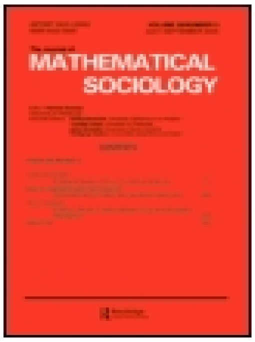

Journal of Mathematical Sociology

ISSN: 0022-250X (Print) 1545-5874 (Online) Journal homepage: https://www.tandfonline.com/loi/gmas20

A DYNAMIC MODEL OF RESIDENTIAL SEGREGATION

Junfu Zhang

Journal of Mathematical Sociology, 28: 147–170, 2004

ISSN: 0022-250X print/1545-5874 online
DOI:10.1080/00222500490480202

A DYNAMIC MODEL OF RESIDENTIAL SEGREGATION

Junfu Zhang
Public Policy Institute of California
San Francisco, California, USA

Using dimes and pennies on a checkerboard, Schelling (1971, 1978) studied the link between residential preferences and segregational neighborhood patterns. While his approach clearly has methodological advantages in studying the dynamics of residential segregation, Schelling's checkerboard model has never been rigorously analyzed. We propose an extension of the Schelling model that incorporates economic variables. Using techniques recently developed in stochastic evolutionary game theory, we mathematically characterize the model's long-term dynamics.

Keywords: Potential game, Stochastic stability, Agent-based simulation, Residential segregation

1. INTRODUCTION

Using dimes and pennies on a checkerboard, Schelling (1971, 1978) studied how residential preferences at the individual level were translated into segregational neighborhood patterns at the aggregate level. He demonstrated that there was no one-to-one correspondence between individual preferences and neighborhood configurations. In particular, Schelling showed that modest racial preferences of individuals could be amplified into high degrees of residential segregation due to dynamic feedback effects.

Schelling's checkerboard model was one of the earliest examples of what today would be called an agent-based model (Epstein and Axtell, 1996). A typical agent-based model consists of a large number of agents with heterogeneous preferences and locations. It begins in out-of-equilibrium

I would like to thank Peyton Young for his guidance. I am also grateful to Carl Christ, Bruce Hamilton, Joe Harrington and seminar participants at the Brookings Institution and the Santa Fe Institute for helpful comments.

Address correspondence to Junfu Zhang, Public Policy Institute of California, 500 Washington Street, San Francisco, CA 94111. E-mail: zhang@ppic.org, Tel: (415) 291-4454.

conditions and agents sequentially optimize based on local conditions. In the Schelling model, one person's move generates externalities that induce other people to move. This results in a dynamical system with feedback effects. Although the model is well-known and its behavior has been studied using computer simulations (see, e.g., Chapter 6 in Epstein and Axtell, 1996), it has yet to be analyzed rigorously.

Young (1998) was the first to recognize that the theory of stochastic dynamical systems could be employed to analytically study the Schelling model. In his book, Young proposed a simple segregation model on a one-dimensional circle, and pointed out that residential segregation patterns are more frequently observed in Schelling-type simulations because they are "stochastically stable."

In this paper, we analyze the two-dimensional case, which is considerably more complex. We also extend Schelling's framework substantially by incorporating an endogenously determined price for housing. Using techniques recently developed in evolutionary game theory (Foster and Young, 1990; Blume, 1993; and Young, 1998), we characterize the model's long-run dynamic properties. We show that even a slight asymmetry in residential preferences between the two groups is enough to induce endogenous segregation.

In addition to our mathematical results, we run agent-based simulations to illustrate the evolutionary dynamics of such a process. Taking the typical residential pattern of the 1960s as a starting point, our model is able to generate patterns of segregation and changes in racial housing price and vacancy rate differentials that are broadly consistent with trends in urban areas in the United States over the past four decades.

The remainder of the paper is organized as follows. Section 2 presents the dynamic model. Section 3 uses agent-based simulations to explore the dynamics of segregation and show how the model relates to recent trends. Section 4 concludes with some remarks. Mathematical proofs are given in the Appendix.

2. A DYNAMIC MODEL

2.1 Basic Setup

Consider a lattice graph with a periodic boundary condition, i.e., a torus. Each vertex is thought of as a residential location. Given any location i, its neighboring locations are defined as the H vertices around i. Here H is a fixed integer. For example, in agent-based computational models, the most commonly used definitions of neighborhood are the "Von Neumann neighborhood" and the "Moore neighborhood." In a Von Neumann neighborhood on a two-dimensional lattice, an agent considers the 4 immediately adjacent agents as neighbors, and hence $H = 4$. In a Moore neighborhood, an agent's neighbors include eight surrounding agents, and therefore $H = 8$ (see Epstein and Axtell, 1996, p. 40). The results of this model are independent of any specific definition of neighborhood. All we need is the "local" property of neighborhood. That is, neighborhood must be considerably smaller than the global area.

A residential location may be occupied by a black agent, a white agent, or may simply be vacant. The quality of housing is identical for each location; the only difference is the endogenous price.

Let P_i be the price of housing (rent) at location i, W_i the number of white neighbors of location i, and B_i the number of black neighbors of location i. We assume that every agent earns the same income, Y. This is an innocuous assumption because even if each person i earns a different Y_i, all results remain the same.

The price of housing is determined by a simplified "market mechanism" by which prices respond to excess demand. There is a similarity between the labor market and the housing market. In the labor market, a fraction of people are unemployed at any given time even though there are job vacancies. This so-called "natural unemployment rate" is necessary for job turnover. Similarly, in the housing market, there are nearly always vacant housing units available, which facilitates housing-market turnover. The idea of a "natural vacancy rate" has been proposed for the housing market (Blank and Winnick, 1953; Rosen and Smith, 1983). We define our pricing mechanism based on this hypothesis.

