Effects of Heterogeneity in Residential Preferences on an Agent-Based Model of Urban Sprawl

Daniel G. Brown[1] *and* *Derek T. Robinson*[1]

ABSTRACT. The ability of agent-based models (ABMs) to represent heterogeneity in the characteristics and behaviors of actors enables analyses about the implications of this heterogeneity for system behavior. The importance of heterogeneity in the specification of ABMs, however, creates new demands for empirical support. An earlier analysis of a survey of residential preferences within southeastern Michigan revealed seven groups of residents with similar preferences on similar characteristics of location. In this paper, we present an ABM that represents the process of residential development within an urban system and run it for a hypothetical pattern of environmental variation. Residential locations are selected by residential agents, who evaluate locations on the basis of preference for nearness to urban services, including jobs, aesthetic quality of the landscape, and their similarity to their neighbors. We populate our ABM with a population of residential preferences drawn from the survey results in five different ways: (1) preferences drawn at random; (2) equal preferences based on the mean from the entire survey sample; (3) preferences drawn from a single distribution, whose mean and standard deviation are derived from the survey sample; (4) equal preferences within each of seven groups, based on the group means; and (5) preferences drawn from distributions for each of seven groups, defined by group means and standard deviations. Model sensitivity analysis, based on multiple runs of our model under each case, revealed that adding heterogeneity to agents has a significant effect on model outcomes, measured by aggregate patterns of development sprawl and clustering.

Key Words: *complex systems; social surveys; spatial modeling; urban sprawl.*

INTRODUCTION

Geographers and economists, dating back to Ricardo (1821) and Thünen (1826), have developed theories of location and land use to describe the process of settlement and its resulting spatial patterns. Although these theories focused initially on agricultural land rents and the appropriate site selection for different agricultural products, subsequent work has extended these theories and models to consider intercity site selection (Marshall et al. 1961), social influences on site selection (Hurd 1903), and residential land uses (Alonso 1964). Each of these formal models has been based on assumptions of homogenous, rational agents acting on a featureless plain. The assumptions limit the usefulness of these models for investigating human-environment interactions or feedbacks known to affect land-use patterns, e.g., through human perception, cognition, and evaluation of the

landscape (Nassauer 1995). Furthermore, human perceptions, e.g., cognitive/mental maps, of landscapes vary (Golledge and Stimson 1987), as do the preferences and approaches firms and households have for selecting sites for businesses or residences within a landscape.

Agent-based modeling (ABM) is a contemporary approach that can be used to represent agents that are heterogeneous, adaptive, and interactive (Holland and Miller 1991, Hong and Page 2004), important characteristics of complex systems that create tractability problems for analytical models. ABM is distinguished from statistical modeling approaches in its focus on the ways in which macro-scale spatial patterns, e.g., urban settlement patterns, result from processes and behaviors of microscale actors, e.g., households and firms, and by its ability to represent nonlinear interactions (Epstein and Axtell 1996, Axelrod 1997, Gilbert and

[1]University of Michigan

Troitzsch 1999, Gimblett 2002, Parker et al. 2002, 2003). Though cellular automata (CA) have also been used to represent these dynamics, ABM permits mechanisms to be assigned to objects other than locations on the landscape, whereas CA rules refer to locations (Benenson and Torrens 2004). Agent-based models have been used to provide "proofs of existence" (Waldrop 1990) of spatial patterns resulting from the actions of individual agents with very simple behavioral rules. For example, Parker and Meretsky (2004) and Sasaki and Box (2003) demonstrated how spatial patterns like those described by Von Thünen can result from economically rational agent behaviors, and Schelling (1969, 1978) demonstrated how patterns of residential segregation can result when individuals have only a small preference to be near people like themselves.

Important issues in the use of ABMs are how to appropriately represent the heterogeneity of agents and their environment as software objects in ways that accurately reflect the actual heterogeneity of " real-world" objects, and what effects heterogeneity has on the outcomes of the models. In individual-based models of ecological systems, the degree of heterogeneity among and within species (Lomnicki and Sedziwy 1989, Uchmanski 2000) has clear effects on the viability and behavior of populations. Many abstract models of land-use dynamics (Sanders et al. 1997, Otter et al. 2001, Cioffi-Revilla and Gotts 2003, Parker et al. 2003, Brown et al. 2004) have demonstrated how complex interactions can give rise to observed land-use patterns without data on actor-level heterogeneity. Models of land-use that use data on agent characteristics have tended to (1) use aggregate data about agents or their environment (e.g., ILUTE, Miller et al. 2004), or (2) use ranges of acceptable values as defined in the literature and randomly assign values within those ranges (e.g., LUCITA, Deadman et al. 2004).

We explored the sensitivity of a model of residential location (called SOME, Brown et al. 2004, 2005) to heterogeneity in resident preferences. We introduced two different types of heterogeneity of agent preferences to the model, and represent that heterogeneity through analysis of survey data on residential preferences in southeastern Michigan (Marans 2003, Fernandez et al. 2005). The first type of heterogeneity, referred to here as "variability," reflects continuous variation in agent characteristics across the entire population or within a single agent type. We introduce this variability into the SOME model by defining normal distributions of residential preferences for each of a number of preference factors. The second type of heterogeneity, referred to here as "categorization," introduces multiple types or groups of individuals with similar preferences. An important difference between these approaches to representing heterogeneity is that variability assumes that various agent characteristics, e.g., preferences on a number of different factors, are independent, e.g., uncorrelated. On the other hand, categorization allows for correlation among agent characteristics, through the definitions of the various categories of agents. Correlation among preferences raises the possibility for nonlinear effects from categorization. We investigated categorization both with and without representing uncorrelated variability among agents within groups, whose characteristics may exhibit correlation. Under a series of experimental settings involving variability and categorization of agent preferences, we measured model outcomes in terms of spatial patterns of development, the distributions of agent utility and, for experiments that used categorization, relationships between group characteristics and their achieved spatial patterns and utility levels. Our goals were not to determine the level of variability that is most realistic, because we assume that populations have variable preferences. Instead, we are exploring the implications of various approaches to representing that variability in a model of urban sprawl, including not representing it at all.

In the next section we summarize an analysis of data from the Detroit Area Study (DAS) survey and describe the SOME model as a platform for experimentation and exploration in relation to the survey data. Next, we describe how we related survey questions to agent preferences, how we summarized the distributions for input to the model, and our approach to sensitivity analysis. We then present the results and conclude with a discussion of the results and some lessons learned about the implications of agent heterogeneity in our model of urban sprawl.

Table 1. Preference factors from the analysis of the survey data (Fernandez et al. 2005), with percent of variance explained by each. Numbers in parentheses indicate the loading of strongly related variables, i. e., those loading 0.5 or higher, on each factor. All variables not loading greater than 0.5 on any one variable are listed under "other factors." All variables used in the analysis are listed once.

Factor	Factor Label	Variance Explained	Strongly Related Preference Variables
1	Social Comfort	16.1 %	familiar with area (0.81) close to family and friends (0.75) people similar to you (0.60)
2	Openness/Naturalness	12.6 %	openness and spaciousness (0.85) close to natural areas (0.85) lots of recreational opportunities (0.50)
3	Residential Aesthetics	12.8 %	attractive appearance of neighborhood (0.75) appearance and layout of the dwelling (0.73)
4	Schools and Work	11.0 %	close to work (0.74) good schools (0.56)
Other factors			housing costs and good value convenient to shopping and schools community size

METHODS

The Survey data

What follows is a brief summary of the survey data; for a complete description, see Marans (2003). The survey was conducted in 2001 through both face-to-face interviews and mail questionnaires collected from residents of Livingston, Macomb, Monroe, Oakland, St. Clair, Washtenaw, and Wayne Counties in Michigan. The results presented in this paper are based on 592 respondents. The selection of respondents for our analysis reflects our interest in modeling the patterns of development based on initial movement into exurban areas. We included only those respondents who identified themselves as having purchased a single-family detached house during the past 10 yr and who lived in the exurban parts of the metropolitan area. To tie our goals to available information about the respondents of the survey, we defined exurban locations to include towns with populations less than 15,000 and all unincorporated areas, i.e., townships. Our analysis provides no basis for making claims about distributions of preferences outside the southeastern Michigan region. Respondents were asked about the importance of variables influencing their decision to move to their current neighborhoods (Table 1). A four-point importance scale ranging from "very important" to "not at all important," recoded to a numeric scale from 1 to 4, was used for the preference variables.

