will be further analyzed in the next section.

### 5.1. Mechanisms of school segregation

Focusing on the subset of most-influential parameters according to the GSA, an LSA is conducted. Results when one or two parameters are varied, while keeping the others fixed at their nominal values, are presented in Figures 3 and 4. Before discussing these in detail, the first thing to note is the sharp decrease in the level of school segregation when households only value school compositions ($\alpha = 1$) when $t = 0.5$ (Figure 4). As households start with a randomly allocated school in which the composition likely reflects the municipality shares (62%–38%) (20,748 and 12,717 households respectively), for $t = 0.5$, both groups can obtain more utility by switching to a school that is closer to 50%–50%. However, this is not possible in every school as the Q5 group only consists of 38%. In this case, the algorithm is not able to find a solution in which the households stop switching. If households' optimal fractions are set to their municipal shares (62%–38%) instead of 50%–50%, this numerical artifact disappears (Figure A6) while the rest of the dynamics are qualitatively similar as in Figure 3. In the following paragraphs, the rest of the results are discussed.

When households increase their optimal fraction ($t$) they explicitly require more of their own group in a school to acquire the maximum utility and hence the level of school segregation rises. Increasing the utility for homogeneous schools ($M$) means reducing the penalty on segregating behavior. Prioritizing composition more over distance (increasing $\alpha$) also leads to segregation levels

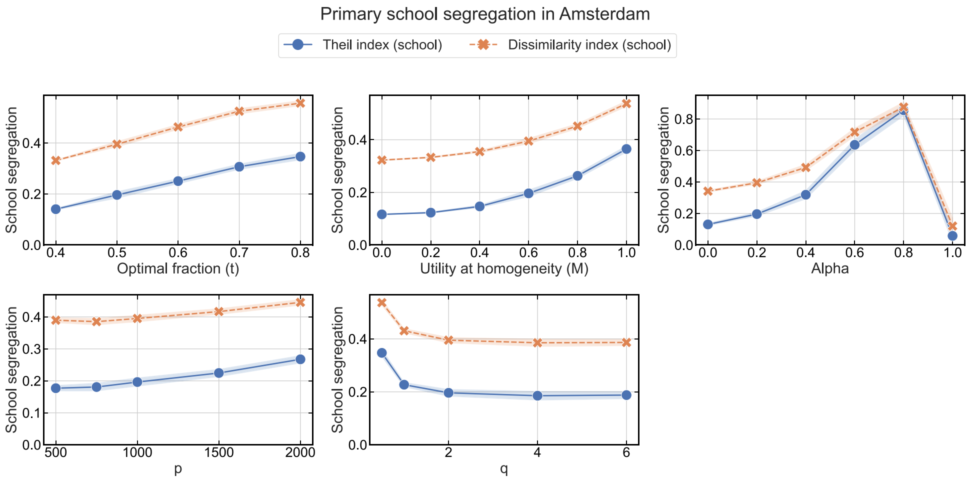


**Figure 3.** Varying one parameter at a time and two measures of school segregation. The plotted level of segregation is the average over 10 model runs, the bands represent the plus and minus the standard deviation. Nominal values of the parameters that are not varied can be found in Table 1.

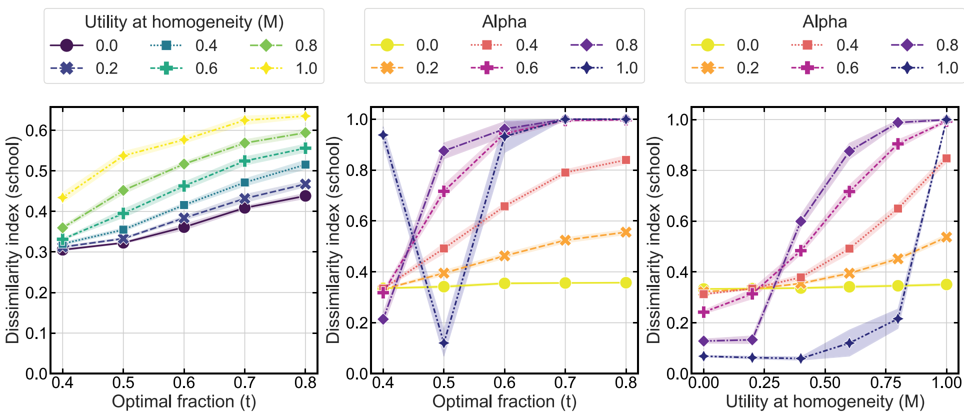


**Figure 4.** School segregation as measured by the Dissimilarity index, whilst varying two parameters at a time. The plotted level of segregation is the average over 10 model runs, the bands represent the mean plus and minus the standard deviation. In the first plot, the optimal fraction ($t$) and penalty for homogeneity ($M$) are varied. The second varies the weight of composition over distance (alpha, $\alpha$) and $t$. The last varies alpha and $M$. Nominal values of the parameters that are not varied can be found in Table 1.

rising which is not surprising. Figure 4 shows that increasing two of these three parameters simultaneously only increases the level of school segregation even further. However, when the penalty of homogeneous schools is very low and $\alpha$ is high, households are really forced to their optimal fractions. This reduces segregation even under the level of what would be expected if everyone chooses the closest school ($\alpha = 0$). These results are consistent with other, more theoretical ABMs of school choice (Bernardi, 2014, pp. 0,0). Reported levels of school segregation in Amsterdam are around 0.6–0.7 for the Q1-Q5 income groups when measured by the Dissimilarity index (Onderwijsinspectie, 2022).

