Technische Universiteit Delft

Modelleren 2A

Segregation

C. Barendrecht
G. Jin
S. Moerman
N. Snijder

Under supervision of:
Dr. P. Cirillo

This page is left intentionally blank.

Abstract

We built an extended version of 'Schelling Tipping Model' to study segregation, in which n individuals of m different types are placed randomly on an l x b board and each individual will move to a new location if less than a q-fraction of her neighbours share his/her type and if he/she is able to find a place that better meets this requirement. Other extensions are random displacement of the individuals and the ability to switch types.

In this report, we mainly studied the standard case in which $n = 40$, $m = 2$, $l = b = 8$ and $q = \frac{1}{3}$ and we looked at the effect of type switching on the system. Our goal was to investigate whether such system will reach an equilibrium, and if it does, in how many generations on average (see formal definition at Introduction) and how this average (denoted as Y) is distributed. Also, we were interested in the fraction of the individual that lives in a homogenous environment (with all neighbours of his/her types) and the average number of generations it takes to reach a certain fraction of homogeneity.

We found that equilibrium is not always reached in the standard case (which also implies the more general case), but it is practically always reached. We saw that average of Y is a non-decreasing function of q, and when type switching is allowed, it requires significantly less generations to reach equilibrium. We used the chi-squared test to conclude that Y is not Poisson, binomially or negative binomially distributed. This is also the case when type switching is allowed. With the Kolomogorov-Smirnov test, we concluded that Y is distributed differently when q is varied.

For various values of q, we tested whether 60%, 80% and full (100%) segregation occured. As expected, this strongly depends on q. The corresponding histograms, did not have any common distribution. We found that it took on average $3.09, 3.83$ and 5.46 generations, in order to reach respectively 60% segregation, 80% segregation and 100% segregation.

When switching types is allowed, the average number of generations until equilibrium is reached, will drop, whereas the happiness at equilibrium will strongly increase.

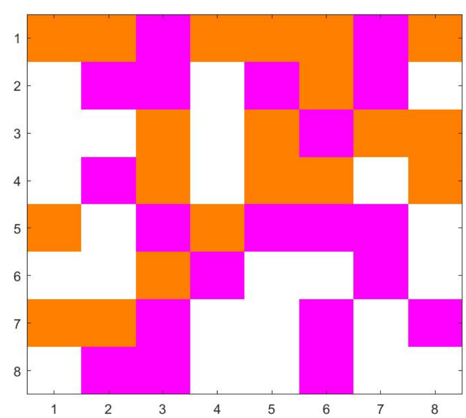
Contents

1 Introduction

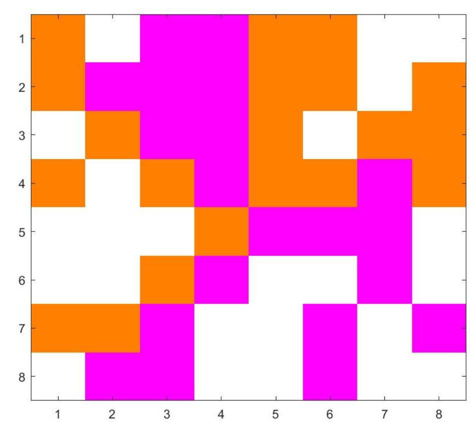
In 1978, Thomas C. Schelling developed his tipping model by placing pennies and dimes on a chess board and moved them according to various rules. By viewing the pennies and dimes as two types of people, the rule of moving as a preference for the individuals, and the chess board as a city, he soon discovered that segregation takes place on the board, even when the preference of an individual is very subtle.

Based on this idea, we created a basic model which consists of an 8×8 board with 40 individuals that are divided evenly into two types. The individuals are moved according to their 'Happiness' in the current place. For the basic model, an individual is considered happy if $\frac{1}{3}$ of his/her second order neighbours (a person has at most 8 neighbours) shares his type. Otherwise, an individual is considered unhappy and will move to the nearest place where his/her happiness is strictly higher. This will be referred to as the 'Happiness Rule'.

An example of the basic model is given below. It is clear that not every individual in figure 1a is happy. After segregation the individuals have relocated themselves as in Figure 1b, there every individual is happy.



(a) Situation before segragation



(b) Situation after segregation

Figure 1: An example of segregation in our basic model: type 1 is orange and type 2 is pink

Thereafter, we extended the basic model by changing the parameters such as the size of the population, the boardsize and the number of types. We also included an option for random displacement: an individual which is not happy will be moved to an empty location chosen at random (with homogenous probability).

For both the basic and the extended model, we ran 500 simulations multiple times and investigated to what extent certain parameters affected the segregation pattern. In order to formulate our research goals precisely, the following definitions are introduced:
1. **Generation**: A population is said to have entered the next generation if the happiness of every individual has been checked once.
2. **Equilibrium after g generations**: The population is said to have reached an equilibrium after g generations if no individual has moved during the $g+1$-th generation.
3. **Homogeneity**: A person is said to live homogenous if all of his/her neighbours share his type.
4. **Segregation time at $p\%$**: The segregation time at $p\%$ is defined as the number of generations after which $p\%$ of the population is homogenous.

For the following sections, when it is not specified whether some parameters has been changed, one may always assume that these parameters are of the standard models.

For the analysis, we focused on the following main questions:

1. Does the population always reach an equilibrium? How many generations on average does it take to reach an equilibrium? How does the Happiness Rule affect the number of generations until equilibrium? What's the probability distribution of the number of generations to reach an equilibrium?

2. What is the average segregation time at 60%, 80% and 100% segregation. How is this segregation time distributed with a Happiness Rule of 1 and how does the Happiness Rule affect the average segregation time?

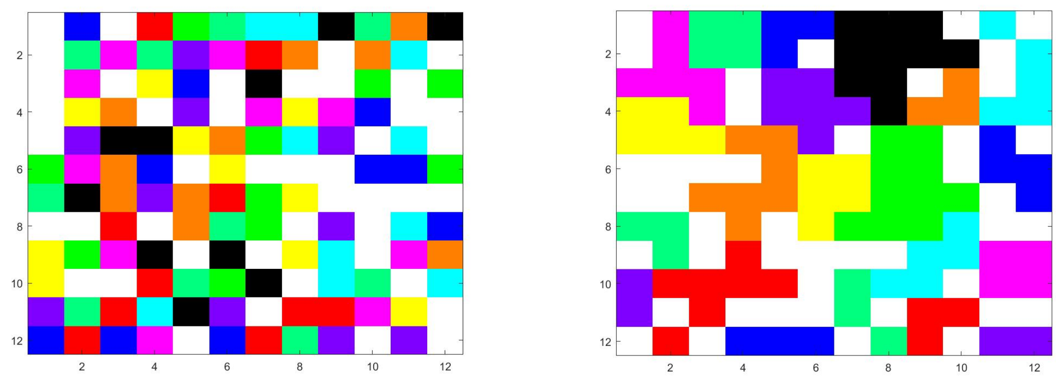
3. If an individual is able to switch to another type with a probability that depends on the types of his/her neighbours, how does this affect the average number of generations (until equilibrium is reached) and to what extent does it impact the corresponding distribution and the other questions posed above?

2 Model description and statistical methods

2.1 Our Model

As mentioned in the introduction, we will mostly be investigating the basic model: segregation of 20 individuals of type 1 and 20 individuals of type 2 on an 8×8 board. In this model an individual is considered unhappy if less than $\frac{1}{3}$ of his/her neighbours shares his/her type. If an individual is unhappy, this individual will move to the nearest place where he/she would be happier. Any person without neighbours is by definition considered unhappy (so his/her happiness is 0), since people prefer living with others over living alone. Furthermore, the basic model uses the second order neighbourhood; the maximum number of neighbours a person can have is 8. The model however can be changed by preference in several ways.

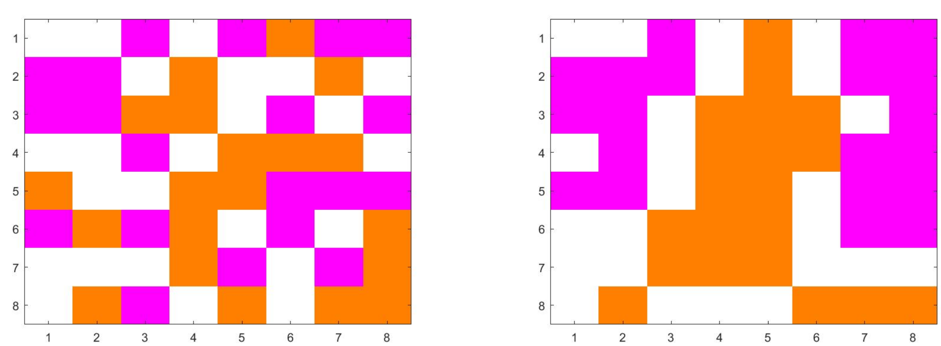
First of all, the board size can be varied. If no other parameters are altered, this will only imply that the board becomes fuller or emptier. Secondly, we can change the number of types on the board. For instance, one can consider a 12×12 board with 10 types of individuals and 10 individuals per type: see Figure 2. As one can see, far more individuals have been moved to another place on the board and segregation takes place.



(a) Situation before segregation (b) Situation after segregation

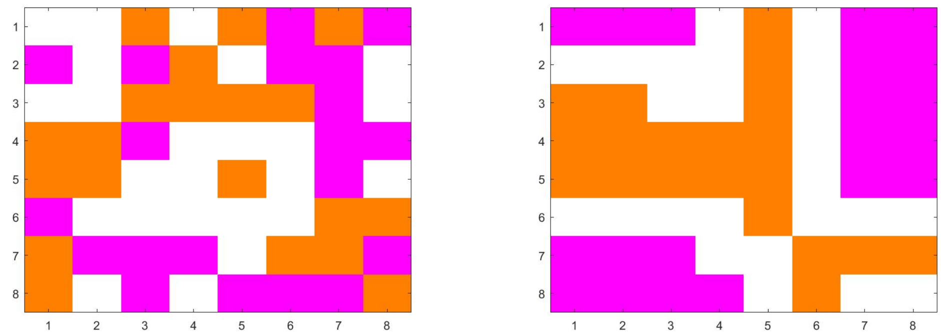
Figure 2: An example of segregation on a 12×12 board with 100 individual of 10 different types

Subsesquently, we can vary our 'Happiness Rule': the minimal fraction of neighbours that has to share an individuals type, in order to be considered happy. The expectation is that the time to reach equilibrium increases as the Happiness Rule is strengthened. This makes sense as every individual needs to be surrounded by relatively more neighbours of his/her own type. In consequence, larger homogenous groups will be formed after segregation with a higher Happiness Rule. In Figures 3 and 4 the effect of a Happiness Rule of respectively $\frac{2}{3}$ and 1 on the board after segregation is illustrated.



(a) Situation before segregation (b) Situation after segregation

Figure 3: An example of segregation in a model with Happiness Rule $\frac{2}{3}$



(a) Situation before segregation (b) Situation after segregation

Figure 4: An example of segregation in a model with Happiness Rule 1

In Figure 4b it is illustrated that every individual lives homogenous if a Happiness Rule of 1 is applied. We will elaborate on the effect of the Happiness Rule on the model in sections 4 and 5.

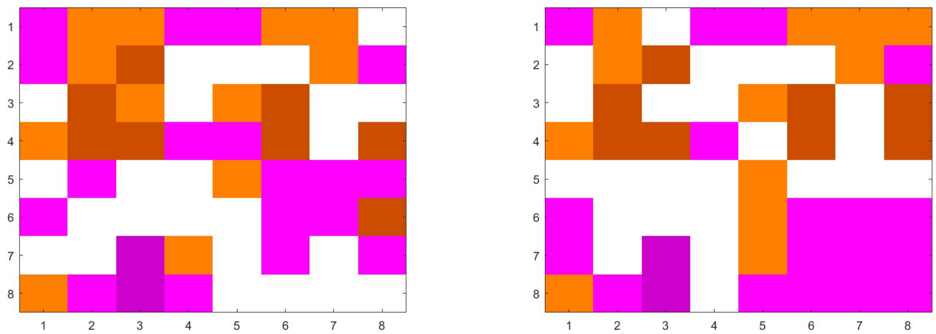
In addition, the type of neighbourhood can be changed. The neighbourhood is by default a second order neighbourhood, in which both the adjacent and diagonal people are considered neighbours. There exists only one smaller type of neighbourhood: the first order neighbourhood. Only adjacent neighbours are accounted for this type of neighbourhood; a person has up to four neighbours. The way the different types of neighbourhood are defined is illustrated in Figure 5. Here, 0 represents the individual himself and every other place displays which order of of neighbourhood it belongs to. We will not examine the effect of the type of neighbourhood on the segregation in our model in this report. However, some interesting graphs are included in the appendix.

5	4	3	4	5
4	2	1	2	4
3	1	0	1	3
4	2	1	2	4
5	4	3	4	5
8	7	6	7	8

Figure 5: Neighbourhood order

Moreover, we included the possibility for individuals to move to a random place on the board if he/she is not happy, rather than moving to the closest position. In the basic model he/she will move to the nearest place where he/she will become happier. With random displacement, we expect that it takes longer to reach an equilibrium. As above, we will not discuss the impact of the random displacement in our report, but we added some results in our appendix.

