# Complex Systems 270 - Agent Based Modelling
# Project 2 - Expanding the Schelling Model

Valliappa Chockalingam, Brandon Kittle, Jay Soni

January 26, 2017

# Contents

## History

In 1971, Thomas Schelling started with a simple experiment aboard an airplane when he found himself with nothing to read. He started by drawing X's and O's to represent people of different color, and erasing the X's and O's or redrawing them based on whether they were happy with their neighborhood. Later, he explained another method whereby two types of coins (nickels and pennies) are placed on a chessboard at random and the coins are moved around based on the proportion of neighboring squares with the same type of coin. Through these experiments, Schelling showed that mild preferences for color or ethnicity of neighbors could lead to total segregation and that the simple rules used could lead to complex emergent behavior.

Later, Schelling authored a book titled "Micromotives and Behavior" that talked about ideas like the "tipping point" where people even with a preference for a mixture of racial backgrounds "up to some limit" could lead to total segregation regardless of their motives, malicious or not.

As the case with most revolutionary ideas, there were indeed those who had different ideas about segregation and how preferences affect social "macrobehavior." Elizabeth Bruch and Rob Mare in their paper "Neighborhood Choice and Neighborhood Change" challenged Schelling's conclusion, stating that "high levels of segregation occur only when individuals' preferences follow a threshold function."

At any rate, these experiments and conclusions are certainly interesting, and we will be talking about the Schelling and Bruch & Mare models of segregation and expanding on them. Today, Thomas Schelling is a professor of foreign affairs, national security, nuclear strategy, and arms control at the School of Public Policy at University of Maryland, College Park, Elizabeth Bruch is an Assistant Professor in Sociology and Complex Systems, and an Affiliate of the Population Studies Center at the Institute for Social Research at the University of Michigan, Ann Arbor, and Robert Mare is a Professor of Sociology and Founding Director of the California Center for Population Research at UCLA.

**Abstract**

In this paper, we begin with a discussion about how the preferences of the agents affects the final percent similar values. We observe, much like Schelling did, that mild preferences can lead to total segregation. Through a graphical analysis, we also found that there is the famous "tipping point" at about 68% of percent-similar-wanted. Finally, we see that the trend seems to be best represented by a cubic function, however, this may be due to overfitting, and the true relation between percent-similar-wanted and percent-similar may well be represented by some other mathematical function, probably logarithmic or quadratic given the appearance of the curve.

We then expanded the model by giving the two kinds of agents different levels of preference, so each type had its own level of percent-similar-wanted. We observed that the overall percent similar was the highest when both groups had a percent similar wanted between 50% and 75%.

Next, we added another agent to the model. So, instead of the two types of agents we had previously, we now had three types of agents. In this model, an interesting phenomenon was observed. Particularly, the lowest percent similar occured when all three agents had a percent similar wanted level that was close to 80%. This is likely due to instability being caused by the high percent similar wanted values that made each group perpetually unhappy, therefore causing the model to run almost indefinitely.

To then observe the affects that a "gradual" function would have over the threshold function we were using previously, we repeated the above experiments with the levels of preference for each agent being set based on a normal distribution with the mean and standard deviation for the percent similar wanted values of each agent set by the user. This gradualist model is more realistic than a threshold model in that in an example with a percent similar wanted of 50% in a threshold model, all of the agents of a type are happy at 50%, but become unhappy at 50.1%. In a gradualist model, the standard deviation allows the percent similar wanted at a certain moment in time to vary. We observed how different values of the mean and standard deviation values in this model cause different percent similar values, and how the "instability effect" observed in the threshold models also show up in the gradualist model.

We then tried many different expansions on the general model. We first determined how majorities and minorities affect the results, i.e., with different ratios of the number of each agent we observed how the percent similar was the lowest when the number of each agent was the same. Later, we tried a few very interesting experiments. First, we started out simple by giving each agent a list of Boolean qualities and removed the color "factor" to simply observe how the model would be affected by a simple threshold function whereby an agent's happiness is determined by the number of agents around it with at least x number of qualities in common. Then, we included a real number to represent each quality, gave a weightage for each quality and observed the results again. Finally, we included the color factor in and also allowed for changes to the "non-color" qualities after a certain amount of time at random. To conclude our research into this topic, we ended with making a more user friendly model where a user could make "events" happen whereby a certain turtle's preference level could change or the preference levels of some type of agent could be changed globally. This is to model the affect that global and local events can have.

In conclusion, through the various models we used, we found many interesting results and realized some rather unintuitive "effects." Many of these can indeed be applied to real life. It however must be noted that there are obvious limitations of using these models. As an example, to model stochastic or probabilistic activity, we used "random numbers." In reality, computers cannot generate truly random numbers, and it is likely that there is some bias or effects caused by the fact that the random numbers chosen were in fact pseudorandom. Another limitation that we faced was that of time and computation. Given limited computational resources and time, we could not study many relationships in great detail, however, we tried compensating this by observing things that we thought might be interesting. To that extent, we believe that many of the results we have found are useful and can certainly be incorporated into understanding segregation, integration, and social macro-behavior.

# 1 Introduction

We started with a base model from the NetLogo models library under "Social Science" and titles "Segregation." The model looks as follows on setup.

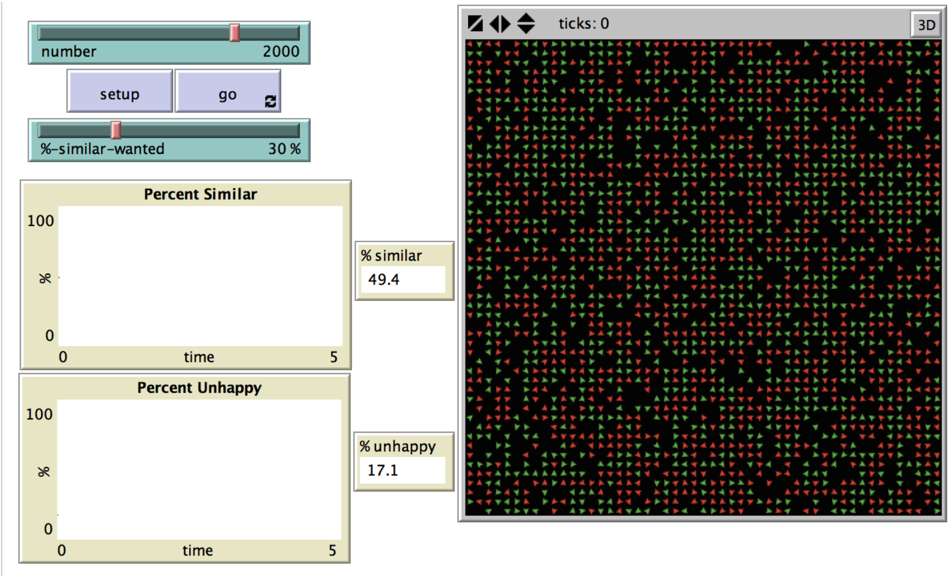



Figure 1: Initial Setup

As can be seen on the image above, NetLogo creates a set of turtles on the black canvas that represent the world. The turtles (NetLogo synonym for agents) are of two colors, red and green to represent the fact that the agents have one of two colors, green or red, associated with them. There is a slider for the number of agents and a slider for the %-similar-wanted. This is a value that represents *every* agent's preference level. In the image, %-similar-wanted is set to 30, this means that every agent is happy when at least 30% of it's neighboring agents have the same color. Finally, looking at the code for this model, it becomes apparent that the model stops when every agent is happy (or runs indefinitely as will be seen later).

Running the model with the initial setup shown on the previous page, we get the the following final result:

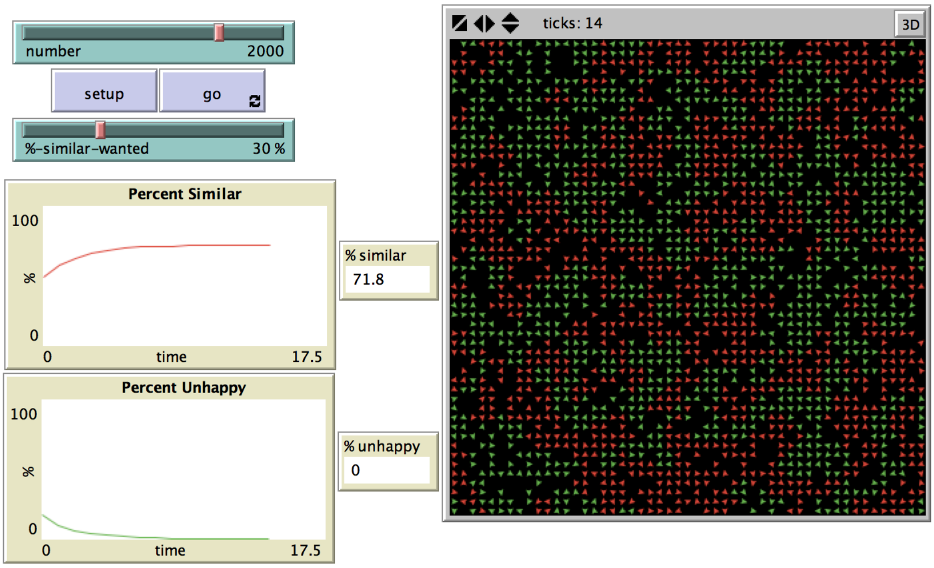



Figure 2: On completion of "Go Forever" with the setup configuration shown before

From the image above it can be seen that the model does halt. After 14 ticks, all the agents are happy (%unhappy is 0). Importantly though, The "Percent Similar" graph and the final %similar value show something very interesting. With %-similar-wanted set to 30%, the final %similar is 71.8%. This agrees clearly with Schelling's results because the rather mild preference every agent possesses to be near agents of the same color leads to near total segregation. 100% Percent Similar is total segregation, but it very hard to reach given the fact that the map size does not really allow too much space for "gaps" between the agents. An example to illustrate the fact that if a lower number of agents were used, the %similar would increase, can be seen in the Appendix.