We posit a natural vacancy rate v^* that is just enough to facilitate "normal" housing market turnover, and let v_i be the actual vacancy rate of the neighborhood where housing unit i is located. In our model, it is natural to assume that all local markets are governed by the same v^*. For simplicity, we consider a linear pricing rule $P_i = a - b(v_i - v^*)$, where a and b are positive constants. It simply says that if location i is in a neighborhood where the actual vacancy rate is high, then its price will be low.[1]

Remember, any agent has H neighboring housing units. Thus the total number of housing units in a neighborhood is $H + 1$. Define V^* as the number of vacant units such that $\frac{V^*}{H+1} = v^*$. Then we can rewrite the pricing rule as $P_i = a - b(v_i - v^*) = a - b\left(\frac{H-B_i-W_i}{H+1} - \frac{V^*}{H+1}\right)$.[2] Rearranging terms,

[1]This corresponds to the "wage curve" in the labor market, which describes the negative relationship between wages and local unemployment (see, for example, Blanchflower and Oswald, 1994).

[2]Writing $H - B_i - W_i$ as the total number of vacancies, we are assuming that location i is always considered occupied when its price is calculated. We need this assumption to avoid "dynamic inconsistency". Without this assumption, location i commands different prices before and after an agent moves in, which seems odd. This assumption is innocuous especially when H is large.

we have $P_i = \frac{a(H+1)-b(H-V^*)}{H+1} + \frac{b(B_i+W_i)}{H+1}$. By choosing a proper currency unit, we can multiply the righthand side of the previous equation by a constant $\frac{H+1}{b}$ and get $P_i = \frac{a(H+1)-b(H-V^*)}{b} + (B_i + W_i) = c + (B_i + W_i)$. Notice here $c = \frac{a(H+1)-b(H-V^*)}{b}$ is a constant, which is the portion of the housing price that every agent has to pay no matter where he/she lives. Therefore, a model where every agent earns Y and pays $P_i = c + (B_i + W_i)$ is equivalent to a model in which every agent earns $Y - c$ and pays $P_i = B_i + W_i$. That is, it is innocuous to normalize c to 0. As we proceed, we will work with the simplest form of linear pricing without loss of generality:

$$P_i = B_i + W_i \tag{1}$$

There is no legal restriction on any agent's residential choices, but all else being equal, a white agent is assumed to prefer to live near white agents rather than near black agents. Since housing is homogeneous in terms of quality, only the price of housing enters the utility function. We assume that a white agent living at location i has utility $U_{wi} = Y + \pi W_i - P_i = Y + (\pi - 1)W_i - B_i$, where $\pi > 0$. A positive value of π implies that the more white neighbors (the fewer black neighbors) a white agent has, the happier he or she is. $Y - P_i$ can be interpreted as the dollar value of all non-housing goods consumed. Letting $\theta = \pi - 1 > -1$, we then have

$$U_{wi} = Y + \theta W_i - B_i \tag{2}$$

Assume that, unlike whites, blacks are color neutral. (Some survey data, e.g., the General Social Survey, support this asymmetric assumption. See also Farley et al., 1978; and Farley, Fielding, and Krysian, 1997). They care about the price of housing but not about their neighbors' color. So a black agent living at location j has utility $U_{bj} = Y - P_j = Y - W_j - B_j$. This is simply the dollar value of all non-housing goods consumed:

$$U_{bj} = Y - W_j - B_j \tag{3}$$

Agents have opportunities to move. In each period of time, a pair of locations is chosen randomly. If both locations are vacant, nothing will happen. If one location is vacant and the other is occupied, the agent in the occupied location may choose to move to the other. If both locations are occupied, the two agents may want to exchange residential locations (see Young, 1998, for a similar setup). Of course, agents make their decisions according to their utilities.

There are 5 types of moves:

1. A black agent moves to a vacant location;
2. A white agent moves to a vacant location;
3. A black agent exchanges residential location with a white agent;

4. A black agent exchanges residential location with a black agent;
5. A white agent exchanges residential location with a white agent.

2.2 A Potential Function

Let SW be the sum of all agents' utilities:

$$SW = \sum_i U_i(Y, W_i, B_i),$$

then we have the following observation.

Claim 1: *The change in SW is always twice as much as the change in the moving agent's (agents') utility.*

Claim 1 is straightforward. Consider a white agent who is moving. The number of white (black) neighbors he or she leaves behind is exactly equal to the number of whites (blacks) who lose a white neighbor because of the move; similarly, the number of white (black) neighbors the mover has in the new neighborhood is equal to the number of whites (blacks) who get a new white neighbor. A similar statement can be made for a black mover. Thus, if a move is advantageous for the moving agent, then positive externalities exceed negative externalities, and the move is also advantageous for the agent's former and current neighbors as a group. See the proof in the Appendix.

Suppose the total number of vertices of the lattice is N. We define a state x as an N-vector, with each element $x_i \in \{black, white, vacant\}$ describing the situation at vertex i. We use ρ to denote $SW/2$:

$$\rho = \frac{1}{2} \sum_i U_i(Y, W_i, B_i).$$

Thus ρ is a function defined on the set of all states X.

A move may involve one or two agents. If only one agent is involved, let u be the agent's utility; if two are involved, let u be the sum of their utilities. When agents get the chance to move, each of their action sets is $\{A_1, A_2\} = \{move, don't\ move\}$. If agents move, they may change the state (residential pattern).

Claim 1 simply tells us that

$$u(\cdot|A_1) - u(\cdot|A_2) = \rho(\cdot|A_1) - \rho(\cdot|A_2). \tag{4}$$

Definition 1 (Monderer and Shapley, 1996): Let Γ be an n-person game with finite strategy sets $Z_1, Z_2, \ldots, Z_n$. The payoff function of player i is $u^i : Z \to R$, where $Z = Z_1 \times Z_2, \ldots, Z_n$ is the set of strategy profiles. A game Γ is a *potential game* if there exists a function $\lambda : Z \to R$ such that for every i and for every $z_{-i} \in Z_{-i}$,

$$u^i(x, z_{-i}) - u^i(y, z_{-i}) = \lambda(x, z_{-i}) - \lambda(y, z_{-i}), \quad \text{for every } x, y \in Z_i.$$

λ is called a ***potential function*** of this game.