Furthermore, we expanded the model with the so-called 'criminals'. One can choose to have c out of the n individuals behave as if they were a criminal. Any person who is not a criminal himself, does not like criminals in his neighbourhood, and thus his/her happiness will be 0, overruling any positive effect of friendly neighbours, if a criminal lives in his/her neighbourhood. As a result, all criminals should end up living together. In Figure 6, one can see an example of the basic model with 10 criminals among the 40 individuals. It is visible that our expectation is not always the outcome. In this case, equilibrium is reached, because every unhappy individual cannot move to another place, since every empty location is in the neighbourhood of a criminal. So the unhappy individuals will not become happier on these places.



(a) Situation before segregation

(b) Situation after segregation

Figure 6: An example of segregation in the basic model with a total number of 10 criminals (the slightly darker places: i.e. the individuals on (3,8) and (3,2))

Last, we made it possible for individuals to switch to another type. In our model there are two versions of switching. The first version is only possible if there are only two different types present on the board. In this version of switching, every person has a chance p of switching to the other type before he moves. In the other version, every unhappy individual will switch type depending on the type of his/her neighbours. This is further elaborated on in section 7. Individuals without neighbours will not switch types, since they have no neighbours who extert their influence on them. We will investigate the effect of the second version of switching further in section 7.

2.2 The applied statistical test

The two sample t-test:
For the following sections, we will mostly use the two sample t-test (ttest2) from Matlab for comparing the means of two distributions. We consider two means to be comparable when they are equal. Otherwise, they are not comparable.

This t-test tests the null-hypothesis $H_0 : \mu = \nu$ with μ, ν respectively the mean of the distribution of $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_n)$, against $H_1 : \mu \neq \nu$. It uses the test statistics:

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{m} - \frac{S_Y^2}{n}}}$$

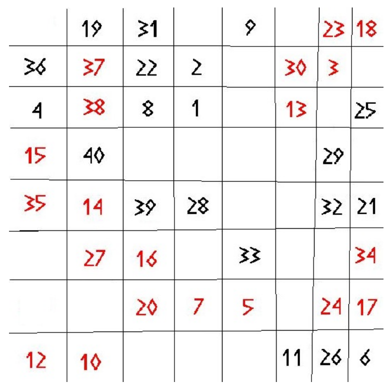
This test assumes that the mean difference $\mu - \nu$ is normally distributed, and since our data are from the discrete random variables, this assumption would not make sense. However, according to the central limit theorems, we may make this assumption only for a sufficiently large sample size. Although exact calculations are required to determine the sample size, we simply chose a sample size of 5000. In section 7.1, we even chose a sample size of 50000, due to the asymmetric distribution of our data. Also, for simplicity, we usually took $m = n$. Furthermore, we have adjusted the test, such that it does not assume the two data sets have an equal variance for the data distribution (this is achieved by setting 'Vartype' as unequal in Matlab).

Other tests:
For testing the goodness of fit of our hypothesized distribution in sections 4 and 7, we have used the two-sample kolomogorov-smirnov test and the chi-squared test in Matlab. More information about these tests can be found on Mathworks.

3 Equilibrium

As mentioned in the introduction, a board has reached quilibrium after g generations if no person has moved in the $g+1$-th generation. In this parapgraph a portion of the first research question will be answered. Namely, will a board always reach an equilibrium in finite time? Based on intuition, this is expected to be true. An individual who moves, will move to a place where relatively more neighbours share his/her type, intuitively, this should imply that the average happiness of the population increases during each generation. In the basic model however, this is not always true. Even though equilibrium is nearly always reached in finite time, there exists cases in which it allows a periodic solution. An example of a scenerio in which this occurs is given below.

(a) The start situation. Individual 37 will move.



(b) Individual 37 is moved to the nearest location that better meets his/her desires.

(c) Accordingly, 38 will move next and is moved to the nearest location that better meets his/her desires.

(d) A generation has passed, 37 is next to move and will do so accordingly. The roles of 37 and 38 have interchanged, which leads to a periodic solution.

Figure 7: A possible configuration of the standard board in which the equilibrium will never be reached. The numbers indicate the turn order of an individuals. The two different colour indicates the two types. The equilibrium is not reached due to the periodic movements of individuals 37 and 38. The movements of inviduals 37 and 38 are tracked and shown in subfigure a to d in a chronological order.

In Figure 7, a possible configuration in which equilibrium will not be reached is shown. The numbers represent the turn order (1 is selected first followed by 2, etc.). Red and black indicate the two different types. It is easily observed that that individuals 1 to 36 satisfy the happiness condition. 37 however does not. During 37's turn, we note that his/her happiness equals 0. The closest empty location has a happiness of $\frac{1}{7}$. And thus 37 will move to the given location, which is indicated in Figure 7b. Next in line to move is

38. Individual 38 has a happiness of $\frac{2}{7}$, which is less than the required $\frac{1}{3}$ and will thus have to move. The closest spot with greater happiness is the nearby corner spot with happiness $\frac{1}{3}$. So 38 will move to that location, which is shown in Figure 7c.

The indviduals 1 to 36 will remain pleased and will thus remain in place. Thus 37 is the first candidate to be moved. 37 has a happiness $\frac{1}{7} < \frac{1}{3}$. The closest empty spot has happiness $\frac{1}{6} > \frac{1}{7}$, thus 37 will move to this location. Since this location was te former spot of 38, we conclude that 37 and 38 have swapped positions, as is shown in Figure 7d.
Since 37 and 38 are of the same type, these 3 moves will continue periodically and thus we find a periodic solution.

Now what went wrong?
After the first move, both 37 and 38 will gain happiness, but as 37 moves away, 38 will lose all of his/her happiness. An endless loop is formed. Note that, the same scenario can be constructed for larger boards. Since we can implement this exact board in a larger board and fill out the remaining locations to satisfy an equilibrium.

Remark:
Although equilibrium is not always reached, we have observed that it is practically always reached. After performing 10000 simulations, there were only on average 2 or 3 cases in which equilibrium is not reached within 10000 generations. This makes the other research questions for which equilibrium is assumed still worth studying.

4 Average number of generations until equilibrium

There are numerous parameters that may affect the average number of generations until equilibrium is reached. We chose to focus on the effect of the Happiness Rule (HR) on the average number of generations. For each HR ranging from 0 to 1 with increment 0.01, we ran 500 simulations and calculated the average, maximal and minimal generations it took to reach equilibrium. The results are shown in figure 8.

From Figure 8, we conclude that the average number of generations (blue graph) is increasing with the Happiness Rule, which is to be expected since a higher HR implies a higher need for neighbours of the same type, and thus a lower probability that a selected individual is happy, making it more likely that he/she will move. We also note that starting at a HR of aproximately 0.7, the average number of generations appears to be nearly constant. This can be explained noting that the required number of neighbours will practically be the same. For example, if the HR were 0.8, a person with 3, 5 or 8 neighbours would require respectively 3, 5 or at least 7 of the same type. This requirement hardly changes if a HR of 0.9 is applied. The same argument explains why the average number of generations is constant at very low HR or if we raise the HR with a sufficiently small amount like 0.01.

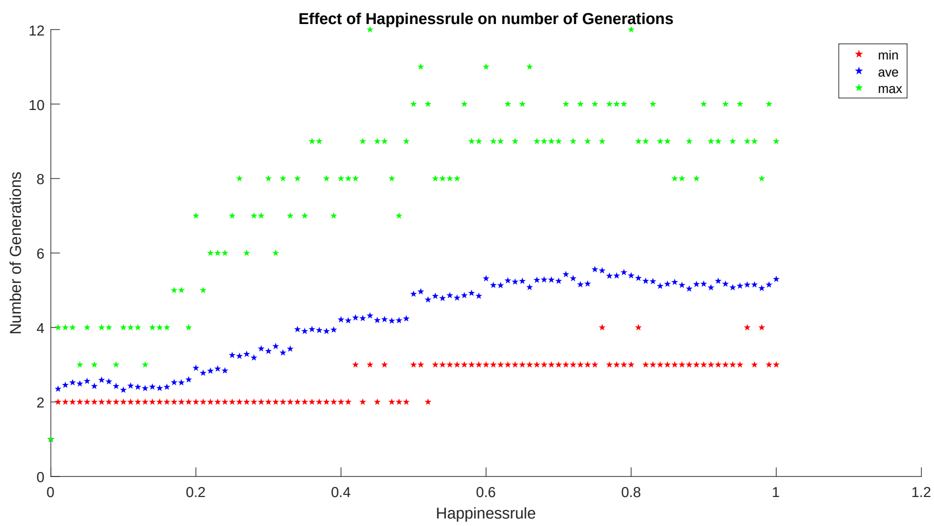
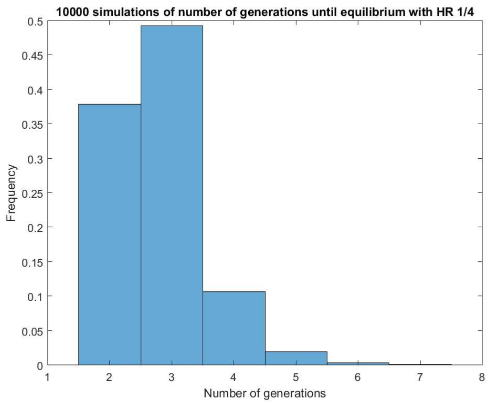
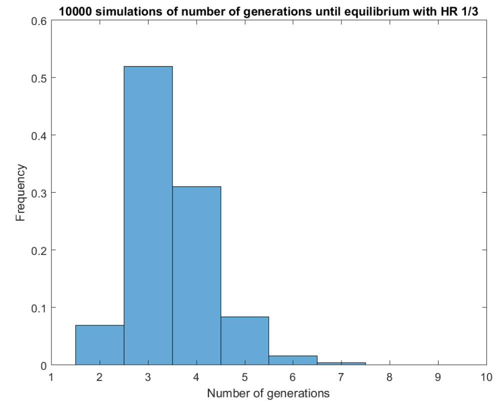


Figure 8: Effect of the happiness rule on the number of generations until reaching equilibrium. The green graph represents the maximal number of generations, the blue graph the average number of generations and the red graph displays the minimal number of generations

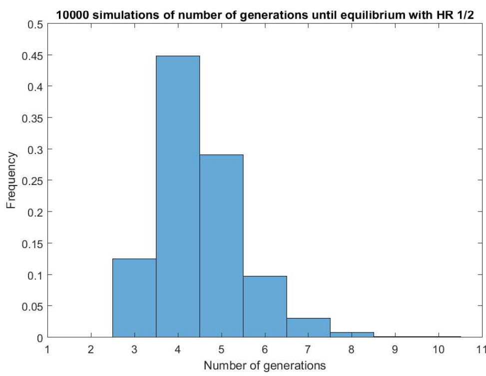
It is also interesting to investigate how the random variables Y_j, which we denote as the number of generations it takes to reach the equilibrium under HR j, is distributed, and how their distributions are effected by the HR. For this purpose, we ran 10000 simulations for with Happiness Rules of $\frac{1}{4}, \frac{1}{3}, \frac{1}{2}$ and 1 and plotted several histograms (Figure 9). We chose bin size 1 for each histogram given that Y_j is a random variable that only takes integer values. Therefore, a histogram with a bin of 0.5 will not give any additional information, but there would be a severe loss of information if bin size 2 is chosen.



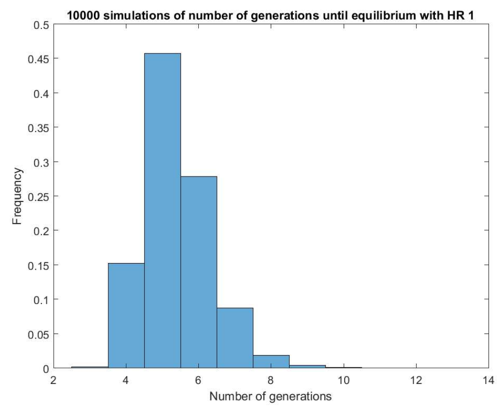
(a) 10000 simulations of number of generations until equilibrium with HR 1/4



(b) 10000 simulations of number of generations until equilibrium with HR 1/3



(c) 10000 simulations of number of generations until equilibrium with HR 1/2



(d) 10000 simulations of number of generations until equilibrium with HR 1

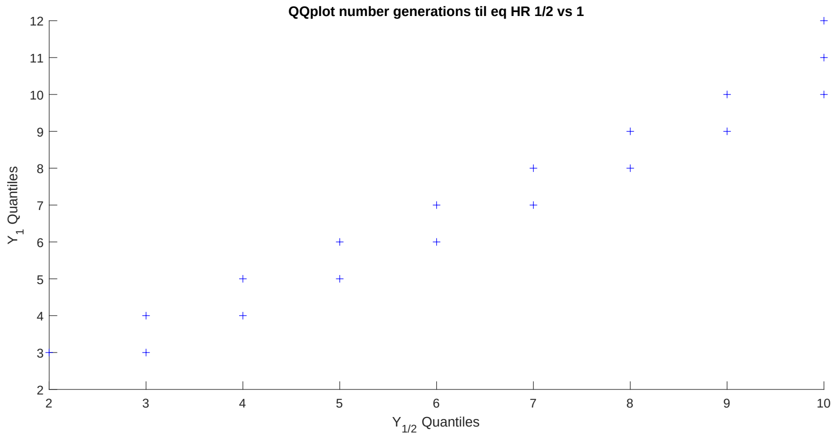
Figure 9: The probability distribution of Y (the number of generations until reaching the equilibrium) is approximated with a histogram of bin size 1. For each HR ($\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$ and 1), 10000 simulations were ran.