To illustrate a common case where the model runs indefinitely, given below is the final result of a simulation with 1500 agents and 99%-similar-wanted stopped at 500 ticks:

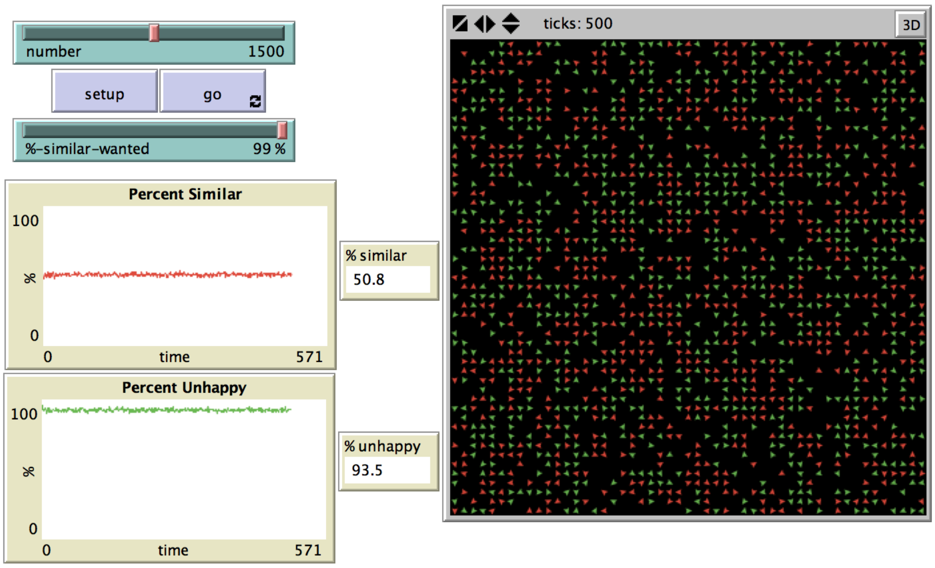



Figure 3: An Indefinite Run

Clearly, the model does not seem to be reaching a point where the agents are all happy. In fact almost all agents seem to be unhappy even after 500 ticks, while the previous run took only 14 ticks to finish. The reason for this is simply that none of the agents can seem to match their preference level, i.e., no agent can seem to find a location where almost all of the neighboring agents are of the same color. Also, another interesting point to note is the constant near 50% Percent Similar. This shows that the agents are really almost randomly configured. This is characteristic of many simulations we attempted, and so, it must be noted now that all the simulations were run with some "stopping condition" whereby the simulation would stop at say 100 ticks.

## 2 Simulations and Graphical Analysis

With some general trial and error to see how the model works, we then set to observe how the model generally behaves. So, we used NetLogo's BehaviorSpace feature to run multiple simulations. These were the settings used:

| Variable | Lower Limit or Constant Value | Increment | Upper Limit |
|---|---|---|---|
| number | 1500 | - | - |
| %-similar-wanted | 10 | 5 | 75 |

The %similar was then measured as the dependent variable and the mean of ten runs was chosen for data analysis. Also, a stopping condition of 100 ticks was chosen after seeing how long most simulations that eventually end take which is usually about a few tens of ticks.

With these settings, we obtained the following graph:

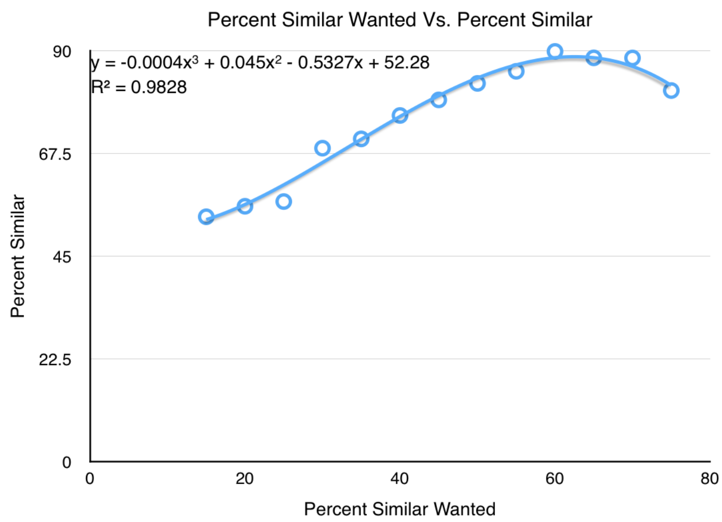



Figure 4: Graph of Percent Similar Wanted Vs. Percent Similar

As can be seen above, the graph seems to show some interesting properties. Firstly, as an explanation of the graph, the trendline used is cubic. This maybe overfitting, but, after trying different trendlines (polynomial of other degrees, logarithmic, power, and linear in particular), a cubic trendline seems to do a good job of modeling the general trend. Secondly, there seems to be a clear increase in the Percent Similar as the Percent Similar wanted. This is quite an intuitive result because as the agents want to be around agents of their own color more strongly, they get the result they want, more agents have neighbors of their own color.

However, after about 68% of Percent Similar Wanted (corresponding to a near 100% Percent Similar), found through minimization of the cubic curve (taking the derivative of the cubic equation and setting it to 0), there seems to be a decrease. This may seem unintuitive, but we believe this result is seen as the model becomes unstable when the preference levels of the agents is very high and their expectations cannot be met, much like what is explained along with Figure 3 where the model runs indefinitely and agents can't seem to find a location that "matches" their preference levels.

Now, the cubic or more generally, polynomial trendline can be explained through the last point. Since there is a peak value for Percent Similar and the general trend seems to be a monotonous increase in the Percent Similar till the peak followed by a monotonous decrease in the Percent Similar possibly due to the instability that arises, a polynomial trendline makes sense. A "simple" linear, power, or logarithmic trendline would not work well as they are either approximately constant or monotonically increasing or decreasing. Only polynomial curves, and maybe some others, like sinusoidal curves allow for such changes. In fact, a sinusoidal curve or a curve that describes a normal distribution might be a better fit. The reason for this is that the experiment was initially done by tossing coins, and as will be explained later, a skewed normal distribution can explain a lot of the results we get.

## 3 Different Preference Levels

To continue our analysis, we considered the fact that each type of agent (by color) might have different preference levels. So, we added modified the program to use two sliders, one slider for the Percent Similar Wanted for type 1 (red agents) and another slider for the Percent Similar Wanted for type 2 (green agents). The code remained mostly the same, with the exception of the update-turtles procedure which after modifications looks as follows:

```
to update-turtles ;; procedure for updating turtles' happiness
  ask turtles
 [
   ;; in next two lines, we use "neighbors" to test the eight patches
  ;; surrounding the current patch
  set similar-nearby count (turtles-on neighbors)
    with [color = [color] of myself]
  set other-nearby count (turtles-on neighbors)
    with [color!= [color] of myself]

  set total-nearby similar-nearby + other-nearby

  if color = red
  [
    set happy? similar-nearby >=
      (%-similar-wanted-type1 * total-nearby / 100)
  ]

  if color = green
  [
    set happy? similar-nearby >=
      (%-similar-wanted-type2 * total-nearby / 100)
  ]
 ]
end
```

Visually, on the interface, the only main difference is that there is another extra NetLogo slider, so that the preference levels of both the types of agents can be changed (can be seen in the appendix).

Now, with these "new variables" to manipulate, we used NetLogo's BehaviorSpace once again. This time, we used the following assignments:

| Variable | Lower Limit or Constant Value | Increment | Upper Limit |
|---|---|---|---|
| number | 1500 | - | - |
| %-similar-wanted-type1 | 0 | 10 | 100 |
| %-similar-wanted-type2 | 0 | 10 | 100 |

It must be noted that, once again, the Percent Similar was measured as the dependent variable and a stopping condition was used, this time with 200 ticks since some simulations seemed to end a few tens of ticks after the 100 ticks mark.

The table with data is given in the Appendix, and the graphical analysis will be explained below.

Using a CAS, we can get many different representations of the data, one "simple" representation is simply a 2-dimensional graph of each of the variables:

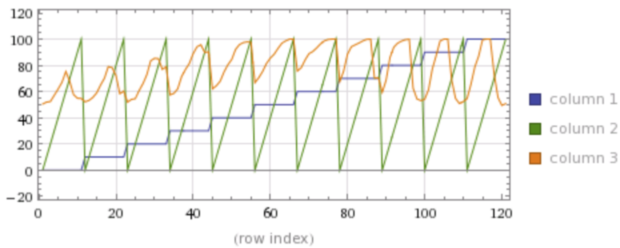



Figure 5: Simple graph of the preference levels and the percent similar obtained

The graph above basically graphs each variable at each simulation. So, the x-axis really represents the simulation number while the y-axis just represents different possible values for each of the variables. Column 1, as can be seen from the table in the Appendix represents the Percent Similar Wanted for the Type 1 agents (red turtles), Column 2 represents the Percent Similar Wanted for the Type 2 agents (green turtles) and finally Column 3 represents the Final Percent Similar wanted seen, either at the point when all the agents are happy or at 200 ticks when the simulation is stopped. Now, while this graph can help see how the variables are related, it does not give a very good representation.

