Dissertation RP

In partial preparation of submission ...

Rocco Paolillo
Bremen International Graduate School of Social Sciences
University of Bremen & Jacobs University Bremen
Mary-Somerville-Str. 9, 28359 Bremen, Germany
rpaolillo@bigsss-bremen.de

Any note 13/08/21 15 01

Executive Summary

WHAT

Society is made up of people belonging to different groups. Residential segregation is a crucial issue to social policies, interested in reducing spatial segregation and foster integration. From a theoretical point of view, much researc has been interested in developing quantitative measures of spatial segregation, qualitative research on the residential preferences of people and how these translate into spatial segregation. A critical interest in the social sciences is to understand the mechanisms and dynamics of how individual residential preferences can translate into spatial segregation. This dissertation is rooted in this area of research. Agent-based modelling and analytical sociology, social complexity are widely recognized as a valid tool to understand such phenomena.

ORIGINAL FRAME Rocco: definition of tolerance: Schelling with threshold "we don't want to stay in minority" Thomas Schelling proposed in the last 60's a noteworthy agent-based modelling to show how individual preferences could translate into spatial segregation. In particular, the model builds on threshold preferences, i.e. the minimal fraction of similar ones in neighborhood composition to remain in a neighborhood. The main contribution of the model is to show how spatial segregation can result from even low preferences for similar ones through cascades effect of individual relocations. The model has a wide application in describing residential ethnic segregation.

BUT

However, societies today shows some features that clash with the original assumptions of Schelling's model and its application in ethnic segregation literature in Western societies. First, society is inertly diverse, the paradigm majority-minority does not work anymore as much as minority-minority. A growing stream of literature shows how other forms of spatial segregation have emerged, e.g. socio-economic status. Other forms of categories can drive homophily preferences in the residential decisions. This is first a theoretical opportunity that can be applied to Schelling's model, as well as seems an empirical evidence of current, diverse societies. WHAT IS NEW?

This dissertation focuses on this aspect: definition of similarity based on other cross-ethnic categories, and the heterogeneity of members of the same group for one or either category.

HOW AND WHAT?

Rocco: tolerance first paper: we don't care In a first paper, we focus on ethnic-oriented vs value-oriented agents. Results show how the introduction of value oriented agents diminishes ethnic segregation compared to the original Schelling's model, though a new form of value segregation emerges. We identify critical points were all agents can be satisfied with neighborhoods. Rocco: tolerance: weight of dominant >= secondary preference In a second paper, we translate the model into a discrete choice random utility model. This not only updates the research outcome to more recent literature in Schelling's literature, but also serves different aims. First, discrete choice random utility models allow to model decision making in residential segregation, modeling the weight for each preference, which translates into randomness of behavior and test robustness of emerging phenomena. Additionally, we tested a mix of preferences with both ethnic and value preference for each type of agent and imposing a dominant and a secondary preference, along with different structural condition in terms of ethnic and value sizes. We specifically focus on how segregation scenarios would differ if a minority population would become more or less liberal. Results show how a tendency towards segregation remains a main outcome of relocation decision, through ethnic integration can be reached as an intermediate state due to a combination of higher randomness and relative group sizes. The most robust result is value segregation of conservatives, as in the ACS paper: that it is robust since appears both in threshold model and linear function. Rocco: tolerance: taste for diversity/push towards integration A third and final paper defines a different level of tolerance. The paper focuses on the scenario where agents maximize their dominant preference and show more tolerant or integrationist attitudes for the secondary preference. LIMITS

The aim of this contribution is

Contents

List of Figures

List of Tables

1 Introduction

1.1 Residential segregation

Rocco: the aim here should be ultimately to justify the focus on micro-macro dynamics, i.e. social complexity as a paradigm and agent-based modelling as a tool -> the same research question is on the dynamics of segregation

- What is it, why important
- Lines of research:
  - quantitative measurements (Duncan: eveness,dissimilarity etc.)
  - qualitative: individual relocation decisions and life course
  - reasons and processes:
    * Spatial Assimilation Theory (focus on minority, gaining economic power move to more affluent (historically white) neighborhoods)
    * Place stratification Theory (local stakeholders (e.g. real estates): planned exclusion of minority)
    * But the main question remains: how the dynamics work? See Van Ham et al. (2012) Rocco: this justifies more abm as a method than Schelling's contribution themselves. I would rather add a paragraph on need of analytical sociology/social complexity as need, and abm as a tool
    * An alternative model: Schelling's model

1.2 Schelling model

Rocco: I think a connection is feasible and I think worthy. I'm not saying Schelling's did his work because of the new legislation, I don't remember something similar, but he mentions segregation by color as main concern '69 paper.

