

Linköping Studies in Arts and Sciences No. 809
Institute for Analytical Sociology Dissertation Series No. 1

Local Social Exposure and Inter-Neighborhood Mobility

Àlex Giménez de la Prada

Local Social
Exposure and
Inter-Neighborhood
Mobility

Àlex Giménez de la Prada



Linköping Studies in Arts and Sciences No. 809
Institute for Analytical Sociology Dissertation Series No. 1
Faculty of Arts and Sciences
Linköping 2021

Linköping Studies in Arts and Sciences • No. 809

At the Faculty of Arts and Sciences at Linköping University, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in Arts and Sciences. This thesis comes from the Institute for Analytical Sociology at the Department of Management and Engineering.

Distributed by:
Department of Management and Engineering
Linköping University
581 83 Linköping

Àlex Giménez de la Prada
Local Social Exposure and Inter-Neighborhood Mobility

Edition 1:1
ISBN 978-91-7929-630-8
ISSN 0282-9800

©Àlex Giménez de la Prada
Department of Management and Engineering 2021

Printed by: LiU Tryck, Linköping, 2021



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

<https://creativecommons.org/licenses/by-nc/4.0/>

Abstract

Studies on ethnic residential segregation analyze how the inter-neighborhood mobility of individuals shapes their spatial distribution across cities. This literature has shown that the residential choices of households partly depend on the ethnic composition of their neighborhoods: higher in-group shares promote the presence of more in-group members, and vice versa. However, and in spite of the remarkable contributions, it remains unclear what exactly these studies refer to as “the neighborhood,” and how alternative definitions could challenge previous findings.

A large majority of studies have primarily adopted an administrative definition of the neighborhood due to limitations in the data collection process. Nevertheless, this definition has typically forced researchers to hold unrealistic assumptions about how households collect the information about the other individuals (Crowder and Krysan, 2016), and to treat the heterogeneity of social processes of large district areas as being homogeneous (Hipp, 2007). More generally, the large extensions of administrative areas have prevented an accurate description of how inter-group exposure affects the mobility dynamics of the individuals at more granular scales, and an assessment of the sociological concept of “the neighborhood” to analyze residential mobility dynamics.

This thesis studies the inter-neighborhood mobility patterns of Westerner households for the years 1998-2017 in Sweden. In particular, it aims at analyzing in detail how close and how permanent inter-group contact and exposure must be in order to prompt native out-mobility and, consequently, ethnic residential segregation. In the first study, I examine how the spatial distance between Westerner and ethnic

minorities moderates the salience of the minority presence and contributes to drive Westerner out-mobility. In the second study, together with Eduardo Tapia, I focus on examining how the previous out-mobility decisions of individuals foster further out-mobility of the in-group neighbors: the social influence effect on residential mobility. In the last study, I examine how the refugee crisis of 2015 has contributed to shaping natives' out-mobility through two modalities of local exposure on Westerners: asylum centers and refugees choosing their own accommodation.

The Analytical Sociology approach (Hedström and Bearman, 2009) informs the research design of the thesis, which seeks to unravel the interdependent aspect of segregation processes whereby the previous mobility actions of individuals may trigger further mobility responses. By applying a counterfactual design (Woodward, 2003) and utilizing Swedish register data, I analyze native out-mobility following the exposure to ethnic growth near the residences of Westerners. This analytical strategy enables me to overcome common limitations of random sampling studies and capture the spatial interaction between individuals using a causal inference approach (Coleman, 1986).

Results described in the above-mentioned studies provide empirical evidence showing the importance of the physical and social environment of Westerners to understanding their mobility patterns and the dynamics of segregation. Study 1 shows that growth in the minority presence in small areas centered on Westerners' home locations is capable of prompting native out-mobility. The closer the groups are to one another, the more likely it is to observe native out-mobility. These findings suggest that neighborhoods defined as administrative areas undermine the measurement of these kinds of interaction effects.

Study 2 confirms the previous findings by showing a greater propensity to move out following Westerners moving out the closer they previously were to other Westerners' residential locations. Moreover, results also show that the higher the number of out-movers and the better the visibility of these out-movers in low population density areas, the greater the likelihood of Westerners of out-moving. By adjusting for

theoretically relevant factors known to affect residential mobility, this study goes beyond out-group exposure and proposes a new alternative mechanism that partially drives the residential mobility of Westerners.

Finally, Study 3 shows that the perceived temporal duration of ethnic change might also influence the mobility decisions of Westerners. More concretely, this study shows that temporary asylum centers do not prompt native out-mobility despite markedly increasing the visibility of out-group salience in the area where this temporary asylum center is established, not even for Westerners living in ethnically mixed areas. Conversely, the absence of this temporary restriction for refugees entering the housing market and self-selecting into Westerner-based areas positively increases native out-mobility, even despite refugees moving in produce overly lower increases in out-group salience. Moreover, native out-mobility is greater when the exposure to new refugees occurs in areas that are already inhabited by other non-Westerners.

Dedication

I would like to especially dedicate this thesis to my mother, Miriam de la Prada García-Cortaire, and my father, José Maria Giménez Arnau, as science can be done with great effort, but it is nothing when it is detached from any love. Showing kindness is as important as knowing your methods. This teaching is thanks to them, but also to my siblings, Guillem and Aitana, two stars that I love and respect more than anything in this world, and my other relatives, including those who departed.

I would also like to dedicate this thesis to my partner, Joel Flores, whose unconditional love and support during these years have been key in overcoming many of the hard times encountered. Lastly, I would like to dedicate this thesis to my friends, who are always supportive when I am back in Barcelona. If anyone has access to good and valuable human capital through their strong ties, that is undoubtedly me.

Acknowledgments

Perhaps because of my sociological training, I tend to see this thesis as a collective effort rather than my own. I would like to thank my colleagues at the Institute for Analytical Sociology for their unhesitating predisposition to discuss any topic and answer any question I had, either during formal seminars or during a gather meeting, or in any of the dozens informal conversations held during lunch, fika, coffee breaks, even in the corridors. I would like to thank Petri Ylikoski, Christian Steglich, Marc Keuschnigg, Maria Brandén, Carl Nordlund, Benjamin Jarvis, Jacob Habinek, Karl Wennberg, Sarah Valdez, Chanchal Balachandran, Jutta Kawalerowicz, Anders Hed, Flóra Samu, Károly Tákacs, and Daniel Barkoczi for their valuable support during this long period. I would also like to thank Peter Hedström for having granted me the opportunity to develop my skills as a modern sociologist in such a high tier Institute as IAS.

If there is one person I am most gratefully thankful of, this is Eduardo Tapia. If this thesis has managed to harbor safe and sound it is thanks to Edu's unconditional support during these long years. His contribution in countless meetings inside and outside the Institute improved the thesis by tiny nudges and helped it becoming what it is today. I would also like to thank Anders Hjorth-Trolle for his support along these years, not the least during the last stages of the thesis. If the methods of the thesis look sufficiently well implemented, it is without doubt thanks to his careful attention and (always kind) comments.

Fortunately for me, I have had the opportunity to be one of the first PhD students at the Institute, together with Selcan Mutgan, Martin Arvidsson, and

Niclas Lövsjö. My former years sitting in the same room with them were, without a doubt, the best ones for me. Miriam Hurtado Bodell and Emanuel Wittberg soon joined the team and enriched our discussions with their own perspectives. I look forward to exploring new topics with you for as much time as possible.

Finally, I would also like to thank Madelene Töpfer and Åsa Arnoldson for their support every time I encountered a difficulty in the landing process in Sweden or in surfing the university's bureau. Life could have not gone easier without you in the team.

Index

List of Figures	11
List of Tables	18
1 Kappa	21
1.1 Introduction	21
1.2 Segregation	27
1.2.1 Segregation dynamics	33
1.3 Analytical Sociology	34
1.3.1 Segregation dynamics and Analytical Sociology	35
1.4 Sociological theories of residential mobility	38
1.4.1 Spatial assimilation	38
1.4.2 Place stratification and Ethnic preferences	40
1.4.3 Social structural sorting model	44
1.4.4 Life-course events and residential mobility	45
1.5 The approach employed in this thesis	46
1.5.1 Data	46
1.5.2 Causal inference	48
1.5.3 Agent-based modeling	54
1.6 Conclusion	56
2 The Importance of Neighborhood Size for the Study of White Flight and Ethnic Residential Segregation	61
2.1 Introduction	61
2.2 Literature review	65
2.2.1 Ethnic composition and native out-mobility	65
2.2.2 Administrative neighborhoods and white flight	67
2.2.3 Neighborhoods size and uncertainty constraints	68
2.3 Data and Methods	70
2.3.1 Data	70
2.3.2 Analytical strategy	71

2.4	Results	74
2.4.1	White flight	74
2.4.2	Segregation	78
2.5	Conclusion	81
3	If you move, I move: The Social Influence Effect on Residential Mobility	85
3.1	Introduction	86
3.2	Literature review	89
3.2.1	Neighborhood mobility	89
3.2.2	The Social Influence effect	91
3.3	Data and Methods	93
3.4	Results	96
3.4.1	The SI effect on residential out-mobility	96
3.4.2	The strength of the SI effect on residential out-mobility	98
3.5	Conclusion and Discussion	102
4	Exit or Voice (and When)? Refugee Exposure, Westerners' Residential Mobility, and Radical-Right Support	105
4.1	Introduction	106
4.2	Literature review	109
4.2.1	The refugee crisis in Sweden	109
4.2.2	Ethnic minority exposure, and natives' radical-right support and residential mobility	110
4.2.3	Ethnic minority visibility and superficial contact	111
4.3	Data and Methods	113
4.3.1	Data	113
4.3.2	Analytical strategy	115
4.4	Results	120
4.5	Conclusion	123
	Appendix	127
A	Chapter 2	128
B	Chapter 3	138
C	Chapter 4	145
	Bibliography	150

List of Figures

- Figure 1.1: Segregation as two interrelated dimensions: (1) the extent to which members of one ethnic group are isolated/exposed to members of another group (x-axis), and (2) the extent to which ethnic groups are evenly/unevenly distributed across space (y-axis). 28
- Figure 1.2: Dot distribution map (1 dot = 1 person) of Westerners (sky blue) and non-Westerners (red) for Stockholm municipality. Non-colored areas within the municipality represent inhabited areas. Each dot depicts an individual plotted in his/her residential location, with an accuracy to the nearest $100\text{m} \times 100\text{m}$ square. The more intense the color at a given point, the higher the density of that group at that point. The same information is displayed for two points in time, 1990 and 2017, which correspond to the first and last years of available information at the time this kappa was written. 30
- Figure 1.3: Segregation in four municipalities in Sweden, Stockholm (purple), Göteborg (red), Malmö (cyan), and Linköping (green), for each year between 1990 and 2017. (LEFT) The spatial version of Theil's H index (\tilde{H}), which quantifies the uneven distribution of ethnic groups across space. (RIGHT) The spatial version of the Exposure index (\tilde{P}). The plot shows the extent to which Westerners meet non-Westerners as neighbors (straight line), and the extent to which non-Westerners meet Westerners as neighbors (dashed line). Each measure uses a kernel smoother that follows an exponential decay function to compute the population density at each point, obtained at the level of $100\text{m} \times 100\text{m}$ square (see Reardon and O'Sullivan, 2004). 31
- Figure 1.4: The unfolding of segregation dynamics, adapted from the approach described in Arvidsson and de la Prada (*forthcoming*). 36

Figure 2.1: The importance of spatial proximity for white flight. **(A)** Moving-out probabilities after matching: (1) natives exposed to a proportional increase in minorities ($X = 1$, black), and (2) natives exposed to no change in ethnic composition ($X = 0$, white) (products indicate neighborhood dimensions). **(B)** Difference between predicted probabilities of moving-out among groups in A across neighborhood definitions. **(C)** Marginal differences between natives exposed to a proportional minority increase only in smaller area ($X = 1$, upper label) or outside a bigger area ($X = 0$, lower label). **(D)** Effect heterogeneity of groups in A with different initial minority percentages across one-hundred-by-one-hundred-meter squares. Predicted probabilities, differences between groups, confidence intervals, and p-values in parentheses are gauged using a weighted linear probability model estimated for the matched sample. The straight horizontal line indicates no effect. Upper bars indicate whether differences between models are statistically significant. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$ 75

Figure 2.2: The importance of neighborhood size for ethnic residential segregation. **(A)** Segregation levels generated as a function of neighborhood size. Segregation lines are plotted for three configurations of ethnic minority share: (1) 50%-50% (solid line); (2) 70%-30% (short-dashed line); and (3) 90%-10% (long-dashed line). **(B)** Segregation levels as a function of time (“ticks,” in thousands) for the same three share configurations as in A. Numbers displayed on the right in each subplot show the neighborhoods’ radii. **(C)** Heatmap showing different levels of segregation (the greater the level of segregation, the darker the tile) generated as a function of both neighborhood radius (x-axis) and the proportion of natives in the system (y-axis). Segregation is measured using Theil’s H Index. 79

Figure 3.1: The SI effect on residential out-mobility. Difference in the probability of moving-out between natives exposed to out-movers and no in-movers ($X = 1$) and natives exposed to no change (neither in- nor out-movers) ($X = 0$). The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-value in parenthesis. The straight horizontal line indicates no effect. 97

Figure 3.2: The marginal effect of the number of previous out-movers on the strength of SI. Difference between natives who are exposed only to 2, 3, or 4 or more previous out-movers ($X = 1$, left), or fewer than this ($X = 0$, right), respectively. The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. ‘***’ $p\text{-value}<0.001$, ‘**’ $p\text{-value}<0.01$, ‘*’ $p\text{-value}<0.05$, ‘NS’ $p\text{-value}>0.05$ 99

Figure 3.3: The effect of residential density on the strength of SI. Effect heterogeneity between natives exposed to out-movers ($X = 1$) and natives exposed to no change ($X = 0$) in $100\text{m} \times 100\text{m}$ residential areas with different numbers of inhabitants: (1) less or equal to 15; (2) between 16 and 30; (3) more than 30. The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. ‘***’ $p\text{-value}<0.001$, ‘**’ $p\text{-value}<0.01$, ‘*’ $p\text{-value}<0.05$, ‘NS’ $p\text{-value}>0.05$ 100

Figure 3.4: The marginal effect of distance on the strength of SI. Difference between natives being exposed only to out-movers leaving from a point close to their own residential location ($X = 1$, upper) or from farther away ($X = 0$, lower), respectively for three different spatial levels: (1) $100\text{m} \times 100\text{m}$; (2) $200\text{m} \times 200\text{m}$; (3) $300\text{m} \times 300\text{m}$. The maximum area is bounded by the $400\text{m} \times 400\text{m}$ square at all levels. The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$ 101

Figure 3.5: The SI effect across residential areas with varying ethnic compositions. Effect heterogeneity between natives exposed to out-movers ($X = 1$) and natives exposed to no change ($X = 0$) in $100\text{m} \times 100\text{m}$ residential areas across three types of ethnic composition: (1) only natives (left); (2) less than 10% non-natives (middle); and (3) more than 10% non-natives (right). The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$ 102

Figure 4.1: The refugee crisis in Sweden. (*LEFT*) Number of asylum petitions (2000-2019), with colors indicating the six countries producing the largest numbers of asylum petitions during the refugee crisis. (*RIGHT*) Number of asylum centers (2007-2019). Source: Migrationsverket. 114

Figure 4.2: The importance of the type of refugee exposure for radical-right voting behavior. Differences in the estimated vote-share of the radical-right Sweden Democrat party (SD) in 2018 produced by OLS using the weights from the Synthetic Control Method. $X = 1$ indicates polling districts either receiving new refugees who self-selected into areas (left), or new asylum centers (right). $X^* = 0$ indicates the estimate calculated for the synthetic control from polling districts that neither experienced an increase in refugees presence or new asylum centers. There is one estimate per type of electoral outcome, with results displayed for municipal (white), county (black), and national (grey) elections. Bars indicate standard errors at the 95% level. P-values are in parentheses. The horizontal line indicates no effect. 121

Figure 4.3: The importance of the type of refugee exposure for native out-mobility. **(A)** Differences in the probability of moving-out from the residential square estimated using OLS employing the weights from Coarsened Exact Matching. Estimates are presented for natives exposed to a proportional increase in refugees self-selecting into areas (left) and natives exposed to new asylum centers within the municipality (right) ($X = 1$), each compared to no change in the refugee population or in new asylum centers respectively ($X = 0$). **(B)** The effect of refugee growth across areas with different levels of ethnic minority presence. **(C)** The same as (B), for the effect of new asylum centers across areas with different levels of ethnic minority presence. Bars indicate standard errors at the 95% level. P-values are in parentheses. The horizontal line indicates no effect. Upper bars indicate significance tests between models. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$ 122

Figure A1: Performance of CEM as measured by three different indicators, one per row: (*TOP*) Standardized difference in means for numerical covariates, or difference in proportions for each category for qualitative covariates, for each year. Each value is shown before and after CEM, linking each covariate/category with a line. (*MIDDLE*) \mathcal{L}_1 measure before and after CEM. (*BOTTOM*) Proportion of treated (empty, black-enveloped bars) and control (grey, empty-enveloped bars) cases pruned. The performance of CEM is checked for each pair of treatment groups and each three-year trial analyzed in the main manuscript, displaying the year corresponding to when covariates are measured (i.e., the first one in each trial). . . . 134

Figure A2: Different radii employed in the agent-based model. (*TOP*) Overlapping areas of radii 2, 4, and 6. (*BOTTOM*) Overlapping areas of radii 8 and 10. . . . 137

Figure A3: Performance of CEM as measured by three different indicators, one per row: (*TOP*) Standardized difference in means for numerical covariates, or difference in proportions for each category for qualitative covariates, for each year. Each value is shown before and after CEM, linking each covariate/category with a line. (*MIDDLE*) \mathcal{L}_1 measure before and after CEM. (*BOTTOM*) Proportion of treated (empty, black-enveloped bars) and control (grey, empty-enveloped bars) cases pruned. The performance of CEM is checked for each pair of treatment groups and each three-year trial analyzed in the main manuscript, displaying the year corresponding to when covariates are measured (i.e., the first one in each trial). . . . 142

Figure A4: Diagnostics for CEM and SCM. (*TOP*) Standardized difference in means of each covariate before and after CEM. Each covariate is repeated four times, one per each time CEM is applied through the period 2012-2015. $X = 1$ indicates a native exposed to a new asylum center/refugee growth in the $100\text{m} \times 100\text{m}$ residential square, whereas $X = 0$ indicates the absence of such a new asylum center/no refugee growth. The dashed horizontal line on 0.2 indicate the threshold commonly used to assess the sufficiently appropriate similarity between the treatment groups and which should not be surpassed (Stuart, 2010). (*BOTTOM*) Difference in the mean trajectory of each covariate of the treated polling districts and the one from the synthetic control for the period before the treatment assignment occurs (2006-2014). $X = 1$ indicates a polling district exposed to a new asylum center/refugee growth in the area, whereas $X^* = 0$ indicates the synthetic control. The dashed line on zero indicates where the trajectory of groups are exactly the same. 147

List of Tables

Table 2.1: Basic descriptives for natives exposed to a proportional increase in minority presence and natives exposed to no change in their neighborhood ethnic composition in Stockholm County (1998-2017). Rows show the mean value of each covariate, prior to (Full sample) and after matching. Numerical covariates are presented in terms of their own scale (or logged), and qualitative covariates are shown as proportions. The total number of cases in each group is smaller in the matched sample due to pruning. The values shown may of course change slightly depending on the definition of neighborhood employed. This table only displays the values for neighborhoods defined as one-hundred-by-one-hundred-meter squares due to space constraints, but a complete overview is available in the Appendix. 73

Table 3.1: Basic descriptives for the entire sample of natives exposed to neither in- nor out-movers, and of natives exposed to at least one person previously having moved-out from their residential area and no in-movers. Each row presents the mean value of a covariate, either in terms of its numerical (logged) scale or as a proportion in the case of categorical covariates. Each value is reported for the entire sample prior to matching (Full sample) and after matching. The smaller number of cases in each group after matching is due to pruning. See Appendix for a more detailed report on the matching performance. . . . 96

Table 4.1: Descriptives for natives exposed to a proportional increase in refugees who had sought their own accommodation and natives exposed to no change (2013-2016). Rows show the mean value for each covariate prior to (Full sample) and after matching. Natives are also matched by county, not shown here due to space constraints. Numerical covariates are presented in terms of their own scale (or logged), and qualitative covariates are shown as proportions. The total number of cases declines due to pruning. Similarity is also achieved for natives exposed to a new asylum center and natives not exposed to a new center, omitted here due to space constraints (see Appendix for more information about the performance of CEM). 118

Table 4.2: Descriptives for polling districts in which a new asylum center was established at any point between 2015-2017, and those with no new asylum center during the same period (the donor pool). The table shows the mean value for covariates whose trajectory has been matched annually for the period 2006-2014, save for those election-related covariates that are only matched following each new election (i.e., once per period 2006-2009, 2010-2013, and 2014). The values for polling districts with no new asylum center are shown prior to applying the Synthetic Control Method (Full sample) and after (Synthetic). Cases are also matched by county, not shown here due to space constraints. Similar results are achieved for polling districts in where there is a proportional increase in refugees who had sought their own accommodation and polling districts with no such change, omitted here due to space constraints (see Appendix for more information about the performance of the SCM). 120

Table A1: Statistical estimates per analytical setting. The differences between groups are the beta coefficients of the linear probability model applied on the full sample (Raw) and on the matched sample (matched-LPM) using ordinary-least squares. The models consist of the native out-mobility outcome (binary) for each exposure group being analyzed (also binary). Predicted probabilities, standard errors, t-values and p-values are gauged using the matched sample. 129

Table A2:	Statistical estimates per analytical setting. The differences between groups are the beta coefficients of the linear probability model applied on the full sample (Raw) and on the matched sample (CEM-LPM) using ordinary-least squares. The models consist of the native out-mobility outcome (binary) for the exposure group being analyzed (also binary). Predicted probabilities, standard errors, t-values and p-values are gauged using the matched sample.	139
Table A3:	Statistical estimates per analytical setting. The differences between groups are the beta coefficients of the linear probability model applied on the full sample (Raw) and on the matched sample (CEM-LPM) using ordinary-least squares. The models consist of the native out-mobility outcome (binary) for the exposure group being analyzed (also binary). Predicted probabilities, standard errors, t-values and p-values are gauged using the matched sample.	144
Table A4:	Statistical estimates from models. The differences between groups are the beta coefficients of the linear probability model using the ordinary-least squares method after applying coarsened exact matching (Out-moving, a binary) and a synthetic control method (SD pr. vote share, numeric). The former analyzes a sample of natives, while the latter analyzes a sample of polling districts.	146

Chapter 1

Kappa

1.1 Introduction

Sociology is the scientific study of groups and societies. The basic components of human societies are the human individuals who constitute them, each with their own specificities and characteristics. This is what makes sociology a *social* science. What separates sociology from other social sciences, however, is the fact that sociologists engage with societies as a whole. That is, the focus is directed at *macro*-properties relating to entire groups, properties that no individual could ever exhibit because they simply do not occur on the scale of single individuals. This is what makes macro-properties special and unique in the world. For example, when we look at the total number of connections linking relatives, friends, and acquaintances to one another, we observe a network of relationships whose shape, structure, and properties belong to no single individual but to all of them at once.

Segregation is the macro-property that constitutes the focus of this thesis. Specifically, segregation describes how members of different groups are distributed across certain elements in space. Residences in neighborhoods, schools in districts, and workplaces within companies, are all common examples of such elements. Segregation scholars, then, consider the distributions of groups across such spatial dimensions.

The term ‘groups’ here refers to social categories that exist in the world and

that are, in some manner, meaningful. Common examples include gender, age, and socioeconomic status. The category under study in this thesis is *ethnicity*. Typical groups used within this category include, among others, ‘white,’ ‘black,’ ‘Hispanic,’ and ‘Asian.’ Even though we may perceive ethnicity as an objective trait, the reality is that scholars do not always agree on how it should be interpreted (e.g., Smaje, 1997). For instance, some scholars have argued that the way we look at ethnic groups should be more closely informed by how individuals express and use the ethnic categories with which they identify themselves (e.g., McDermott and Samson, 2005). These scholars emphasize individuals’ subjective experiences as a key factor for understanding ethnic identities. Although these discussions are relevant because they touch simultaneously upon both theory and the way we measure the social world, the conclusion typically reached is that ethnic identities are continually evolving, which produces problems for quantitative sociologists wishing to arrive at meaningful conclusions.

One common way around this dilemma is to measure ethnicity using individuals’ countries of birth. Studies employing this approach would, for instance, assign Swedish ethnicity to individuals born in Sweden. The advantage of this measure is that it is objective and helps reduce the wide variety of possible groups to just a handful. In this thesis, I will focus on the group ‘Westerners,’ a term commonly used to denote the ethnic majority groups in Western countries. The reason for this focus on Westerners is that the data that are available for this group allow me to analyze their mobility choice better than for any other ethnic group. For the purposes of this study, I define Westerners as individuals born either in Sweden or in any other Western European country (E.U.-15), the U.S. or Canada. Individuals born elsewhere, or those born in a Western country to a parent or parents who were not, I classify as ‘ethnic minorities,’ or simply non-Westerners.

The general object of study in the thesis is *ethnic residential segregation* (ERS), i.e., the study of how members of different ethnic groups distribute themselves across residences in a metropolitan area. Of the many aspects associated with ERS, I focus

on the role played by the physical environment in influencing how individuals move and change their home residences. More concretely, I work within a sociological tradition that hypothesizes that physical space structures many of the relevant social relations that influence residential mobility (Logan, 2012). For this reason, for example, Robert E. Park (1924), one of the founders of the Chicago School and a pioneering ERS scholar, suggested that spatially proximate individuals tend to exert the most influence on one another. Similarly, Grodzins (1957) observed that Anglos are more sensitive to the presence of African-Americans living in the same neighborhood rather than some blocks away. Schelling (1971) showed that spatially embedded groups that want to live near a small presence of other co-ethnics can precipitate very high levels of segregation.

By focusing on three interrelated sub-questions, I study the role of the physical environment in determining how proximate and how permanent exposure to other individuals' ethnicity needs to be in order to affect Westerners' residential mobility. The first of these sub-questions asks: *How does spatial distance between ethnic minorities and Westerners influence how salient the ethnic minorities are to Westerners, and how does this modify the Westerners' mobility patterns?* Since the "administrative neighborhood" constitutes the spatial measure most commonly used to capture ethnic mobility dynamics, this first sub-question addresses the utility of this measure, by carefully mapping the strength of out-group exposure at varying geographical scales to see which is capable of prompting Westerners to move out of their current residential area.

The second sub-question asks: *To what extent do previous decisions to leave a neighborhood prompt additional out-mobility among Westerner neighbors?* This second sub-question moves beyond the role of out-group exposure and other factors known to impact residential mobility, and places a new emphasis on the propensity and the conditions whereby individuals may in part imitate former neighbors' past residential mobility, i.e., the so-called social influence effect.

The third and final sub-question asks: *How much does the manner in which*

Westerners encounter a non-Westerner population matter for their propensity to change residential location? I pose this last question to distinguish between two modes through which the refugee crisis of 2015 may have affected residential mobility: asylum centers, which markedly increase the salience of out-group visibility in urban areas for a limited period of time, and refugees with permanent residence permits, whose self-selection into neighborhoods produces less salience but which may increase the appeal for other refugees to move in.

My approach to answering these questions begins by studying Westerners in their physically delimited local environments, and how their *exposure* to relevant social changes in those local environments modifies their mobility patterns. Following a counterfactual design (see Woodward, 2003), I seek to identify the exposure effect by contrasting Westerners who experience the relevant change in their local environments to others who do not experience this change. Using Swedish register data, I embrace a causal inference approach employing different techniques to adjust for confounders and remove any association between exposure and mobility that is not due solely to the exposure.

Each of these questions is addressed in a separate chapter in the thesis. Chapter 2 contrasts patterns of out-mobility following a growth in out-group presence within concentric areas of various sizes centered on Westerners' residential locations, and compares mobility estimates across geographical scales while retaining the administrative neighborhood as the outer bound for the unit of reference. Using matching and linear probability models (see Ho et al., 2007), the analyses show that a growth in out-group presence is capable of producing positive out-mobility to the extent that it occurs within an area bounded by a three-hundred-by-three-hundred-meter square from Westerners' home residences, beyond which there is little to no effect. The shorter the distance at which this growth occurs, the greater the positive effect. Furthermore, the chapter analyzes how these micro-level decisions aggregate to produce segregation patterns using agent-based modeling, and elucidates how space interacts with the total presence of groups to produce greater segregation as agents'

self-defined neighborhoods increase in size. This chapter provides empirical evidence to support those who advocate questioning the use of administrative neighborhoods for social spatial analysis (Logan, 2012), and underscores the importance of using socially and cognitively meaningful areas in order to detect these kinds of interaction effects.

Chapter 3, co-authored with Eduardo Tapia, assesses the social influence effect by focusing on Westerners' out-mobility following exposure to prior out-mobility only among other Westerners. Following Shalizi and Thomas (2011), we find a positive increase in outward mobility among Westerners who have experienced previous out-mobility from their residential area, in contrast to similar Westerners who have experienced neither in- nor out-mobility. Furthermore, we examine in detail how the following four factors contribute to shaping the strength of social influence: (1) the distance from ego, (2) the number of previous departures, (3) the visibility of these departures as measured in terms of the area's population density, and (4) the area's ethnic composition. In line with the analyses presented in the previous chapter, we find that only proximate distances matter, and that greater distances monotonically decrease this effect. A greater number of previous departures and greater visibility resulting from low population density also contribute to increasing the potency of social influence, but we find no support for our hypothesis suggesting stronger effects in ethnically mixed areas. This chapter contributes to studies in the field of ERS by moving beyond out-group exposure and showing the role of social influence as a partial driver of residential mobility.

Finally, Chapter 4 examines the impact of the 2015 refugee crisis, and disentangles the issue of whether asylum seekers contribute to Westerner out-mobility in ways that are not typically found in relation to non-asylum-seeking migrants. The chapter begins by analyzing Westerners' out-mobility first following the establishment of a new asylum center in their local area, and second following refugee in-mobility resulting from their self-selection of accommodation in Westerner-majority local areas. Results show positive effects on Westerner out-mobility only for the latter, and

suggest that the limited period during which asylum centers remained established may have counterbalanced their high out-group salience, finally resulting in no effect on the mobility of Westerners. The absence of this temporary restriction among refugees self-selecting into majority-based areas may have increased the area’s appeal for other refugees (and other non-asylum-seeking migrants) to move into (Clark, 1991), thus retaining the ethnic composition mixed and increasing the likelihood of Westerners to leave their residence. In addition, the chapter uses synthetic control methods (Abadie et al., 2010) to analyze how both modes of exposure affect support for the political far-right. In contrast to the mobility analyses, a significant increase in the far-right vote share is only found in areas with new asylum centers. Overall, the analyses presented in this chapter underscore the contribution made by asylum seekers and out-groups in producing lasting changes in the ethnic composition of an area as a key factor in understanding behavioral outcomes among natives following influxes of asylum seekers in their local contexts.

The results presented in these studies are in line with expectations based on the aforementioned assumption that spatial distance structures social relations with regard to residential mobility and ethnic segregation. The ethnic composition of local environments, the behavior of neighboring co-ethnics, and the proximity and perceived duration of out-group in-mobility are all elements shown by this thesis to partially shape out-mobility patterns among natives.

This thesis makes important contributions to academic and policy aspects of segregation. First, it advances the knowledge base of sociology, a science that is arguably still at a kind of “pre-Newtonian” stage, by describing in unprecedented detail how individuals affect each other and produce complex mobility dynamics. Increasing our understanding of how social dynamics work helps us to develop beyond our current stage, and build toward successful explanations by describing the social cogs and wheels of how segregation works (Hedström and Ylikoski, 2010).

In addition to these scientific aspects, segregation research is important because existing research shows that ethnic minorities in segregated areas tend to suffer with

regard to important welfare indicators (see Sharkey and Faber, 2014). Some of the disparities that have been documented include lower health status, lower second-generation educational attainment, greater exposure to violence, and others. These inequalities constitute a profound problem, not only because they are unfair, but also because they hinder the integration and assimilation of minorities into host societies. By making explicit natives' mobility dynamics on a more detailed scale, this thesis enhances our capacity to make predictions with regard to potential interventions that seek to implement accessible housing or the establishment of temporary asylum centers. The results of this thesis provide a first assessment of how the allocation of measures might be optimized, and of how to prevent an unintended increase in the level of segregation based on the anticipated reactions of natives.

The organization of the rest of this kappa is as follows. There follows a brief exposition of what segregation consists in, which is in turn followed by a discussion of how I study segregation dynamics from an analytical sociological perspective, an exposition of current sociological theories regarding residential mobility in the field of ERS, and an introduction to the methodological aspects of this thesis. The kappa ends with some concluding remarks on the general research question addressed by the thesis.

1.2 Segregation

What is segregation? Reardon and O'Sullivan (2004) proposed a definition based on two dimensions: (1) the extent to which members of different ethnic groups are surrounded by members of any other group throughout space, and (2) the extent to which members of one group are also in touch with members of another group throughout space. The first dimension is known as the *even/uneven* dimension, a measure that focuses on *all* groups and which describes how they are spatially distributed. The second dimension is called the *exposure/isolation* dimension, and describes, separately for each ethnic group, the extent to which members of *one* group meet neighbors from another group. Thus, segregation levels are reckoned

to be high when levels of unevenness and of isolation are both high. The opposite applies for low levels of segregation.

One can picture these two dimensions as a sort of checkerboard, something like Figure 1.1. The Figure depicts four representations of the same checkerboard table, each portraying an ideal representation of segregation along the isolation-exposure dimension (x-axis) and the evenness-unevenness dimension (y-axis). Each square in each checkerboard is black or white, signifying a member of either the Black group or the White group.

