Paper notes

Rocco Paolillo

1 Component of probability of the choice

- Object of choice: neighborhood
- Choice: relocate to a neighborhood
- Utility: payoff one is likely to accept to make the choice, i.e. to relocate to one neighborhood (or stay on current one). In Zhang (2004) it is made up of:

  - deterministic component U: how much people desire/value one preference in order to make the choice as ideal payoff, e.g. degree of ethnic concentration (e.g. max utility = 1 at 100% similars). Zhang (2004) uses a single-peaked function (vs maximization).
  - random term ϵ that refers to other characteristics of the option, i.e. the neighborhood, that influence the choice but independent of utility payoff. In regression models this is the random error terms. Random utility models in abm recreate with the exponential function, so that not always the best option according to the deterministic component U is chosen.

- Probability of the choice: $\beta u(.) + \epsilon$, which is conceptually different but related to payoff u

  - β is a parameter indicating how much the characteristic of the option is important in the probability to make a choice, or the random term ϵ. It is modeled as a positive constant: the lower is β, the higher is random term ϵ, i.e. other random characteristics, are relevant to choose one neighborhood. The higher is β, the lower will be the importance of random term ϵ and higher the importance of deterministic utility component to make a choice, i.e. moving to the neighborhood.

2 Aim of the paper

- Independent of the ideal payoff u people have for one characteristic of the neighborhood, my question is why someone should give more importance to that characteristic over the others, that is:

  why should β vary among individuals?

  An answer in Esser (2010) is social, educational, economic capital and how it translates into the investment of people into receiving context/ethnic context and his general suggestion. For instance:

  - The lower an agent speaks a language, the higher the need for aggregating with co-ethnics would be, so to increase β_{ethnic} independently of what is the ideal payoff for ethnic composition of the neighborhood;
  - the higher is social mobility intention either as investment or value in the own community, the higher the wealth (socio-economic status) of the neighborhood would be β_{status}
  - the higher the need to associate with people sharing the same values of local community (receiving context in Esser (Esser 2010) (e.g. political orientations, norm beliefs...) even as strategy for crossing ethnic boundaries (Wimmer 2013), the higher the composition of value-orientation would become with increase of β_{value}

- Costs of relocation can be included as costs that hinder relocation despite the choice. This can be one application of Esser (2010)'s definition of costs as limit to the investment on the receving context option. In Esser (2010) costs is loosely derived from social, economic and cultural capital milieu and are basically conceptualized upon the concrete choice actors (i.e. migrants) want to engage in, costs of relocation can be one of many.

- In the ACS paper (Paolillo and Lorenz 2018) we focused on the interplay of different homophily preferences in Schelling's threshold model and diffent type of segregation that emerge. With this direction the next paper could link to the ACS paper (Paolillo and Lorenz 2018) by:

  - addressing the reason why people would have preference for one characteristic over the other
  - what's the effect of people holding different preferences at same grades, while in the ACS they were exclusive tolerant: value, intolerant: ethnicity
  - extension to the random utility model as in Zhang (2004) would distinguish between the ideal payoff u for a characteristic and how important is that characteristic to people β
  - an additional point to random utility models would be to link β to needs and resources of people as in Esser (2010). This has some advantages to me:

    * relative group size and ethnic boundary making in Esser (2010) are supposed to influence probability to make a choice.In the ACS paper we included symbolic boundaries through the definition of similarity and relative group size, now we would include the overall frame of the model with need $\to$ change in β) and costs of relocation
    * as in the original Schelling (Schelling 1969, 1971) and other literature, homophily *per se* is a mechanism which describes the aggregation dynamics, but it does not provide causal mechanism on why people should aggregate based on one dimension over the others. This frame of the paper could fill this point
    * foster the linkage between Esser (2010)'s speculation and Schelling's model, so to benefit the long term agenda of the project
    * deeper definition of utility frame and rational choice, linking to the literature in abm and concretisize Esser (2010)'s model itself, which was elaborated as a summa of general and abstract rational choice

- Linking β to needs and including costs, $\beta u(.) + \epsilon$ should become something like

$$N_e U_e + N_v U_v + N_s U_s - C$$

where N = needs, C = costs and e, v, s the ethnic, value and socio-economic status dimension. The random term ϵ is implemented through the exponential function.

