AGENT-BASED RESIDENTIAL SEGREGATION: A HIERARCHICALLY STRUCTURED SPATIAL MODEL[1]

David O'Sullivan, The Pennsylvania State University, University Park[*]
James Macgill, The Pennsylvania State University, University Park
Chaoqing Yu, The Pennsylvania State University, University Park

ABSTRACT

In this paper we present a variation on Schelling's model of residential location dynamics that combines two concepts of neighborhood: the continuous neighborhood and the bounded neighborhood. Schelling's model is described in outline. The structure of the current model is described in detail. The effect on model behavior of varying the size of bounded neighborhoods while also varying the balance between local and regional level effects on agent behavior is explored, and preliminary results are reported. The scale of bounded neighborhoods considered by agents in making residential location decisions has important impacts on overall model outcomes. A range of possibilities for further work is discussed.

INTRODUCTION: SCHELLING'S MODELS OF RESIDENTIAL SEGREGATION

Schelling's simple model of residential segregation dynamics (Schelling 1969, 1971, 1978) is rightly regarded as a seminal example of multi-agent simulation in social science (Macy and Willer 2002). In the fifteen years from 1988 to 2002, Schelling's 'Dynamic models of segregation' (1971) has been cited 125 times, 70 of these occurring from 1999 to 2002 (ISI 2003). The model's persistent popularity derives from its simplicity, and its compelling demonstration of the emergence of stable, aggregate, socio-spatial patterns from local interactions between household agents. In Schelling's model, households of two types, make decisions to remain at or leave their current residential location depending on dissatisfaction with that location. Dissatisfaction arises from a household having too many neighbors of the opposite type, or too few neighbors of its own type. Using this framework, Schelling shows that strongly segregated large-scale residential patterns can arise even when two groups are relatively tolerant of one another's presence.

Neighborhoods are conceived in two different ways in Schelling's work: *continuous neighborhoods* and *bounded neighborhoods*.

In the *continuous neighborhood* case, households occupy locations on a lattice or 'checkerboard' (Sakoda 1971) and decisions are made with regard to the types of households in adjacent locations on this lattice. Agents demand that some fraction of their immediate

---

[1] Please cite as O'Sullivan, D., J. R. Macgill and C. Yu. 2003. Agent-based residential segregation: a hierarchically structured spatial model. Presented at Agent 2003 'Challenges in Social Simulation', October 2-4, University of Chicago and Argonne National Laboratory, IL.

[*] *Corresponding author address*: David O'Sullivan, The Pennsylvania State University, Department of Geography, 302 Walker Building, University Park, PA 16802. e-mail: dbo2@psu.edu

neighbors on the lattice is of the same type as themselves. Agents that are unhappy under these criteria move to a nearby location where their residential preference requirements are satisfied. Unsurprisingly, when households demand many neighbors similar to themselves, the result is dramatic segregation of the lattice into large regions occupied exclusively by households of only one type. Schelling assesses the extent of segregation by counting the average number of like neighbors in the final stable pattern that results (Schelling 1971, 157–8). He finds that “the resulting segregation [is] a rapidly rising function of demands [for like neighbors] in the range from about 35% to 50%.” (Schelling 1971, 159). This occurs, because increasing demand for like neighbors leads to increased discontent in random initial patterns, to a greater likelihood that moving households will displace previously content households, and also to increased concentration of households in regions of the checkerboard already densely populated by other households.

Schelling also investigates the effect of different combinations of household tolerance profiles in two populations, with respect to *bounded neighborhoods*. A bounded neighborhood is a ‘container’ populated by a number of households. Schelling uses the bounded neighborhood concept to analyze the relationship between residential preferences in the population, and the stability properties of different combinations of numbers of households of each type in a *single* bounded neighborhood in isolation. By analyzing plausible—albeit hypothetical—tolerable ratios between two groups, he demonstrates that the only stable states in many cases are exclusive neighborhoods where all the residents are from one group or the other.

Comments on Schelling’s models

A number of aspects of these models deserve comment in the current context.

First, in the continuous neighborhood case, household behavior is governed by a demand for like neighbors, whereas in the bounded neighborhood case, antipathy toward different neighbors is the driving force. Although these mechanisms can be combined, it is difficult to do so without introducing numerous arbitrary parameters (demand for like, tolerance of different, and so on). In the present model, we have adopted the tolerance/antipathy approach for household behavior at both local and regional scales. Thus, it is the presence of too many households of a different type whether locally or in a larger bounded neighborhood that causes household decisions to relocate.

Schelling’s experiments in the continuous neighborhood case seem to have been conducted by hand, although this is not clear. The most important effect of this on the operation of the model is vagueness about the order in which households are considered for relocation. Thus, in Schelling’s description of the rules of movement, he says, “Identify the discontents […] and, *in some order* move them to where they are content.” (Schelling 1971, 156). This is followed, almost immediately, by an acknowledgement that this makes a difference in detail, but not in general: “The *particular* outcome will depend very much on the order in which discontented [households] are moved, the *character* of the outcome not very much.” (Schelling 1971, 156)

Similar comments apply to vagueness in the rules of movement such that a household relocates to the “nearest” vacant spot with a neighborhood that is acceptable. Vagueness on these points makes it impossible to replicate Schelling’s simulations in a computational simulation. We therefore use random ordering both of household relocation decisions, and of consideration of equidistant vacant locations to minimize effects that seem likely to arise from any more structured sequencing of relocation events.