Historical/academic background: years of end of racial laws and planned segregation (late 60's). Schelling was an economist and wanted to demonstrate how some forms of segregation could be "unplanned," emerging from the interaction of individual preferences.

Original aim: explain unplanned segregation as aggregating outcome of individual preferences in limited spatial constraints.

Description of the model and main contribution

The basic condition of Schelling model is that similarity of agents is defined as a "twofold, exhaustive and recognizable distinction" (Schelling, 1969, p. 488). What I describe here is known as the spatial proximity, or checkerboard version of the model. Consider two groups of agents who reside on a lattice grid (a checkerboard) that represents a city, so that each individual agent occupies a cell of the grid and considers the set of the eight surrounding agents as its own 'neighborhood.' Some spots are empty on the checkerboard. Agents are divided by color[1]. At each time-step, one agent can choose whether to stay or to leave. Its residential moves depend on the concentration of similar agents in the neighborhood, quantified by a threshold. If the concentration of desired similar agents in the local neighborhood falls below the desired threshold, the agent leave to a random sport. The contribution of the model is to describe how the interaction of individuals subscribing to a similar behavior can translate into high levels of segregation Rocco: and high levels even with mild preferences, check if add here or later Imagine one red agent starts. If the percentage of other red agents in the neighborhood falls below the threshold, the agent relocates to one of the empty spots on

[1]This is a description of the model adapted from León-Medina (2017), Rocco: for the red/blue scenario, delete in the end if any

the checkerboard, so to leave its previous spot in the old neighborhood empty. By doing so, the red agent changes the composition of both its previous neighborhood and its new one. In particular, the percentage of same-color agents decreases for other red agents left in the previous neighborhood, while it increases for other red agents in the new neighborhood. In this way, the previous neighborhood becomes less attractive for the remained red agents, who could leave when it is their turn to decide. On the contrary, the new neighborhood becomes more attractive both to other red agents who already live there (who are then more prone to stay) and potentially to unsatisfied red agents living in other neighborhoods. The same works for blue agents. Through the change in neighborhoods' compositions, therefore, the relocation decision of each agent has unwillingly a cascade effect on the next agent to decide. This enables a self-reinforcing dynamic so that some neighborhoods become progressively less attractive for agents of one group and more attractive for those of the other group. In other words, the concentration of one group in the neighborhood reinforces its attractiveness for other in-group members when they decide to relocate. This, in turn, can lead to high segregation—even with relatively tolerant agents.

- The impact of the model in literature: considered one of the first agent-based modelling, though not in the intention of Schelling. Influence in the agent-based literature. Known also in empirical studies of spatial segregation
- Focus on agent-based modelling literature: BRIEFLY: extensions, application to other methods, e.g. GIS. In details in each section of the papers
- What remains as a stable result: tendency towards segregation as a stable outcome

1.3 BUT

Rocco: here it must be a paragraph driving the reader to what is the research question/justification of the dissertation, then in details with the literature

- More recent literature on spatial segregation in Western societies shows some results that contrast Schelling's assumptions and scenarios
- Results: ethnic segregation is decreased, while other forms of segregation have emerged (e.g. socio-economic status)
- Societies are inherently different. The paradigm majority-minority most literature on ethnic segregation was built on (see School of Chicago and spatial assimilation) not stand. Different paradigms emerge (see Crul (2016), Vertovec (2007))

Rocco: here it becomes tricky: diverse type of segregation as the aspect that clashes with Schelling's framework

1.4 New forms of Residential segregation

- As an outcome: residential segregation experienced on diverse forms at at different levels for members of the same ethnic group or social class or other categories
- New residential preferences alternative to ethnic homophily:
  - Increased preference for diversity for some members of the same social category
  - Diverse ethnic tolerance
  - Preference for other forms of similarity

