

# Title

*Author*

## 1 Introduction and Questions

Zhang (2004) cites and refers to Schelling (1978) for his prominent influence on threshold and tipping model, still then elaborates on a model based on random utility with single-peaked function and  $\beta$ . How much can a rational model which uses peak of utility and probability of choice be considered to rely on Schelling's threshold model?

What is the relation between random utility models and threshold models?

- Zhang (2004) assumes an entire world, not empty patches, so that *utility* = 1 at 50% is because 4/8 of agents are of similar ethnicity (plus the peak function). To replicate his model there should be not empty space. Bruch & Mare (2006) models with 15 empty patches. Should our world have not empty patches, so to maximize utility? This can allow also to test density effects on satisfaction and not bias on concentration calculation.

### 1.1 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 (which implies to stay on the same neighborhood). In Zhang (2004) it is made up of:
  - deterministic component  $U$ : how much people desire/value one preference in order to make the choice, 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  $\epsilon$  that refers to other characteristics of the option, i.e. the neighborhood, that influence the payoff to choose the neighborhood although not related to the deterministic component. 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 is taken.
- Probability of the choice:  $\beta u(.) + \epsilon$   
 $\beta$  is a parameter indicating how much the random term is important modeled as a positive constant. The lower is  $\beta$ , the higher is random term  $\epsilon$ , i.e. other random characteristics, are relevant to choose one neighborhood. The higher is  $\beta$ , the lower will be the importance of random term  $\epsilon$  and higher the importance of deterministic utility component to make a choice, i.e. moving to the neighborhood.
- Given the existence of an ideal payoff to make a choice  $u$ , why should someone give more importance to that dimension over the other, that is, why should  $\beta$  vary among individuals? An answer in Esser (2010) is social, educational, economic capital. For instance:
  - The lower an agent speaks a language, the higher for ethnic homophily would be  $\beta_{ethnic}$ ;
  - the higher is social mobility intention, the higher the wealth (socio-economic status) of the neighborhood would be  $\beta_{status}$
  - 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...), the higher

- Costs of relocation can be included as costs that hinder relocation despite high probability to choose. This can be one application of Esser (2010)'s definition of costs as limit to the investment on the receiving context option (although in the ethnic population, but it can be extended). In Esser (2010) costs is loosely derived from social, economic and cultural capital milieu, e.g. risk of social exclusion from in-group, but economic costs of relocation can be one simple implementation now.

## 2 Aim of the paper

- The conceptual difference between  $u$  and  $\beta$  should be  $u$ : what is the ideal point (payoff) to make the choice and  $\beta$  how important is that component to people to make a choice. The paper could focus on the interplay between different utility functions at ethnic group level and the individual importance for each of them. This would allow to test how segregation can emerge from the interplay of different homophily preferences.
  - The model could be initialized with correlations from data, as Andreas suggested, plus literature (e.g. right-wing and anti-diversity/migrant attitude)
- Given the interplay of homophily preferences, to compare the difference between threshold models and random utility models with single peaked function. It makes sense to consider  $\beta$  would apply equally to threshold models and single-peaked function, as well as the inclusion of randomness in a pure Schelling-type model. Check Bruch & Mare (2006) who should have had similar interest.
  - This aim/study would make a straight connection with the ACS paper and its conclusion
- The model can address the question why  $\beta$ , the importance of one characteristic of the neighborhood, would change depending on the needs (characteristics) of the individuals (households)
- The model can address the question of how costs interact with needs and preferences (economic capital) of individuals.
  - Both points above (need  $\rightarrow$  change in  $\beta$ ) and costs of relocation can be linked to Esser (2010) focus on capital theory. Relative group size and ethnic boundary making in Esser (2010) are supposed to influence probability to make a choice. As linking to the ACS paper, they realize into relative group size and symbolic boundary making through homophily preferences. This would foster the linkage Schelling-Esser of the general project agenda.
- As in the Schelling's model (Schelling, 1969, 1971), there is not question on why preferences for similar exist. While in the ACS paper we extended the model by including different dimensions of similarity. Extension to rational models as this would benefit understanding of different homophily preferences through their explicative reasons.
- The formula  $\beta u(.) + \epsilon$  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  $\epsilon$  is implemented through the exponential function.