By equation (4), our spatial game is a potential game, and ρ is a potential function.

This potential function allows us to ignore the details of the dynamics. By keeping track of this potential, we will have enough information about how the game is played. If the potential increases, we know some agents have taken utility-improving moves; if it decreases, we know some agents have made bad decisions, diminishing their utilities. As we shall see later, the potential function greatly simplifies our analysis of the dynamic properties of the game.

2.3 The Log-linear Behavioral Rule

We assume that agents are boundedly rational; in particular, they sometimes make mistakes and take utility-decreasing moves. Suppose the probability that agents choose "move" (A_1) when they have the chance is determined by the following equation:

$$\Pr(A_1) = \frac{e^{\beta u(\cdot|A_1)}}{e^{\beta u(\cdot|A_1)} + e^{\beta u(\cdot|A_2)}}, \quad \beta \gg 0. \tag{5}$$

A standard justification of this rule follows from the usual interpretation of the logit model in econometrics: Agents' utilities are subject to shocks that follow an i.i.d. extreme value distribution (McFadden, 1973; Brock and Durlauf, 1999). Alternatively, it can be interpreted as "perturbed" decision making in which agents occasionally deviate from playing their "best response" (Blume, 1993, 1997; Young, 1998). Clearly, agents become very unlikely to make utility-decreasing moves as $\beta \to \infty$.

2.4 Main Results

We define x^t as the state at time t, so we have a finite Markov process. Let's use P^β to denote the Markov process (its transition probability matrix). We call it a perturbed process because the system is subject to shocks and agents do not always make "correct" decisions. Small values of β imply big perturbations; the perturbation vanishes as β approaches infinity. P^β is ***irreducible*** because there is a positive probability of moving from any state to any other state in a finite number of periods. P^β is ***aperiodic*** because the process can travel from any state x to x itself in any finite number of periods. Hence, by elementary Markov Chain theory,[3] P^β has a unique

[3]Karlin and Taylor (1975) is a standard reference.

stationary distribution μ^β satisfying the equation $\mu^\beta P^\beta = \mu^\beta$. Moreover, $\mu^\beta(x)$ is the cumulative relative frequency with which state x will be observed when the process runs for a long time. It is also the probability that state x will be observed at any time t given that t is sufficiently large.

Definition 2 (Foster and Young, 1990): A state $x \in X$ is *stochastically stable* relative to a perturbed process P^β if $\lim_{\beta\to\infty} \mu^\beta(x) > 0$. The *stochastically stable set* is the smallest set that contains all the stochastically stable states.

A stochastically stable state will be observed much more frequently than a state that is not stochastically stable. As $\beta \to \infty$ and $t \to \infty$, it is likely that the system will be in the stochastically stable set.

Since N is finite, we know there must exist a state x that is associated with the maximum SW. Let S be the set of all such states that maximizes SW:

$$S = \{x|\rho(x) \geq \rho(y), \quad \forall y \in X\}.$$

With the potential function and the probabilistic decision rule, we can prove the following proposition.

Proposition 1: *S is stochastically stable under the perturbed process. That is, in the long run, we will see a state in S almost all the time given that β is large.*

This result immediately follows Theorem 6.1 in Young (1998).

Proposition 1 is fairly intuitive. If x is a state in S, then we know x has the highest potential. Let's say y is a state outside S and y differs from x at exactly two vertices. Obviously, there are many such y's. If the two differing locations are chosen by chance, then with one move (or switch), y can change into x and vice versa. Since y is not in S, it must have a lower potential. We know for certain that the move/switch changing y into x is a "good" move, because it increases potential and hence increases the mover's utility. By our decision rule, if the move increases the mover's utility, it happens with a very high probability. On the contrary, the move/switch changing x into y decreases potential and the mover's utility, so it happens with a much lower probability. Therefore, it is likely that y becomes x, but unlikely that x becomes y. That is, it is easier to fall into the set S, but very difficult to get out of it. As a matter of fact, it is always easier to evolve to a state with higher potential than to a state with lower potential. Remember that S contains all those states with the highest potential. No matter where we start, if β is large and enough time passes, we will eventually arrive at a state in S and then stay in S for a long time. Even if we get out of it by chance, we will eventually fall back into it again. In mathematical terms, this means:

$$\lim_{\beta\to\infty}\lim_{t\to\infty}\Pr\{x^t \in S\} = 1. \tag{6}$$

This is precisely the defining property of the stochastically stable set.

Claim 2: *If the number of vacant locations is small enough relative to N, then every $x^* \in S$ is a state in which white agents are clustered together, and black agents are scattered all over the rest of the landscape. In particular, vacancies tend to be in black areas rather than white areas.*

This seems obvious given the utility functions of the two groups of agents. Therefore, blacks and whites are segregated in a state x^*.

Let's say that not many locations are vacant. Then proposition 1 and Claim 2 lead to the following proposition.

Proposition 2: *If blacks are color-neutral and whites have a slight preference for like-color neighbors, then, in the long run: (i) residential segregation is observed most of the time; (ii) the rate of vacancy is higher in black neighborhoods than in white neighborhoods; and (iii) whites pay more than blacks do for equivalent housing.*

Everything here follows from the previous proposition and Claim 2. As we will see in the simulation presented in the following section, given an initial condition similar to the situation in the 1960s, Proposition 2 explains what happened after the Fair Housing Act put a ban on racial discrimination in the housing market.

This model is an enriched version of Schelling's checkerboard model. Schelling's original model assumed that both groups of agents prefer to live with like-color neighbors (although it has been shown that the preference need not be too strong to achieve segregation). In this enriched version, an important improvement is that one group's preference for like-color neighbors is sufficient for segregation, because the higher price of housing keeps the other group away.