Looking at the histograms, all Y_j's ($j \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\}$) might come from the same distribution, but with different parameters, since their sample means are obviously different.

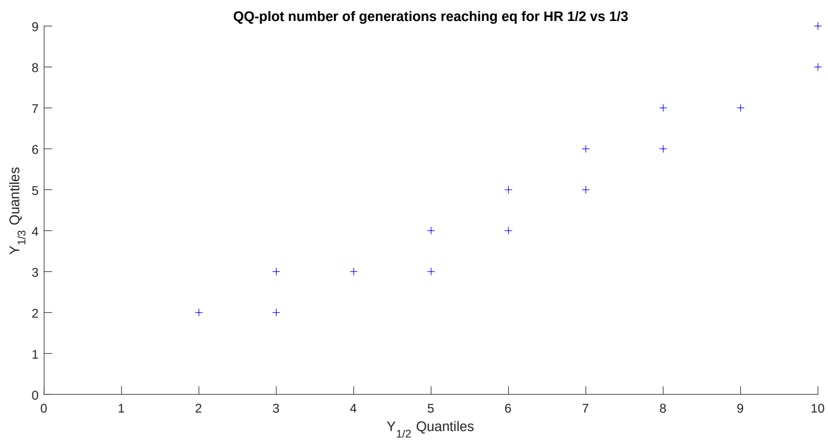
For simplicity, we chose to mainly investigate the distribution of $Y_{1/3}$, since other Y_j's might follow the same family distribution as suggested by the histograms. We suspected that $Y_{1/3}$ might be Poisson distributed. So with this assumption, we first used the 'fitdist'-function of Matlab to calculate the maximum likelihood estimator $\hat{\lambda}$ for the real Poisson parameter λ. Then, we performed the chi-squared test using the chi2gof function in Matlab, with null-hypothesis $H_0 : F_{1/3} = F_{P,\hat{\lambda}}$, for which F is the unknown cumulative distribution function (CDF) from the data and $F_{P,\hat{\lambda}}$, the CDF from Poisson($\hat{\lambda}$). The alternative hypothesis is $F_{1/3} \neq F_{P,\hat{\lambda}}$. Furthermore, we chose 8 bins for the chi-squared test (otherwise the test is not accurate due to low expected counts in some bins). The test rejected the null-hypothesis and with significance level $\alpha = 0.05$.

With the same method, we also tested whether $Y_{1/3}$ could be binomially or negative binomially distributed. But once again, they have been rejected by the chi-squared test. So we can only conclude that $Y_{1/3}$ is not Poisson, binomially or negative binomially distributed.
We have also performed this statistical analysis on other Y_j's (for instance Y_1), and obtained the same result as for $Y_{1/3}$.

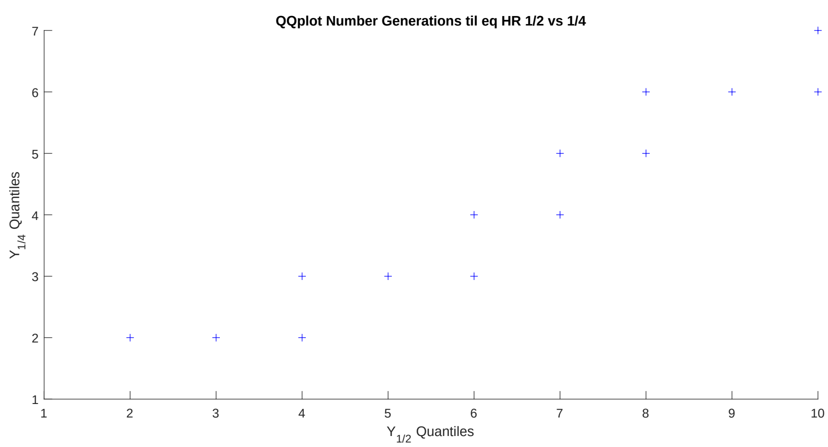
In order to make a better comparison of the distribution between $Y_{1/4}, Y_{1/3}, Y_{1/2}$ and Y_1, several QQ-plots were made. Thich is shown in Figure 10.



(a) QQ-plot of Y_1 vs $Y_{1/2}$



(b) QQ-plot of $Y_{1/3}$ vs $Y_{1/2}$



(c) QQ-plot of $Y_{1/4}$ vs $Y_{1/2}$

Figure 10: QQ-plot of Y_1 vs $Y_{1/2}$, $Y_{1/3}$ vs $Y_{1/2}$ and $Y_{1/4}$ vs $Y_{1/2}$

As expected, the QQ-plots have a stairwise pattern. This is because the data came from a discrete random variable. The plots appear to be linear, especially the plot of Y_1 vs $Y_{1/2}$. The plot indicates that these 4 Y_j's might belong to the location-scale family of the same distribution.

With the two-sample kolomogorov-smirnov test (kstest2) of significance level $\alpha = 0.05$ in Matlab, which has null-hypothesis $H_0 : F(X) = G(Y)$ for F, G the CDF of X and Y, it can be concluded that these Y_j's do not come from the same distribution with the same parameters. Unfortunately, we were not able to find a test that tests whether these Y_j's comes from the same family distribution (i.e. same distribution but different parameter values).

5 Is every individual happy after equilibrium?

As shown in section 3, equilibrium is not always reached within 10000 generations, stronger even, equilibrium is occasionally never reached at all. But when it does, it is natural to question whether every individual is happy and how happy they are. To answer this question, we ran 500 simulations for each HR ranging from 0 to 1 with increments of 0.01 and we calculated the average, maximal and minimal happiness of the population for each happiness rule. We ran the simulations om the standard board and the results are shown in Figure 11.

From Figure 11, we observe that the minimal happiness is almost always equal to or greater than the happiness rule, which means that every individual is in fact happy in the equilibrium. We also see that maximal happiness is always 1, which means there is always at least one invidividual who lives homogenous. Furthermore, we see that the average happiness is always closer to the maximal happiness than to the minimal happiness. At a HR of approximately 0.8, we see that every individual has happiness 1. This implies that happiness 0.8 guarantees nearly full segregation. Complete segregation can only be assured at a HR greater than 0.875, Namely, assume a person does not live in a homogenous neighbourhood, then at least 1 out of at most 8 neighbours does not share his/her type. In other words, the happiness of this person is at most equal to $1 - \frac{1}{8} = 0.875$.

It is not very surprising that every individual is happy after the equilibrium, since it would otherwise mean that one individual is unable to find a place that better meets his/her desire, which seems rather unlikely on a board with 24 empty spaces. Another remarkable result is that Figure 11 shows that the minimal happiness is higher than the HR and that the average happiness is nearly 1 if the happiness rule higher than 0.8. At first glance, this might seem impossible because the heavy requirement of being happy, but on the other hand, it shows that the strong 'need' for segregation actually leads to greater happiness.

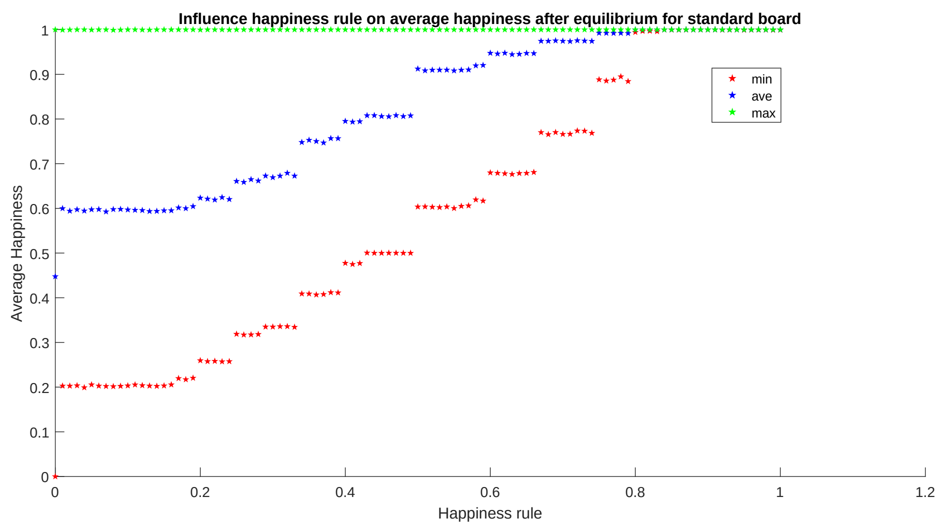


Figure 11: Effect of happiness rule on the happiness of the individuals. The green graph shows the maximal happiness, the blue graph shows the average happiness and the red graph shows the minimal happiness

To illustrate that the variance of every above calculated average happiness is very low, we selected Happiness Rule 1/3 as an example and made a histogram in Figure 12, showing that the probability that any individual is happy is practically 1.

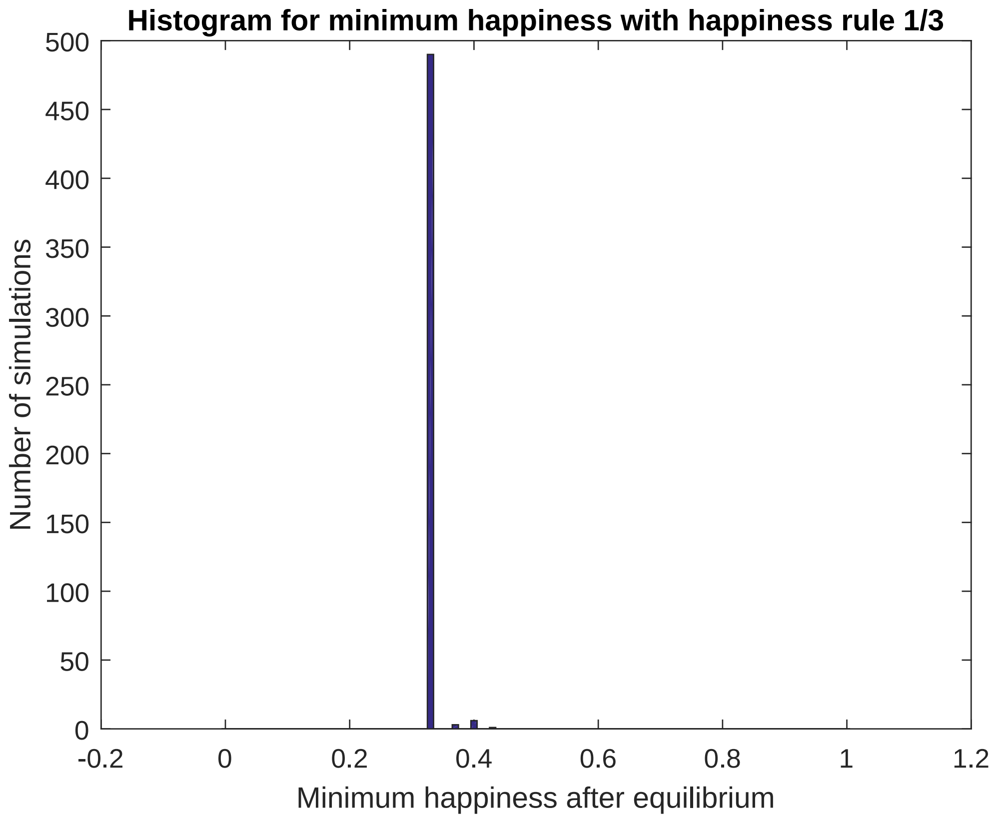


Figure 12: Histogram for minimum happiness after 500 simulations with HR=1/3

6 Average segregation time

As mentioned in the introduction, the segregation time at $p\%$ is defined as the number of generations until at least $p\%$ of the population on a board lives in homogenous groups. Where person i is said to live homogenous if for any neighbour j of i, we have $\text{Type}(j) = \text{Type}(i)$. This gives immediate rise to questions concerning the relation between the choice of p and the average segregation time at $p\%$. Furthermore, it is unclear if segregation at $p\%$ is guaranteed before a board reaches an equilibrium and what the effect is of the happiness boundary on the existence of a segregation time.

To study any of the given questions, we will first have to formalise our choices of board as well as the questions proposed.

6.1 Formalisations

Prior to starting any test or properly formalising our research questions however, we note that segregation at $p\%$ is not guaranteed to take place. For example, if we consider $p = 100$ on the standard board with Happiness Rule 1/3, we will nearly never have total segregation before the board reaches an equilibrium. Therefore one might instead consider the average fraction of segregation at equilibrium, for any given happiness fraction.

Furthermore, the average segregation time as function of the segregation fraction should theoretically be a strictly increasing function since for any given board we have:

$$\begin{aligned}&n\% \text{ lives in homogenous groups after } k \text{ generations } \Rightarrow\\ &m\% \text{ lives in homogenous groups after } k \text{ generations, for any } 0 \leq m \leq n\end{aligned}$$

Having noted these facts, we can now properly formalise the research questions.
The following questions are proposed:

1. What is the relation between the average segregation time and p.
2. What is the average segregated fraction of the population after a board reaches equilibrium for given choices of happiness.
3. For any Happiness Rule ranging from 0 to 1, how often do we reach at least 60%, at least 80% or even full (100%) segregation?
4. For 60% and 80% segregation, what distribution do we get for the segregation time, for a HR of 1? Are these comparable?