1.5 Superdiverse societies

- Society is more diverse compared to the historical context Schelling built on and to the theoretical assumption of similarity based on ethnicity as an exclusive dimension:

- Sociodemographic changes:
  - the paradigm majority-minority (e.g. white vs black) does not work anymore, rather an amalgama of ethnicities (see Vertovec (2007)) as effect of new generations of migrants' descendents and faster international mobility
  - correlation ethnicity/social class: members of minority can be in upper class, members of majority being in the lower class Rocco: though majority still holds a numeric majority and in the upper class, it is changing
- In such a complex scenario, the process of defining boundaries of membership become relevant: Ethnic boundary making

1.6 Ethnic Boundary Making

- dynamic process of definition of inclusion
- ethnicity not an exclusive category

2 Aim of dissertation

- Given these observations on modern diverse societies and different complex scenarios of spatial segregation, how can we bring Schelling to these scenarios? In particular the question evolves into:
- What are the consequences of different definition of similarity?
- What different types of spatial segregation would emerge? And what other, if any, unexepcted consequences?
- How can spatial segregation being reduced out of these extensions?
- Main scenario investigated: definition of similarity based on ethnic membership vs cross-ethnic supposedly inclusive category (in the model description value)
- Different degree of preference or tolerance for either similarity based on the own group membership Rocco: the next point is crucial and should be well developed, summarizing what specific niche/scenario the dissertation proposes as advance to Schelling's framework and research question(s), and then move with specific literature
- Out of such diversity, other categories people belong to can matter in the definition of similarity and neighborhood preferences, independent on and complementary to ethnic membership. People belonging to different groups can hold different degree of preferences for ethnic membership or alternative categories.
- This is both a theoretical possibility and an empirical evidence as growing empirical literature shows
- It can contribute to lower ethnic segregation and show more complex scenarios of segregation

3 Paper 1

- Aim: exploration of what are the consequences of different definition of similarity; effect of relative group size
- Extension: different definition of similarity, remaining into a deterministic, deterministic model and threshold function
- Literature associated? More focused on ethnic boundary making and homophily preferences
- Model description: tolerant agents value-oriented vs intolerant ethnicity-oriented
- Measures: exposure, density
- Experiments: parameter sweeping
- Results: decrease ethnic segregation due to tolerant agents, spillover effect

4 Paper 2

- Aim: Understand better mechanisms associated with different homophily preferences and heterogeneity for value/ethnic similarity, how integration for both or either dimensions can be reached?
- Explore consequences of a mix of preferences for ethnic and value similarity in both types of agents, mantaining dominant and secondary preference; effect of ethnic ratio and value distribution in minority
- Extension: probabilistic model and linear utility function random utility model: to represent the decisional process, a weight for each dimension and robustness of emerging results (randomness); ethnic ratio and value distribution of minority
- Rocco: here it risks to be tricky: a theoretical section, or mention briefly? Here literature on discrete choice random utility: utility models (empirical science McFadden...), utility function vs probability function, what differs between abm and regression models. Literature in abm: Bruch-Van Rijt discussions...

Random utility models for discrete choice have a long history in housing research (Frankhauser & Ansel, 2016) and in recent years they have been applied in the agent-based modeling framework (E. E. Bruch & Mare, 2006, p. bruch2012methodological). Stemming from the utility maximization paradigm, these model assume that the decisional process underlying the choice of economic actors is unknown, and it can be deduced by observed preferences, i.e. how selection of respondents differ for attributes of the options available (Hess et al., 2018), e.g. different neighborhood composition. So, aim of regression models comparing choices of the sample is to estimate vector parameters that quantify the likelihood to select one option over the other depending on the difference in their attributes (Manski, 1977). Utility in this context is defined as the attractiveness for each characteristic the options differ for and it is based on the response curve of the respondent (E. Bruch & Atwell, 2015, p. train2009discrete) imposed based on the theoretical model of the analyst. Nevertheless, random utility models divide between a systematic component of utility, i.e. observable differences between options based on their utility, and a random term, representing all unknown factors associated with selection of that options, might they depend on other characteristics of the option, characteristics of the selector or an interaction of both. Compared to our model, random utility for a generic neighborhood is:

$$U = \beta_e U_e + \beta_v U_v + \epsilon \tag{1}$$

where:

