A COGNITIVELY PLAUSIBLE MODEL OF NEIGHBORHOOD CHOICE

Elizabeth E. Bruch
Department of Sociology & Population Studies
University of Michigan

Joffre Swait
Centre for the Study of Choice
University of Technology, Sydney

ABSTRACT: Despite recognition that neighborhood environments are important contexts shaping health outcomes and health behaviors, we know very little about how existing neighborhoods got to be the way they are, and how they stay that way or change. Although there have been efforts in recent years to study the linkages between individual-level residential mobility patterns and macro-level changes in spatial inequality and segregation, such studies use implausible behavioral models of how people decide whether, when and where to move. This hampers both our understanding of the individual process of residential mobility and neighborhood attainment, as well as the inferences possible from simulations linking individual mobility with macro-level outcomes. Our work aims to: (1) develop and estimate cognitively plausible models of residential choice that allows for a decision-maker with incomplete information, heuristic (simplified) strategies for both search and screening, and a multi-stage decision process; (2) use agent-based models to explore the aggregate consequences of these behavioral models for patterns of residential segregation.

MOTIVATION: Within the United States, residential segregation is an enduring feature of the urban landscape (Logan 2011). Given that roughly 15 percent of the population moves each year, residential mobility is a key mechanism through which neighborhood patterns are produced and reproduced. Over the past several decades, sociologists and demographers have devoted considerable attention to describing patterns of residential choice. In the 1990s this work mostly took the form of "locational attainment" models (e.g., Logan and Alba 1993), which regress socio-demographic attributes of Census households onto univariate measures of neighborhood composition (e.g., median income, proportion black, etc.). In the 2000s, Crowder and colleagues used more explicit measures of mobility from the Panel Study of Income Dynamics to examine correlations between respondents' wealth, income, and race/ethnicity and the probability of moving into or out of a neighborhood characterized by its racial or economic composition (Crowder, South, and Chavez 2006; South, Crowder, and Pais 2011). Most recently, a line of work has advocated using discrete choice models (c.f. Mare and Bruch 2003; Bruch and Mare 2012), which allow neighborhood to be characterized based on multiple attributes (e.g., racial composition, median income, and housing prices).[1]

However, in all cases, the work emphasizes statistical significance and magnitude of coefficients rather than the underlying behaviors of individual actors. For example, discrete

[1] The models require an explicit characterization of the "choice set"—that is, the set of potential neighborhoods—and coefficients are estimated conditional on the choice set. In residential applications thus far, the choice set is assumed to be all potential residential neighborhoods within the metro area.

choice models assume a fully informed, utility maximizing rational actor who can easily weigh the relative merits of hundreds if not thousands of potential destinations within a metro area (c.f., McFadden 1974, 1978, 1999). But as we know from residential search studies, most people only consider a very small number of units prior to the selection of a new residence, and only look within a small geographic area (Rossi 1955; Clark and Smith 1982; Barrett 1976). This suggests that some kind of underlying screening process is at work, in which home-seekers first identify which areas or units are worthy of consideration, and then examine that small subset more closely. A choice model in which individuals first eliminate almost all potential options based on some simple cutoff (e.g., low crime, or no more than 30 minutes commute to work) and then examine the remaining options more closely, will have very different aggregate consequences than a choice model in which housing seekers give each potential destination the same level of scrutiny. In addition, it is a fact that omission of such a screening process from the model of choice if it actually occurs in observed data leads to significantly biased parameter estimates (Swait 1984; Swait and Ben-Akiva 1987a), which in turn lead to misunderstanding of the individual decision process, and ultimately to mis-directed and erroneous policy formulation.

To negate these undesirable outcomes, our goal is therefore to develop a statistical model of residential choice that better reflects the underlying decision process through which people select neighborhoods.