We then tried using another 2-dimensional representation, the bubble map:

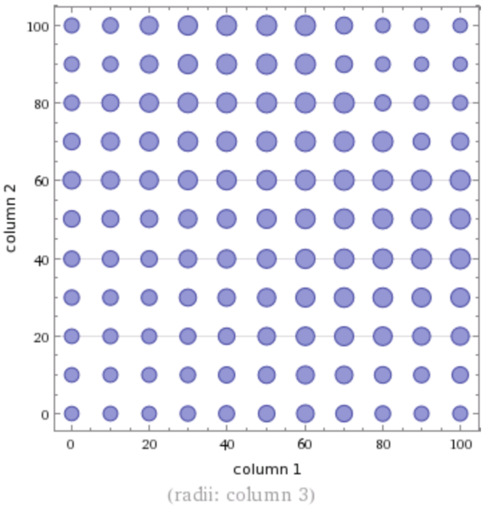



Figure 6: Bubble Map for the two types of agents with different preference levels case

Now, the representation is much more clear. By looking at a particular coordinate, say x = 40, y = 30, we can see the radius of the circle at that coordinate. It has a relatively medium radius compared to all the other circles ("bubbles") which suggests that when the red turtles have a 40% preference to be around turtles of the same color and the green turtles have a 30% preference to be around turtles of the same color, there appears to be a "relatively medium" Percent Similar. These particular conclusions were some that we drew from this graph and found particularly noteworthy:

- A low preference level for either the red or green agents seems to restrict or constrain the final Percent Similar obtained. Graphically, this can be seen by going along one of the "rows" or "columns" of "bubbles" above say for example along x = 10 or y = 0 and observing the radii of the "bubbles." They are approximately the same.

- There are a few lines along which the observations we and Schelling made before can be seen. For example, along y = x (when each agent has the same preference level), we can clearly see that there first appears to be an increase in the radii of the circles as we go along the line and then the radii starts shrinking "after" x = y = 70.

  The line y = x is really a line we have already studied, but some other lines like y = 80 for example show pretty much the same characteristic. y = 80 in particular seems to be particularly interesting since the radius seems to increase rather suddenly and decrease in a similar manner, suggestive of Schelling's "tipping point."

- Apart from low preferences, there are other scenario's where the radius doesn't really change, meaning the line is like a restriction (radius is independent of y or x). Examples of this are the lines x = 50 and y = 50. On both these lines, the radii of the circles do not appear to change based on the "other' variable. That is, on the line x = 50, the value of y doesn't seem to change the radii of the circles, and similarly, on the line y = 50 the value of x doesn't seem to change the radii of the circles. This means that when one of the agents has a 50% preference for having agents of their own color around them, the Percent Similar is determined solely by the fact that they have this 50% preference. The preference level for the other agent type doesn't seem to really matter.

Now, as can be seen from this explanation of the graph, the strength of this representation is that the general trend can be seen quite well, however, the actual results with different levels of preferences are hard to see.

Having said all of the above, the representation of the data is best done through a 3 dimensional graph as we have two independent variables (%-similar-wanted-type1 and %-similar-wanted-type2) and one dependent variable (Percent Similar) rather than a single independent variable and a single dependent variable as the case with the previous graphical analysis and set of simulations where we had a single Percent Similar Wanted. With this said, given below are two 3-dimensional graphs in different perspectives. More graphs in different perspectives are given in the appendix.

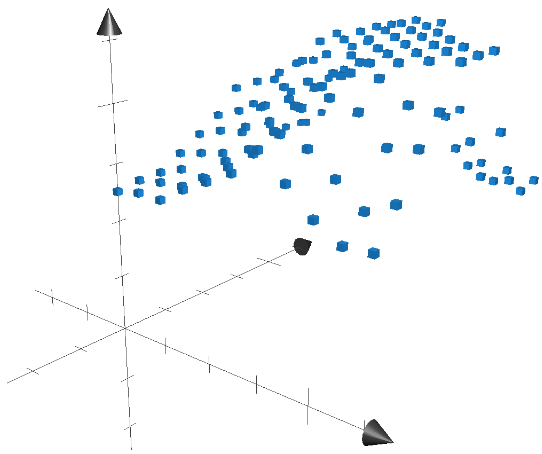


Figure 7: 3-dimensional graph from a perspective where the 3 axes can be seen clearly

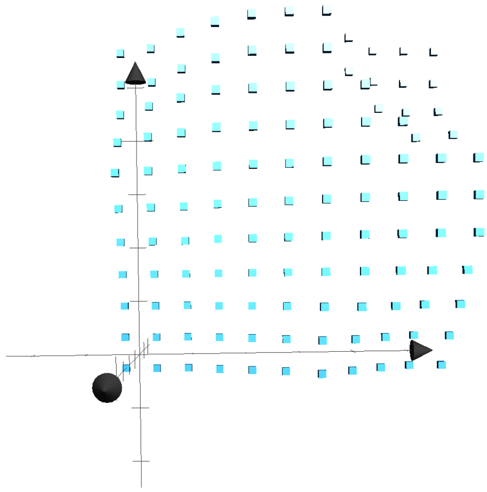


Figure 8: 3-dimensional graph from a top-down perspective where the sharp descent can be seen clearly

From the two graphs it becomes more apparent what the behavior is like. Just like the case where both the types of agents had the same preference level, there seems to be a decline in the percent similar after some point. In the above top-down perspective graph, about 9 points seem to be much "deeper" than the others. These are the points that describe the "unstable" behavior that was described before. When either the green turtles or the red turtles have very high preference levels, the model becomes unstable and the percent similar remains somewhere around 50% throughout the simulation (the agents are really randomly "arranged" and so each agent has about half of it's neighbors with the same color and the other half of it's neighbors have the "opposite" color).

More interestingly, we can see which "combinations" of preferences leads to the most percent similar and least similar. Sorting the table by column 3 in decreasing order, we see that we get 100% Percent Similar when one type has a 50% or 60% preference and the other group has a 90% or 100% preference. So, as an example, if the green turtles have a 100% preference to be around turtles of their own color, and the red turtles have a 50% preference to be around turtles of their own color, we get the final Percent Similar to be 100%, total segregation. This leads to the unintuitive result that the conditions optimal for segregation seem to be when one "group" has a moderate or mild preference to be around agents in it's "group" and the other "group" has a very strong preference to be around those of it's "group." On the other hand, the combinations that provide the lowest final Percent Similar have both the preference levels very low (a few tens) or very high (near 90 or 100). In fact, it appears that when both the types of agents have very high preference levels, the Percent Similar is often lower than when both the types of agents have very low preference levels. So, extreme preference levels (very low or very high) lead to very low Percent Similar values, likely due to the instability that has been talked about before.

To conclude this research into how different preference levels can change the Percent Similar values when we have two types of agents, we found some intriguing statistics, given below:

| - | Column 1 (%-similar-wanted for red agents) | Column 2 (%-similar-wanted for green agents) | Column 3 (Final Percent Similar) |
|---|---|---|---|
| mean | 50 | 50 | 77.009 |
| s.d. | 32 | 32 | 16.982 |
| min | 0 | 0 | 49.384 |
| median | 50 | 50 | 76.921 |
| max | 100 | 100 | 100.000 |

As can be seen from the table, the mean, standard deviation and some other statistics were calculated for each of the columns. The column we are really worried about is the rightmost column. The mean of the Percent Similar across all the simulations is approximately 77%. What this means is that on average, whatever the preferences the two types of agents have, we can guess that the Final Percent Similar will be about 75%. The fact that the mean of the Final Percent Similar is not 50% and is higher suggests that, as Schelling showed, mild preferences really can have a big affect on Segregation. It seems that just by virtue of varying preferences people might have for being around those of their own color/ethnicity, on average 75% segregation occurs. Furthermore, the standard deviation of about 15% seems to mean that by "changing" the preferences of people, the Percent Similar can only really be affected by about 15% both upwards and downwards. This gives a range of about 60% to 90% segregation using $\pm 1 s.d.$

# 4 Three Agent Types

So far we have only used 2 agents in the model. In real life, there are usually more than two types of agents in any scenario. So, we added another agent to the model, and given below are a brief description of the model, results and discussion.

First, we tried a model where there were 3 types of agents, identified as blue agents, red agents and green agents, and every agent had the same preference level. This is analogous to the first model we used where we had two agent types and one preference level for all the agents, the only difference being that now there were three agent types and one preference level for all of them. So, also the interface for this model is identical to the interface for the first model, with the exception of there now being blue colored agents. Now, using BehaviorSpace, and graphing software, we plotted the following graph:


Figure 9: Graph for the Three Agent Types Model with Homogeneous Preferences

The trendline plotted definitely seems to over-fit the data, however, lower order polynomial trendlines seemed to have a very low Coefficient of Determination ($R^2$). Nevertheless, what we are really looking for is the general behavior. In that respect, the graph above is very similar to the one given on Page 7 (Figure 4). In fact the two graphs look almost identical. There still appears to be a decline in the Percent Similar after a peak value, this time at a lower Percent Similar Wanted value, about 60% (solving derivative of trendline $= 0$). We believe the reason for this is that the "instability factor" becomes important at a lower Percent Similar Wanted value when we have an extra agent type because higher levels of preference are harder to achieve with three agents rather than two (having almost all neighbors of the same type is less likely).

So, having an extra agent doesn't really seem to affect the behavior of Percent Similar as a function of Percent Similar Wanted other than the fact that the curve's peak has shifted to the left slightly.