The parameters controlling the distance utility, $p$ and $q$ have a substantially smaller effect. However, note that this is for a fixed optimal fraction $t = 0.5$, $M = 0.6$ and $\alpha = 0.2$. Nevertheless, segregation increases if $q$ becomes very small. In this scenario, small changes in distances lead to large differences in utility, which might not be very realistic as common sense indicates most people would be indifferent between a school 500 versus one 600 m away. However, as a lot of households have quite some schools close to their home (Figure A1), there is enough opportunity to segregate based on composition preferences and regular Schelling dynamics unfold (Bernardi, 2014; Schelling, 1971). In this context, this dynamic entails that children move out of schools even though they strive for a mixed school, triggering a feedback mechanism which causes others to move as well. When $p$ increases, schools further away still receive substantial utility and hence are also attractive choices, leading to composition preferences driving the level of school segregation once again.

### 5.2. The role of residential segregation and geographic assignment policies

One major benefit of the model is that one can simulate hypothetical scenarios that might be difficult or even impossible to observe in real life. For example, to study the effect of residential segregation in schools, one would want to observe what schools households would choose when there is residential segregation present and when there is not. While impossible in reality, in the model, households can be randomly allocated to a residential location (i.e. residential segregation is zero or at least very low). The level of school segregation that emerges when residential segregation is basically zero can then be compared to the simulations where one starts from a residentially segregated state. The difference in the observed levels can then be ascribed to residential segregation for this model specifically. Moreover, one can analyze the potential effect of policy in the form of geographic assignment mechanisms, such as the priority at your eight closest primary schools in Amsterdam (Gemeente Amsterdam, 2020). Do they strengthen the link between residential and school segregation, what would happen if one gives priority at the four closest and not eight, does this increase or decrease the level of school segregation?

Figure 5 shows the level of school segregation with (blue) and without (orange) residential segregation, as well as the difference between the two (green). For all parts of the parameter space, the green line is above zero, indicating a nonzero effect of residential segregation on school segregation. This effect appears to be larger when school segregation is low, leaving enough room for distance preferences to be of substantial influence. In empirical studies, researchers often use the level of school segregation when assigning children to their closest school (black lines) as proxy for the effect of

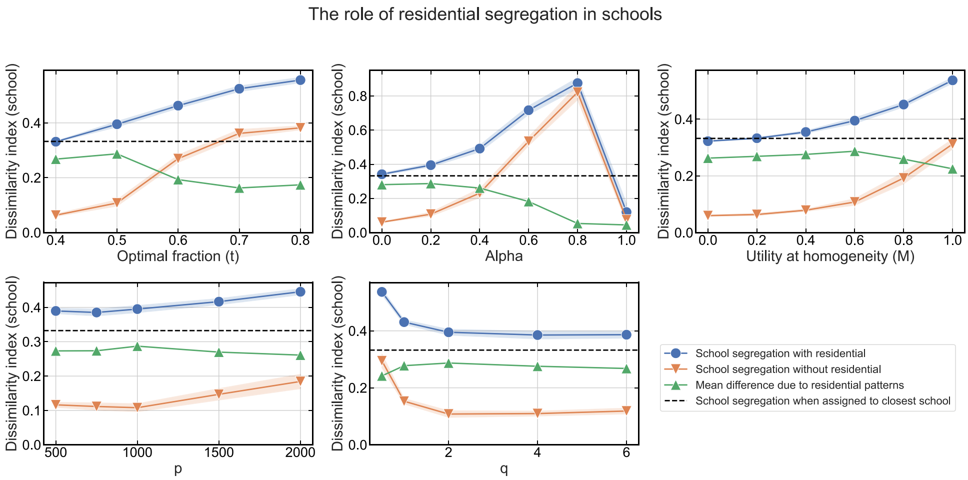


**Figure 5.** The estimated effect of residential segregation in schools while varying one parameter at the time. School segregation with residential patterns (blue), without (orange) and their difference (green). School segregation when assigning children to their closest school (Euclidean) is in black. Both the difference and the closest school allocation are based on the means of 10 model runs. Nominal values of the parameters that are not varied can be found in Table 1.

residential segregation. Although the closest school allocation is quite close to the actual effect of residential segregation in this model, it seems to overestimate it, especially when schools already segregate without residential patterns. This happens when households are less tolerant (high optimal fraction or small penalty for homogeneous schools) or emphasize composition over distance ($\alpha \geq 0.5$). Although $p$ and $q$ control the distance utility function, they do not really affect the role of residential segregation in schools. This suggests there are sufficient schools available very close to home for the majority of households to segregate (see also Figure A1).

As the municipality of Amsterdam has the policy to give priority at the eight closest primary schools given your residential location, it is interesting to see how it affects school segregation if this number is changed. Hence, an experiment is conducted where only the $n \in \{1, 2, 4, 8\}$ closest schools are considered. These schools are very close on average to the majority in both the Q1 and Q5 group, so only composition preferences are assumed to matter ($\alpha = 1$). Note that this is not exactly a priority scheme, but a strict assignment mechanism, although 86% attends a priority school in Amsterdam (Breed Bestuurlijk Overleg, 2022).

If only one school is considered, everyone attends their closest school and the level of school segregation does not change when the parameters do (Figure 6). However, as soon as households have more schools available, school segregation starts to emerge given their preference. Although the effects are more pronounced for the optimal fraction than for the utility at homogeneity, segregation almost exclusively rises when the number of