Empirical studies have suggested that income inequalities between blacks and whites play a minor role in explaining residential segregation (Taeuber and Taeuber, 1965; Massey and Denton, 1993). This model helps us better understand why income disparity may not be a crucial determinant. In the model, blacks clearly can afford to live in any neighborhood that whites can afford, because they earn the same income. However, they choose not to. The reason is simple economics: if people have better alternatives, then they may not buy something even if they can afford it. In this scenario, whites are willing to pay a premium to live with other whites. This preference bids up the price of housing in white neighborhoods. Blacks have no reason to pay the premium for living with whites, so they simply choose the less "crowded" (in a topological sense but not in a Euclidean sense) and less expensive residential areas left by whites.

The model describes a world where whites flee from blacks. However, we do not deny the existence of a different world, especially during the first half of the twentieth century, in which whites kept blacks away by discriminatory housing prices, by restrictive covenants, or even by more hostile actions. We do not expect Proposition 2 to hold up in such a world.

It is worth noting that any variation of Schelling's "checkerboard" model, like ours, has a welfare implication. Segregation is such a robust result that it is compatible with various assumptions at the individual level (See Schelling, 1971, and Zhang, 2004). In general, segregation could be an optimal or suboptimal outcome depending on individuals' preferences. Therefore, the welfare implication of any segregation model should be read with caution. The basic setup of our model implies social optimality of segregation. Yet, simulations show that all the results of the model still hold if we introduce some non-linearity into the utility functions. In those cases, segregation is no longer socially optimal. Such non-linear models are extremely messy to work with, if not entirely intractable.

3. AGENT-BASED SIMULATIONS

In contrast to Proposition 2, which is a limiting result, this section examines the dynamic properties of the model with finite parameter values. We shall also explore some variations of the model.[4] All these are done with agent-based simulations on a 30×30 landscape. We use the Moore neighborhood definition, in which each agent considers the eight surrounding locations as neighboring locations.

3.1 Properties of the Model: The First Simulation

This first simulation demonstrates how the segregation model works. Figures 1–3 present the outcomes from a typical run with $\theta = 1$, $\beta = 2$, and $Y = 10$. A residential location is painted dark gray if it is occupied by a black agent, light gray if occupied by a white agent, and left blank if it is vacant. "APB" and "APW" are the average prices of housing paid by blacks and whites, respectively.

We start our simulation in a random state: 400 black agents and 400 white agents are randomly distributed on the 30×30 landscape and hence 100 locations are vacant (Figure 1). In this initial state, blacks and whites are mixed together and pay roughly the same price for housing. As time goes on, black and white clusters begin to emerge, and eventually are

[4]Interested readers may want to try out the simulations presented here and many other variations. The Java Applet is available from the author upon request.

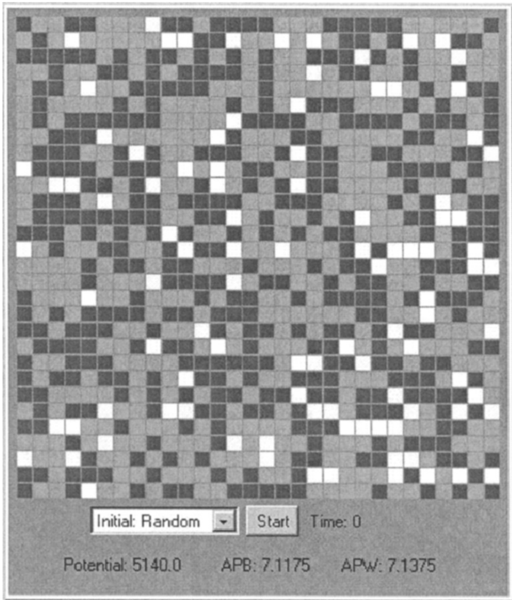

FIGURE 1 A random initial state.

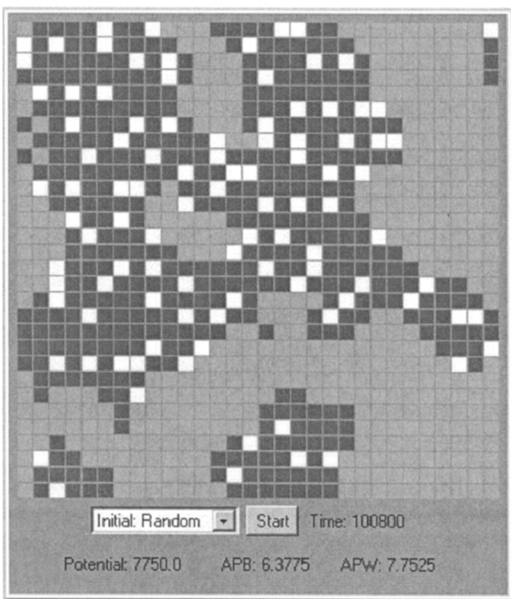

FIGURE 2 A snapshot in the short run.

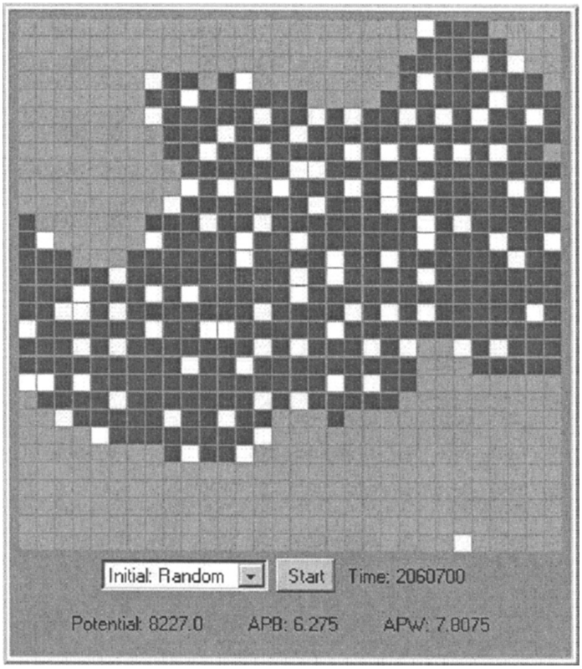

FIGURE 3 Segregation in the long run.

completely segregated. Prices paid by blacks and whites start to differ shortly after the start. In the long run, the price differential is substantial, and vacancies are much more likely in black neighborhoods (Figure 3).