To establish results regarding these questions, we consider different setups in testings. We will be testing two different boards. The first board to be analysed is the standard board. The second board is a larger "4-Type" board. The details are specified in Table 1:

Table 1: The specifcations of the two considered boards

	Standard Board	4-Type Board
Number of types	2	4
Length	8	10
Width	8	10
Happiness	1	1
Population per type	20	16

The 4-Type board is constructed to maintain the same ratio of inhabited and uninhabited locations as the standard board. The choice of happiness on these boards is 1 unlike the usual $\frac{1}{3}$. This guarantees that for any $p \leq 100$, segregation at $p\%$ takes place prior to the board reaching an equilibrium. To observe the average segregation time $p\%$ for any p, 500 simulations were ran per board and averaged out in order to give an approximation for the average segregation time at $p\%$. Likewise the average segregated fraction will be estimated by the average of the segregated fraction of an equilibrium from 500 simulations with given happiness q.

6.2 Results

6.2.1 Question 1

The results regarding the first question, namely what is the relation between the average segregation time and p are shown below:

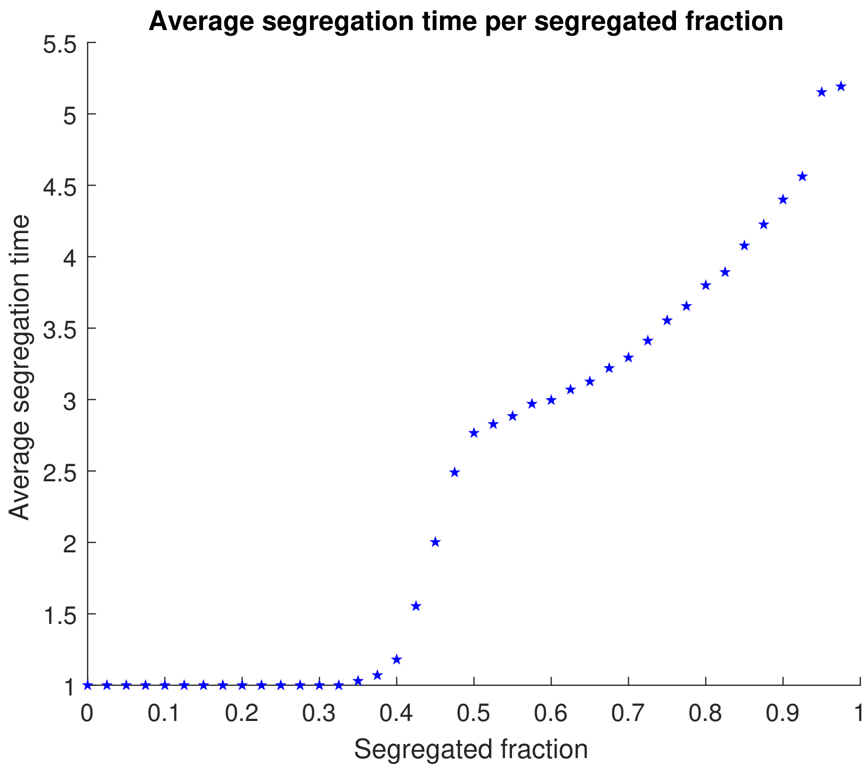


Figure 13: Average segregation time on the standard board

Most notable of Figure 13 is that the relation between the average segregation time and p, opposite to expectations, is neither linear nor exponential. Instead, it appears to be partially exponential and partially lineair. From the figure, we note that the average segregation time increases fastest between 0.4 and 0.5. Which is to say that it takes three times longer for 50% of the population to live in homogenous groups than for 40%. Also note that the 'lift off' is approximately at $\frac{1}{3}$.

Partial explanation

The first part of this graph is easily understood. When the board is initially formed, the odds are fairly high that several persons might already live homogenous. Due to the HR of 1, any non homogenous person will move, resulting in a further segregation. As the boundary of 10% is fairly low, it is likely that this will be achieved in less than 1 generation. As mentioned above, the second interesting segment of this graph is the segment between 0.4 and 0.5, as mentioned, the figure shows that it takes considerably longer to reach a segregation of 50% than it takes to reach 40%. This can be argued from a sociological perspective as follows. Some people are in general unhappy with their lives and will thus move to a new location anyways. The process of segregation, is a fairly lenghty process this way. However once people discover that their close neighbours are moving, they in turn decide to move as well. This does not neccessarily result in more moves, however as one type leaves a neighboorhood, the other type is to remain in place, since the neighbours of this type will tend to be more homogenous when their opposite type leaves. Thus once 50% segregation has been reached, it becomes easier to segregate even further. Since the repositioning of one individual may result in three people becoming homogenous. This explains the part between 50% and 70%. It might appear reasonable to believe that this increase will continue and that it that it would take at most two more generations to reach equilibrium but after 70% segregation, it becomes harder for individuals to move to a homgenous location. This is also heavily impacted by the laws defined in the model, a person will not move to the best location but to the closest better location, which might lead to a person moving from one disfavourable location to another.

Next, we consider the same question however this time on the 4-Type board (Figure 14).

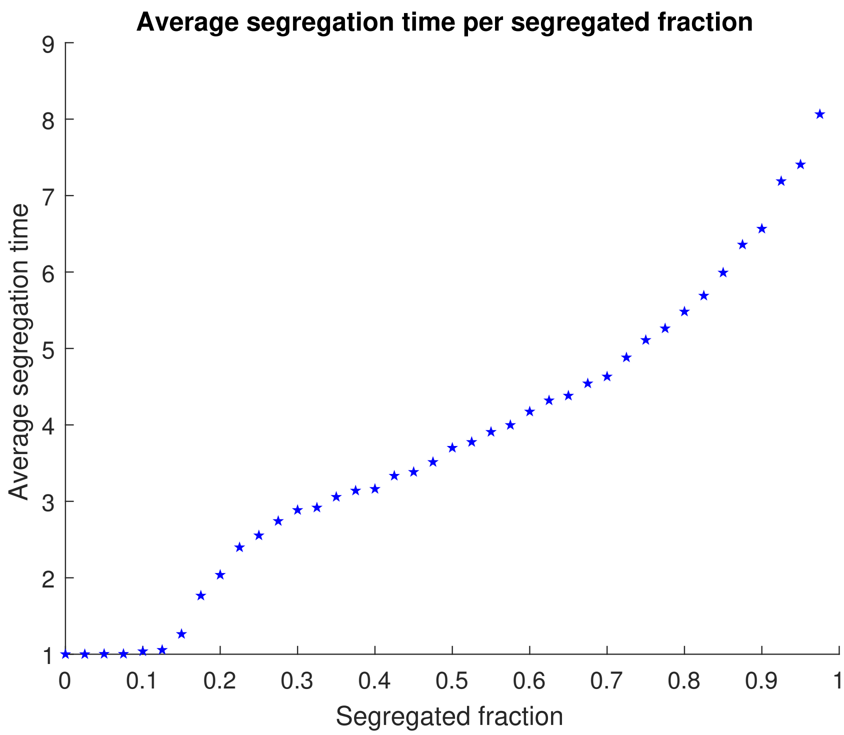


Figure 14: Average segregation time on the 4-Type board

The most notable difference between Figure 14 and Figure 13 is that the 'lift off' happens much earlier on the 4-Type board than in Figure 13. Although this difference feels intuitive, we will give a formal explanation. Consider the probability that any individual i has a neighbours of the same type, and no neighbours of another type, given $k-a$ empty spots, where k is the maximum number of possible neighbours for the location of i. In the basic model, this probability equals

$$p_b = \frac{19}{39}\frac{18}{38}...\frac{19-a}{39-a} = \frac{\frac{19!}{(19-a)!}}{\frac{39!}{(39-a)!}} = \frac{19!(39-a)!}{39!(19-a)!}$$

In the 4-Type model, this probability is

$$p_4 = \frac{15!(63-a)!}{63!(15-a)!}$$

which is less than half of p_b.

Also note that an individual is more likely to be homogenous if he/she is placed in a corner or on an edge. This is because a corner or edge spot has less possible neighbours than an interior spot, thus increasing the probability of only having neighbours of the same type. The 4-Type board has relatively more interior locations and fewer edge positions than the standard board. This effect also decreases the fraction of segregation prior to displacement of individuals. However the people/boardspace ratio is maintained in both models and thus should not have too much of an impact on the results.

Another effect is the rise of the number of generations, that is, the time it takes before the board reaches full segregation. Interestingly, this effect is less strong than one might initially expect. On the standard board, it takes on average 5.5 generations to reach full segregation. On the 4-Type board however, it takes 8.5 generations.

The remaining part of Figure 14 is quite similar to Figure 13: a quick rise, followed by a slightly decreasing ascending relation. The explanation is similar to that of the basic model.

6.2.2 Question 2

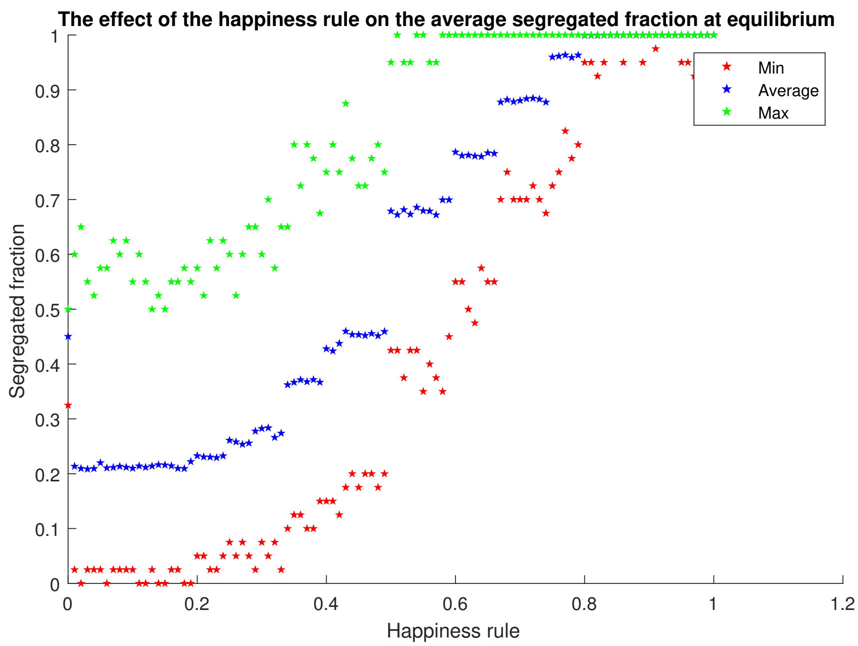


Figure 15: Average segregated fraction as a result of happiness on the standard board. The red and green dots represent the minimum and maximum segregrated fraction of 500 boards for the given happiness.

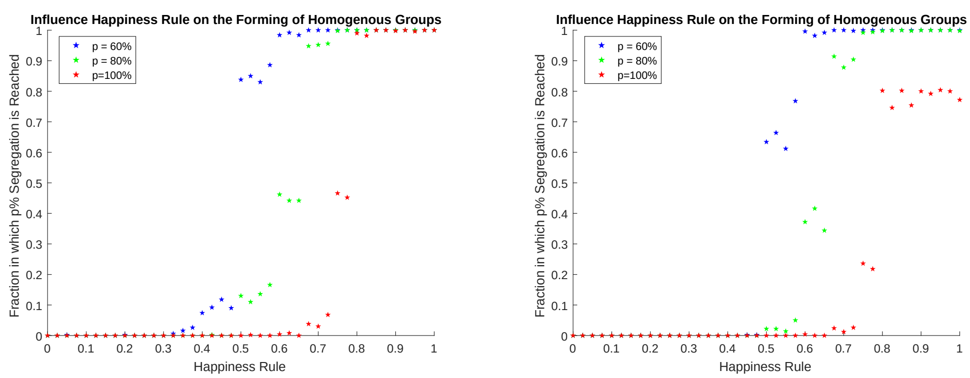
Figure 15 displays the average segregated fraction, the minimum segregated fraction and the maximum segregated fraction ar equilibrium, as a function of the Happines Rule. Notable in this Figure is that the average segregated fraction is greater than the happiness. Another point worth noting is that the segregated fractions tend to appear in different groups seperated by relatively large percentages.

Figure 16: Average (blue) segregated fraction as a result of happiness on the 4-Type board. In green the respective maximum segregated fraction is shown, in red the respective minimum segregated fraction is shown

Figure 16 shows that in contrast to the standard board, the 4-Type board does not respect the property that the average segregated fraction is always greater than or equal to the happiness. A point of interest however, is that the averages of both Figures 15 and 16 seem to cluster. This is because, as mentioned earlier, the effect of the Happiness Rule is discrete when using second order neighbourhood.

6.2.3 Question 3

In order to answer the third question, for any given Happiness Rule, how often is 60%, 80% and 100% segregation reach, we ran 500 simulations for each HR ranging from 0 to 1 with increment 0.025.
The results are shown in Figure 17.



(a) The standard board.

(b) The 4-Type board.

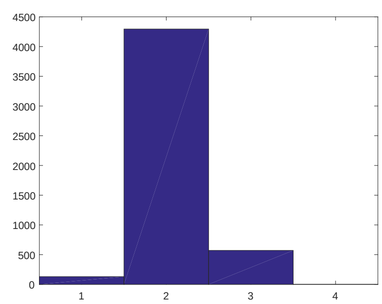
Figure 17: For each Happiness Rule ranging from 0 to 1 with increments 0.025, the fraction of the simulations in which 60% (blue), 80% (green) and 100% (red) segregation is reached.