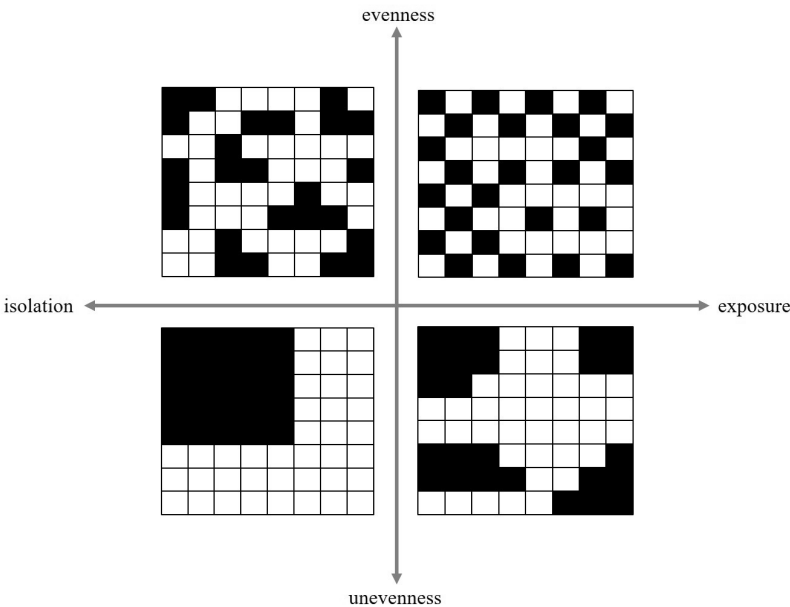


Figure 1.1: Segregation as two interrelated dimensions: (1) the extent to which members of one ethnic group are isolated/exposed to members of another group (x-axis), and (2) the extent to which ethnic groups are evenly/unevenly distributed across space (y-axis).

One key observation, focusing only on the un/evenness axis, is that black squares tend to be more clustered in the uneven case than in the even case. The same is true for white squares. In the even case (top), both blacks and whites appear to be surrounded by members of both their own group and the other group, producing a pattern closely resembling the typical checkerboard used in chess and other games. The opposite is true in more uneven cases (bottom), in which members

of the same group tend chiefly to surround one another (hence the clusters).

The Figure also indicates important differences in the checkerboard patterns captured along the isolation-exposure dimension. This second dimension points to important variations in the extent to which members of each group tend to meet neighbors from the other group along the un/evenness axis. Greater exposure (right) implies members of each group meeting more individual members of the other group, whereas higher isolation (left) entails more exposure mainly to members of the same group. Hence, despite the fact that the checkerboards below the y-axis share a high degree of unevenness, we would conclude that the checkerboard with higher levels of isolation is the most segregated, since, in this case, in addition to being unevenly distributed, group members mainly encounter neighbors from the same group as themselves.

Imagining segregation as a checkerboard is useful because it depicts what we understand heuristically as segregation in highly simplified, abstract terms. However, real systems look rather different from a checkerboard, and we are, after all, interested principally in explaining real systems. Figure 1.2 shows the residential location of each individual living in the municipality of Stockholm on the basis of their membership of one of two ethnic categories, ‘Westerner’ (sky blue) or ‘non-Westerner’ (red), based on the definition presented earlier. The Figure depicts two points in time, 1990 and 2017, providing some sense of how the residential distribution of these two groups has changed over time.

As we can see, in 1990 Stockholm municipality was already ethnically mixed, with Westerners clearly constituting the great majority of residents. Westerners and minorities appear to have been fairly evenly distributed across the municipality, but with more uneven, isolated clusters found around the edge of the municipality, particularly to the north. By 2017, we observe a substantial increase in the share of minorities across the entirety municipality. For instance, we can observe that many of the points with a substantial minority presence in 1990 had an even greater minority presence by 2017, especially in the areas around the periphery of the municipality, in

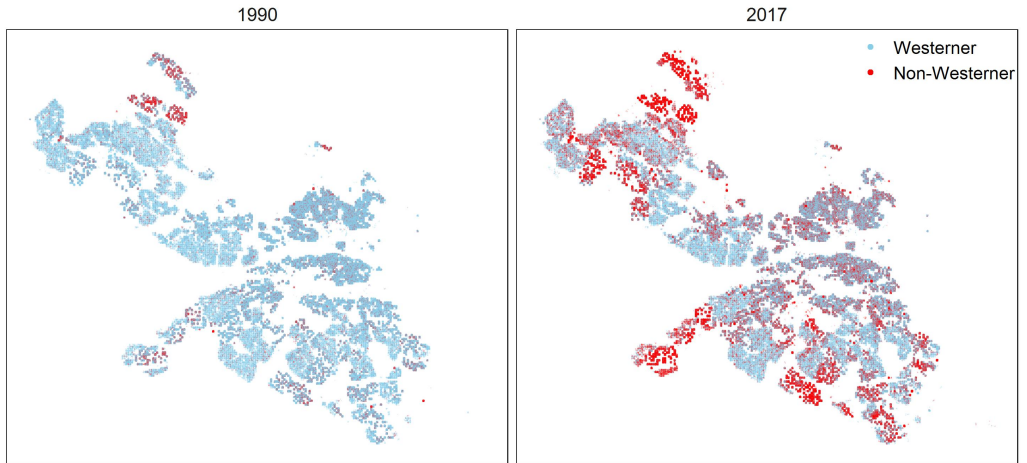


Figure 1.2: Dot distribution map (1 dot = 1 person) of Westerners (sky blue) and non-Westerners (red) for Stockholm municipality. Non-colored areas within the municipality represent inhabited areas. Each dot depicts an individual plotted in his/her residential location, with an accuracy to the nearest $100\text{m} \times 100\text{m}$ square. The more intense the color at a given point, the higher the density of that group at that point. The same information is displayed for two points in time, 1990 and 2017, which correspond to the first and last years of available information at the time this kappa was written.

both the north and the south. At the same time, the map also reveals the presence of larger ethnically “mixed” areas, that alternate with areas of mainly Westerners, around the center of the municipality, which were not present in 1990.

Maps can be particularly useful for studying segregation in that they provide an opportunity to obtain a sense of the level of segregation in a society at a single glance. At the same time, the static character of the information they provide constitutes a fundamental limitation: comparing the situations depicted between one map and another quickly becomes arduous, and the way information is presented in maps leaves them vulnerable to deceptive manipulation by interested parties (Monmonier, 2005). We intend, however, to summarize the information plotted in our map in a more precise and quantifiable way in order to show how segregation differs along each dimension, and how each has evolved over time.

Although scholars have proposed numerous means to this end, only two have been more commonly employed, the Index of Dissimilarity and Theil’s H Index. Both

quantify the uneven distribution of groups, thereby focusing on the first dimension of segregation. Although segregation scholars employ the Index of Dissimilarity more than any other (Taeuber and Taeuber, 1976), Reardon and Firebaugh (2002) have convincingly argued that this index suffers serious shortcomings in relation to measuring segregation. For instance, the index, which primarily facilitates the study of the spatial distribution of two groups (such as Anglos and African-Americans), is highly sensitive to neighborhood size and the size of the minority population used to compute it (Winship, 1977). At the same time, the authors show that Theil's H index not only avoids these problems, but is also capable of being extended in interesting directions.

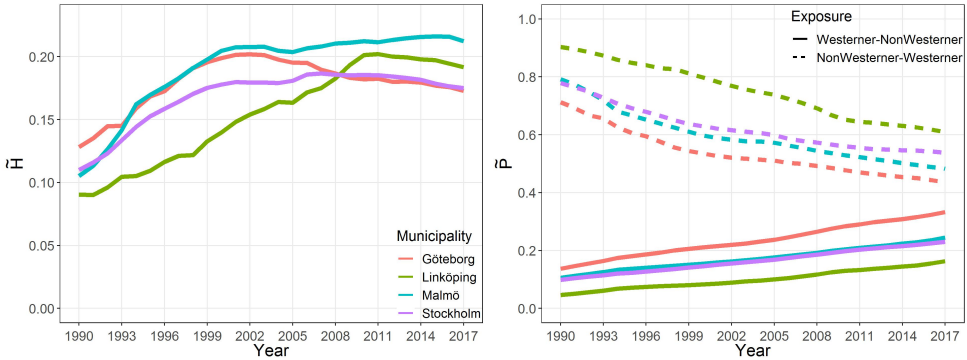


Figure 1.3: Segregation in four municipalities in Sweden, Stockholm (purple), Göteborg (red), Malmö (cyan), and Linköping (green), for each year between 1990 and 2017. (LEFT) The spatial version of Theil's H index (\tilde{H}), which quantifies the uneven distribution of ethnic groups across space. (RIGHT) The spatial version of the Exposure index (\tilde{P}). The plot shows the extent to which Westerners meet non-Westerners as neighbors (straight line), and the extent to which non-Westerners meet Westerners as neighbors (dashed line). Each measure uses a kernel smoother that follows an exponential decay function to compute the population density at each point, obtained at the level of $100\text{m} \times 100\text{m}$ square (see Reardon and O'Sullivan, 2004).

Figure 1.3 shows annual segregation levels in four Swedish municipalities for 1990-2017. Each subplot in the Figure uses a measure proposed by Reardon and O'Sullivan (2004). The plot on the left shows the spatial version of Theil's H index (\tilde{H}), quantifying the extent to which Westerners and minorities are surrounded by members of their own and other groups within each municipality, thus focusing on the y-axis of Figure 1.1. The closer the index is to 1, the more uneven the groups' spatial distribution, and the nearer to 0, the more even. The plot on the right shows

the spatial version of the Exposure Index \tilde{P} , the x-axis of Figure 1.1, which measures the extent to which members of each group encounter members of the other group across space, displayed separately for Westerners (solid line) and minorities (dashed line). Values close to 1 indicate more exposure between groups, while lower values indicate greater isolation.

First and foremost, the plot shows increasing levels of unevenness over the last three decades, not only for Stockholm but also for the second (Göteborg), third (Malmö) and fifth (Linköping) most densely populated municipalities in Sweden. At the same time, these municipalities exhibit distinctive growth patterns. Linköping municipality (green) differs most significantly from the others. Concretely, the patterns of uneven distribution in the Linköping municipality increased almost monotonically during the period 1990-2010, which was then followed by a mild attenuation until the end of the time series. The other three municipalities instead saw larger increases, primarily during the 1990s, followed by sustained or even decreased levels during the 2000s and 2010s. The main exception would be the municipality of Malmö (cyan), whose uneven levels continued increasing almost monotonically for the entire period (although with a flatter slope than that of Linköping).

Two key patterns seem to characterize the trends in exposure between groups for each municipality. First, Westerners seem to have increased their levels of exposure to minorities monotonically, indeed in a rather linear fashion and with no periods of stagnation. Second, the exposure of minorities has declined over the years, with this decline accelerating somewhat in the 1990s, and then continuing at a slower rate. Overall, Figure 1.3 confirms the trend observed in the maps presented earlier, and shows an increase in ERS in Sweden over recent decades, primarily as a result of a growth in the uneven spatial distribution of groups and in the levels of isolation of ethnic minorities.

At this point, we now have some idea about what segregation is, what it looks like in real social systems, and even how it can change over longer periods of time. We soon realize, however, that these impressions are rather unsatisfactory: we have

described what segregation is but not how it works. In order to advance sociological inquiry, we must now study in detail *how* the individuals living in the system move, and how their movements aggregate to produce the levels of segregation we observe. In other words, we must investigate the dynamics of segregation.

1.2.1 Segregation dynamics

The generic name applied to the multiple ways in which a metropolitan area may grow and/or sustain its level of segregation is *segregation dynamics*. One particular aim of this thesis is to contribute to unraveling and improving our understanding of the logic of these dynamics within the field of ERS.

Since segregation quantifies the spatial distribution of groups, we should expect levels of segregation to change as these groups relocate. Thus, examining residential mobility patterns is key to unraveling segregation dynamics.¹ Common explanations in the sociological literature point to the role of physical factors, or some element(s) of the physical environment, in partly determining these residential movements. These include the appearance of nearby buildings, the quality of neighborhood schools, the distance from the city center, or the levels of crime and violence in the area.

In addition, research has focused considerable attention on the *ethnic composition* of neighborhoods as another aspect of the way in which the physical environment modifies groups' patterns of residential mobility (Grodzins, 1957; Goering, 1978). For instance, research in the U.S. has shown that the probability of out-mobility among Anglos increases with the presence of African-Americans in the neighborhood (South and Crowder, 1998; Crowder, 2000; Quillian, 2002; Crowder et al., 2006; Card et al., 2007). Studies that have adopted a multi-ethnic approach (Pais et al., 2009; Crowder et al., 2012) and examined the mobility patterns of natives in Western European countries have produced similar results (Brâmă, 2006;

¹Although these movements also entail individuals moving out/away from other municipalities and even countries, the movements that have received the most attention in ERS are those between dwellings within metropolitan areas. I therefore mainly refer to this type of movement in my discussion of mobility in segregation dynamics.

Schaake et al., 2010; Hedman et al., 2011; Andersson, 2013; Aldén et al., 2015; Boschman and van Ham, 2015; Müller et al., 2018).

The current thesis aligns itself with this strand of research and investigates how Westerners' exposure to ethnicity differences among neighbors in their local physical environments can influence their residential movements. There now follows a general description of the analytical framework employed by the study.

1.3 Analytical Sociology

Hedström and Bearman (2009) define Analytical Sociology (AS) as a conceptual approach to understanding the social world through *mechanism-based* explanations. According to this perspective, we may consider a macro-phenomenon to be explained when the mechanism(s) responsible for its change over time are well defined (Hedström, 2005). This requires a detailed account of the phenomenon's essential constituents, how these parts may interact, and how these interactions can bring about change at the social level over time.

Mechanism-based explanations overcome the shortcomings of other widely used modes of constructing explanations, which are mainly comprised of the statistical model and the so-called covering law model. Sociologists typically employ these means to describe relations between two or more macrophenomena, such as the relationship between high levels of ethnic segregation in a city and disparities in ethnic assimilation. Briefly, these explanatory modes have been criticized for being black-box explanations that quantify the relationships between variables of interest but that offer no account whatsoever of the process(es) that brought the relationship about in the first place (see also Goldthorpe, 2001). Conversely, mechanism-based explanations provide a detailed and clear account of how individuals' actions aggregate to produce such relationships.

There exist several definitions of 'mechanisms' within sociology and the philosophy of science (for a review, see Hedström and Ylikoski, 2010). For the purposes of this kappa, I will adhere to the "minimal" definition of a mechanism articulated by

Machamer et al. (2000) as consisting of individuals and their actions, organized so as to bring about change in a social system. On the basis of this definition, individuals and their actions are just as responsible for *generating* phenomena as the sequential organization of these actions over time (León-Medina, 2017b).

This definition may be preferable to others for two key reasons. First, as Hedström (2005) argues, it merges assumptions regarding mechanisms that are common to other definitions without losing generalizability. Such assumptions include: the importance of actions as the root of causal power in the social world, the role of individuals as carriers of that power, and the chain of sequences of actions and interactions that lead to a social event (Glennan, 2017).

The second reason is that departing from this minimal definition makes important concepts in AS more readily assessable. For instance, the so-called principle of methodological individualism (Elster, 1982) may be reassessed as the principle whereby individuals, their relations, and their actions constitute the main sources of causal power driving social mechanisms. Additionally, it may be possible to restate the Mertonian concept of middle-range theory (Merton, 1949) more explicitly, as generalizing any particular mechanism (involving individuals, their actions, and the specific organization of the mechanism) to be broadly applicable to other contexts.

1.3.1 Segregation dynamics and Analytical Sociology

One important aspect of previous studies on ERS is their reliance on the *percentage of ethnic minorities in the area* as the main independent variable affecting the residential mobility of ethnic groups (the dependent variable) (e.g., Crowder, 2000). Although such analyses are very useful in unraveling behavioral patterns, seeking mechanism-based explanations requires more detail than that provided by the above-mentioned relationship, since it requires a description of individuals, their actions, and their relations with other individuals.

In this thesis, I study how exposure to changes in the physical social environment produced by individuals' prior mobility actions may produce further mobility

among Westerners. Endeavoring to be more precise in describing how this occurs, I develop a framework, based on a forthcoming paper by Arvidsson and de la Prada, that is well-suited to studying segregation dynamics from a mechanism-based perspective. Figure 1.4 presents a sketch of my approach.

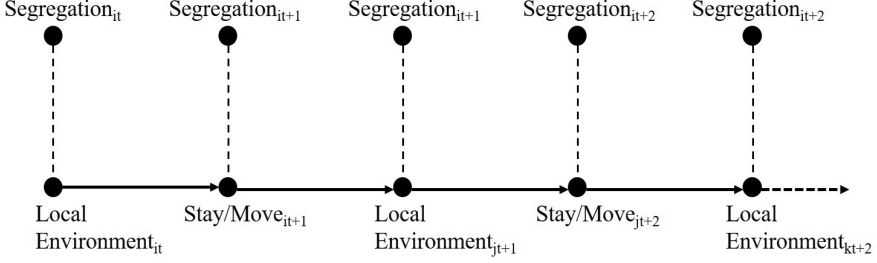


Figure 1.4: The unfolding of segregation dynamics, adapted from the approach described in Arvidsson and de la Prada (*forthcoming*).

Let me begin by briefly describing the components described in this picture. You may note nodes, arrows, and discontinuous lines. The nodes are time-indexed, following from the direction of the arrow, and capture different kinds of information. The topmost nodes encode values of segregation, such as those shown in Figure 1.3. The bottom nodes are of two types: *action* nodes and *local environment* nodes. Action nodes represent individuals' mobility actions, such as moving out or staying. Local environment nodes encode social elements in individuals' local contexts, such as the ethnic composition of the area, that are hypothesized to affect these individuals' residential mobility patterns. Finally, each of these nodes has a subscript indicating an individual, which reflects each individual possessing his or her own unique local environment and taking his or her own actions.

The dynamics of the diagram are represented by the arrows, which encode causal relationships between the bottom nodes in the following way. When we begin studying a social system, we observe a certain macro-state encoded by the level of segregation. At the same time, we observe an individual i from an ethnic group moving out of her dwelling, in this case as a result of the level of co-ethnic presence in i 's local environment being too low. This movement has thereby resulted in a tiny

change in the ethnic composition of the local environments of i 's former neighbors. Let us call one of these neighbors j . The movement of i has lowered the presence of other co-ethnics in j 's local environment, which motivates j to move. This produces another tiny change in the local environments of j 's prior neighbors, who will now decide whether to stay or move. From this point onward, the diagram shows a concatenation of individual actions and minute changes in the ethnic composition of individuals' local environments accumulating over time, gradually shifting towards segregation as a result.

This very simple model of residential mobility illustrates a crucial point in segregation dynamics, the fact of individuals' *spatial interconnection*. When no one moves, the situation viewed from the perspective of any neighbor remains the same, but whenever anyone moves out or moves in, the resulting tiny change in their local environments necessarily alters the local environments of their neighbors. This also illustrates the uniqueness of individuals' local environments throughout space, as each individual's perspective on her physical environment will be unique and will differ slightly from that of any and each of her neighbors.

I thus follow the general logic of segregation dynamics described above to empirically test how the changes in local environments produced by different ethnic groups' prior moving actions may produce further changes in mobility patterns. Despite testing different kinds of local environments and settings, the basic approach used throughout the thesis essentially consists in comparing two very basic situations: (1) a local environment having *changed* in some relevant respect, and (2) another local environment remaining *unchanged* in the same relevant respect. The logic of comparing change with non-change reflects the definition of causation as counterfactual manipulation found in Woodward (2003), quantifying the effect of one variable upon another by outlining the differences between situations in which that cause is present and absent respectively.

By focusing on changes in local environments, this approach allows for a straightforward adjustment for variables that may influence the exposure effect in observa-

tional studies. Before discussing how I implement this approach in practice, however, I will first introduce the theories of residential mobility that have motivated my own research questions.

1.4 Sociological theories of residential mobility

1.4.1 Spatial assimilation

Spatial assimilation theories assign a central role to the way differences in ethnic groups' human capital differentially facilitate access to better-off neighborhoods, with this serving as a basic cause driving a high degree of segregation (see Charles, 2003). From this perspective, improved levels of wealth, education, and income attainment among the minority population translate into lower segregation levels following in-mobility by ethnic minorities into majority-based neighborhoods.

Assimilation is a concept that is related to integration, although not precisely the same. Most importantly, integration has a policy connotation that is not necessarily pertinent to assimilation, which rather seeks to examine to what extent an immigrant population differs substantively from a native population in some relevant way(s). More concretely, assimilation studies seek differences between immigrants and the native population along three general dimensions, (1) the *socioeconomic* dimension, which is basically concerned with educational and wage attainment, (2) *residential patterns*, which are concerned with ethnic segregation, and (3) the *cultural* dimension, which emphasize language acquisition, intermarriage and friendship patterns between members of native and immigrant populations, and immigrants' subjective feelings of belonging to the host society (Waters and Jiménez, 2005; Drouhot and Nee, 2019). An ethnic group's assimilation, then, is reckoned to be high when it draws close to the native population along these dimensions.

Assimilation studies reveal a general, progressive erosion of ethnic barriers in educational and wage attainment for second-generation immigrants, although this process starts with the previous generation coming into the city. According to the

“invasion-and-succession” model (Park, 1936; Aldrich, 1975), immigrants first take residences in so-called *ethnic enclaves*, which are very concentrated areas mainly inhabited by other migrants and that serve as “ports-of-arrival” and facilitate assimilation (Logan et al., 2002). These ports are generally viewed as being transitional, in that immigrants tend to leave as soon as their assimilation allows them to (Brâmă, 2008).

In spite of this, first-generation immigrants tend still to trail natives with regard to many socioeconomic and residential indicators. For instance, they tend to report higher levels of unemployment (Koopmans, 2010), and lower wage attainment (Dancygier and Laitin, 2014), especially those living in highly segregated areas (Thomas and Moye, 2015); they end up with lower-status jobs than natives, even when they have similar labor market qualifications (Constant and Massey, 2005), and they maintain approximately constant wage-levels over their life trajectories (Wessel et al., 2017). More recently, Andersson et al. (2019) revealed that refugees in Sweden arriving at neighborhoods with higher levels of co-ethnic presence showed lower levels of labor-market integration than their refugee counterparts who lived in majority-based neighborhoods five years after arrival.

Thus, the second stage of assimilation consists in seeing to what extent second-generation immigrants can overcome the obstacles faced by the previous generation and become more prosperous than their parents. Although important differences still exist among some groups, the overall conclusion of assimilation studies in the U.S. and Western Europe is that there is a gradual erosion of ethnic, racial, religious, and other differences between immigrants and ethnic majority “native” populations (Drouhot and Nee, 2019). Some studies in the E.U. have documented a greater parity between second-generation immigrants and natives in terms of wages and educational attainment up to secondary school, but a slightly lesser degree in relation to higher education (Jonsson and Rudolphi, 2011; Hermansen, 2016).

Other studies have shown a decrease in ethnic segregation following second-generation assimilation. In the U.S., this has been observed primarily among Asians

(Nee and Holbrow, 2013) and in part among Hispanics (Tran and Valdez, 2015) after reaching parity in educational and occupational attainment with their native contemporaries (Reardon et al., 2009). Furthermore, studies analyzing ethnic groups' mobility patterns have more explicitly documented the impact of human capital on residential mobility, clearly showing that higher levels of education and socioeconomic status increase the propensity of ethnic minorities to move into majority-based areas (South and Crowder, 1998; Quillian, 1999; Charles, 2003; Crowder et al., 2006). Quillian (2012) in particular has shown how ethnic residential segregation follows income segregation in the U.S., with more affluent members within ethnic groups tending to live in areas with fewer co-ethnics who have income levels below their own.

1.4.2 Place stratification and Ethnic preferences

Although scholars agree that levels of segregation have decreased in many metropolitan areas due partly to second-generation immigrants settling in less segregated neighborhoods (Ottensmann, 1995), some ethnic groups, such as African-Americans in the U.S. and Muslims in E.U., still show high levels of segregation after years of supposed assimilation. There are two models that focus on accounting for this incongruence. The first, the so-called *place stratification model* (Massey and Denton, 1993), suggests that Westerners' aversion to sharing residential spaces with minority neighbors reinforces discriminatory practices among real estate agents, landlords, mortgage lenders, and neighbors, which prevent minorities gaining access to housing in majority-based areas, thereby sustaining segregation (Crowder and Krysan, 2016). Prior research has largely reported that ethnic minorities tend to receive fewer callbacks from landlords and have a higher probability of being denied a mortgage (see Riach and Rich, 2002; Ahmed and Hammarstedt, 2008).

Discrimination can also sustain segregation by limiting minorities' assimilation prospects. For instance, Arrow (1973) has shown that minorities underinvest in their human capital, believing discrimination will continue no matter what they do

following previous experiences of discouragement in situations involving “sporadic” discrimination based on distaste. Arrow argues that this form of “statistical discrimination” produces a negative feedback process, so that it becomes rational for employers and mortgage lenders to conclude that immigrant populations tend to be less qualified (Booth et al., 2012), or less capable of paying a mortgage (Dancygier and Laitin, 2014). This negative-feedback process produces job-market segmentation, which exacerbates ethnic disparities and sustains segregation through the disproportionately lower occupational attainment of minorities by comparison with similar natives (Constant and Massey, 2005).

A second hypothesis aimed at explaining patterns of persistent segregation is the so-called *ethnocentric model* (Grodzins, 1957; Schelling, 1971). This model emphasizes individuals’ in-group affinity (Krysan and Farley, 2002) or out-group aversion (Pettigrew, 1998; Zick et al., 2008) as the primary cause of continuing segregation.

Observing ethnic preferences is not a feasible option for social scientists, in much the same way as observing the mind is not feasible for psychologists. Notwithstanding this, some survey studies in the U.S. have found large between-group differences in revealed ethnic preferences. More concretely, Anglos seem to report a far lower tolerance of non-Anglos than other ethnic groups, who seem to be more open to living in ethnically mixed areas (Clark, 1991, 2002). Other mobility studies have indicated that Anglos move out as soon as the proportion of African-Americans residents in a neighborhood reaches 5% (Card et al., 2007; Aldén et al., 2015). On the other hand, African-Americans have been shown to hold preferences for ethnically mixed areas (Clark, 1991). Quillian (2002) has found evidence for this by showing that part of the out-mobility found among Anglos is due to larger influxes of African-Americans into Anglo-based areas. Moreover, Clark (2009) finds that ethnocentric preferences were higher among Anglos of lower socioeconomic status than among others, with higher status Anglos expressing more tolerance for living in ethnically mixed areas. Crowder et al. (2006) arrived at a similar conclusion by analyzing Anglos’ residential mobility patterns based on different levels of wealth.

Further, studies in the U.S. suggest that, despite the importance of socioeconomic status as a predictor of mobility, the likelihood of Anglos' leaving a neighborhood still increases with higher presence of African-Americans in the neighborhood (Crowder, 2000; Quillian, 2002; Card et al., 2007; Bader and Krysan, 2015). A multi-ethnic approach has yielded similar results (Pais et al., 2009; Crowder et al., 2012). More recently, Hall and Crowder (2014) also find evidence of native flight (among both Anglos and African-Americans) in cities that have not traditionally been considered gateway cities for newly arrived immigrants, but which have become or are becoming cities of this kind.

Ethnic mobility studies in Western Europe show similar results to those found in the U.S. Due to the wide diversity of ethnic groups within the native population of Western Europe, the native population of each Western European country occupies the role of a generic "Westerner" population, while non-Western immigrants occupy the role of the minority population. Thus, natives tend to leave their neighborhoods as the presence of non-Westerners increases, even after adjusting for socioeconomic status (Bråmås, 2006; Schaake et al., 2010; Hedman et al., 2011; Andersson, 2013; Aldén et al., 2015; Boschman and van Ham, 2015; Malmberg et al., 2014; Müller et al., 2018).

Despite abundant evidence in support of the ethnic preference hypothesis, methodological limitations render conclusions about the correctness of the hypothesis difficult (if not impossible) to reach. Two considerations explain this. First, the fact that we observe natives leaving as soon as the presence of non-natives exceeds a relatively low threshold does not imply that other ethnic minorities follow similar behavioral rules. For instance, low levels of African-Americans moving into majority-based areas could also reflect discriminatory processes impeding access to those areas.

The second reason relates to the so-called *racial-proxy hypothesis*. This hypothesis states that the propensity of natives to leave neighborhoods as the presence of minorities increases reflects neither a preference for living with other co-ethnics nor

an avoidance of minority groups, but rather avoidance of undesirable lower-class characteristics that are often associated with minority groups (Harris, 2001). This hypothesis highlights the current association between ethnic minorities and the concentration of poverty in segregated areas as an important motive for natives not moving into minority areas.

Numerous studies in the so-called “neighborhood effects” literature have documented the relationship between areas in which minorities are clustered and concentrations of low-social-class characteristics. For instance, children living in neighborhoods with lower-than-median incomes for the city in question tend to present lower levels of school test scores and self-esteem (particularly among African-Americans) (López Turley, 2003; Quillian, 2014), higher drop-out rates from school and a higher risk of teenage pregnancy (Harding, 2003), greater and more lasting levels of verbal disability (among African-Americans) (Sampson et al., 2008); and an increased risk of exposure to or reporting of violence in the streets (Hipp, 2007; Harding, 2009), particularly in ethnically-mixed areas (Legewie and Schaeffer, 2016). Ethnic minorities tend also to settle in areas characterized by worse environmental conditions, such as higher levels of air or noise pollution (Downey, 2006; Crowder and Downey, 2011; Sharkey and Faber, 2014). More generally, minorities living in areas with a greater minority presence exhibit a higher risk of hospitalization (Björkegren, 2018).

The so-called ‘Moving-to-Opportunity’ (MTO) experiment represents a unique study on the health effects of living in distressed areas. Randomly selected families living in neighborhoods below the median income level in the U.S. received vouchers allowing them to move to wealthier neighborhoods. Since its inception in 1994-1998, the experiment has largely confirmed that living in a wealthier neighborhood substantively diminishes feelings of unsafety, reported housing problems, and calls to emergency services, while promoting markedly lower levels of stress (Ludwig et al., 2008, 2012). Furthermore, Chetty et al. (2016) have reported that children younger than thirteen years old of age at the time of the move reported higher college attendance and higher earnings than those who were older when they moved.

This strong association between minority clusters and ethnic disparities is important because it can lead to an entire populace learning to associate ethnic minorities with undesirable characteristics of social class. This means that limitations associated with the manipulation and observation of ethnic preferences preclude the design of any experiment capable of differentiating neatly between the two (i.e., the role of ethnic preferences and the role of racial proxy beliefs) with regard to their effects on the propensity of Westerners to change their residential mobility patterns. If this reasoning is correct, then we currently have no way of determining which hypothesis is correct, since any experiment may capture an inseparable admixture of the two. Of course, this limitation applies only to the extent that we suspect such beliefs to be pervasive within the population (which may not hold true for all societies).

1.4.3 Social structural sorting model

Sociologists typically treat the hypotheses I have summarized as competitive, mutually exclusive explanations of residential mobility. However, as Crowder and Krysan (2016) emphasize, they need not be mutually exclusive: in a given society all might hold to some extent. It is on the basis of this view that Crowder and Krysan have formulated their *social structural sorting model* (see Krysan and Crowder, 2017).

This hypothesis underlines the importance of studying the decision process whereby individuals seek new dwellings as a means of understanding how high levels of segregation *persist* rather than diminishing. The authors hypothesize that individuals' social networks markedly influence the search for and access to new information about new dwellings and neighborhood quality. When the majority of network peers share the same ethnicity (see Lazarsfeld and Merton, 1954; McPherson et al., 2001; Leszczensky and Pink, 2019), and these network peers live in spatial proximity to one another (a well-documented tendency, e.g., Hipp et al., 2012; Preciado et al., 2012; Verdery et al., 2012; Smith et al., 2016; Small and Adler, 2019), then the information transmitted along these channels potentially describes areas

that are largely inhabited by members of the same group, reinforcing segregation patterns in the absence of either an explicit choice to inhabit areas with greater concentrations of co-ethnics or the need for any discrimination.

This model also distinguishes between two separate stages in the moving decision. First, individuals primarily decide which neighborhoods or areas are suited to their living preferences and which are not. Second, individuals select dwellings in the areas regarded as suitable, focusing on their concrete characteristics. Although this hypothesis is rather new and there is little supporting empirical evidence, Bruch and Swait (2019) have shown that dividing the decision process into these two stages, for example by first reducing the entire choice set using some heuristic rule and then analyzing the factors that influence residential mobility, can substantially modify the conclusions researchers might draw from data by comparison with the conclusions that would be drawn if they did not separate the decision process in this way. Similarly, a vignette study by Bader and Krysan (2015) has shown that Anglos tend primarily to discard highly non-Anglos areas before considering exactly where they wish to move.

1.4.4 Life-course events and residential mobility

Finally, I would like to mention the role of life-course events in shaping many households' moving decisions. Demographers have championed this approach, hypothesizing that life-course events play an important role in shifting individuals' space-related housing needs, and increase their likelihood of moving whenever those needs change in some relevant way. Studies have shown that events such as having a child, marrying, or reaching adulthood, can serve as important triggers in motivating the search for new dwellings (Clark and Onaka, 1985; Clark and Dieleman, 1996; van Ham and Clark, 2009).

1.5 The approach employed in this thesis

1.5.1 Data

Throughout the thesis, I use Swedish register data for the years 1998-2017, a unique dataset that is suited to the study of mobility dynamics for three reasons. First, it contains individual-level census information for everybody who has lived in Sweden at any time within the period under study. Second, its unique longitudinal coverage allows for the detailed study of how the impact of a certain type of exposure has changed over time. The greater the available number of repeated observations over time, the more confident I can be that the accuracy of the statistical inference is reliable. Finally, the registers are not only comprehensive in their coverage and provide data for a long period of time, they are also very detailed. The quantity of data to which I have been granted access allows me to track not only the extent to which individuals change residential locations, but also important variables that are central to the analyses presented in the thesis, such as life-course covariates, socioeconomic status (e.g., number of years of education and disposable income (including welfare aids from the Swedish state)), country of birth, ownership and type of tenure tenancy of the dwelling, and more.

Most importantly, these features allow me to study spatial interaction effects in ways that are typically unfeasible on the basis of standard random-sampling studies, which constitute a more common approach in quantitative social science research. As noted by Coleman (1986), random-sampling techniques have enabled social scientists to study individual behavior rigorously, to the extent that most social science research has shifted from its original focus on social to individual behavior. However, the data obtained from random sampling lacks precise measures of individuals' interactions and the influence they exert on one another, which is necessary for the study of social behavior.² For this reason, the census-based nature

²At the same time, other sampling methods could circumvent this difficulty via the use of random sampling among clusters of people, as proposed by Sobel (2006), thus overcoming violations of the “no interference” assumption of randomized controlled trials. Sampson (2008) proposes something similar within the framework of *ecometrics*, i.e., applying treatments to entire neighborhoods

of the registers and their high granularity, provide a unique opportunity to study potential interactions and influences among physically proximate individuals in ways that traditional surveys simply cannot.