The continuous neighborhood formulation of Schelling's work has been widely acknowledged in the multi-agent social simulation community. This is perhaps because this approach is suggestive of common devices in contemporary multi-agent work, particularly the grid-based space in which agents interact (see, for example, the Sugarscape model, Epstein and Axtell 1996). The continuous neighborhood approach is also consistent with notions from complexity science about the efficacy of purely local interactions in producing larger global structures.

Schelling's work is extremely insightful and thought provoking. The important finding in the bounded neighborhood case that stable racially integrated neighborhoods are unlikely for many combinations of tolerance profiles has been confirmed based on empirical data (Clark 1991).

COMBINING CONTINUOUS AND BOUNDED NEIGHBORHOODS IN A HIERARCHICAL MODEL

We present a hierarchical version of the Schelling model that combines his two neighborhood types. Household agents consider the type of immediately neighboring households in a lattice of residential locations (the *continuous* or *local neighborhood*), but also consider the aggregate nature of the *bounded neighborhood* (or *district*) containing their residential location. The model consists of a number of bounded neighborhoods, each containing a number of residential locations at points on a lattice. This structure is illustrated in figure 1.

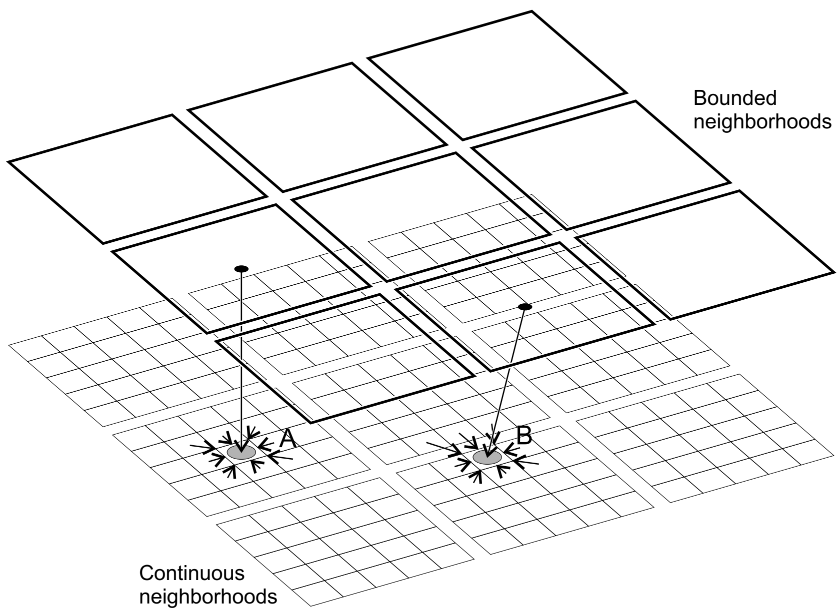

FIGURE 1 Structure of the hierarchical Schelling model. The small grid cells represent residential locations with a *continuous neighborhood* structure. Residential locations are contained in *bounded neighborhoods*, whose aggregate state is also considered by agents in residential decision-making behavior. Arrows show relations of influence on decision-making.

Decisions of agent A, at the center of a bounded neighborhood are affected by immediately neighboring agents, and by the aggregate state of the containing bounded neighborhood. An agent B, at the edge of its containing bounded neighborhood, is affected by neighbors both inside and outside the bounded neighborhood, but only by the containing bounded neighborhood at the aggregate level.

As far as possible, we retain the simplicity of Schelling's formulation of households' rules of movement. In assessing the level of 'happiness' with a current or potential location, household agents determine *local happiness*, h_L and *regional happiness*, h_R. This assessment is made with respect to a single agent parameter called *tolerance*, T, which varies from agent to agent. Tolerance is a real number between 0 and 1, indicating the fraction of occupied neighboring locations whose agents may be of a different type from the agent without negatively affecting its happiness, and prompting it to seek an alternative location. Local happiness is the difference between an agent's tolerance and the fraction of occupied locations in the agent's continuous neighborhood occupied by agents of a different type. Similarly, regional happiness is the difference between agent tolerance and the fraction of occupied locations in the agent's bounded neighborhood occupied by agents of a different type. Formally, an agent, A_i, has tolerance $T(A_i)$, and type $t(A_i)$, where

$$0 \le T(A_i) \le 1 \tag{1}$$

and

$$t(A_i) \in \{\text{RED}, \text{BLUE}\} \tag{2}$$

Color is a convenient visual marker for agent type, but any discrete valued variable will suffice.

If we denote the set of agents in the local (continuous) neighborhood by N_L, and the set of agents in the district (or region or bounded neighborhood) by N_R, then we can determine local and regional happiness for the agent from

$$h_L(A_i) = T(A_i) - \frac{\left\| \{A_j : A_j \in N_L \wedge t(A_j) \ne t(A_i)\} \right\|}{\|N_L\|} \tag{3}$$

and

$$h_R(A_i) = T(A_i) - \frac{\left\| \{A_j : A_j \in N_R \wedge t(A_j) \ne t(A_i)\} \right\|}{\|N_R\|} \tag{4}$$

It is a simple matter to combine happiness values using a single model-wide parameter *local-regional balance*, b_{LR}, to determine an overall happiness h for the agent, according to

$$h(A_i) = (1 - b_{LR})h_L(A_i) + b_{LR}h_R(A_i) \tag{5}$$

where b_{LR} = 0 results in agent happiness depending only on agents in the continuous neighborhood, while b_{LR} = 1 means that happiness is dependent only on agents in the same bounded neighborhood.