It is notable that the two graphs are fairly similar, with one key difference. In contrast to the standard

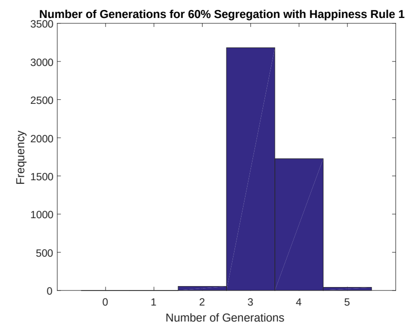
board, on the 4-Type board, full segregation is not guaranteed by a Happiness Rule of 1 (or 0.8). This implies that the board reached an equilibrium state in which no person is able to move to a better location, despite the fact that the boards have the same inhabited to uninhabited location ratio.
We will not discuss these graphs in detail.

6.2.4 Question 4

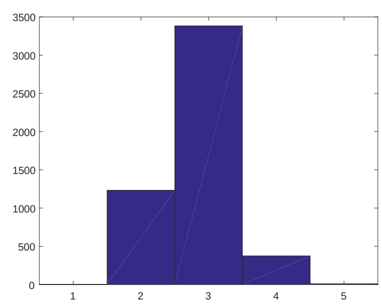
We will now focus on a more relevant question, namely what are the distributions for the segregation time at 60% and 80% segregation, with a HR of 1. To approximate these distributions, we ran 5000 simulations each and plotted them on both the standard board and the 4-Type board in Figures 18a to 18d. Figure 19 displays the QQ-plots of the 60% histograms versus the 80% histograms on the two different boards.



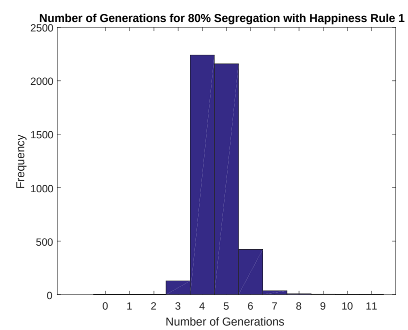
(a) On the standard board with a HR of 1, this histogram displays the number of generations it took until 60% segregation occured, based on 5000 simulations.



(b) On the 4-Type board with a HR of 1, this histogram displays the number of generations it took until 60% segregation occured, based on 5000 simulations.

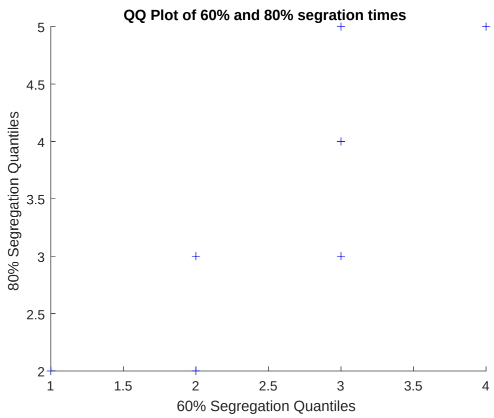


(c) On the standard board with a HR of 1, this histogram displays the number of generations it took until 80% segregation occured, based on 5000 simulations.

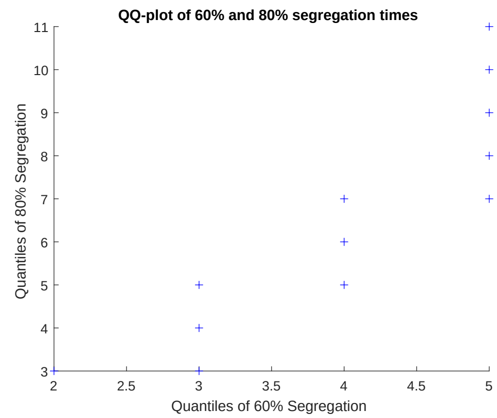


(d) On the 4-Type board with a HR of 1, this histogram displays the number of generations it took until 80% segregation occured, based on 5000 simulations.

Figure 18: The histograms of 5000 simulations of the segregation time at 60% and 80% segregation on the standard board (left) and the 4-Type board (right)



(a) A QQ plot of the 60% histogram 18a versus 80% 18c on the standard board



(b) A QQ plot of the 60% histogram 18b versus 80% 18d on the 4-Type board

Figure 19: QQ-plots of the 60% segregation histograms versus the 80% on the standard board (left) and the 4-Type board (right).

Based on the histograms, one might expect that for both boards, the distribution of the 60% and 80% might come from the same location-family of distributions. However, the QQ-plots indicate that we don't have enough information to confirm this, so we may not assume this to be true.

In Figures 18a and 18c, we see that the number of values for the number of generations that actually appear, is relatively small. This makes it hard to perform statistical tests on the actual distribution of this stochastic. However, we did test whether the means of the two samples for the standard board are different with a significance of 95%. For simplicity, we did not perform this test for the 4-Type board.

To test whether the average generations $\mu_{60\%}$ it takes to reach 60% homogenous group is equal to the average generations $\mu_{80\%}$ for 80%, we performed the ttest2 on these two averages. The null-hypothesis is $H_0 : \mu_{60\%} = \mu_{80\%}$, and the alternative hypothesis is $H_1 : \mu_{60\%} \neq \mu_{80\%}$. For this test, we estimated $\mu_{60\%}$ and $\mu_{80\%}$ with the sample means of sample size 5000. The results are shown in table 2.

Table 2: Sample means of number of generations until 60% and 80% homogenous group is reached with HR 1.

Fraction homogenous group (%)	60	80
Sample Mean	2.0892	2.8322
Sample Variance	0.1349	0.3013

The test has rejected the null-hypothesis, which is expected because the difference of the sample means are around 0.8, which should exceeds the critical value of the test.

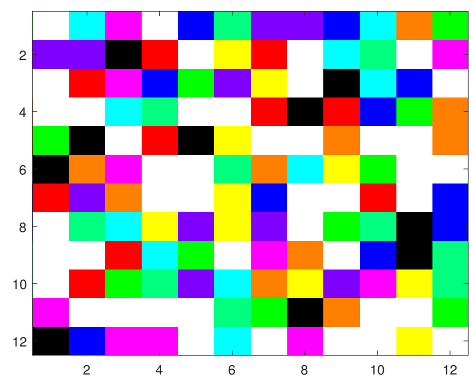
7 The effect of switching types

Segregation is a phenomenon taking many different forms. Rather than limiting the scope to racial segregation, we decided to broaden our view. For example, one might delve into the segregation in studies with respect to friendship. In this case, for any person k, the neighbours of k represent k's friends or the persons with whom k spends most of his time. In this case, k's type, can represent either the classes he is currently taking, or the kind of sports person k is doing, or more realistically, k's political beliefs. In any of these cases, the type of k is not necessarily set indefinitely.
In these cases, the type of k might switch, depending on the types of his friends. This is a sociological process known as conformation and plays a major role in the everyday behaviour of people.

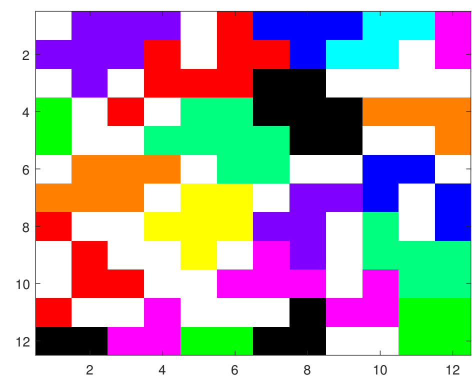
This gives rise to a new modification of the model. For any person k, before moving to a new location, k has a catagorical probability with to switch to a different type, where

$$\mathbb{P}(\text{NewType}(k) = t) = \begin{cases} \frac{\#\text{Neighbours of type } t}{\text{Total number of neighbours}} & \text{if total number of neighbours} > 0 \\ \mathbb{1}_{\{t=\text{Type}(k)\}} & \text{if total number of neighbours} = 0 \end{cases}$$

Note that this is well defined, since if total number of neighbours > 0, we have $\mathbb{P}(\text{NewType}(k) = t) \geq 0$ and $\sum_{i=0}^{\text{types}} \mathbb{P}(\text{NewType}(k) = t) = 1$. An example of how this may effect the equilibrium, is illustrated below:



(a) Situation before segregation



(b) Situation after segregation

Figure 20: An illustration of the effect of switching types on the 12×12 Board

A crucial difference between equilibrium with and without switching, is that if switching is allowed, the concentration of the types may differ between the starting board and the equilibrium status. This can be seen in Figure 20 as the number of people of the green type decreased from 10 to 8, the yellow type even decreased from 10 to 6. As a matter of fact, allowing switching might lead to the extinction of several types.

This is not generally the case for lower bounds on the happiness rule, but if the happiness rule is increased, several types might cease to exist. Despite this being quite interesting, the extinction of types is not included in this report. What will be studied in this report, is the effect of switching, on the given research questions. In other words, how will the results discussed in this report, vary when switching types is allowed.

7.1 Average number of generations until equilibrium

First, the effect of switching on the average number of generations until equilibrium will be discussed. From Figure 8, we deduced that on average, a board reaches equilibrium in only a few generations. This however decreases even more drastically when switching goes in effect:

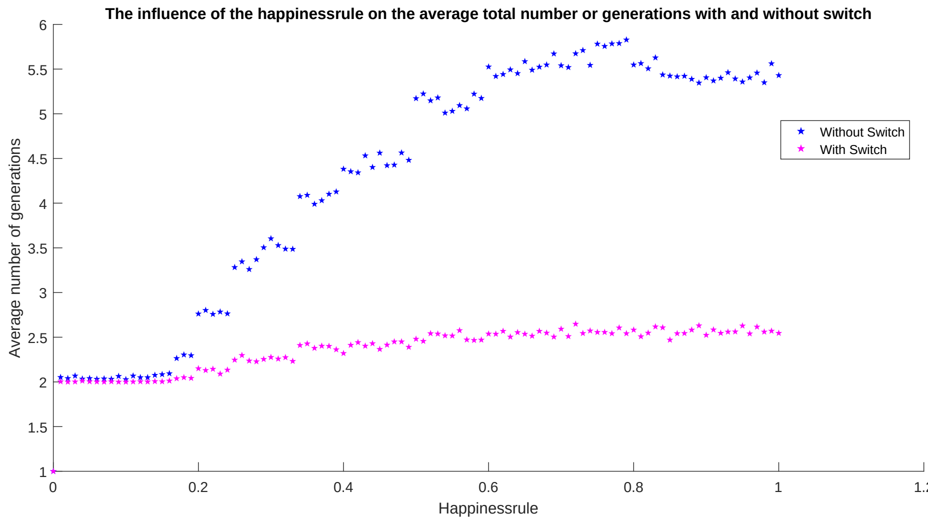


Figure 21: Number of generations until equilibrium on an 8x8 board with standard settings, with and without type switching

In Figure 21, we see a comparison of the average number of generations for with and without type switching. Clearly, when type switching is applied, it takes much less generations to reach the equilibrium. This is expected, since the probability of becoming the same type as the neighbours is more favored in this model, which results in an increase of happiness. Or, in a sociological view, the person is willing to adapt to the environment rather than moving on.

From Figure 21, we observed that the average number of generations to reach equilibrium is nearly constant for any HR in the interval $(0.6, 1)$. Just to get an idea whether this statement is justified, we selected three different HR values, namely 0.6, 0.8 and 1 and performed the two sample t-test (for more description about this test, see section 2.2, Model and Statistical Methods) in Matlab. Because the histograms show that the data might follow an asymmetric distribution and this t-test requires the difference of the means to be normally distributed, we chose a sample size of 50000 to ensure normality through the Central Limit Theorem.

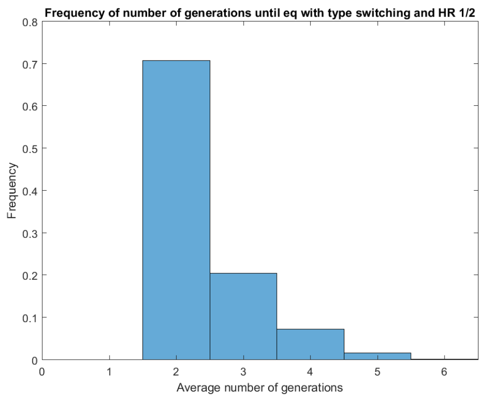
The sample means are shown in Table 3:

Table 3: Sample means of number of generations to reach equilibrium simulated with HR 0.6, 0.8 and 1, with 50000 sample size for each HR.

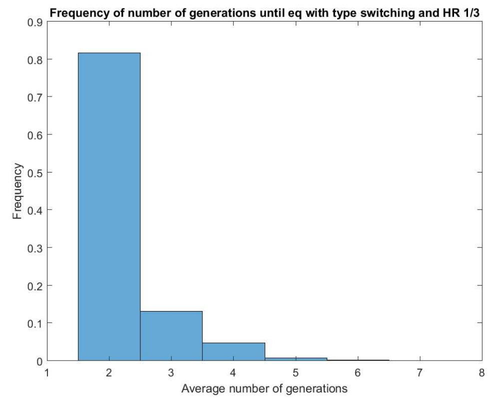
HR	0.6	0.8	1
Sample Mean	2.5213	2.5616	2.5711

We applied the t-test twice. First time comparing the mean of HR 0.6 and 0.8. Second time for comparing the mean of 0.8 and 1. The first test rejected the null-hypothesis, the second did not. So with 95% confidence level, we may conclude that the average generations are unequal for HR 0.6 vs 0.8, but equal for HR 0.8 vs 1.