The goal of a previous analysis of these data was to summarize the major factors affecting residential location decisions, and to identify variation among residents on these various factors (Fernandez et al. 2005). We used a factor analysis to identify four preference factors with eigenvalue scores over 1.0, indicating significant explanatory power, with each

factor explaining over 10% of the total variance and 52.2% of the total variance accounted for by the four factors (Table 1). Because the statistical analysis is described in detail elsewhere (Fernandez et al. 2005), the results are summarized here and the reader is referred to the earlier paper for more details.

The loadings on Factor 1, social comfort, pointed to the importance of social networks and other social factors in the selection of residential locations (Table 1). This is a factor we had not included in previous versions of our model, but one that could alter the model dynamics enough that we decided to represent it in the model. As a result, we constructed a measure of neighborhood similarity, which is described in the section on The SOME model, with the specific intent of representing the tendency of some residents to value nearness to people more like themselves than others. We interpreted Factor 2, openness/naturalness, as describing the aesthetic quality associated with broadscale visual amenities on the landscape, such as like open space and rolling terrain, that residential homebuyers consider in their choice of a residential location. This factor was represented in the model as a general "aesthetic quality" component. Factor 3, residential aesthetics, was related to the aesthetic characteristics of the particular dwelling and neighborhood, which we interpreted to operate at a scale that is too finely grained for our modeling goals, recognizing that neighborhoods and homes can be designed in a variety of ways within a given setting. We, therefore, did not use the scores on Factor 3 in any of the model-based analyses. Factor 4, schools and work, represented availability of a range of urban services and employment opportunities provided by commercial, institutional, industrial, and retail developments. Because of the high loading of the nearness to work question on Factor 4, along with good schools and the loading, albeit weak, of convenience to shopping and schools (Fernandez et al. 2005), we represented Factor 4 in the model using an indicator of distance to services.

The SOME model, described in the next section, was designed to represent decision making about residential location that includes evaluation of landscape characteristics that are related to those identified in the factor analysis of survey respondents, with the exception of Factor 3. A subsequent cluster analysis of survey respondents was used to identify groups of agents with similar preference characteristics and to evaluate the effects of these categories on spatial settlement patterns. The resulting seven categories of residents, which represented heterogeneity through categorization, were created based on similarity of scores on the four preference factors, though only three of the factors were used in the subsequent model analyses (Fernandez et al. 2005) and summarized as means and standard deviations of factor scores within each cluster (Table 1). Individuals in Cluster 1 assigned greater importance to Social Comfort, Residential Aesthetics, and Schools and Work issues, indicated by the low mean values for these factor scores. Residents in Cluster 2 assigned greater importance to Openness/Naturalness, Residential Aesthetics and Schools and Work, while those in Cluster 3 assigned greater importance to Social Comfort and Openness/Naturalness. Those in Cluster 4 assigned greater importance to Schools and Work, and Cluster 5 put more weight on Residential Aesthetics and, to a lesser extent, Social Comfort. Cluster 6 residents found Openness/Naturalness and Residential Aesthetics most important, and Cluster 7 residents assigned greater importance to Openness/ Naturalness and Schools and Work.

The model

The first model from our project entitled Spatial Land-Use Change and Ecological Effects (SLUCE), which we have named SLUCE's original model for exploration (SOME), was designed specifically as a residential location model that would be linked to the results obtained from the Detroit Area Survey (Fernandez et al. 2005) to simulate residential land-use patterns. Initial implementations of the model incorporated distance to service centers and aesthetic quality as drivers of residential location. These initial mechanisms were analytically validated, while studying the effects of a greenbelt near a growing city (Brown et al. 2004). Also, using the initial version of SOME, we determined that SOME could generate distributions of development cluster sizes that compared well with the structural form of real-world cities (Batty and Longley 1994) using different types, e.g., additive vs. multiplicative, of utility functions as representations of residential location decision making (Rand et al. 2003). The version described here as well as a version for pedagogical purposes is available on the Internet (http://www.cscs.umich. edu/sluce/).

A wide range of factors driving residential location

Fig. 1. Screen capture from a typical simulation run of the SOME model. The spatially autocorrelated and heterogeneous landscape is represented by shades of green, and brighter green represents higher aesthetic quality. The initial service center is located in the center of the landscape in yellow, and all other service centers are depicted in red. Residents are depicted using dark black squares. Last, the violet circle represents the boundary used in our sprawl measurements. The labels and circle were added for graphic illustration and are not displayed by typical model runs.
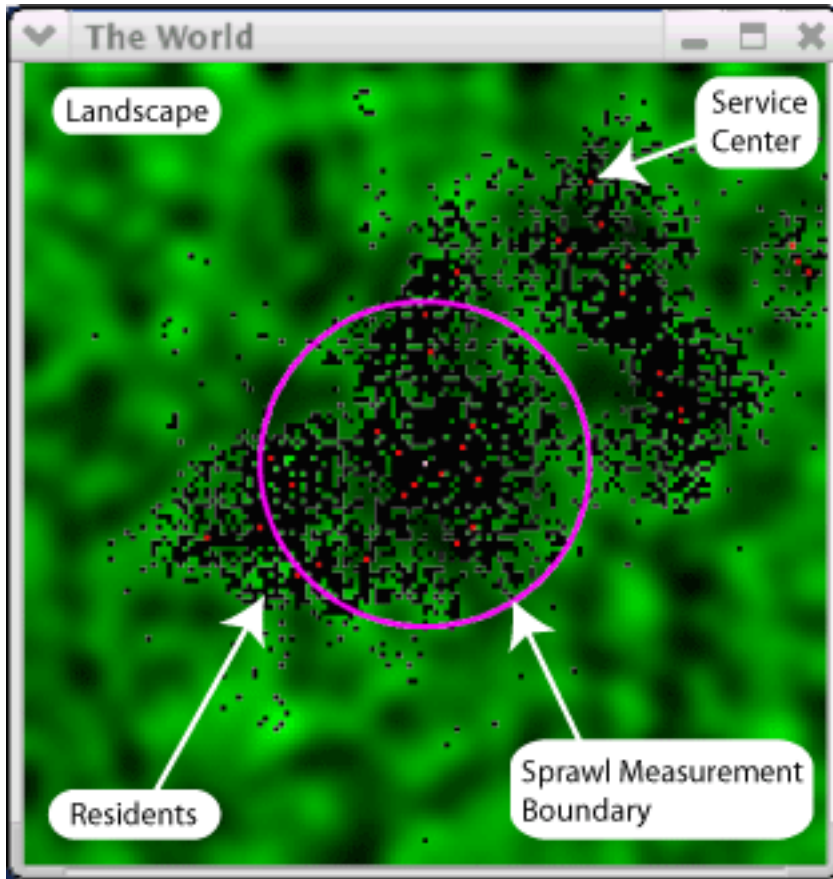
choices have been described in the literature, including the cost of travel and travel time to work (Block and Dupuis 2001), proximity to and quality of open space and urban amenities (Irwin 2002), housing characteristics (Geoghegan 2002), agglomeration effects (Krugman 1993, Weisbrod et al. 1980), sociodemographic factors (Weisbrod et al. 1980), and government policies. Because our focus was on linking an agent model to household survey data to investigate the influence of residential household heterogeneity on development patterns, we implemented only those drivers identified in the household survey. Thus in the current version of

SOME, described in more detail in Appendix 1, the residential agents use distance to service centers, related to Factor 4, above, aesthetic quality (Factor 2), and neighborhood similarity (Factor 1) to select residential locations, the first two of which are illustrated in Fig. 1. The agents use a modified Cobb-Douglas utility function (Appendix 2) to evaluate the utility they would attain from a number of randomly selected locations, using the following form:

Fig. 2. Standard deviation scaling for the social comfort factor. Distributions for each of the seven groups were derived using the mean and standard deviation factor scores for each group. Distributions were cut off at ± a) 1.5, b) 2, and c) 3 standard deviations.