3 Notes on the model

- Now agents' characteristics of ethnicity, value and ses are nested. I think it would be worth to have a change in u and β for any intersection, but with risk to increase the number of parameters. The simplest solution I found so far was to link it to the value-orientation of agents as in the ACS paper (Paolillo and Lorenz 2018).

- A way to reduce the number of parameters and make the model more realistic might be to have u changing as function of the identification agents hold for the three categories (value, ethnicity and ses), changing the categories to a gradient $-1...+1$, e.g.:

  - value: extreme right-wing $\longleftrightarrow$ extreme leftwing, or pro local norms $\longleftrightarrow$ vs local norms.
  - ses: extrem low $\longleftrightarrow$ extreme high
  - ethnicity: dychotomous group 1 vs group 2 This way would maybe allow to explore more with fewer parameters:
  - ideal payoff function u as in the single-peaked function could derive from how people positionate along the continuum, as well combination of different homophily preferences. Assuming value as tolerance towards diversity, people extremely conservative would likely want in their neighborhood people both of the same ethnic group, but also of the same value orientation. People in the middler positions would care less about the value-orientation of others (as in the current ACS paper)

- the model would be relaxated which would justify the single-peaked function, linking u to shift in M and affect the right slope: people who aim at 100% similar neighborhood would be more likely to choose a 90% similar neighborhood than 30%, as well as people aiming at a mixed 50% neighborhood would be more likely to choose a 60% of similar, as in Zhang (Zhang 2004)
- different and more realistic scenarios, and examples in literature could be compared (Van Ham et al. 2012) or what data show. This supports the aim of the abm/project to clarify how different scenarios of integration/segregation are possible in an unified model and theory building

I implemented in the model the single-peaked function as in Zhang (2004) since I think it allows to explore more articulated scenarios. I think it would be worth to question about as above.

$$U = \begin{cases} \frac{x}{np}, & \text{if } x \leq np, \\ (2Z - M) + \frac{(M-Z)x}{np}, & otherwise \end{cases}$$

Assuming $Z = 1$ as peak of utility at desired fraction P, then the equation becomes as follow. This is how I implemented for the three utilities (ethnicity, value, ses). Did not understand to be honest how the second equation Zhang (2004) (otherwise) is reduced to the one used in Andreas Flache's slides and Sage's model (how 2Z-M disappears). I adapted their equation, but I understand it and it works as I tested.

$$U = \begin{cases} \frac{x}{np}, & \text{if } x \leq np, \\ M + \frac{(1-(x/n))*(1-M)}{(1-p)}, & otherwise \end{cases}$$

where M = constant for the right slope
x = number of similar ones in Moore neighborhood
n = number of agents in the Moore neighborhod $\rightarrow$ x/n: fraction similars
p = desired concentration

Fig. 1 replicates Zhang (2004) for desired fraction $p = 0.5$ in an 8 patches Moore neighborhood, $Z = 1$ and $M = 0.6$.
Fig 2 shows how the function will not change for the same proportion within a different number of patches occupied in the neighborhood, here with occupied patches equal to 6.
Fig. 3 shows how for M=1 utility becomes constant once the ideal payoff is reached, kind similar to threshold models
Fig. 4 shows how for M=0 the right slope is symmetrical with left one.

Fig. 5 shows the condition where desired concentration $p = 0$ in a 8 patches neighborhoood, with $M = 0.6$, which means agents actively do not want any similar in the neighborhood. This means that although utility = 1 at 0 is not possible, the function becomes linear decreasing.
Fig. **??** shows how this translates in the model with $\beta = 100$. Seems to me it increases stable integration, and it is not an error: as Zhang (2004) describes, with $\beta = 0$ agents just move randomly, with equal chance to move to alternative patch or stay in neighborhood. Here with $\beta = 100$ the characteristic of the neighborhood is relevant to them, although the payoff is 0 and $\beta * u = 0$, correct I think conceptually and as the stable patterns of emerged segregations and average utility show. To check if this can be related to Bruch and Mare (2006)'s error cited by Van de Rijt, Siegel, and Macy (2009)