Figure 4 depicts the trajectory of the potential function starting from a random state. While the potential function exhibits a rising trend over time, it does decrease from time to time, which reflects the fact that people make utility-decreasing moves occasionally. The speed of the increase in the potential is declining: it starts with a steep upward trend and flattens out eventually. Corresponding to this dynamic is the evolution of residential patterns: starting from a random state, patches of whites and blacks soon appear on the landscape as individuals move, which drives up the potential quickly. Clusters of whites expand as other whites find such neighborhoods attractive, and at the same time, nearby blacks are pushed away by high prices in white clusters. At this stage, the rise of the potential slows down. Little by little, white clusters join each other, and eventually all white clusters merge into a single white area. This final step takes a long time as the boundaries of clusters shift back and forth. Although the system only approaches the stochastically stable state and is not exactly in it with finite

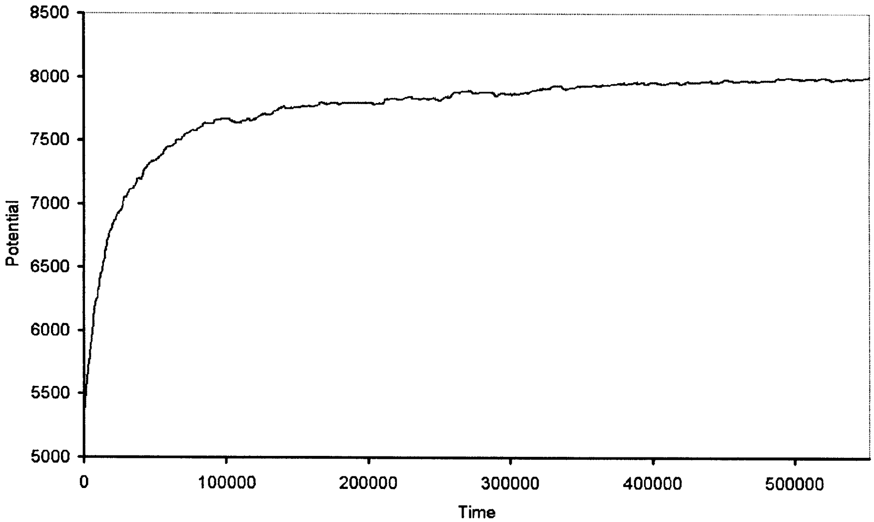

FIGURE 4 Evolution of potential starting from a random state.

parameter values, highly segregated residential patterns emerge long before the dynamical system comes close to a stochastically stable state.

Proposition 2 holds while β approaches infinity, and thus we want to know how finite values of β affect the dynamics of the model. Notice that, since any utility function is invariant to linear transformation, the value of β is meaningful only relative to the unit of utility. Following the literature, we define "waiting time" as the time it takes for the dynamical system to reach a certain state for the first time. Again, starting from random initial states with $\theta = 1$ and $Y = 10$, we examine how the waiting time varies with β until the potential rises above 7800. As shown in Figure 5, the expected waiting time is a decreasing function of β. From the behavioral rule we know that $\beta = 0$ means that agents move randomly without considering racial composition in neighborhoods. In that case, a segregated neighborhood can only emerge by chance, but the chance is so small that we almost never see it. So the expected waiting time is infinitely large when β is close to zero. The simulation shows that the expected waiting time drops sharply as β increases from 0. Before the value of β reaches 1, the waiting time hits a plateau and turns flat. Beyond that point, small increments in the β value have negligible impacts on waiting time.

We also tried other variations, such as different relative population sizes and different proportions of vacancies. Segregation as well as price and vacancy differentials emerge under all alternative parameterizations. In general, such alternative parameterizations also influence expected waiting time. However, comparative statics in those cases produce few insightful

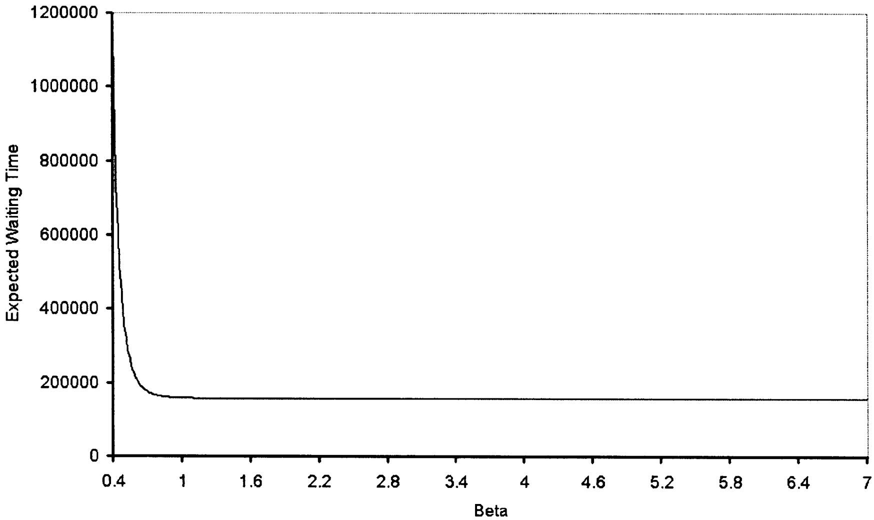

FIGURE 5 Expected waiting time as a function of beta.

results. Nonetheless, it is worth noting that the proportion of vacancies has limited impact on the speed of segregation. This is only because we allow agents to exchange residential locations. If we rule out the possibility of trading locations, a smaller proportion of vacancies significantly reduces the speed of segregation.