UNDERSTANDING THE CHOICE PROCESS*:* How people search within and then choose from a large collection of potential alternatives is a central focus of behavioral decision theory in general, and of several fields of applied economics in particular (transport and marketing stand out in this regard, the former more so than the latter). Scholars have accumulated substantial empirical evidence for the idea that, when choosing from more than a small handful of alternatives, decisions are typically made sequentially, with each stage reducing the set of potential options (Swait 1984; Swait and Ben-Akiva 1987a,b; Roberts and Lattin 1991, 1997; Swait 2001a,b; Swait and Erdem 2007). For a given individual, the set of potential options can first be divided into the set that he or she knows about, and those that he or she is unaware. This "awareness set" is further divided into options the person would seriously evaluate for "purchase", and those that are irrelevant or unattainable. This smaller set is referred to as the *choice set*,[2] and the final decision is restricted to options within that set. Research in consumer behavior suggests that the decision process to include alternatives in the choice set can be fundamentally different from the process of reaching the final choice itself (Shocker et al. 1991). For example, an individual purchasing milk at the supermarket might only consider organic brands, or containers below a given price range. Essentially, people favor less cognitively taxing rules that use a small number of choice attributes earlier in the choice process, but consider a wider range of factors for the final decision (Payne, Bettman, and Johnson 1993).

There is substantial empirical and theoretical support for choice sets in human decision-

[2] The choice set is defined as the subset of presented alternatives that have non-zero probabilities of being chosen (Swait and Feinberg 2013, *forthcoming*). Equivalently, alternatives excluded from the choice set can be thought of as structural zeros in a contingency table.

making. Payne (1976) shows that when comparing two apartments in a housing search, research subjects carefully weighted multiple attributes of each apartment. However, when faced with a larger number of potential alternatives, subjects eliminated many of the available alternatives as quickly as possible, and on the basis of a limited amount of information search and evaluation. Similarly, Bettman and Park (1980) show that the formation of the choice set is linked to one or two attributes only, while the final selection is more holistic. A large body of decision research demonstrates that strategies to screen potential options for study and evaluation are non-compensatory; a decision-maker's choice to eliminate from or include for consideration based on one attribute will not be reversed (or compensated for) based on the value of other attributes (e.g., Dieckmann, Dippold, and Dietrich 2009; Hauser, Ding, and Gaskin 2009). In other words, non-compensatory decision rules act as "dealbreakers" or "dealmakers" that serve to eliminate many potential alternatives from entering the choice set. Once the decision maker has narrowed down his or her options to a few alternatives, the final choice decision may allow different dimensions of alternatives to be compensatory; in other words, a less attractive value on one attribute may be offset by a more attractive value on another attribute.

These micro-details of decision-making might be solely of psychological interest, except for the fact that they have important implications for macro-level social patterns of inequality. A multistage decision process that eliminates many potential alternatives in the initial stage using a non-compensatory decision rule, and then only later allows for a more holistic evaluation will likely have very different aggregate implications for social inequality and social differentiation than a single-stage decision rule that assumes a more holistic evaluation and decision approach. For example, in residential segregation an important mechanism for stable integration is the cumulation of small, unlikely mobility decisions (Bruch and Mare 2006, 2009). If individuals using a screening rule *never* consider neighborhoods above some threshold number of black residents or a given poverty rate, regardless of the other amenities of the neighborhood, this cumulation will never occur.

Figure 1 contrasts the modeling framework used in conventional models of individual choice (Panel A) with the framework used in the more "cognitively plausible" choice set formation (CSF) choice models proposed in this article (Panel B). In conventional choice models (e.g., McFadden 1974; Ben-Akiva and Lerman 1985; Bruch and Mare 2012), the analyst assumes that individuals evaluate all $k = 1,2 \dots K$ salient attributes of *all* potential choice alternatives $j = 1,2 \dots J$ (denoted by X_{ij}^{k}) in a single decision stage, typically by computing a weighted sum of the attributes of each alternative and then choosing the alternative with the highest overall value. This assumes a compensatory evaluation process in which a less desirable value on one attribute may be compensated for by a more desirable value on another attribute. In contrast, CSF models allow for choices to be made in multiple stages, with each stage reducing the number of alternatives under consideration and different rules governing decisions at each stage. Typically only a small subset of the K attributes is processed at the screening stage (see Figure 1, Panel B). The screening rules may be compensatory or non-compensatory, and different screening rules can be hypothesized and evaluated empirically.

Note that in many cases—including our own—the researcher does not observe the choice set. We only observe the final choice outcome. However, we can treat the choice set as a latent

construct and test alternative hypotheses about what attributes led to its creation. We anticipate that allowing for choice sets and multistage decision-making will dramatically improve the fit of our models. We will compare the fit of standard discrete choice models with our more "cognitively plausible" specification.