Distributions of Population Groups
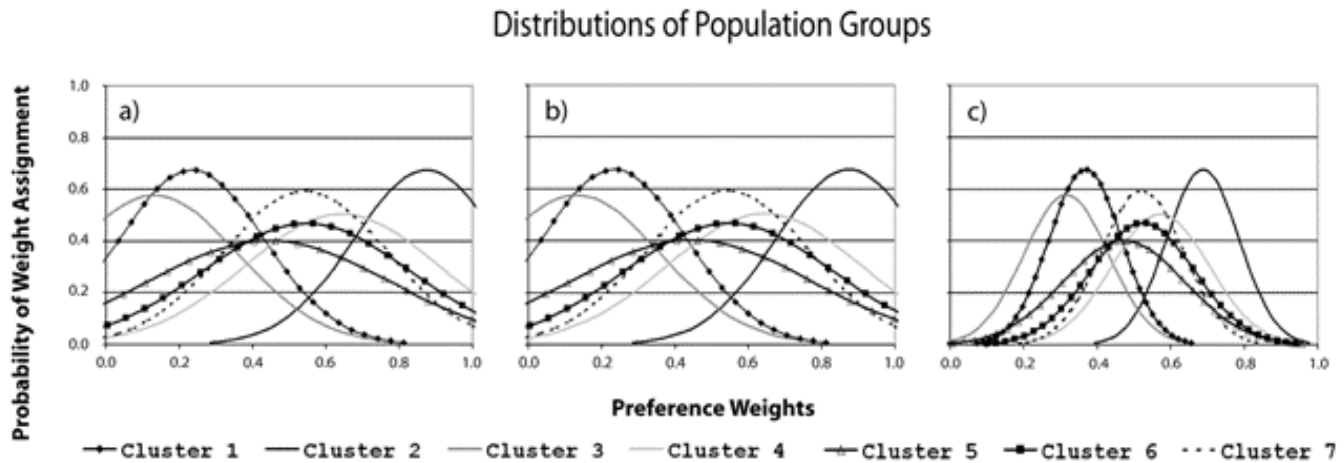
$$u_{r(x,y)} = \prod_{i=1}^{m} \left(1 - \left|\beta_i - \gamma_{i(x,y)}\right|\right)^{\alpha_{ir}} \quad (1)$$

where $u_{r(x,y)}$ is the utility of location (x,y) for resident r; α_{ir} is the weight the resident r places on factor i; β_i is the preferred value on component i and assumed constant for all residents, i.e., all residents desire the most aesthetic quality and shortest distance to service centers; $\gamma_{i(x,y)}$ is the value of component i at location (x,y), and m is the number of components evaluated, i.e., three. Every agent must have a preference weight for each component, i.e., α_{ir}, and the preference weights across the three components were constrained to sum to one, for reasons outlined in Appendix 2. We assigned preference weights to the agents based on our analysis of the survey data.

The first component in the utility function, distance to service centers, was calculated using straight-line distances from each location to the nearest service center. The values were inverted and scaled to the range [0,1], such that the maximum possible distance value was assigned a value of 0 and a cell immediately adjacent to two service centers was 1. This value was recalculated each time a new service center was added to the landscape.

For the aesthetic quality component, each location in the landscape was assigned a value from a hypothetical landscape, which was a surface of spatially autocorrelated values drawn from a normal distribution with values in the range [0,1] and a mean value of 0.5 (Fig. 1). We held the map of aesthetic quality values constant across all experiments and over time within a given model run.

The third component in the utility function incorporates an evaluation of social similarity by the active resident with the presettled resident agents in the neighborhood of a location. Residents measured their neighborhood similarity by using agent preference weights as surrogates for visual cues of wealth and lawn manicure, as well as distance preferences, which also suggest similar lifestyles tastes and preferences. Every time an agent evaluated a cell, that agent's preference weights for all three factors, aesthetic quality, distance to services, and neighborhood similarity, were compared to those of any other agents that were already located in the vicinity, i.e., the eight cell neighborhood of the cell being evaluated. If there were no residents in the neighborhood, the neighborhood similarity score for that location was set to a neutral value, i.e., 0.5 in a possible range of 0-1. For all residents in the vicinity, the composite similarity of preferences is calculated as

Table 2. Means and standard deviations of preference weights across the entire survey population and for each group resulting from the analysis in Fernandez et al. (2005).

| Cluster	1	2	3	4	5	6	7	Population
Aesthetic Quality								
Mean	0.37	0.41	0.90	0.69	0.33	0.33	0.76	0.54
S.D.	*0.25*	*0.25*	*0.56*	*0.31*	*0.32*	*0.29*	*0.22*	*0.39*
Distance to Service Centers								
Mean	0.33	0.38	0.77	0.34	0.77	0.79	0.47	0.55
Std. Dev.	*0.22*	*0.26*	*0.43*	*0.30*	*0.34*	*0.33*	*0.26*	*0.37*
Neighborhood Similarity								
Mean	0.30	0.78	0.22	0.60	0.45	0.53	0.53	0.49
Std. Dev.	*0.26*	*0.26*	*0.31*	*0.35*	*0.44*	*0.38*	*0.30*	*0.37*
Evenness	>0.99	0.95	0.88	0.96	0.94	0.95	0.98	>0.99

$$\gamma_{3r(x,y)} = \frac{n}{\sum_{j=1}^{n}\sqrt{[(\alpha_{1r} - \alpha_{1j})^2 + (\alpha_{2r} - \alpha_{2j})^2 + (\alpha_{3r} - \alpha_{3j})^2}} \quad (2)$$

where $\gamma_{3r(x,y)}$ is the neighborhood similarity value at location *(x,y)* for agent *r*; α_{1r}, α_{2r}, and α_{3r} are the preference weights for the three factors in the utility function Eq.1 for agent *r*, the one evaluating the location; α_{1j}, α_{2j}, and α_{3j} are the corresponding preference weights of agent *j* located at one of the eight locations neighboring location *(x,y)*; and *n* is the number of agents occupying neighboring locations. The neighborhood similarity values were then scaled to the range [0,1].