Westerners comprise the main study population examined in the thesis. This group represents the largest ethnic group in the metropolitan areas under study, and the detailed information available for this group allows me to adjust for factors that previous research has identified as affecting out-mobility.

I also classify ethnic minorities as non-Westerners, including second-generation immigrants. This approach is common in ethnic mobility studies in Western Europe (e.g., Müller et al., 2018), which often involves the inclusion of second-generation immigrants as a result of their still-in-progress assimilation, since this precludes them from being equally comparable to their native-Westerner counterparts.

The main independent variable employed throughout the thesis is a binary variable (0/1) indicating the absence/presence of a relevant change in Westerners' local environments. The main dependent variable is Westerners' out-mobility patterns. My decision to study out-mobility rather than in-mobility follows a belief, in line with that of Grodzins (1957), that this more clearly captures when a certain exposure in the local environment really matters to Westerners. It also greatly simplifies the complexity of the dependent variable, by also reducing it to a binary indicating remaining or moving.

I apply causal inference methods to the register data, which, together with agent-based modeling, constitute the main methodological approaches employed in the thesis.

and studying their effectiveness by means of comparisons with other neighborhoods that are chosen as controls. Interestingly, this is basically the reasoning that underlies the role of “worlds” in the design of macro-experiments, in which each randomly allocated social group represents a single data point representing a unique instantiation of a collective behavior (Hedström, 2006; Salganik et al., 2006).

1.5.2 Causal inference

One important methodological aspect of this thesis is its general approach to causal inference. Generally stated, causal inference comprises a set of statistical tools devised to extract meaningful information about causal relationships from data that are not collected using a randomized controlled trial (Hernán and Robins, 2020). Causal inference is therefore useful for drawing conclusions about the potential causal chains whereby exposure to certain changes in Westerners' local environments may increase their likelihood of moving out, particularly in comparison to having been exposed to no change.

Most importantly, causal inference provides the means to which this comparison may fairly be applied. The approach employed in this thesis mainly follows the strategy of adjusting for confounder variables (covariates) (see Pearl, 2010). This involves removing the associations between independent and dependent variables that are not due to the independent variable alone, but that instead arise from common factors that causally determine both the independent and the dependent variables. The research literature that I have presented in this kappa is thus essential to determining the factors that need to be adjusted for. Following the adjustment procedure, and provided we believe the assumptions that inform this procedure (Heckman, 2005), we will have increased the plausibility that the reported effects truly are actually closer to what we would observe in a randomized controlled trial (Dunning, 2008).

I will now briefly sketch the three adjustment methods that are used to gauge causal estimates: coarsened exact matching, the synthetic control method, and the weighted linear probability model, and why I think they may assist in this research.

Coarsened exact matching

In broad terms, *matching* is a general adjustment strategy devised to ensure that the groups under comparison (typically encoded on the basis of the independent variable) also look similar in terms of the covariates that are adjusted for (Stuart,

2010). Many have argued that matching is a more efficient means of adjusting for covariates than regression adjustment alone because the presence of groups that are overly dissimilar, perhaps as a result of some form of non-linear relationship between the covariates and the independent variable, prevents the possibility of making any fair comparison between them (Deaton and Cartwright, 2018).

We might discover a situation like this, for example, if we wished to test the efficacy of a new drug by *observing* individuals who took it and others who did not. Imagine that we observe a mainly elderly population taking the drug, and a young population who mostly do not take the drug. If we proceeded by simply comparing the two groups using regression alone, adjusting for age, our estimates would suffer from (1) larger standard errors and greater uncertainty (Rosenbaum and Rubin, 1983), (2) model dependence (i.e., the level of dependence on the particular functional relationship assumed to gauge the estimates would be very high) (Ho et al., 2007), and, in the case of a non-linear relationship between covariates and the independent variable, (3) lack of robustness (i.e., the model estimates would not approach the real estimate as the sample size increases).

Conversely, if we pruned young individuals from our dataset while matching similar elderly people who did or did not take the drug, our estimates would approach the real estimate with increasing sample size (even with a non-linear relationship between covariates and the independent variable) due to a weakening of model dependence (i.e., the estimates produced using different functional assumptions would yield similar results).

Any matching method proceeds in three stages (Stuart, 2010). First, the researcher determines a proper measure of *distance* to determine the scale at which two observations might be closer to or farther from one another. The second stage consists in applying a *matching algorithm*. Although the details vary between algorithms, in this stage of the process the matching algorithm takes the measure of distance into account to determine whether or not a given match is good. Finally, the third stage consists in assessing *balance*, that is, checking whether the match-

ing did a good job in improving the degree of similarity between the groups to be compared in relation to the independent variable.

Several matching methods are available to the researcher. One widely used method is propensity score matching (Rosenbaum and Rubin, 1983). This method first computes the propensity for each observation to fall into any of the groups of the independent variable given the observation’s values on the covariates, and then matches it with observations with a similar propensity. A clear advantage of this method is that it reduces the number of covariates to only one, the propensity score, which allows for the matching of individuals even in settings with hundreds of covariates.

However, some scholars have criticized propensity score matching for trying to emulate *full* randomized controlled trials. More concretely, their criticism is not that propensity score matching emulates a randomized trial, but rather that *blocked* randomized controlled trials are a much more efficient means of computing causal effects, as their design ensures a random allocation to groups that also look similar in respects that are hypothesized to have a significant impact on variations in the causal estimate between different subgroups of a given sample (Deaton and Cartwright, 2018).

King and Nelson (2018) have shown that this distinction may be important in relation to causal inference. By relying on computational experiments, these scholars have shown that propensity score matching may become inefficient at some point, by discarding observations in a way that actually *increases* dissimilarity between groups of the independent variable. This is because propensity score matching guarantees an expectation of group similarity, rather than similarity in any given sample (Iacus et al., 2011). Deaton and Cartwright (2018) also raise this important point in connection with randomized controlled trials, and underline the role of blocked designs as a potential solution for ensuring *in-sample* similarity (as opposed to *expected* similarity) for any given trial.

To overcome these issues, Iacus et al. (2012) devised *coarsened exact matching*

(CEM). In a nutshell, CEM is a non-parametric matching method from the family of Monotonic Imbalance Bounding (MIB) methods (Iacus et al., 2011). This family generalizes from the so-called Equal Percent Bias Reducing (EPBR) methods introduced by Rubin (1974), which rather than attempting to improve the expected imbalance between groups of the independent variable, instead seek to reduce the actual in-sample imbalance between groups on each covariate separately and within different regions of the covariate space, as chosen *ex-ante* by the scientist. This class of methods guarantees that the addition of one variable or a change in the variable’s coarsened values will not alter the maximal imbalance on the remaining variables.

The non-parametric nature of CEM allows us to avoid reliance on assumptions about the functional relationship between covariates and the assignment to groups, which can be useful when such relationships are highly non-linear. Another clear advantage of CEM is its great simplicity. In a nutshell, it begins by applying the binning function, which coarsens each covariate separately. In the next step, the algorithm generates the multivariate cross-tabulation of all the covariates. Finally, CEM assumes that observations within each multidimensional cell are equal, save for which group they fall into, and applies a weight to each observation. Cells with observations belonging only to one group (treated or control) receive a weight of 0 (i.e., exclusion from the analysis), while the remainder receive a positive non-zero weight.

CEM is an essential tool in the context of this thesis. To assess the performance of CEM on each occasion, I report standard measures found in the literature to gauge the improvement in similarity produced by the matching procedure, such as the so-called \mathcal{L}_1 index to measure the similarity between groups across the entire multi-dimensional space of covariates (see Iacus et al., 2012), and the standardized difference in means, which examines the improvement in similarity separately for each covariate (Stuart, 2010).

Synthetic control method

In addition to CEM, I apply another technique called the *synthetic control method* (SCM) (Abadie, 2003), an improved version of the so-called difference-in-difference method from the econometric literature. As such, applications of SCM mostly involve aggregated data, such as for census areas, municipalities, or even countries, at least one of which happens to have received an intervention whose effects we wish to quantify. The primary strategy used for these kinds of settings is to apply a difference-in-difference approach, with another set of aggregated areas that did not receive the intervention serving as controls to compute the effect on the treated area. To ensure that areas are comparable to one another, this method relies on the so-called *parallel-trends assumption*, which states that the treatment area would have behaved in the same way as the control areas in the absence of the intervention. To satisfy the parallel-trends assumption, the trajectory of the dependent variable for the aggregated areas in the control group must resemble that of the treated area fairly well, at least up to the time of intervention. However, as Abadie et al. (2010) have noted, the criteria used by researchers to select a set of control areas typically lack complete transparency. Moreover, it turns out to be extremely difficult to find two or more aggregated areas that look largely the same regarding the trajectory of the dependent variable (and also that of their covariates) so as to satisfy the parallel-trends assumption.

It was in order to overcome this limitation that Abadie et al. devised the SCM. This method is, in essence, very similar to the difference-in-difference method, but with the important qualification that it ensures that the trajectory of the dependent variable (and the covariates) in the area receiving the intervention is as close a match as possible to the trajectory in the aggregated areas that do not receive the intervention. This is achieved by devising a “synthetic” control, an observation (that does not properly exist in reality) that is constructed on the basis of information relating to aggregated areas that did not receive the intervention (the donor pool). In practice, this involves reweighting each donor so that the sum of all the donors

reflects the same trajectory as that of the treated area up until the time of the intervention.

This new synthetic control ensures a safe comparison between the intervention aggregated area and another that did not receive the intervention. The main distinctive feature of SCM relative to the difference-in-difference method, then, is the computation of the synthetic control against which the effect of the intervention will be compared.

An important limitation of SCM as devised by Abadie et al. is that the method only works when the intervention is applied to only one area. In this thesis, however, I apply SCM to a scenario in which more than one municipality received the intervention, and I therefore follow Robbins et al. (2017) and utilize the information from all treated areas to compute a synthetic control that matches the mean trajectory for all of the intervention areas (see Chapter 4).

Weighted linear probability model

In all papers, I gauge the estimated effect of the independent variables on the dependent variable by fitting a *weighted linear probability model* subsequent to CEM or SCM using the ordinary-least-squares method. Linear regression is typically appropriate when the dependent variable is numerical. For binary outcomes, such as that used in this study, logistic regression is, at least in sociology, the method that is usually recommended for fitting a statistical model. One important reason for this is that logistic regression, unlike linear regression, never makes predictions outside the range 0-1. However, logistic regression has a notable drawback that linear regression does not suffer from, which renders logistic regression less desirable in practice. Stated briefly, the drawback is that each coefficient from a logistic regression is standardized according to the true variance (which is presumed to be fixed). This means that only a *correctly* specified model will produce coefficients that, when scaled by the true variance from the regression fit, equal the true coefficient (Mood, 2010).

This scaling factor would not be a problem were we analyzing data from, say, a randomized controlled trial. In such a case, because the expectation is that unobserved confounders are excised from the equation by design, the model’s error term will not be correlated with the independent and dependent variables, and thus we would get the same estimate using logistic or linear regression. However, in observational data (as in any particular trial of a randomized controlled experiment) unobserved heterogeneity is likely to affect the independent and dependent variables in some way, perhaps due to the existence of unanticipated confounders, high measurement error, attrition, or even because groups might be too dissimilar to one another. In the presence of unobserved heterogeneity, therefore, the variance at which coefficients are standardized is not the true variance, implying that coefficients from the logistic regression fit will carry some unknown amount of misinformation.

An important resultant practical problem is that different model specifications are likely to include either more or less unobserved heterogeneity relative to any other model fit. In practice, this implies that we cannot compare estimates produced using logistic regression across different specifications, samples, within samples, or even over time, as each model’s unique baggage of unobserved heterogeneity is likely to affect the size of the coefficients in different ways. Nevertheless, one of the important analytical tasks in this thesis involves being able to draw meaningful conclusions on the basis of different model specifications and within-sample variation. Since the goal is mainly that of comparing the dependent variable over groups relative to the independent variable rather than making predictions, fitting a linear probability model is more desirable than the use of logistic regression, given that linear regression does not suffer from the shortcomings of the scaling problem associated with logistic regression (*ibid.*).

1.5.3 Agent-based modeling

CEM, SCM, and linear regression are tools devised to produce causal estimates based on observational data. Causal inference is not, however, the only methodology

available, and the literature on AS in fact encourages the use of other methods, especially *agent-based modeling* (ABM), to study mechanism-based explanations (see Hedström, 2005).

The ABM method differs greatly from causal inference, and its use among sociologists has increased in the recent decades (Bianchi and Squazzoni, 2015). Rather than operating in terms of “variable”-based thinking, as when we work with independent and dependent variables, ABM uses a language that describes behaviors and interactions among a set of artificial agents (Macy and Willer, 2002). More concretely, ABM is a simulation methodology that allows us to program artificial agents with certain behavioral rules. The agents then act in accordance with these rules and interact with one another, producing changes in their artificial environment. Beyond this point, the details of any ABM model will depend on the research at hand, in terms of both the rules agents follow and the outcome under analysis.

This approach to social research complements causal inference in several ways. First of all, some have argued that agent-based models are particularly suited to sociological understanding (Ylikoski, 2014), mainly as a result of their ability to study social mechanisms using the computer as an artificial “lab” to analyze social interaction (Hedström and Bearman, 2009). This is in part due to ABM’s great flexibility, which allows for agents’ behavioral heterogeneity to be encoded in ways that are typically unfeasible in standard differential-equation modeling (Axtell, 2000) while still using a formal language (Simon, 1978). At the same time, ABM can incorporate insights from experiments and causal-inference research, such as when the estimates obtained from a survey are used to inform agents’ behavioral rules (Bruch and Mare, 2006). This allows for analysis of the dynamics of a system under the complete control of researchers, and for the scrutiny of how systems change as some relevant parameter value varies. This type of “possible”-counterfactual analysis is unfeasible on the basis of causal inference alone (cf. Pearl, 2013), although the empirical focus of the causal-inference approach can be used to test predictions derived from ABM (Centola and Baronchelli, 2015).

The use of ABM in this thesis serves mainly as a means of obtaining information about the way agents whose mobility depends on an area's ethnic composition contribute to producing segregation, and how segregation changes as the size of such areas also changes. This macro-level analysis then complements the empirical analysis of how an area's actual size is estimated to affect individuals' micro-level decisions (see Chapter 2).

1.6 Conclusion

Does the physical environment and the social elements of this environment play a role in influencing Westerners' mobility? The analyses performed in the studies presented in this thesis provide empirical evidence that supports this proposition. More concretely, the thesis shows how prior mobility decisions are capable of producing additional mobility by elucidating how the physical components of local environments moderate the effects of individuals' exposure to neighbors' ethnicity to produce these interaction effects.

Two main *social* aspects of Westerners' local environments are shown to produce out-mobility: (1) the presence of ethnic minorities, and (2) the previous mobility of in-group members. One clear result shown in this thesis is that ethnic composition partly drives individuals' mobility and determines their predisposition to stay in or move out of an area. In line with previous research, analyses in this thesis show that a greater presence of non-Westerners is capable of producing native out-mobility. The effect of non-Westerners moving in is substantially mediated by the prior ethnic compositions of areas, and the analyses show that a growth in the presence of ethnic minorities is capable of sparking native out-mobility when the minority population reaches 5%-10% or more. Little to no effect is shown for areas below this value. The seemingly high susceptibility around this critical value suggest the presence of discontinuous or "tipping" dynamics in native mobility, which has previously not been identified (Ottensmann, 1995) due to a reliance on the use of administrative neighborhoods as the spatial unit of analysis (Grodzins, 1957).

The same conclusions would apply to asylum seekers from non-Western countries entering an area's housing market, as compared to non-asylum-seekers migrants, although not necessarily to those who are living temporarily in an asylum center. The higher out-group salience produced by those centers would not be capable of producing out-mobility even for those natives living in ethnically mixed areas, and thus remain apart from the conclusions reached for non-Westerners (both asylum seekers and non-asylum-seekers migrants) choosing their own accommodations.

Beyond the role of ethnic composition, the thesis shows that in-group influence also contributes to partly driving the out-mobility of co-ethnics. At least in situations in which Westerners' previous exposure is limited to out-mobility only among Westerner-neighbors, the probability that they will imitate this behavior is greater than in situations in which no change occurs (everything else remaining constant). While this result is clear for areas in which the current non-Western population is null, it is also possible that this effect will interact with ethnic minority exposure in ethnically mixed areas, potentially accelerating ethnic change in those areas as a result of some combination of social-influence and ethnic-preference dynamics.

As regards the *physical* component of local environments, three factors are shown to describe the role of space in producing out-mobility: (1) spatial proximity, (2) visibility salience, and (3) temporal duration. Spatial proximity increases the capacity of two individuals to influence each other, as long as their residences are located within three hundred meters of one another. Beyond this point, individuals exert negligible influence on each other's mobility dynamics, which are instead affected by others located at closer proximity. Overall, the analyses show a monotonically decreasing effect the larger the distance between two individuals, and call into question the use of administrative neighborhoods as a means of capturing residential mobility dynamics.

In addition to spatial proximity, individuals' visibility is also affected by other factors that are equally important in producing out-mobility. Thus, the lower an area's population density and the higher the number of individuals who move out

of the area, the greater the “strength” of individuals’ influence on others who are spatially proximate. Finally, the analyses also reveal an interaction between these factors, moderating the salience of ethnic groups and the duration of the period for which an ethnic change is perceived as being likely to last. Thus, even large out-group concentrations may exert a negligible influence on other individuals if it is perceived that this change will not last over time. This logic also applies in reverse, in that low out-group exposure may be sufficient to affect natives’ out-mobility if other factors, such as a moderate non-Westerner presence in the area or a high influx of out-members, contribute to producing a belief that an area’s ethnic composition will remain mixed.

This thesis also contributes to the development of AS by detailing a simple analytical framework that may provide insights into mechanisms and facilitate their discovery on the basis of empirical data. Although AS is mainly concerned with outlining processes that bring about social phenomena (Hedström and Swedberg, 1996), these processes, as they are typically conceived in sociology (e.g., Coleman, 1984), tend to involve some causal effect from a relevant social factor or social interaction to an action produced by individuals. This framework allows us to detail what these key interactions are as a series of ‘social environment \rightarrow action’ relationships that can be tested empirically as regular causal effects. Provided that on the one hand sufficient information about the social system is available on a longitudinal basis, and on the other that the problem at hand allows for the application of causal inference or the design of an experiment, this analytical framework may be of utility for the formulation of theoretical mechanisms in terms of falsifiable causal relationships, thus facilitating their empirical tractability.

Furthermore, this framework is sufficiently general to be able to accommodate different classes of social interactions and social settings. As the thesis shows, this analytical framework has enabled the study of no less than three distinct “types” of local social environments (i.e., based on ethnic composition, in-group out-mobility, and asylum seeker migrants), and there is nothing that would preclude its application

to other social settings. Moreover, the framework is flexible enough even to be able to describe more complex social interactions, such as, for instance, the effect of a sequence of two or more changes in individuals' local environments on some behavioral outcome (see also Pearl and Robins, 1995).

The thesis also has some important limitations. The most important of these regards the implementation of causal inference and the requirements underlying the so-called “ignorability” assumption (Rubin, 1974). This not only requires that the main confounder variables have been properly adjusted for, but also that there have been no violations regarding the principle of “no interference” (Sobel, 2006), or the so-called Stable Unit Treatment Value Assumption (SUTVA). Generally stated, this assumption means that individuals do not interact within or across groups, so that differences observed in the dependent variable are due solely to the assigned group. It is likely that this is not the case in social systems based on observational data, however, since information about ethnic composition and mobility plans can flow without restriction among individuals embedded in social networks. This poses a serious limitation on the extent to which researchers might be able to identify complex social relationships using solely observational data. The design and implementation of macro-experiments (Salganik and Watts, 2009), or the advancement of methods capable of estimating the amount of error due to violations of SUTVA, might represent two possible solutions to this problem.

One important aspect of segregation dynamics—which is overlooked here—involves the careful depiction and dissection of the complete sequences of events that affect segregation. The focus here has been on analyzing one key aspect of such processes at a time, thus setting aside a comprehensive study of the concatenation and chain of sequences of events as they unfold over time. Storytelling descriptions about processes typically take the place of careful descriptions of these dynamics in the sociological literature (León-Medina, 2017a). However, more is needed. The ambition of AS is that of making more detailed and precise descriptions of these dynamics a mandatory requisite for explaining social phenomena and fostering so-

ciological insight.

Chapter 2

The Importance of Neighborhood Size for the Study of White Flight and Ethnic Residential Segregation

Abstract

The concept of “neighborhood” is crucial for understanding mobility dynamics in ethnic residential segregation, but it is highly unclear what spatial unit the concept refers to. Prior research has relied on census areas, but these large areas imply unrealistic individual-level cognitive assumptions, seriously limiting their utility. In this first empirical chapter, I assess the relationship between mobility patterns and levels of segregation using different neighborhood sizes and register data for Stockholm County (1998-2017). The empirical results show that the estimated size of the contextual effect is highly dependent on the neighborhood definition employed. Only events taking place in close proximity to individuals’ homes have any measurable impact on residential mobility.

2.1 Introduction

Many residential neighborhoods in contemporary urban landscapes display patterns of segregation along ethnic lines (Musterd, 2005; Lee et al., 2008). That is, West-

erners and ethnic minorities tend to live in communities composed mainly of their own ethnic group. Numerous studies have shown that this separation into distinct neighborhoods partly contributes to driving ethnic disparities along important welfare dimensions (Massey, 2016) such as health status (Ludwig et al., 2008; Crowder and Downey, 2011; Chetty et al., 2016), wage and income (Reardon and Bischoff, 2011; Ludwig et al., 2012; Thomas and Moye, 2015), and educational attainment (Harding, 2003, 2009; Sampson, 2008).

The neighborhood is widely accepted as a fundamental unit in relation to understanding many social processes that are markedly influenced by space (Sampson et al., 2002), including segregation and clustering processes. For instance, in analyzing inter-neighborhood mobility, Grodzins (1957) assumed that the neighborhood would constitute the space in which Anglos were most reactive to an increased presence of African-Americans. A number of studies, both in the U.S. (South and Crowder, 1998; Crowder, 2000; Quillian, 2002; Card et al., 2007; Bader and Krysan, 2015) and in Western Europe (Bråmă, 2006; Schaake et al., 2010; Andersson, 2013; Aldén et al., 2015; Müller et al., 2018), have since empirically documented the role of neighborhoods in determining natives' mobility responses to changes in ethnic composition.

Although this research has increased our understanding of how segregation can arise and endure, it is highly unclear what spatial unit the “neighborhood” exactly refers to. Data limitations have led many researchers to report inter-neighborhood mobility patterns for neighborhoods defined primarily in terms of administrative boundaries or census areas (Crowder and Krysan, 2016). However, census areas tend to be large, and their definition most often reflects criteria that are quite distinct from the interests of sociological research (Galster, 2001). Since we as yet know little about the impact of neighborhood definitions, using census areas may potentially introduce bias in unknown ways.

For our purposes, for example, treating census areas as neighborhoods entails ascribing the inhabitants of these areas, not entirely realistically, the almost unlim-

ited cognitive power necessary to follow and analyze the ethnic composition of these very large urban spaces (Crowder and Krysan, 2016; Bruch and Swait, 2019). The use of census areas, then, could lead to an underestimation of contextual effects on the mobility patterns of natives who may care more about the local ethnic composition at more proximate scales, and may ignore or be unaware of changes farther away within the “neighborhood” (Todd and Gigerenzer, 2003; Bruch and Feinberg, 2017).

Furthermore, the way we define neighborhoods significantly determines our assumptions about what ethnic change may look like over time (Fowler, 2016). As Logan (2012) has argued, any definition of the neighborhood implies assumptions about how spatial distance moderates inter-group exposure, about how contiguous neighborhoods are treated (Reardon and O’Sullivan, 2004), and about how social processes aggregate over large geographical areas (O’Sullivan, 2009). These assumptions stem from the so-called modifiable areal unit problem (MAUP), whereby processes that are heterogeneous at lower scales are treated as being homogeneous at higher scales (Hipp, 2007). Most importantly for inter-neighborhood mobility, this implies that the definition of “neighborhood” both determines our estimations of mobility patterns and constrains how the processes that drive neighborhood ethnic change will unfold over time, e.g. whether the ethnic composition of neighborhoods “tips” suddenly (Grodzins, 1957; Goering, 1978), or transitions smoothly (Molotch, 1969; Ottensmann, 1995).

The main contribution of this chapter involves assessing how mobility estimates change with changing definitions of neighborhood, and the consequences of this with regard to how we explain segregation. To assess the effect of neighborhood definitions on mobility estimates, I study the propensity of natives to move from their home residence following exposure to a proportional increase in non-Westerners in “neighborhoods” of varying size. I apply matching (Stuart, 2010) to register data for all natives living in Stockholm County (1998-2017). Finally, I use an agent-based modeling approach to detail how different neighborhood definitions affect our

understanding of the role of the processes that generate segregation.

Empirical results indicate three main findings. First, the analyses reveal no selective out-mobility by natives from neighborhoods defined as census areas. Second, as soon as the definition of neighborhood encompasses noticeably smaller areas, the models show out-movements by natives who are exposed to an increased presence of non-Westerners. More concretely, the effect appears strongest in neighborhoods defined as encompassing individuals' most proximate environments, here defined as one-hundred-by-one-hundred-meter and three-hundred-by-three-hundred-meter squares. Increases in non-Westerners beyond this point appear to have little to no effect in prompting native out-mobility. Overall, the models show a monotonically decreasing effect on out-mobility the greater the distance between natives' residential locations and the location of out-group growth, and identify the limit of out-group influence as lying somewhere between three hundred and five hundred meters.

Finally, the analyses also indicate that the effect of non-Westerner growth in prompting out-mobility is also highly moderated by the ethnic composition of the proximate area in which this growth takes place. Thus, increases in non-Westerners do not appear to be capable of producing out-mobility when the level of the ethnic minority presence remains below 10%. Conversely, the models reveal a visible positive effect in areas where minority presence is equal to or above this value, and indicate that areas with a 15%-20% minority presence produce the greatest effects.

The simulation results also show three main findings. First, levels of segregation rise quickly as soon as the neighborhood size of the agents increases. Second, high segregation levels occur only when the ethnic share in the system is balanced (50%-50%). When there is imbalance, the system does not reach an equilibrium, producing low segregation levels as a result of the inability of the minority group to settle and stop moving. Most importantly, the distribution of differences in levels of segregation as a function of both neighborhood size and ethnic share in the system is abrupt and discontinuous. Finally, medium-to-high levels of segregation always

occur in neighborhoods whose boundaries are defined as enclosing agents' immediate surroundings, regardless of the total ethnic proportions of the system.

Many have argued for the importance of adopting a multi-scalar approach to understanding the role of neighborhoods, especially with regard to the issue of how segregation varies at different geographical scales (Lee et al., 2008; Reardon et al., 2009; Fowler et al., 2016). The results presented in this chapter would also suggest the use of this approach when measuring residential mobility, while they also call into question the usefulness of defining neighborhoods on the basis of administrative boundaries or census areas when examining the selective mobility of natives (Grodzins, 1957).

This chapter also has important implications for policy makers. By detailing how natives might react to increases in out-group mobility, the chapter provides a first assessment of how to strategically allocate the placement of accessible housing and temporary asylum centers to areas in which natives are less likely to move-out. The analyses in this chapter can therefore inform policy solutions aimed at improving the assimilation of minorities and at reducing segregation.

2.2 Literature review

2.2.1 Ethnic composition and native out-mobility

A common finding in most empirical research on ethnic residential segregation (ERS) is that residential mobility is influenced not only by the quality of one's current dwelling or of a prospective new dwelling, but also by the ethnic composition of the neighborhood in which the dwelling is located.

One prominent explanation attributes ethnic change in residential neighborhoods to selective movements by ethnic majorities (natives) in response to neighborhood ethnic composition, with this occurring via two mechanisms: white flight and white avoidance. *White flight* refers to a process whereby the ethnic majority moves out from a neighborhood due to an influx of ethnic minorities. *White*

avoidance refers to a process whereby the ethnic majority avoids moving into neighborhoods with a large minority presence. Although both mechanisms are important, this chapter focuses on the role of the neighborhood in prompting white flight.

Two hypotheses have been advanced to account for observed levels of white flight (Charles, 2003). The ethnic-preference hypothesis states that natives, like any other group, share a propensity for living near other co-ethnics (Crowder and Krysan, 2016). This propensity could result from either genuine prejudice (Pettigrew, 1998) and distaste (Bertrand and Duflo, 2017) towards others, or simply in-group affinity (Krysan and Farley, 2002). The racial-proxy hypothesis, on the other hand, states that white flight reflects neither natives' preferences for living near other co-ethnics nor their aversion to minority groups, but rather an avoidance of undesirable social class characteristics with which minority groups tend to be associated (Harris, 2001).

The evidence is less clear as to the precise motivations that drive white flight. Findings from vignette (Bader and Krysan, 2015) and experimental (Krysan et al., 2009) studies, as well as from surveys on natives' revealed preferences (Clark, 1991), have indicated a role for ethnocentrism. However, others have suggested that neighborhoods may enter a negative feedback-loop of degradation that reinforces associations between ethnic minorities and undesirable social class characteristics (Logan, 1978), which makes it difficult to disentangle the two. Quillian (2012) has found the process of ethnic segregation to be thoroughly entangled with within- and between-group separation on the basis of income levels (see also Quillian, 2014). Downey (2006) finds that areas with a higher African-American suffered from higher levels of air pollution. Legewie and Schaeffer (2016) find violence to be generally more predominant in situations where ethnic boundaries are less clear. Finally, other studies suggest that mistrust between neighbors increases with the neighborhood's ethnic heterogeneity (Dinesen and Sønderskov, 2015; Koopmans and Schaeffer, 2016), although the evidence is unclear (see also Abascal and Baldassarri, 2015).

2.2.2 Administrative neighborhoods and white flight

One essential premise of white flight dynamics is that individuals must somehow acquire accurate information about their neighborhood's ethnic composition in order for this to influence their mobility behavior. Accordingly, the belief that such information is leveraged at the level of neighborhoods has received widespread acceptance (Grodzins, 1957). Primarily as a result of data limitations, researchers have typically examined this by equating "neighborhoods" with administrative boundaries (or census areas) in order to capture the influence of ethnic composition on the propensity of natives to leave.

Mobility records in the U.S. combined with census-tract data constitute one notable source for evidence of white flight. These studies suggest that even though socioeconomic status is an important predictor of mobility, the likelihood of native out-mobility increases with the presence of African-Americans (South and Crowder, 1998; Crowder, 2000; Quillian, 2002; Crowder et al., 2006; Card et al., 2007). A multi-ethnic approach (Pais et al., 2009; Crowder et al., 2012) has yielded similar results. More recently, Hall and Crowder (2014) also find evidence of native flight (among whites and African-Americans) in cities that were not traditionally considered gateway cities for newly arrived immigrants, but which have become or are becoming cities of this kind. Finally, other studies have found no effect that substantively raises the likelihood of native out-mobility following a growth in the presence of African-Americans in adjacent neighborhoods (Steinnes, 1977; Schwab and Marsh, 1980; Crowder and South, 2008; Crowder et al., 2011).

Studies focused on the mobility of natives in Western Europe have yielded similar results. As in the U.S., Westerners tend to leave their (census) neighborhoods when the presence of non-Westerners increases (Bråmă, 2006; Schaaake et al., 2010; Hedman et al., 2011; Andersson, 2013; Aldén et al., 2015; Boschman and van Ham, 2015; Malmberg et al., 2018; Müller et al., 2018).

Despite the extensive use of census areas in studies of neighborhood change and other spatial processes (Entwisle, 2007), defining neighborhoods as census tracts

implies the use of rather large districts whose boundaries tend to encompass areas that are relatively homogeneous with regard to the needs addressed by each country's Statistical Bureau. However, it is also known that the needs addressed by census areas will not necessarily align with the definition of neighborhood that is most relevant to the purposes of a particular researcher (Galster, 2001).

An important argument against defining neighborhoods as census areas when studying inter-neighborhood mobility is the unrealistic cognitive assumptions that doing so imposes in relation to individuals. For instance, there is an assumption that individuals have complete information about the ethnic composition of the whole census area, no matter how large this area is (Crowder and Krysan, 2016).

Moreover, there is an assumption that all individuals care equally about the ethnic composition at every point within the census area, without distinguishing places that are nearer or farther relative to their own position (Logan, 2012). At the same time, there is also an assumption that individuals located to one another on either side of the boundary between neighboring areas will be oblivious of each other, or will think only in terms of their respective census areas rather than possessing their own "range" centered upon the specific location of their place of residence (Reardon and O'Sullivan, 2004). Finally, the use of census areas overlooks the role of urban architecture in spurring or obstructing social interactions across space (Grigoryeva and Ruef, 2015).

2.2.3 Neighborhoods size and uncertainty constraints

One current limitation of studies on inter-neighborhood mobility is the unknown potential bias that may result from choosing a particular definition of neighborhood. A natural solution to this issue would consist in observing how mobility patterns vary as the definition of neighborhood also varies. However, neighborhoods can vary on the basis of a wide range of features, such as their size, shape, demographic composition, or delineation in relation to contiguous neighborhoods. This chapter follows some advances made in the field of cognitive science with regard to decision-

making to argue that the significance of human cognitive limitations favors a focus on neighborhoods that are smaller than typical census areas in terms of the total space they encompass.

More concretely, cognitive science studies on decision-making have shown that individuals tend to determine their behavior by relying on “heuristics,” or “rules of thumb,” when information costs or uncertainty are high (Todd and Gigerenzer, 2003; Katsikopoulos, 2011; Mousavi and Gigerenzer, 2014). Under such conditions, heuristics exploit cognitive capabilities that limit the amount of information analyzed in making a decision, both easing and increasing the speed of the decision process (Gigerenzer and Gaissmaier, 2011). Others have argued that these cognitive shortcuts also operate on the basis of social cues (Hertwig and Herzog, 2009; Garcia-Retamero et al., 2010).

Given the large uncertainty involved in a moving decision, it might be expected that heuristics could play a role in limiting the amount of relevant information gathered by individuals in order to decide whether to stay or move. Indeed, uncertainty is a hallmark of many decisions that are typically dependent on social information (Hedström, 1998), residential mobility undoubtedly included. The generally large number of alternative dwellings that are in play, which typically makes investigating and pondering relevant consequences for each of them unfeasible (Bruch and Swait, 2019), clearly exemplifies a situation that is rife with uncertainty (Savage, 1954) and where we should thus expect to find heuristics being used (Crowder and Krysan, 2016).