4 Doubts

- It is still not very clear to me what you can get out of a regression model for what concerns utility. Some material I am reading describes dependent variable y as the probability to choose option x, and assumes
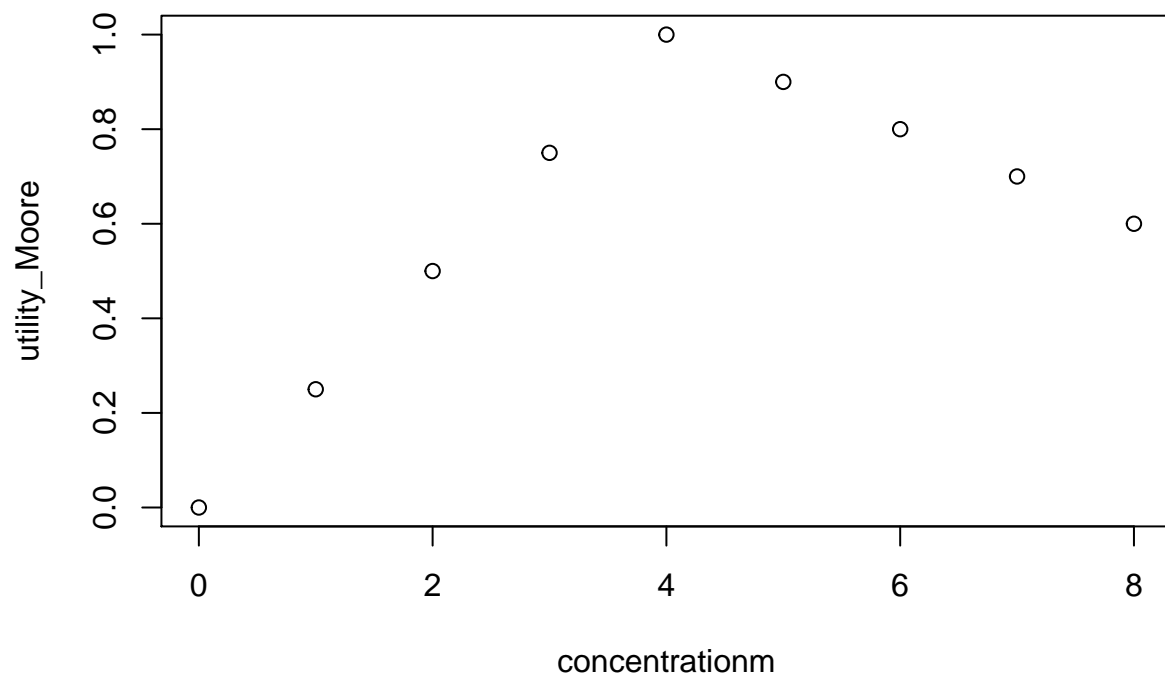
Figure 1: utility in Moore 8 patches; desired fraction 50%, M=0.6, Z=1