The neighborhood similarity component distinguishes this version of the SOME model from earlier published versions (Brown et al. 2004, 2005). In addition, this new component represents decision making that is similar to that found in the residential segregation model by Schelling (1969, 1978), in that

Fig. 3. The frequency of agents in each cluster according to the results of the analysis of the Detroit Area Study (Fernandez et al. 2005).
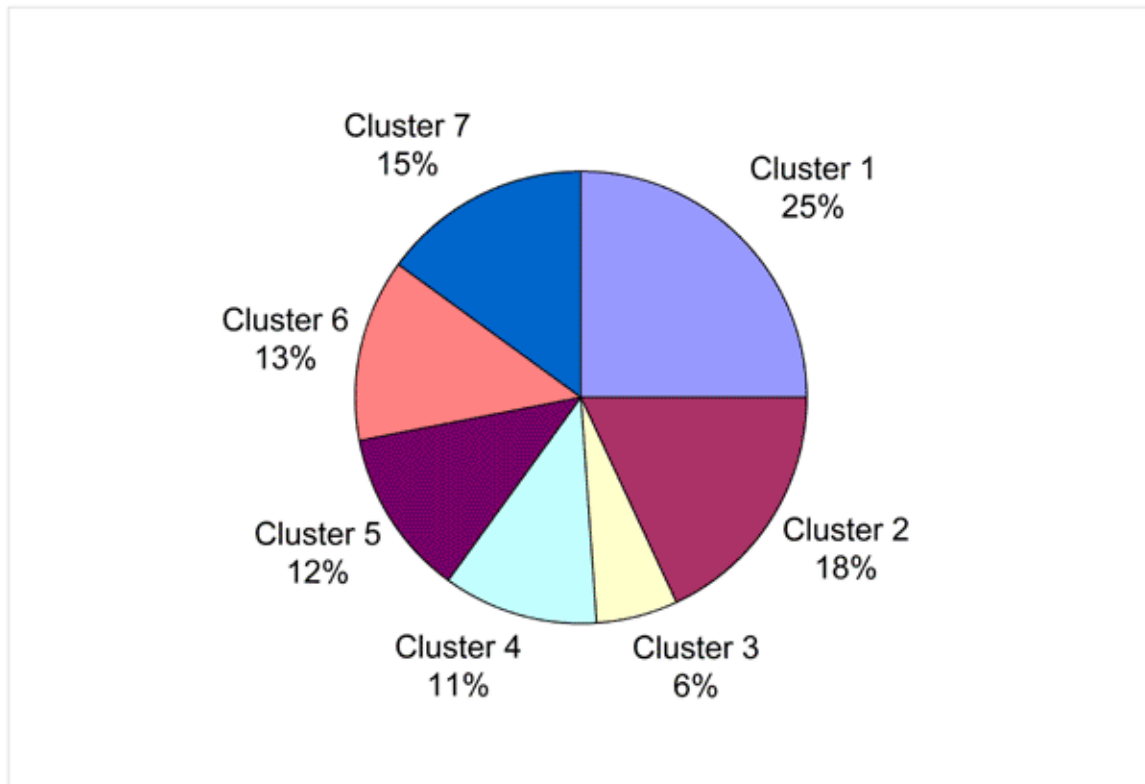
agents use similarity with neighbors as a consideration in where to locate. For this reason, it might be possible to use this version of the SOME model to extend Schelling's work on segregation using continuous measures of similarity and situations where agents consider more than just similarity, i.e., aesthetic quality and distance to services. Focused as we are on the effects of heterogeneity in preferences on urban settlement patterns, our goals do not extend to performing these comparisons.

We initialized the model with a single service center, located at the center of a 151 x 151 landscape map. Residents enter the world at a constant rate, i.e., 10/ time step, and a new service center locates near every 100th resident. Each resident selects the location that maximizes their utility function from a set, i.e., 15, of randomly selected sites. The limited number of sites evaluated is intended to mimic the effects of incomplete information and bounded rationality on the part of residential homebuyers. Once residents settle at a location, they remain at that location throughout the model run. The area of

Table 3. Measures used to describe model results.

Measure	Abbr.	Units	Description
Largest Patch Index	LPI	%	% of developed area taken up by largest patch
Mean Patch Size	MPS	# Cells	Size of average developed patch
Edge Density	ED	# Cell Sides / Cell	Developed/non-developed boundaries
Mean Nearest Neighbor	MNN	Cells	Ave. distance between developed patches
Developed Outside Radius	DOR	# Cells	# developed cells outside 31 cell radius
Proportion Outside Radius	POR	proportion	Proportion of residents of a given type outside 31 cell radius
Mean Resident Utility	MRU	Utility Units	Ave. utility value at end of model run
Variance Resident Utility	VRU	Utility Units	Variance of utility values at end of run
Gini Coefficient of Resident Utility	GINI	N/A	Disparity of utility across residents

developed land is about 15% of the total landscape. Landscape size and proportion of development values were chosen to minimize edge effects. Additional details on the sequence of model processes are given in Appendix 1, which provides a pseudo-code description of the model.

Populating the model with survey-based agents

Agent preference weights were empirically defined using the factor scores derived from the analysis of household survey data. Effectively, the factor analysis distilled the essential drivers of residential location by loading the array of survey questions onto four factors, of which we used three. Scores on Factor 1 were used to represent preference for neighborhood similarity, whereas scores on Factor 2 were used to represent preference for aesthetic quality and scores on Factor 4 were used to represent preference for nearness to services. In order to place the factor scores in the appropriate range for use in the model [0,1], we used a standard deviation scaling. The values of the mean factor score minus and plus a multiple of the standard deviation were set to zero and one, respectively. All values below/above the mean minus/plus the multiple of the standard deviation were set to 0/1. We ran all simulations and analyses with the preference weights scaled using three different multiples of standard deviations: 1.5, 2, and 3 (Fig. 2). We report only the numeric results for the case of 2 standard deviations, but the results are qualitatively similar in the other cases. Because the preference weights

Table 4. Means and standard deviations (S.D.) of metrics describing spatial patterns of development and utility distributions, calculated across 30 runs of the SOME model with agent preferences input in each of five experimental settings.

Metric	Statistic	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
		Uniform	Homo-genous	Normal	Group Means	Group Normals
LPI	Mean	9.90	12.19	8.78	11.04	9.72
	S.D.	*1.90*	*1.52*	*2.07*	*2.08*	*2.20*
MPS	Mean	9.92	16.04	10.20	13.70	10.56
	S.D.	*0.56*	*1.29*	*0.70*	*0.78*	*0.72*
ED	Mean	27.28	23.49	27.29	24.37	27.03
	S.D.	*0.80*	*0.66*	*0.61*	*0.50*	*0.83*
MNN	Mean	167.57	169.57	173.24	170.11	175.00
	S.D.	*8.77*	*11.42*	*11.13*	*11.08*	*11.34*
DOR	Mean	2086.40	1870.27	2126.37	1937.47	2069.00
	S.D.	*177.30*	*196.65*	*217.15*	*190.83*	*186.73*
MRU	Mean	34.40	25.10	28.90	28.70	29.70
	S.D.	*0.50*	*1.20*	*0.60*	*0.60*	*0.50*
VRU	Mean	3.30	1.60	3.10	2.00	3.10
	S.D.	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*
GINI	Mean	0.30	0.27	0.34	0.27	0.33
	S.D.	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*

for a given agent were constrained to sum to one (Appendix 2), the model proportionally rescaled drawn preference weights whenever they summed to something other than one. This step modified the distributions of preference weights somewhat, but was necessary given the assumptions of the utility function. Also, although the preference data collected from residents were in discrete form, i.e., indicators from strongly agree to strongly disagree, the numeric results of this approach to estimating preferences for input to the model are continuous, i.e., values between 0 and 1. Because our goal was to use the survey sample to estimate the distribution of the population, which was created through the model simulation, such conversion of form was reasonable.

We ran the SOME model with five different experimental settings, each summarizing the heterogeneity in the agent preferences differently. The "Uniform" case introduced no information from the survey and was our null model for comparison purposes. Agent preferences on each factor were drawn randomly from a uniform distribution. The "Homogeneous" case involved assignment of the mean value over the entire population for each preference weight to all agents (Table 2), creating a population of identical agents. In the "Normal" case, the agent preferences were drawn randomly from a normal distribution described by the overall mean and standard deviation, introducing variability into the agents' preference weights. The "Group Means" case involved assigning the mean preference weights from each of the seven clusters of residents. This case introduced categorization without variability. Agents with the mean preference of each group were created in the model in proportion to the corresponding numbers observed in the survey data, weighted by socioeconomic characteristics of the respondents (Fig. 3). Finally, the "Group Normals" case combined categorization and variability, by drawing preference weights randomly from one of seven distributions, described by the mean and standard deviation of the normal distribution of preference weights on each factor for each cluster derived from the survey data.