Further, the high costs associated with acquiring complete information about the ethnic composition of even moderately large areas could further constrain individuals to rely on heuristics that favor social information about smaller local environments rather than larger ones (Bruch and Feinberg, 2017). Information about the ethnicity of neighbors living at close proximity is not only more salient and accessible to individuals on a daily basis, but, as has been shown by psychological studies, may even affect individuals’ experiences of what constitutes their neighborhood. These

studies have reported that individuals’ so-called self-defined neighborhoods tend to encompass markedly smaller areas than administrative neighborhoods (Coulton et al., 2013; Charreire et al., 2016), which has also been shown to lead individuals chiefly to overestimate the presence of out-group members in their census-based neighborhood (Clark, 2002; Semyonov et al., 2008; Wong et al., 2012).

Most importantly with regard to the study of mobility patterns, this implies that increases in the presence of out-group members at greater spatial distances from natives’ local environments should yield no noticeable effects on their probabilities of moving-out (Park, 1924). In line with this, studies have shown the importance of so-called “t-communities,” i.e. urban structures that are fully accessible at every point using only pedestrian streets (Grannis, 1998). These communities, that are typically delimited by non-pedestrian roads and that are smaller in size than census neighborhoods, “naturally” bound the areas in which interactions with other neighbors, including out-groups, are more likely to occur on a daily basis and which are thus more likely to affect residential mobility (Grannis, 2005).

2.3 Data and Methods

2.3.1 Data

I use register data for Stockholm County (1998-2017). This dataset, provided by Statistics Sweden, includes all natives living in the region with almost no missing data. It contains complete information regarding their residential mobility and location, with the latter being measured both at the so-called “SAMS”-level, the census boundaries defined by Statistics Sweden, and at the one-hundred-by-one-hundred-meter residential square-level (henceforth “residential square”). While the boundaries defined by the former measure create relatively large and homogeneous areas, the dimensions of the residential square typically encompass a large residential building and reflect more precisely the residents’ geographical location. This yields a total of 84,208 unique residential squares and 897 unique SAMS.

Dependent variable

The main outcome examined is natives' moving-out behavior, a binary variable measuring either a stay or a move between residential squares between two consecutive years (e.g. Crowder, 2000).

Independent variables

The registers also record important factors that are typically held to influence residential mobility. These factors are life-course related, such as the natives' age, civil status, and family type (Clark and Huang, 2003; Clark and Withers, 2007); socioeconomic factors, including disposable income (logged), years of education, type of tenure tenancy and ownership, and the area's median disposable income (e.g. Crowder et al., 2006); as well as ethnicity-related factors, mainly the country of birth of natives as either Sweden or E.U.-15/U.S./Canada (Jarvis et al., 2017), and the proportion of the area's population that is comprised of minorities, defined as those whose country of birth (or whose parents' country of birth) lies outside Sweden/E.U.-15/U.S./Canada. Finally, I include the number of individuals living in the area (which has also been shown to affect out-group visibility) (Valdez, 2014).

2.3.2 Analytical strategy

The high level of granularity available in the census registers allows me to use the variation in natives' exposure to a growing minority presence to measure the impact of minority presence on increasing the likelihood of natives to out-mobility. More concretely, I compare natives exposed to a proportional increase in minorities with other natives who are exposed to no change in the ethnic composition of their residential area. I view this "treatment," although it is not perfect, as capturing the basic level of out-group exposure that is necessary to produce white flight, compared to the "control" case in which this cause is absent (all else remaining constant) (Woodward, 2003).

Most importantly, I repeat this analysis separately for four distinct definitions

of neighborhoods in terms of the total area encompassed: (a) one-hundred-by-one-hundred-meter squares; (b) three-hundred-by-three-hundred-meter squares; (c) five-hundred-by-five-hundred-meter squares; (d) census areas (SAMS). These configurations permit a comparison of the effect of defining neighborhoods as being less extensive than census areas while census areas are used as reference.

Table 2.1 shows the mean value of each covariate and the number of observations in each group of the full sample. One notable finding from this analytical strategy is that groups can look very different across certain observed covariates. In this scenario, Ho et al. (2007) have argued that regression adjustment alone may be a less efficient strategy for estimating differences between groups, mainly as a result of the method’s increased dependence on the model’s specific functional form and the larger standard errors that result when the groups examined are too dissimilar to one another (Rosenbaum and Rubin, 1983). I address these undesired drawbacks by adjusting for the covariates through matching (Stuart, 2010). The matching method I use is Coarsened Exact Matching (CEM) (Iacus et al., 2012), a non-parametric technique argued to be more reliable than other matching methods in improving in-sample similarity (King and Nelson, 2018) thanks to its resemblance to a blocked-experimental design (Deaton and Cartwright, 2018). As Table 2.1 also indicates, CEM achieves remarkable improvement in similarity across groups (see Appendix for a much broader account of the performance of CEM).

To ensure that covariates follow a strictly causal order and that their measurement always precedes the “allocation” of natives into groups (Pearl, 2010), they have been matched dynamically (Aral et al., 2009). I divide the time series into overlapping trials of three consecutive years. For each trial, the first year determines the measurement of the covariates used in the matching, the second year the binary group into which natives fall, and the third their mobility outcomes. This approach ensures that natives are only matched to other similar natives within each trial, and also conveniently rematches them when any of their covariates changes over time. Once matching is complete for each trial, I gauge the effect of exposure

Covariate	Full sample		Matched sample	
	No change	Pr. increase	No change	Pr. increase
<i>Age</i>	50.42	51.05	53.24	53.12
<i>Years of education</i>	12.14	11.87	11.65	11.70
<i>Disposable income (log)</i>	7.43	7.35	7.37	7.38
<i>Single</i>	0.26	0.41	0.34	0.34
<i>Married</i>	0.61	0.38	0.45	0.45
<i>Divorced</i>	0.13	0.22	0.22	0.21
<i>With children</i>	0.55	0.37	0.39	0.39
<i>No children</i>	0.27	0.20	0.23	0.23
<i>Sweden</i>	0.84	0.80	0.87	0.87
<i>EU-15/US/Canada</i>	0.16	0.20	0.13	0.13
<i>Renter</i>	0.03	0.41	0.29	0.29
<i>Owner</i>	0.04	0.38	0.31	0.31
<i>Neigh. pr. non-Westerners</i>	0.03	0.18	0.12	0.13
<i>Neigh. median disp. income (log)</i>	7.44	7.37	7.39	7.40
<i>Neigh. N inhabitants</i>	17.16	194	94.16	112.68
<i>N</i>	4,121,178	7,990,769	3,872,889	3,722,574

Table 2.1: Basic descriptives for natives exposed to a proportional increase in minority presence and natives exposed to no change in their neighborhood ethnic composition in Stockholm County (1998-2017). Rows show the mean value of each covariate, prior to (Full sample) and after matching. Numerical covariates are presented in terms of their own scale (or logged), and qualitative covariates are shown as proportions. The total number of cases in each group is smaller in the matched sample due to pruning. The values shown may of course change slightly depending on the definition of neighborhood employed. This table only displays the values for neighborhoods defined as one-hundred-by-one-hundred-meter squares due to space constraints, but a complete overview is available in the Appendix.

using a weighted linear probability model (LPM) estimated for the entire sample using the ordinary-least squares method. The model employed consists only of the outcome, the variable determining natives' group, and the matching weights. This allows me not only to compare differing outcomes between groups without including the scaling factor in logistic regression (Mood, 2010), but also to compare estimates across models employing different neighborhood definitions. I follow Clogg et al. (1995) in calculating significance tests for differences between the estimates from different models.

For the simulation analysis, I depart from the classic checkerboard model of Schelling (1971) to study the levels of segregation that are generated as the neighborhood definition changes. The sizes tested range from agents' most immediate surroundings to areas that are as big as census areas (see Appendix). These neigh-

neighborhoods overlap with the neighborhoods of others, reflecting self-defined neighborhoods. My methodology here is an agent-based model (ABM) (Macy and Willer, 2002), an alternative modeling approach that is gaining popularity in sociology (Bianchi and Squazzoni, 2015) and that allows for the study of counterfactual scenarios in which artificial agents interact following certain programmed behavioral rules, giving rise to complex patterns exclusively from the “bottom up” (Ylikoski, 2014). The analyses presented follow standard sensitivity analysis (Bruch and Atwell, 2015) on the basis of neighborhood size and the ethnic share of the minority group in the system, since previous analyses addressing the role of neighborhood size have not tested this configuration (see Laurie and Jaggi, 2003; Fossett and Waren, 2005).

2.4 Results

2.4.1 White flight

Figure 2.1 shows the empirical probabilities of moving-out across different neighborhood definitions. Subplot A displays the likelihood of residential mobility for each group and neighborhood definition, whereas subplots B-D display the difference between the predicted probabilities, 95% confidence intervals, and p-values (in parentheses) from the LPM for the matched sample. Upper horizontal bars show whether differences between the estimates from different models are statistically significant.

Starting with neighborhoods defined as census areas (SAMS), the model in subplots A-B clearly shows no increase in the probabilities of moving-out between having been exposed to an increased minority presence and no change. In fact, the model seems to outline a negative effect, clearly indicating no white flight whatsoever, in line with previous findings using similar data and equating SAMS with neighborhoods (e.g. Bråmă, 2006; Müller et al., 2018).

However, we obtain different results when we change the definition of neighborhoods to cover areas much smaller than census areas. All these models clearly

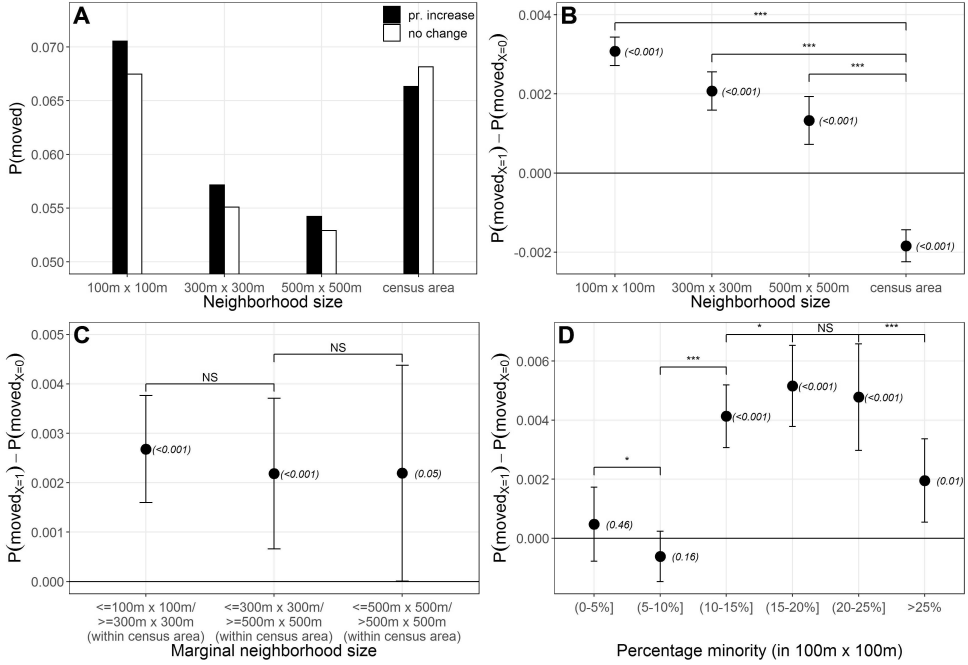


Figure 2.1: The importance of spatial proximity for white flight. **(A)** Moving-out probabilities after matching: (1) natives exposed to a proportional increase in minorities ($X = 1$, black), and (2) natives exposed to no change in ethnic composition ($X = 0$, white) (products indicate neighborhood dimensions). **(B)** Difference between predicted probabilities of moving-out among groups in A across neighborhood definitions. **(C)** Marginal differences between natives exposed to a proportional minority increase only in smaller area ($X = 1$, upper label) or outside a bigger area ($X = 0$, lower label). **(D)** Effect heterogeneity of groups in A with different initial minority percentages across one-hundred-by-one-hundred-meter squares. Predicted probabilities, differences between groups, confidence intervals, and p-values in parentheses are gauged using a weighted linear probability model estimated for the matched sample. The straight horizontal line indicates no effect. Upper bars indicate whether differences between models are statistically significant. ***: $p\text{-value} < 0.001$, **: $p\text{-value} < 0.01$, *: $p\text{-value} < .05$, 'NS' $p\text{-value} > .05$.

exhibit a positive and statistically significant increase in the probability of natives moving-out following an increase in out-group presence. In particular, the effect seems higher in those models in which neighborhoods are defined as the spaces most proximate to the immediate surroundings of the natives' dwellings, as exemplified in the form of two informative clues. First, the models show a monotonically decreasing effect the larger the dimensions encompassed by the neighborhood definition ($\beta \approx 0.003, 0.002, 0.001$, respectively for one-hundred-meter, three-hundred-meter and five-hundred-meter squares; these estimates are highly statistically significant).

Second, the largest estimates are found for the models focused on the smallest areas analyzed in this study, i.e. the one-hundred and three-hundred-meter squares. Moreover, these models also produce the highest t-values, indicating that these neighborhood definitions encode the largest differences between groups (t-value \approx 16.72, 8.38, 4.29, respectively for the models based on one-hundred-meter, three-hundred-meter and five-hundred-meter squares).

Subplot C in Figure 2.1 shows a tendency similar to that seen in the previous subplots, but at a slightly different angle. The plot reveals marginal increases in the probabilities of white flight with each increment in neighborhood size for the sample of natives who are only exposed to out-group growth. The groups compared indicate whether minority growth occurred only in the smaller neighborhood (upper label in each category in the plot), or only on the “outside” of the bigger neighborhood (lower label in each category). The census area serves as the upper bound for all models, so the maximum distance at which growth may occur is within the diameter of a census area. As can be seen, all models show a positive increase in moving-out following exposure to an increase in out-group presence in the most proximate area. This effect is statistically significant, however, only for the models in which neighborhoods are defined as one-hundred-meter and three-hundred-meter squares. Conversely, the model in which neighborhoods are defined as five-hundred-meter squares yields differences that are not statistically significant in relation to those of census areas. In addition, the distance between estimates across models seems not to be statistically significant, which primarily means that it is inconclusive whether the models based on smaller neighborhood sizes are qualitatively different from one another.

Results so far seem to indicate that the models employing census areas as the units for neighborhood size do not properly capture white flight dynamics, whereas definitions of neighborhoods below the level of five-hundred-meter squares do. However, both the absolute and the relative out-mobility estimates seem rather low, indicating a much higher propensity to stay than to move. This is likely to be due

to various factors. First, the models reflect the estimated effect one year subsequent to out-group growth, a time-window that may be too restrictive to yield large effects. Second, as has been noted by other scholars (Bruch and Swait, 2019), the higher tendency to stay may also reflect the importance of “life-course events” in triggering out-mobility (Clark and Huang, 2003; Clark and Withers, 2007).

A further possible reason is that a growth in minority presence might exert a varying influence depending on the neighborhood ethnic composition prior to this growth. To outline the effect of heterogeneity across neighborhoods with different levels of minority presence, subplot D in Figure 2.1 again shows the estimated difference in out-mobility between the conditions of exposure to out-group growth and no change. Unlike the other plots, this one computes the effect separately for very small neighborhoods (defined as one-hundred-meter squares) whose prior proportion of non-Westerners varies in increments of 5%. The models clearly show no white flight in native-based areas. In contrast, the white flight effect appears as soon as the percentage of minorities surpasses 10%, and reaches a peak for neighborhoods where the minority presence lies at 15%-20%. In fact, in areas where the presence of non-Westerners is higher than 10%, the effect size seems almost to be double that of the estimate shown in subplot B. For small neighborhoods with a larger minority presence, the size of the estimate seems to diminish, partly due to the low number of cases of natives living in these neighborhoods that is observed in the data. Most importantly, the distance between the model estimates showing the jump from no effect to a positive effect appears to be both qualitatively and statistically significant. These results seem to be in line with previous research suggesting abrupt transitions in the probabilities of out-mobility as minority presence increases (Grodzins, 1957; Schelling, 1971; Card et al., 2007; Aldén et al., 2015), but they diverge from this research in locating the presence of these dynamics mainly in neighborhoods defined as being markedly smaller than census areas, which was the definition typically employed by these other authors.

Overall, the models indicate several results. Most importantly, defining neigh-

borhoods as census areas seems not to yield estimates that are in line with expectations derived from the white flight hypothesis. Rather, the results seem to indicate that neighborhood sizes that are smaller than census areas are the only ones capable of producing qualitatively and statistically significant positive white flight estimates. In addition, the models seem to indicate higher tendencies toward out-mobility when exposure to an increasing minority presence occurs closer to the residential location of natives, although the effect sizes do not seem to be statistically significantly different between the models in which neighborhoods are defined as one-hundred-meter or three-hundred-meter squares. Finally, the models also show important effect heterogeneity across small neighborhoods with differing initial levels of ethnic minority presence. Thus, neighborhoods with a minority presence below 10% do not seem to trigger white flight dynamics, while higher levels of minority presence seem to increase the propensity to move-out in a rather abrupt way.

2.4.2 Segregation

Figure 2.2 presents results from the ABM simulations analyzing how segregation, measured using Theil's H Index (Reardon and Firebaugh, 2002), varies as a function of two dimensions: neighborhood size (here defined as circles with different radii within which each agent self-defines its own exposure area, which are equal for everyone), and the total ethnic share of the two groups in the system. Of course, the larger the radius, the greater the neighborhood size. Moreover, I show results that display all possible combinations of the two dimensions (Bruch and Atwell, 2015).

Subplot A of Figure 2.2 shows the segregation levels generated given different neighborhood sizes (indicated by their radii). The simulations show that when a neighborhood's ethnic makeup is equally divided, segregation increases with radii up to a level of 10. These results accord with previous findings (Laurie and Jaggi, 2003; Fossett and Waren, 2005). However, when the ethnic ratio is unbalanced, segregation levels decrease markedly once the radius has reached a certain critical

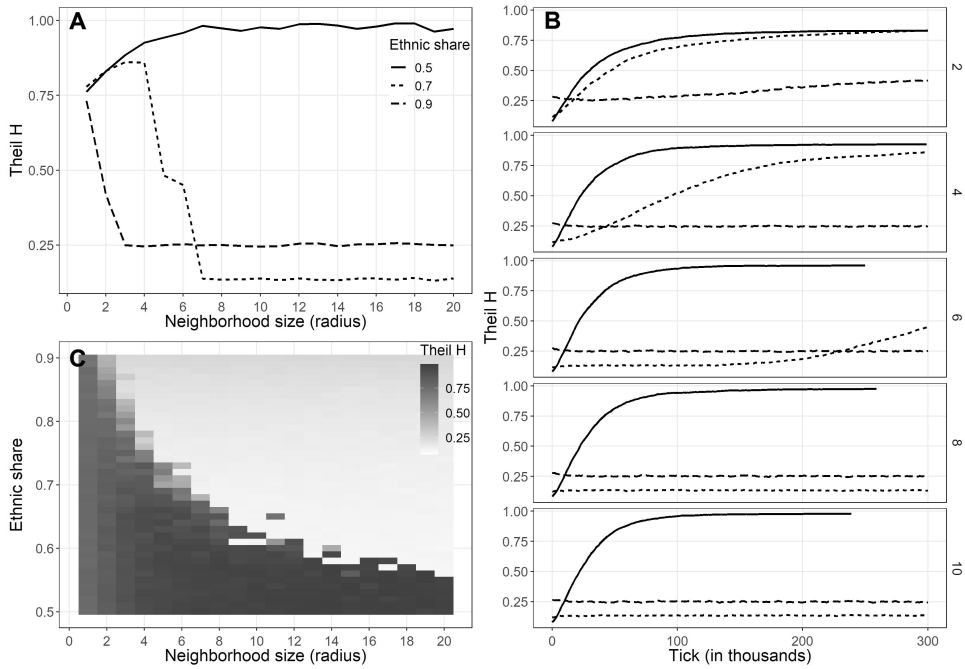


Figure 2.2: The importance of neighborhood size for ethnic residential segregation. **(A)** Segregation levels generated as a function of neighborhood size. Segregation lines are plotted for three configurations of ethnic minority share: (1) 50%-50% (solid line); (2) 70%-30% (short-dashed line); and (3) 90%-10% (long-dashed line). **(B)** Segregation levels as a function of time (“ticks,” in thousands) for the same three share configurations as in A. Numbers displayed on the right in each subplot show the neighborhoods’ radii. **(C)** Heatmap showing different levels of segregation (the greater the level of segregation, the darker the tile) generated as a function of both neighborhood radius (x-axis) and the proportion of natives in the system (y-axis). Segregation is measured using Theil’s H Index.

point. The only difference across configurations is found precisely where this critical point occurs: when the ethnic distribution is 70/30, segregation starts to decline once neighborhoods have a radius of 4, reaching a minimum beyond a radius of 7. Likewise, when the ethnic distribution is 90/10, segregation levels start to diminish rapidly from the start, quickly reaching a low point beyond a radius of 3. Conversely, high levels of segregation tend to occur in simulations involving neighborhoods of small radii regardless of the ethnic share.

Subplot B shows the same tendency described above as a function of time, with each subplot showing the case for a specific radius. In the balanced scenario, the plot always shows high segregation levels, which rise as the radius of the neighborhoods

increases. In fact, segregation increases more steeply the greater the radius, reaching higher levels sooner. For more unbalanced group distributions, however, it takes considerably longer to reach higher levels of segregation, and these increases are only found up to a certain radius, beyond which segregation remains low.

Finally, subplot C shows levels of segregation as a continuous function of neighborhood radius (x-axis) and ethnic share (y-axis). Darker areas indicate higher levels of segregation, lighter areas the opposite. The plot seems to outline a concave-up, decreasing curve separating two radically distinct segregation outcomes. Below the line demarcating the boundary, the shaded area marks very high levels of segregation, whereas above this line segregation decreases substantially. The plot shows clearly how extremely high levels of segregation only exist given the following two conditions: (1) neighborhood size must be large (even as large as census areas) and (2) ethnic proportions must remain approximately balanced. Beyond certain points along these two dimensions, however, the simulations consistently show lower levels of segregation.

The simulations produce these results as a result of the way in which the behavioral rule interacts with neighborhood radius. Since the model departs from Schelling (1971), agents are initially randomly scattered across space, and only move when the share of co-ethnics in their neighborhoods falls below 50%. Because larger neighborhoods cover larger fractions of the available area, they also include far more other agents. This, in turn, increases the likelihood that agents will grow dissatisfied and keep moving. Under the scenario with balanced groups, however, groups tend to “satisfy” their behavioral rule in the long run and find suitable spots to stay, producing large levels of segregation as a result, although the larger the area, the longer it takes to arrive at this point. Nevertheless, as soon as an imbalance appears in the ethnic share, only the members of the predominant group are able to satisfy their behavioral rule and stop moving, and this will happen more quickly as the probability of encountering members of the other group diminishes. The remaining places that are unoccupied by the majority group, which are randomly scattered

across the total space, are insufficient for minorities to satisfy their behavioral rule, and this is exacerbated as neighborhood radius increases. This effect, combined with a lower level of ethnic diversity in the system, produces artificially low levels of segregation. Interestingly, however, the simulations show that the transition from high to low levels of segregation across the parameter space is discontinuous rather than smooth.

2.5 Conclusion

Even mild levels of intolerance can lead to high levels of segregation. Thomas Schelling's (1971) seminal work highlighted the interdependence of agents in influencing one another. Physical space constrained the range of this influence, as agents could only influence those located nearby. Hence, early movers could affect later movers via sequences of changes in consecutive local environments. Many others have since followed a largely similar strategy while equating neighborhoods, the spaces in which agents can influence each other, exclusively with census areas. This study has sought to move beyond the limitations that are typically associated with the data collection strategy employed in these studies, and show by means of empirical analysis and simulation how changing our definition of neighborhood can change our expectations about the inter-neighborhood mobility of natives and the consequences of this mobility for producing segregation.

For the empirical analyses, I employed matching based on register data for Stockholm County (1998-2017), a longitudinal dataset with a high degree of granularity, which allowed me to exploit variations in natives' exposure to increases in the level of minority presence to observe the differences that followed from defining neighborhoods of different sizes. For the simulation, I turned to an agent-based model to map in detail the implications of neighborhood size for the generation of different levels of segregation.

The empirical results have shown that defining neighborhoods as census areas does not yield estimates in accordance with the expectations of the white flight

hypothesis. However, neighborhoods whose dimensions encompassed areas smaller than census areas yielded positive white flight. Moreover, the models depicted a rapid decay in white flight as soon as the neighborhood includes areas that lie farther than one hundred meters from natives' residential locations, although neighborhoods defined as including distances of up to three hundred meters also seem to yield positive estimates. Finally, no effect appears to emerge following exposure to an increase in ethnic-minority presence beyond three hundred meters, which marks an important limit beyond which individuals are instead influenced by others located at a closer proximity to themselves. In conclusion, the evidence presented would call into question the accuracy and reliability of white flight dynamics that are observed in neighborhoods defined as census areas, since such definitions tend to produce areas that are too large, with distances that do not allow individuals to influence one another. The closer individuals live to one another within a distance of three-hundred meters, the greater the influence they may exert on one another (Park, 1924).

At the same time, the analyses show important effect heterogeneity for white flight across smaller neighborhoods with a varying ethnic composition. Thus, white flight occurs mainly when the proportion of the neighborhood population comprised of ethnic minorities exceeds approximately 10%, peaking in areas where this proportion reaches 15%-20%.

Simulation analysis also shows important differences in the levels of segregation that are generated. When agents are assessing the ethnic composition of larger neighborhoods, segregation is greater by comparison with situations in which the assessment is made in relation to smaller neighborhoods. However, these results hold only when there is balance between the ethnic shares of the two groups in the system. Otherwise, members of the predominant group tend to spread through larger neighborhoods, pushing minorities to keep moving due to scarce, sparse vacancies that are located randomly, artificially lowering segregation levels as a result. Moreover, this transition from higher to lower levels of segregation as a function of

both neighborhood size and ethnic share is abrupt and discontinuous. In contrast, in smaller neighborhoods, medium-to-high segregation levels tend to appear regardless of the relative size of the different ethnic groups. According to the assumptions of the model, we should always expect medium-to-high segregation levels as minorities move into a metropolitan area, as individuals will tend to find dwellings near other co-ethnics within small areas.

This chapter also has important limitations. Since I still measure changes in local ethnic composition on the basis of some form of “common neighborhood,” it is likely that varying locations within small residential squares also affect individuals’ exposure to minorities, thus the analyses will still be affected by the MAUP. For instance, using residential squares hampers distinguishing between the influence of “same-building-exposure” and “next-building-exposure,” which might involve important differences in how individuals get to know their neighbors. Similarly, while this study, in order to reduce overall complexity, did not take other elements in the urban environment into account, such as walls, parks, or the similarity of building dispositions, these are likely to have some form of notable impact (Grigoryeva and Ruef, 2015).

Another important limitation of this chapter involves the way its scope is restricted only to addressing the behavior of natives. As other studies have shown, other groups may change their mobility logic and thus produced different effects on segregation to those observed here, since the simulation analyses presented above have assumed that they would behave in the same way as natives (Clark, 1991).

Further, studying interactions between groups using observational data entails limitations of its own. For instance, one assumption that is implicit in the models, and thus untested, is that individuals do not contact each other within/between exposure groups. However, we know that social networks may lead them to interact and so modify the effect of treatment itself, violating the so-called stable unit treatment value assignment assumption (Sobel, 2006). Finally, the current state of the segregation literature makes it difficult to verify with certainty the directions of

the causal paths that constitute underlying assumptions for this study (Elwert and Winship, 2014).

Future research could advance on this study by addressing how “small neighborhoods” interact with “official” census neighborhoods (Sampson, 2008). For instance, one might use the insights from this chapter to study how ethnic change unfolds in real administrative neighborhoods, assuming that individuals focus on smaller neighborhoods in assessing their relevant ethnic composition. Real-world coordinate data could be used to address this while at the same time overcoming the limitations of the data employed in this study, which could help to disentangle the effects of exposure more accurately, for example by incorporating other elements in the urban environment.

Chapter 3

If you move, I move: The Social Influence Effect on Residential Mobility

Abstract

¹ There are many theories that account for why households move between residential areas. In this second empirical chapter, we advance on this by formulating a new mechanism whereby a household's probability of leaving a neighborhood is informed by the number of other households who have previously left that neighborhood. We call this mechanism: the social influence (SI) effect. By applying matching to Swedish register data for Stockholm County (1998-2017), and after adjusting for theoretically relevant confounders from the existing literature, we find that SI has a significant effect on neighborhood out-mobility. Furthermore, we find that the SI effect is moderated by the visibility with which others' behaviors is observed, measured as the number of previous out-movers, the distance to ego, and its salience in the social environment. Our study also discusses some ways in which SI might be entangled with other mechanisms, and outlines future directions from which studies of residential segregation dynamics might be approached.

¹This chapter is co-authored with Eduardo Tapia.

3.1 Introduction

Either ethnically or socioeconomically colored, segregation scholars suggest that the spatial separation of groups has negative consequences for the welfare of segregated families, as it fosters an unequal distribution of opportunities across residential areas. For example, households that are overrepresented in economically disadvantaged neighborhoods (Logan, 2011) are at increased risk of ending up with lower earnings and of suffering from industrial hazards (Sampson et al., 2002; Ludwig et al., 2008; Crowder et al., 2011; Quillian, 2012; Thomas and Moye, 2015).

Social scientists agree that households' selective mobility patterns—leaving and moving into different neighborhoods—constitute the main motor of residential segregation. Previous studies have reported these mobility patterns to be determined by distinct factors, including households' life-course events, housing market characteristics, and households' neighborhood preferences (Clark and Dieleman, 1996; Green, 1997; South and Crowder, 1998; Clark and Huang, 2003; Crowder and Krysan, 2016). However, previous research has largely overlooked the possibility that individuals' mobility decisions are affected by others' past mobility behavior: the social influence (SI) effect.

Studying how individuals' behavior is influenced by the past behavior of their peers has been a cross-sectional topic of inquiry in the social science literature. It has allowed researchers to better understand individual choices beyond personal preferences and socioeconomic characteristics. In addition, SI-related studies provide a theoretical framework for the sociological understanding of how some macro events are linked to micro behaviors.

The generalizability of this framework has permitted its application to a wide range of sociological phenomena (Granovetter, 1978; Cialdini and Goldstein, 2004; Flache et al., 2017). For instance, Hedström (1994) has observed how people previously having joined unions increases the chances of others joining, allowing trade unions to spread spatially. In the context of cultural markets, Salganik et al. (2006) show that prior information on the popularity of songs, compared to a blinded sit-

uation, increases the unpredictability and inequality of their market share (see also Muchnik et al., 2013; van de Rijt et al., 2014). Despite the substantial amount of empirical evidence showing the effect of SI in relation to a wide variety of issues, to our knowledge, no attempt has been made to investigate whether SI plays a role in neighborhood choice, and, consequently, in macro-patterns of segregation.

Building on social theories of influential behavior as a belief-formation process (Hedström, 1998) and self-attention theory (Mullen, 1987), we hypothesize that, after adjusting for factors that have previously been identified as significant in the literature on neighborhood mobility, SI will play a role in whether households leave or stay in their current neighborhoods.² More concretely, our theoretical argument states that when a household leaves a neighborhood, it sends a negative signal (leaving the neighborhood) to other neighbors about the neighborhood that is being left. As a result, this negative signal contributes to a negative change in neighbors' beliefs about the desirability of the current neighborhood, which in turn increases their own likelihood of moving-out.

As in previous studies of SI (Salganik et al., 2006), we also hypothesize that the strength of this effect will be moderated by the ease with which the signal is perceived by the receivers in the following respects: (1) the number of households leaving the neighborhood; (2) the residential density of the area; (3) the physical distance between the signal and the receiver, and (4) the proportion of non-natives living in a neighborhood.

This study examines how the earlier residential mobility behavior of households influences the moving-out behavior of current residents, adjusting for previous explanations found in the literature on neighborhood mobility. Our analysis focuses on register data for Stockholm County. Apart from providing long time-series in relation to immigration processes, the granularity of Swedish registers provides a unique opportunity to study the effect of SI, since it allows us to control for other important factors relating to neighborhood mobility. However, despite the richness of the

²We are aware that SI can also influence in-mobility. However, studying this would require a different approach to that taken here, and also a different dataset.

data, the lack of a significant number of non-native cases for all of our confounder variables has forced us to restrict our focus to studying the native population.

By applying Coarsened Exact Matching (CEM), we exploit variation in natives' exposure only to out-movers or to an absence of movers in order to assess the effect of SI on out-mobility behavior, and also to evaluate how the visibility of influential signals tempers, or exacerbates, the SI effect. Our results show that SI plays a key role in relation to residential out-mobility. Specifically, we find that households are more likely to leave neighborhoods when other households have previously left, compared to when there has been no such change. In addition, we also find that this effect is stronger when previous out-movers leave lightly populated areas, are physically closer, and greater in number. We found no evidence for an effect of neighborhood ethnic composition.

This study contributes to the literature on residential segregation in two ways. First, the study of social influence allows us to evaluate the effect of one of the most intriguing and relevant topics in social science research in a novel context where it has not previously been tested. Second, we are not only proposing a new causal mechanism that drives neighborhood mobility and, consequently, residential segregation patterns, but also advancing the theory of social influence itself by providing new empirical evidence on its occurrence and intensity.

The structure of this chapter is as follows. First, we introduce current theories that account for residential mobility, and this is followed by an introduction on SI theories. We then introduce our analytical design and explain how we apply CEM to Swedish register data. Finally, we present and discuss our results. We conclude this chapter with a discussion.

3.2 Literature review

3.2.1 Neighborhood mobility

Residential segregation is an important issue in modern societies. Scholars have repeatedly documented its detrimental effects in a wide variety of contexts (Sampson et al., 2002; Charles, 2003; Ludwig et al., 2008; Crowder et al., 2011; Logan, 2011; Sharkey and Faber, 2014; Thomas and Moye, 2015; Chetty et al., 2016).