Overall, an agent's rule of movement is to determine overall happiness, based on neighboring agent types both locally and in its bounded neighborhood. If the agent has an overall negative happiness score, then it is unsettled and tries to relocate. This involves examining available vacant locations at successively greater distances in the lattice until one is found where the agent happiness score would be positive. Potential locations at the same distance from the current location are considered in random order to ensure no directional bias in agent relocation. As soon as a suitable location is found, the agent moves to that location. It is possible that no suitable location is available, in which case the agent does not resettle.

Agents are considered for relocation one at a time in random order in one 'sweep' through the agent population. Each sweep of the population occurs in a different random order.

Implementation details

The model described above was implemented in the *RePast* agent modeling toolkit (Collier no date). Repast provides a simple bridge to the *GeoTools* open source package for displaying and analyzing geographically referenced datasets, and given our interest in exploring the impact of geographic perceptions on models of segregation behavior, this was a natural choice.

To facilitate future investigation of more complex spatial patterns, the model's geographic structure is initialized by reading two geographic information system (GIS) files, one representing residential locations in the continuous neighborhood layer, and a second representing bounded neighborhoods. Geographical processing is applied to determine both the continuous neighborhood relations among residential locations (i.e., the lattice structure), and the nesting of residential locations in the continuous neighborhood layer within containing bounded neighborhoods. The resulting neighborhood relationships are stored in a graph data structure that records *adjacency* relations between locations, and *containment* relations between bounded neighborhoods and locations. This allows agents to retrieve information about the numbers of agents of their own or opposite type in their continuous neighborhood, and in the bounded neighborhood. This approach will allow future investigation of model dynamics with irregularly shaped locations and districts, although such examples are not considered in this paper.

RESULTS

Model input parameters

The values of the various model parameters used in the reported experiments are summarized in table 1.

TABLE 1 Summary of model parameter settings

Parameter name	Settings used	Comment
Local-regional balance	0.0, 0.1, 0.2, … 0.9, 1.0	Varied through full range to study effect of variation in local *versus* regional behavior
Bounded neighborhood sizes	4×144 locations 9×64 locations 16×36 locations 36×16 locations	Varied to study impact of different bounded neighborhood sizes on behavior
Tolerance	0.2 to 0.4	Agents initialized from a random uniform distribution
Occupancy rate	0.75	Fixed
Fraction blue	0.5	Fixed

Model output or measurement parameters

A number of summary statistics are used to track model progress.

The *dissimilarity index* D is reported with respect to the set of bounded neighborhoods. D is a measure of residential segregation for population count data reported for zones, which indicates the extent to which two population groups are not similarly distributed among the zones (Duncan and Duncan 1955, Taeuber and Taeuber 1965, Taeuber and Taeuber 1976). Given two population groups with total populations R and B, the counts of the groups living in each zone i may be denoted r_i and b_i. These values are combined across all n zones, to give

$$D = 0.5\sum_{i=1}^{n}\left|\frac{r_i}{R} - \frac{b_i}{B}\right| \tag{7}$$

D has value 0 if two populations are distributed identically across a set of zones. It has a value of 1, if they are completely segregated, that is, if all blues are located in zones that contain no reds, and *vice versa*.

Because D is calculated with respect to a set of bounded neighborhoods, it is possible even when D indicates little segregation, for agents to be locally segregated such that agents have neighbors in their continuous neighborhood predominantly of the same type as themselves. Such local segregation is measured using an *average fraction of like neighbors* statistic, S_L (for locally similar). For each agent, the fraction of occupied neighboring locations whose occupying agents are of the same type is averaged across all agents.

Both D and S_L are pattern measures calculated at any point during a model run. The remaining two output parameters are cumulative measures of model dynamics over each sweep through all the agents. The *fraction of agents unsettled*, p_U, and *fraction of agents resettled*, p_R, record respectively, the fraction of all agents in the model that were unsettled and tried to relocate, and the fraction of all agents in the model that successfully resettled, during a sweep of the whole agent population. Note that p_R is less than p_U by definition, since only unsettled agents attempt to relocate.

Final stable patterns

Initially, we observe final stable patterns in the model to see how different are outcomes relative to the patterns of Schelling's continuous neighborhood case. As expected, with the local-regional balance parameter set to 0, outcomes are identical to Schelling's examples (see figure 2), and the bounded neighborhoods make no difference.

When the local-regional balance is increased to 0.5 final configuration is reached, illustrated in figure 3 for two different sets of bounded neighborhoods. Agent responses to the composition of bounded neighborhoods result in a bimodal distribution of bounded neighborhoods—either predominantly red or predominantly blue. The change in agent priorities also means that dissimilar agents may be tolerated as immediate neighbors in the continuous neighborhood, along boundaries between bounded neighborhoods with different majorities. Examples of the opposite effect are also apparent: single isolated agents of the 'wrong' type are found in some bounded neighborhoods because the tolerable (empty) configuration of their continuous neighborhood allows them to ignore the majority of unlike agents in the bounded neighborhood.