β_e = weight parameter for ethnic similarity, with $\beta_e[0, \infty)$

U_e = ethnic utility of neighborhood

β_v = weight parameter for value similarity, with $\beta_v[0, \infty)$

U_v = value utility of neighborhood

ϵ = random term

While parameters β_e and β_v can be estimated through regression models, the random term ϵ remains unknown. The conditional logit model introduced by (McFadden, 1994) is a specific type of discrete choice model that allow to quantify the effect of systematic utility over random component, though remaining the latter unknown. Assuming the random term ϵ follows a type I extreme value distribution, e.g. Gumbel distribution, the probability to select neighborhood j out of options k in choice set C is:

$$P_{j \in C} = \frac{exp(\beta_e U_j^e + \beta_v U_j^v)}{\sum\limits_{k \in C} exp(\beta_e U_k^e + \beta_v U_k^v)} \tag{2}$$

In this computation of probability, parameters β_e and β_v become the weight of how much the systematic component of utility for that dimension matters in the selection of the option compared to the random component represented by the unknown random term. The higher β_e or β_v, the higher the likelihood that

the option with highest utility for that dimension will be selected, the lower β_e or β_v, the higher the chance that an option is selected randomly for that dimension. With both β_e and β_v equal 0, the all options have equal probability to be selected, since the choice is totally stochastic, i.e. dependent on random term.

Implementation of discrete choice in agent-based modelling has a number of peculiarities compared to empirical regression models. First and above all, estimated preferences in regression models depend on a the comparison of a limited set of observed cases that can profoundly affect results. Moreover, utility remains a random variable estimated through parameters β and computation of probability to select one option. Probability to select one option over the others as behavioral response to change in their attributes is the equivalent to utility. As (E. E. Bruch & Mare, 2009) stress out, agent-based modeling are deeply different in this aspect, allowing to model independently and quantify the elements of the conditional logit based on a theoretical model. Researchers can impose different combinations of utility function for all ranges of neighborhood characteristics and parameters β in the relocation decisions of agents to observe their aggregated results. Additionally, modellers can include other elements that contribute to the dynamics of emerging spatial segregation, such as diverse heterogeneous distributions of preferences (see (Xie & Zhou, 2012)) or population structures that can influence neighborhood composition (E. E. Bruch, 2014).

Also agent-based modelling can benefit from the implementation of discrete choice models, not only for the formalization of the decisional process of agents and calibration with parameters β estimated (E. E. Bruch & Mare, 2006), but also for the inclusion of the stochastic random term. Randomness is useful to both test robustness of observed phenomena in complex systems and increase their realism through inclusion of random fluctuations against deterministic behavior. A traditional way to include randomness in the dynamics of an agent-based model is as an external noise, for instance in Schelling model, with a percentage of agents, or additional agents, forced to randomly relocate by the researcher. Implementation of discrete choice allows to include randomness as an endogenous component in the relocation decision of individual agents through the random term compared to parameters of determinism. Useful to our interest in the interplay of value and ethnic preferences, the parameter β can vary as a local variable of agents, so to sort out differences in the ratio between deterministic and random relocation depending on characteristics of agents. We build on this peculiarity of random utility models to test how random behavior of a specific type of agents for value orientation, ethnicity or a combination of both can influence the emergence of stable equilibria of hybrid segregation. We can moreover explore the interdependence between different types of agents and how determinism of agents would react to different distributions of the two characteristics of ethnicity and value orientation in the population. In the next section we describe our extension of (Paolillo & Lorenz, 2018) and how we implemented random utility discrete choice.

- Model description
- Measures: exposure, spatial clutering for asymmetric conditions
- Experiments: parameter sweeping, more sensitivity analysis
- Results: by-product value segregation of conservatives (ex spillover effect) better understood, as effect of more robust/denser neighborhoods of liberals, integration can be reached as an intermediate state towards segregation, due to a mix of lower randomnes, ethnic and value size. Effect of ethnic ratio on effect of weights: By-product of majority size over minority, specific scenario of liberals minority.