Previous studies have noted that patterns of segregation are driven by the ways in which households move around cities. Depending on the scientific field, several explanations have been provided to account for why and how households move between neighborhoods. For instance, demographic explanations focus on households' dwelling requirements and the shifting role of these requirements in the life-trajectories of families. These explanations argue that life-course events play an important role for mobility because these events shift and unbalance households' space-related housing needs, motivating the search for new dwellings (Clark and Dieleman, 1996). Thus, changes in family composition, such as getting married or having children, have been found to be important predictors of mobility (Clark and Onaka, 1985; van Ham and Clark, 2009). Similarly, age, socioeconomic status, and tenure position have also been found to have important effects on neighborhood mobility. Hence, younger, renters, and affluent households are more likely to move out than their counterparts (Green, 1997; Clark and Huang, 2003).

On the other hand, sociological explanations instead direct their focus at the interaction between neighborhoods' environments and households' socioeconomic-ethnic attributes as a trigger of residential mobility. A significant part of this literature points to the ethnic composition and socioeconomic status of neighborhoods as the main drivers of neighborhood mobility. These explanations state that households hold ideal preferences, in terms of neighborhoods' ethnic composition and socioeconomic status, with regard to where they want to reside. Mobility is thus a response to the desire to fulfill these preferences (Grodzins, 1957; Schelling, 1971; South and

Crowder, 1998; Charles, 2003; Quillian, 2012; Crowder and Krysan, 2016).

Within this preference-based explanatory approach, natives typically prefer living in native-dominated neighborhoods whereas non-natives search for more racially mixed areas (Clark, 1991). As a consequence, residential segregation increases as natives leave and avoid neighborhoods that experience an increase in the share of non-natives (e.g., Quillian, 1999; Crowder, 2000; Crowder et al., 2006; Card et al., 2007; Pais et al., 2009; Schaake et al., 2010; Crowder et al., 2011; Hedman et al., 2011; Hall and Crowder, 2014; Aldén et al., 2015; Boschman and van Ham, 2015).

In addition to the processes outlined above, scholars have also documented the effect of neighborhoods' socioeconomic status on households' mobility. For example, it has been found that improvements in the socioeconomic status of a neighborhood foster the out-migration of low-affluence families (Atkinson, 2000) and, by contrast, a decline would promote the flight of more affluent families (South and Crowder, 1998; Quillian, 1999). At the same time, a decline in the quality of buildings or schools has also been documented to produce more high-status households to move-out (Quillian, 2012). Either way, it is high-status households who have greater opportunities to react to changes in their social environments.

While all previous explanations offer different micro-mechanisms to account for moving-out behavior, it remains unclear whether the prior out-mobility³ of residents can increase the likelihood of out-mobility among former neighbors after adjusting for all of the existing mobility explanations described above. We call this the SI effect. The aim of this study is thus to propose the SI effect as an alternative and complementary explanation for residential mobility.

³There are other explanations related to inter-neighborhood mobility which have also noted out in previous studies, but which nevertheless are not directly related to the scope of this study. For example, scholars have also shown the presence of discriminatory practices in the housing market (e.g., Massey and Denton, 1993; Ahmed and Hammarstedt, 2008), which bias the in-flow of non-native families into certain neighborhoods.

3.2.2 The Social Influence effect

The long-standing relevance of SI in the sociological literature has been due to the way it complements existing models of individual decision-making in a wide range of contexts and, most importantly, as a result of its suitability in disentangling how the effects of such decisions lead to macro-social patterns of interest. For example, Granovetter's threshold model (1978) states that the propensity of individuals to participate in a riot is a function of the number of persons already involved. The implication of this premise is that macro results are highly dependent on how micro actions unfold over time as a consequence of previous actions, bringing about different aggregate results as a function of the sequence in which things happen (Anderson, 1972; Salganik and Watts, 2009).

In its most general conceptualization, SI refers to the effect of previous individual actions on the future behavior of other individuals who are aware of those actions. Essentially, we claim to observe SI whenever an individual modifies her expected behavior after observing the behavior of others; for example, when the likelihood of X doing A instead of B increases as a result of X having perceived others doing A. This example reveals three elemental properties of SI. (1) At the individual level, people usually follow certain motivations in order to adjust their behavior to the behavior of others. (2) In any SI pattern, there is always a specific action that is first perceived and then imitated. (3) SI takes place as long as others' behavior is perceived.

Empirical patterns of the type described in the above example can be driven by several mechanisms at the individual level. Distinct psychological mechanisms have been argued to account for why individuals follow others' behaviors, such as conformity (Asch, 1955); compliance (Cialdini and Goldstein, 2004); social learning (Akers et al., 1979); rational imitation (Hedström, 1998); or legitimacy (DiMaggio and Powell, 1983), to name a few examples. In this chapter, we conceive of SI as a belief-formation process (Hedström, 1998). Concretely, we argue that an individual who leaves a neighborhood sends a negative signal to other neighbors about the

desirability of the neighborhood, which in turn undermines the current favorable belief of neighbors regarding the area in question, thus increasing their likelihood of moving-out.

Previous studies have found similar patterns in other domains. For example, in the context of labor market mobility, Felps et al. (2009) find that the probability of a worker quitting her job is higher when other co-workers have previously quit. Likewise, we hypothesize that the probability of a household leaving a neighborhood is greater when (an)other household(s) has/have previously left, as compared to a situation in which nobody has left.

While the previous hypothesis is related to the existence of SI in relation to neighborhood mobility, the following section elaborates on how the strength of SI can vary depending on how likely it is that the receivers receive the signal. More concretely, in this study we identify four factors that moderate the ease with which these signals are received: (1) the number of sources; (2) the residential density of the area; (3) the physical distance to the sources; and (4) the ethnic composition of the area.

The first factor fuels the following hypothesis: the larger the number of previous movers, the greater the probability that others will leave the neighborhood. The rationality of this hypothesis is not only based on the fact that a greater number of movers are more visible than a smaller number, it also captures the idea that a greater number of movers may exert a greater reinforcing impact on the negative signal, as has been shown in several studies on the adoption of innovations (e.g., Granovetter, 1978; Salganik et al., 2006; Salganik and Watts, 2008; Centola and Macy, 2007; Aral et al., 2009; Centola, 2010; Muchnik et al., 2013; van de Rijt et al., 2014).

The second factor, residential density, also builds on the visibility of out-movers. Thus, we hypothesize that in poorly populated areas the salience of out-movers' behavior is greater compared to areas of high population density.

The third factor, on the other hand, holds that the greater the distance between

the origin of the signal (households leaving the area) and the receiver, the lower the probability that the receivers will leave the area. This factor captures the degree of interaction among individuals distributed in space, and how they influence each other (e.g., Reardon and O’Sullivan, 2004; Logan, 2012). In concrete terms, individuals exert more influence on others the closer they are to them in space (Park, 1924; Gould, 1991; Hedström, 1994).

Finally, Self-Attention Theory (Mullen, 1987) serves as the basis for the final factor affecting the strength of SI. This theory postulates that in-group attention varies as a function of group composition. More concretely, the lower the proportion of in-group members in the group, the higher the salience of in-group members, and, consequently, the greater the level of attention paid to them (Mullen, 1991). Transferring this to the neighborhood mobility context: natives moving-out from a native neighborhood are less salient to other natives than natives leaving highly segregated neighborhoods. Accordingly, we formulate the following hypothesis: the greater the proportion of non-natives in the neighborhood, the higher the effect of SI among natives.

3.3 Data and Methods

Our analysis is based on longitudinal Swedish register data for Stockholm County (1998-2017). Besides offering precise individual census information on demographic and socio-economic aspects, this unique dataset tracks the residential mobility of all residents with almost no missing data. The spatial residential location of households is measured at the $100\text{m} \times 100\text{m}$ square-level (henceforth ‘residential area’). Our main outcome, moving-out, is a binary indicating a change in the residential area between two consecutive years.

The basis of our analytical strategy involves analyzing the probabilities of moving-out by comparing on the one hand natives exposed to one or more native out-movers and no native in-movers, and on the other hand, natives exposed to no change (i.e. neither in- nor out-movers). The level of granularity offered by

the registers thus allows us to capture in a simple way the basic intuition behind our SI hypothesis, whereby having previously observed others moving-out (provided there are no in-movers) increases the probability of moving-out in comparison to the scenario in which there are neither in- nor out-movers.

Most importantly, this design allows us to adjust for relevant confounders that are likely to modify the effect of SI across individuals and areas. In particular, we follow Shalizi and Thomas (2011) (translated to the neighborhood context) and argue that the likelihood of following previous out-movers is confounded by factors that account for the high degree of similarity between neighbors within a particular area, which can also cause moving-out behavior. This latter property, generally known as homophily (McPherson et al., 2001), is a well-documented fact in residential segregation, especially along socioeconomic and ethnic lines (Charles, 2003; Quillian, 2012). Hence, adjusting for factors that account for neighborhood mobility will allow us to differentiate the SI effect from scenarios in which similar movers may show a higher likelihood of moving as a result of some shared attribute.

As described in the previous section, we adjust for four main types of confounding factors that influence residential mobility behavior. First, we include life-course covariates. In particular, we adjust for natives' age, type of family, civil status, and type of tenancy tenure (Clark and Huang, 2003). Second, we include socioeconomic factors related to the individuals and the residential areas, such as the number of years in education, their disposable income (logged), and the median disposable income (logged) of the residential area (South and Crowder, 1998). Third, we account for ethnic preferences by including natives' country of birth according to their parents' country of birth (either both of Swedish origin, otherwise from E.U.-15/North-America), and the proportion of non-natives in the residential area (Crowder, 2000). Finally, we adjust for the number of inhabitants in the residential area to account for the visibility of out-movers (Valdez, 2014).

One potential drawback of our analytical strategy is that the groups to be compared might differ too greatly on a given covariate. Table 1 shows the situation for

the Full sample. As can be seen, some individual covariates are somewhat imbalanced across groups, such as the proportion of natives with children, the proportion of renters and owner-occupiers, or the ethnic composition, socioeconomic status, and number of inhabitants in the residential area. Scholars have previously argued that standard regression adjustment alone might be an inefficient way to estimate causal effects as a result of larger standard errors (Rosenbaum and Rubin, 1983) and increased dependence on the functional form presumed to gauge effects (Ho et al., 2007). To overcome these issues, we apply Coarsened Exact Matching (CEM) to adjust for the outlined covariates (Iacus et al., 2012). CEM is a non-parametric adjustment technique that is also intended to improve the degree of similarity between groups given a set of covariates (Stuart, 2010). As can also be seen in Table 3.1, CEM produces a remarkable improvement in the average value of all covariates between the groups (see Appendix for a more complete overview of the performance of CEM).

Finally, to ensure that covariates are always measured before natives are exposed to out-movers (Pearl, 2010), we dynamically rematch individuals over the available years (Aral et al., 2009) by splitting the temporal range into overlapping chunks of three consecutive years. The first year determines the measurement of the covariates, the second the treatment exposure, and the third the outcome: the probability of a household leaving its current residential area. By applying this dynamic matching, we ensure that natives are always matched with other similar natives within any given trial, and also that they are conveniently rematched in the case of any of their covariate values changing over time.

Once CEM has been applied to each trial, we gauge our SI estimates by applying a weighted linear probability model (LPM). The model consists solely of the binary moving outcome, the variable indicating which group natives are in, and the CEM weights. The choice of this approach rather than the more conventional use of logistic regression for the analysis of binary outcomes is because we are concerned with estimating the difference between groups rather than on predicting, and also

	Full sample		Matched sample	
	No change	Out-movers	No change	Out-movers
<i>Age</i>	50.58	49.96	49.84	49.96
<i>Years of education</i>	12.13	12.4	12.4	12.41
<i>Disposable income (log)</i>	7.44	7.48	7.48	7.49
<i>Single</i>	0.25	0.25	0.23	0.23
<i>Married</i>	0.63	0.64	0.67	0.67
<i>Divorced</i>	0.12	0.11	0.1	0.1
<i>With children</i>	0.55	0.62	0.64	0.64
<i>No children</i>	0.29	0.25	0.26	0.26
<i>Sweden</i>	0.84	0.83	0.86	0.86
<i>EU-15/North-American</i>	0.16	0.17	0.14	0.14
<i>Renter</i>	0.01	0.03	0.01	0.01
<i>Owner</i>	0.01	0.03	0.02	0.02
<i>Neigh. pr. non-Westerners</i>	0.06	0.09	0.07	0.07
<i>Neigh. median disp. income (log)</i>	7.45	7.47	7.47	7.48
<i>Neigh. N inhabitants</i>	12.11	24.54	18.76	20.46
<i>N</i>	1,300,050	2,977,420	1,106,555	2,604,240

Table 3.1: Basic descriptives for the entire sample of natives exposed to neither in- nor out-movers, and of natives exposed to at least one person previously having moved-out from their residential area and no in-movers. Each row presents the mean value of a covariate, either in terms of its numerical (logged) scale or as a proportion in the case of categorical covariates. Each value is reported for the entire sample prior to matching (Full sample) and after matching. The smaller number of cases in each group after matching is due to pruning. See Appendix for a more detailed report on the matching performance.

with comparing the estimates across non-nested models, which is not feasible using logistic regression without making what are typically very strong and unverifiable assumptions (Mood, 2010). To gauge the statistical significance of differences between estimates from different models, we follow Clogg et al. (1995).

3.4 Results

3.4.1 The SI effect on residential out-mobility

Figure 3.1 presents the results for our first hypothesis: the social influence effect on neighborhood mobility. The figure shows the SI effect after matching (CEM+LPM, black) as captured by the beta coefficient of the LPM. Error bars represent confidence intervals (CI) at the 95%, with p-values in parenthesis (See Appendix for results and statistics in tabulated form). To give a sense of the adjustment made

by CEM, we compare it to the estimate made before matching (Raw, white), thus without incorporating any of the covariates.

As can be observed in the Figure, the model shows a clearly positive estimate for moving-out after being previously exposed to out-movers by comparison with exposure to no change. The same estimate with no adjustment made for any covariates seems to notably overestimate the effect by comparison with the estimate after CEM. The difference between the groups is highly statistically significant (p-value<0.001). At the same time, the estimates for moving-out seem to be rather low, partly due to the much higher predisposition not to move, which has also been noted by other mobility scholars (Clark and Onaka, 1985; Clark and Dieleman, 1996). In addition to factors already noted by those authors, we hypothesize that this is also partly due to the strength with which the SI signal is received by the individuals concerned, which we analyze below.

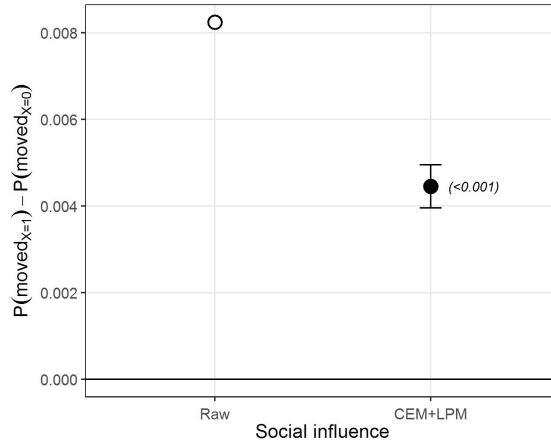


Figure 3.1: The SI effect on residential out-mobility. Difference in the probability of moving-out between natives exposed to out-movers and no in-movers ($X = 1$) and natives exposed to no change (neither in- nor out-movers) ($X = 0$). The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-value in parenthesis. The straight horizontal line indicates no effect.

3.4.2 The strength of the SI effect on residential out-mobility

While the previous result provides evidence for the effect of SI on residential out-mobility, the following section evaluates how the strength of the signal made by out-movers leaving residential areas is affected in four specific domains: the number of previous out-movers, the density of the residential area, the spatial distance from previous out-movers, and the ethnic composition of the residential area.

We start by analyzing the marginal effect on the strength of SI by comparing natives who are exposed to a larger (left) or smaller (right) number of previous out-movers, for each binary configuration respectively (see Figure 3.2). In line with our expectations, the plot clearly shows that being exposed to a larger number of out-movers increases the probability of moving-out compared to a lower exposure, as indicated by the positive sign and high statistical significance of the estimates. The largest increase in this probability seems to occur when the previous number of out-movers increases from 2 to 3, as indicated by the statistically significant difference between the two. An increase of 4 or more out-movers also seems to yield an increase in the probability of moving-out compared to a lower number, although the estimate does not seem to be either qualitatively nor statistically significantly different from an increase of 3 or more out-movers. Again, the model without covariates appears to markedly overestimate the effects by comparison with the estimate after CEM.

We continue by analyzing the effect of SI across residential areas with different total numbers of inhabitants (see Figure 3.3). As can be seen from the plot, CEM estimates seem to show a downward trend in the effect on moving-out as the area becomes more densely populated. We observe that SI seems to have a marked effect in residential squares with relatively low population density, here defined as those with up to 15 inhabitants. The effect remains positive and statistically significant for areas with a medium number of inhabitants, between 16 and 30, although the effect is clearly weaker than for the lower density case, with the difference between this and the lower density scenario being statistically significant. Finally, the model seems to indicate no effect after crossing the threshold of 30 inhabitants, partly

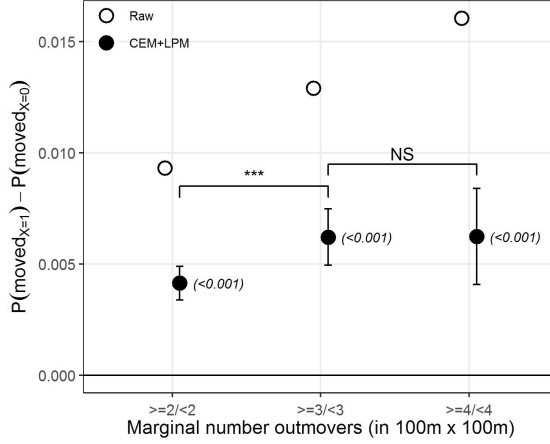


Figure 3.2: The marginal effect of the number of previous out-movers on the strength of SI. Difference between natives who are exposed only to 2, 3, or 4 or more previous out-movers ($X = 1$, left), or fewer than this ($X = 0$, right), respectively. The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$.

because such cases are much less common in the data. These results are also in line with our expectations, with lower density areas allowing previous out-movers to be more easily noticed than higher density areas. Interestingly, the estimates yielded by CEM show an important correction in relation to higher density areas by comparison with the model without adjustment, which instead shows a larger effect that is most likely capturing out-moving behavior due to other factors in more high-density urban areas.

Figure 3.4 shows the marginal effect on SI of increasing the spatial distance between previous out-movers and the location of natives. To gauge this effect, we compare natives exposed to out-movers that either left from “closer” distances from their residential location (indicated by the upper product in each category) or from “farther away” (the lower product in each category), bounded by the $400\text{m} \times 400\text{m}$ square. As expected, the plot indicates a greater SI effect when out-movers moved from more proximate areas compared to more distant ones. As can be seen from the Figure, the estimates are greatest when the comparison is made between exposure to out-movers either within or beyond an area of $100\text{m} \times 100\text{m}$, after which the

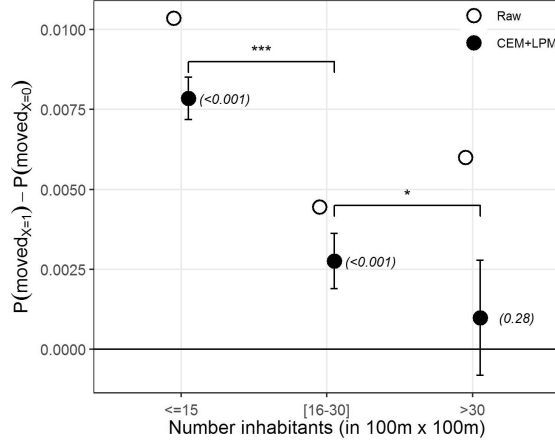


Figure 3.3: The effect of residential density on the strength of SI. Effect heterogeneity between natives exposed to out-movers ($X = 1$) and natives exposed to no change ($X = 0$) in $100\text{m} \times 100\text{m}$ residential areas with different numbers of inhabitants: (1) less or equal to 15; (2) between 16 and 30; (3) more than 30. The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$.

effect is weak or not statistically significant at all. The 100m model also seems to yield an estimate that is statistically significantly different from the 200m model. Thus, the models indicate the SI effect to be strongest when previous out-movers left from the most proximate area, as indicated by the $100\text{m} \times 100\text{m}$ residential area, to then diminish rapidly the greater the distance separating ego from previous out-movers, yielding no effect beyond the $300\text{m} \times 300\text{m}$ residential area.

Finally, Figure 3.5 shows the effect of the ethnic composition of residential areas on the strength of SI. As can be seen from the plot, the estimated effect is clearly positive and statistically significant in areas where the presence of non-natives is null, suggesting evidence for the presence of SI even in native-only areas. Nevertheless, the effect remains positive and statistically significant as areas become more ethnically mixed, although the size of the effect declines, thus leading us to reject the direction of our final hypothesis, that natives moving-out from non-native areas would be more salient—more influential—for other natives than natives leaving native-dominated areas.

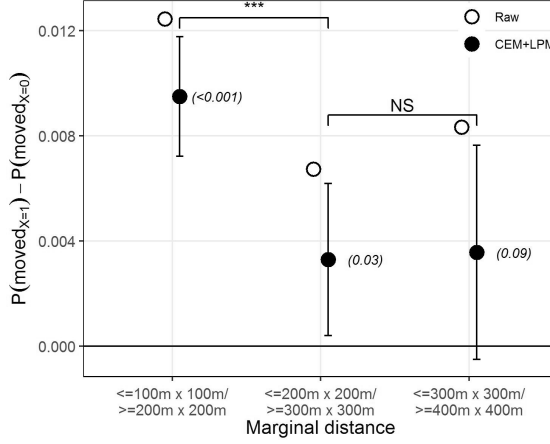


Figure 3.4: The marginal effect of distance on the strength of SI. Difference between natives being exposed only to out-movers leaving from a point close to their own residential location ($X = 1$, upper) or from farther away ($X = 0$, lower), respectively for three different spatial levels: (1) $100\text{m} \times 100\text{m}$; (2) $200\text{m} \times 200\text{m}$; (3) $300\text{m} \times 300\text{m}$. The maximum area is bounded by the $400\text{m} \times 400\text{m}$ square at all levels. The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. '***' $p\text{-value} < 0.001$, '**' $p\text{-value} < 0.01$, '*' $p\text{-value} < 0.05$, 'NS' $p\text{-value} > 0.05$.

Figure 3.5 highlights two important aspects of the role of SI and ethnic composition on neighborhood mobility. First, it demonstrates that SI processes may partly drive residential mobility independently of other ethnic preference-based explanations, since previous out-movers in native-only areas do not change the ethnic composition by increasing the presence of ethnic minorities. Second, the fact that the SI effect is smaller in ethnically mixed areas might potentially suggest that ethnic preference-based mobility may be stronger than SI in these areas, or that the two processes are entangled in non-linear ways, making it difficult to distinguish them from one another, and potentially making them impossible to identify using only observational data. This indicates a problem that has previously been largely unnoticed in the literature on segregation and residential mobility as a result of having overlooked SI processes of the type defined here, namely that ethnic-based processes may be entangled with SI processes and that the two may be indistinguishable from each other, which at the same time suggests a potential for the study of new mechanisms that combine the role of ethnic preferences and social influence in order to

identify new dynamics in residential mobility.

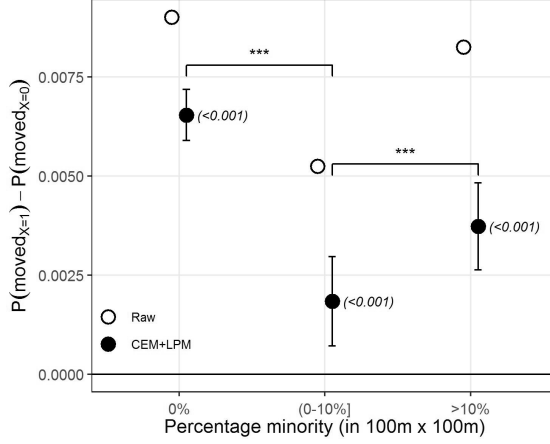


Figure 3.5: The SI effect across residential areas with varying ethnic compositions. Effect heterogeneity between natives exposed to out-movers ($X = 1$) and natives exposed to no change ($X = 0$) in $100\text{m} \times 100\text{m}$ residential areas across three types of ethnic composition: (1) only natives (left); (2) less than 10% non-natives (middle); and (3) more than 10% non-natives (right). The estimates are the β coefficient from the LPM on the raw data (in white) and after applying CEM (in black). Error bars indicate 95% confidence intervals, p-values in parentheses. The straight horizontal line indicates no effect. The upper bars indicate whether coefficients across models are statistically significantly different from one another. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$.

3.5 Conclusion and Discussion

SI is a ubiquitous phenomenon in the social sciences. Understanding how SI operates not only provides a better understanding of how individuals take decisions in several areas of their lives, but also provides a first foundation stone for accounting for the development of macro social patterns. The primary goal of this chapter has been to make a first exploratory attempt to analyze the presence of a new mechanism, the SI effect, on residential mobility. To achieve this overarching goal, our analysis has focused on answering two main research questions. First, are there SI processes that affect the likelihood of moving-out from a neighborhood after having observed others doing so? Second, is the strength of this effect moderated by the ease with which the signal is perceived?

In order to isolate the effect of SI on moving-out behavior, we applied CEM and

adjusted for several previously tested factors that influence neighborhood mobility. Our results show that after adjusting for theoretically relevant confounders, SI has a significant effect on neighborhood out-mobility and, furthermore, that the strength of this effect is stronger to the visibility of others' behavior. We measured the visibility of this signal on the following dimensions: signal size, distance to ego, and the signal's salience given the social environment.

Our work is not exempt from limitations. First, even though we controlled for several factors that the mobility literature has identified as affecting neighborhood mobility, our data cannot rule out the effect of neighborhood reputation on mobility (Bryant and Oliver, 2009). Second, another important limitation of the registers employed and of our identification strategy is that we cannot adjust for the ethnic composition after natives are exposed to out-movers (Ho et al., 2007). This could be problematic because, as has repeatedly been stressed in the segregation literature, natives tend to leave as soon as areas start to become mildly mixed (Grodzins, 1957). It is possible that this problem may potentially persist even after adjusting for the ethnic composition of the area prior to the exposure occurring, as the change produced by out-movers might theoretically still push natives to move-out as a result of an increase in the area's non-native share, rather than as a result of SI per se. This means that SI and ethnic-related motivators for moving-out behavior (Charles, 2003) may be confounded in an essential way, at least in ethnically mixed areas.

As regards the size effect of our results, they should be considered in the light of three main characteristics of residential mobility. First, as previous studies have suggested, moving is a highly costly behavior. Evidence shows that only a small fraction of the population moves each year (Krysan and Crowder, 2017). Thus, given the strong tendency for inertia and for staying in the same neighborhood, finding even small SI effects in this context says something about the strength of SI effects. Second, residential segregation is an inherently dynamic process (Schelling, 1971). As such, careful attention should be paid to small effects that might easily scale-up over time in a non-obvious and unpredictable way (Merton, 1988), as it is the case

of reinforced behaviors due to SI dynamics (Salganik et al., 2006). And third, our model implicitly assumes that all the neighbors are equally able to perceive others' residential movements. However, even though this perfect information assumption does not hold in real-life situations (even in small areas, there is a chance of not seeing others' people moving out), we found the SI effect. This means that our results underestimate the effect of SI since there will be situations in which receivers will not receive the signal despite the fact others might move out close to them. Accordingly, our estimations are at the very least conservative of the magnitude of the SI effect.

Our results provide clear evidence of an effect of SI for the native population with regard to out-mobility. However, further research is needed to study the complex ways in which the SI mechanism of neighborhood mobility proposed here interacts with households' ethnic preferences. Another natural extension of this study would be to evaluate the effects of SI on moving-in behavior, in addition to evaluating whether non-native households show similar SI patterns of mobility. Finally, our study also raises some interesting questions in relation to studies of residential segregation. For example, it establishes new factors with which to understand the dynamic aspects of ethnic segregation. This could potentially complement and enrich current models of segregation that have only taken the effect of households' neighborhood preferences into account.

Chapter 4

Exit or Voice (and When)? Refugee Exposure, Westerners' Residential Mobility, and Radical-Right Support

Abstract

¹ The refugee crisis of 2015 became a major issue of both national and pan-European debate. Behavioral reactions among natives in the form of support for radical-right parties or leaving neighborhoods following influxes of non-Westerners are well documented, but a detailed account of how asylum seekers contribute to these dynamics remains elusive. In this final empirical chapter, I study how asylum centers and refugees choosing their own residences prompt each of these two behavioral outcomes using register data for the whole of Sweden (2013-2018). The analyses show a divergence depending on the particular type of refugee exposure experienced and the specific behavior under analysis. Only increased radical-right support is observed following the establishment of a new asylum center, whereas greater native out-mobility is found following refugees self-selecting into native-based areas.

¹I would like to thank Felix Lennert for his contribution in providing the data on asylum centers in Sweden by Migrationsverket.

4.1 Introduction

The refugee crisis of 2015 and its aftermath, reflected in numerous European countries facing an influx of large numbers of asylum seekers, became a major issue of both national and pan-European debate. The crisis has highlighted the deficiencies of current asylum systems following decades of restrictions being implemented across these countries (Dustmann et al., 2017; Hatton, 2017; Bernhard and Kaufmann, 2018), which have contributed to precluding the assimilation of refugees as a result of long processing times for asylum petitions (Hainmueller et al., 2016; Hvidtfeldt et al., 2018), and have increased the risk of living in residentially segregated areas (Campesi, 2018; Andersson et al., 2019).

Recent research has shown that natives of host societies may change their behavior as the refugee population increases. In particular, a greater refugee population has been documented to increase levels of support for radical-right parties (Vasilakis, 2017; Dinas et al., 2019), and to perpetuate the concentration of poverty at the neighborhood level through a devaluation of housing prices near asylum centers (Dröes and Koster, 2019; Kürschner Rauck, 2020). Although this research has increased our understanding of how the refugee crisis has impacted natives of Western European societies, an accurate description detailing how this occurs remains elusive.

Most importantly, previous research on far-right support (Valdez, 2014) and residential mobility (Grannis, 2005) has shown that the salience of ethnic minority visibility and spatial proximity are key factors that are necessary to influence the behavior of natives. These factors underline the importance of the way in which natives encounter refugees in their local contexts, rather than the total refugee population, as a primary condition for increasing far-right support and neighborhood out-mobility (see Logan, 2012). However, due mainly to limitations in the available data, most research has relied on the share of refugees in large areas to account for natives' responses, and has thereby underestimated the outcomes produced by refugee presence by assuming equal salience throughout these large areas, whose size

means that refugee visibility may in fact be too low to produce any effect (Park, 1924).

This chapter aims to fill this gap by analyzing natives' behavioral outcomes following two qualitatively different modalities of local refugee exposure: asylum centers and refugees self-selecting into the housing market. While both produce an increase in the total share of refugees, they entail radically different ways in which natives may experience new refugees moving into their local areas. For instance, asylum centers markedly increase visibility salience by means of a substantial influx of asylum seekers and their placement in highly limited spaces. At the same time, the high volatility that characterizes the presence of these asylum seekers, as their applications are dealt with and asylum centers are gradually dismantled over time, greatly limits the period during which they may exert any influence on the native population. This is in stark contrast to the situation in which asylum seekers self-select into native-based areas, with these asylum seekers producing a sparser degree of exposure compared to asylum centers, but an exposure that may potentially be longer-lasting due to the asylum seekers remaining in the same area upon being granted permission to stay in the host country, and that may also increase the area's appeal for other refugees (and non-asylum seeker ethnic minorities) to move in (Clark, 1991).

I study how these two modes of exposure may have influenced natives' propensity to move-out and to support radical-right parties following the establishment of a new asylum center or incoming refugees self-selecting into native-based areas. I employ Coarsened Exact Matching (Iacus et al., 2012) and a Synthetic Control Method (Abadie et al., 2010) to analyze each behavioral outcome respectively, using Swedish register data (2013-2018) for two reasons: the first being that Sweden is among the countries that have received the largest proportion of asylum seekers in the E.U. (UNHCR, 2015), which ensures a large sample of natives who have been exposed to neighboring refugees; the second being that the information provided by the registers on actors with refugee status allows me to disentangle the role of

asylum seekers and non-asylum seekers in relation to the non-Westerner exposure effect (FitzGerald and Arar, 2018).

One could summarize the results as follows. The analyses show a divergence depending on the particular type of refugee exposure and the specific behavior under analysis. More concretely, results show higher radical-right support in areas following the establishment of a new asylum center, but not following an increase in refugees self-selecting into neighborhoods. In contrast, the models indicate a positive increase in native out-mobility following refugees self-selecting into native-based areas, but not after the establishment of a new asylum center. This latter effect is moderated by the prior ethnic composition of the area, and increases as the ethnic minority presence becomes greater. However, this does not apply to the establishment of new asylum centers, which produce no out-mobility effect regardless of the area's ethnic composition.

The analyses presented in this chapter advance our understanding of how the 2015 refugee crisis has impacted Western European countries by showing the importance of detailing the way in which natives encounter a refugee population in order to understand their behavioral reactions. The results described suggest that the type of refugee exposure on the one hand, and the type of behavioral outcome on the other, interact and determine how natives who are discontented by influxes of asylum seekers in their local contexts may express their discomfort (Hirschman, 1970): while new asylum centers might increase radical-right support, only refugees self-selecting into accommodation in a given neighborhood contribute to forming a perception that the area's ethnic composition may remain mixed, thus increasing the likelihood of natives out-mobility from the neighborhood in question.

This chapter also makes important policy-related contributions. By describing in greater detail how natives react to different modes of refugee dispersal within the host country, the chapter provides a first assessment of how the allocation of new asylum centers and affordable housing might be optimized, which can inform policy solutions aimed at maximizing refugee assimilation by lowering their impact on the

native population (Black, 2001).

4.2 Literature review

4.2.1 The refugee crisis in Sweden

Sweden's extensive acceptance of asylum seekers and refugees throughout its recent history is well-documented (Migrationsverket, 2021). While the country has witnessed several peaks in the number of asylum seekers since the 1980s, it was only during the 2015 refugee crisis that the country experienced such a massive, hitherto unforeseeable growth in the number of arrivals. Levels first peaked in 2014, when the Swedish Migration Agency received over 80,000 asylum seekers mainly from Syria, followed by Eritrea and by stateless persons. Germany and Sweden accounted for thirty and thirteen percent respectively of all asylum claims made in the E.U. in 2014 (UNHCR, 2015), with Sweden guaranteeing permanent residence permits to all refugees from Syria. Just one year later, Sweden saw a doubling of the size of this peak, receiving a further 163,000 asylum seekers, this time primarily from Syria and Afghanistan (including 35,000 unaccompanied minors). By 2016, increased border controls and changes in Sweden's migration laws had made it more difficult to receive residence permits and to reunite with family, contributing to a reduction of the total number of asylum seekers to 29,000.