Next, we gave each of the types of agents different levels of preference. The interface for this model can be seen in the Appendix and the major part of the code used is given below:

```
to update-turtles ;; procedure for updating turtles' happiness
  ask turtles
  [
    ;; in next two lines, we use "neighbors" to test the eight patches
   ;; surrounding the current patch
   set similar-nearby count (turtles-on neighbors)
     with [color = [color] of myself]
   set other-nearby count (turtles-on neighbors)
     with [color!= [color] of myself]

   set total-nearby similar-nearby + other-nearby

   if color = red
   [
     set happy? similar-nearby >=
       (%-similar-wanted-type1 * total-nearby / 100)
   ]

   if color = green
   [
     set happy? similar-nearby >=
       (%-similar-wanted-type2 * total-nearby / 100)
   ]
   if color = blue
   [
     set happy? similar-nearby >=
       (%-similar-wanted-type3 * total-nearby / 100)
   ]
  ]
end
```

The BehaviorSpace settings we used were:

| Variable | Lower Limit or Constant Value | Increment | Upper Limit |
|---|---|---|---|
| number | 500 | - | - |
| %-similar-wanted-type1 | 0 | 20 | 100 |
| %-similar-wanted-type2 | 0 | 20 | 100 |
| %-similar-wanted-type3 | 0 | 20 | 100 |

Now, it must be noted that in these simulations we have three independent variables (the preference level of each of the "group") and one dependent variable (the final Percent Similar). This brings to question how to interpret and present the data, as the ideal way to represent the data would be to use a 4-dimensional plot. However, a 4-dimensional plot cannot be easily represented on a 3-dimensional plot, let alone a 2-dimensional plane. The method we decided to use was a 3-dimensional graph with the x-axis representing the Preference Level for agents of type 1, y-axis representing the Preference Level for agents of type 2, and the z-axis representing the Preference Level for agents of type 3, with color to represent the final Percent Similar. The graph is shown below from a 2-axis perspective, other perspectives for the same graph can be found in the Appendix:

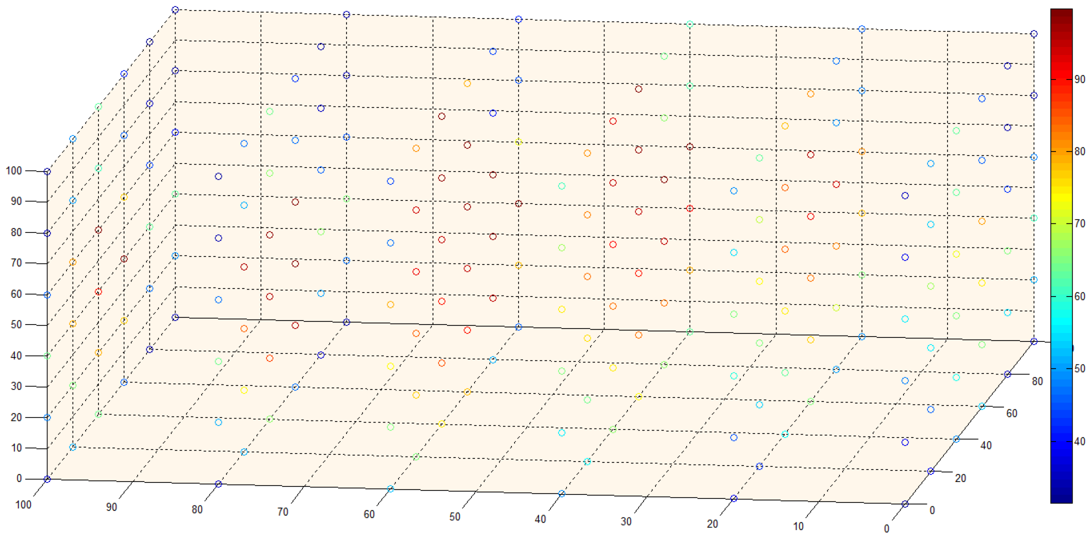



Figure 10: 3-axis perspective of Three Agent Types each with different Preference Levels

The above graph and the ones given in the Appendix all certainly give a representation of the data, however, they are not really good representations for understanding the data and having more than two independent variables generally means that graphical representations aren't quite the useful way to decipher the patterns behind the data. So, rather than continuing to graphically represent the data, we chose to get statistics for the data and use the table itself to understand the data.

First, we can understand which combinations of Preference Levels lead to higher Percent Similar values simply by arranging the table in a descending order based on the column 4 values (Percent Similar Column). Doing so, we see that combinations, represented by (Preference Level for Type 1 Agents, Preference Level for Type 2 Agents, Preference Level for Type 3 Agents) like: (60, 80, 60), (80, 60, 60), (40, 60, 100), & (60, 60, 80) give near 100% Percent Similar. On the other end, combinations like: (80, 100, 100), (80, 80, 100), (0, 100, 100), & (0, 80, 100) give about 30% Percent Similar values.

On looking at the statements given above and the sorted table not many conclusions can be made, in fact even a multivariate regression analysis did not yield much. We tried a multivariate regression analysis since we have three independent variables, and we got the following equation:

$$y = \alpha + \vec{x} \cdot \vec{\beta} + \epsilon, \alpha \approx 64.60, \beta_1 \approx -0.00709, \beta_2 \approx 0.00038, \beta_3 \approx -0.00521$$

The equation turns out to have a Coefficient of determination of approximately $2.16x10^{-4}$, adjusts to -0.0139. This is certainly very low; the near zero value in fact suggests that the equation obtained has almost no correlation with the data.

We then tried using the mean, median and mode of the Preference Levels, the mean seemed to give a better picture. The graph is given below for the Mean of the Preference Levels Vs. the Final Percent Similar.

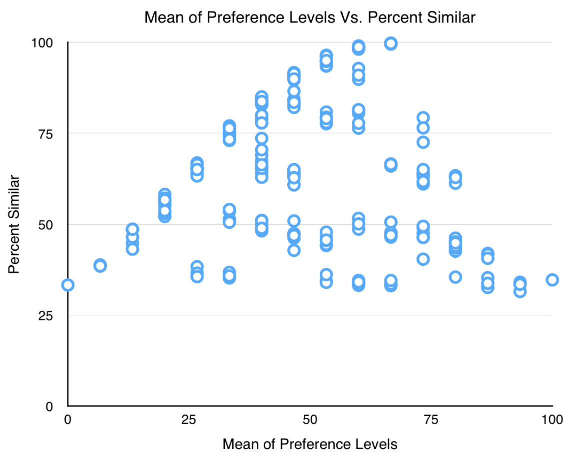



Figure 11: Mean Preference Level for the three agent types Vs. Final Percent Similar

At first, the above graph doesn't seem to really illuminate many results, however, on closer analysis, some interesting facts emerge. Firstly, it must be noted that the graph above in fact has many similar characteristics to the graphs plotted before. Looking at the "topmost" values for every x-axis "point" we observe the familiar increase followed by peak followed by decrease. This indeed suggests that the instability factor plays a role here too. The Percent Similar increases until a peak after which the model becomes unstable. Next, it must be noticed that each of the x-axis "values" corresponds to some set of points all with the same x-coordinate but different y-coordinates. Looking at x = 0 for example leads to a single point at x = 0 and approximately y = 35. Similarly, looking at say x = 20 gives about 5 points all around y = 40. So, this suggests that with a mean preference level of 20, we get somewhere around a 40% Percent Similar. x = 60 on the other hand leads to a much more widespread possibilities for y. So, with a 60% as the mean of the three "percent similar wanted's", we can get from somewhere around 20% to 100% Percent Similar.

## 5 Gradualist models

In all the previous models, we have used a threshold function. The real world may really be best modeled through a gradualistic model as suggested by Bruch & Mare. To this extent, we started by first using a two agent model with the percent similar wanted being set not by a single slider, but by a set of two sliders; one for the mean of the Percent Similar wanted and the other for the standard deviation of the Percent Similar Wanted. The interface thus looked as follows (at setup):

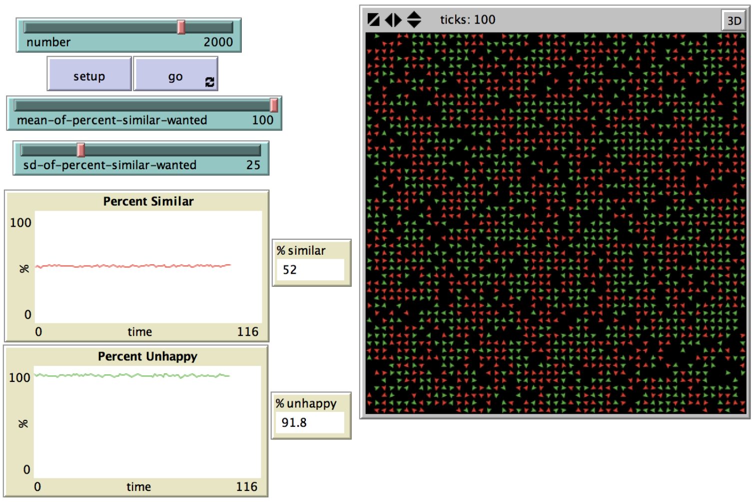



Figure 12: Interface for the Two Agent Gradualistic Model

We used BehaviorSpace with the following settings to do analysis on the model:
["sd-of-percent-similar-wanted" [0 5 25]]
["mean-of-percent-similar-wanted" [0 10 100]]
["number" 2000]

We then looked at graphing the data obtained. The obvious way to represent two independent variables with one dependent variables is to use a three-dimensional graph. So, we first did that. The graph obtained follows:

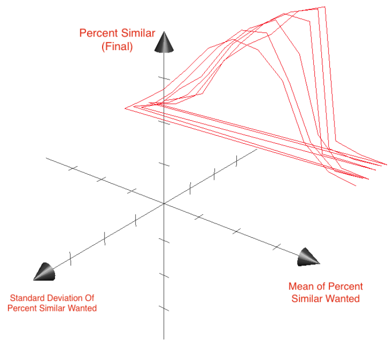



Figure 13: 3D visualization of Two Agent Type Gradualistic Model

As the case with most 3-D representations, some interesting conclusions can be made, but, we are better off using a 2-dimensional representation. So, we ended up "slicing" the graph using some xz and yz planes to get two-dimensional graphs.