Model sensitivity analysis

We used five measures to describe the spatial patterns of development and three to describe the distribution of agent utilities after 340 time steps, summarized across 30 runs of the model for each of the five experimental settings (Table 3). Four landscape metrics were calculated using Fragstats (McGarigal et al. 2002) to describe the spatial patterns based on patches of development: largest patch index (LPI), mean patch size (MPS), edge density (ED), and mean nearest neighbor (MNN). The four additional measures were calculated within the SOME code. A circle with a radius of 31 cells around the initial service center, encircling approximately 18% of the lattice area, was used to count the number of developed cells occurring outside this radius (DOR), as a measure of settlement dispersion or sprawl (Fig. 1). Because the total number of residents varies by type, we computed the proportion of residents outside the radius (POR) to compare the results by resident type. The mean and variance of resident utility values (MRU and VRU) at the end of each model run was calculated across all agents and for each agent type separately. The Gini coefficient (GINI) was calculated to describe the disparity among agent utility levels (Sen 1973). The Gini coefficient approaches 1 when a large difference exists between the observed distribution of utility values and that of a population with evenly distributed utility.

We compared average values for each of these eight measures among the different experimental settings. Because standard errors for such comparisons are determined to a large extent, and can be manipulated, by changing the number of model runs, we do not report significance test results. Instead, we report the standard deviations for the 30 simulations we ran and interpret the differences of means considering the standard deviation values.

The experiments involving the seven groups of agents, i.e., Group Means and Group Normals, facilitated an exploration into the causes of spatial patterns of development and distributions of utility. First, we compared the mean values of the different measures calculated for each group separately. Then, we examined the relationships between group mean preference weights and spatial location by group (POR). Finally, to evaluated the degree to

Fig. 4. Representative map results from a single run of the model under each of the experimental settings:
A) Group Means, B) Group Normals, C) Homogenous, D) Normal, and E) Uniform.
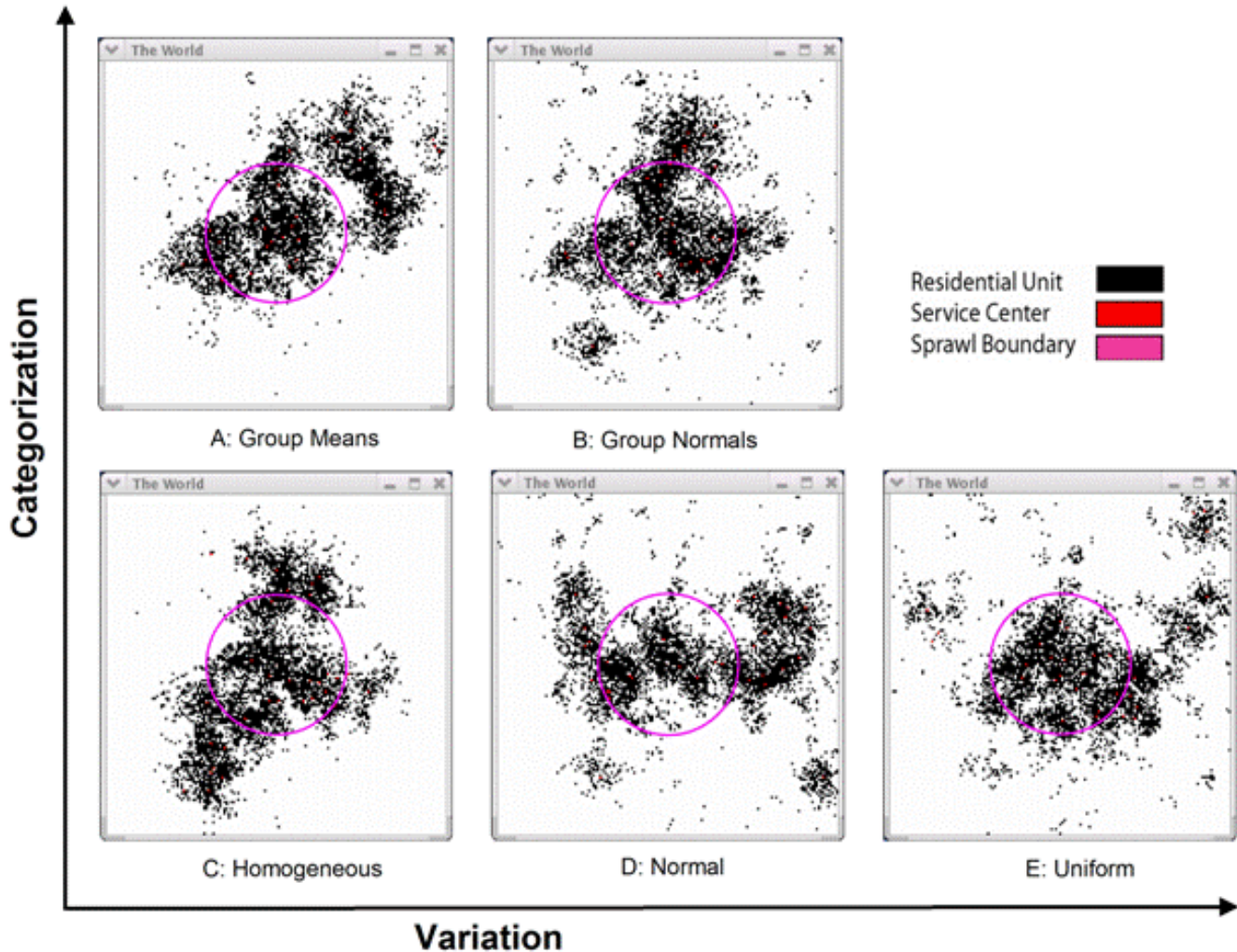
which the generality of agents' preferences affected
their overall welfare, we compared MRU and GINI
by category with the Shannon evenness index of the
average preference weights for those categories
(Table 1), calculated as

$$E_k = \frac{-\sum_{i=1}^{S} p_{ki} \cdot \ln p_{ki}}{\ln S} \quad (3)$$

where E_k is the evenness value for the kth group of
agents, p_{ki} is the average weight that group places
on the ith preference factor, rescaled so that sum of
average weights across all factors is one, and S is

Table 5. Means and standard deviations of metric values by agent category when model was run with mean values and normal distributions for each category.

