**Using Exact Sciences Modeling Tools to Understand Social Phenomena**

**Course #: 55772**

Homework #1: The Schelling model – **points for discussion**

Parameters and variables

* In the next lecture we will define the following terms:

Parameters / Variables / dependent variables / constant

* **Programming** variablesvs **model** variables: When describing a model you should not describe your internal variables (e.g. “current iteration” is not a model variable!)

Pseudo-code

* Since your code is in English - write your pseudo-code in English.
* The art of pseudo-code: between too much code and natural English.

Consistency – The Classic Schelling model:

* Some students modified the original Schelling model. It is okay if the student explicitly stated the modification and the reasons for his/her decision. The mistake is much more problematic (and more points were reduced) if there was inconsistency between model description / pseudo code / code / insights.

In such a stochastic model, small modification can totally change the results.

Here are some common mistakes - and you can think on cases when/why the following models differ from the classic Schelling model:

* + looking for unhappy residents from top-left to bottom-right, instead of randomly.
  + moving unhappy resident to a location that makes her happy instead of to a random empty location.
  + Moving a bunch of unhappy residents at once.
  + “number of iteration” – some people defined “iteration” in a way different from the Schelling model, e.g. iterating over the whole city, or all the unhappy people - which clearly affect the results and missed some of the important insights.

Simulations and figures:

* If the results are too noisy you can smooth e.g. by running the simulation with exactly the same variables 100 times and average.
* When running a simulation checking a specific variable (e.g. ratio of blue-red residents) you have to state clearly the other fixed variables (e.g. number of empty locations in the city)
* Vary only the parameters of interest. An elegant way of varying multiple variables at once is to use a full factorial experimental design, where you test all the combinations of the parameters. Then, you can average over the parameters you are not using, or show a representative example.

Figures:

* Some students plotted a linear interpolation between the points. Clearly, the experiment was not continuous – so how did you choose the sample points and how did you interpolate? i.e., did you take 1:8:24, or was it 1:1:24? This should be explicit in your figure and in the description.
* Graphs with no units, or worse: two graphs with different units and different scales on the same axes.
* When comparing several figures, the axis must have the same range. Comparing figure with y-axes 1e4 to a graph with y-axes 160 is misleading.
* Graphs that you “cut” without explanation. E.g., Students didn’t show the graphs for very dense cities, as the simulation never converged. This is okay IF you clearly refer to it in the text. This is especially important in this experiment since this is what we are interested in – convergence! Moreover, even if the simulation did not halt, the segregation value did usually converge. If you are “cutting” graph – you must explain why!
* How to show 2D functions: few chosen 1D slices, or 3D figure.
* Drops / peaks / local optima in the graph must be referred to, **even if it unexplained** your scientific instinct is expected to not miss weird behavior in your simulations – might you had a scientific a discovery? ☺

General:

* Technical report VS homework: we do not expect you to print 30 figures. Some students had excellent job, but it was hard to separate the sheep from the goats and the beautiful insights were lost. This is part of a scientific work.
* Graphs and results without insights: This is not a programming course; We do expect your insights!

**Task 1: Recalling the basic Schelling model**

Think about the Schelling model studied in class. Describe what the model's building blocks are: purpose, assumptions (that's tricky!), parameters, and variables. Be accurate and comprehensive as possible.

**Task 2: Implementing the basic Schelling model**

Please implement the Schelling model as studied in class. For the sake of uniformity use a "town" of 30X30. Start with two populations "yellow" and "blue" with symmetric preferences (yellow prefer blue in the same way that blue prefer yellow). Start with a random allocation of people in the town.

* The "threshold" is the percentage of neighbors like you, below which a person is unhappy and wants to move.
* Define the level of segregation as the number of people surrounded only by people from their type (ignore the empty places).
* Make sure your model counts the number of time periods till it converges.

1. Test the model to see you obtain results similar to what we saw in class.
2. Write a pseudo code describing how you implemented the model
3. Run the model multiple times and see how the level of segregation and the time to convergence depend on the following parameters:
4. Number of empty places.
5. The threshold.
6. The initial proportion between yellow and blue units.

Draw graphs to describe each of these dependencies and explain your insights.

**Task 3: Expanding the basic Schelling model**

Think of one way to expand the model, based on a realistic behavior. For example, you can choose asymmetric preferences, people who can randomly relocate also if they are happy, relocation in a certain radius, or any other behavior that might be relevant and interesting.

1. Describe the extension and why it is important.
2. Implement the extension in the code.
3. Run the code for multiple times and explain your insights. Think of interesting and meaningful ways to present the results. What do you see? What are the managerial implications?

Be creative, use your modeler's mind, and have fun!