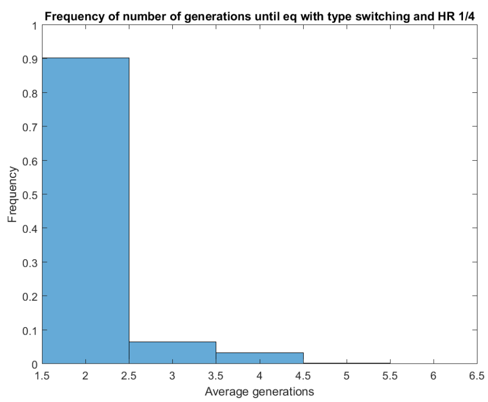
Similar to section 4, in order to get an idea about the distribution of the random variable of $Y_{j,s}$ (referring to the notation in section 4, here s means type switching is applied), we plotted again histograms for HR $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$ and 1 in Figure 22.



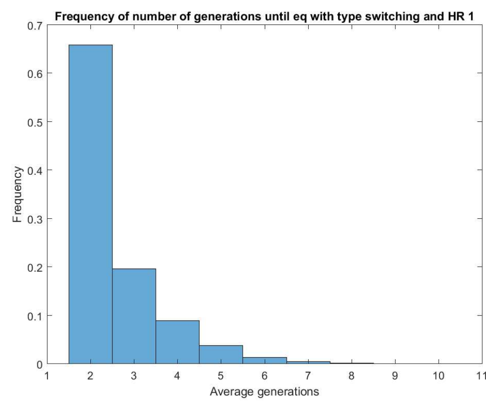
(a) 5000 simulations of number of generations until equilibrium with HR $\frac{1}{4}$ with switch



(b) 5000 simulations of number of generations until equilibrium with HR $\frac{1}{3}$ with switch



(c) 5000 simulations of number of generations until equilibrium with HR $\frac{1}{2}$ with switch



(d) 5000 simulations of number of generations until equilibrium with HR 1 with switch

Figure 22: The probability distribution of Y (the number of generations until reaching the equilibrium) is approximated with a histogram of bin size 1. For each HR ($\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$ and 1), 5000 simulations were ran.

Again, all $Y_{s,j}$'s appear to be following the same distribution just as without switch (section 4). The histograms suggest that the $Y_{j,s}$'s might follow a geometrical distribution, which is a negative binomial with $r = 1$. But we could not perform the same statistical analysis as in section 4, because the maximum likelihood estimators for the negative binomial could not be calculated through the fitdist or other functions. This is due to the fact that the sample means exceeded the sample variance. So from this we could conclude that $Y_{j,s}$'s cannot be geometrically (or NB) distributed.

It is much more interesting to compare the distribution of $Y_{j,s}$ with Y_j. A remarkable observation is that the histograms of $Y_{j,s}$ completely lost it's symmetry compared to the histograms of Y_j. Apparently, type switching stimulates segregation so well that it only takes 1 generation to reach equilibrium (remark: we always start with 1 generation, so the fastest segregation is reached at the second generation).

To further examine how type swithcing affects generations until equilirbium, we chose to compare $Y_{1,s}$ with Y_1 in a QQ-plot:

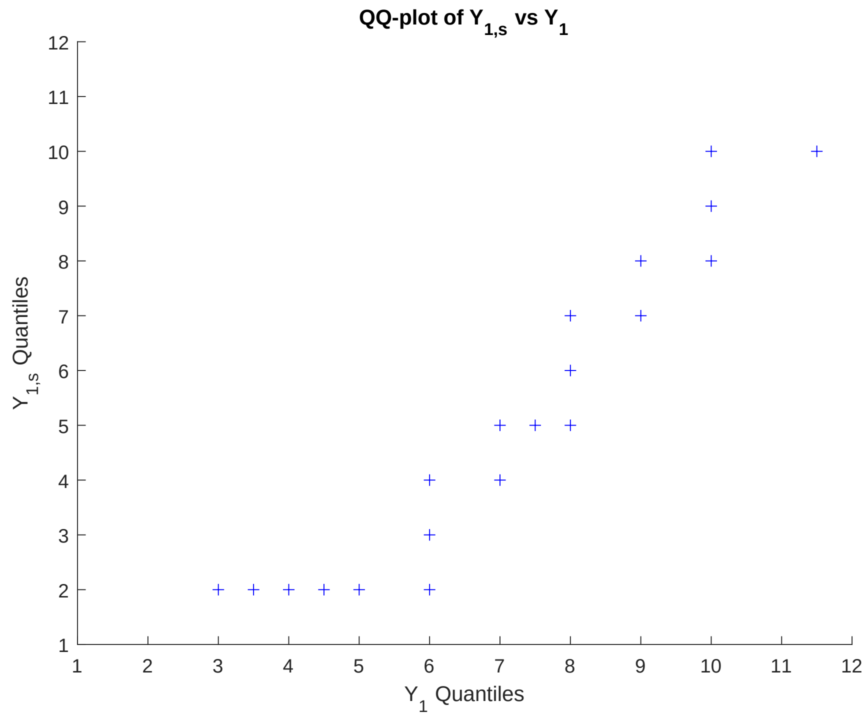


Figure 23: QQ-plot of $Y_{1,s}$ vs Y_1

Starting from quantile 3 of Y_1, the line is flat, which is expected because the most quantile of $Y_{1,s}$ is concentrated on 2. After the flat line we see a linear pattern, which suggests that the 'tail' part of $Y_{1,s}$ and Y_1 are not too differently distributed. This is also seen by comparing the Figure 22 with Figure 9.

7.2 Happiness at equilibrium

Secondly, we investigate the effect of the ability to switch type on the mean happiness after the equilibrium is reached. We examine this effect for different values of the Happiness Rule as done in section 5. In section 5, we found that the mean happiness at equilibrium increases as the Happiness Rule gets higher. Intuitively, we expect a higher happiness when individuals can switch type, because the chance to become happier by switching is really big. This follows directly from the probability distribution we defined earlier in this section. The comparison is shown in Figure 24

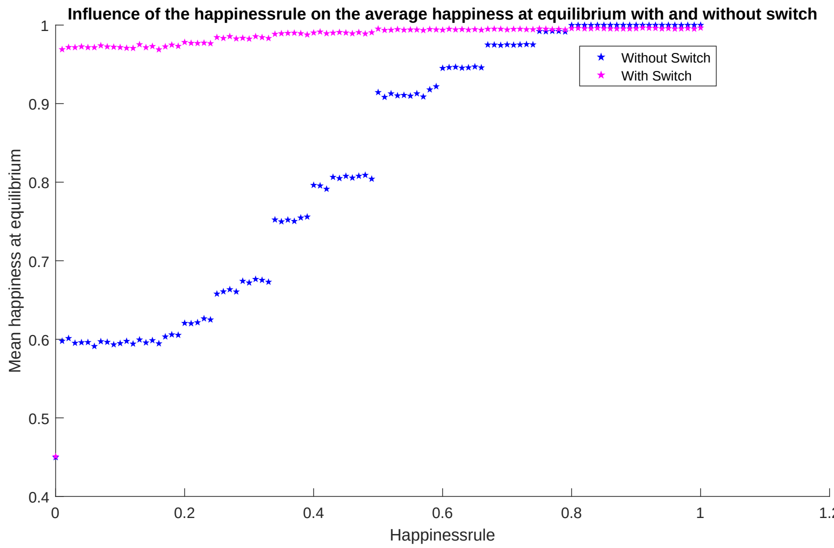


Figure 24: The effect of the ability to switch types (pink) and the Happiness Rule on the mean happiness at equilibrium), on the standard board. Blue shows the realtionship without switching types

It seems that the possibility to switch types causes a much higher average happiness at equilibrium than expected. However, for Happiness Rules higher than 0.8, we see that the mean happiness on the board is slightly lower, if the individuals can switch type. This may be be explained by the number of generations. We already saw that the mean number of generations is less if the possibility to switch types is included. This results in fewer chances for each individual to improve his/her happiness. Thus, the mean happiness will be lower. For the lower Happiness Rules, the switch has a much higher influence, because the chance to get a high happiness is significantly higher than keeping a lower happiness.

7.3 Average segregation time

The final topic to be discussed in this report, will be the effect of conformation on the (average) segregation time. For this purpose, we will compare the results obtained in section 6 with the results if switching were allowed. We will only consider the results for the standard board.
For any $p \in [0, 100]$, we will consider the time until which $p\%$ of the population lives in homogenous groups. With a chosen HR of 1, we obtain the results shown in Figure 25.

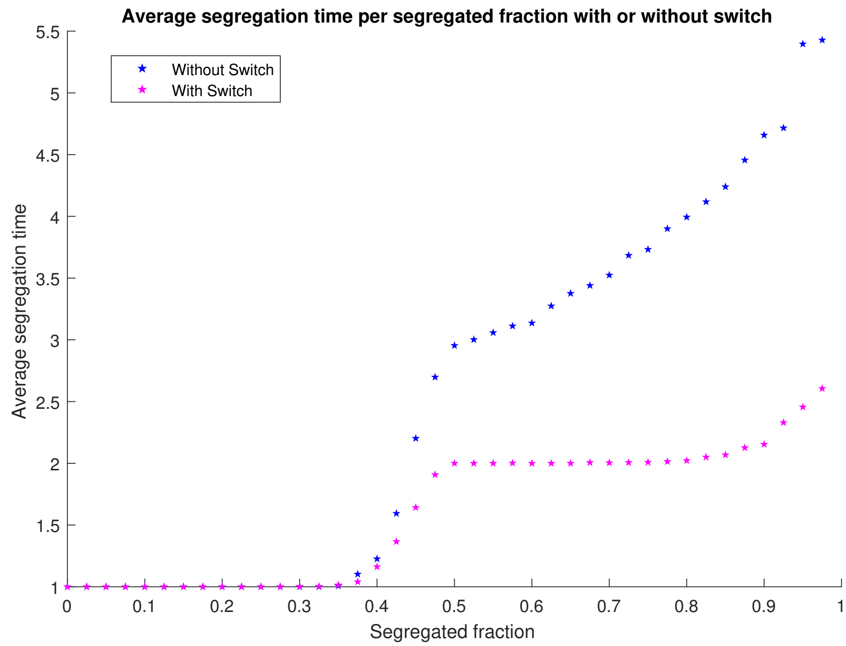


Figure 25: The average segregation as a function of the segregation fraction with (pink) and without (blue) the abbility to switch types

From Figure 25, the hypothesis that the average segregation time is lower for every fraction, if switching is allowed, is accepted. A less trivial result of conformity, is that the average segregation time is nearly constant between the fractions 0.5 and 0.8. To test whether it is justified to make this claim, the t-test as described in section 2.2 was conducted. We will test if the segregation time at 60% and 80% have the same mean value if 5000 simulations are ran each. The t-test, rejected the null-hypothesis that 60% segregation and 80% segregation share the same mean value (the respective means were 2 and 2.01 with th ability to switch). As mentioned in the introduction, a mean of 2 generations implies here, that the board has reached 60%, in at most 1 generation, i.e. every individual has moved or switched type only once at most. We conclude that on the standard board with HR 1, segregation at p% takes place in less than one generation, for any $p \leq 70\%$.

It is interesting to note that both graphs appear constant for any $q \leq \frac{1}{3}$. This can be explained as follows. In both scenarios, the segregated fraction has been reached, by the boards default setup and thus no person will move before segregation has been reached. Another remarkable similarity between both graphs, is that they tend to grow with the same order between a fraction of 0.3 and 0.5.

8 Conclusion

8.1 Equilibrium is not always reached

Coming back to set of the first research questions, we have seen that equilibrium is not always reached on the standard board in finite time. This is because a periodic movement of individuals may occur. Also, equilibrium might not always be reached on a larger board (with equivalent settings), since we can just implement this periodic solution in the larger board.

8.2 Average number of generations until equilibrium

For the standard model, we have seen that equilibrium is reached on average at the third generation, and that the average is a non-decreasing function of the happiness rule in the model. When type switching is allowed, the equilibrium is reached on average at the second generation for the standard board. And again, the average is a non-decreasing function with respect to the happiness rule. Moreover, we have observed that type switching strongly stimulates segregation.

We conclude that Y_j and $Y_{j,s}$, which respectively denotes the number of generations it takes to reach the equilibrium without or with type switching, is neither Poisson, binomially, geometrically, or negative binomially distributed, for $j \in I = \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\}$. The QQ-plot suggests that Y_j and Y_k, for $j \neq k$, $j, k \in I$, might come from the location-scale family of the same distribution. This however, could not be confirmed by a statistical test. Instead, with the Kolmogorov-Smirnov test, it was concluded that Y_j and Y_k are not distributed with the same parameter. Furthermore, the QQ-plot appeared linear for Y_1 and $Y_{1,s}$, which could be an indication that these random variables originate from the same distribution.

8.3 Average segregation time

Segregated fraction

The third research question was: "How is this segregation time distributed with a Happiness Rule of 1 and how does the Happiness Rule affect the average segregation time?".
We conclude that the average segregation time is 3.09 generations at 60%, 3.83 generations at 80% and 5.46 generations at 100%. We have seen that the relation of the average segregation time depending on the segregated fraction considered, is neither linear nor exponential. For the standard board, it takes three times as long to reach 50% segregation as it takes for 40% segregation. After this barrier has been breached, segregation takes place relatively easily. On the standard board, full segregation is, on average, reached after 5.4 generations. For the 4-Type board this number is 8.5.