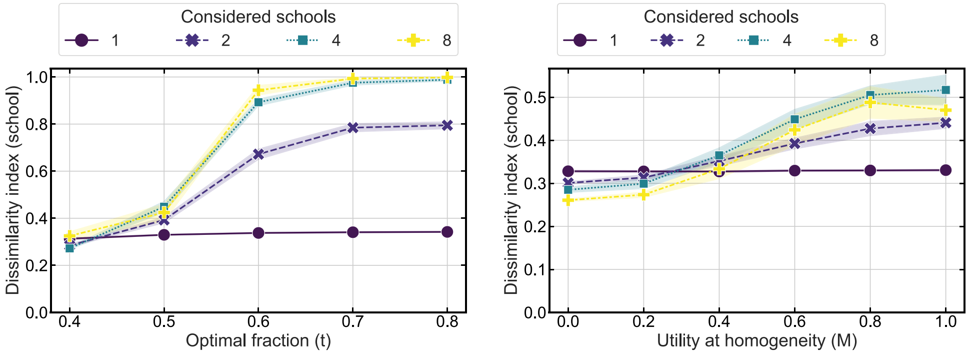


**Figure 6.** Varying the number of schools considered together with the optimal fraction or utility at homogeneity. The plotted level of segregation is the average over 10 model runs, the bands represent the 10% and 90% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

schools is increased, even for moderate preferences. Additionally, the increase in segregation from 2 to 4 schools considered is larger than from 4 to 8. Hence, in this model, more school choice leads to more segregation, which is also a consistent finding in empirical research (Boterman et al., 2019b; Wilson & Bridge, 2019). Also relevant here is that at some point more choice does not mean anything anymore (schools are fully segregated). This could reflect itself in the very weak effect of the introduction of the 8 priority schools in Amsterdam, if this had been 4 the effect could have been bigger (Boterman, 2022a).

Note that in the part of the parameter space where households strive to be a minority ($t < 0.5$, $M < 0.4$), although very small, more schools actually allow for less segregation to emerge compared to when households only consider the closest schools, resonating earlier results from stylized models (Boterman, 2021b). Both increasing the optimal fraction and the utility at homogeneity still show the same increasing trend as in the original simulations. This suggests the final segregation that emerges is a delicate balance between preferred and locally available compositions (i.e., within "reasonable" distance); moreover, this interaction is highly non-linear. This means understanding how distance is considered is vitally important to understand the potential for segregation in the city and also offering the number of priority schools can have significant effects on the resulting segregation levels. If one prefers more of their own group and residential segregation is strong: more schools might not help to desegregate, but if one prefers a smaller fraction of similar households and residential segregation is strong, then more choice could lead to less segregation.

The steep increase from one to two schools raises the question if this is mainly due to having only two groups. Because, for two groups, two schools are sufficient for segregation dynamics to unfold. Hence, the same experiment

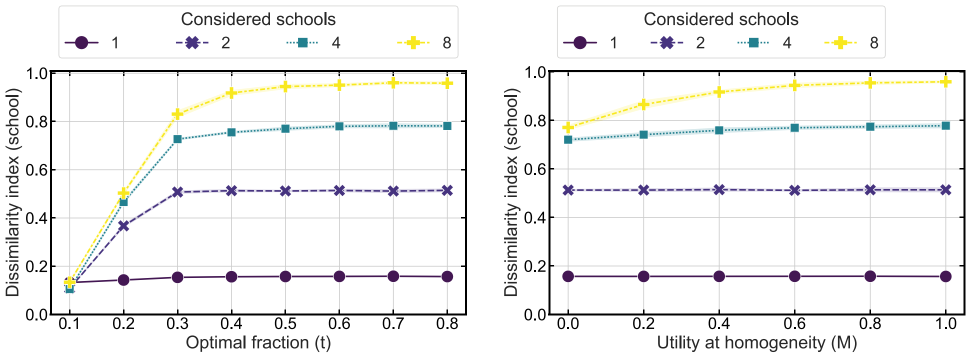


**Figure 7.** Varying the number of schools considered for all five income groups together with the optimal fraction or utility at homogeneity. The plotted level of segregation is the average over 10 model runs, the bands represent the 10% and 90% percentiles. Nominal values of the parameters that are not varied can be found in Table 1.

is repeated for all five income groups, resulting in roughly 72,000 households with the following shares in the municipality: Q1: 27%, Q2: 21%, Q3: 18%, Q4: 15%, Q5: 19%. Note that all groups are equally distant from one another as they consider only their own group and "others" in the utility calculations. In real-life, Q4 and Q5 might see each other as more similar than Q1 and Q5, for example. Looking at Figure 7 one can see that going from one to two schools still leads to a large increase in school segregation. However, it is clearly limited to a certain level. Only if they are allowed to consider more schools than groups (i.e. eight schools) and without distance preferences, the system is able to fully segregate. Note that for two groups this already happens when four schools are considered. The utility at homogeneity does not seem to have a large effect, which can be explained by the fact that the optimal fraction is still fixed at 0.5 for this figure and $M$ only controls the composition utility between 0.5 and 1. Households from the specific groups are very unlikely to find a school with a composition of more than 50% of their own group. This hints at an important link between the number of groups, how similar other groups are deemed in society and the availability of schools. Further research could shed more light on this complex interplay of factors.

## 6. Conclusion

This study developed one of the first empirically calibrated ABMs of school choice, based on the Amsterdam primary school context. This has allowed us to disentangle some of the complex interplay of distance and composition preferences, as well as the interaction of school choice driven by these preferences with policy in an empirical context. More specifically, using publicly available data on the neighborhood ("buurt") level, residential locations of low- and high-income households, 0–20% (Q1) versus 80–100% (Q5) respectively, of the countrywide after-tax income distribution are approximated. Euclidean distances between residential and accurate school locations allowed for the simulation of school choices under various composition- and distance preferences of households.

The direct and interaction effects in the local sensitivity analysis show that, in this model, substantial school segregation emerges even when the households are relatively tolerant to diversity. Reiterating findings from stylized models of school choice (Bernardi, 2014, pp. 0,0). For example, requiring only 50% ($t \geq 0.5$), of similar pupils in a school and explicitly wanting some diversity (i.e. homogeneous schools are penalized, $M \leq 0.6$), can lead to empirically observed levels of school segregation (i.e. Dissimilarity index of 0.6–0.7). A result found in interviews with middle-class households in Amsterdam as well (Manzo, 2013). However, this is conditional on having at least some preference for the composition of a school ($\alpha \geq 0.2$), otherwise residential segregation and very strong distance preferences restrict school segregation.