The refugee crisis of 2015 had a marked impact on the ethnic composition of Sweden's population, with Syria displacing Finland (after more than seventy years) as the country with the largest foreign population. The experience of this rapid ethnic change over such a short period contributed to directing a strong media focus at Sweden's response to the crisis, sparking further debates over the refugee issue that already permeated both the public and political discourse. The radical-right party Sweden Democrats (Sverigedemokraterna), for instance, doubled their share of the electoral vote in 2010 (Rydgren and Ruth, 2013) and continued to increase this vote share in 2014 (Müller et al., 2014), achieving their best results since their

inception in 1988.

4.2.2 Ethnic minority exposure, and natives' radical-right support and residential mobility

Hirschman (1970) has highlighted two key alternate ways through which discontented members of an organization may express their discomfort: “exiting” and “voicing.” I follow this model and investigate how natives who are discontented following influxes of refugees in their local areas may express their discomfort, either by changing their residential mobility patterns by moving-out of neighborhoods (*exiting*), and/or by increasing their level of support for radical-right parties (*voicing*).

Research has shown that an increase in the presence of ethnic minorities in an area is positively correlated with increased support for radical-right parties (Rydgren, 2007). Similar results have been found regardless of whether the minorities under study are immigrants (Knigge, 1998; Swank and Betz, 2003; Golder, 2003), native minorities (Rydgren and Ruth, 2011, 2013; Valdez, 2014; Dustmann et al., 2019), or asylum seekers (Lubbers and Scheepers, 2001; Arzheimer, 2009; Kenny and Miller, 2020). Recent studies analyzing the 2015 refugee crisis have confirmed these results by showing higher radical-right support in areas with significant exposure to asylum seekers (Vasilakis, 2017; Dinas et al., 2019).

In a similar line of inquiry, previous research on residential mobility has documented that natives show a higher tendency to move out of and avoid neighborhoods with large influxes of ethnic minorities (Grodzins, 1957). These results confirm that, although individuals' mobility rates are determined by their socioeconomic status (DeLuca et al., 2019) and their self-selection into neighborhoods that they can afford, the larger the neighborhood share of ethnic minorities, the greater the likelihood that natives will move out (South and Crowder, 1998; Quillian, 1999, 2002; Crowder, 2000; Crowder et al., 2006; Card et al., 2007; Pais et al., 2009; Schaake et al., 2010; Andersson, 2013; Aldén et al., 2015; Müller et al., 2018).

4.2.3 Ethnic minority visibility and superficial contact

One prominent cause accounting for the observed levels of radical-right support (Rydgren, 2007) and out-mobility patterns (Massey and Denton, 1993) in ethnically mixed societies is found in the level to which natives hold anti-immigrant attitudes and prejudices (Pettigrew, 1998; Zick et al., 2008).

One explanation to account for changes in inter-group attitudes is the so-called contact hypothesis (Allport, 1954). This hypothesis posits that greater levels of inter-group contact may serve to decrease prejudice provided the following conditions apply: (1) groups share equal status in their contacts; (2) they share common goals; (3) they engage in cooperative tasks; and (4) authorities, the law, etc. provide external support. Despite abundant experimental evidence confirming the predictions of this hypothesis (for a review, see Pettigrew and Tropp, 2006), results analyzing observational data have primarily found higher levels of anti-immigrant attitudes the larger a community's minority population (Blalock, 1957; Scheepers et al., 2002; Schneider, 2008), including asylum seekers (Hangartner et al., 2019). Thus, contrary to expectations derived from the contact hypothesis, these studies have concluded that neighborhood and even school contact are not sufficient to decrease anti-immigrant attitudes (McLaren, 2003) (although cross-group friendship can, see Turner et al., 2007).

Valdez (2014) has accounted for this incongruence by showing that natives who do not necessarily engage “meaningfully” with other ethnic minorities, but who nevertheless encounter out-group members and engage with them sporadically in daily activities, are more likely to support a radical-right party. This type of “superficial contact,” moderated by the salience of out-groups and their spatial proximity to natives, is capable of reinforcing anti-immigrant attitudes (Semyonov et al., 2008) by increasing the perception that the share of out-group members in a given area is greater than it actually is (Wong et al., 2012; Craig and Richeson, 2014), as well as by highlighting self-group salience among natives. In a laboratory experiment, for example Paolini et al. (2010) have shown that natives tend to report

higher levels of self-group identity following negative experiences of sporadic inter-group contact.

Most importantly, superficial contacts can bring anti-immigrant attitudes to the fore as a politically salient issue. Through an increased salience of out-group visibility, political dimensions such as multiculturalism and integration may gain more relevance for voters than less salient issues in specific electoral contests (List and Dietrich, 2011). In line with this, Hopkins (2010) has found that electoral campaigns that politicize the issue of immigration tend to increase the vote share of the radical right by comparison with campaigns that occur in a context of high immigration, but where immigration is not a hot political issue. Rydgren and Ruth (2013) have found increased support for radical-right parties in municipalities in which out-group presence is greatest, while Dustmann et al. (2019) have documented significant growth in such support mainly in rural areas that are exposed to asylum centers. Likewise, others have found higher levels of anti-immigrant sentiment (Hangartner et al., 2019) and radical-right support (Vasilakis, 2017; Dinas et al., 2019) among natives living in territories that are more directly adjacent to neighboring countries from which asylum seekers are more likely to arrive.

Evidence for the role of the salience of out-group visibility and of superficial contact has also been found in research on residential mobility (Logan, 2012). Adopting a similar perspective, scholars have shown that natives' residential mobility decisions are markedly shaped by interactions that occur mainly inside local communities that are internally connected by pedestrian streets and delimited by non-pedestrian roads (Grannis, 1998, 2005). This has also been highlighted by studies showing that increased minority presence is incapable of changing the residential mobility of natives living in adjacent neighborhoods (Crowder and South, 2008; Crowder and Downey, 2011). Finally, others have recently found that house values decrease with proximity to asylum centers, especially in locations, such as streets containing markets, that are visited by both natives and asylum seekers, where inter-group contacts are greatest (Dröes and Koster, 2019; Kürschner Rauck, 2020).

Building on the evidence showing the role of out-group salience in changing the behaviors of natives via superficial contact, I hypothesize that differences in the interaction between the salience of out-group visibility and the type of refugee exposure will prompt different reactions by natives. Following Valdez (2014), I expect that the higher salience produced by asylum centers will primarily prompt radical-right support, while the temporary nature of such centers will not convey the impression that the increased out-group presence associated with them will translate into permanent ethnic change in the area, undermining the reinforcement necessary to spark out-mobility (Ottensmann, 1995). At the same time, the more sparse exposure to asylum seekers associated with their self-selecting into native-based areas will not be sufficient to increase radical-right support (Valdez, *ibid.*), but may exert a stronger influence on the residential mobility of natives via an increase in the appeal of these areas to other refugees (and non-seeker migrants) who might then also move in (Clark, 1991). Therefore, incoming refugees will exert a greater influence on the likelihood of native out-mobility the greater the level of ethnic minority presence in the natives' local contexts.

4.3 Data and Methods

4.3.1 Data

The analyses employ Swedish register data, a dataset collected by Statistics Sweden that contains records of all individuals living in Sweden, with virtually no missing data. The data employed take two different forms, depending on the outcome examined. For the residential mobility analysis, the registers provide micro-data for all Swedes/Western Europeans (natives) for the period 2013-2017, and include information on their residential locations at the level of the one-hundred-by-one-hundred meter residential square (henceforth “residential square”), and the year in which they moved in. For the voting analysis, the registers provide data aggregated at the level of polling districts, and include the vote share of the political parties in any of

the three main types of election (municipal, county, and national), which have been conducted in Sweden every four years during the period 2006-2018.

In order to gather information about the type of refugee exposure natives may have experienced, I combine the registers with information on asylum centers provided by the Swedish Migration Agency. On the one hand, the registers contain precise information on refugees granted permanent residence permits in Sweden, including the year in which a decision was made on their asylum petition. Aside from this, the registers contain the same data for refugees as for others, including their residential location and country of birth. Most importantly, this not only allows for native exposure to non-Westerner asylum seekers to be distinguished from exposure to non-Westerners who are not asylum seekers (FitzGerald and Arar, 2018), but also ensures that these refugees are not counted as residing in an asylum center, and have therefore chosen their own accommodation. This is due to Sweden’s policy to provide temporary housing to asylum seekers who need such accommodation during the period in which their petitions are being processed. The data on asylum centers is not in the registers and has been provided to this study by the Sweden Migration Agency, which contain the location of all asylum centers established at the municipal level and the number of asylum seekers living in them during the period 2007-2019.

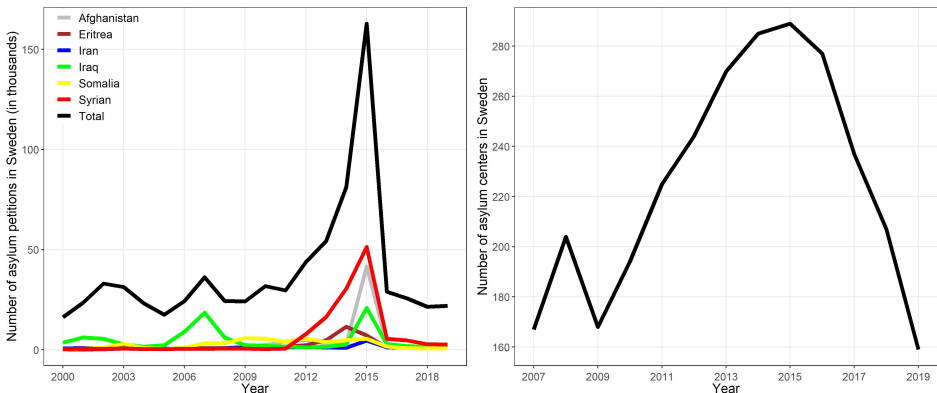


Figure 4.1: The refugee crisis in Sweden. (*LEFT*) Number of asylum petitions (2000-2019), with colors indicating the six countries producing the largest numbers of asylum petitions during the refugee crisis. (*RIGHT*) Number of asylum centers (2007-2019). Source: Migrationsverket.

Figure 4.1 shows relevant information about asylum seekers in Sweden. The subplot on the left shows the annual number of asylum petitions for the period 2000-2019. The colored lines represent the six countries which produced the largest number of asylum petitions during the refugee crisis of 2015. The figure clearly shows that the largest number of asylum petitions were made by refugees with Syrian citizenship (albeit closely followed by other groups). The refugee crisis of 2015 represented a clear spike in the historical trend in the number of asylum petitions in Sweden, necessitating the opening of additional asylum centers to handle the increased number of petitions (see the subplot on the right). The number of centers increased monotonically after 2009, with this increase accelerated during the period 2011-2013, and then peaking in 2015 when more than 280 centers were in use across the country. The total number of active asylum centers then quickly diminished, falling to 160 centers nationwide in 2019, a number not seen since the previous decade.

4.3.2 Analytical strategy

In order to disentangle the propensity of natives either to move-out or support a radical-right party following different types of refugee exposure, I perform two separate analyses following a counterfactual design. First, I compare natives exposed to an increase in the presence of refugees whose residence applications had been granted with natives exposed to no change. Second, I compare natives exposed to the establishment of a new asylum center to natives who were not exposed to a new asylum center. This approach allows for the study of natives' reactive behaviors by comparing the situation in which a given type of refugee exposure is present with the situation in which that same type of exposure is absent (see Woodward, 2003).

Further, the approach enables me to adjust for variations among natives that are likely to impact their likelihood of experiencing any refugee exposure and their propensity to move-out or support the radical-right (Pearl, 2010). I follow a different strategy for each behavioral outcome under analysis, reflecting the differing available

datasets. To adjust for confounders that affect natives' propensity for out-mobility, I apply Coarsened Exact Matching, and to do the same at the level of polling districts, I apply a Synthetic Control Method. Following Ho et al. (2007), the weights resulting from the application of these methods are included in separate linear regression analyses, which are used to gauge the effect of refugee exposure on each behavioral outcome using the ordinary-least square (OLS) method, respectively.

Residential mobility

The availability of individual-level data and the granularity provided by the registers allows me to apply Coarsened Exact Matching (CEM) (Iacus et al., 2012), which some have argued constitutes a more efficient means of adjusting for confounders (covariates) than regression adjustment alone (Stuart, 2010). This non-parametric matching strategy ensures that the groups to be compared are as similar as possible given data on a set of observed covariates. The resultant increase in similarity benefits the estimation by reducing its uncertainty (Rosenbaum and Rubin, 1983) and lowering its reliance on the specific functional form of the model employed to gauge it (Ho et al., 2007). I follow standard practice and ensure that the standardized difference in means for each covariate is below 0.2 (Stuart, *ibid.*).

The literature on neighborhood mobility has highlighted four main types of covariates that shape the propensity to move. (1) First, the authors of existing studies have highlighted the relevance of life-course events, such as having offspring, getting married, or going to university, as important triggers of mobility (Clark and Huang, 2003; Clark and Withers, 2007). I therefore include natives' age, civil status, and family type. (2) Socioeconomic factors also moderate residential mobility, and I thus include individual disposable income (logged), their type of tenure tenancy and ownership, years of education, and neighborhood socioeconomic status (which I measure here as the median disposable income (logged) within the residential square) (South and Crowder, 1998; Crowder et al., 2006). (3) To adjust for the tendency of individuals to cluster in neighborhoods with high proportions of other

co-ethnics, I adjust for the non-asylum seeker ethnic minority presence in residential squares and the natives' country of birth (either their own, or that of their parents, categorized either as Sweden, or else from E.U.-15/U.S./Canada). (4) Finally, I adjust for the total number of inhabitants in the residential squares, the county of residence, the previous level of refugee, and whether there are any asylum centers in the municipality.

Finally, to ensure that the measurement of covariates precedes the natives' categorization into the exposure groups, I split the period examined available into overlapping trials of three consecutive years. Each native in each trial then has a measure for each of the covariates corresponding to the first year of the trial, the exposure group into which they fall as recorded in the second year, and their mobility outcome in the third year. This approach ensures that each native is matched to similar natives at any time point, and conveniently rematches them should any of their covariates change over time (Aral et al., 2009).

Table 4.1 shows the mean values of each covariate for each group exposed to an increased presence of refugees prior to (Full sample) and after matching. The table shows how CEM produces similarity in each covariate value across groups, indicating, as expected, that the groups look more similar with regard to the observed covariates after matching (complete information about the performance of CEM is available in the Appendix).

Far-right support

The available voting data aggregated at the polling-district level reduces the efficacy of the matching method due to an increase in the difficulty of finding reasonably good matches. To overcome this limitation, I employ a Synthetic Control Method (SCM) (Abadie et al., 2010). This method extends the traditional difference-in-differences (DiD) approach by ensuring that polling districts that receive an intervention are comparable to other polling districts that did not receive it (the "donor pool"). This is achieved via the estimation of a "synthetic" control from the donor pool that

	Full sample		Matched sample	
	No change	Refugee increase	No change	Refugee increase
<i>Age</i>	53.28	53.05	54.38	54.58
<i>Years of education</i>	11.66	11.75	11.65	11.65
<i>Disposable income (log)</i>	7.51	7.48	7.51	7.51
<i>Single</i>	0.30	0.41	0.32	0.32
<i>Married</i>	0.55	0.37	0.48	0.48
<i>Divorced</i>	0.14	0.23	0.20	0.20
<i>With children</i>	0.47	0.34	0.40	0.40
<i>No children</i>	0.31	0.22	0.27	0.27
<i>Sweden</i>	0.90	0.85	0.90	0.90
<i>EU-15/US/Canada</i>	0.10	0.15	0.10	0.10
<i>Renter</i>	0.04	0.38	0.20	0.20
<i>Owner</i>	0.03	0.35	0.16	0.23
<i>Neigh. pr. non-Westerners</i>	0.02	0.06	0.04	0.04
<i>Neigh. pr. refugees</i>	0.02	0.13	0.10	0.09
<i>Neigh. median disp. income (log)</i>	7.52	7.50	7.53	7.53
<i>Neigh. N inhabitants</i>	12.83	148.58	42.8	51.59
<i>N</i>	7,637,003	4,452,037	6,398,544	1,778,813

Table 4.1: Descriptives for natives exposed to a proportional increase in refugees who had sought their own accommodation and natives exposed to no change (2013-2016). Rows show the mean value for each covariate prior to (Full sample) and after matching. Natives are also matched by county, not shown here due to space constraints. Numerical covariates are presented in terms of their own scale (or logged), and qualitative covariates are shown as proportions. The total number of cases declines due to pruning. Similarity is also achieved for natives exposed to a new asylum center and natives not exposed to a new center, omitted here due to space constraints (see Appendix for more information about the performance of CEM).

closely resembles the average trajectory of the intervened polling districts (Robbins et al., 2017), a recurrent limitation in DiD applications that SCM overcomes.

I use SCM to analyze radical-right electoral outcomes in 2018 in polling districts that received increases in refugees/new asylum centers at any point during the period 2015-2017. I employ the remaining available years prior to the “intervention” (i.e., 2006-2014) to create the synthetic control, using other polling districts that did not experience the relevant changes during the period 2015-2017 as the donor pool.

The literature on radical-right voting has shown that young, unemployed, working-class males with lower levels of education are more likely to manifest support for radical-right parties (Evans, 2005; Elisabeth, 2005; Arzheimer and Carter, 2006; Arzheimer, 2009). I therefore include the following covariates whose trajectories are matched for each year: ethnic minority presence measured as the percentage of non-Westerners in the polling district; the number of individuals in the district; the

median disposable income (logged); the percentage of asylum-seekers; the percentage of individuals with a university degree; the percentage of renters; the percentage married; the percentage younger than forty years of age; the percentage unemployed; the county in which the district is located; the presence/absence of asylum centers in the municipality surrounding the polling district; the percentage of votes cast for a party of the left; and the percentage of votes cast for the radical-right Sweden Democrat party. The latter two covariates were of course matched following each election (i.e., once per period 2006-2009, 2010-2013, and 2014).

Table 4.2 shows the mean value across the entire trajectory for each covariate for polling districts with new asylum centers and the synthetic control generated by aggregating poll districts without new asylum centers. As with the matching procedure, the table shows the mean value for those covariates whose trajectory has been matched annually from 2006 until 2014, save for those electoral-related covariates whose value is only matched following each election. The mean value for each covariate for polling districts that received no new asylum centers clearly approximates the mean value for the polling districts that received new centers, showing the improvement produced by the synthetic control in the level of similarity with the intervened polling districts (complete information about the performance of SCM is available in the Appendix).

	No center (Full sample)	No center (Synthetic)	New asylum center
<i>Pr. SD vote (municipal)</i>	0.023	0.015	0.019
<i>Pr. SD vote (county)</i>	0.022	0.017	0.022
<i>Pr. SD vote (national)</i>	0.028	0.025	0.028
<i>Pr. Left vote (municipal)</i>	0.328	0.326	0.325
<i>Pr. Left vote (county)</i>	0.334	0.337	0.332
<i>Pr. Left vote (national)</i>	0.335	0.332	0.325
<i>Asylum center (municipal)</i>	0.534	0.148	0.099
<i>Pr. refugees</i>	0.076	0.048	0.045
<i>N inhabitants</i>	1,301.341	1,205.866	1,208.21
<i>Pr. non-Westerners</i>	0.046	0.028	0.036
<i>Pr. with university</i>	0.294	0.262	0.264
<i>Pr. renters</i>	0.261	0.164	0.167
<i>Pr. married</i>	0.437	0.464	0.516
<i>Pr. unemployed</i>	0.463	0.439	0.45
<i>Pr. with less 40 years</i>	0.366	0.335	0.335
<i>Pr. female</i>	0.506	0.488	0.506
<i>Median disp. income (log)</i>	7.292	7.284	7.309
<i>N</i>	4,886	4,886	41

Table 4.2: Descriptives for polling districts in which a new asylum center was established at any point between 2015-2017, and those with no new asylum center during the same period (the donor pool). The table shows the mean value for covariates whose trajectory has been matched annually for the period 2006-2014, save for those election-related covariates that are only matched following each new election (i.e., once per period 2006-2009, 2010-2013, and 2014). The values for polling districts with no new asylum center are shown prior to applying the Synthetic Control Method (Full sample) and after (Synthetic). Cases are also matched by county, not shown here due to space constraints. Similar results are achieved for polling districts in where there is a proportional increase in refugees who had sought their own accommodation and polling districts with no such change, omitted here due to space constraints (see Appendix for more information about the performance of the SCM).

4.4 Results

I now present and discuss the behavioral responses of natives following each type of refugee exposure. Figure 4.2 shows the vote share for the radical-right Sweden Democrat party (SD) in the 2018 elections using OLS after SCM. The Figure shows one estimate per type of election: municipal (white), county (black), and national (grey). The estimates are obtained by contrasting the synthetic control ($X^* = 0$) with the average for polling districts that experienced an increase in refugee presence (left plot), and with the average for polling districts with at least one new asylum center (right plot) ($X = 1$). As can be seen, the models produce noticeably

divergent estimates depending on the type of refugee exposure. In particular, the models showing the effect of being exposed to refugees who have self-selected into the housing market (left) seem to yield estimates that are neither statistically nor qualitatively different from those of the synthetic control.

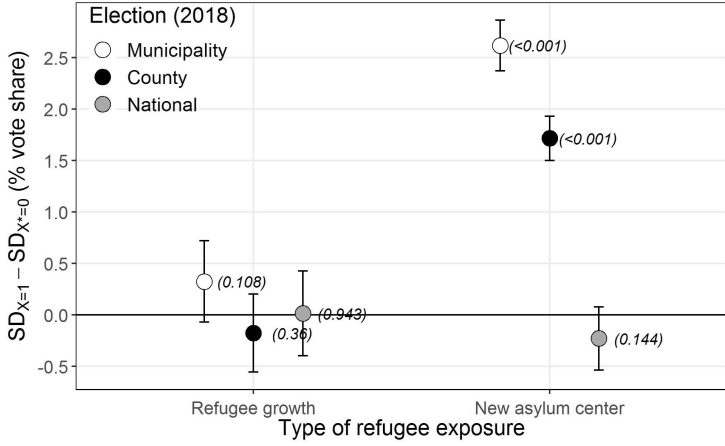


Figure 4.2: The importance of the type of refugee exposure for radical-right voting behavior. Differences in the estimated vote-share of the radical-right Sweden Democrat party (SD) in 2018 produced by OLS using the weights from the Synthetic Control Method. $X = 1$ indicates polling districts either receiving new refugees who self-selected into areas (left), or new asylum centers (right). $X^* = 0$ indicates the estimate calculated for the synthetic control from polling districts that neither experienced an increase in refugees presence or new asylum centers. There is one estimate per type of electoral outcome, with results displayed for municipal (white), county (black), and national (grey) elections. Bars indicate standard errors at the 95% level. P-values are in parentheses. The horizontal line indicates no effect.

The opposite is found for the models analyzing the impact of new asylum centers on radical-right support (right). In this case, the models yield a positive increase in the radical-right vote share when polling districts that receive new asylum centers are contrasted with the synthetic control of those that did not. More concretely, the estimates are positive for the elections conducted at the municipality- and county-level, for which increases are estimated to be around 2.5% and 1.8% respectively. These results are highly statistically significant. At the same time, the models predicting radical-right vote differences at the national election level seem not to yield significant differences in relation to the synthetic control. According to the models, then, it is only exposure to refugees through the placement of asylum centers that led to an increase in radical-right support in the 2018 elections, and only at

the municipal and country election level.

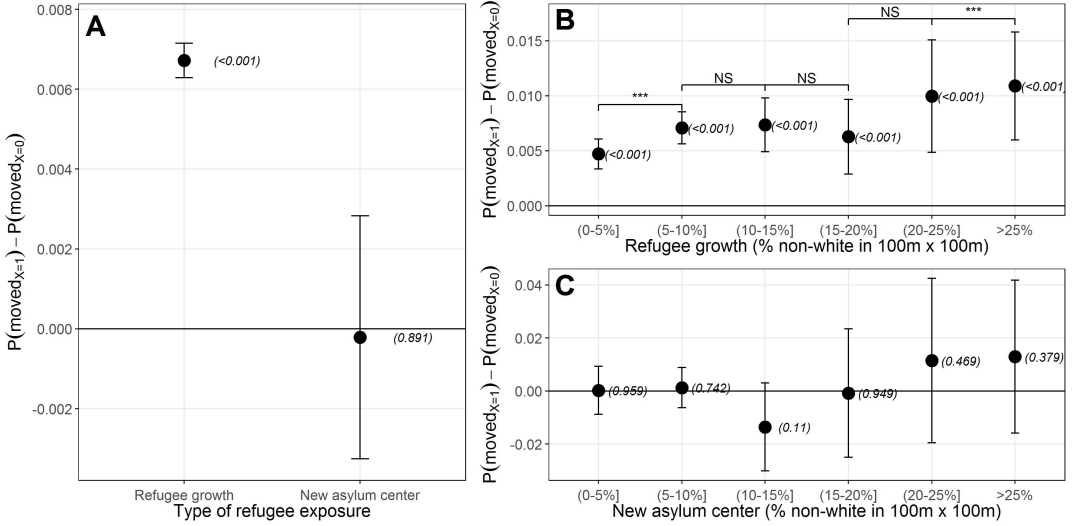


Figure 4.3: The importance of the type of refugee exposure for native out-mobility. **(A)** Differences in the probability of moving-out from the residential square estimated using OLS employing the weights from Coarsened Exact Matching. Estimates are presented for natives exposed to a proportional increase in refugees self-selecting into areas (left) and natives exposed to new asylum centers within the municipality (right) ($X = 1$), each compared to no change in the refugee population or in new asylum centers respectively ($X = 0$). **(B)** The effect of refugee growth across areas with different levels of ethnic minority presence. **(C)** The same as (B), for the effect of new asylum centers across areas with different levels of ethnic minority presence. Bars indicate standard errors at the 95% level. P-values are in parentheses. The horizontal line indicates no effect. Upper bars indicate significance tests between models. ‘***’ $p\text{-value} < 0.001$, ‘**’ $p\text{-value} < 0.01$, ‘*’ $p\text{-value} < 0.05$, ‘NS’ $p\text{-value} > 0.05$.

Figure 4.3 presents estimates similar to those in Figure 4.2, but now centering on natives’ residential mobility patterns.² As before, the models seem to diverge depending on the type of exposure at hand. Conversely, however, in this case the models show estimates in the opposite direction from those found in relation to voting. More concretely, as Figure 4.3 (A) shows, the only model that seems to indicate an increased probability of out-mobility is that which contrasts natives exposed to an increased presence of refugees with no change. This result is also highly statistically significant. Conversely, no effect is found for natives exposed to a new asylum center.

²I follow Clogg et al. (1995) to gauge the statistical significance between regression coefficients from non-nested models.

Subplots (B)-(C) in Figure 4.3 show the same effect across areas with different levels of ethnic minority presence, as indicated by the share of non-Westerners in the residential square prior to natives experiencing any new refugee exposure. Again, the model that consistently yields positive estimates is that focused on natives exposed to refugees self-selecting into their areas, in which the effect size seems to increase with the non-Westerner population share. In line with the literature, these models seem to yield statistically significant increases where the proportion of non-Westerners is greater than 5%, remaining more or less stable up to a 25% non-Westerners share. Beyond this point, the probability of moving-out seems to increase even more. In contrast, the models show consistent results indicating that among natives, being exposed to new asylum centers does not prompt out-mobility, even as the minority share in the area increases.

According to these models, then, natives seem to react differently depending on the behavioral outcome and the type of refugee exposure. Thus, natives were found to express increased support for radical-right parties following a change producing high out-group visibility, in the case of new asylum centers, but not following refugees self-selecting into native-based areas. At the same time, only increases in the presence of refugees who had self-selected into native-based areas seemed sufficient to produce an increase in the likelihood of native out-mobility, and this increase in the likelihood of out-mobility was particularly marked in areas that were already ethnically mixed.

4.5 Conclusion

The refugee crisis of 2015 represents an important exogenous event that has realigned discussions around the refugee issue in many Western European countries. In this third empirical chapter, I have shown that natives facing influxes of refugees near where they live may react in different ways depending on how they encounter the refugee population.

In particular, I have relied on Swedish register data for the whole of Swe-

den (2013-2017) to show the need to differentiate between (1) asylum seekers who find their own accommodation (displaying a “sparse-but-persistent” concentration of refugees) and (2) those who reside in asylum centers (displaying a “high-but-temporary” concentrations of refugees). This chapter has thus studied whether these two types of refugee exposure might have changed natives’ propensities to move from their home residences, and/or to support a radical-right party.

By relying on Coarsened Exact Matching to analyze natives’ residential mobility patterns, I have shown that only exposure to refugees who have chosen their own accommodation seems sufficient to prompt natives to move-out. This effect is stronger in areas where the presence of non-Westerners was greater prior to the influx of refugees. At the same time, the results show no increase in levels of out-mobility following the establishment of a new asylum center when compared to areas with no new asylum center, even in areas with an ethnically mixed population.

In addition, I used a Synthetic Control Method to analyze radical-right support at the polling-district level during the electoral contest of 2018. The analyses showed that levels of radical-right political support did not increase in polling districts that received influxes of refugees who had self-selected into those areas. Conversely, a positive increase in radical-right support was observed following the establishment of a new asylum center. This effect is observed for the elections at the municipal- and the county-level, but not at the national-level.

In conclusion, the results presented in this chapter suggest that the visibility of refugees and their spatial proximity to the native population partly determine how natives may react to increases in the refugee population in their local contexts. When their overall visibility in an area is high but for a limited period of time, there is sufficient reinforcement to produce new far-right support, but not to prompt native out-mobility. Conversely, when the visibility of an increased refugee presence is low, but the ethnic composition of an area has a higher likelihood of remaining mixed, for example, due to a greater non-Westerner presence prior to the increased refugee presence, new far-right support does not occur, but there is an increase in native

out-mobility.

This chapter is also subject to notable limitations. First, the data on the location of asylum centers at the municipal level should, ideally, be more precise in order to reduce the uncertainty introduced by the unknown spatial distance between the centers and the natives' residential locations. As a result of data limitations regarding the precise locations of these centers, new asylum centers might be located anywhere within the natives' municipality, a fairly large area that is highly likely to reduce out-group visibility for most natives. Thus, the analyses may have underestimated the effect of this modality of refugee exposure as a result of being unable to adjust for the distance to a new asylum center.

A second important limitation is that, at least in Sweden, asylum seekers awaiting the resolution of their asylum petition may also opt to seek their own accommodation. This may for example have been the case for many asylum seekers who were reunited with their families upon arrival. Our own data on refugees who had chosen their own accommodation was restricted to refugees who had already been granted permission to stay, introducing a bias that may also lead to an underestimation of their effect on natives' out-mobility.

In addition to these data constraints, I have not included any meaningful analysis of the media's role in making the visibility of asylum centers and refugees even more salient. The media has clearly played a determinative role in conditioning Swedish discourses on the integration issue, impacting native behavior in ways this study leaves unexplored.

Finally, as others have noted (Sobel, 2006), an intrinsic limitation of causal analyses of social interactions at the level of neighborhoods, particularly when these analyses are primarily based on observational data, is the almost certain violation of the so-called stable unit treatment value assignment-assumption. This is due to the existence of social networks, which can produce contacts between individuals exposed to new asylum seekers and others not exposed to this group, potentially confounding the exposure effect.

Future research, in addition to overcoming many of these limitations, could also improve on this study by investigating other meaningful ways in which natives who are discontented by an influx of refugees might express their discomfort. Voicing their opinions in the media or via social platforms, attending “anti-integration” demonstrations, or even moving their children from schools attended by high numbers of refugees, all constitute examples of other behavioral reactions that have not been explored here. In order to provide more clues on how to optimize the establishment of asylum centers, future research could also unravel the relevance of various different aspects, such as the asylum centers’ concrete shape, size, or the urban and demographic characteristics of the place in which they are established, to minimize the impact of new asylum centers on natives’ behavioral outcomes.

Appendix

A Chapter 2

	Difference (Raw)	Difference (Matched-LPM)	$P(\text{moved}_{X=1 Z})$	$P(\text{moved}_{X=0 Z})$	Std. Error	T-value	P-value
<i>Neighborhood definition</i>							
100m x 100m	0.03	0.003	0.071	0.067	0.0002	16.718	<0.001
300m x 300m	0.026	0.002	0.057	0.055	0.0002	8.377	<0.001
500m x 500m	0.026	0.001	0.054	0.053	0.0003	4.29	<0.001
census area	0.006	-0.002	0.066	0.068	0.0002	-8.912	<0.001
<i>Marginal neighborhood size</i>							
$\leq 100\text{m} \times 100\text{m} / \geq 300\text{m} \times 300\text{m}$	0.025	0.003	0.062	0.059	0.0006	4.844	<0.001
$\leq 300\text{m} \times 300\text{m} / \geq 500\text{m} \times 500\text{m}$	0.018	0.002	0.053	0.051	0.0008	2.808	<0.001
$\leq 500\text{m} \times 500\text{m} / \geq 500\text{m} \times 500\text{m}$	0.02	0.002	0.052	0.05	0.0011	1.969	0.05
<i>Percentage minority (100m x 100m)</i>							
(0-5%]		0	0.064	0.063	0.0006	0.745	0.46
(5-10%]		-0.001	0.071	0.071	0.0004	-1.416	0.16
(10-15%]		0.004	0.076	0.072	0.0005	7.613	<0.001
(15-20%]		0.005	0.074	0.069	0.0007	7.368	<0.001
(20-25%]		0.005	0.073	0.068	0.0009	5.195	<0.001
>25%		0.002	0.069	0.067	0.0007	2.711	0.01

Table A1: Statistical estimates per analytical setting. The differences between groups are the beta coefficients of the linear probability model applied on the full sample (Raw) and on the matched sample (matched-LPM) using ordinary-least squares. The models consist of the native out-mobility outcome (binary) for each exposure group being analyzed (also binary). Predicted probabilities, standard errors, t-values and p-values are gauged using the matched sample.