## 3 Model

- Notes: at the moment the distribution of values or ses is at the ethnic group level, it is not nested. E.g. it would be possible to extend with distribution of tolerant/intolerant within high ses class and independent distribution of tolerant/intolerant within low ses class. I think this would allow to better address the correlation between the dimensions at a more realistic rate, but increase the number of parameter.

- 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 category to a gradient  $-1... +1$ , e.g.:
  - value: extreme right-wing  $\longleftrightarrow$  extreme leftwing, or pro loval norms  $\longleftrightarrow$  vs local norms
  - ses: extrem low  $\longleftrightarrow$  extreme high
  - ethnicity: dychotomous group 1 vs group 2 People who recognize as extremely leftwing or extremely rightwing might be more likely to attribute higher value to  $\beta_v$  than people who are on the average position. Other dynamics that might be associated can be different homophily/heterophily behavior:
  - people at the lowest continuum for socio-economic status might be willing to move to richer neighborhoods (i.e. heterophily behavior) than people higher in the scale who would relocate to neighborhoods with similar status (i.e. homophily)
  - inter-relation between identification and homophily preferences: if values is considered as tolerance towards diversity as in the ACS paper, people at the middle position should not care about ethnicity, those at the extreme intolerant position devalue the neighborhood, thus increasing the ethnic homophily preference. A way could be considering as similar those who fall into a  $+0.10 \setminus 0.10$  range. The slope of utility including M can be derived by the position of the agent along the continuum, starting from 0.5 for those in the average position. E.g. 0 value would have a threshold of 0.5 and progressively losing it. An extreme agent at 1, with utility 100% of those considered as similar, the same for  $-1$  with  $u = 1$  at 100% similar (extremely liberal agents). Also less than 50 would mean others not similar, including extreme ones would be in the neighborhood.

The function below works with the modelization of utility. In Zhang (2004), it is assumed people have a preference for mixed neighborhood ( $u = 1$  at 50%), depending on number of agents in neighborhood. As formulated now, the number of desired agents  $n$  is substituted with the desired concentration.

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

Assuming  $Z = 1$  as peak of utility at desired fraction  $P$ , then the equation becomes:

$$U = \begin{cases} \frac{x}{np}, & \text{if } x \leq np, \\ (2 - M) + \frac{(M-1)x}{np}, & \text{otherwise} \end{cases}$$