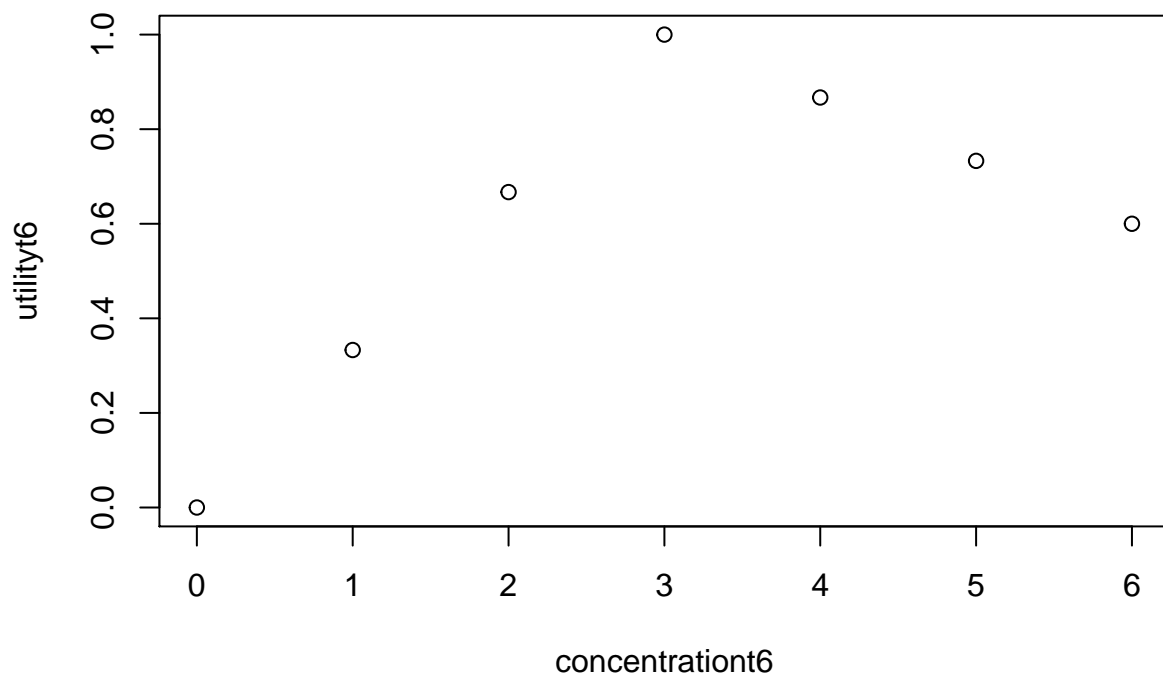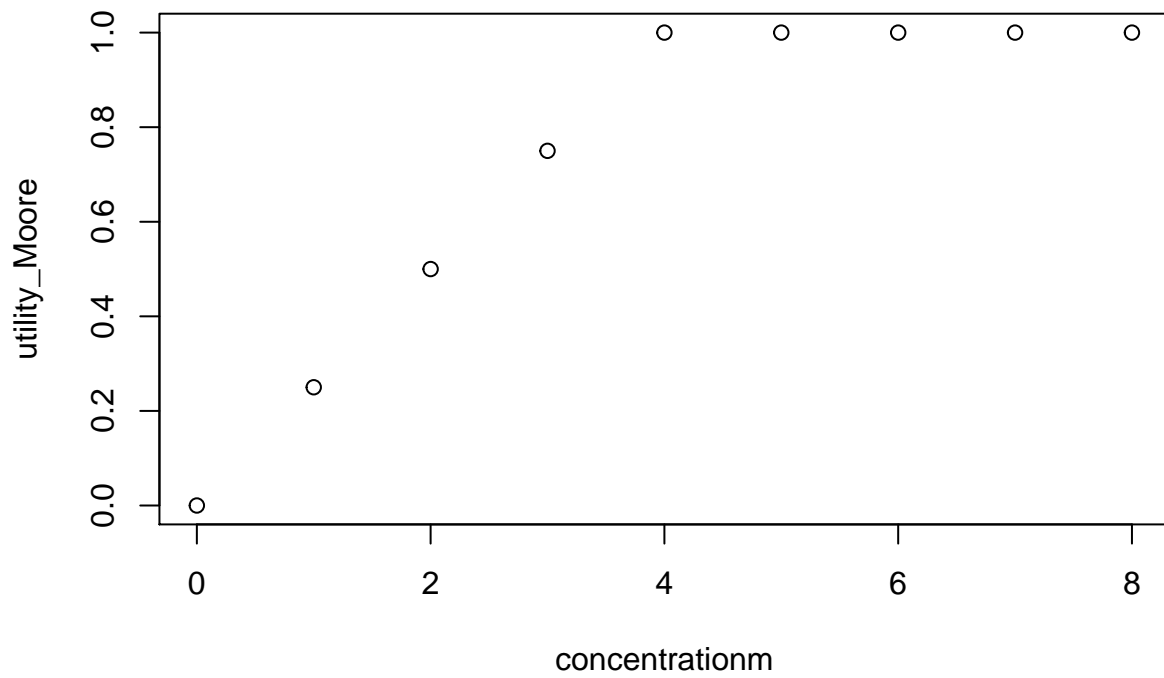Figure 2: utility in Moore 6 patches; desired fraction 50%, M=0.6, Z=1