Cluster		POR		MRU		GINI	
		Means	*Normals*	*Means*	*Normals*	*Means*	*Normals*
1	Mean	0.56	0.62	41.00	33.30	0.15	0.31
	S.D.	*0.06*	*0.06*	*0.70*	*0.60*	*0.01*	*0.01*
2	Mean	0.55	0.61	30.50	29.70	0.19	0.32
	S.D.	*0.07*	*0.06*	*0.80*	*0.70*	*0.01*	*0.01*
3	Mean	0.57	0.60	15.50	27.00	0.33	0.35
	S.D.	*0.08*	*0.07*	*1.00*	*1.60*	*0.02*	*0.01*
4	Mean	0.60	0.63	30.00	30.00	0.18	0.32
	S.D.	*0.06*	*0.08*	*0.90*	*0.90*	*0.01*	*0.01*
5	Mean	0.54	0.58	20.80	27.60	0.32	0.35
	S.D.	*0.08*	*0.07*	*0.80*	*0.90*	*0.02*	*0.01*
6	Mean	0.55	0.58	19.80	27.40	0.33	0.35
	S.D.	*0.06*	*0.06*	*0.70*	*1.00*	*0.01*	*0.01*
7	Mean	0.59	0.60	24.40	28.20	0.23	0.33
	S.D.	*0.05*	*0.06*	*0.90*	*0.80*	*0.01*	*0.01*

Fig. 5. Relationships of evenness of agent preferences by group with A) average utility and B) Gini coefficient of disparity in utility.
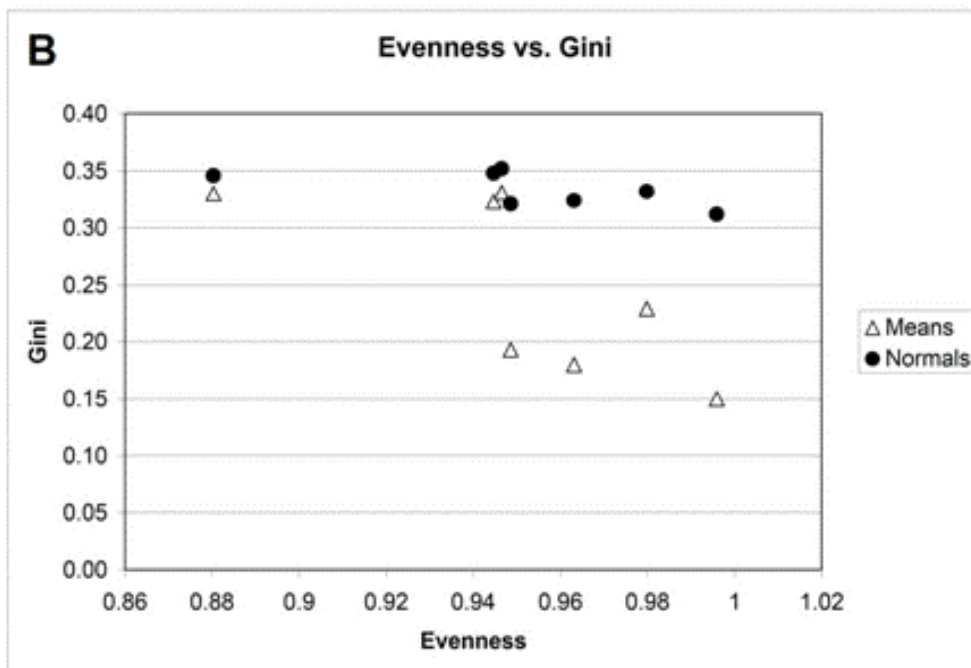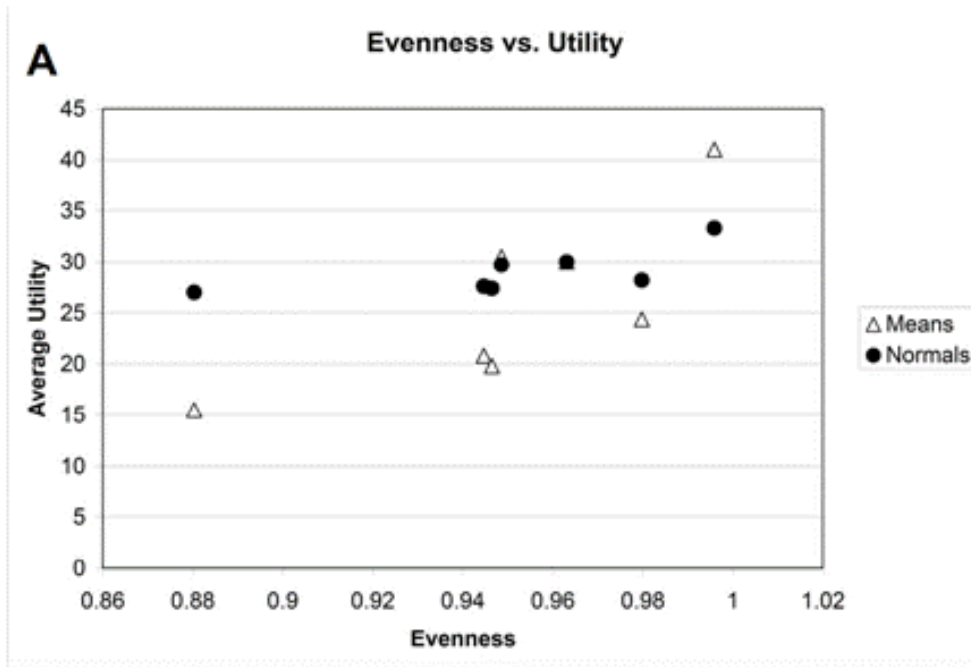
the number of preference factors, in this case 3. The index reaches a maximum of 1 when all preferences factors have equal weight. Lower values indicate a more specialized group.

RESULTS

Aggregate results

The results from the Uniform case indicated how the model would perform in the absence of survey data and without knowledge of residential preference weights. Effectively this represented complete variability, because each agent had unique preference weights or combinations thereof, and preference weights may have any value between the range 0-1. Not surprising, then, is the fact that patches were small on average, i.e., lowest MPS, oddly shaped, i.e., high ED, and well distributed across the landscape (high MNN, Table 3, Fig. 4e). The agents also had a wide range of satisfaction levels (highest VRU) and, as a population. had the highest overall utility (MRU) among the various treatments (Table 4). Other measurements fell in intermediate ranges among the other four experiments.

The Homogeneous case was most different from the Uniform case (Table 3, Figure 4c). When all residents had the same preferences weights, the resulting development pattern was most compact, i. e., highest LPI and MPS, lowest ED and DOR. The compaction, due to all agents preferring the same locations, led the Homogeneous agent population to be the most unsatisfied, i.e., lowest MRU and lowest VRU. However, the Homogeneous population also had the lowest disparity among agent utility values (GINI). Similar to the Homogeneous case, the Group Means case had the second lowest disparity (GINI). A slightly higher VRU led to greater satisfaction levels among the population, i.e., higher MRU, in the Group Means case. Also like the spatial patterns produced in the Homogeneous case, the Group Means case had compact development clusters, i.e., high LPI and MPS, and a low ED and DOR (Fig. 4a).

Using a Normal distribution to represent agent variability led to slightly larger patch sizes (LPI and MPS) and similar aggregate spatial results, but much lower agent satisfaction levels (MRU) with greater disparity (GINI) also occurred. If we think of the Normal case as adding variability to the

Homogenous case, we see that the effects of variation are much greater than those of categorization as described above in our comparison between the Homogeneous and Group Means cases (Table 4). The Normal case (Fig. 4d) produced results with more fragmented patterns, with smaller patches, i.e., lower MPS and LPI, more edge, i.e., higher ED, and greater spread from the center, i.e., higher DOR, though the latter was only marginally higher relative to the standard deviations. Also, agents had higher levels of utility (MRU) but greater variance and disparity (VRU and GINI) than the Homogenous case.

The results from the Group Normals case (Figure 4b) were virtually indistinguishable from the Normal case on all measures, but very different from the Group Means case on at least the size of the largest patch and amount of edge (LPI and ED), and the variance and disparity in utility (VRU and GINI, Table 4). Since the combination of categorization and variability in the Group Normals case was little different from the Normal case, it may be more useful at the aggregate to focus strictly on factors of variability rather than those of categorization. These results illustrate that it is more useful at the aggregate level to focus strictly on variability as a type of heterogeneity than categorization. It should also be noted that results from the standard deviation scaling that altered the degree of variability, as illustrated in Fig. 2, resulted in very similar model outcomes. Therefore the existence of variability was more influential than the degree of variability.