5 Paper 3?

- Aim: definition of tolerance as integrationist preferences: push towards integration. Would this reduce by-product value segregation that appeared both in threshold and discrete choice version, and reduce ethnic segregation?
- Model description
- Experiments
- Results

6 Conclusions

- What can we say about segregation dynamics?
- Effect of cross-ethnic preferences vs ethnic preferences: what is the difference, through neighborhood density
- Effect of relative weights (discrete choice implementation)
- Effect of relative group and value size
- Intermediate levels of integration as effect of randomness, preference and relative size
- What unexpected results were reached? How strong segregation emerges even if integrationist attitudes?

7 Discussion

- Difference between deterministic threshold model vs discrete choice model: what changes? Theta 0 = beta 0, high determinism small fluctuations due to randomness
- How a threshold and a discrete choice model differ? How robut results are across different implementations?
- Comparison with discrete choice models in empirical science used e.g. to estimate parameters determinism β and utility function:
- observed preference (e.g. from census data): equilibria are already emerged and other factors (or dynamics) context-specific could cause segregation
- stated preference (e.g. survey): other factors could intervene, so that the actual segregation could be not derived from individual stated preference
- In the end, they reply to a different question: The aim in regression models is to estimate and validate individual preferences from data. The aim of abm is to observe aggregated results and formalize their dynamics, plus the possibility to model separately utility function and probability function (see E. E. Bruch & Mare (2009)) and impose any range of parameters Rocco: connects with limits/future agenda: what is the use of theoretical models vs data-driven models for spatial sciences

8 Limits and Future Research Agenda

- Limits
- Focus/contribution for modelling/complexity literature over empirical literature in residential studies
- Theoretical models vs data-driven models
- However, contribution of abm is to understand mechanisms, which justifies theory-driven models.
- Data
- Validation in the future, we know the dynamics and unexpected consequences of individual preferences based on diverse definition of similarity, and sociodemographic conditions
- comparsion with empirical data: what type of data? During the Phd different surveys and aggregate neighborhood data were explored which could have fit the model -> need to have joint probability -> micro data
- Additional layers: other socio-demographic characteristics, and resources associated

9 Appendix

References

Bruch, E., & Atwell, J. (2015). Agent-based models in empirical social research. *Sociological Methods & Research*, *44*(2), 186–221.

Bruch, E. E. (2014). How population structure shapes neighborhood segregation. *American Journal of Sociology*, *119*(5), 1221–1278.

Bruch, E. E., & Mare, R. D. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology*, *112*(3), 667–709.

Bruch, E. E., & Mare, R. D. (2009). Preferences and pathways to segregation: Reply to van de rijt, siegel, and macy. *American Journal of Sociology*, *114*(4), 1181–1198.

Crul, M. (2016). Super-diversity vs. Assimilation: How complex diversity in majority–minority cities challenges the assumptions of assimilation. *Journal of Ethnic and Migration Studies*, *42*(1), 54–68.

Frankhauser, P., & Ansel, D. (2016). *Deciding where to live: An interdisciplinary approach to residential choice in its social context*. Springer.

Hess, S., Daly, A., & Batley, R. (2018). Revisiting consistency with random utility maximisation: Theory and implications for practical work. *Theory and Decision*, *84*(2), 181–204.

León-Medina, F. J. (2017). Analytical sociology and agent-based modeling: Is generative sufficiency sufficient? *Sociological Theory*, *35*(3), 157–178.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3), 229–254.

McFadden, D. (1994). *Conditional logit analysis of qualitative choice behavior* (P. Zarembka, Ed.; pp. 105–142). New York, NY:Academic Press.

Paolillo, R., & Lorenz, J. (2018). How different homophily preferences mitigate and spur ethnic and value segregation: Schelling's model extended. *Advances in Complex Systems*, *21*(06n07), 1850026.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, *59*(2), 488–493.

Van Ham, M., Manley, D., Bailey, N., Simpson, L., & Maclennan, D. (2012). Neighbourhood effects research: New perspectives. In *Neighbourhood effects research: New perspectives* (pp. 1–21). Springer.

Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, *30*(6), 1024–1054.

Xie, Y., & Zhou, X. (2012). Modeling individual-level heterogeneity in racial residential segregation. *Proceedings of the National Academy of Sciences*, *109*(29), 11646–11651.