One underlying reason is that most households from both groups have ample schools in close proximity to let composition preferences (i.e. $t, M, \alpha$) dominate the school choice, allowing for regular Schelling dynamics with feedback effects to emerge (Bernardi, 2014; Schelling, 1971). This means that although households might be relatively tolerant and/or explicitly want some diversity, they end up with segregated schools. When some households from a particular group might move schools, this triggers other households (potentially from both groups) to switch as they can now obtain a higher utility elsewhere, resulting in cascading effects (Merry & Boterman, 2022c). This reasoning is substantiated by the "restricting school choice" experiment, because as soon as households go from having only one school to choose from, to two or more, a substantial increase in school segregation can be observed.

Another benefit of this model is that one can simulate hypothetical scenarios that are hard or even impossible to implement and monitor in real-life. By randomly shuffling households' residential locations and letting them choose schools, the level of school segregation that can be expected with neutral residential patterns is simulated. Comparing this neutral scenario with the level of school segregation resulting from empirically realistic residential patterns, we can then accurately estimate the level of school segregation due to residential segregation. This experiment shows that residential segregation always has an influence in this model; however, it affects school segregation most when composition preferences are not too strong. In case of strong composition preferences (i.e. intolerance), schools will segregate substantially regardless of residential patterns. In a second experiment, the distance preferences are replaced by letting households only consider the $n$-closest schools ($n \in \{1, 2, 4, 8\}$), as a substantial proportion of households have up to this number of schools within walking/cycling distance in Amsterdam. Reiterating the ample choice argument: as soon as households are allowed to have two choices instead of one, segregation increases drastically, even for tolerant households ($t \leq 0.5$); this increase is even more severe from moving from one to four or from one to eight schools. These effects are also visible when decreasing the penalty for homogeneity ($M$), but to a lesser extent. Interestingly, if all five income groups are considered, full segregation only occurs when more schools than groups are considered. This makes sense as sufficient schools are needed to be able to segregate, but points to an important relation between the groups and availability of schools. However, this likely depends on the assumption that households distinguish only between their own group and others. If they would deem more similar others more desirable, a Q1 household might feel more similar to Q2 than Q5 households, the effect would probably be smoother. Further research could identify what groups are important in a specific context, how similar others are and how the number of schools reasonably available to households affect school segregation.

The findings of our empirically calibrated simulations suggest that it is important to obtain more precise empirical knowledge on what composition preferences households have, the number of schools within feasible distance and levels and patterns of residential segregation. Moreover, we demonstrated that interaction effects of these variables on school segregation can be strongly nonlinear. In our model, restricting the subset of schools households can choose from can significantly affect resulting segregation levels. These results speak to existing empirical findings from school choice research. Empirical research from England, for example, also links a higher population density and hence school density (more choice), to higher levels of school segregation (Burgess et al., 2005). Additionally, in a comprehensive literature review, Wilson and Bridge (2019) link an increase in school choice across various contexts around the world (some of them policy-related), to an increase of school segregation. If parents prefer more of their own group and residential segregation is strong, more schools might not help to desegregate. On the other hand, if parents prefer a smaller fraction of similar households and residential segregation is strong, then more choice could lead to less segregation. That distance-related policy can have an effect is shown in the historical integrated busing policy in the United States. In this policy cities with substantial residential segregation were allowed to bus children to schools to districts further away, increasing the number of schools within feasible distance and thereby mitigating the influence of residential on school segregation (Billings et al., 2014). However, this might have had adverse effects as well, sparking large numbers of White households to move out of neighborhoods (White Flight), avoiding these policies and less desired school compositions (i.e. having implications for school compositions preferences) (Orfield, 2001). It should be noted that busing was also combined with local policies and political/jurisdictional issues, and hence its effect is hard to reduce to just distance/composition preferences in a more realistic model (Logan et al., 2008). It does show that understanding how distance and composition are considered and how they interact is vitally important to understand the potential for school segregation in a city and possible counteracting policies. Empirically calibrated agent-based modeling offers a powerful tool for this that future research could employ. Moreover, often employed methodologies to estimate school choice preferences, such as discrete choice models, do not allow to incorporate these interactions and feedback mechanisms explicitly. This might lead to incorrect inference and ineffective policies (Maes, 2021), estimating households to be less tolerant than they actually are, for example, as feedback mechanisms might be part of the driving force besides intolerance.

As our model is one of the first empirically calibrated ABM of school choice, the incorporated model household behavior is very much a simplified version of reality. Several simplifications point to avenues for future extensions of our work. For example, households are shown to choose schools not only based on distance and composition, but also consider school quality, profile and rely on their social network (Boterman, 2022b). Furthermore, these factors might be weighted non-linearly, instead of linearly which is what is assumed in this model. In addition, they are subject to certain allocation mechanisms and might use walking/cycling distance instead of Euclidean. Moreover, schools themselves could be gatekeeping or other institutional behaviors can be at play (Boterman et al., 2019b). Some households might even go as far as moving neighborhoods for schools (Boterman, 2021a), which is a bidirectional relationship this model has not implemented. Households are also found to have heterogeneous preferences (Hailey, 2022), meaning that they almost certainly do not all have the same real-life counterparts of $t, M, \alpha, p, q$ as they do in this model. Additionally, the only household characteristic used here is parental income. While this is shown to be important for school choices and resulting segregation, households might identify with multiple groups and/or along other lines (e.g. ethnicity, educational level) (Van Noord et al., 2019). Also, households can be classified to belong to a certain group according to data, but they might not identify with that group themselves. Input for that can come from more qualitative studies, for example. Furthermore, in the five income groups experiment, the different groups are treated as equidistant from each other, an assumption that is likely violated in real-life. There are also reasonable expectations for heterogeneous preferences between groups. For example, a scenario where Q1 might prefer more diversity, while Q5 prefers Q5 homogeneity. In the model, Q5 would not have a high penalty for homogeneous schools, making it more likely that substantial school segregation arises, even if Q1 is relatively tolerant. However, to properly model this, it would increase the number of parameters in the model among others and is left for future work.