Forming of Homogenous Groups

On the standard board, 60% segregation is almost always reached when a Happiness Rule > 0.6 is chosen. If we increase the segregation to 80%, the Happiness Rule must be greater than 0.7. For full segregation, a HR of 0.8 is required. On the 4-Type board, these values lie slightly higher, except for full segregation, which is not guaranteed on the 4-Type board.

60% and 80% segregation times

We found that on average, for the standard board with Happiness Rule 1, it took approximately 3.09 generations to reach 60% segregation. 3.83 generations for 80%. The distribution of these segregation times was hard to determine, but the earlier mentioned averages can be considered statistically different.

8.4 The Effect of switching types

The second research question stated: "If an individual is able to switch to another type with a probability that depends on the types of his/her neighbours, how does this affect the average number of generations (until equilibrium is reached) and to what extent does it impact the corresponding distribution and the other questions posed earlier?"

From Figure 21 it can clearly be concluded that it will take at most 2.5 generations on average until equilibrium is reached if it is possible to switch types. Whereas it may take up to 5.5 generations to reach equilibrium, if type switching is not allowed. We also conclude that the corresponding histograms (with and without switching), do not share the same distribution. We noticed a strong increase in the average happiness at equilibrium when switching is allowed and the average segregation time never exceeds 2 generations for any $p < 80\%$.

9 Discussion

9.1 Average number of generations with and without type switching

As mentioned in the conclusion, we could not find a proper fit for the distribution of Y_j and $Y_{j,s}$, since it does not follow any standard discrete distribution like poisson, binomial, geometric or any other standard distribution. We doubt that the distributions of Y_j or $Y_{j,s}$ are ever recorded in the literature. But in our opinion, the histograms (Figure 9) provide a good approximation for the distributions. A remarkable observation was that the histograms, which can be defined as a random variable $h_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{a_{j-1}<x\leq a_j}$, for $x \in (a_{j-1}, a_j]$ (x a realisation of a random variable), did not appear very random. What we mean here, is that when we plotted the histograms of several samples from the same Y_j, they were always nearly identical, which is not expected from a random variable. Although it was not confirmed by a statistical test, the observation points to a persumable very peculiar distribution of Y_j.

Another remark about the histograms, is that unfortunately, there is a small convention error for the number of generations (an error which could not be corrected in time). Our model generation counters always started at 1 and thus by defnition, equilibrium can never be reached at the first generation. This convention is actually confusing because the counter would give 2 if 1 generation has passed. So in our model, we actually assumed that one generation has always passed. However, this error does not seem to affect the histogram more than just a shift of the mean.

9.2 Average segregation time

In line with section 9.1, a distribution of the 60% and 80% segregation times could not be found. We should note that, when running a simulation that takes 500 instances for a given happiness rule, some uncertaincy still remains. This is best shown in Figure 16, where, due to the discrete effect of the happiness rule, the average segregated fraction should be a perfect stairs pattern, but the figure shows some distortion. Of course, this is expected since 500 may be considered a 'large' number, but it still results in distortions. One could argue whether the step size for the happiness rule dependence should have been chosen larger. Since for second order neighbourhood, the number of possible happiness values is quite limited. However, due to the earlier mentioned inaccuracies, setting this step size too large decreases the accuracy of the figures that include happiness rule dependence.

9.3 The effect of switching types

As seen in Figure 25, if conformity is allowed, the population will be nearly 80% homogenous after one generation. This gives rise to a critical note for this extension with respect to the standard board. It is our belief namely, that the standard board is too limited in size, number of types and population density. As a result the standard board will reach equilibrium in no more than 2 generations on average. On extremely rare occasions, it might take 7 generations to reach equilibrium (Figure 22).

Appendix

Here are some of our results we did not further use in our report, because we did not investigate them. However, these graphs may be really interesting to investigate further.

First, we show our results on the effect of the Happiness Rule on the total number of moves.

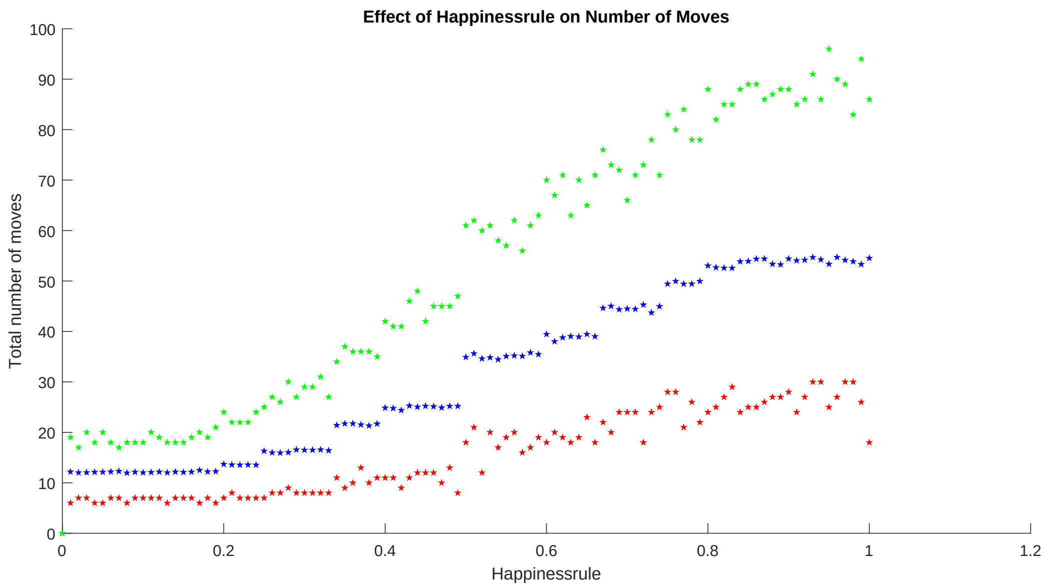


Figure 26: Effect of the Happiness Rule on the total number of moves with the standard settings

Figure 27 shows the effect of the random move on the number of generations.

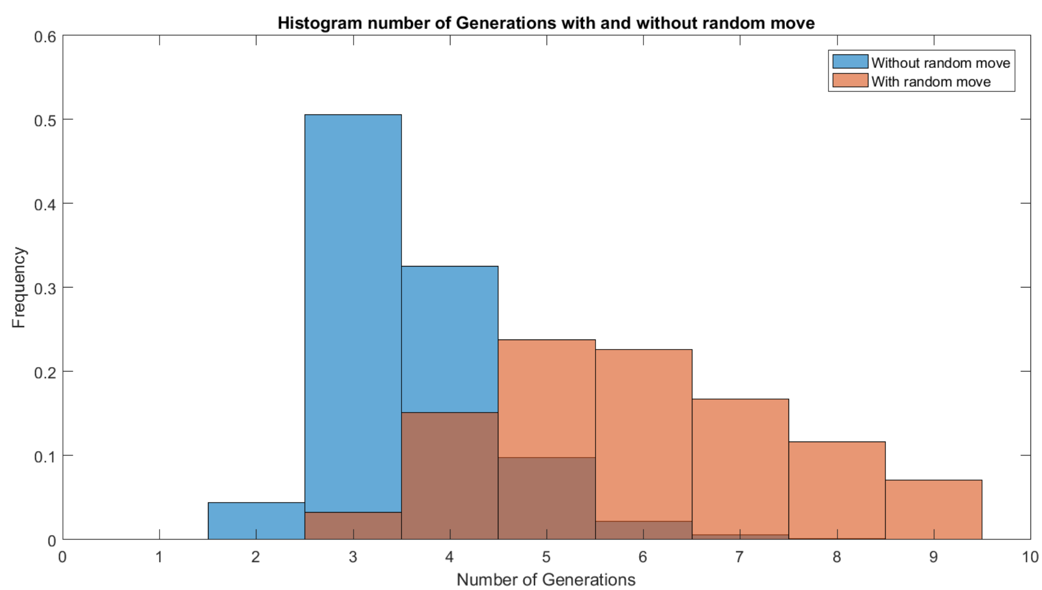


Figure 27: Effect of the random move on the number of generations with the standard settings

Figure 28 shows the effect of the order of neighbourhood on the number of generations, the number of moves, the segregration fraction and the average happiness of all individuals in equilibrium.

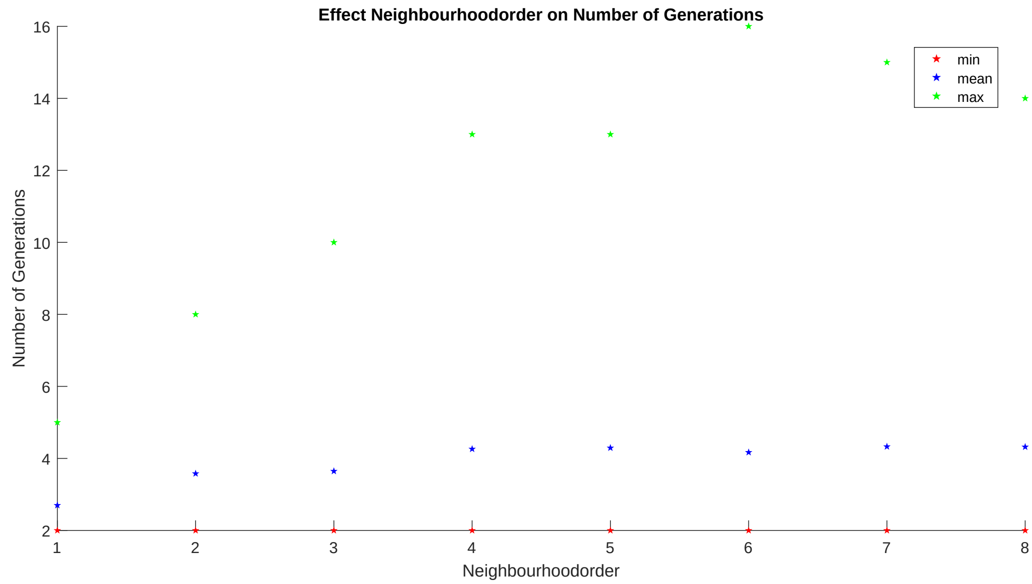


Figure 28: Effect of the Neighbourhood order on the number of generations on the standard board

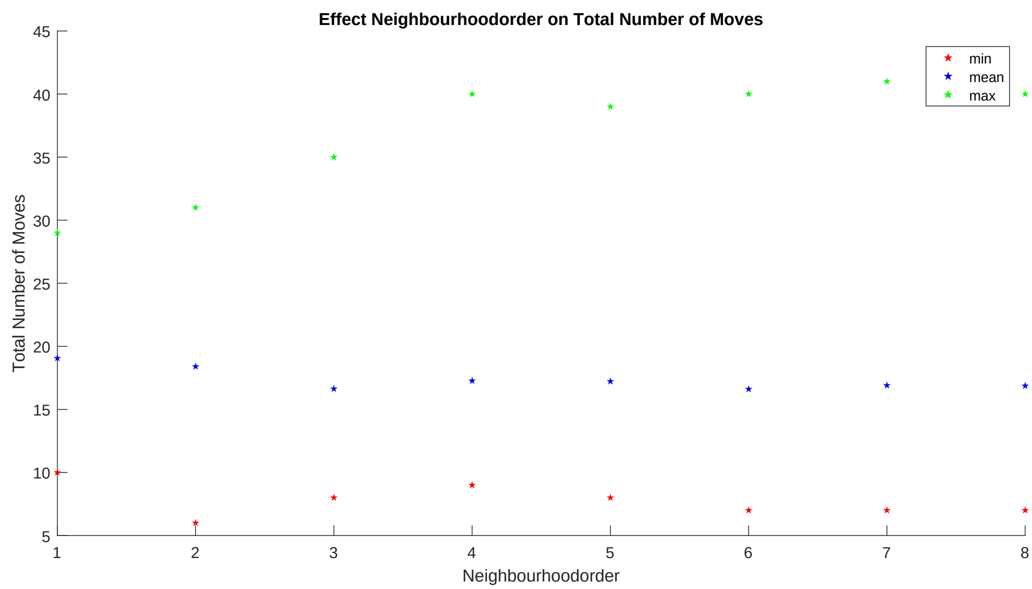


Figure 29: Effect of the Neighbourhood order on the total number of moves on the standard board