3.2 Recent Trends of Segregation: The Second Simulation

Here we briefly discuss a parallel of several trends of segregation and show how our model helps explain them.

Since the Fair Housing Act of 1968, housing in urban America has been characterized by the persistence of racial segregation, the expansion and consolidation of black ghettos, and substantial changes in racial housing price and vacancy rate differentials.

The dramatic formation of black ghettos began in the early twentieth century. The Fair Housing Act was intended to eliminate racial discrimination in the housing market. Since its passage, the level of residential segregation has dropped slightly. Though this declining trend is robust, the decreases are modest, and the level of segregation remains very high (Cutler, Glaeser and Vigdor, 1999; Farley and Frey, 1994).

Since 1970, there has also been a trend toward black suburbanization, which has effectively expanded black ghettos and pushed the color line into the suburbs. Ghettos, in the meantime, sprawled in such a way that previously distinct black communities merged with one another. As a

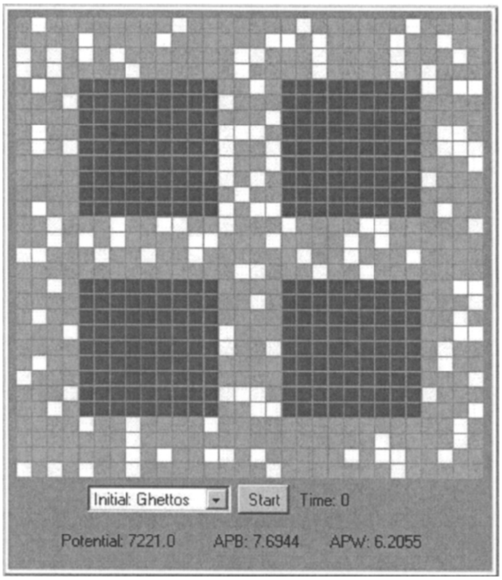

FIGURE 6 An initial state with ghettos.

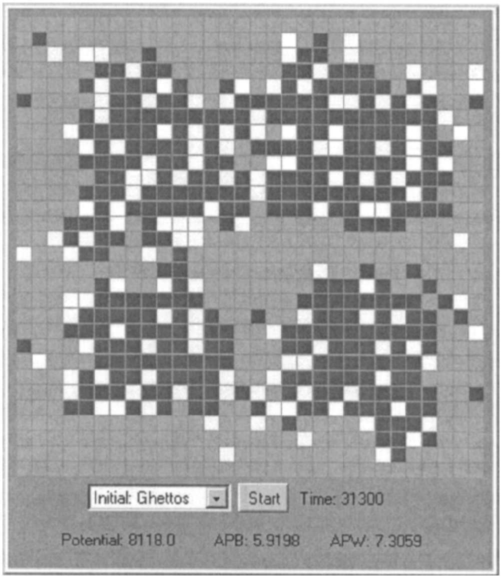

FIGURE 7 A transitional state.

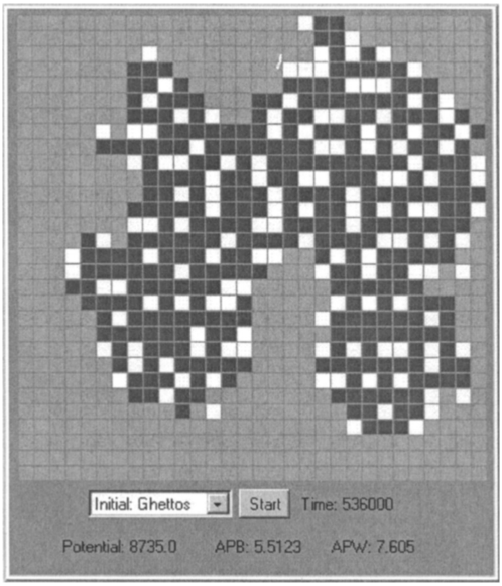

FIGURE 8 Segregation persists with reversed price and vacancy differentials.

result, blacks are not only highly separated from whites, but also concentrated in a few large geographic areas that span city and suburb (Clark, 1981; Connolly, 1974).

There has also been a shift in the relative price of housing in black and white neighborhoods. In the mid-twentieth century, blacks tended to live in lower-vacancy neighborhoods and paid a higher price for equivalent housing than whites (see, e.g., Weaver, 1948; Duncan and Duncan, 1957; King and Mieszkowski, 1973; and Yinger, 1978). At the end of the twentieth century, black neighborhoods were characterized by higher vacancy rates and lower housing prices than white neighborhoods (see, e.g., Chambers, 1992; Reifel, 1994; and Zhang, 2001).

The existing literature does not provide an adequate account of these trends. This second simulation shows how our model helps explain the recent history of urban development. We start with a state in which segregated enclaves are already established and free mobility is suddenly introduced, similar to what happened in the 1960s. Until that time, legally sanctioned discriminating practices, such as segregational zoning ordinances, restrictive covenants, and racial steering, were common in urban housing markets throughout the U.S., which greatly limited the supply of housing for blacks. Consequently, in cities such as Chicago, Cleveland, and Detroit, the price of housing paid by blacks was artificially high. For the same reason, housing vacancy rates were extremely low in black neighborhoods. Our initial state in Figure 6 reflects the situation at that time: a high level of segregation, high housing prices and low vacancies for blacks. (Of course, black ghettos need not be so square and regular.)

Suppose that racial discrimination is now banned by law, and the simulation starts with $\theta = 1$, $\beta = 2$, and $Y = 10$. As time goes on, with no discriminating forces, blacks begin to "invade" white neighborhoods (Figure 7). This shows a promising sign of desegregation. However, as blacks appear, many whites choose to move to other predominantly white neighborhoods, leaving blacks behind. Eventually, the "invasion" of blacks results in a retreat by whites. The color line is shifted, ghettos expand and segregation remains at a high level (Figure 8). Moreover, we see in Figure 8 that whites now pay higher prices for equivalent housing and that vacancy rates are lower in white neighborhoods, contrary to the initial state. This is exactly how urban development occurred in the U.S. over the past four decades.