On a more methodological level, inference using agent-based modeling can be compared with, for example, regression techniques, which are still used extensively in school choice/segregation studies. Comparisons of model predictions in future research can then show if one method works better than another, and if so, when. Lastly, and perhaps most importantly, our model is not validated as there is no calibration with actual school choices of households in Amsterdam. While our model allows for this possibility and corresponding techniques are available, this remains for now an area for future research. The only form of validation shown here, is that the model is able to recreate similar levels of school segregation as observed in real-life. This does not guarantee that household behavior and their school choices within the model mimic that of reality however. Additionally, combining model calibration with an injection of empirical realism in the form of more household/ school characteristics (i.e. heterogeneity) the model might shed more light on true underlying mechanisms of school segregation. This can pave the way for more realistic policy simulations that take the interactions within/between factors and different groups/households into account and hopefully lead to the development of more effective measures that can reduce levels of school segregation in educational systems all over the world.

## 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 third 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). The work was supported by the Ministerie van Onderwijs, Cultuur en Wetenschap [024.003.025]; Nederlandse Organisatie voor Wetenschappelijk Onderzoek [464.18.112]; the Dutch Inspectorate of Education and the City of Amsterdam.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Eric Dignum http://orcid.org/0000-0002-9560-8005
Willem Boterman http://orcid.org/0000-0002-8908-5842
Andreas Flache http://orcid.org/0000-0002-8276-8819
Mike Lees http://orcid.org/0000-0002-5457-9180

## References

Boterman, W. R. (2020). Socio–spatial strategies of school selection in a free parental choice context. *Transactions of the Institute of British Geographers, 46*(4), 882–899.

Boterman, W. R. (2021a). The role of geography in school segregation in the free parental choice context of Dutch cities. *Urban Studies, 56*(15), 3074–3094.
Boterman, W. R. (2021b). School segregation in the free school choice context of Dutch cities. *Understanding School Segregation: Patterns, Causes and Consequences of Spatial Inequalities in Education*. London: Bloomsbury, 155–178.
Boterman, W. R. (2022a). Dealing with diversity: Middle-class family households and the issue of 'black' and 'white' schools in Amsterdam. *Urban studies, 50*(6), 1130–1147.
Boterman, W., Musterd, S., Pacchi, C., & Ranci, C. (2022b). School segregation in contemporary cities: Socio–spatial dynamics, institutional context and urban outcomes. *Urban Studies, 56*(15), 3055–3073.
Bernardi, F. (2014). Compensatory advantage as a mechanism of educational inequality: A regression discontinuity based on month of birth. *Sociology of Education*, *87*(2), 74–88.
Billings, S. B., Deming, D. J., & Rockoff, J. (2014). School segregation, educational attainment, and crime: Evidence from the end of busing in charlotte-mecklenburg. *The Quarterly Journal of Economics*, *129*(1), 435–476. https://doi.org/10.1093/qje/qjt026
Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), 7280–7287.
Borghans, L., Golsteyn, B. H., & Zolitz, U. (2015). Parental preferences for primary school characteristics. *The BE Journal of Economic Analysis & Policy*, *15*(1), 85–117. https://doi.org/10.1515/bejeap-2014-0032
Boterman, W. R. (2019). The role of geography in school segregation in the free parental choice context of Dutch cities. *Urban Studies*, *56*(15), 3074–3094.
Boterman, W., Musterd, S., Pacchi, C., & Ranci, C. (2019b). School segregation in contemporary cities: Socio-spatial dynamics, institutional context and urban outcomes. *Urban Studies*, *56*(15), 3055–3073.
Breed Bestuurlijk Overleg. *Placement of children in Amsterdam primary schools 2020–2021*. (2021). Retrieved July, 2022, from https://bboamsterdam.nl/wat-we-doen/toelatingsbeleid/uitvoering-van-stedelijk-toelatingsbeleid-in-cijfers/
Bruch, E., & Atwell, J. (2015). Agent-based models in empirical social research. *Sociological Methods & Research*, *44*(2), 186–221. https://doi.org/10.1177/0049124113506405
Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2015). What parents want: School preferences and school choice. *The Economic Journal*, *125*(587), 1262–1289.
Burgess, S., Wilson, D., & Lupton, R. (2005). Parallel lives? ethnic segregation in schools and neighbourhoods. *Urban Studies*, *42*(7), 1027–1056. https://doi.org/10.1080/00420980500120741
Butler, T. (2003). Living in the bubble: Gentrification and its' others' in north london. *Urban Studies*, *40*(12), 2469–2486. https://doi.org/10.1080/0042098032000136165
Candipan, J. (2020). Choosing schools in changing places: Examining school enrollment in gentrifying neighborhoods. *Sociology of Education*, 0038040720910128. https://doi.org/10.1177/0038040720910128
Cordini, M., Parma, A., & Ranci, C. (2019). 'White flight'in milan: School segregation as a result of home-to-school mobility. *Urban Studies*, *56*(15), 3216–3233. https://doi.org/10.1177/0042098019836661
Dienst Uitvoering Onderwijs. *Primary school addresses*. (2022). Retrieved May, 2022, from https://duo.nl/open onderwijsdata/primair-onderwijs/scholen-en-adressen/
Doepke, M., Hannusch, A., Kindermann, F., & Tertilt, M. (2022). *The economics of fertility: A new era. Technical report*. National Bureau of Economic Research.
Dronkers, J. (2016). Islamic primary schools in the netherlands. *Journal of School Choice*, *10*(1), 6–21.