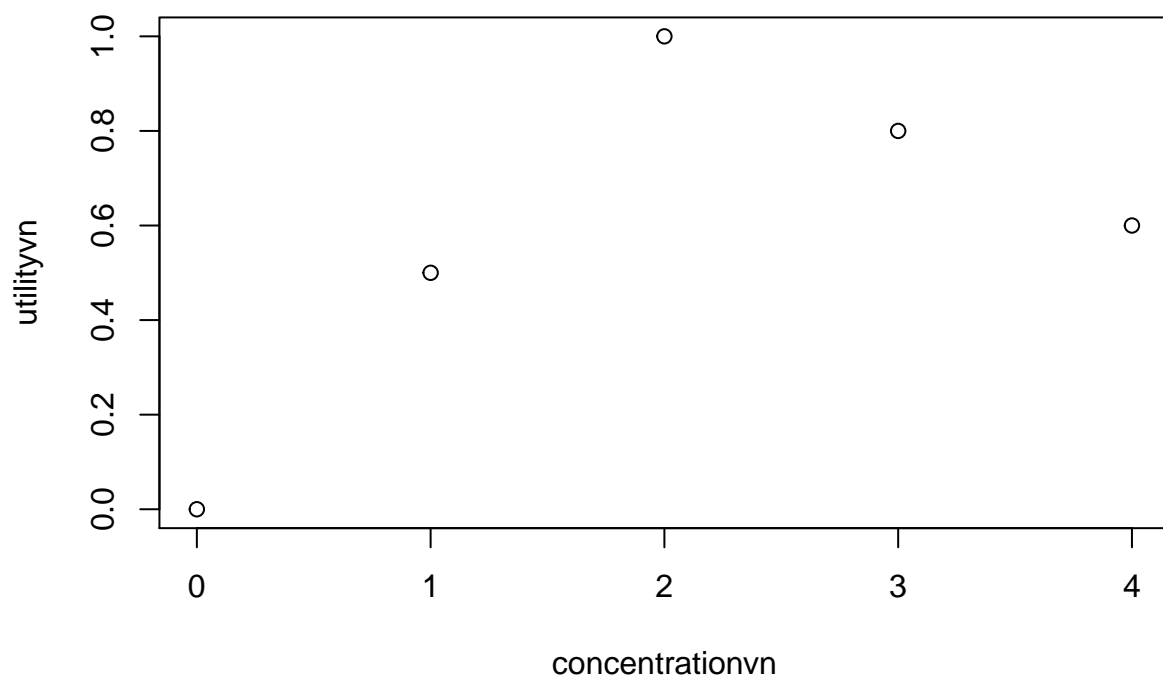
The first 3 steps compares to Zhang (2004) results,  $Z = 1$  and  $M = 0.6$  the last one is to test with a neighborhood of 8 patches. With  $m = 0$ , utility is constant once the ideal payoff is reached as in a threshold model, with  $m = 1$ , symmetrical peaked function results. To edit in table<sup>1</sup>

\* With  $p = 0.2$  the right-side of peaked function goes below 0, in sum conditions as  $p = 0.8$  and  $p = 0.3$  utility is approximated above .9 but does not reach 1.

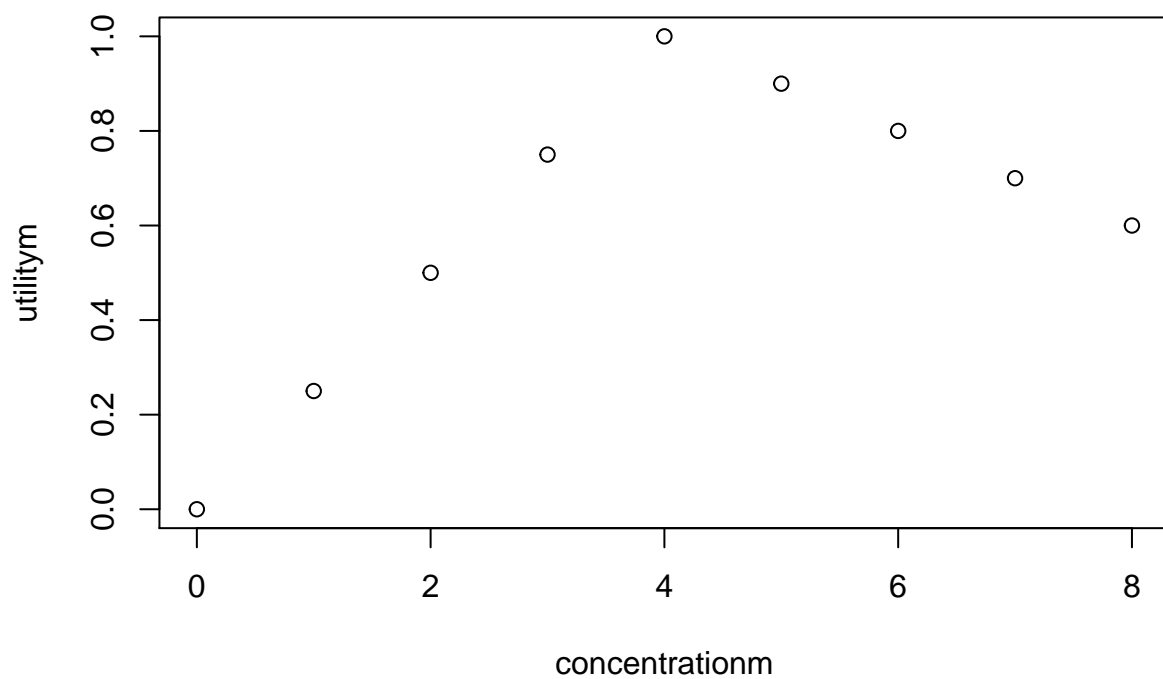
Test for *utility* = 1, utility = 0 means that agents want 0 similars in their neighbborhood, thus neighborhoods with not any similar one will have max utility and literally decreasing for neighborhoods with more similar ones. With  $\beta = 0$ , agents don't care about the composition of the neighborhood, so that both current location and alternative patch have prob. equal to zero, selecting alternative or current one randomly, sometimes relocating. Increasing  $\beta$ , the utility attributed to the patch due to neighborhood composition is multiplied times  $\beta$ , meaning those neighborhoods less inhabited by similar ones will higher probability to be chosen, thus increasing integration. It is not agents don't care, but agents care that similar ones are not there. See if this is related to Bruch & Mare (2006)'s error cited by Van de Rijt, Siegel, & Macy (2009)