Other settings of the local-regional balance parameter result in different balances in the outcome patterns between the tendency to local segregation on the one hand, and to bounded neighborhood segregation accompanied by 'tolerance' for unlike neighbors across district boundaries on the other, as illustrated in figure 4.

When the local-regional balance is set to 1 (completely regional), greater variation in outcomes is observed as discussed in more detail in the next section.

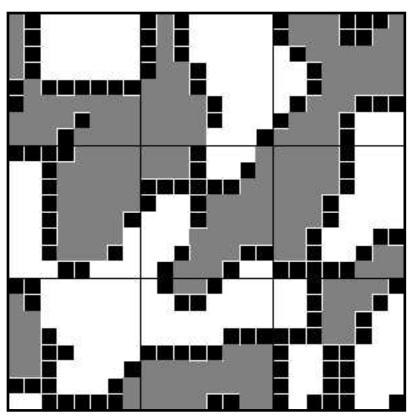

FIGURE 2 Typical outcome with the local-regional balance parameter set to 0 (fully local). Agent states are shown in gray and white. Black locations are vacant.

FIGURE 3 Two outcomes with the local-regional balance parameter set to 0.5. The left-hand case shows a 3-by-3 grid of bounded neighborhoods, each with 64 locations; the right-hand case shows a 4-by-4 grid of bounded neighborhoods, each with 36 locations.

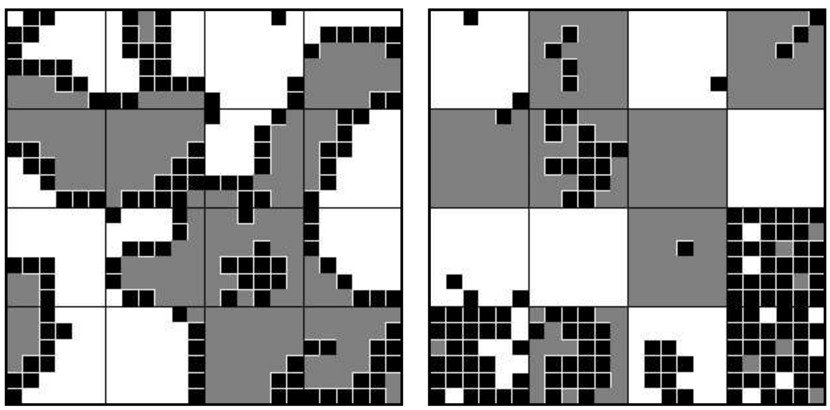

FIGURE 4 Typical outcomes with a 4-by-4 grid of districts and the local-global balance parameter set to 0.25 (left-hand case) and 0.9 (right-hand case). Note that all the images in figures 2 through 4 are based on the same initial random number generator seed setting of 1061992058618.

Varying bounded neighborhood size and local-regional balance

In this section we report preliminary findings from experiments where the size of bounded neighborhoods was varied and the local-regional balance parameter was varied. Results are shown in figures 5 through 8 as the local-regional parameter is varied from 0 to 1 in increments of 0.2.

The first point to make about these figures is that there is considerable continuity in the model behavior through all the results shown. The dominant behavior is for the model to segregate, and to do so rapidly. When the model stabilizes, agents are (usually) no longer unsettled, and are content to stay where they are. Segregation behavior occurs in almost all cases.

Two differences are observed as the local-regional balance parameter is increased. First, the final stable state exhibits patterns that are increasingly segregated as measured by the dissimilarity index, and decreasingly segregated as measured by the average fraction of like neighbors. This is the phenomenon already noted above whereby increasing emphasis on the bounded neighborhood allows agents to have dissimilar immediate neighbors across district boundaries. This is a direct result of the presence of bounded neighborhoods 'steering' local segregation to fit inside the boundaries, so that a higher dissimilarity index is observed.

Second, as the local-regional parameter increases, the time taken for the model to stabilize decreases. Thus in the first row of all diagrams with b_{LR} set to 0 or 0.2 there is relocation activity over about four sweeps of the agent population; in the second row (b_{LR} equal to 0.4 or 0.6) relocation activity occurs for only around three sweeps; and, in the third row (b_{LR} equal to 0.8 or 1.0) stabilization happens after only two sweeps of the population, in most cases. This is a result of the combination of bounded neighborhoods and regional level behavior causing agents to relocate to districts that are tipping into a state where all agents are of the same type. Once settled in such locations, agents will not move again. When more local considerations are dominant, it is possible for an agent to initially resettle in a location that is locally congenial, but which subsequently becomes less desirable as the bounded neighborhood starts tipping into the opposite type of agent.

Clearly, this relatively neat picture of the model's behavior breaks down in the last plot on figure 6 (with a 36-location bounded neighborhood, and $b_{LR} = 1.0$), and is similarly inadequate for high values of b_{LR} in both the 64- and 144-location bounded neighborhood cases (figures 7 and 8). With these combinations of settings, wide variation in outcomes across the sets of random runs is evident. In a significant fraction of cases the model gets stuck in a configuration where bounded neighbors are incompletely sorted so that the dissimilarity index D is not near 1. In these situations, many household agents remain unsettled, but are unable to find preferable locations and so do not resettle. At present it is unclear if any statistic anticipates this outcome, although large numbers of unsettled households failing to resettle during the first agent population sweep is a promising candidate predictor.