Two of the "Mean Percent Similar Wanted Vs. Percent Similar" graphs obtained follow (with the rest being in the Appendix):

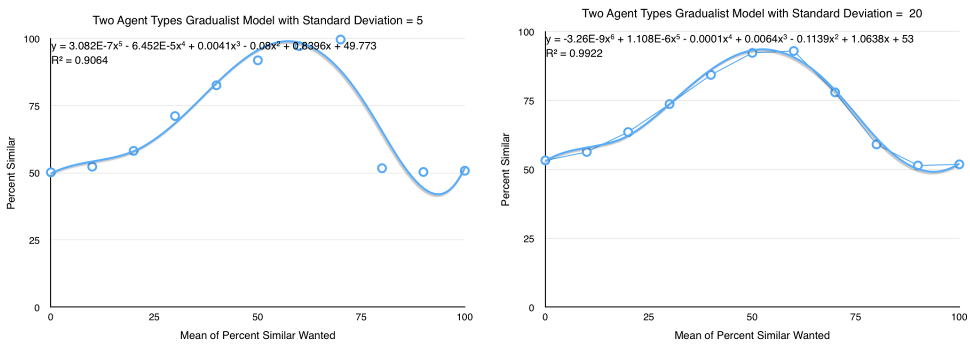



Figure 14

Two graphs can be seen above. The graph to the left plots Mean of the Percent Similar Wanted values to the Final Percent Similar with the standard deviation of the Percent Similar wanted being fixed at 5%. The second graph to the right shows the plots Mean of the Percent Similar Wanted values to the Final Percent Similar with the standard deviation of the Percent Similar wanted being fixed at 20%.

Both the graphs have the familiar monotonous increase, followed by peak and monotonous decrease. There also appears to be a slight shift left of the peak, but it seems negligible. The important thing to note is however how the "discontinuities" disappear when going from a standard deviation of 5% to 20%. The "discontinuity" in particular that we are talking about is the sharp fall after the peak in the s.d. = 5% graph which doesn't seem to appear in the graph with s.d. = 20%. The increase in the standard deviation seems to smooth out the graph. The reason for this smoothing out effect appears to simply be caused by the fact that the Percent Similar Wanted values in this model are not the same for everyone. Some agents have a higher Percent Similar wanted while others have a lower Percent Similar Wanted. An increase in the standard deviation increases the range of the Percent Similar Wanted values agents have. With a higher standard deviation, the Percent Similar Wanted values vary a lot more and this causes more agents to have higher (and lower) Percent Similar Wanted values than the mean. With a low standard deviation, more agents have Percent Similar Wanted values close to the mean, and the "instability" affect described before comes into affect at high "Mean of Percent Similar Wanted values." With a high standard deviation, there are more agents with low Percent Similar wanted values (below the mean and low in general). All of these agents with low Percent Similar wanted values cease to move as they are happy with most any configuration of neighbors compared to those agents with high Percent Similar Wanted values. So, there is less movement overall, and the "instability effect" doesn't play as a major role in determining the final Percent Similar value.

Now, the above graphs and explanations all "lie along the xz planes," i.e, the explanations arise from slicing the 3-d plot obtained previously into slices that each represent some standard deviation. To continue the analysis of the model, we can now "slice the graph along the yz plane." What this means is that we observe the Percent Similar as a function of the Standard Deviation. We already have talked about how the standard deviation affects the Percent Similar values, however, an analysis of greater detailed where we go through a sequence of standard deviations we believe is quite useful.

Given below is one the graphs obtained (the rest being in the Appendix). The graph plots the Standard deviation of the Percent Similar Wanted against the Percent Similar obtained with a fixed mean of Percent Similar Wanted at 60%:

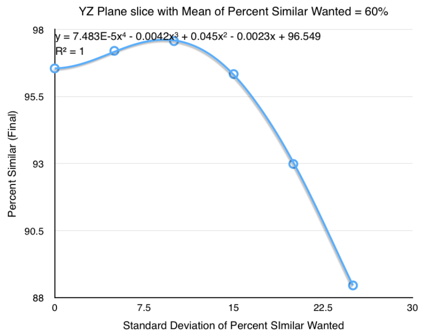



Figure 15: YZ Plane Slice with Mean of Percent Similar Wanted = 60%

Looking at the graph above, we can clearly see that there appears to be the same characteristic tipping point. This is quite a different tipping point though. Interestingly it seems that some standard deviations lead to an increase in the percent similar values while others lead to a decrease. From about 0 to 10% s.d. the percent similar increases as the s.d. increases. This means that with a wider variation in percent similar wanted values, more segregation is caused. However, after about 10%, the percent similar seems to decrease. Looking at the other graphs in the appendix, it becomes clear that this peak s.d. changes with different means, however, the relationship between the two seems almost random as regression analysis and curve fitting do not seem to be able to find a clear relation. So, it seems that for each mean value of percent similar wanted, there is a point until which an increase in the standard deviation in fact increases the percent similar. The reason for this may well be that NetLogo somehow seems to produce random numbers skewed above the mean, however, this is unlikely as we observed this behavior on multiple runs and checked the random numbers being generated by NetLogo. The only other explanation we could find was related to the now oft-talked about "instability affect." When the s.d. increases, there is more variability (a wider range) of the percent similar wanted values. This keeps some agents moving much more than others. There is a s.d. after which there are too many agents who are standing still, and thus stabilizing the model, leading to a decrease in the percent similar.

The implications of the previous conclusions is a very interesting one. Having varied preference levels among a population can actually cause segregation. This seems obvious yet unintuitive. Having more people with lower preference levels (and higher preference levels) leads to more segregation that having everyone at near the same preference level. The reason for this seems complex. Other than the fact that having more agents with lower preference levels leads to more stability, there is also the fact that having homogeneous preferences leads to more "conformity and uniformity," either everyone is happy or unhappy, there is less chance for variations in happiness.

Now, extending the model to other scenarios seems to be redundant. In the model with different preference levels for each type of agent, and the model with three rather than two agents (and the corresponding "different preference levels" model) there do not appear to be many more results that we have not already discussed. So, to avoid redundancy, we just discuss the major conclusions we derived from these two models below.

In the model with different preferences for each agent type, with 2 "mean percent similar wanted values" and 2 "standard deviation for percent similar wanted" sliders we end up having 4 independent variables. Graphical analysis becomes very complex, so, we used a regression analysis graphing different independent variables with one dependent variable. The results we got were that having a larger and more "closer" standard deviation leads to a smoothing out of the percent similar (rather than the sharp decline seen before). Having means that are mid-to-high (near 60-80%) and "closer" leads to higher percent similar values (similar to the result in section 3).

In the model with three agents and a single mean preference slider with a single s.d. preference slider, we get results almost identical in behavior to the two agent single slider for mean and single slider of s.d. case.

In the model with different preferences for each agent type, with 3 "mean percent similar wanted value" sliders and 3 "standard deviation for percent similar wanted" sliders, we get results similar to the two agent types with 2 "mean percent similar wanted values" and 2 "standard deviation for percent similar wanted" sliders case. With 9 independent variables, graphical analysis becomes very complex, so, we used a regression analysis graphing different independent variables with one dependent variable. Again, the results we got were that having a larger and more "closer" standard deviation leads to a smoothing out of the percent similar (rather than the sharp decline seen before) and having means that are mid-to-high (near 60-80%) and "closer" leads to higher percent similar values (similar to the result in section 3).

The aim of all of this research is really to identify potential solutions to real world problems; in this case, "How do we reduce segregation and increase integration?" So, to conclude our research into gradualistic models, we began by using Bruch's segregation model and identifying conditions we would need to reduce segregation. The Bruch model initially, after setup, looks as follows:


Figure 16: Bruch's segregation model after setup

As can be seen above, Bruch's model uses a few extra variables. The important one which we are really worried about is the "utility" drop-down. Using this model and starting with a segregated setup, we can try to find scenarios where we get the least percent similar in the "Threshold" case and in the "Continuous" case.

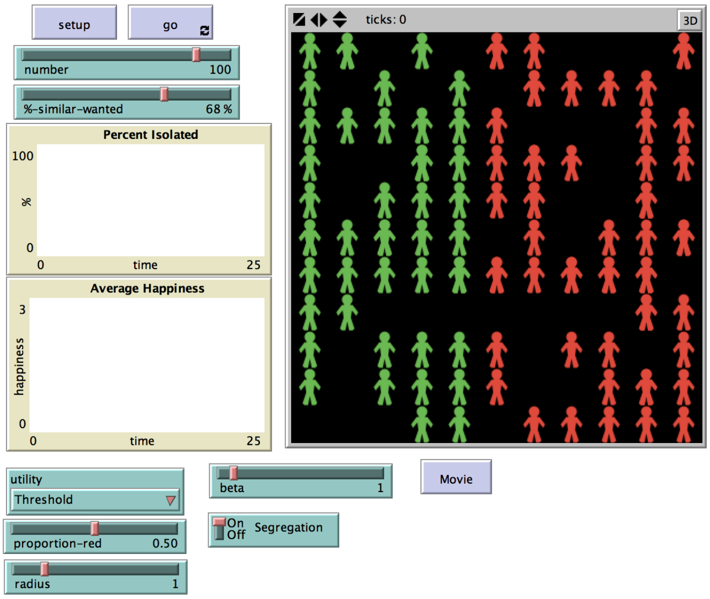



Figure 17: Bruch's segregation model after setup with Segregation Switch set to on

Just to illustrate how the setup looks with the additional switch we added "Segregation" turned on, here is the interface after setup with the switch turned on:

After much trial and error, we came up with the following conclusions, listed as a table:

| Variable | High/Low For Higher Integration with Threshold | High/Low For Integration with Continuous |
|---|---|---|
| Percentage Similar Wanted | Low | Low |
| beta | - | Low |
| Radius | High | High |

Note that beta is a variable that seems to be describing the utility of patches. The higher the beta value, the higher the utility, causing more segregation. beta is not a factor considered in the Threshold simulations, hence the hyphen above. Finally, it must be noted that there is a drop-down to choose the movement type, best vs. probabilistic. Both of the movement types seem to lead to an overall same level of segregation as a probabilistic movement scenario leads to certain agents sticking to their positions if they are happy.

# 6 Extending the Model Further

To continue our research into how preferences of agents can determine the final Percent Similar, we first started with a simple experiment and removed the "color variable."

Each agent is given a list of qualities. Initially, we started with a Boolean to represent each quality. So, an agent might have say some list like [1 0 0 ... 1]. Then, in this simple model, we used a single Percent Similar wanted much like the "initial" model. Also, we varied a variable called "Number of Qualities Similar Wanted."