DATA*:* We will use data from the Los Angeles Family and Neighborhood survey (LA FANS). The LA FANS allow us to characterize the choice set formation process for a single metro area. The LA FANS is a panel study of 3,250 households who were living in one of 64 sampled neighborhoods in Los Angeles County at the time of the initial survey. The first wave of data was collected in 2000-2001, and the second wave in 2006-2008. See Sastry *et al*. (2006) for more details on the survey design. The data include retrospective residential histories for the two years prior to the first wave of data collection, and the 5-8 years between the Wave I and Wave II data collection. We have access to restricted data on block-level geographic identifiers for the neighborhoods lived in by survey respondents, as well as other focal points within the urban landscape (the schools attended by household children, place of work, and the location of any church or other place of worship). We also have a host of socio-demographic attributes (housing tenure, race/ethnicity, work status, and number of children). We use 1990-2010 decennial Census data to derive neighborhood measures for each neighborhood in Los Angeles County. Linear interpolation provides neighborhood measures within Census years. We have also obtained detailed data on transport planning zones for Los Angeles County from the Southern California Association of Governments (SCAG), which are used to develop a matrix of commuting times between areas. Data from the LAPD are used to get measures of crime for each Census tract over the observation period.

ANALYSIS PLAN: The goal is to develop a multistage decision model of residential choice that allows for empirical evaluation of alternative hypotheses about choice set formation. The outcome variable is the neighborhood chosen by the household at a given point in time. Figure 2 illustrates one such model, where the choice process is viewed as a series of decisions about where to live, each decision reducing the number of available alternatives. This figure hypothesizes a process whereby the decision-maker first decides whether to move from the current unit or stay put. If the decision is to move, the mover then determines which sub-regions of Los Angeles to focus on by deciding: (1) which macro-region to search within (Los Angeles County can be divided into seven macro geographic regions, most of which are separated by mountains: West LA, San Fernando Valley, San Gabriel Valley, Antelope Valley, South Bay, Central Los Angeles, and the Gateway cities);[3] (2) which types of areas within the macro region to focus on using some criterion ("Aggregation Behavior") (e.g., areas with affordable housing, shorter commutes, good schools, low crime, and/or other amenities). This produces the set of neighborhoods that "survive" screening, which is likely a far smaller subset than the total number of neighborhoods in Los Angeles County.

[3] A small literature on housing search shows that, when searching for housing, almost all people concentrate their search within a narrow geographic radius. For example, in his examination of housing search patterns among Toronto residents, Barrett (1976) found that 92% of the housing seekers he interviewed confined their search to housing units within a 3-mile radius, and the median distance was less than ¾ of a mile. In his study of prospective home buyers in the San Fernando Valley, Huff (1986) shows that many households' searches tend to cluster around "anchor points"—geographic locations such as workplace or schools that orient the search.

We will test different hypotheses regarding the appropriate specification of both macro regions and aggregation behavior by estimating alternative models and then comparing the goodness of fit. In the LA FANS data, one only observes the final choice of neighborhood. We do not observe what neighborhoods were explicitly considered and subsequently rejected. The unobservability of the choice set requires us to allow for possible choice sets, but assign a probability of occurrence to each. Another challenge in estimating this model is the large number of neighborhood alternatives and potential choice sets in Los Angeles County. For example, in theory one can search for housing in any combination of the 7-macro regions, which results in many possible choice sets in Stage 1. In addition, there are approximately 2000 Census tracts in Los Angeles, each of which represents a unique neighborhood destination. Most choice set applications in marketing have a much smaller number of choice alternatives, generally between 3-7 total items. We will explore different strategies of applying constraints to eliminate highly unlikely choice combinations (for example, a person looking for housing in Beverly Hills is unlikely to also be searching in Watts).

In the next stage of the project, we will simulate the residential dynamics implied by the cognitively plausible choice models by extending on an existing empirical agent-based model that uses real geography and populations (see Bruch 2013: 33-37). We will use this model to explore how each stage of the residential choice process described in Figure 2 matters for aggregate patterns of segregation. Our hope is that over the longer run this behaviorally nuanced model of residential choice will provide subsidies for new theories of segregation, and more targeted policies aimed at reducing segregation.

REFERENCES

Barrett, Frank. 1976. The search process in residential relocation. *Environment and Behavior* 8: 169-198.

Bayer, P, R McMillan, and K S Rueben. 2004. "What drives racial segregation? New evidence using census microdata." *Journal of Urban Economics* 56 (3): 514-535.