Note that the last plot in figure 8 with a large bounded neighborhood (144 locations) and the local-regional parameter set to 1, exhibits the most extreme form of this behavior where no relocation is seen at all. This is because agents only care about bounded neighborhood states, and with only four neighborhoods to choose from, there is a strong probability that randomly initialized bounded neighborhoods will be judged the same. Large bounded neighborhoods make all model locations effectively the same, so that little or no relocation is observed even though virtually all agents are unsettled.

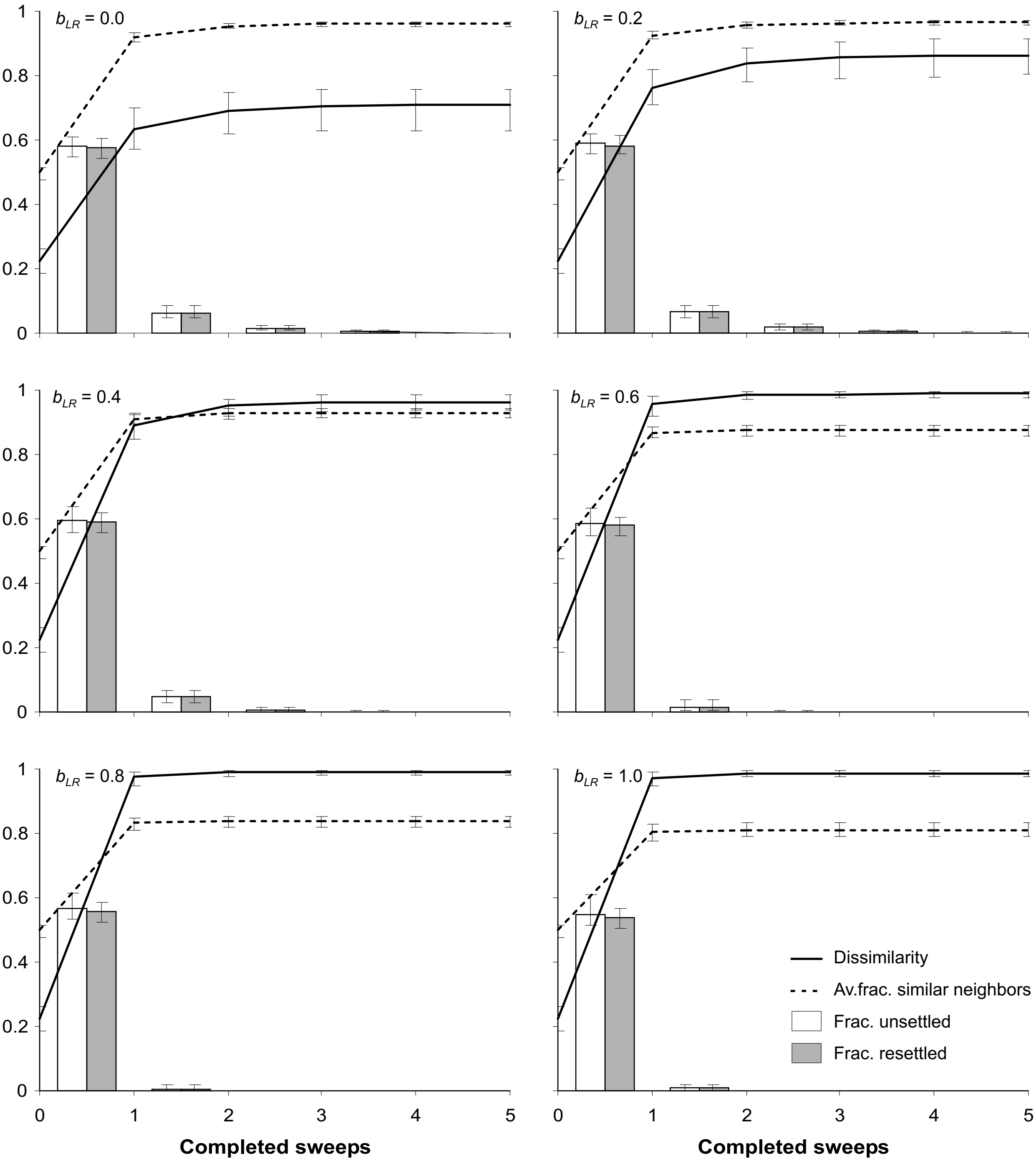

FIGURE 5 Summary results for 16-location bounded neighborhoods as local-regional balance is varied from 0 to 1 in steps of 0.2. Pattern measures (dissimilarity and average fraction of similar neighbors) are shown as line graphs with values recorded at the end of each sweep through all agents. Dynamic summary measures (fraction of agents unsettled and fraction resettled) are shown by bars, and record these values summed over a sweep through all agents. All four statistics are include an 'error bar' indication of the range of values between the 10^{th} and 90^{th} percentile over 100 randomly generated runs.