Now, before discussing the simulations conducted, we believe it is important to bring up a point that we had to think of us: "What exactly are we trying to measure?" In the case where we had a color associated with each agent, the answer was easy, how many of a turtles neighbors on average have the same color as it? In other words, we wanted a measure of how segregated the "final" world is (when all the agents are "happy"). In this scenario where color is not really a variable we are worried about, we questioned what exactly we wanted to measure. We decided that the answer is really how similar those are around you at the end state. So, we measured the average number of qualities in common. So, if we get a value of 1.2 when agents want 50% of their neighbors to be similar to them and agents consider those with two of three qualities in common "similar" to them, this means that with the agents "wants," agents have neighbors with roughly 1 quality in common. Note that though the graphs are labeled "Percent Similar" on the y-axis, a better wording is possibly "Number of Qualities Similar on Average among neighbors." Using this methodology, we obtained the following graphs with Number Of Qualities = 3.

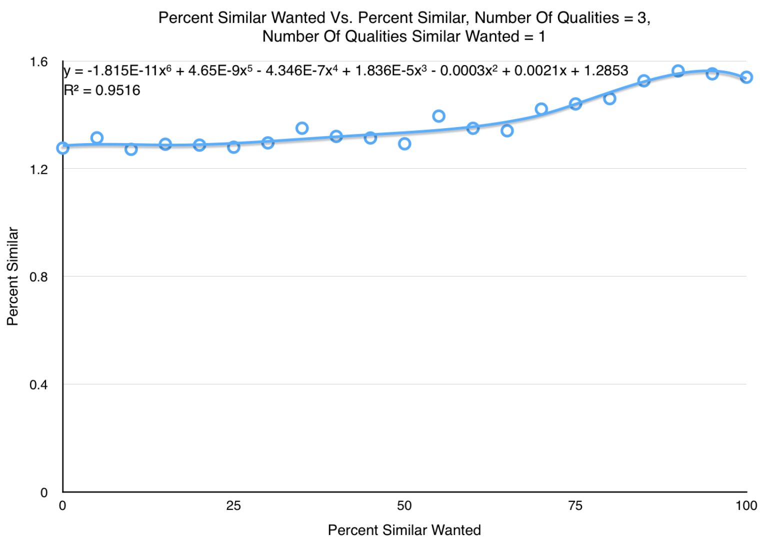



Figure 18: 3 Qualities, # Of Qualities For Similar: 1

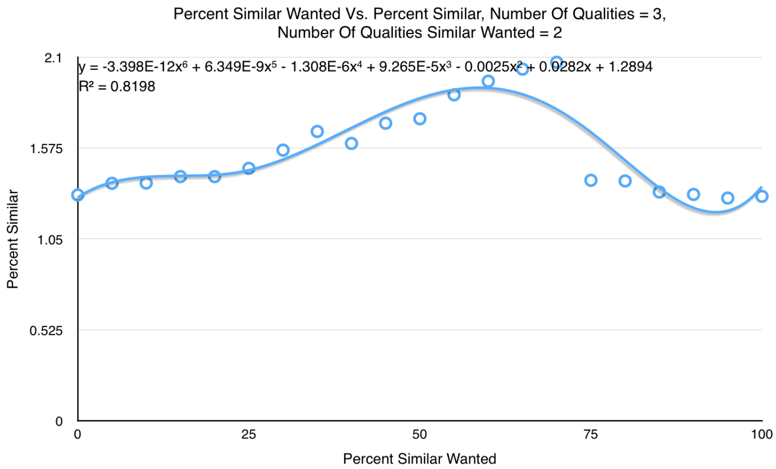



Figure 19: 3 Qualities, # Of Qualities For Similar: 2


Figure 20: 3 Qualities, # Of Qualities For Similar: 3

It is important to understand what the graphs exactly represent. The first graph, on the previous page, represents the Percent Similar obtained for different Percent Similar Wanted values when agents consider those around "similar" when they have at least 1 quality in common. Similarly, the second graph, the top one on this page, represents the Percent Similar obtained for different Percent Similar Wanted values when agents consider those around "similar" when they have at least 2 qualities in common. And without a loss of generality, the same applies to the graph above, replacing the at least condition with 3 qualities.

All the graphs show an increase in the Percent Similar as a function of Percent Similar wanted. However, the first graph seems to have no peak followed by descent while the second and third graphs seem to have both a peak and descent. On a closer look, it can be noticed that in fact, the first graph also has a peak, it is very high though, near percent similar wanted = 90 and there is a shallow descent. So, all graphs have a monotonically increasing portion followed by a peak and then a descent. The interesting feature of each of the graphs is the location of the peak. As the "Number of Qualities to be considered similar increases," the peak shifts to the left and increases in "amplitude."

The increase in amplitude makes sense because as the "Number of Qualities to be considered similar" increases, the Percent Similar (better phrased as the average number of qualities in common among neighbors) should increase in general.

The shifting to the left can be explained through coins as hinted to before. Initially, Schelling conducted experiments using coins on a chessboard as the main model. Using that analogy here, an agent basically gets a set of qualities based on an initial set of coin tosses, in this case exactly three coin tosses. These coin tosses can be thought of as sums. If an agent's qualities are given by three heads, we can consider that to be 3. If an agent's qualities on the other hand is given by three tails, we can consider that to be 0. So, when an agent looks at it's neighbors and compares the Booleans of it's list with the Booleans of the others' lists, it basically is computing a sum. The probability that a neighbor has 1 quality in common is much higher than the probability that an agent has 3 or even 2 qualities in common. So, we are basically talking about a random variable, say X, the sum of common qualities, often simply called the hamming distance. The random variable is distributed along a simple normal distribution. Checking when $X \geq 1$ "describes" the first graph, $X \geq 2$ "describes" the second graph and $X \geq 3$ "describes" the third graph. In other words, the graphs can be described as behaving like "skewed normal distributions."

Now, having said the above, there is another interesting thing to note, when the agents wanted 1 quality in common with their neighbors, they got upto 1.6 qualities in common with their neighbors on average, a higher number. When the agents wanted w qualities in common with their neighbors, they got upto 2.1 qualities in common with their neighbors on average, a higher value (by a small amount). Finally, when the agents wanted 3 qualities in common with their neighbors, they got upto 2.6 qualities in common with their neighbors on average, a lower value. What these conclusions say is that as the number of qualities similar wanted increases, the final outcome of number of qualities similar on average with neighbors even in the "best" case seems to be below the agents' preferences. This seems to bring a point that we have repeatedly visited, high expectations, or specifically preferences, of the agents often leads to lower "similarity measures" maybe due to the instability discussed many times above or indeed some other factors that we might have overlooked.

In real life, there are usually more than three qualities associated with an agent, say a person. To verify our results with a more realistic example, we considered a model where every agent had 50 qualities. The graphs are given in the appendix. The results seem to apply to a large extent, however, there are some facts we thought were worth mentioning.

With 50 qualities, the simulation settings we used originally took a very long time, calculated to be about 9 hours. As the case with most other simulations, we had to reduce the scope and increase the increment in our BehaviorSpace to reduce the time taken. We still managed to only reduce the simulation run time by about 5 hours. Nevertheless, it is worth mentioning that time and computing constraints are indeed an issue we faced. Having said that, the first few hundred simulations finished instantly. This is because with "Number of Qualities to be considered similar" around 10 or 20, the percent unhappy turns out to be 0% (with percent similar hence near 100%) as every turtle seems to have only "similar" turtles around (neighboring turtles all have 10/20 qualities in common). So, the simulations never even runs. On the other end, the last few hundred simulations ran until the stopping condition of 200 ticks that was used. This is because with "Number of Qualities to be considered similar" around 40 or 45, the percent unhappy turns out to be 100% (with percent similar hence near 0%) as every turtle seems to have only "non-similar" turtles around (none of the neighboring turtles ever have 40/45 qualities in common).

Apart from these observations, the results observed in the case with 3 qualities largely apply to the case with 50 qualities; there still appears to be an increasing in the "amplitude" of the peak and a shift left of the peak as the "Number of Qualities to be considered similar" increases.

Before continuing, it is probably a good idea to compare this to "color segregation," as that is what we started with. Though the general characteristic of the graphs appears to be the same, it is important to note that some distinctions can surely be made among the two, however small. Looking back at Figure 4, we see some clear differences. The "color segregation" graph appears to be much smoother having almost no discontinuities. On the other hand, many of the "qualities graphs" seem to have discontinuities. One example of this is Figure 20 which clearly has a discontinuity at about Percent Similar Wanted = 40%. The reason for these discontinuities is not apparent, however, Figure 18 where we have "Number of Qualities Similar Wanted" set to 1 is really very much similar to the "color segregation" model and the Figure 18 doesn't appear to have any discontinuities. This suggests that when the "Number of Qualities For Similar Wanted" is set to 1, the model is really very similar to the "color segregation" model, the exception being that instead of having 1 quality (or color) for agents to be considered similar, we have 1 or more common qualities as being the condition for agents to be considered similar. The discontinuities occur in the 2 and 3 qualities for similar cases probably because of some properties of the "skewed normal distributions" talked about before. This also importantly suggests that increasing the "Number of Qualities For Similar Wanted" plays a major role in determining the final Percent Similar. So, every quality probably does matter in real life too when forming relationships. There is however a question of whether every quality should be considered with the same weight.

To continue our analysis of this "qualities model," we thought about what kind of things the model doesn't really take into account. Going back to the threshold vs. gradualist argument, that becomes important here. In real life, rarely are people influenced in a threshold manner. So, using a gradualist model in conjunction with the "qualities model" become our next aim. We found that there were many ways to implement the gradualist model. We first decided to use a model where an agents probability of moving would be correlated to the average hamming distance of the neighbors' qualities lists to it's qualities list.