Gemeente Amsterdam. (2020). *Schoolwijzer Amsterdam*. Retrieved March 23, 2020, from https://schoolwijzer.amsterdam.nl/nl/po/hoe-werkthet

Hailey, C. A. (2022). Racial preferences for schools: Evidence from an experiment with white, black, latinx, and asian parents and students. *Sociology of Education*, *95*(2), 110–132. https://doi.org/10.1177/00380407211065179

Jongejan, D., & Tijs, J. (2010). Prima, maar niet voor mijn kind': opleidingsniveau en houding ten aanzien van zwarte scholen onder autochtone nederlandse ouders [level of education and atititude toward black schools among native-dutch parents]. *Migrantenstudies*, *26*, 2–20. https://research.rug.nl/en/publications/prima-maar-niet-voor-mijn-kind-opleidingsniveau-en-houding-ten-aa

Karsten, S., Ledoux, G., Roeleveld, J., Felix, C., & Elshof, D. (2003). School choice and ethnic segregation. *Educational Policy*, *17*(4), 452–477. https://doi.org/10.1177/0895904803254963

Ladyman, J., & Wiesner, K. (2020). *What is a complex system?*. Yale University Press.

Lareau, A., & Weininger, E. B. (2003). Cultural capital in educational research: A critical assessment. *Theory and Society*, *32*(5), 567–606. https://doi.org/10.1023/B:RYSO.0000004951.04408.b0

Logan, J. R., Oakley, D., & Stowell, J. (2008). School segregation in metropolitan regions, 1970–2000: The impacts of policy choices on public education. *American Journal of Sociology*, *113*(6), 1611–1644. https://doi.org/10.1086/587150

Maes, M. (2021). *Research handbook on analytical sociology* (pp. 100–134). Edward Elgar Publishing.

Manzo, G. (2013). Educational choices and social interactions: A formal model and a computational test. *In Class and stratification analysis* (Vol. 30, pp. 47–100). Emerald Group Publishing Limited.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Merry, M. S., & Boterman, W. (2022c). Educational inequality and state–sponsored elite education: the case of the Dutch gymnasium. Comparative Education, 56(4), 522–546.

Muncipality of Amsterdam. *Diversity monitory 2015–2016 and 2016–2017*. (2017). Retrieved July, 2022, from https://onderzoek.amsterdam.nl/publicatie/monitor-diversiteit-in-het-basisonderwijs-2015-2016-en-2016-2017

Municpality of Amsterdam. *Basisbestand Gebieden Amsterdam (BBGA)*. (2022a). Retrieved May , 2022, from https://onderzoek.amsterdam.nl/dataset/basisbestand-gebieden-amsterdam-bbga

Municpality of Amsterdam. (2022b). *Maps Amsterdam*. Retrieved May, 2022, from https://maps.amsterdam.nl/opengeodata

Mutgan, S. (2021). *Free to choose? Studies of opportunity constraints and the dynamics of school segregation* [PhD thesis]. Linköping University Electronic Press.

Onderwijsinspectie. *Segregation in Dutch primary schools*. Retrieved June, 2022, from https://www.onderwijsinspectie.nl/onderwerpen/staat-van-het-onderwijs/documenten/rapporten/2020/04/22/technisch-rapport-svho-2020-segregatie

Oosterbeek, H., Sóvágó, S., & van der Klaauw, B. (2021). Preference heterogeneity and school segregation. *Journal of Public Economics*, *197*, 104400. https://doi.org/10.1016/j.jpubeco.2021.104400

Orfield, G. (2001). Schools more separate: Consequences of a decade of resegregation. ERIC. https://escholarship.org/uc/item/5cx838jx#main

Page, S. E. (2015). What sociologists should know about complexity. *Annual Review of Sociology*, *41*, 21–41. https://doi.org/10.1146/annurev-soc-073014-112230

Perry, L. B., Rowe, E., & Lubienski, C. (2022). School segregation: Theoretical insights and future directions. *Comparative Education*, *58*(1), 1–15. https://doi.org/10.1080/03050068.2021.2021066

Reardon, S. F., & Owens, A. (2014). 60 years after brown: Trends and consequences of school segregation. *Annual Review of Sociology*, *40*, 199–218. https://doi.org/10.1146/annurev-soc-071913-043152

Renzulli, L. A. (2005). Organizational environments and the emergence of charter schools in the united states. *Sociology of Education*, *78*(1), 1–26. https://doi.org/10.1177/003804070507800101

Renzulli, L. A., & Evans, L. (2005). School choice, charter schools, and white flight. *Social Problems*, *52*(3), 398–418.

Ruijs, N., & Oosterbeek, H. (2019). School choice in amsterdam: Which schools are chosen when school choice is free? *Education Finance and Policy*, *14*(1), 1–30. https://doi.org/10.1162/edfp_a_00237

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*(2), 143–186. https://doi.org/10.1080/0022250X.1971.9989794

Spaiser, V., Hedström, P., Ranganathan, S., Jansson, K., Nordvik, M. K., & Sumpter, D. J. (2018). Identifying complex dynamics in social systems: A new methodological approach applied to study school segregation. *Sociological Methods & Research*, *47*(2), 103–135. https://doi.org/10.1177/0049124116626174

Thurner, S., Hanel, R., & Klimek, P. (2018). *Introduction to the theory of complex systems*. Oxford University Press.

Ukanwa, K., Jones, A. C., & Turner, B. L., Jr. (2022). School choice increases racial segregation even when parents do not care about race. *Proceedings of the National Academy of Sciences*, *119*(35), e2117979119.

Van Noord, J., Spruyt, B., Kuppens, T., & Spears, R. (2019). Education-based status in comparative perspective: The legitimization of education as a basis for social stratification. *Social Forces*, *98*(2), 649–676. https://doi.org/10.1093/sf/soz012

Wilson, D., & Bridge, G. (2019). School choice and the city: Geographies of allocation and segregation. *Urban Studies*, *56*(15), 3198–3215.

Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: Erg models of a friendship network documented on facebook. *American Journal of Sociology*, *116*(2), 583–642. https://doi.org/10.1086/653658

Zwier, D., Geven, S., Bol, T., & van de Werfhorst, H. G. (2022). Let's stick together: Peer effects in secondary school choice and variations by student socio-economic background. *European Sociological Review*. https://doi.org/10.1093/esr/jcac033

## Appendix A

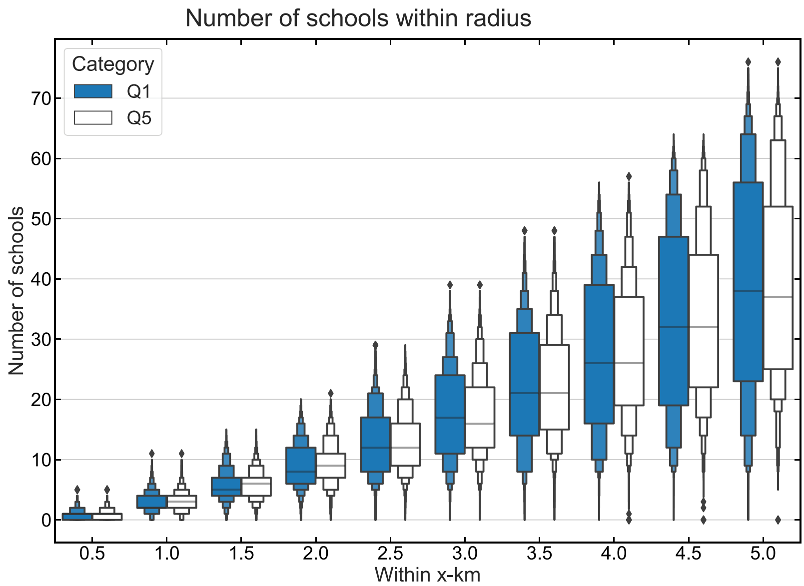


**Figure A1.** Number of schools within a x-km Euclidean distance per household per group.

**Algorithm 1**: Simulation

```
Initialization: approximate residential locations Initialization: allocate agents to random schools
for school_step do
    agents_to_move = random_sample(agents, fraction_to_move)

    foreach agent in agents_to_move do

        if minimum capacity of current school is not reached then
            utilities = calculate utilities(agent, schools)
            normalized utilities = normalize utilities(utilities, temperature)
            ranked schools = rank schools(normalized utilities)

            foreach school in ranked schools do
                if has space(school) then
                    school.add agent(agent)
                    break
                end
            end
        end
    end

    if school process converged then
        break
    end
end
```

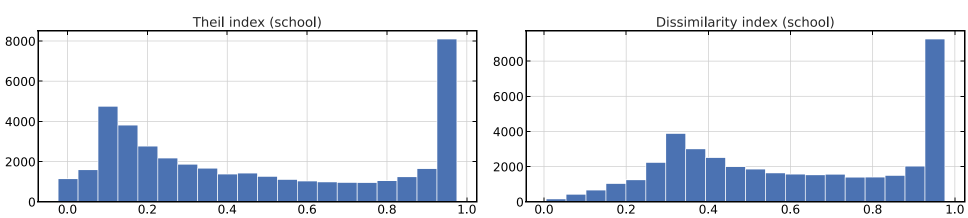


**Figure A2.** Distribution of the two measures of school segregation used as output for the global sensitivity analysis.

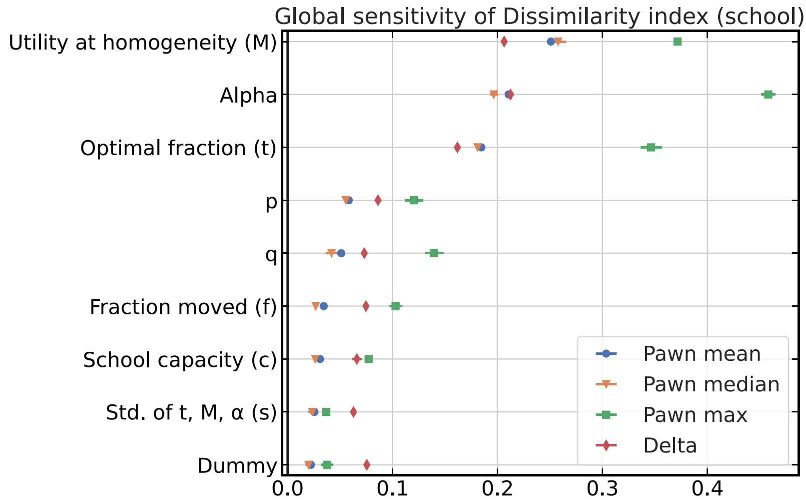


**Figure A3.** Global sensitivity analysis using a Sobol sequence with 2048 samples. The moment independent PAWN mean, median and max indices are plotted, together with the Delta measure for the Dissimilarity index. Confidence intervals are based on 100 bootstrapped samples.

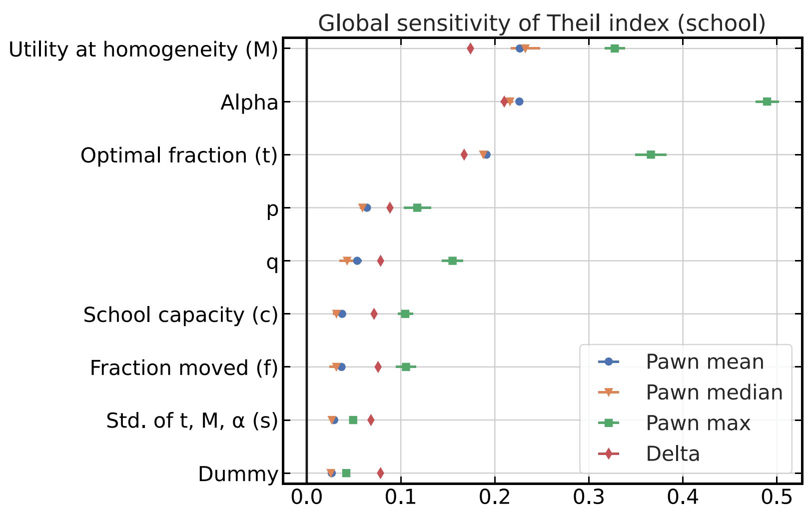


**Figure A4.** Global sensitivity analysis using a Saltelli sequence with 2048 samples. The moment independent PAWN mean, median and max indices are plotted, together with the Delta measure for Theil's measure of segregation. Intervals are based on 100 bootstrapped samples.