3.3 Nonlinear Utilities: The Third Simulation

For simplification and tractability, we have assumed linear utility functions in our mathematical model. While the mathematical model serves as a good benchmark, it is worth trying various alternative setups with the simulation to test the robustness of the results. Simulations shows that asymmetric residential preferences are the key driver of the model; non-linear utilities that deviate somewhat from the setup in the benchmark model are able to produce similar results. Most important, segregation in such cases does not necessarily correspond to optimal or close-to-optimal states.

For example, in a variation, we assume that blacks are color neutral but whites have the following utility function:

$$\begin{aligned} U_{wi} = Y - P_i &+ \pi_1 \min\{W_i, x\} \\ &- \pi_2 \max\{0, W_i - x\}, \quad \pi_1, \pi_2 > 0 \text{ and } 0 < x < H. \end{aligned}$$

It is a kinked utility function that peaks at x. That is, if a white agent i has less than x white neighbors, an additional white neighbor adds to his or her utility; if i has x or more white neighbors, the next white neighbor causes disutilities. It reduces to our benchmark model if $x = H$. Since we use the Moore neighborhood in the simulation, an agent could have 8 neighbors. We choose $x = 5$, which means that a white agent is most happy in a neighborhood that is about 60 percent white (5/8). Other parameters take the following values: $\pi_1 = 2$, $\pi_2 = 0.5$, $\beta = 2$, and $Y = 10$. The asymmetry of the preference is reflected in $\pi_2 < \pi_1$, which implies that a predominantly white neighborhood is more attractive to a white agent than a predominantly black neighborhood, even though predominantly white neighborhoods are not ideal for any white agent.

Figure 9 shows a typical long-run residential pattern. Severe segregation still emerges and price and vacancy differentials are still significant. Starting from a random state, the segregation dynamics evolve as follows: At first, those whites who have few white neighbors start to move to neighborhoods that are more than half white. This has two impacts. On the one hand, the neighborhoods left behind by those white movers will necessarily become "blacker" and even less attractive to other whites; on the other hand, those moves crowd out blacks from the destination neighborhoods and they are naturally directed to predominantly black neighborhoods by the price mechanism. Before long, most neighborhoods where blacks live become so black that no whites would like to reside in them; whites in such black neighborhoods would prefer to move to an all-white neighborhood rather than stay. This process pushes segregation towards its extreme.

While segregation is striking in the long run in this alternative setting, it is not complete. In particular, the white area is not 100 percent white. This is because an all-white neighborhood with one or two vacancies is not attractive to a white agent since it is already too white, but it is quite appealing to some blacks because it is reasonably cheap. For this reason, we always see a few blacks scattered throughout white neighborhoods; but whites will never consider going across the color line because it represents a big utility loss. This asymmetry not only creates price and vacancy differentials, but also makes segregation stable once it is established. As a

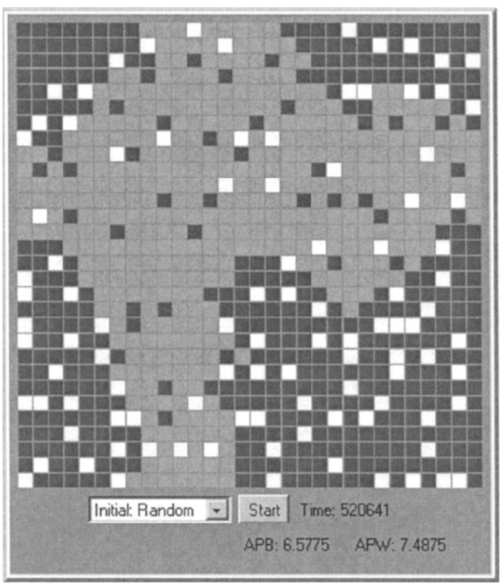

FIGURE 9 Segregation under integrationist preferences.

result, although whites in fact prefer mixed neighborhoods, they end up with segregated ones because of the asymmetric preferences. This provides another illustration of Schelling's insight that macro behavior can deviate substantially from micro motives. Zhang (2004) further shows that segregation could emerge even if both blacks and whites prefer mixed neighborhoods. In Figure 9, only 12.5 percent of whites have 5 white neighbors; 77.3 percent of them have more than 5 like-color neighbors. Not surprisingly, if we choose $x > 5$, the simulation produces more striking segregation results than that seen in Figure 9.

4. CONCLUDING REMARKS

We have shown that the Schelling-type simulation models can be rigorously analyzed using techniques recently developed in evolutionary game theory. For three decades, segregation in the Schelling model has been known as an "emergent and persistent phenomenon." We have now translated the loose description into a precisely defined mathematical concept. By explicity formulating residential moves as a spatial game, we have embedded the analysis of segregation into a game-theoretic framework where various analytical tools are readily available.

We have enriched the Schelling model by adding a simplified housing market. Our model shows that one group's preference for like-color neighbors is sufficient to cause residential segregation. People in this group live together and bid up housing prices. The higher price keeps people away from the other group. The housing market plays an important role in the dynamics of segregation. Without the housing market, segregation will not emerge if individuals are initially assigned to locations randomly. In addition, the housing market has enhanced the power of the Schelling model. Our model not only generates segregation but also produces predictions of the housing market. In particular, taking the situation prior to the Fair Housing Act of 1968 as an initial state, our model is able to account for many observed regularities over the past 40 years, such as the persistence of segregation, the reversal of housing price differentials, and the reversal of housing vacancy differentials.