The user interface for this model is identical to the ...
The major part of code used is as follows:

Now, we used BehaviorSpace with the following settings:

["Number_Of_Qualities" 50]
["number" 1500]
["Number_Of_Qualities_For_Similar_wanted" [10 2 30]]
["%-similar-wanted" [0 10 100]]

The results were very interesting. No matter what setup we used, the final percent similar similar is always about 21% and the model almost never ends. We worked a lot on figuring out why this is the case. What we ended up concluding is that making the "move" probabilistic or stochastic, keeping an unhappy turtle in a fixed spot causes instability leading to a constant percent similar and a model that doesn't end, very much like the scenario of very high percent similar wanted values in Schelling's model.

The next model we used was one with a real number values between 0 and 1 (including) for each agent as the sole quality (and no color variable). The agents would move in correlation to the mean (absolute) distance of it's quality. So, if an agent has 0.5 as it's quality and it's 2 neighbors have 0.8 and 0.1, the value calculated would be avg(abs(0.8 - 0.5) + abs(0.1 - 0.5)) = 0.35. Now, a random number between 0 and 1 would be chosen and if the number is greater than the value obtained and the agent is unhappy, the agent would move. Surprisingly, we got the exact same results. This time, the percent similar similar was always about 20%. Again, keeping an unhappy turtle in a fixed spot causes instability leading to a constant percent similar and a model that doesn't end, very much like the scenario of very high percent similar wanted values in Schelling's model.

Now, the natural thing to do was combine the "color segregation" and "qualities segregation" models. To do this, we added a "weightage" variable where a user could set how much "qualities" matters over "color" for the agents. A weightage of 1 signifies that "qualities" are the sole determiner of whether an agent considers other agents around "similar" to itself. A weightage of 0 sgnifies that "color" is the sole determiner of whether an agent considerd other agents around "similar" to itself. Rather than continue explaining the functions now, we feel that the code really "speaks for itself." A major part of our code is given in the Appendix.

These were the BehaviorSpace settings we used:

```
["Weight_Of_Qualities" [0.2 0.2 1]]
["num_of_qualities_similar_wanted" [20 20 80]]
["mean-time-change" 0]
["mean-turtle-quality-changes" 0]
["number" 1000]
["time-change-sd" 0]
["turtles-quality-change-sd" 0]
["number-of-qualities" 100]
["%-similar-wanted" [10 20 90]]
```

All the graphs obtained are in the appendix. The results obtained will be discussed here.

Firstly, each "set" of graphs contains 4 graphs. All the graphs in one "set" have the same weightage. So, for example, in the first set of graphs, we have weightage set to 0.2. One of the graphs plots Percent Similar Wanted Vs. Percent Similar with "Number of Quantities Wanted for similar" set to 20. Another graph plots Percent Similar Wanted Vs. Percent Similar with "Number of Quantities Wanted for similar" set to 40. Yet another graph plots Percent Similar Wanted Vs. Percent Similar with "Number of Quantities Wanted for similar" set to 60. Finally, another graph plots Percent Similar Wanted Vs. Percent Similar with "Number of Quantities Wanted for similar" set to 80.

Now, most of the graphs (particularly the ones with low weightages) still show all the characteristic monotonous increase, followed by "peak/tipping point," followed by decline. The first two graphs of each set of graphs look the same. Interestingly, the next two graphs almost always show a "Flipping effect." This effect has been observed before, explained as a "shift" to the left of the peak, however, in these sets of graphs, the flipping is very apparent. To emphasize again, this effect is seen repeatedly. We believe that this shows that Number of Qualities Wanted is a major factor in the model. Also, it looks like a Skewed Normal Distribution, as described before, can be thought of using coins being tossed multiple times and calculating the sum.

Looking at the last two sets of graphs, weightage now above 0.5. So, qualities matter more than color at this point. There is also a general direct quadratic/biquadratic variation between percent similar wanted and percent similar. The "Flipping Effect" is still visible here as the weightage is so high making the number of similar qualities wanted matter much more. Another effect, what we call the"Matching Effect" was also observed.I h effect can be described as follows: "if Percent Similar Wanted is "greater" than the P(getting a match), percent similar increases rather "quickly."

So, now, with some major observations stated and "effects observed," here are some general conclusions from these simulations: The Contact Hypothesis seems to be modeled (and verified) as in this model, when agents are not "classified" or set to happy/unhappy solely based on color, Instabilities seem to play a major role in the final outcome. Some rather unintuitive phenomenon like the "flipping effect" and "matching effect" were seen.

Before concluding, it is worth mentioning that we also worked on a few other models. First, we tried to understand what effect minorities/majorities have on all of the models. To this extent, we tried the "original" Schelling model with different proportions of red and green agents. A graph obtained is shown below:

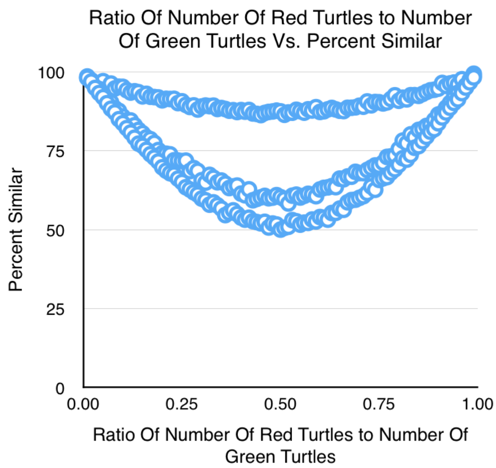



Figure 21

As can be seen above, as the ratio of the number of green turtles to number of red turtles increases, the percent similar decreases. Moreover, the lowest percent similar is at 50% of each type of agent in the population. This is a very intuitive result. However, the unintuitive result is that this relation is quadratic with respect to the Percent Similar Wanted at 50%, but near constant at percent similar wanted 0% and 100%.

Finally, to illustrate another variable we didn't really account for, we decided to make a user friendly model where a user can set off global / local events. The user interface at setup looks as follows:

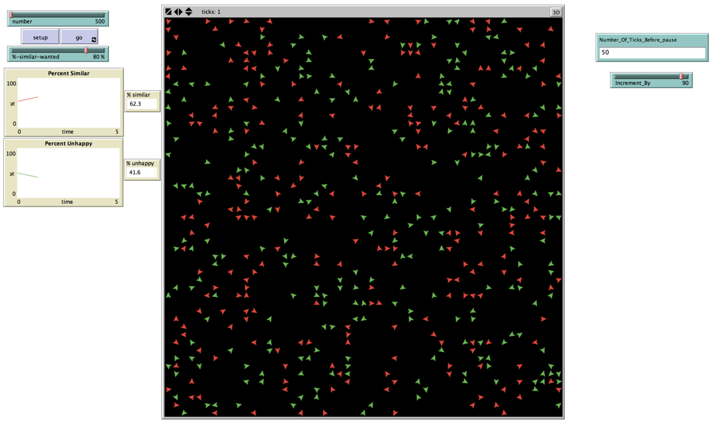



Figure 22

As can be seen above, the user can choose a number of ticks interval to be questioned about creating events. Also, the user can set the "increment by" to increase the percent similar wanted (or decrease it) by some set value.

After these settings are set, the user then clicks "Go Forever" and after the number of ticks interval has passed, a user dialog opens:

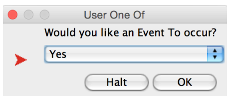
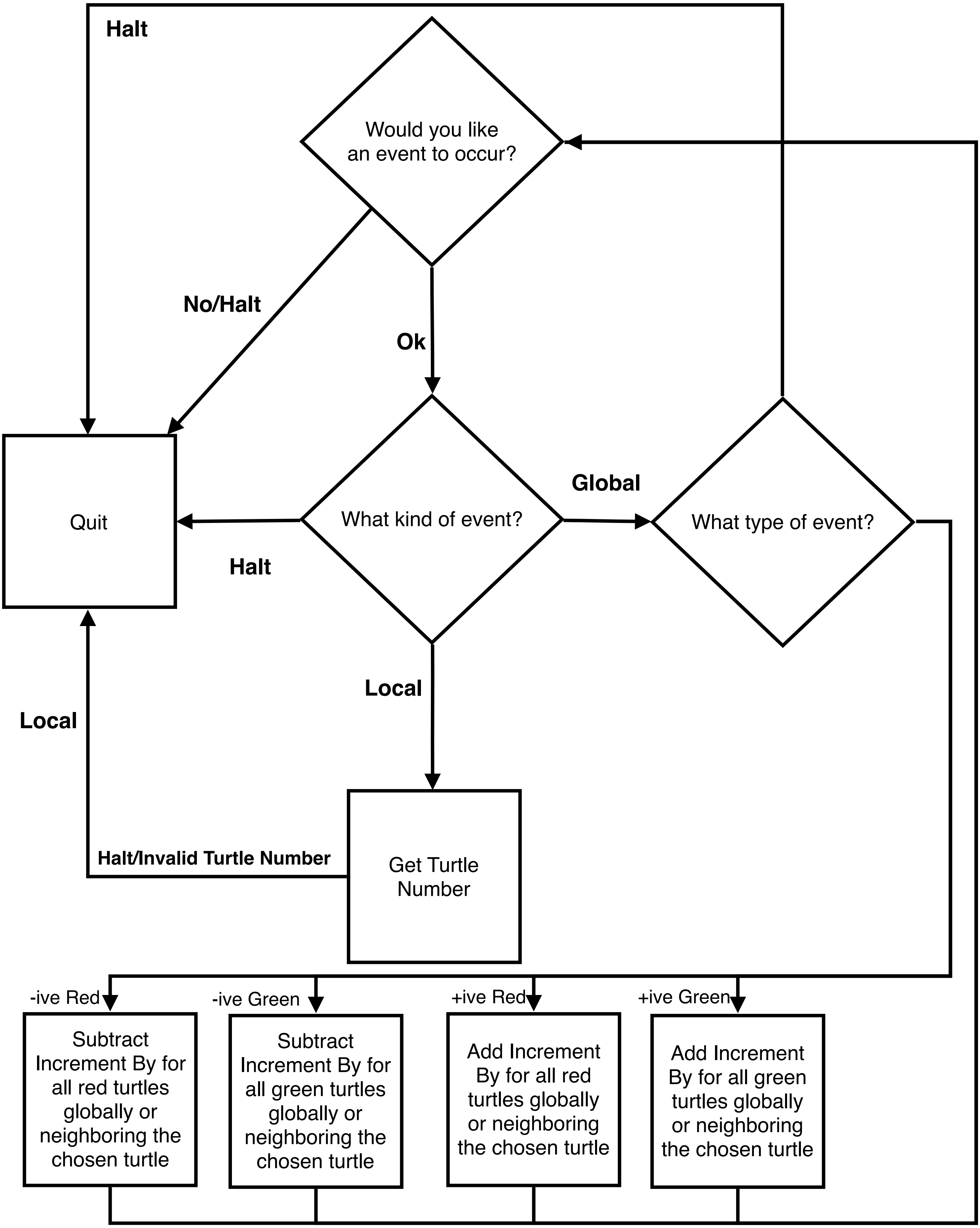



Figure 23

Then, a set of user dialogs follow based on the user responses. A flow chart is given on the following page for this.