Coarsened exact matching

In this thesis, I use matching to ensure *positivity* when estimating the causal effect, i.e., to guarantee that each value of the covariates adjusted for has at least one observation that belongs to each of the treatment groups being analyzed (Hernán and Robins, 2020). More generally, matching are efficient tools to outline causal effects because they adjust for confounders while improving the degree of similarity between the treatment groups (Stuart, 2010). This reduces the dependence on the functional relationship assumed between the treatment variable and the outcome variable that is used to gauge the causal effect (the so-called model dependence, see Ho et al., 2007), and reduces the uncertainty of the causal estimate that results from groups being too dissimilar from one another on the observed covariates (Rosenbaum and Rubin, 1983; Deaton and Cartwright, 2018).

The matching method that I use is Coarsened Exact Matching (CEM) (Iacus et al., 2012). CEM is a non-parametric method that it is very simple to understand and use. In a nutshell, it starts by applying the binning function $H(X_j)$, which discretizes each covariate $H(X_j)$ separately into strata. This discretization procedure is essential, and it should be done following theoretical arguments (Iacus et al., *ibid.*). In the next step, the algorithm generates the Cartesian product of all discretized covariates, i.e., $H(X_1) \times \dots \times H(X_j)$, resulting in the multivariate cross-tabulation $H(\mathbf{X})$. Finally, CEM assumes that observations within each multidimensional stratum are equal except for the treatment they fall in, and applies a weight to each observation. Observations in each stratum that only belong to one group in the treatment variable (either treated or control) receive a weight of zero (i.e., are “pruned” from the data), while the remaining ones receive the following:

$$w_i = \begin{cases} 1, & i \in T^s \\ \frac{m_C}{m_T} \frac{m_T^S}{m_C^S}, & i \in C^s \end{cases} \quad (1)$$

Where m_C corresponds to the total number of control cases, and m_C^S to the number of control cases in stratum S . m_T and m_T^S refer to the treated counterparts.

After applying CEM, I apply a linear regression model using the ordinary-least squares method and the weights obtained in the CEM procedure with the binary outcome and a single predictor (another binary variable that indicates the treatment in which observations belong to and which varies according to the exposure groups being analyzed, see the manuscript for details). I used the *cem* package (v. 1.1.9) developed for the statistical package R by Iacus et al. (2009).

CEM performance

As recommended by Stuart (2010), any matching implementation should be followed by diagnostics indicating its performance, usually around the degree of similarity improved after the matching. More concretely, I measure the performance of CEM focusing on three indicators: (1) the standardized difference in means/proportions for each numerical/qualitative covariate before and after matching; (2) the so-called \mathcal{L}_1 measure for the overall multihistogram of covariates; and (3) the proportion of treated/control observations that are pruned. Each measure gives a different look at the data. The first one glances at the degree of similarity improvement separately for each covariate and, in case of qualitative covariates, for each category. This measure yields values close to zero when there is high similarity between groups along each particular covariate, while it yields values close to one when the groups are highly dissimilar. This is by far the most common diagnostic reported in matching studies, and several recommendations have been made for determining what values signify a good matching performance. Stuart (ibid.) generally recommends that a matching procedure has worked well and groups in the treatment variable are comparable when the standardized difference in means for each covariate is below 0.2. Similarly, Austin (2009) situates this threshold to be 0.1 in the case of applying propensity score matching. I therefore follow both criteria and set 0.2 as the maximal “dissimilarity” allowed for any given covariate.

Despite being the diagnostic most used to quantify the performance of a matching procedure, the difference in means for each covariate by definition overlooks how

groups are similar on the entire distribution of each covariate. To do this, another advantage of actually using CEM is the availability of the so-called \mathcal{L}_1 measure, which is defined as:

$$\mathcal{L}_1(f, g) = \frac{1}{2} \sum_{l_1 \dots l_j \in H(X)} |f_{l_1 \dots l_j} - g_{l_1 \dots l_j}| \quad (2)$$

Where $f_{l_1 \dots l_j}$ are the relative frequencies for observations belonging to the stratum with coordinates $l_1 \dots l_j$ of the multivariate cross-tabulation of the treated units, and $g_{l_1 \dots l_j}$ the same for control units. \mathcal{L}_1 outputs values between zero and one, with zero indicating complete overlap between the distributions of both groups (i.e., perfect similarity) and one otherwise. According to Iacus et al. (2012), then, any real improvement in the degree of similarity between the treatment groups should be indicated by a notable reduction in the \mathcal{L}_1 for the matched sample in comparison to the \mathcal{L}_1 for the sample before matching. Finally, the third indicator measures “the price” for the similarity improvement in terms of the proportion of pruned cases that are necessary to reduce the dissimilarity between the treatment groups. For instance, too much pruning could make the matched sample too different from the original population, which could be undesirable for the research. However, a clear advantage of the register data is that the large pool of available observations enables me to prune considerably while still remaining a great bulk of the observations, and thus without substantially changing the external validity of the analyses.

Figure A1 shows each of these three indicators for each exposure group analyzed in the main manuscript, one row per indicator. The first row shows the standardized difference in means/proportions for each numerical covariate/category before and after matching. The straight horizontal line is on 0.1, the conservative level stipulated by Austin. Each covariate is repeated for each of the three-year trial with which I split the range of years available (see the main manuscript for details).

To begin, the Figure shows how some covariates are greatly imbalanced before applying CEM, which makes matching a necessary step. The decreased lines for each covariate show that current CEM bins work well enough to get a standardized

difference below 0.1 across covariates, years, and exposure groups. Although some covariates seem to surpass this threshold for some of the years, they still remain below the threshold 0.2 recommended by Stuart.

The second row also shows the \mathcal{L}_1 before and after matching. Because this is a measure that takes the entire histogram into account, there is one value per three-year trial. As can be seen, groups were highly imbalanced before CEM is applied. The plots also show that the reduction of imbalance is rather remarkable, as Iacus et al. recommend it should look like. Although there is no direct threshold to compare, all the \mathcal{L}_1 values after CEM are very close to zero, indicating that the multihistogram between groups is highly similar.

Finally, the last row shows the proportion of each “control”/“treated” cases pruned. Treated cases (empty, black-enveloped bars) consist in natives exposed to a proportional increase in the minority presence, whereas control cases (grey, empty-enveloped bars) are, for the first four columns, those who are exposed to no change, and for the last three columns, natives exposed to a proportional increase in the minority presence in the outer side of the neighborhood (see main manuscript). As before, results are shown for each three-year trial, displaying the year when the covariates are measured. As can be seen, there is some pruning going on, which heightens almost up to 50% for most cases, and up to 80% in the five-hundred-by-five-hundred meter square model. For each pair of exposure groups, the cases that are mostly pruned are those belonging to the more numerous treatment group, which means that the level of pruning is rather adequate to reduce the dissimilarity between the treatment groups, and at the same time it is not too drastic to remove most of the cases.

In conclusion, the diagnostic measures discussed show a rather good improvement in the similarity of all pair of groups being analyzed, which improves the quality of the causal estimate as a result of pruning a reasonable number of observations that belong to the group with more cases.

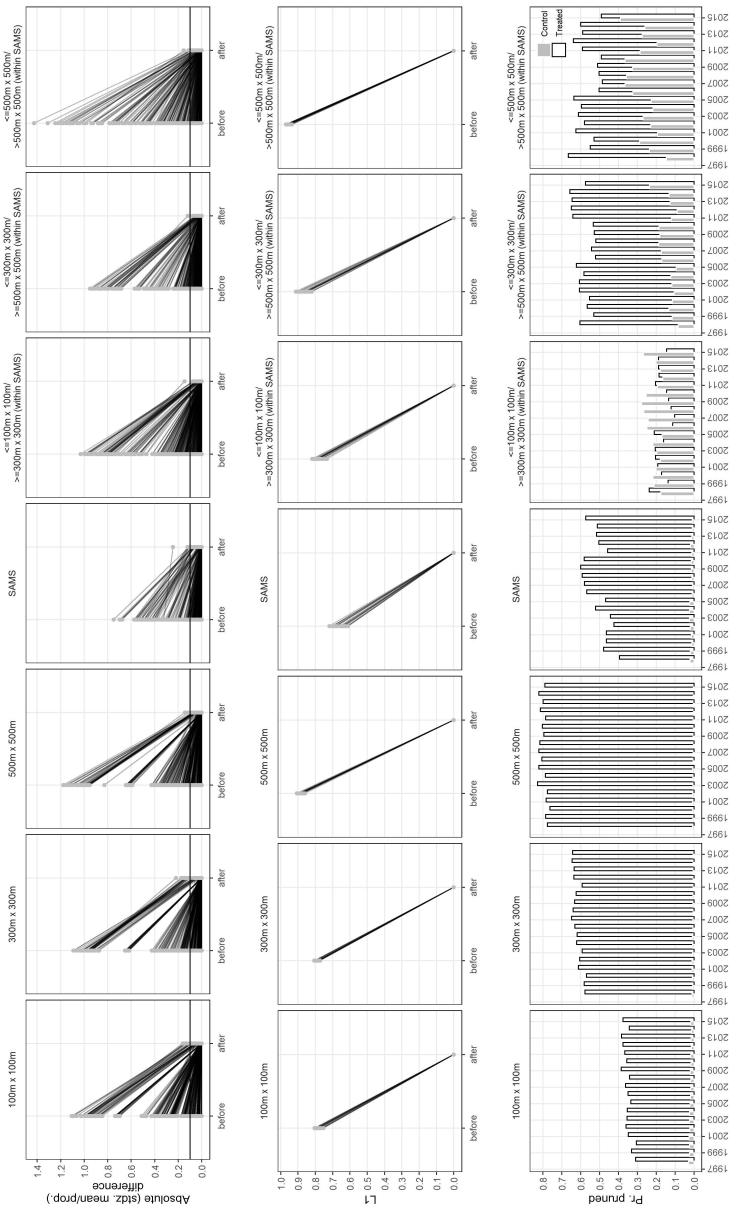


Figure A1: Performance of CEM as measured by three different indicators, one per row: (TOP) Standardized difference in means for numerical covariates, or difference in proportions for each category for qualitative covariates, for each year. Each value is shown before and after CEM, linking each covariate/category with a line. (MIDDLE) L_1 measure before and after CEM. (BOTTOM) Proportion of treated (empty, black-enveloped bars) and control (grey, empty-enveloped bars) cases pruned. The performance of CEM is checked for each pair of treatment groups and each three-year trial analyzed in the main manuscript, displaying the year corresponding to when covariates are measured (i.e., the first one in each trial).

The covariates used for improving the similarity between the treatment groups originally come from the Swedish register data, however I follow my own discretization for creating the bins used by the CEM procedure. They vary slightly depending on two factors: (1) the concrete treatment groups being matched; and (2) the year that matching is being applied for some numerical covariates. In general, the bins used have the following discretization:

- Civil status (qualitative): (1) single; (2) married and registered partners; (3) widow, divorced, or other.
- Family type (qualitative): (1) children under/above 18; (2) no children; (3) other.
- Tenure type (qualitative): (1) renter; (2) owner; (3) cooperative; (4) non-residential.
- Ethnicity (qualitative): (1) Swedish; (2) E.U.-15/U.S./Canada.
- Age (numeric): (1) 18-26; (2) 27-35; (3) 36-41; (4) 42-56; (5) 57-112.
- Years of education (numeric): (1) 5-12; (2) 13-20.
- Disposable income, logged (numeric): taking into account the decile, different for every year.
- Proportion of non-westerners in area (numeric): taking into account the decile, different for every year.
- Median disposable income in area (numeric): taking into account the five quantiles, different for every year.
- Number of individuals living in each area (numeric): taking into account the five quantiles, different for every year.

Agent-based model

Agent-based modelling (ABM) is a type of simulation methodology that consists in programming rules to some artificial agents (Macy and Willer, 2002). Although these agents can be anything in principle, in sociological studies agents represent persons and the rules their behavioral rules, with the objective to let those agents interact and observe the macro-properties that emerge from these interactions. At least since Schelling (1971), ABM has been shown to be a crucial tool to study social

interdependency, or the ability of agents to interact with other agents directly or through their respective social environments. ABMs have been proved an essential theoretical tool for sociologists (Ylikoski, 2014; Bianchi and Squazzoni, 2015) especially for its use to derive conclusions about macro-phenomena from a set of posited behavioral and structural axioms, and also for being a high versatile tool that allows the researcher to use it in combination with other common methods of the social sciences (Chattoe-Brown, 2013).

Because the ABM in this paper consists of only one part of the entire methodology, the decision has been to keep the model as simple as possible. To achieve that, I follow Laurie and Jaggi (2003) and extend directly from the well-known ‘spatial proximity model’ (Schelling, 1971), a landmark model in the history of ABMs in ethnic residential segregation. Thus, many of the key characteristics are from that model:

- The world has the checkerboard structure with two dimensions (98×98 pixels). Each pixel is filled with one house and it is inhabited by at most one agent.
- Agents belong to one of two groups, whose share varies depending on the experimental condition (continuously from 50%-50% to 90%-10%).
- The percentage of vacancies (i.e., houses not filled by any agent) is always 10% across experimental conditions.
- Agents can only do either one of two things: staying where they are, or moving to a random place.
- Agents decide to move only if the share of other co-ethnics within the exposure area is below 50
- When the simulation begins, agents are randomly scattered in the world (see Winship, 1977).

The only meaningful differences across the simulations are (1) the size of the area of exposure the agents use to compute the ethnic composition, which I split into self-defined neighborhoods of different radii, from 2 to 20, and (2) the presence of the ethnic minority in the system, which can range from 50%-50% majority-minority

composition to a 90%-10% composition. To get a sense of the sizes produced by the radii used by the agents to compute the ethnic composition, Figure A2 shows a schematic picture of these areas. I implemented the model manually using NetLogo (v. 6.1.1) (Wilensky and Rand, 2015).

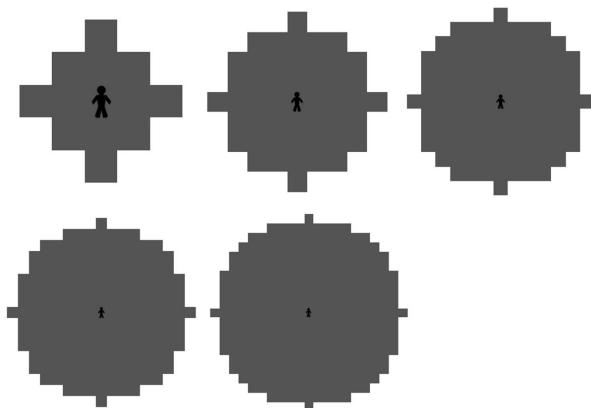


Figure A2: Different radii employed in the agent-based model. (TOP) Overlapping areas of radii 2, 4, and 6. (BOTTOM) Overlapping areas of radii 8 and 10.

B Chapter 3

	Difference (Raw)	Difference (CEM-LPM)	$P(\text{moved}_{X=1} \mathbf{Z})$	$P(\text{moved}_{X=0} \mathbf{Z})$	Std. Error	T-value	P-value
Out-movers/no change	0.008	0.004	0.056	0.052	0.0003	17.51	<0.001
<i>Marginal distance to previous out-movers</i>							
$\leq 100\text{m} \times 100\text{m} / \geq 200\text{m} \times 200\text{m}$	0.012	0.01	0.059	0.05	0.0012	8.19	<0.001
$\leq 200\text{m} \times 200\text{m} / \geq 300\text{m} \times 300\text{m}$	0.007	0.003	0.054	0.051	0.0015	2.24	0.03
$\leq 300\text{m} \times 300\text{m} / \geq 400\text{m} \times 400\text{m}$	0.008	0.004	0.052	0.048	0.0021	1.72	0.09
<i>N previous out-movers</i>							
$\geq 2 / < 2$	0.009	0.004	0.058	0.054	0.0004	10.75	<0.001
$\geq 3 / < 3$	0.013	0.006	0.061	0.055	0.0006	9.57	<0.001
$\geq 4 / < 4$	0.016	0.006	0.062	0.056	0.0011	5.67	<0.001
<i>N inhabitants (in 100m x 100m)</i>							
≤ 15	0.01	0.008			0.0003	23.22	<0.001
[16-30]	0.004	0.003			0.0004	6.28	<0.001
> 30	0.006	0.001			0.0009	1.08	0.28
<i>Per. minority (in 100m x 100m)</i>							
0%	0.009	0.007			0.0003	19.85	<0.001
(0-10%]	0.005	0.002			0.0006	3.2	<0.001
$> 10\%$	0.008	0.004			0.0006	6.67	<0.001

Table A2: Statistical estimates per analytical setting. The differences between groups are the beta coefficients of the linear probability model applied on the full sample (Raw) and on the matched sample (CEM-LPM) using ordinary-least squares. The models consist of the native out-mobility outcome (binary) for the exposure group being analyzed (also binary). Predicted probabilities, standard errors, t-values and p-values are gauged using the matched sample.

CEM performance

As this second empirical study also relies on CEM, we evaluate CEM’s performance using the same measures and criteria explained in the first empirical study. Figure A3 shows each of these measures for each pair of treatment groups analyzed in the main manuscript, one row per measure. The first row shows the standardized difference in means/proportions for each numerical covariate/category before the matching and after. The straight horizontal line is on 0.1, the conservative level stipulated by Austin (2009). Each covariate is repeated for each of the three consecutive years within which we apply the dynamic matching (see the manuscript for details). To begin, the plot shows how some covariates are greatly imbalanced before applying CEM, which makes matching a necessary step for the empirical analyses of this chapter as well. The decreased lines for each covariate show that current CEM bins work well enough to get a standardized difference below 0.1 across covariates, years, and pairs of treatment groups. Although some covariates seem to surpass this threshold for some of the years, they still remain below the threshold 0.2 recommended by Stuart (2010).

The second row shows measures of the \mathcal{L}_1 also before and after matching. Because this is a measure that takes the entire multihistogram into account, there is one measure per each three-year trial. As can be seen, the reduction of imbalance is rather remarkable, as Iacus et al. (2012) recommend it should be. All the \mathcal{L}_1 values after CEM are very close to zero, indicating that the multihistogram between groups is highly similar for all groups.

Finally, the last row shows the proportion of each “control” (grey, empty-enveloped bars) and “treated” (empty, black-enveloped bars) cases pruned, which varies across settings (see the main manuscript). As before, results are offered by each three-year trial, displaying the year when the covariates are measured. As can be seen, there is some pruning going on, which heightens almost up to 50% for most cases, and up to 70% in the setting where natives are either exposed to 4-or-more out-movers or less. For each pair of treatment groups being analyzed, the cases

that are mostly pruned are those belonging to the more numerous treatment group, which means that the level of pruning is rather adequate to reduce the dissimilarity between treatment groups, and at the same time it is not too drastic to go in detriment of the external validity of the estimates.

In conclusion, the diagnostic measures discussed show a rather good improvement in the balance for each treatment group, which is to a great extent granted by the census information and the high granularity that the Swedish register data offer.

As in the previous chapter, the bins used in CEM vary slightly depending on two factors: (1) the concrete pair of treatment groups; and (2) the year that matching is being applied for some numerical covariates. In general, however, the bins used have the following discretization (the original discretization is either from the Swedish registers, or it is computed using information provided in the data):

- Civil status (qualitative): (1) single; (2) married and registered partners; (3) widow, divorced, or other.
- Family type (qualitative): (1) children under/above 18; (2) no children; (3) other.
- Tenure type (qualitative): (1) renter; (2) owner/cooperative; (3) non-residential.
- Ethnicity (qualitative): (1) Swedish; (2) E.U.-15/U.S./Canada.
- Age (numeric): (1) 18-26; (2) 27-35; (3) 36-41; (4) 42-56; (5) 57-112.
- Years of education (numeric): (1) 5-12; (2) 13-20.
- Disposable income, logged (numeric): taking into account the deciles, different for every year.
- Proportion of non-westerners in area (numeric): taking into account the deciles, different for every year.
- Median disposable income in area (numeric): taking into account the quintiles, different for every year.
- Number of individuals living in each area (numeric): taking into account the deciles, different for every year.

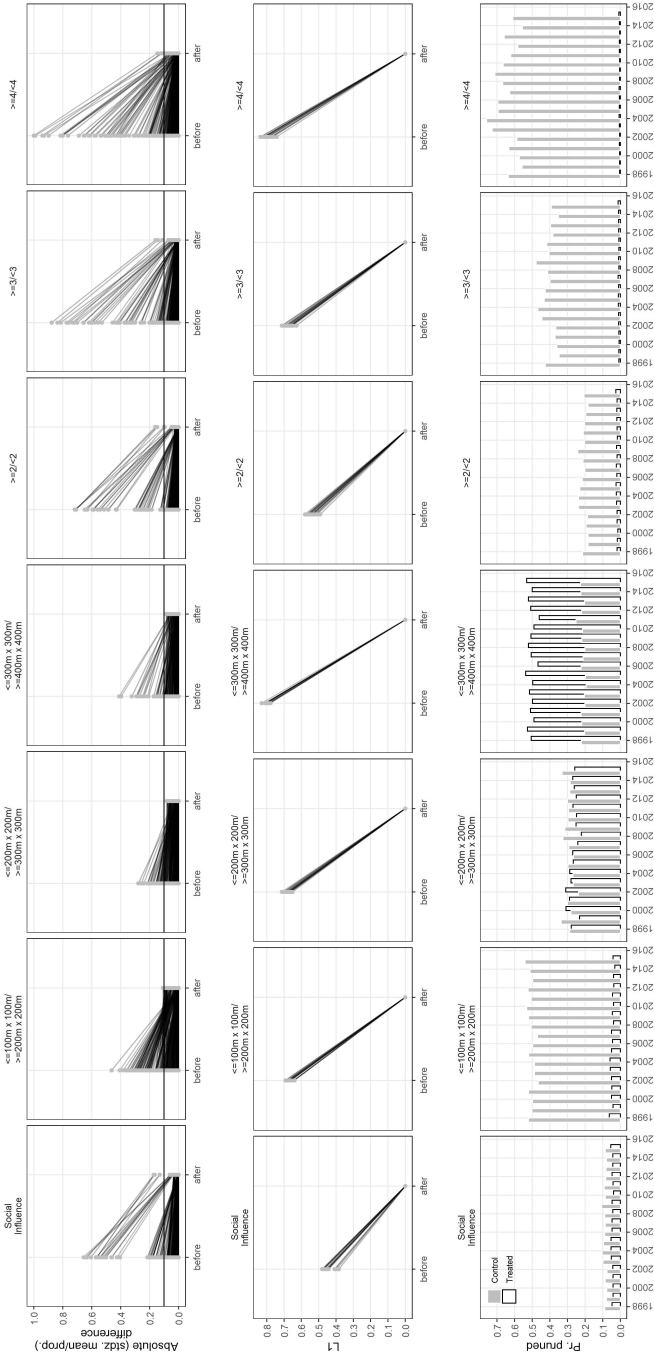


Figure A3: Performance of CEM as measured by three different indicators, one per row: (*TOP*) Standardized difference in means for numerical covariates, or difference in proportions for each category for qualitative covariates, for each year. Each value is shown before and after CEM, linking each covariate/category with a line. (*MIDDLE*) L_1 measure before and after CEM. (*BOTTOM*) Proportion of treated (empty, black-enveloped bars) and control (grey, empty-enveloped bars) cases pruned. The performance of CEM is checked for each pair of treatment groups and each three-year trial analyzed in the main manuscript, displaying the year corresponding to when covariates are measured (i.e., the first one in each trial).

Robustness checks: Family type

A potential problem with our analyses is that our analytical setting for capturing social influence effects in residential mobility may also capture other processes related with the family structure that follow the same behavioral pattern studied here but which not obey to social influence dynamics. Divorced couples that move at different rates from a common dwelling, or siblings moving out one after another as they enter the university, for instance, are processes that involve being exposed to a co-ethnic leaving, although not a neighbor, and leaving thereafter.

In the paper, family type is a covariate we adjust for, which means that we are partly dealing with this issue already. However, the family type in the cases described above changes *after* natives are “exposed to a co-ethnic moving-out.” The problem, however, is that we cannot adjust for changes in the family type after we measure the treatment assignment. This is because we would be adjusting for a post-treatment variable, which would completely invalidate our estimate of the treatment effect since it would most likely open spurious correlations that are not present when we do not condition on that post-treatment variable (see Pearl, 2010).

In spite of this situation, it might still be useful to re-do the analyses while focusing on a subsample of natives who remain their family type *unchanged* along the three-year trial in which the native’s covariates, exposure group, and mobility outcome are recorded. Table A3 shows these results while sharing the structure of Table A2 for comparability reasons. As can be seen, the estimates reduce their size in comparison to those seen in Table A2, but their substantive direction and standard errors remain largely the same.

	Difference (Raw)	Difference (CEM-LPM)	$P(\text{moved}_{X=1} \mathbf{Z})$	$P(\text{moved}_{X=0} \mathbf{Z})$	Std. Error	T-value	P-value
Out-movers/no change	0.004	0.003	0.033	0.03	0.0002	11.73	<0.001
<i>Marginal distance to previous out-movers</i>							
$\leq 100\text{m} \times 100\text{m} / \geq 200\text{m} \times 200\text{m}$	0.010	0.006	0.033	0.028	0.001	5.46	<0.001
$\leq 200\text{m} \times 200\text{m} / \geq 300\text{m} \times 300\text{m}$	0.004	0.003	0.031	0.029	0.0013	2.09	0.04
$\leq 300\text{m} \times 300\text{m} / \geq 400\text{m} \times 400\text{m}$	0.005	-0.001	0.027	0.027	0.0017	-0.3	0.77
<i>N previous out-movers</i>							
$\geq 2 / < 2$	0.006	0.003	0.034	0.031	0.0003	8.18	<0.001
$\geq 3 / < 3$	0.010	0.006	0.037	0.032	0.0006	9.97	<0.001
$\geq 4 / < 4$	0.014	0.006	0.039	0.033	0.001	6.48	<0.001
<i>N inhabitants (in 100m x 100m)</i>							
≤ 15	0.005	0.005			0.0003	16.78	<0.001
[16-30]	0.003	0.002			0.0004	4.85	<0.001
> 30	0.004	0			0.0008	-0.31	0.75
<i>Per. minority (in 100m x 100m)</i>							
0%	0.004	0.004			0.0003	14.82	<0.001
(0-10%]	0.003	0			0.0005	0.73	0.46
$> 10\%$	0.005	0.002			0.0005	4.74	<0.001

Table A3: Statistical estimates per analytical setting. The differences between groups are the beta coefficients of the linear probability model applied on the full sample (Raw) and on the matched sample (CEM-LPM) using ordinary-least squares. The models consist of the native out-mobility outcome (binary) for the exposure group being analyzed (also binary). Predicted probabilities, standard errors, t-values and p-values are gauged using the matched sample.

C Chapter 4

Type of refugee exposure	Sample	Outcome	Estimate	Std. Error	T-value	P-value
<i>Exiting</i>						
Refugee growth	All sample	Out-moving	0.0067	2e-04	30.601	(<0.001)
	(0-5%]		0.0047	7e-04	6.8001	(<0.001)
	(5-10%]		0.0071	7e-04	9.5116	(<0.001)
	(10-15%]		0.0074	0.0012	5.9041	(<0.001)
	(15-20%]		0.0063	0.0017	3.6312	(<0.001)
	(20-25%]		0.01	0.0026	3.8272	(<0.001)
	>25%		0.0109	0.0025	4.3521	(<0.001)
	New asylum center		All sample	-2e-04	0.0016	-0.1371
(0-5%]		2e-04	0.0046	0.0513	(0.96)	
(5-10%]		0.0013	0.0039	0.3287	(0.74)	
(10-15%]		-0.0135	0.0085	-1.5999	(0.11)	
(15-20%]		-8e-04	0.0123	-0.0638	(0.95)	
(20-25%]		0.0115	0.0158	0.7245	(0.47)	
>25%		0.0129	0.0147	0.8801	(0.38)	
<i>Voicing</i>						
Refugee growth	All sample	SD pr. vote share (muni.)	0.0032	0.002		(0.11)
	All sample	SD pr. vote share (county)	-0.0018	0.0019		(0.36)
	All sample	SD pr. vote share (nat.)	2e-04	0.0021		(0.94)
New asylum center	All sample	SD pr. vote share (muni.)	0.0262	0.0013		(<0.001)
	All sample	SD pr. vote share (county)	0.0172	0.0011		(<0.001)
	All sample	SD pr. vote share (nat.)	-0.0023	0.0016		(0.14)

Table A4: Statistical estimates from models. The differences between groups are the beta coefficients of the linear probability model using the ordinary-least squares method after applying coarsened exact matching (Out-moving, a binary) and a synthetic control method (SD pr. vote share, numeric). The former analyzes a sample of natives, while the latter analyzes a sample of polling districts.

Diagnostics

I mainly evaluate the performance of Coarsened Exact Matching (CEM) by the so-called standardized difference in means of each covariate before and for each treatment group being analyzed in the main manuscript before and after applying CEM. As recommended by Stuart (2010), the maximum standardized difference allowed is 0.2, although others recommend that groups are only sufficiently similar if the standardized difference for any covariate is below 0.1 (Austin, 2009). On the side of the Synthetic Control Method (SCM), the main recommendation is that the mean trajectory of each covariate of the synthetic control is sufficiently close (i.e., not exact) to the mean trajectory of each covariate of the treated group (Abadie et al., 2010). In addition to the *cem* package (Iacus et al., 2009), I have also employed the *microsynth* package (v. 2.0.31) developed for the statistical package R by Robbins et al. (2017).

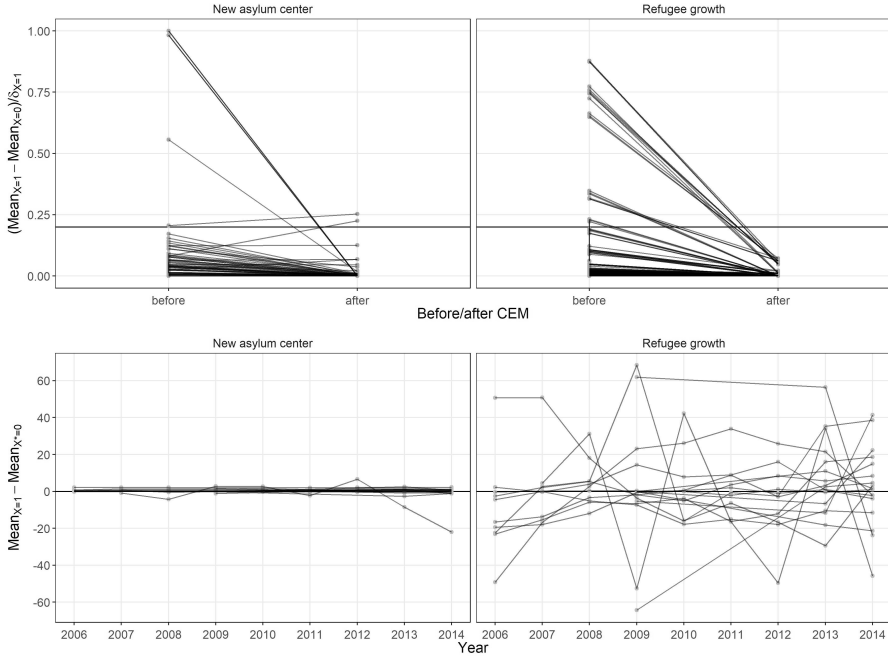


Figure A4: Diagnostics for CEM and SCM. (*TOP*) Standardized difference in means of each covariate before and after CEM. Each covariate is repeated four times, one per each time CEM is applied through the period 2012-2015. $X = 1$ indicates a native exposed to a new asylum center/refugee growth in the $100\text{m} \times 100\text{m}$ residential square, whereas $X = 0$ indicates the absence of such a new asylum center/no refugee growth. The dashed horizontal line on 0.2 indicate the threshold commonly used to assess the sufficiently appropriate similarity between the treatment groups and which should not be surpassed (Stuart, 2010). (*BOTTOM*) Difference in the mean trajectory of each covariate of the treated polling districts and the one from the synthetic control for the period before the treatment assignment occurs (2006-2014). $X = 1$ indicates a polling district exposed to a new asylum center/refugee growth in the area, whereas $X^* = 0$ indicates the synthetic control. The dashed line on zero indicates where the trajectory of groups are exactly the same.

I show the diagnostics for CEM and SCM in graphical form in Figure A4. The top of the Figure shows the standardized difference in means for each covariate before and after CEM. The number of covariates shown appears every time that matching is applied for the period 2012-2015. The dashed line on 0.2 indicates the threshold of maximal dissimilarity allowed stipulated by Stuart (2010). As can be seen, most covariates are below the threshold of 0.1, and those that surpass it do not go beyond 0.2. Some covariates slightly surpass the threshold to reach a value of 0.25, although this is only occurs for a two covariates in total. This indeed is

an indication that groups exposed to an asylum center/refugee growth or not are sufficiently similar. The so-called \mathcal{L}_1 (Iacus et al., 2012), which looks at the overall multihistogram of the covariates, also shows high similarity between groups and covariates. This performance is achieved partly thanks to the availability of the registers for the whole population of Sweden, which consequently allows CEM to prune more than 90% of observations not exposed to the center. (Both the \mathcal{L}_1 and the percentage of cases pruned are available in tabular form upon request.)

On the bottom side of the same Figure, there is plotted the difference between the mean covariate value of the treated polling districts and the mean value of the same covariate of the synthetic control, one per each type of exposure to refugees. The plots show that SCM is highly successful in finding weights that highly resemble the average value of the true treated polling districts, although the similarity seems to be best in the analysis of polling districts being exposed (or not) to a new asylum center.

In conclusion, the diagnostic measures discussed show a rather good improvement in the balance for the matching to improve balance and successfully resemble the mean trajectory of treated poll districts.

Finally, I show the bins used for CEM. I follow standard recommendations and use my own bin discretizations (Iacus et al., 2012):

- Civil status (qualitative): (1) single; (2) married and registered partners; (3) widow, divorced, or other.
- Family type (qualitative): (1) children under/above 18; (2) no children; (3) other.
- Tenure type (qualitative): (1) renter; (2) owner/cooperative; (3) non-residential.
- Ethnicity (qualitative): (1) Swedish; (2) E.U.-15/U.S./Canada.
- Asylum center (qualitative): (1) 1 (indicating existence); (2) 0 (indicating absence).
- Age (numeric): (1) 18-26; (2) 27-35; (3) 36-41; (4) 42-56; (5) 57-112.
- Years of education (numeric): (1) 5-12; (2) 13-20.
- Proportion of refugees (numeric): (1) 0-0.025; (2) 0.025-1.