Results by category

In this section, we describe the results among the seven categories used in the Group Means and Group Normals experiments. The Group Normals case resulted in more spread away from the initial service center (POR) than did the Group Means case, but the difference was small relative to the standard deviations across model runs (Table 5, Figs. 4a and 4b). There were relatively large differences between the resident categories in utility (MRU) and disparity (GINI). The disparity in utility a group experienced tended to decrease linearly with its increasing average utility ($R^2 = 0.86$ and 0.82 for the Group Means and Group Normals cases, respectively). Also, MRU and GINI both tended to be higher in the Group Normals case than in the Group Means case. Under both experimental conditions, Groups 1, 2, and 4 achieved the highest

MRU, with correspondingly low GINI. These substantial differences in utility distributions among groups provided an opportunity to explore the preference characteristics that affected utility.

The amount of spatial sprawl a group experienced (POR, Table 5) increased with its average preference for aesthetic quality, $R^2 = 0.22$ and 0.71 for the Group Means and Group Normals cases, respectively, and decreased with its average preference for proximity to service centers ($R^2 = 0.55$ and 0.14). Strength of preference for neighbors like themselves had no effect on POR ($R^2 < 0.01$ in both cases).

The evenness values across the different categories ranged from 0.88 to >0.99 (Table 2), indicating a tendency for all groups to have relatively even weights across the different preference factors. Nonetheless, we observed relationships between evenness and both average (MRU) and disparity (GINI) in utility (Fig. 5). The relationships of evenness with MRU and GINI were stronger in the Group Mean case ($R^2 = 0.61$ and 0.51, respectively) compared to the Group Normals case ($R^2 = 0.49$ and 0.41, respectively), but they were consistent. Increases in group evenness corresponded to increases in utility and decreases in disparity. Because the evenness values refer to those of the average agent of each type, it more accurately represents the Group Means case than the Group Normals case.

DISCUSSION

Although assumptions of homogeneity or uniform distributions of preferences have proven useful in our initial models of land-use change for pedagogical purposes (Brown et al. 2004), they are simplified representations of reality. By coupling survey data with a simple model of residential location (SOME), we investigated the effects of two types of heterogeneity, i.e., categorization and variation, on aggregation spatial patterns of urban settlement. There is evidence from our experiments that variability in preference weights had a much stronger influence on results than did categorization. What evidence there is for an effect of categories, e.g., lower MPS and higher MRU for the Group Means vs. Homogenous cases suggests that using Group Means alone introduces some of the same effects that variability had on the results, but only more weakly. As more and more categories are introduced, the results to begin to approximate the effects of variability.

The analyses presented here corroborate findings from an earlier version of SOME (Rand et al. 2002) by showing that the introduction of variable agents results in more sprawl, regardless of the amount of variability. The Uniform case, which incorporated no a priori information, i.e., no survey data, into the model, exhibited spatial patterns and utility levels that were not distinguishable from the Normal and Group Normals cases. Conceptually, the Uniform case represents the upper limit on variation. Compared to a Homogenous population, the experiments with variation and categorization produced settlement patterns that were more fragmented and sprawling, and agents were able to achieve higher and more uniform utility levels (Table 4).

The influence of heterogeneity on development patterns resulted from the interaction of preference weights within the multiplicative utility function. As weights were allowed to vary, different residents found different types of settings more satisfying. For example, one resident weighted aesthetic quality high and distance to services low and, therefore, was more satisfied with a location far from the original center than a resident that had a high preference for proximity and low preference for aesthetic quality. As different agents selected locations on the basis of variable preferences, they tended to spread themselves out on the landscape more than residents did with identical preferences; the latter tended to cluster near each other because they liked the same things.

The results of our analysis by group suggest that, by allowing different agents to prefer and occupy different parts of the landscape, the competition for particular sites is reduced and, as a result, agent welfare is increased on average with the introduction of heterogeneity. More variability in the agent population resulted in higher average agent utility (Table 4). Spatial dispersion of agents (POR) in the different groups was weakly and negatively related to the group-average preferences for aesthetic quality and positively related to the proximity of service centers. Therefore, agents with a preference for aesthetic quality contributed to sprawling development patterns. Additionally, categories of agents that were characterized by relatively even preference weights across the three preference factors, i.e., generalists, achieved higher

levels of utility. Although the increases in aggregate average utility with increasing agent variability suggest that specialization of agents improved utility, relationships at the group level suggest that agents with less specialized preferences achieved highest utility, perhaps because they were more likely, through their imperfect sampling process, to find a location that satisfied their preferences. Thus, having different agents prefer different locations, i. e., more specialization, increased average utility, but those agents that were the most indifferent, i.e., more generalist, had the highest individual utilities.

The various calculations used to convert survey responses into preference weights in the utility model mean that the translation from survey data to model input is imperfect. First, populating models with data from social surveys requires that the actors surveyed be the same as, or at least adequate informants for, agents represented in the model, and the questions asked of the survey respondents must be designed to adequately gather information on actor decision making that can be mapped to agent behavior. In this case, there was good correspondence between the conceptual agents in the model and the households surveyed. Aside from the well-known limitations of using stated preferences, in comparison with revealed preferences (e.g., as studied in economics, Murphy et al. 2005), the meanings of questions about specific preferences need to match their meaning in a model. Furthermore, questions about preferences are not adequate for validating that a specific model decision-making process is correct. Second, because our utility model required preference weights to sum to one, distributions of preferences drawn from the survey data were necessarily modified in the model. Nonetheless, the relative weights and amounts of variability were consistent enough to produce results that comported with expectations when preference weights varied. For example, agent groups with more weight on distance to services sprawled less than those with less weight on distance to services. Because of the mutual dependence of the three preference-weight distributions, additional efforts to ensure that the model input distributions matched the distributions in the survey data would require a multi-criteria optimization, which was beyond the scope of the current study. Third, if a model runs over a long period, we need to consider possible dynamics in preferences, for which survey data may or may not be available. In our case, survey data were only available for one point in time. Despite these

challenges, surveys can be useful for a number of purposes. First, they may reveal factors used in decision making that are not included in the model, e.g., our social comfort factor. In addition, models may highlight areas that need to be reexamined using additional survey questions. In this paper, we described our decisions and rationale for relating survey responses to agent preferences in our model, and focused on using the survey data to evaluate the effects of alternative approaches to representing heterogeneity.

As a sensitivity analysis, our study provides important information about the degree to which spatial settlement patterns vary depending on the amount of variability in preferences represented in the residential population. The results suggest that there are clearly biases towards a more compact and less sprawling pattern when homogeneity is assumed. For this reason, it is important to attempt to represent the heterogeneity of the population. Our goals were not to perform, and our data did not support, a validation exercise, in which the model results were compared with multiple settlement patterns to evaluate the relative truthfulness of each model outcome. Tests of this sort are important next steps for this research theme.

CONCLUSIONS

Social surveys can serve as a source of information about the heterogeneity present in agents that are being represented in an agent-based model, provided the survey questions relate directly to the agent attributes in the model. Our experimental results, generated using an agent-based model of residential development that was populated with heterogeneous agent preferences that represented those observed in a survey of location preferences, indicated that introducing variability increased the amount of agent dispersion, or sprawl, the model produced. These findings provide critical insight into the limitations of models that assume homogenous populations. They also suggest that we can understand sprawl as, at least partially, a process driven by variability in preferences. Relationships between groups of similar agents indicated that agent preferences, and their distributions across various factors, affect spatial patterns of development and the utility achieved by agents. Agents preferring aesthetic quality to proximity to services dispersed more than those preferring proximity. Generalist agents achieved highest

average utility levels. Despite the fact that categorization had only weak effects on the results, the possibility remains that, in some models and for some populations, significant correlations structures among preferences might produce substantial effects from categorization.