## [1] 0.0 0.5 1.0 0.8 0.6

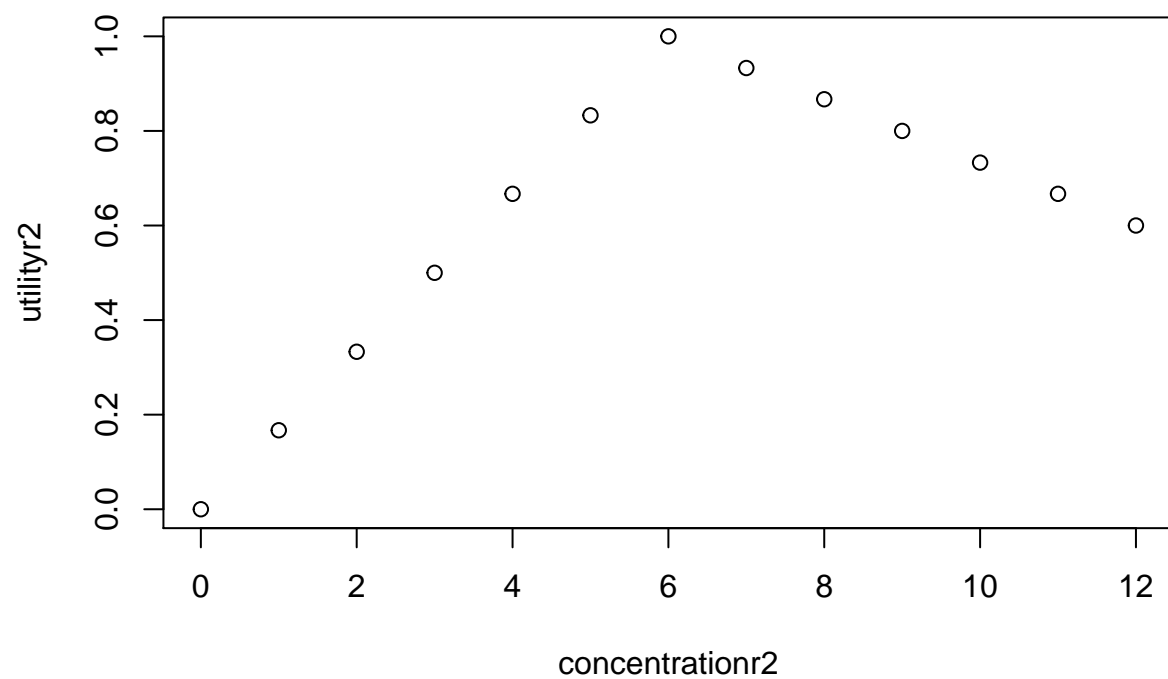
<sup>1</sup>Not managed to implement Flache's formula on slides, either not understood correctly what it would translate on my formula:  $p = n$  number of agents;  $f = P$  desired concentration.