Ben-Akiva, M and S Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT press.

Bettman, J. R., and C. W. Park. 1980. "Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis." *Journal of Consumer Research* XX:234–248.

Bruch, ED and RD Mare. 2006. "Neighborhood choice and neighborhood change." *American Journal of Sociology* 112 (3): 667-709.

Bruch, Elizabeth. 2013. "Residential Sorting by Race and Income." Conditionally accepted to the *American Journal of Sociology.*

Clark, W A V and T R Smith. 1982. "Housing market search behavior and expected utility theory: 2. The process of search." *Environment and Planning A* 14: 717-737.

Crowder, K, S J South, and E Chavez. 2006. Wealth, race, and inter-neighborhood migration. *American Sociological Review* 71: 72-94.

Dieckmann, Anja, Katrin Dippold, and Holger Dietrich. 2009. "Compensatory versus Non-compensatory models for predicting consumer preferences." *Judgment and Decision Making* 4: 200-213.

Hauser, J R, M Ding, and S P Gaskin. 2009. "Non-compensatory (and compensatory) models of consideration-set decisions." In *Proceedings of the Sawtooth Software Conference.*

Huff, James. 1986. "Geographic Regularities in Residential Search Behavior." *Annals of the Association of American Geographers* 76: 208-227.

Logan, John. 2011. "Separate and Unequal: The Neighborhood Gap for Blacks, Hispanics, and Asians in Metropolitan America." New York: Russell Sage.

Logan, J R and R D Alba. 1993. Locational returns to human capital: Minority access to suburban community resources. *Demography* 30: 243-268.

McFadden, Daniel. 1974. "Conditional logit analysis of qualitative choice behavior." Chapter 4 in *Frontiers in Econometrics*. New York: Academic Press.

_____. 1978. "Modelling the choice of residential location." Institute of Transportation Studies, University of California.

———. 1999. "Rationality for economists?" *Journal of risk and uncertainty* 19 (1): 73–105.

Mare, Robert, and Elizabeth Bruch. 2003. "Spatial Inequality, Neighborhood Mobility, and Residential Segregation." California Center for Population Research Working Paper No. 003-03.

Payne, J. W. 1976. "Task complexity and contingent processing in decision making: An information search and protocol analysis." *Organizational behavior and human performance* 16: 366–387.

Payne, J. W., J. R. Bettman, and E. J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.

Roberts, J. H., and J. M. Lattin. 1991. "Development and testing of a model of consideration set composition." *Journal of Marketing Research* XX:429–440.

———. 1997. "Consideration: Review of research and prospects for future insights." Journal of Marketing Research XX:406–410.

Sastry, N, B Ghosh-Dastidar, J Adams, and A R Pebley. 2006. The design of a multilevel survey of children, families, and communities: The Los Angeles family and neighborhood survey. *Social Science Research* 35: 1000-1024.

Shocker, A. D., M. Ben-Akiva, B. Boccara, and P. Nedungadi. 1991. "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions." *Marketing Letters* 2: 181–197.

South, Scott, Kyle Crowder, and Jeremy Pais. 2011. Metropolitan structure and neighborhood attainment: Exploring intermetropolitan variation in racial residential segregation. *Demography* XX*:*1-30.

Swait, J 1984. *Incorporating Constraints in Discrete Choice Models of Transportation Demand.* Doctoral Thesis, Dept. of Civil Engineering, Massachusetts Institute of Technology.

Swait, Joffre. 2001a. Choice Set Generation Within the Generalized Extreme Value Family of Discrete Choice Models, *Transportation Research Part B*, 35(7):643-666.

Swait, Joffre. 2001b A Non-Compensatory Choice Model Incorporating Attribute Cutoffs, *Transportation Research Part B*, 35(10):903-928.

Swait, J. and M. Ben-Akiva. 1987a. Incorporating Random Constraints in Discrete Choice Models of Choice Set Generation, *Transportation Research Part B,* 21, 91-102.