Responses to this article can be read online at:
http://www.ecologyandsociety.org/volXX/issYY/artZZ/responses/

LITERATURE CITED

Alonso, W. 1964. *Location and land use: toward a general theory of land rent.* Harvard University Press, Cambridge, Massachusetts, USA.

Axelrod, R. 1997. *The complexity of cooperation: agent-based models of competition and collaboration.* Princeton University Press, Princeton, New Jersey, USA.

Batty, M., and P. Longley. 1994. *Fractal cities: a geometry of form and function.* Academic Press, San Diego, California, USA.

Benenson, I., and P. M. Torrens. 2004. *Geosimulation: automata-based modeling of urban phenomena.* Wiley, New York, New York, USA.

Block, D., and E. M. Dupuis. 2001. Making the country work for the city: von Thünen's ideas in geography, agricultural economics and the sociology of agriculture. *American Journal of Economics and Sociology* **60**(1):79-98.

Brown, D. G., S. E. Page, R. Riolo, and W. Rand. 2004. Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling and Software* **19**:1097-1109.

Brown, D. G., S. E. Page, R. Riolo, M. L. Zellner, and W. Rand. 2005. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science* **19**:153-174.

Chiang, A. C. 1984. *Fundamental methods of mathematical economics.* Mcgraw-Hill, Montreal, Canada.

Cioffi-Revilla, C., and N. Gotts. 2003. Comparative analysis of agent-based social simulations: GeoSim and FEARLUS models. *Journal of Artificial Societies and Social Simulation* **6**(4). Available online at: http://jasss.soc.surrey.ac.uk/6/4/10.html.

Cobb, C. W., and P. H. Douglas. 1928. A theory of production. *The American Economic Review* **18**(1):139-165.

Deadman, P. J., D. T. Robinson, E. Moran, and E. Brondizio. 2004. Colonist household decision making and land-use change in the Amazon rainforest: an agent-based simulation. *Environment and Planning* **31**(5):693-709.

Elster, J., and J. E. Roemer. 1991. *Interpersonal comparisons of well-being.* Cambridge University Press, Cambridge, UK.

Epstein, J. M., and R. Axtell. 1996. *Growing artificial societies: social science from the bottom up.* The Brookings Institution, Washington, D.C., USA.

Fernandez, L. E., D. G. Brown, R. W. Marans, and J. I. Nassauer. 2005. Characterizing location preferences in an exurban population: implications for agent based modeling. *Environment and Planning B* **32**:799-820.

Geoghegan, J. 2002. The value of open spaces in residential land use. *Land Use Policy* **19**:91-98.

Gilbert, N., and K. G. Troitzsch. 1999. *Simulation for the social scientist.* Open University Press, Philadelphia, Pennsylvania, USA.

Gimblett, R. H., editor 2002. *Integrating*

geographic information systems and agent-based modeling techniques for simulating social and ecological processes. Oxford University Press, New York, USA.

Golledge, R. G., and R. J. Stimson. 1987. *Analytical behavioral geography.* Croom Helm, London, U.K.

Holland, J. H., and J. H. Miller. 1991. Artificial adaptive agents in economic theory. *The American Economic Review* **81**(2):365-370.

Hong, L., and S. E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of USA* **101**:16385-16389.

Hurd, R. M. 1903. *Principles of city land values.* The Record and Guide, New York, New York, USA.

Irwin, E. G. 2002. The effects of open space on residential property values. *Land Economics* **78**(4):465-480.

Krugman, P. R. 1993. On the relationship between trade theory and location theory. *Review of International Economics* **1**(2):110-122.

Lomnicki, A., and S. Sedziwy. 1989. Do individual differences in resource intake without monopolization cause population persistence and stability? *Journal of Theoretical Biology* **136**:317-326.

Marans, R. W. 2003. Understanding environmental quality through quality of life studies: The 2001 DAS and its use of subjective and objective indicators. *Landscape and Urban Planning* **65**(1-2):75-85.

Marshall, A., C. W. Guillebaud, and Royal Economic Society (Great Britain). 1961. *Principles of Economics.* Macmillan for the Royal Economic Society, London, England, UK.

McGarigal, K., S. A. Cushman, M. C. Neel, and E. Ene. 2002. *FRAGSTATS: spatial pattern analysis program for categorical maps.* University of Massachusetts, Amherst, Massachusetts, USA. Available online at: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

Miller, J. E., D. J. Hunt, J. E. Abraham, and P. A. Salvini. 2004. Microsimulating urban systems.

Computers, Environment and Urban Systems **28**:9-44.

Murphy, J. J., P.G. Allen, T. H. Stevens, and D. Weatherhead. 2005. A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resources Economics* **30**:313-325.

Nassauer, J. I. 1995. Culture and changing landscape structure. *Landscape Ecology* **10**(4):229-237.

Otter, H. S., A. van der Veen, and H. J. de Vriend. 2001. ABLOoM: location behaviour, spatial patterns, and agent-based modelling. *Journal of Artificial Societies and Social Simulation* **4**(4). Available online at: http://jasss.soc.surrey.ac.uk/4/4/2.html.

Parker, D. C., and V. Meretsky. 2004. Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems, and Environment* **101**:233-250.

Parker, D. C., T. Berger, and S. Manson, editors. 2002. *Agent-based models of land-use and land-cover change. Report and review of an international workshop* (Irvine, California). LUCC Report Series No. 6, LUCC Focus 1 Office, Anthropological Center for Training and Research on Global Environmental Change Indiana University, Bloomington, Indiana, USA.

Parker D. C., S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman. 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers* **93**:314-337.

Rand, W., M. Zellner, S. E. Page, R. Riolo, D. G. Brown, and L. E. Fernandez. 2002. The complex interaction of agents and environments: an example in urban sprawl. Pages 149-161 *in* C. Macal and D. Sallach, editors. *Proceedings, Agent 2002: Social agents: ecology, exchange, and evolution* (Chicago, 2002). University of Chicago and Argonne National Lab, Chicago, Illinois, USA.

Rand, W., D. G. Brown, S. E. Page, R. Riolo, L. E. Fernandez, and M. Zellner. 2003. Statistical validation of spatial patterns in agent-based models. *In* J. P. Muller, editor. *Proceedings of Agent Based Simulation 4* (Montpellier, 2003). CIRAD, Montpellier, France.

Ricardo, D. 1821. *On the principles of political economy and taxation.* J. Murray, London, England, UK.

Sanders, L., D. Pumain, H. Mathian, F. Guérin-Pace, and S. Bura. 1997. SIMPOP: a multiagent system for the study of urbanism. *Environment and Planning B* **24**:287-305.

Sasaki, Y., and P. Box. 2003. Agent-based verification of von Thünen's location theory. *Journal of Artificial Societies and Social Simulation* **6**(2). Available online at: http://jasss.soc.surrey.ac.uk/4/4/2.html.

Sen, A. 1973. *On economic inequality.* Clarendon Press, Oxford, England, UK.

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

Schelling, T. C. 1978. *Micromotives and macrobehavior.* Norton, New York, New York, USA.

Thünen, J. H. V. 1826 *Die Isolierte Staat in Beziehung auf Landwirtshaft und Nationalökonomie.* Perthes, Hamburg, Germany.

Uchmanski, J. 2000. Individual variability and population regulation: an individual-based approach. *Oikos* **90**:539-548.

Waldrop, M. M. 1990. Asking for the moon. *Science* **247**(4943):637-638.

Weisbrod, G., Ben-Akiva, M., and S. Lerman. 1980. Tradeoffs in residential location decisions: transportation versus other factors. *Transportation Policy and Decision-Making* **1**(1):13-16.

Appendix 1. SOME Model Description and Pseudo Code

Please click here to download file 'appendix1.pdf'.

Appendix 2. Modified Cobb-Douglas Utility Function

Please click here to download file 'appendix2.pdf'.