- Disposable income, logged (numeric): taking into account the decile, different for every year.
- Proportion of non-westerners in area (numeric): taking into account the decile, different for every year.
- Median disposable income in area (numeric): taking into account the decile, different for every year.
- Number of individuals living in each area (numeric): taking into account the decile, different for every year.

Bibliography

- Abadie, A. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(3):113–132.
- Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s Tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.
- Abascal, M. and Baldassarri, D. (2015). Love thy neighbor? Ethnoracial diversity and trust reexamined. *American Journal of Sociology*, 121(3):722–782.
- Ahmed, A. M. and Hammarstedt, M. (2008). Discrimination in the rental housing market: A field experiment on the Internet. *Journal of Urban Economics*, 64(2):362–372.
- Akers, R. L., Krohn, M. D., Lanza-Kaduce, L., and Radosevich, M. (1979). Social learning and deviant behavior: A specific test of a general theory. *American Sociological Review*, 44(4):636.
- Aldén, L., Hammarstedt, M., and Neuman, E. (2015). Ethnic segregation, tipping behavior, and native residential mobility. *International Migration Review*, 49(1):36–69.
- Aldrich, H. (1975). Ecological succession in racially changing neighborhoods. *Urban Affairs Review*, 10(3):327–348.
- Allport, G. W. (1954). *The Nature of Prejudice*. Addison-Wesley.
- Anderson, P. (1972). More is different. *Science*, 177(4047):393–396.
- Andersson, R. (2013). Reproducing and reshaping ethnic residential segregation in stockholm: The role of selective migration moves. *Geografiska Annaler, Series B: Human Geography*, 95(2):163–187.
- Andersson, R., Musterd, S., and Galster, G. (2019). Port-of-entry neighborhood and its effects on the economic success of refugees in Sweden. *International Migration Review*, 53(3):671–705.

- Aral, S., Muchnik, L., and Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549.
- Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter, O. and Rees, A., editors, *Discrimination in Labor Markets*, pages 3–33. Princeton University Press.
- Arvidsson, M. and de la Prada, A. G. (forthcoming). Causal inference for social processes. *Forthcoming*.
- Arzheimer, K. (2009). Contextual factors and the extreme right vote in Western Europe, 1980-2002. *American Journal of Political Science*, 53(2):259–275.
- Arzheimer, K. and Carter, E. (2006). Political opportunity structures and right-wing extremist party success. *European Journal of Political Research*, 45(3):419–443.
- Asch, S. (1955). Opinions and social pressure. *Scientific American*, 193(5):31 – 35.
- Austin, P. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched. *Statistics in Medicine*, 28:3083–3107.
- Axtell, R. (2000). Why agents? On the varied motivations for agent computing in the social sciences. *Center on Social and Economics Dynamics Working Paper*, 17:1–23.
- Bader, M. D. and Krysan, M. (2015). Community attraction and avoidance in Chicago: What’s race got to do with it? *Annals of the American Academy of Political and Social Science*, 660(1):261–281.
- Bernhard, L. and Kaufmann, D. (2018). Coping with the asylum challenge: Tightening and streamlining policies in Western Europe. *Journal of Ethnic and Migration Studies*, 44(15):2506–2523.
- Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. In Duflo, E. and Banerjee, A. V., editors, *Handbook of Field Experiments*, chapter 8, pages 309–393. North Holland.
- Bianchi, F. and Squazzoni, F. (2015). Agent-based models in sociology. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(4):284–306.
- Björkegren, E. (2018). Neighborhoods and youth health: Everybody needs good neighbors? *IFAU Working Paper*.
- Black, R. (2001). Fifty years of refugee studies: From theory to policy. *International Migration Review*, 35(1):57–78.

- Blalock, H. (1957). Per cent non-white and discrimination in the South. *American Sociological Review*, 22(6):677–682.
- Booth, A. L., Leigh, A., and Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics*, 74(4):547–573.
- Boschman, S. and van Ham, M. (2015). Neighbourhood selection of non-Western ethnic minorities: Testing the own-group effects hypothesis using a conditional logit model. *Environment and Planning A: Economy and Space*, 47(5):1155–1174.
- Bråmås, Å. (2006). ‘White flight’? The production and reproduction of immigrant concentration areas in Swedish cities, 1990–2000. *Urban Studies*, 43(7):1127–1146.
- Bråmås, Å. (2008). Dynamics of ethnic residential segregation in Göteborg, Sweden, 1995–2000. *Population, Space and Place*, 14(2):101–117.
- Bruch, E. and Atwell, J. (2015). Agent-based models in empirical social research. *Sociological Methods and Research*, 44(2):186–221.
- Bruch, E. and Feinberg, F. (2017). Decision-making processes in social contexts. *Annual Review of Sociology*, 43:207–227.
- Bruch, E. and Mare, R. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology*, 112(3):667–709.
- Bruch, E. and Swait, J. (2019). Choice set formation in residential mobility and its implications for segregation dynamics. *Demography*, 56:1665–1692.
- Bryant, J. and Oliver, M. B. (2009). *Media effects: Advances in theory and research*. Routledge.
- Campesi, G. (2018). Between containment, confinement and dispersal: the evolution of the Italian reception system before and after the ‘refugee crisis’. *Journal of Modern Italian Studies*, 23(4):490–506.
- Card, D., Mas, A., and Rothstein, J. (2007). Tipping and the dynamics of segregation. *NBER Working Paper Series*, (13052).
- Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197.
- Centola, D. and Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, 112(7):1989–1994.

- Centola, D. and Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734.
- Charles, C. Z. (2003). The dynamics of racial residential segregation. *Annual Review of Sociology*, 29:167–207.
- Charreire, H., Feuillet, T., Roda, C., Mackenbach, J. D., Compernelle, S., Glonti, K., Bárdos, H., Le Vaillant, M., Rutter, H., Mckee, M., De Bourdeaudhuij, I., Brug, J., Lakerveld, J., and Oppert, J. M. (2016). Self-defined residential neighbourhoods: Size variations and correlates across five European urban regions. *Obesity Reviews*, 17(Suppl. 1):9–18.
- Chattoe-Brown, E. (2013). Why sociology should use agent based modelling. *Sociological Research Online*, 18(3):1–28.
- Chetty, R., Hendren, N., and Katz, L. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, 106(4):855–902.
- Cialdini, R. and Goldstein, N. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621.
- Clark, W. (1991). Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model. *Demography*, 28(1):1–19.
- Clark, W. (2002). Ethnic preferences and ethnic perceptions in multi-ethnic settings. *Urban Geography*, 23(3):237–256.
- Clark, W. (2009). Changing residential preferences across income, education, and age: Findings from the multi-city study of Urban inequality. *Urban Affairs Review*, 44(3):334–355.
- Clark, W. and Huang, Y. (2003). The life course and residential mobility in British housing markets. *Environment and Planning A: Economy and Space*, 35(2):323–339.
- Clark, W. and Onaka, J. L. (1985). An empirical test of a joint model of residential mobility and housing choice. *Environment and Planning A: Economy and Space*, 17(7):915–930.
- Clark, W. and Withers, S. (2007). Family migration and mobility sequences in the United States Spatial mobility in the context of the life course. *Demographic Research*, 17:591–622.
- Clark, W. A. and Dieleman, F. M. (1996). *Households and housing: Choice and outcomes in the housing market*. Center for Urban Policy Research.

- Clogg, C., Petkova, E., and Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100(5):1261–1293.
- Constant, A. and Massey, D. S. (2005). Labor market segmentation and the earnings of German guestworkers. *Population Research and Policy Review*, 24:489–512.
- Coulton, C., Jennings, M. Z., and Chan, T. (2013). How big is my neighborhood? Individual and contextual effects on perceptions of neighborhood scale. *American Journal of Community Psychology*, 51(1-2):140–150.
- Craig, M. A. and Richeson, J. A. (2014). More diverse yet less tolerant? How the increasingly diverse racial landscape affects white Americans’ racial attitudes. *Personality and Social Psychology Bulletin*, 40(6):750–761.
- Crowder, K. (2000). The racial context of white mobility: An individual-level assessment of the white flight hypothesis. *Social Science Research*, 29(2):223–257.
- Crowder, K. and Downey, L. (2011). Interneighborhood migration, race, and environmental hazards: Modeling microlevel processes of environmental inequality. *American Journal of Sociology*, 115(4):1110–1149.
- Crowder, K., Hall, M., and Tolnay, S. E. (2011). Neighborhood immigration and native out-migration. *American Sociological Review*, 76(1):25–47.
- Crowder, K. and Krysan, M. (2016). Moving beyond the big three: A call for new approaches to studying racial residential segregation. *City and Community*, 15(1):18–22.
- Crowder, K., Pais, J., and South, S. J. (2012). Neighborhood diversity, metropolitan constraints, and household migration. *American Sociological Review*, 77(3):325–353.
- Crowder, K. and South, S. (2008). Spatial dynamics of white flight: The effects of local and extralocal racial conditions on neighborhood out-migration. *American Sociological Review*, 73(5):792–812.
- Crowder, K., South, S., and Chavez, E. (2006). Wealth, race, and inter-neighborhood migration. *American Sociological Review*, 71:72–94.
- Dancygier, R. M. and Laitin, D. D. (2014). Immigration into Europe: Economic discrimination, violence, and public policy. *Annual Review of Political Science*, 17:43–64.
- Deaton, A. and Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, 210:2–21.
- DeLuca, S., Wood, H., and Rosenblatt, P. (2019). Why poor families move (and where they go): Reactive mobility and residential decisions. *City and Community*, 18(2):556–593.

- DiMaggio, P. J. and Powell, W. W. (1983). The Iron Cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2):147.
- Dinas, E., Matakos, K., Xefteris, D., and Hangartner, D. (2019). Waking up the Golden Dawn: Does exposure to the refugee crisis increase support for extreme-right parties? *Political Analysis*, 27(2):244–254.
- Dinesen, P. T. and Sønderskov, K. M. (2015). Ethnic diversity and social trust: Evidence from the micro-context. *American Sociological Review*, 80(3):550–573.
- Downey, L. (2006). Environmental racial inequality in Detroit. *Social Forces*, 85(2):771–796.
- Dröes, M. and Koster, H. (2019). A world divided: Refugee centers, house prices, and household preferences. *Tinbergen Institute Discussion Paper*, TI 2019-075/VIII.
- Drouhot, L. G. and Nee, V. (2019). Assimilation and the Second Generation in Europe and America: Blending and segregating social dynamics between immigrants and natives. *Annual Review of Sociology*, 45:177–199.
- Dunning, T. (2008). Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly*, 61(2):282–293.
- Dustmann, C., Fasani, F., Frattini, T., Minale, L., and Schönberg, U. (2017). On the economics and politics of refugee migration. *Economic Policy*, 32(91):497–550.
- Dustmann, C., Vasiljeva, K., and Piil Damm, A. (2019). Refugee migration and electoral outcomes. *The Review of Economic Studies*, 86(5):2035–2091.
- Elisabeth, I. (2005). The vulnerable populist right parties: No economic realignment fuelling their electoral success. *European Journal of Political Research*, 44(3):465–492.
- Elster, J. (1982). Marxism, functionalism, and game theory. *Theory and Society*, 11:453–482.
- Elwert, F. and Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40:31–53.
- Entwisle, B. (2007). Putting people into place. *Demography*, 44(4):687–703.
- Evans, J. (2005). The dynamics of social change in radical right-wing populist party support. *Comparative European Politics*, 3(1):76–101.
- Felps, W., Mitchell, T. R., Hekman, D. R., Lee, T. W., Holtom, B. C., and Harman, W. S. (2009). Turnover contagion: How coworkers’ job embeddedness and job search behaviors influence quitting. *Academy of Management Journal*, 52(3):545–561.

- FitzGerald, D. S. and Arar, R. (2018). The sociology of refugee migration. *Annual Review of Sociology*, 44(1):387–406.
- Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., and Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4):2.
- Fossett, M. and Waren, W. (2005). Overlooked implications of ethnic preferences for residential segregation in agent-based models. *Urban Studies*, 42(11):1893–1917.
- Fowler, C. (2016). Segregation as a multiscalar phenomenon and its implications for neighborhood-scale research: The case of South Seattle 1990-2010. *Urban Geography*, 37(1):1–25.
- Fowler, C., Lee, B., and Matthews, S. (2016). The contributions of places to metropolitan ethn racial diversity and segregation: Decomposing change across space and time. *Demography*, 53(6):1955–1977.
- Galster, G. (2001). On the nature of neighbourhood. *Urban Studies*, 38(12):2111–2124.
- Garcia-Retamero, R., Takezawa, M., and Galesic, M. (2010). Simple mechanisms for gathering social information. *New Ideas in Psychology*, 28(1):49–63.
- Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62:451–482.
- Glennan, S. (2017). *The New Mechanical Philosophy*. Oxford: Oxford University Press.
- Goering, J. (1978). Neighborhood tipping and racial transition: A review of social science evidence. *Journal of the American Planning Association*, 44(1):68–78.
- Golder, M. (2003). Explaining variation in the success of extreme right parties in Western Europe. *Comparative Political Studies*, 36(4):432–466.
- Goldthorpe, J. (2001). Causation, statistics, and sociology. *European Sociological Review*, 17(1):1–20.
- Gould, R. (1991). Multiple networks and mobilization in the Paris Commune, 1871. *American Sociological Review*, 56(6):716–729.
- Grannis, R. (1998). The importance of trivial streets: Residential streets and residential segregation. *American Journal of Sociology*, 103(6):1530–1564.
- Grannis, R. (2005). T-Communities: Pedestrian street networks and residential segregation in Chicago, Los Angeles, and New York. *City and Community*, 4(3):295–321.

- Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6):489–515.
- Green, A. (1997). *Housing, family and working lives: A report prepared for the Department of the Environment, Transport and the Regions*. University of Warwick, Institute for Employment Research.
- Grigoryeva, A. and Ruef, M. (2015). The historical demography of racial segregation. *American Sociological Review*, 80(4):814–842.
- Grodzins, M. (1957). Metropolitan segregation. *Scientific American*, 197(4):33–41.
- Hainmueller, J., Hangartner, D., and Lawrence, D. (2016). When lives are put on hold: Lengthy asylum processes decrease employment among refugees. *Science Advances*, 2:e1600432.
- Hall, M. and Crowder, K. (2014). Native out-migration and neighborhood immigration in new destinations. *Demography*, 51(6):2179–2202.
- Hangartner, D., Dinas, E., Marbach, M., Matakos, K., and Xefteris, D. (2019). Does exposure to the refugee crisis make natives more hostile? *American Political Science Review*, 113(2):442–455.
- Harding, D. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology*, 109(3):676–719.
- Harding, D. (2009). Collateral consequences of violence in disadvantaged neighborhoods. *Social Forces*, 88(2):757–784.
- Harris, D. (2001). Why are Whites and Blacks averse to black neighbors? *Social Science Research*, 30(1):100–116.
- Hatton, T. (2017). Refugees and asylum seekers, the crisis in Europe and the future of policy. *Economic Policy*, July:447–496.
- Heckman, J. (2005). Rejoinder: Response to Sobel. *Sociological Methodology*, 35:135–162.
- Hedman, L., van Ham, M., and Manley, D. (2011). Neighbourhood choice and neighbourhood reproduction. *Environment and Planning A: Economy and Space*, 43(6):1381–1399.
- Hedström, P. (1994). Contagious collectivities: On the spatial diffusion of Swedish Trade Unions, 1890-1940. *American Journal of Sociology*, 99(5):1157–1179.

- Hedström, P. (1998). Rational Imitation. In Hedström, P. and Swedberg, R., editors, *Social Mechanisms. An Analytical Approach to Social Theory*, pages 306 – 329. Cambridge University Press.
- Hedström, P. (2006). Experimental macro sociology: Predicting the next best seller. *Science*, 311:786–787.
- Hedström, P. and Bearman, P. (2009). What is Analytical Sociology all about? An introductory essay. In Hedström, P. and Bearman, P., editors, *The Oxford Handbook of Analytical Sociology*, pages 3–24. Oxford University Press.
- Hedström, P. and Swedberg, R. (1996). Social mechanisms. *Acta Sociologica*, 39(3):281–308.
- Hedström, P. and Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, 36:49–67.
- Hedström, P. (2005). *Dissecting the Social. On the Principles of Analytical Sociology*. Cambridge University Press.
- Hermansen, A. S. (2016). Moving up or falling behind? Intergenerational socioeconomic transmission among children of immigrants in Norway. *European Sociological Review*, 32(5):675–689.
- Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC.
- Hertwig, R. and Herzog, S. M. (2009). Fast and frugal heuristics: Tools of social rationality. *Social Cognition*, 27(5):661–698.
- Hipp, J., Faris, R., and Boessen, A. (2012). Measuring ‘neighborhood’: Constructing network neighborhoods. *Social Networks*, 34(1):128–140.
- Hipp, J. R. (2007). Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review*, 72(5):659–680.
- Hirschman, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. London: Harvard University Press.
- Ho, D., Imai, K., King, G., and Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236.
- Hvidtfeldt, C., Schultz-Nielsen, M. L., Tekin, E., and Fosgerau, M. (2018). An estimate of the effect of waiting time in the Danish asylum system on post-resettlement employment

- among refugees: Separating the pure delay effect from the effects of the conditions under which refugees are waiting. *PLoS ONE*, 13(11):e0206737.
- Iacus, S. M., King, G., and Porro, G. (2009). cem: Software for Coarsened Exact Matching. *Journal of Statistical Software*, 30(9):1–27.
- Iacus, S. M., King, G., and Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493):345–361.
- Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24.
- Jarvis, B., Kawalerowicz, J., and Valdez, S. (2017). Impact of ancestry categorisations on residential segregation measures using Swedish register data. *Scandinavian Journal of Public Health*, 45(17 Suppl.):62–65.
- Jonsson, J. O. and Rudolphi, F. (2011). Weak performance-strong determination: School achievement and educational choice among children of immigrants in Sweden. *European Sociological Review*, 27(4):487–508.
- Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice. *Decision Analysis*, 8(1):10–29.
- Kenny, P. D. and Miller, C. (2020). Does asylum seeker immigration increase support for the far right? Evidence from the United Kingdom, 2000–2015. *Journal of Ethnic and Migration Studies*, 0(0):1–18.
- King, G. and Nelson, R. (2018). Why propensity scores should not be used for matching. *Working Paper*.
- Knigge, P. (1998). The ecological correlates of right-wing extremism in Western Europe. *European Journal of Political Research*, 34(2):249–279.
- Koopmans, R. (2010). Trade-offs between equality and difference: Immigrant integration, multiculturalism and the welfare state in cross-national perspective. *Journal of Ethnic and Migration Studies*, 36(1):1–26.
- Koopmans, R. and Schaeffer, M. (2016). Statistical and perceived diversity and their impacts on neighborhood social cohesion in Germany, France and the Netherlands. *Social Indicators Research*, 125:853–883.
- Krysan, M., Couper, M. P., Farley, R., and Forman, T. A. (2009). Does race matter in neighborhood preferences? Results from a video experiment. *American Journal of Sociology*, 115(2):527–559.

- Krysan, M. and Crowder, K. (2017). *Cycle of Segregation: Social Processes and Residential Stratification*. Russell Sage Foundation.
- Krysan, M. and Farley, R. (2002). The residential preferences of blacks: Do they explain persistent segregation? *Social Forces*, 80(3):937–980.
- Kürschner Rauck, K. (2020). “Not in My Backyard!” The 2015 refugee crisis in Germany. *SSRN Electronic Journal*.
- Laurie, A. J. and Jaggi, N. K. (2003). Role of ‘vision’ in neighbourhood racial segregation: A variant of the schelling segregation model. *Urban Studies*, 40(13):2687–2704.
- Lazarsfeld, P. and Merton, R. K. (1954). Friendship as social process: A substantive and methodological analysis. In Berger, M., editor, *Freedom and Control in Modern Society*, chapter 2, pages 18–66. New York: Van Nostrand.
- Lee, B., Reardon, S., Firebaugh, G., Farrell, C., Matthews, S., and O’Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review*, 73(5):766–791.
- Legewie, J. and Schaeffer, M. (2016). Contested boundaries: Explaining where ethnoracial diversity provokes neighborhood conflict. *American Journal of Sociology*, 122(1):125–161.
- León-Medina, F. (2017a). Analytical sociology and agent-based modeling: Is generative sufficiency sufficient? *Sociological Theory*, 35(3):157–178.
- León-Medina, F. (2017b). Analytical sociology and its ‘syntax’ for explanation. *Social Science Information*, 56(4):502–525.
- Leszczensky, L. and Pink, S. (2019). What drives ethnic homophily? A relational approach on how ethnic identification moderates preferences for same-ethnic friends. *American Sociological Review*, 84(3):394–419.
- List, C. and Dietrich, F. (2011). A model of non-informational preference change. *Journal of Theoretical Politics*, 23(2):145–164.
- Logan, J. (1978). Growth, politics, and the stratification of places. *American Journal of Sociology*, 84(2):404–416.
- Logan, J. (2011). Separate and unequal: The neighborhood gap for Blacks, Hispanics and Asians in metropolitan America. Technical report, US2010 Project.
- Logan, J. (2012). Making a place for space: Spatial thinking in social science. *Annual Review of Sociology*, 38(1):507–524.

- Logan, J., Zhang, W., and Alba, R. (2002). Immigrant enclaves and ethnic communities in New York and Los Angeles. *American Sociological Review*, 67(2):299–322.
- López Turley, R. N. (2003). When do neighborhoods matter? The role of race and neighborhood peers. *Social Science Research*, 32:61–79.
- Lubbers, M. and Scheepers, P. (2001). Explaining the trend in extreme right-wing voting: Germany 1989–1998. *European Sociological Review*, 17(4):431–449.
- Ludwig, J., Duncan, G., Gennetian, L., Katz, L., Kessler, R., Kling, J., and Sanbonmatsu, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science*, 337(September):1505–1510.
- Ludwig, J., Liebman, J., Kling, J., Duncan, G., Katz, L., Kessler, R., and Sanbonmatsu, L. (2008). What can we learn about neighborhood effects from the Moving to Opportunity Experiment? *American Journal of Sociology*, 114(1):144–188.
- Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1):1–25.
- Macy, M. W. and Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, 28(1):143–166.
- Malmberg, B., Andersson, E., and Bergsten, Z. (2014). Composite geographical context and school choice attitudes in Sweden: A study based on individually defined, scalable neighborhoods. *Annals of the Association of American Geographers*, 104(4):869–888.
- Malmberg, B., Andersson, E., Nielsen, M., and Haandrikman, K. (2018). Residential segregation of European and non-European migrants in Sweden: 1990–2012. *European Journal of Population*, 34(2):169–193.
- Massey, D. (2016). Residential segregation is the linchpin of racial stratification. *City and Community*, 15(1):4–7.
- Massey, D. and Denton, N. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.
- McDermott, M. and Samson, F. L. (2005). White racial and ethnic identity in the United States. *Annual Review of Sociology*, 31:245–261.
- McLaren, L. (2003). Anti-immigrant prejudice in Europe: Contact, threat perception, and preferences for the exclusion of migrants. *Social Forces*, 81(3):909–936.
- McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.

- Merton, R. (1949). On sociological theories of the middle range. In *Social Theory and Social Structure*, pages 39–53. Simon and Schuster, The Free Press.
- Merton, R. (1988). The Matthew Effect in science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS*, 79(4):606–623.
- Migrationsverket (2021). Swedish Migration Agency. Statistics about asylum applications. <https://www.migrationsverket.se/English/About-the-Migration-Agency/Statistics/Asylum.html>. Accessed: 09-04-2021.
- Molotch, H. (1969). Racial change in a stable community. *American Journal of Sociology*, 75(2):226–238.
- Monmonier, M. (2005). Lying with maps. *Statistical Science*, 20(3):215–222.
- Mood, C. (2010). Logistic regression: Why we cannot do what We think we can do, and what we can do about it. *European Sociological Review*, 26(1):67–82.
- Mousavi, S. and Gigerenzer, G. (2014). Risk, uncertainty, and heuristics. *Journal of Business Research*, 67(8):1671–1678.
- Muchnik, L., Aral, S., and Taylor, S. (2013). Social influence bias. *Science*, 341:647–651.
- Mullen, B. (1987). Self-attention theory: The effects of group composition on the individual. In *Theories of group behavior*, pages 125–146. Springer.
- Mullen, B. (1991). Group composition, salience, and cognitive representations: The phenomenology of being in a group. *Journal of Experimental Social Psychology*, 27(4):297–323.
- Müller, T., Grund, T., and Koskinen, J. (2018). Residential segregation and ‘ethnic flight’ vs ‘ethnic avoidance’ in sweden. *European Sociological Review*, 34(3):268–285.
- Müller, T., Hedström, P., Valdez, S., and Wennberg, K. (2014). Right-wing populism and social distance towards Muslims in Sweden – Results from a nation-wide vignette study. *Working Paper, Institute for Analytical Sociology, Linköping University*, pages 1–18.
- Musterd, S. (2005). Social and ethnic segregation in Europe: Levels, causes, and effects. *Journal of Urban Affairs*, 27(3):331–348.
- Nee, V. and Holbrow, H. (2013). Why Asian Americans are becoming mainstream. *Daedalus*, 142(3):65–75.
- O’Sullivan, D. (2009). Changing neighborhoods-neighborhoods changing: A framework for spatially explicit agent-based models of social systems. *Sociological Methods and Research*, 37(4):498–530.

- Ottensmann, J. (1995). Requiem for the tipping-point hypothesis. *Journal of Planning Literature*, 10(2):131–141.
- Pais, J. F., South, S. J., and Crowder, K. (2009). White flight revisited: A multiethnic perspective on neighborhood out-migration. *Population Research and Policy Review*, 28(3):321–346.
- Paolini, S., Harwood, J., and Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin*, 36(12):1723–1738.
- Park, R. (1924). The concept of social distance as applied to the study of racial attitudes and racial relations. *Journal of Applied Sociology*, 8:339–344.
- Park, R. (1936). Human ecology. *American Journal of Sociology*, 42(1):1–15.
- Pearl, J. (2010). An introduction to causal inference. *The International Journal of Biostatistics*, 6(2):Article 7.
- Pearl, J. (2013). Structural counterfactuals: A brief introduction. *Cognitive Science*, 37(6):977–985.
- Pearl, J. and Robins, J. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 444–453.
- Pettigrew, T. (1998). Reactions toward the new minorities of Western Europe. *Annual Review of Sociology*, 24:77–103.
- Pettigrew, T. and Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5):751–783.
- Preciado, P., Snijders, T. A., Burk, W. J., Stattin, H., and Kerr, M. (2012). Does proximity matter? Distance dependence of adolescent friendships. *Social Networks*, 34(1):18–31.
- Quillian, L. (1999). Migration patterns and the growth of high poverty neighborhoods, 1970–1990. *American Journal of Sociology*, 105(1):1–37.
- Quillian, L. (2002). Why is Black-White residential segregation so persistent?: Evidence on three theories from migration data. *Social Science Research*, 31(2):197–229.
- Quillian, L. (2012). Segregation and poverty concentration: The role of three segregations. *American Sociological Review*, 77(3):354–379.
- Quillian, L. (2014). Does segregation create winners and losers? Residential segregation and inequality in educational attainment. *Social Problems*, 61(3):402–426.

- Reardon, S. and Bischoff, K. (2011). Income inequality and income segregation. *American Journal of Sociology*, 116(4):1092–1153.
- Reardon, S. F., Farrell, C. R., Matthews, S. A., O’Sullivan, D., Bischoff, K., and Firebaugh, G. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Social Science Research*, 38(1):55–70.
- Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, 32(1):33–67.
- Reardon, S. F. and O’Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, 34(1):121–162.
- Riach, P. A. and Rich, J. (2002). Field experiments of discrimination in the market place. *The Economic Journal*, 112(483):F480–F518.
- Robbins, M., Saunders, J., and Kilmer, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention. *Journal of the American Statistical Association*, 112(517):109–126.
- Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Rubin, D. (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.
- Rydgren, J. (2007). The sociology of the radical right. *Annual Review of Sociology*, 33(1):241–262.
- Rydgren, J. and Ruth, P. (2011). Voting for the Radical Right in Swedish Municipalities: Social Marginality and Ethnic Competition? *Scandinavian Political Studies*, 34(3):202–225.
- Rydgren, J. and Ruth, P. (2013). Contextual explanations of radical right-wing support in Sweden: Socioeconomic marginalization, group threat, and the halo effect. *Ethnic and Racial Studies*, 36(4):711–728.
- Salganik, M., Dodds, P., and Watts, D. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5614):854–856.
- Salganik, M. and Watts, D. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly*, 71(4):338–355.
- Salganik, M. and Watts, D. (2009). Web-based experiments for the study of collective social dynamics in cultural markets. *Topics in Cognitive Science*, 1(3):439–468.

- Sampson, R. (2008). Moving to inequality: Neighborhood effects and experiments meet social structure. *American Journal of Sociology*, 114(1):189–231.
- Sampson, R., Morenoff, J., and Gannon-Rowley, T. (2002). Assessing “neighborhood effects”: Social processes and new directions in research. *Annual Review of Sociology*, 28:443–478.
- Sampson, R., Sharkey, P., and Raudenbush, S. (2008). Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proceedings of the National Academy of Sciences of the United States of America*, 105(3):845–852.
- Savage, L. (1954). *The Foundations of Statistics*. New York: Dover. 2nd ed.
- Schaaake, K., Burgers, J., and Mulder, C. H. (2010). Ethnicity at the individual and neighborhood level as an explanation for moving out of the neighborhood. *Population Research and Policy Review*, 29(4):593–608.
- Scheepers, P., Gijsberts, M., and Coenders, M. (2002). Ethnic exclusionism in European countries: Public opposition to civil rights for legal migrants as a response to perceived ethnic threat. *European Sociological Review*, 18(1):17–34.
- Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2):143–186.
- Schneider, S. (2008). Anti-immigrant attitudes in Europe: Outgroup size and perceived ethnic threat. *European Sociological Review*, 24(1):53–67.
- Schwab, W. and Marsh, E. (1980). The tipping-point model: Prediction of change in the racial composition of Cleveland, Ohio, neighborhoods, 1940-1970. *Environment and Planning A: Economy and Space*, 12:385–398.
- Semyonov, M., Raijman, R., and Gorodzeisky, A. (2008). Foreigners’ impact on European societies: Public views and perceptions in a cross-national comparative perspective. *International Journal of Comparative Sociology*, 49(1):5–29.
- Shalizi, C. and Thomas, A. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, 40(2):211–239.
- Sharkey, P. and Faber, J. (2014). Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects. *Annual Review of Sociology*, 40(1):559–579.
- Smaje, C. (1997). Not just a social construct: Theorising race and ethnicity. *Sociology*, 31(2):307–327.

- Small, M. L. and Adler, L. (2019). The role of space in the formation of social ties. *Annual Review of Sociology*, 45(1):111–132.
- Smith, S., Van Tubergen, F., Maas, I., and McFarland, D. (2016). Ethnic composition and friendship segregation: Differential effects for adolescent natives and immigrants. *American Journal of Sociology*, 121(4):1223–1272.
- Sobel, M. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407.
- South, S. and Crowder, K. (1998). Leaving the 'hood: Residential mobility between Black, White, and integrated neighborhoods. *American Sociological Review*, 63(1):17–26.
- Steinnes, D. (1977). Alternative models of neighborhood change. *Social Forces*, 55(4):1043–1057.
- Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1):1–21.
- Swank, D. and Betz, H.-G. (2003). Globalization, the welfare state and right-wing populism in Western Europe. *Socio-Economic Review*, 1(2):215–245.
- Taeuber, K. and Taeuber, A. (1976). A practitioner's perspective on the Index of Dissimilarity. *American Sociological Review*, 41(5):884–889.
- Thomas, M. and Moye, R. (2015). Race, class, and gender and the impact of racial segregation on Black-White income inequality. *Sociology of Race and Ethnicity*, 1(4):490–502.
- Todd, P. M. and Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, 24:143–165.
- Tran, V. C. and Valdez, N. M. (2015). Second-generation decline or advantage? Latino assimilation in the aftermath of the Great Recession. *International Migration Review*, 51(1):155–190.
- Turner, R. N., Hewstone, M., Voci, A., Paolini, S., and Christ, O. (2007). Reducing prejudice via direct and extended cross-group friendship. *European Review of Social Psychology*, 18(1):212–255.
- UNHCR (2015). Asylum trends 2014 - Levels and trends in industrialized countries. <https://www.unhcr.org/statistics/unhcrstats/551128679/asylum-levels-trends-industrializedcountries-2014.html>. Accessed: 17-09-2020.

- Valdez, S. (2014). Visibility and votes: A spatial analysis of anti-immigrant voting in Sweden. *Migration Studies*, 2(2):162–188.
- van de Rijt, A., Kang, S., Restivo, M., and Patil, A. (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences*, 111(19):6934–6939.
- van Ham, M. and Clark, W. A. (2009). Neighbourhood mobility in context: Household moves and changing neighbourhoods in the Netherlands. *Environment and Planning A: Economy and Space*, 41:1442–1459.
- Vasilakis, C. (2017). Massive migration and elections: Evidence from the refugee crisis in Greece. *International Migration*, 56(3):28–43.
- Verdery, A. M., Entwisle, B., Faust, K., and Rindfuss, R. R. (2012). Social and spatial networks: Kinship distance and dwelling unit proximity in rural Thailand. *Social Networks*, 34(1):112–127.
- Waters, M. C. and Jiménez, T. R. (2005). Assessing immigrant assimilation: New empirical and theoretical challenges. *Annual Review of Sociology*, 31:105–125.
- Wessel, T., Andersson, R., Kauppinen, T., and Andersen, H. S. (2017). Spatial integration of immigrants in Nordic cities: The relevance of Spatial Assimilation Theory in a welfare state context. *Urban Affairs Review*, 53(5):812–842.
- Wilensky, U. and Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. The MIT Press.
- Winship, C. (1977). A revaluation of indexes of residential segregation. *Social Forces*, 55(4):1058–1066.
- Wong, C., Bowers, J., Williams, T., and Simmons, K. D. (2012). Bringing the person back in: Boundaries, perceptions, and the measurement of racial context. *Journal of Politics*, 74(4):1153–1170.
- Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. Oxford University Press.
- Ylikoski, P. (2014). Agent-based simulation and sociological understanding. *Perspectives on Science*, 22(3):318–335.
- Zick, A., Pettigrew, T., and Wagner, U. (2008). Ethnic prejudice and discrimination in Europe. *Journal of Social Issues*, 64(2):233–251.

FACULTY OF ARTS AND SCIENCES

Linköping Studies in Arts and Sciences No. 809
Institute for Analytical Sociology Dissertation Series No. 1
Department of Management and Engineering

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se