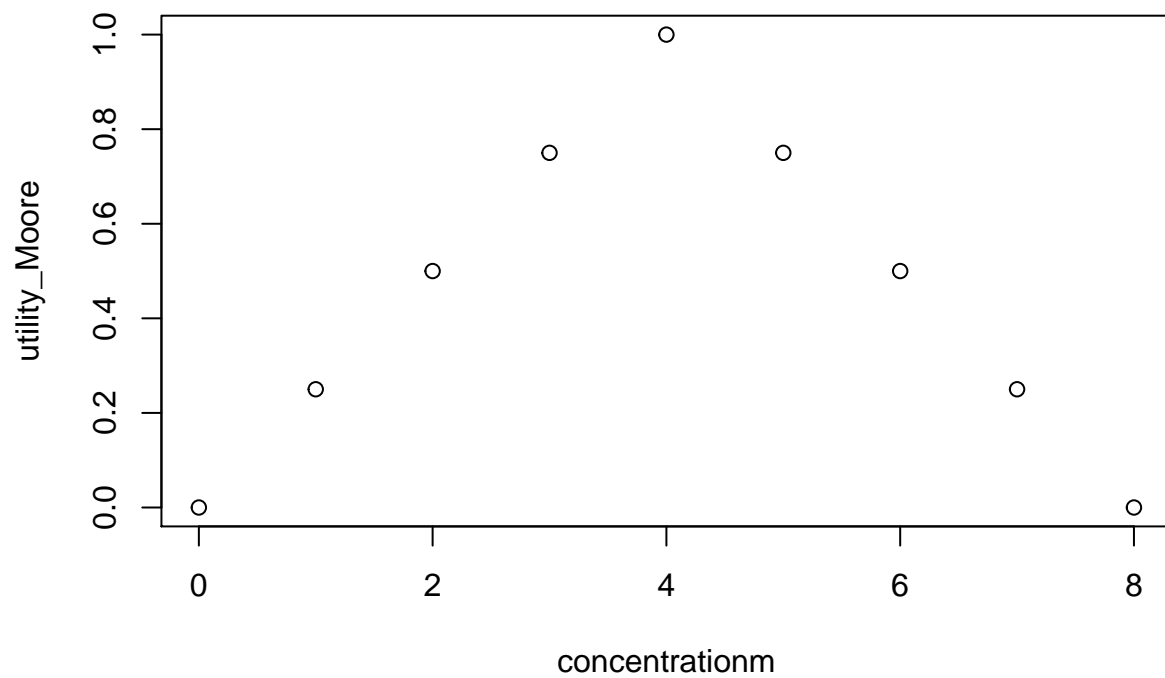Figure 3: utility in Moore 8 patches; desired fraction 50%, M=1, Z=1

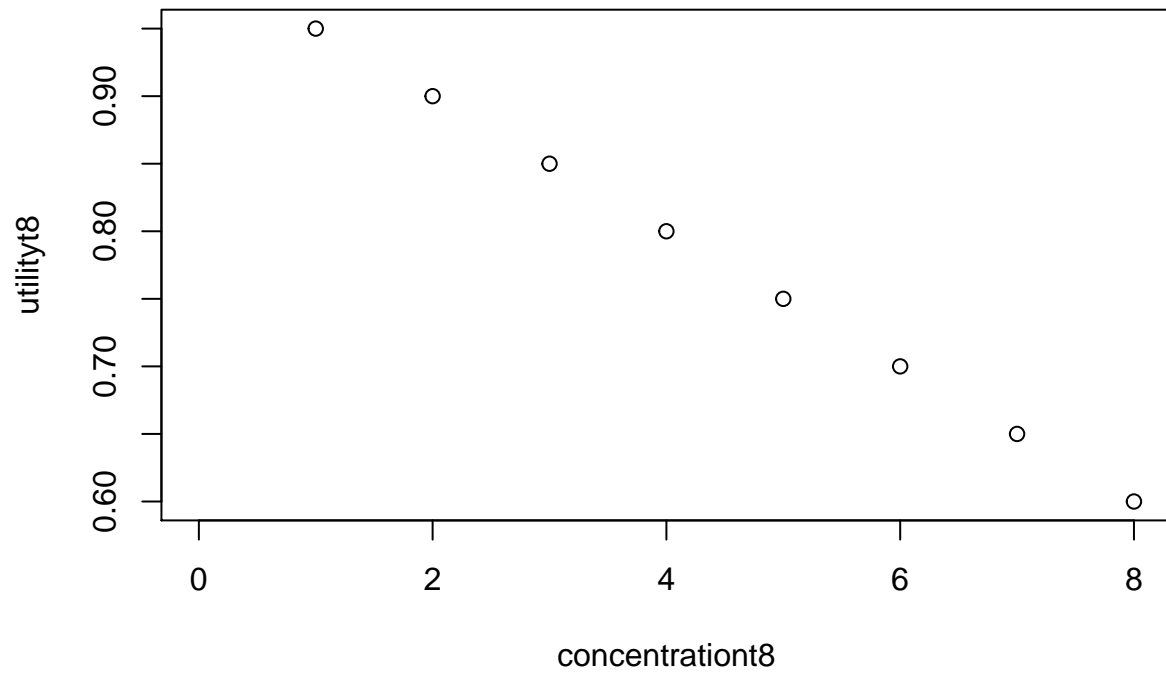Figure 4: utility in Moore 8 patches; desired fraction 50%, M=0, Z=1