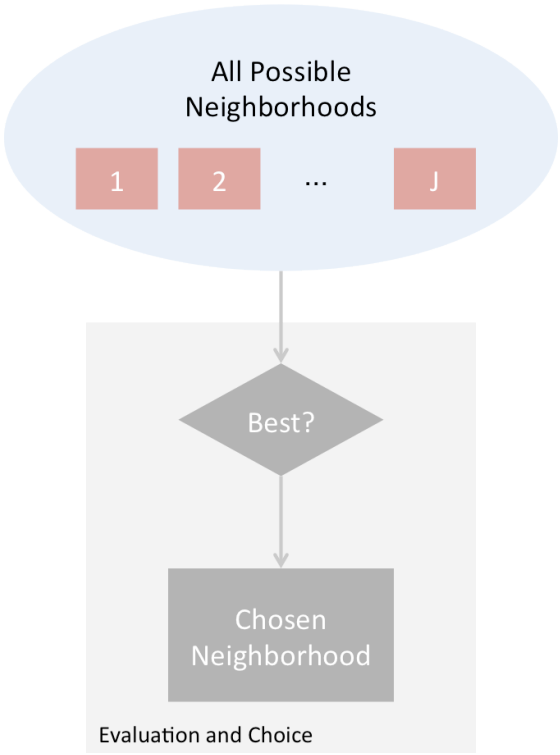
Swait, J. and M. Ben-Akiva. 1987b. Empirical Test of a Constrained Choice Discrete Model: Mode Choice in São Paulo, Brazil, *Transportation Research Part B*, 21, 103-115.

Swait, Joffre and Tulin Erdem. 2007. Brand Effects On Choice And Choice Set Formation Under Uncertainty, *Marketing Science*, 26(5):679-697.

Swait, Joffre and Fred Feinberg (2013) Deciding How to Decide: An Agenda for Multi-Stage Choice Modeling Research in Marketing, Handbook of Choice Modelling, Stephane Hess and Andrew Daly, Editors, Edward Elgar:London, UK, forthcoming.

Fig 1. Contrasting Conventional Choice Models with "Cognitively Plausible" Choice Models

A. Behavioral Model Underlying Standard Discrete Choice Model



B. Behavioral Model Underlying "Choice Set Formation" Models

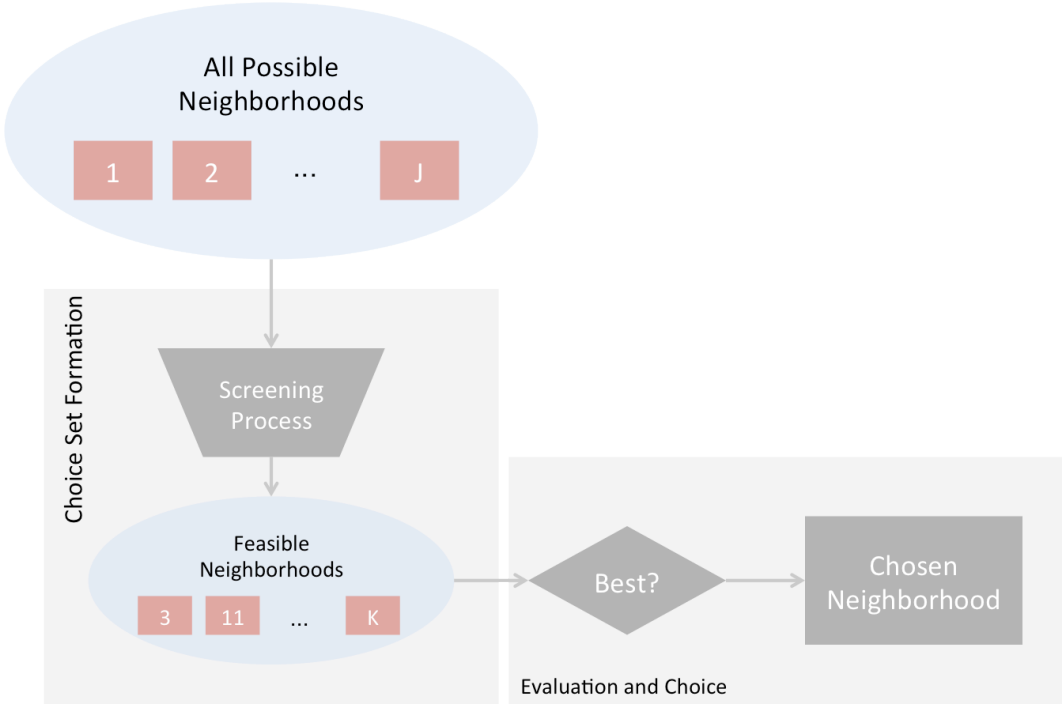


Fig 2. Multi-Stage Model of Residential Choice