Halt
Would you like an event to occur?
No/Halt
Ok
Quit
What kind of event?
Global
What type of event?
Halt
Local
Local
Get Turtle Number
Halt/Invalid Turtle Number
-ive Red
-ive Green
+ive Red
+ive Green
Subtract Increment By for all red turtles globally or neighboring the chosen turtle
Subtract Increment By for all green turtles globally or neighboring the chosen turtle
Add Increment By for all red turtles globally or neighboring the chosen turtle
Add Increment By for all green turtles globally or neighboring the chosen turtle

The flowchart on the previous page describes the algorithm pretty well, however, just to show how local events look to the user, here is a screen shot of what the user sees after creating a local event:

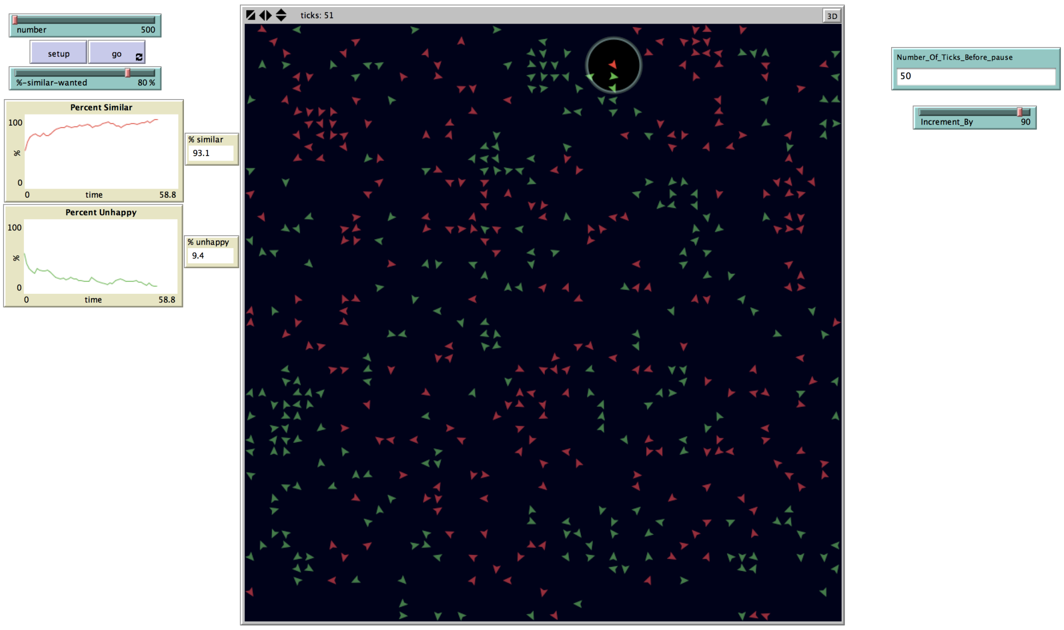



Figure 24: User view after local event

As can be seen above a local event creates a "watch view" on the selected turtle.

# Appendix

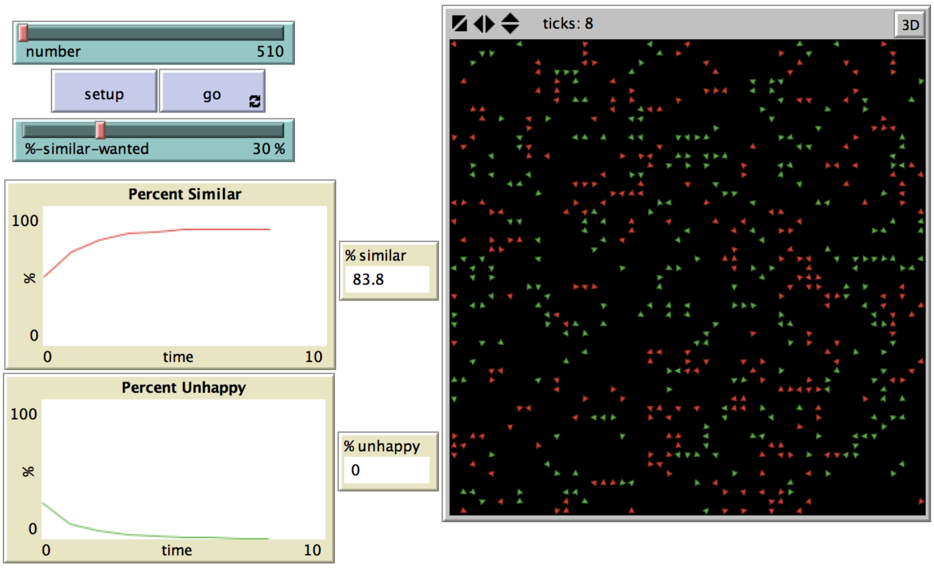



Figure 25: An illustration of how the percent similar increases as the number of agents decreases

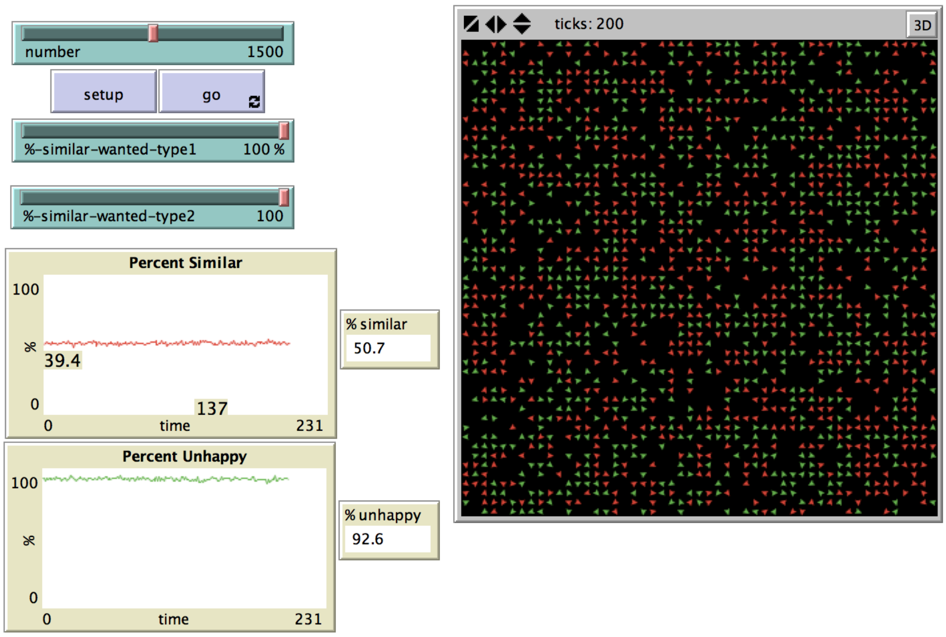



Figure 26: The interface for the Two Agents with different preferences model

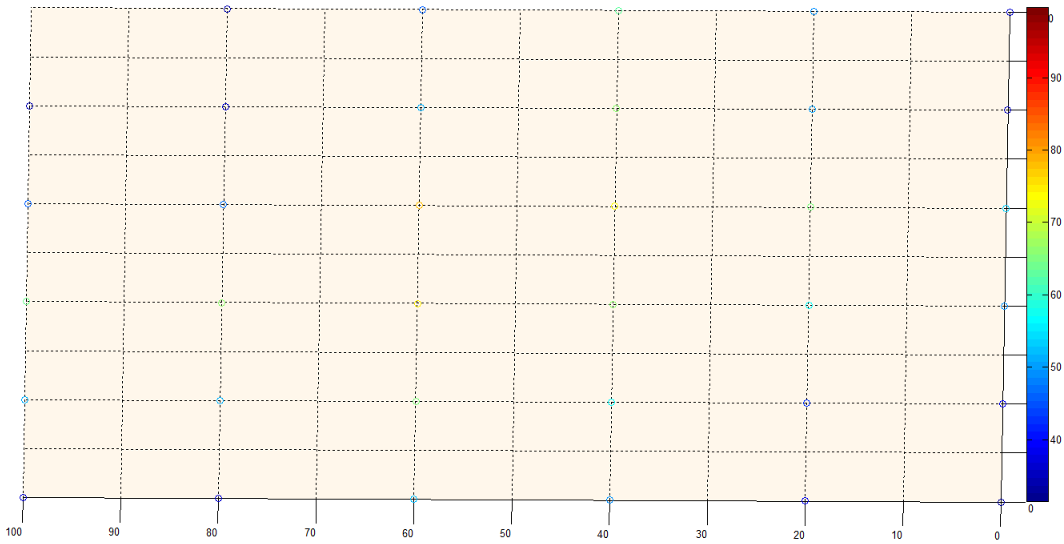



Figure 27: Top-down xy perspective of Three Agent Types each with different Preference Values