Figure 5: Results with desired fraction p = 0 in Moore neighborhood 8 patches, M=0.6 and Z=1

this to be expression of the term utility quantified in beta coefficient times variable(characteristic) for each characteristic. But following Zhang (2004), utility of the option made per se would be already the product of β probability of choice * u ideal payoff. I do not understand if it is possible to disentangle the two components from a regression, although intuitive cases come to mind (e.g. finding negative correlation between costs in Esser (2010) and probability of choice). Borghans, Golsteyn, and Zölitz (2015) use the deviance of betas related to difference within groups, which I understand how to connect to multinomial choice as differences in choice due to characteristics of people, but it is not to my doubt. It is how to disentangle utility in empirical terms which is not very clear to me, and how the modeling choice can be related to empirical data, or if I am overthinking on it. In any case I would need to study more.

- Some models as Zhang (2004) and Bruch and Mare (2006) put random utility models within Schelling's framework. As I understand so far there are different components that collapse as multinomial choice, single-peaked function etc. In general they seems to me to fit within utility perspective, but it is not clear to me how much these utility models conceptually fit in Schelling (1969) threshold model they cite or are innerly different. In my understanding a threshold model would not care about the left-slope of a single-peaked function, and utility would be constant as Bruch and Mare (2006) also describes. There would be not consideration of (random) probabilities but prob = 1 once threshold is guaranteed, despite adding randomness, and not assessment of different options, as in the white flight. I think there is a difference conceptually and it should be marked. I also wonder how much a threshold model and utility models would differ when compared with empirical data and how they could be compared, which many abm paper do not focus on. I would like to study this as well, would make sense considering the ACS paper is a threshold model and now shifting to utility, but I think it would be beyond this paper. In general for my PhD I have to wite a paper of my onw.

References

Borghans, Lex, Bart HH Golsteyn, and Ulf Zölitz. 2015. "Parental Preferences for Primary School Characteristics." *The BE Journal of Economic Analysis & Policy* 15 (1): 85–117.

Bruch, Elizabeth E, and Robert D Mare. 2006. "Neighborhood Choice and Neighborhood Change." *American Journal of Sociology* 112 (3): 667–709.

Esser, Hartmut. 2010. "Assimilation, Ethnic Stratification, or Selective Acculturation?. Recent Theories of the Integration of Immigrants and the Model of Intergenerational Integration." *Sociologica* 4 (1): 0–0.

Paolillo, Rocco, and Jan Lorenz. 2018. "How Different Homophily Preferences Mitigate and Spur Ethnic and Value Segregation: Schelling's Model Extended." *Advances in Complex Systems* 21 (06n07): 1850026.

Schelling, Thomas C. 1969. "Models of Segregation." *The American Economic Review* 59 (2): 488–93.

———. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1 (2): 143–86.

Van de Rijt, Arnout, David Siegel, and Michael Macy. 2009. "Neighborhood Chance and Neighborhood Change: A Comment on Bruch and Mare." *American Journal of Sociology* 114 (4): 1166–80.

Van Ham, Maarten, David Manley, Nick Bailey, Ludi Simpson, and Duncan Maclennan. 2012. "Neighbourhood Effects Research: New Perspectives." In *Neighbourhood Effects Research: New Perspectives*, 1–21. Springer.

Wimmer, Andreas. 2013. *Ethnic Boundary Making: Institutions, Power, Networks*. Oxford University Press.

Zhang, Junfu. 2004. "Residential Segregation in an All-Integrationist World." *Journal of Economic Behavior & Organization* 54 (4): 533–50.