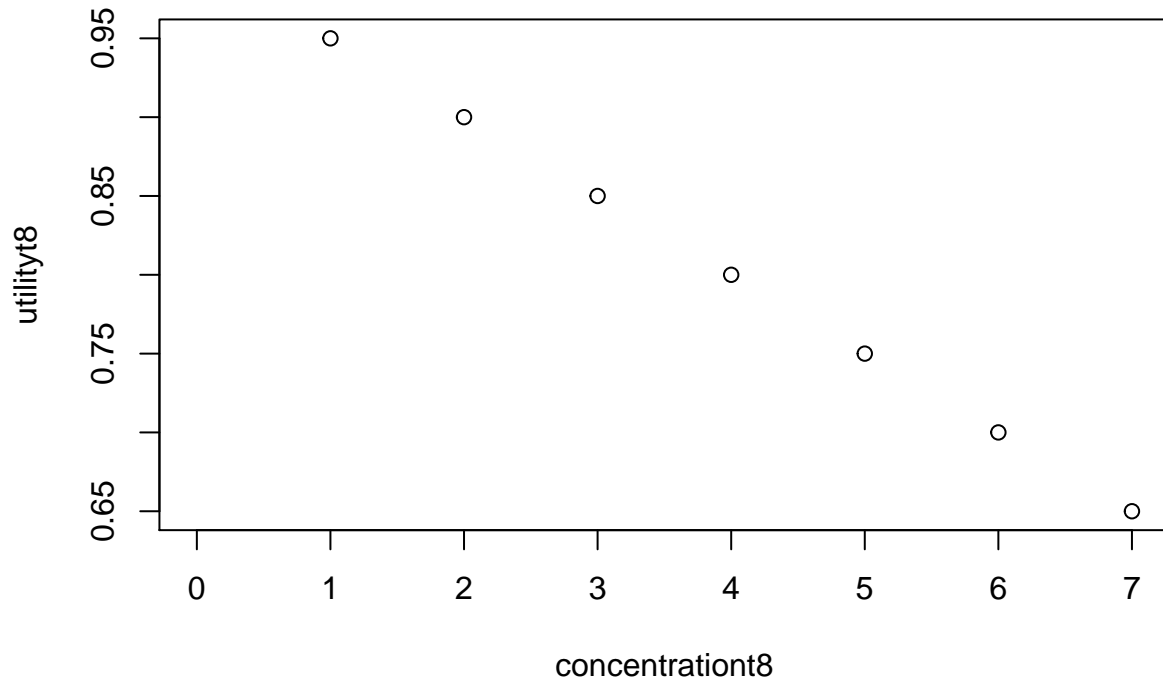
```
## [1] 0.00 0.25 0.50 0.75 1.00 0.90 0.80 0.70 0.60
```



```
## [1] 0.000 0.167 0.333 0.500 0.667 0.833 1.000 0.933 0.867 0.800 0.733
## [12] 0.667 0.600
```



```
## [1] NaN 0.95 0.90 0.85 0.80 0.75 0.70 0.65
```



## 4 Limits

- Both  $u$  and  $\beta$  can vary across individuals, but the model would explode in its complexity
- Even if we consider households as agents who make a choice, the decision to relocate can be the product of negotiation of people in that household. Too complicated to be included

## References

- Bruch, E. E., & Mare, R. D. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology*, 112(3), 667–709.
- Esser, H. (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.
- Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59(2), 488–493.
- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2), 143–186.
- Schelling, T. C. (1978). *Micromotives and macrobehavior*. WW Norton & Company.
- Van de Rijt, A., Siegel, D., & Macy, M. (2009). Neighborhood chance and neighborhood change: A comment on bruch and mare. *American Journal of Sociology*, 114(4), 1166–1180.
- Zhang, J. (2004). Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization*, 54(4), 533–550.