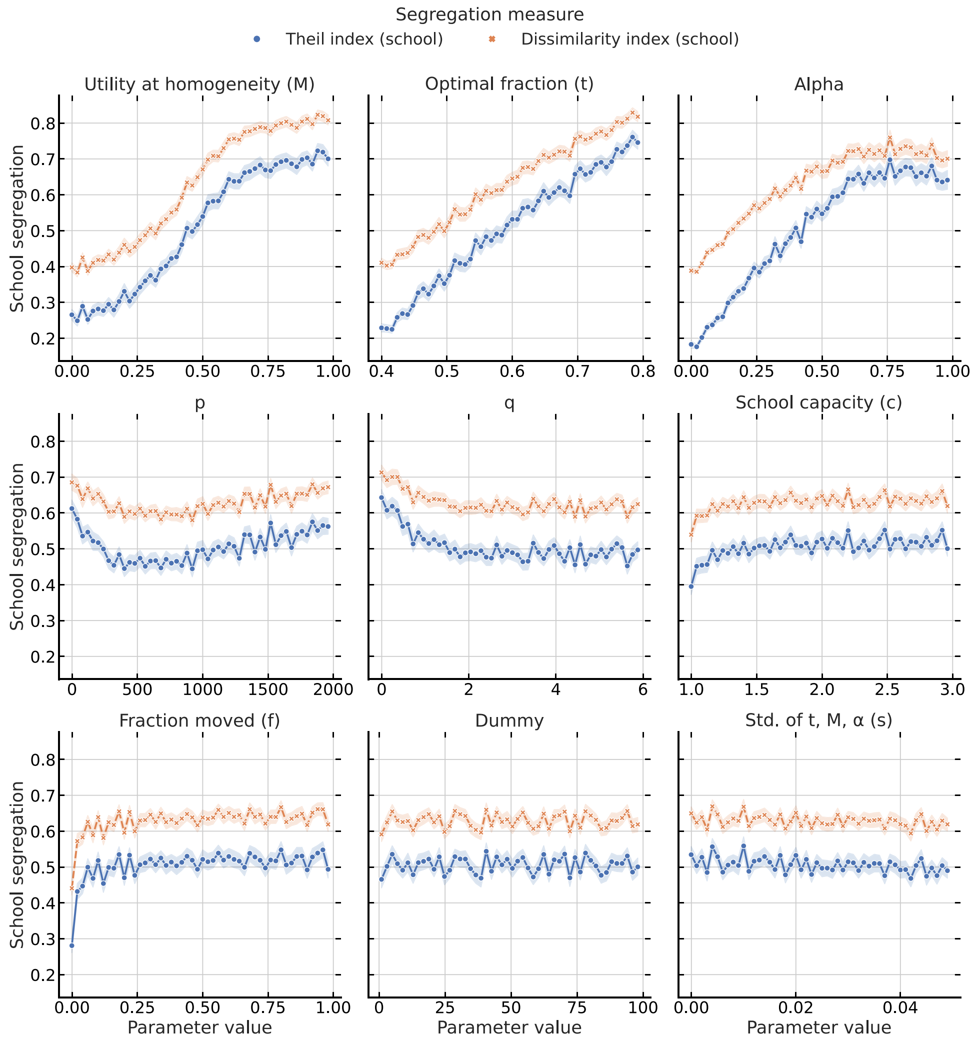


**Figure A5.** Average effects of varying parameters using the Saltelli sequence of the GSA. The four measures of school segregation are plotted. Observations are binned using 50 bins, where each of the bins represents the average level of school segregation observed. The parameter of interest is "fixed" at the bin value, while the others are allowed to vary. The bands represent the 5th and 95th percentiles of the observations within a bin.

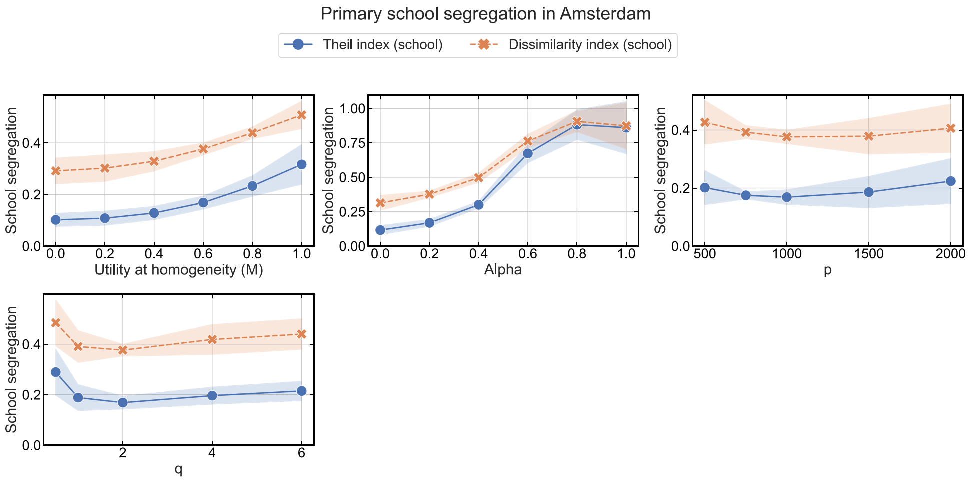


**Figure A6.** Varying one parameter at a time when using relative optimal fractions (0.62 and 0.38) and two measures of school segregation. The plotted level of segregation is the average over 10 model runs, the bands represent the mean plus and minus the standard deviation respectively. Nominal values of the parameters that are not varied can be found in Table 1.