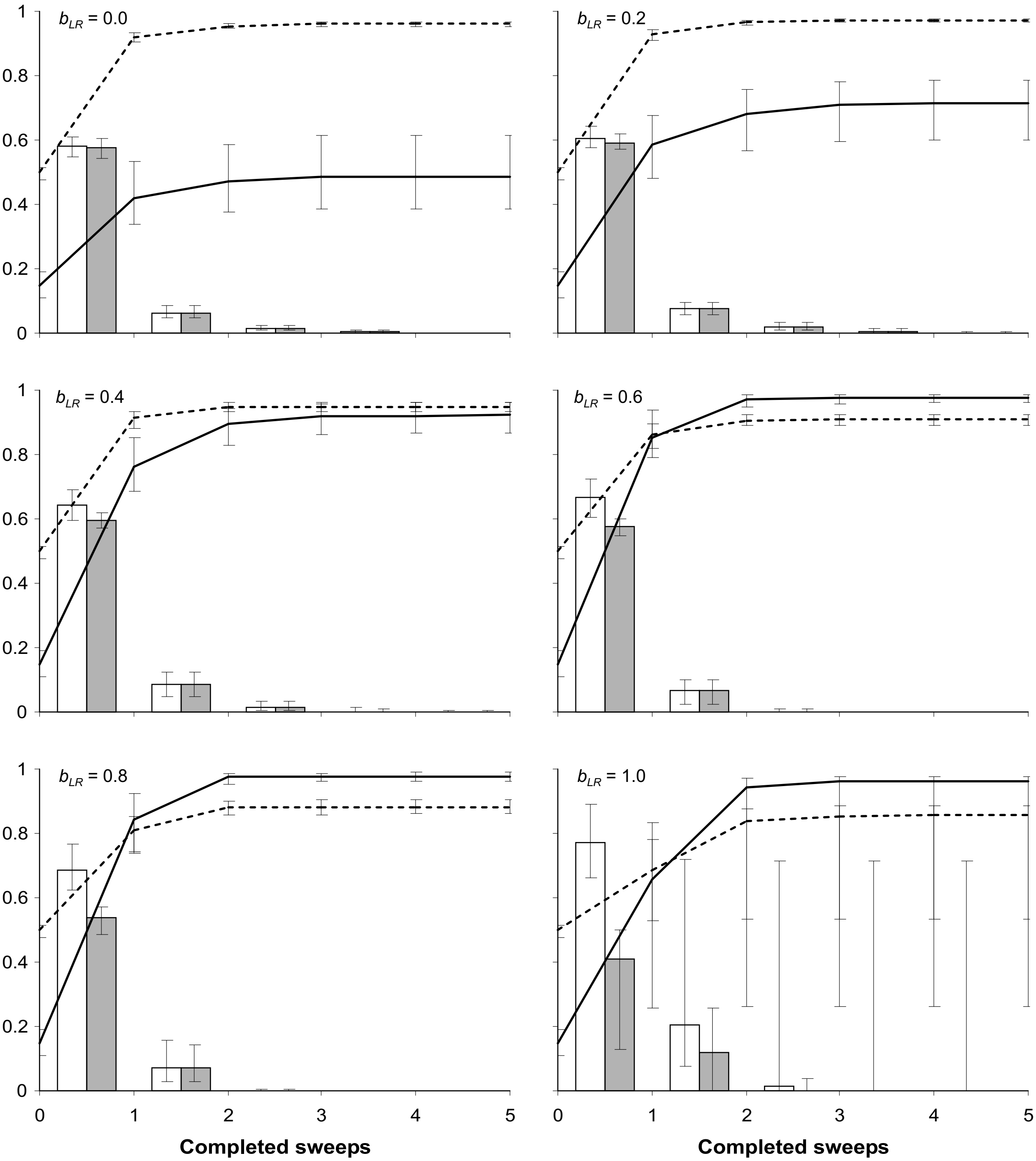

FIGURE 6 Summary results for 36-location bounded neighborhoods as local-regional balance is varied from 0 to 1 in steps of 0.2.