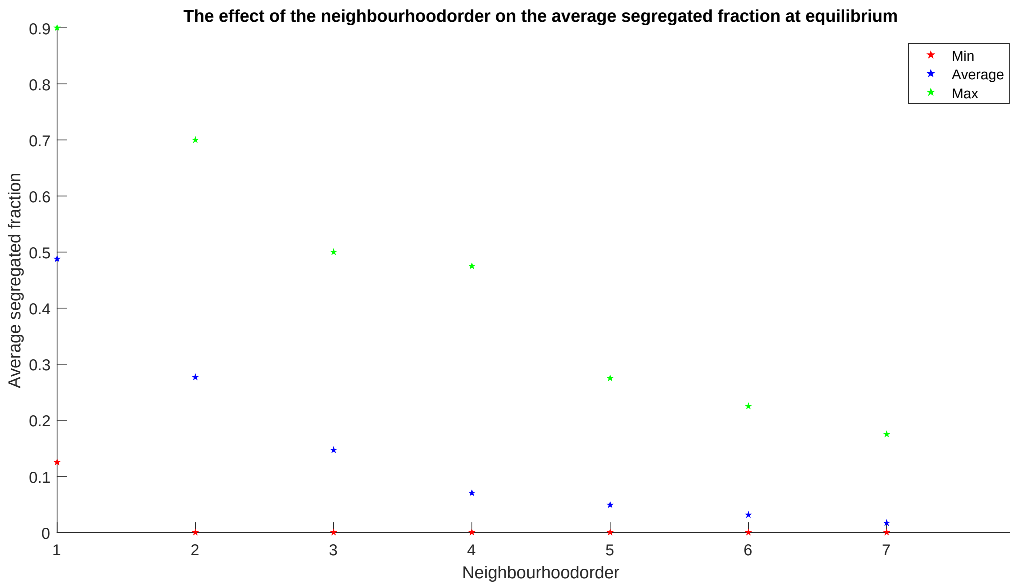


Figure 30: Effect of the Neighbourhood order on the segregated fraction in equilibrium on the standard board

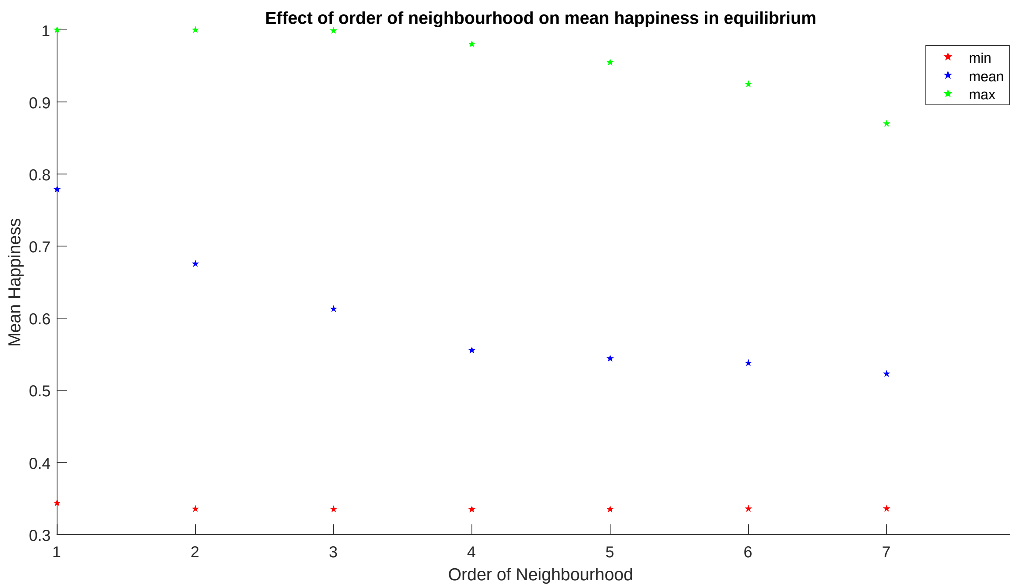


Figure 31: Effect of the Neighbourhood order on the mean happiness in equilibrium on the standard board

Furhtermore, the effect of the criminals is really interesting. Figures 32 to 36 show the effects of the number of criminals on the standard board (The maximum number of criminals is 40).

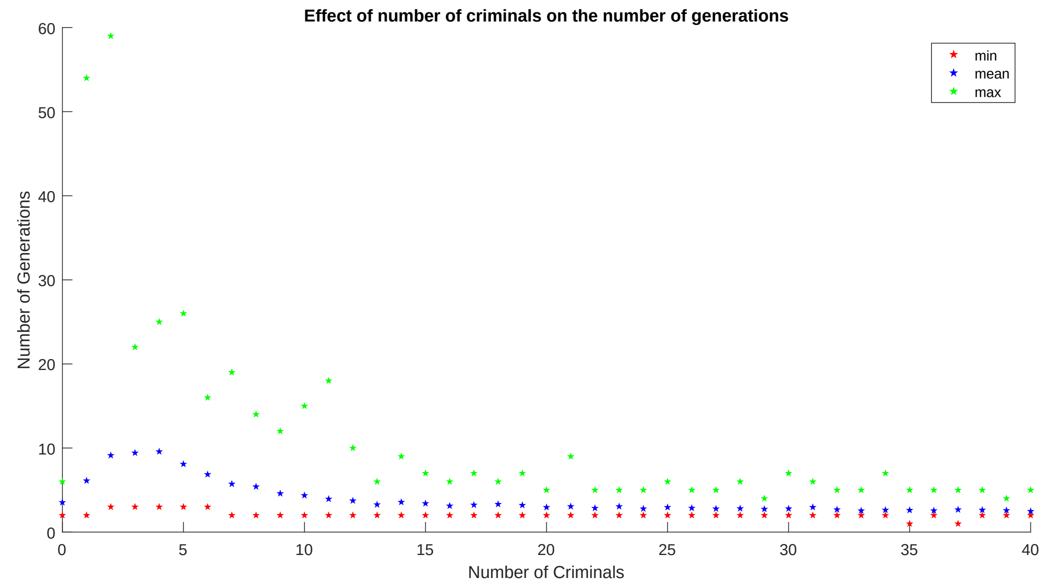


Figure 32: Effect of the number of criminals on the number of generations on the standard board

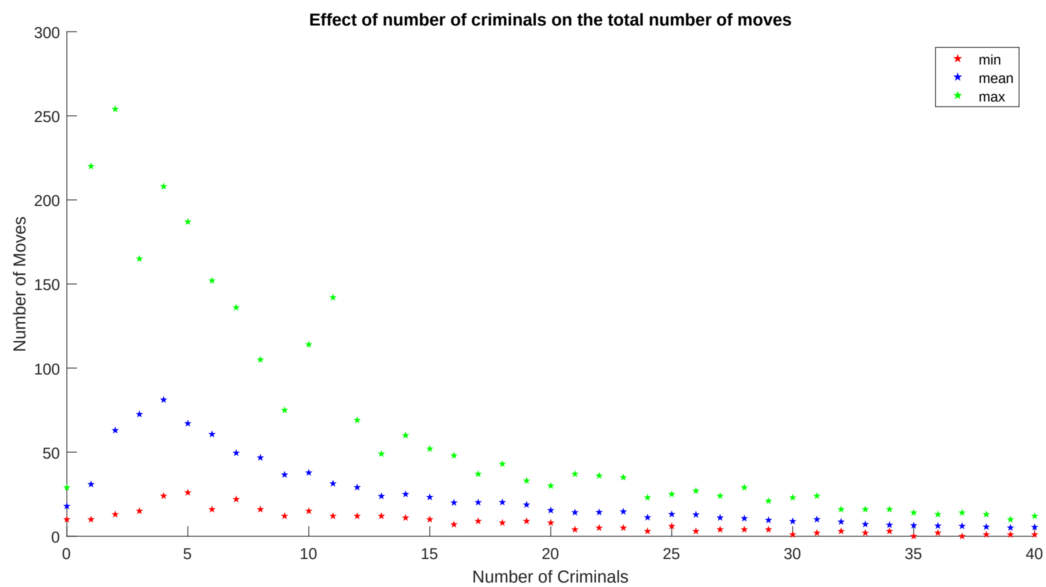


Figure 33: Effect of the number of criminals on the total number of moves on the standard board

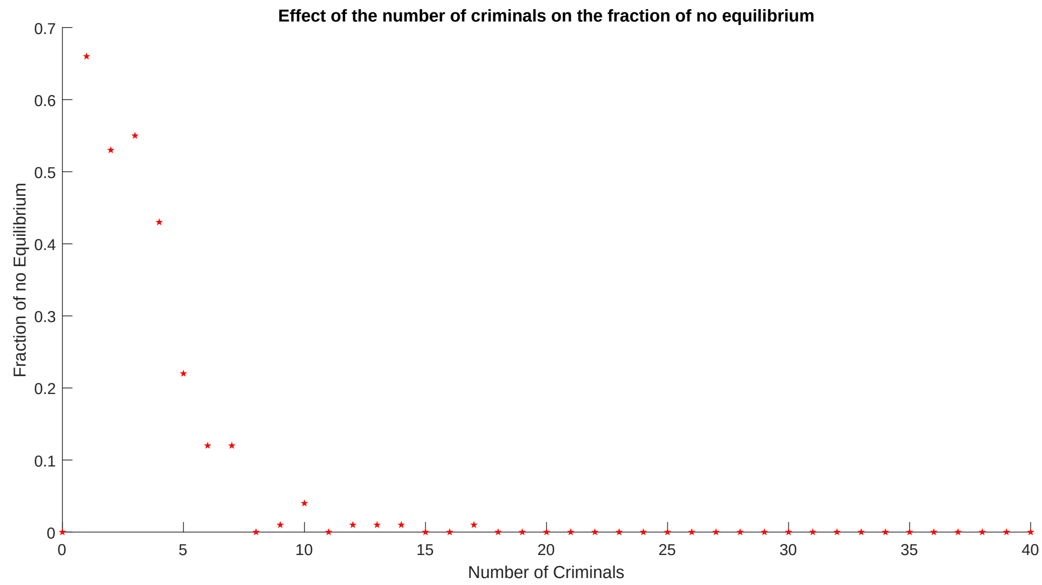


Figure 34: Effect of the number of criminals on the relative number of times with no equilibrium on the standard board

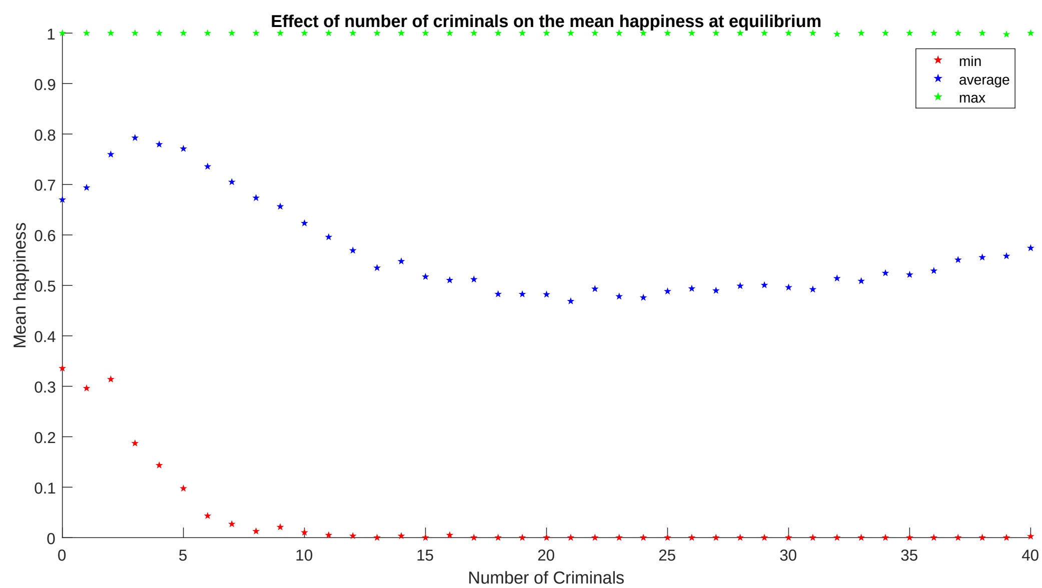


Figure 35: Effect of the number of criminals on the mean happiness at equilibrium on the standard board

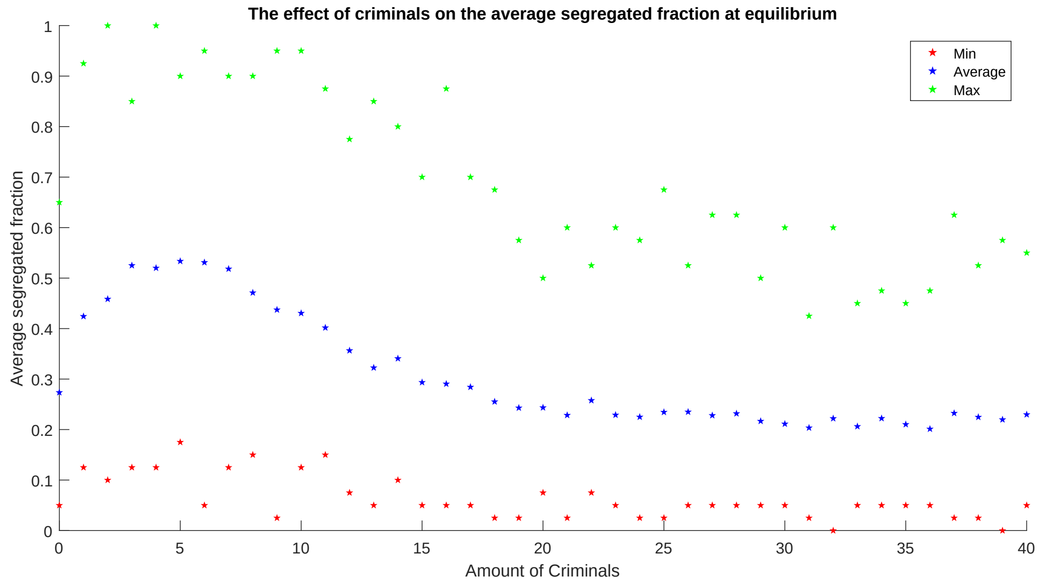


Figure 36: Effect of the number of criminals on the mean segregation on the standard board