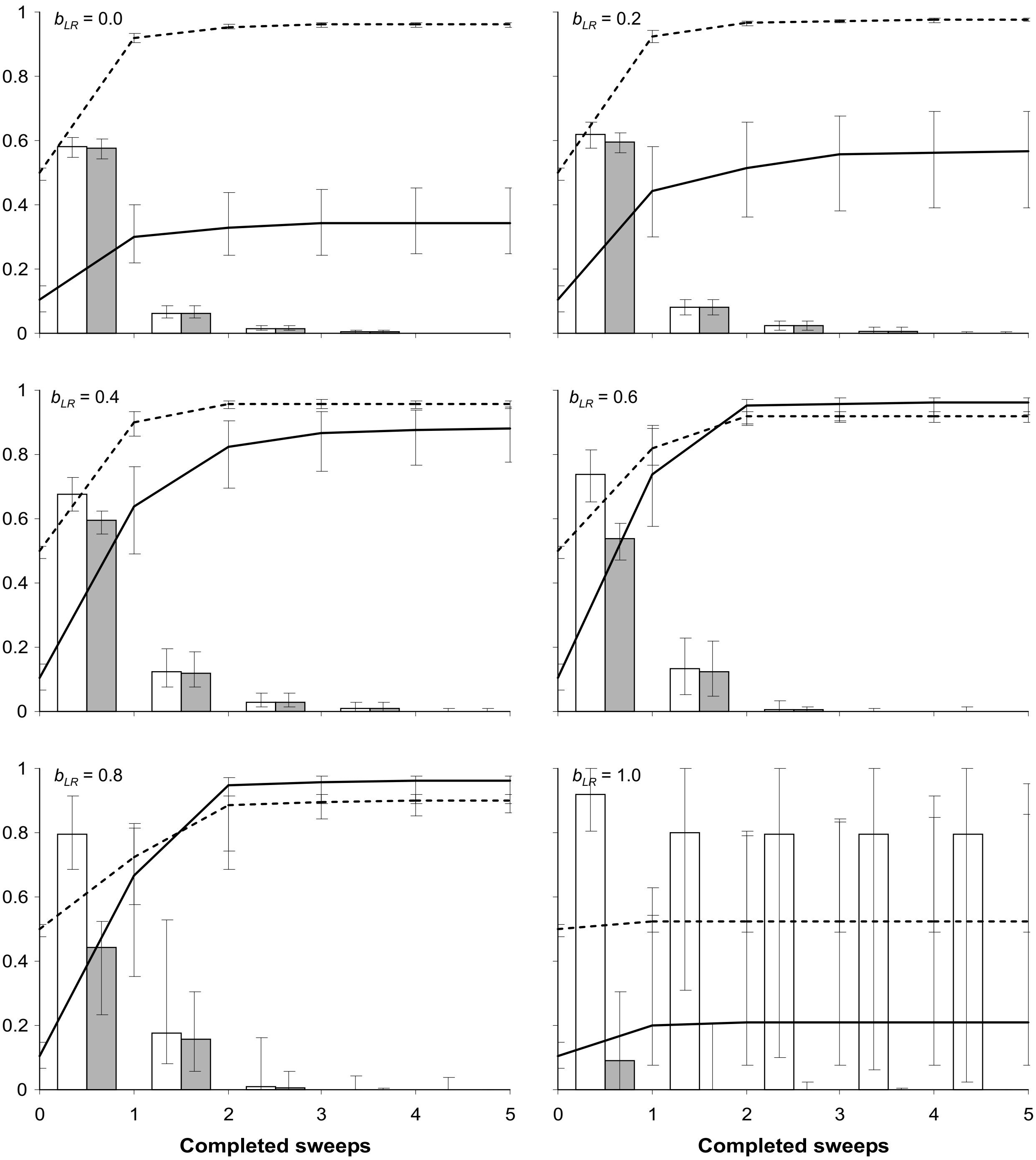

FIGURE 7 Summary results for 64-location bounded neighborhoods as local-regional balance is varied from 0 to 1 in steps of 0.2.

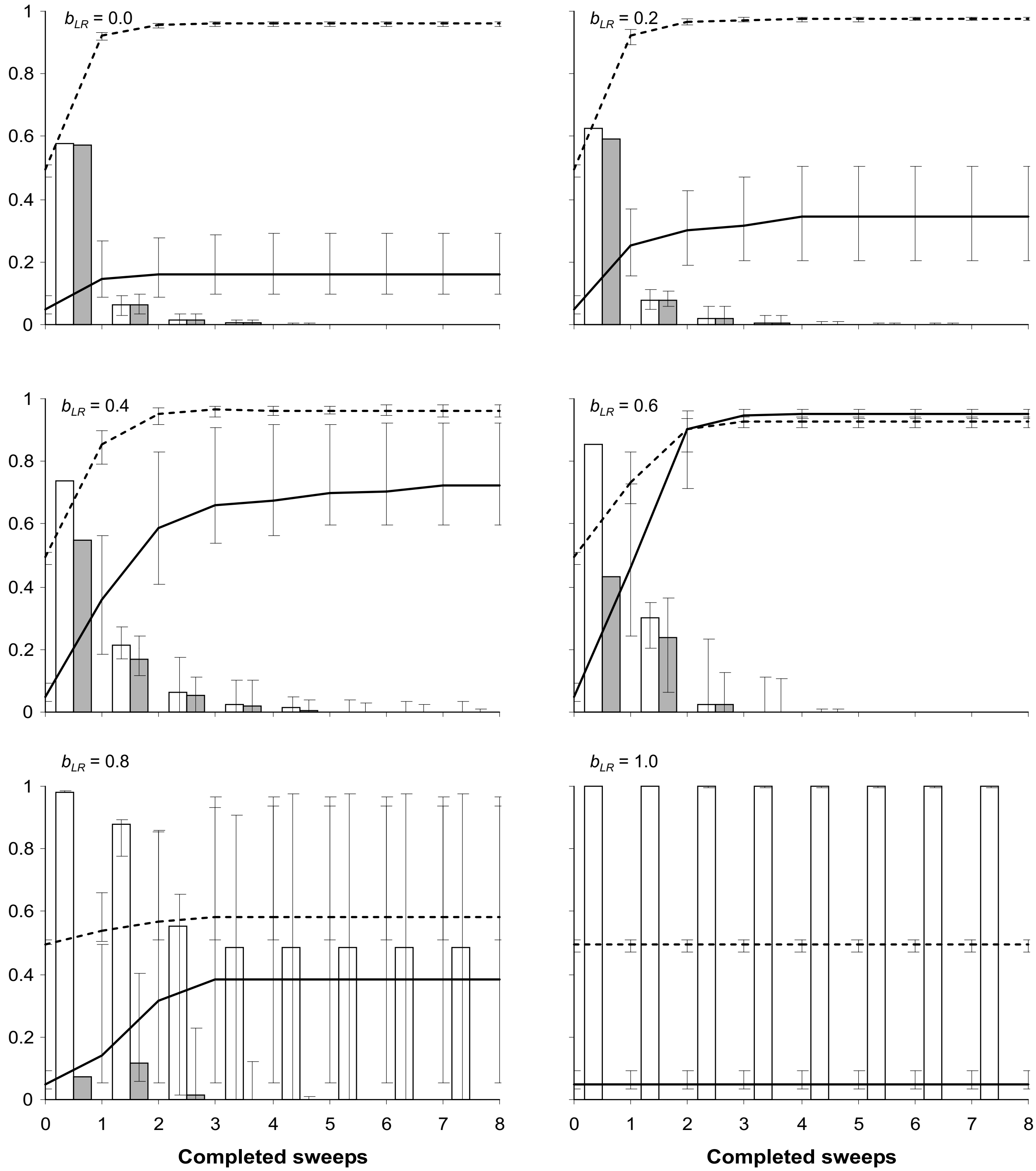

FIGURE 8 Summary results for 144-location bounded neighborhoods as local-regional balance is varied from 0 to 1 in steps of 0.2. Note the longer time sequences in these cases. Also note that these results are based on running only 50 random sequences.

CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

From the foregoing discussion, we conclude that both bounded neighborhoods and variation in agent local-regional behavior have significant effects on Schelling-type model dynamics and on the resulting stable spatial patterns that are observed.

Small bounded neighborhoods have relatively limited impact on the model, except to alter the details of final stable patterns if agents are attentive to bounded neighborhood effects. However, as agent attention to bounded neighborhoods is increased, the speed with which the model stabilizes increases, as a result of preferential movement into neighborhoods that are tipping into exclusive occupation by one or the other type of agent. This suggestion could be partially confirmed by measuring the average distance moved by relocating agents, to see if agents move further as the local-regional balance parameter is increased.

For larger bounded neighborhoods it is possible for the model to get stuck in a configuration where agents are unsettled but are unable to relocate because no alternative location is judged preferable. This appears to be a result of initial preferential relocation into tipping neighborhoods leaving unsettled agents with a choice of locations in the remaining neighborhoods, which are all judged similar to one another. These effects occur only when the local-regional balance is tilted toward regional effects, because no local preferences enable habitable 'niches' to be established by a series of local relocations.

Clearly, this model is extremely abstract, so that interpretation of these findings is tricky. Perhaps the most useful way to think about the results is in terms of communication processes among agents. When local behavior is dominant segregation is slower (but surer) as agents only attend to nearby locations and local niches can be established gradually that enable eventual complete segregation. When regional scale behavior is dominant, segregation is more rapid (but less sure) because preferred new locations are rapidly identified. However, depending on the scale—the bounded neighborhood size—over-attention to only larger scale neighborhoods can prevent segregation from occurring completely.

Seen in this way, the model seems likely to be a useful vehicle for exploration of the important role of information in residential location decision making. In the current version of the model, bounded neighborhoods are fixed, but it is intended that this limitation will be removed in future developments, so that the dynamics of emerging local property markets can be explored. Modeling of the behavior of other agents operating at different spatial scales in the residential location context (realtors and banks, in particular) is also planned. In fact, in the current implementation agent relocation is handled by a 'global realtor' class, to facilitate exploration of these aspects in revised versions of the model.

Remaining with the theme of further software developments, abstraction from the current model to a more general class of geographical agent models using the *RePast* architecture is planned. The difficulty of understanding the behavior of even this relatively simple model may be greatly reduced in future by closer integration with dynamic visualization environments. This is a direction we intend to pursue in the medium term.

REFERENCES

Clark, W. A. V. 1991. Residential Preferences and Neighborhood Racial Segregation - a Test of the Schelling Segregation Model. *Demography*, **28**(1) 1-19

Collier, N. no date. RePast: An Extensible Framework for Agent Simulation. Social Science Research Computing, University of Chicago http://repast.sourceforge.net/docs/repast_intro_final.doc [accessed June 24, 2002].

Duncan, O. D. and B. Duncan. 1955. A methodological analysis of segregation indexes. *American Sociological Review*, **20** 210-17

Epstein, J. M. and R. Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Press & MIT Press: Washington, DC.

ISI. 2003. Web of Science. [accessed May 24, 2003,

Macy, M. W. and R. Willer. 2002. From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology,* **28** 143-66

Sakoda, J. M. 1971. The checkerboard model of social interaction. *Journal of Mathematical Sociology,* **1** 119-32

Schelling, T. C. 1969. Models of segregation. *American Economic Association Papers and Proceedings*, **59**(2) 488-93

Schelling, T. C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology*, **1** 143-86

Schelling, T. C. 1978. *Micromotives and Macrobehavior*. Norton: New York.

Taeuber, K. E. and A. F. Taeuber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change*. Aldine Publishing: Chicago, IL.

Taeuber, K. E. and A. F. Taeuber. 1976. A practitioner's perspective on the index of dissimilarity. *American Sociological Review,* **41** 884-89