We recognize that there exist alternative explanations of residential segregation. For example, differential preferences for local public goods, such as school quality, could produce similar segregation results and housing price discrepancies. Our model proposes a competing hypothesis that invites further investigation to determine which theory better accounts for the empirical data. The model has shown that, without any discriminatory behavior in the housing market, a slight preference for like-color neighbors in one race can give rise to a high level of residential segregation and cause it to persist. While institutionally sanctioned discrimination might have been sufficient to create segregation historically, out model suggests that the elimination of discrimination may not be sufficient to achieve desegregation.

5. APPENDIX: PROOFS

Proof of Claim 1

Proof: We first consider the case in which the two locations chosen are not neighboring ones.

Type-1 move: Suppose black agent k moves from location i to a vacant location j. Then her individual gain is $(Y - W_j - B_j) - (Y - W_i - B_i) = (W_i + B_i) - (W_j + B_j)$. Each neighbor of i (white or black) gains 1 unit of utility, and each neighbor of location j (white or black) loses 1 unit of utility. So the net social gain is agent k's gain plus her former and current neighbors' gain: $(W_i + B_i - W_j - B_j) + (W_i + B_i - W_j - B_j) = 2(W_i + B_i - W_j - B_j)$, which is twice as much as the individual gain.

Type-2 move: Suppose white agent t moves from location p to a vacant location q. Then her individual gain is $(Y + \theta W_q - B_q) - (Y + \theta W_p - B_p) = (\theta W_q - B_q) - (\theta W_p - B_p)$. Each white neighbor of p loses θ units of utility, and each black neighbor of location p gains 1 unit of utility, so their net gain is $B_p - \theta W_p$. Each white neighbor of q gains θ units of utility, and each black neighbor of location θ loses 1 unit of utility, so their net gain is $\theta W_q - B_q$. The total net gain is agent t's gain plus her former and current neighbors' gain: $(\theta W_q - B_q - \theta W_p + B_p) + (B_p - \theta W_p + \theta W_q - B_q) = 2(\theta W_q - B_q - \theta W_p + B_p)$, which is again two times the individual gain.

Type-3 move: Suppose a black agent at vertex u chooses to exchange residential location with a white agent at vertex w. This is equivalent to the situation in which the black agent moves to a vacant location w and the white agent moves to a vacant location u. If the black agent's gain is g and the white agent's gain is h, then their individual gain is $g + h$. By our previous argumants, the social gain is always twice as much as the individual gain. So, the change in SW is $2(g + h)$.

Since the other two types of moves involve same-color agents, individual and social gains are both 0. The relationship is still maintained.

It is not difficult to verify that the relationship holds even if the two locations chosen are neighboring locations.

Proof of Proposition 1

Proof: This is essentially a repetition of Young's proof of Theorem 6.1 (Young, 1998).

Let $\mu(x)$ be the stationary distribution of the perturbed process, which specifies the probability that a state x is visited in the long run. We define P_{xy} as the transition probability from state x to y. The *detailed balance condition* states:

$$\mu(x)P_{xy} = \mu(y)P_{yx}, \quad \forall x, y \in X.$$

We claim that the stationary distribution takes the following form:

$$\mu(x) = \frac{e^{\beta\rho(x)}}{\sum_{z\in X} e^{\beta\rho(z)}}, \tag{$*$}$$

where ρ is our potential function.

If $x = y$ or $P_{xy} = P_{xy} = 0$, $(*)$ satisfies the detailed balance condition. If $x \neq y$ and $P_{xy} \neq 0$ or $P_{xy} \neq 0$, then it must be true that x differs from y at only two locations i, j. If the total number of vertices is N, then these two locations will be chosen with probability $1/\{N(N-1)\}$. It follows that:

$$\begin{aligned}
\mu(x)P_{xy} &= \left\{\frac{e^{\beta\rho(x)}}{\sum_{z\in X} e^{\beta\rho(z)}}\right\}\left\{\frac{1}{N(N-1)} \cdot \frac{e^{\beta u(\cdot|move)}}{e^{\beta u(\cdot|move)} + e^{\beta u(\cdot|don't\ move)}}\right\} \\
&= \left\{\frac{e^{\beta\rho(x)}}{\sum_{z\in X} e^{\beta\rho(z)}}\right\}\left\{\frac{1}{N(N-1)}\right. \\
&\qquad \left. \times \frac{e^{\beta u(\cdot|don't\ move)+\beta u(\cdot|move)-\beta u(\cdot|don't\ move)}}{e^{\beta u(\cdot|move)} + e^{\beta u(\cdot|don't\ move)}}\right\} \\
&= \left\{\frac{e^{\beta\rho(x)}}{\sum_{z\in X} e^{\beta\rho(z)}}\right\}\left\{\frac{1}{N(N-1)} \cdot \frac{e^{\beta u(\cdot|don't\ move)} \cdot e^{\beta\rho(y)-\beta\rho(x)}}{e^{\beta u(\cdot|move)} + e^{\beta u(\cdot|don't\ move)}}\right\} \\
&= \left\{\frac{e^{\beta\rho(y)}}{\sum_{z\in X} e^{\beta\rho(z)}}\right\}\left\{\frac{1}{N(N-1)} \cdot \frac{e^{\beta u(\cdot|don't\ move)}}{e^{\beta u(\cdot|move)} + e^{\beta u(\cdot|don't\ move)}}\right\} \\
&= \mu(y)P_{yx}.
\end{aligned}$$

Therefore, (*) satisfies the stationary equation because

$$\sum_{x\in X} \mu(x)P_{xy} = \sum_{x\in X} \mu(y)P_{yx} = \mu(y)\sum_{x\in X} P_{yx} = \mu(y) \cdot 1 = \mu(y).$$

This implies that (*) is a stationary distribution of the perturbed process. Since the process is finite and *irreducible*, it has a unique stationary distribution, which must be defined as (*).

From (*), we immediately conclude that the states that maximize social welfare are stochastically stable, because the perturbed process assigns positive probability to those states when $\beta \to \infty$.

Proof of Claim 2

Proof: Consider the lattice graph on a torus, some of whose vertices are occupied by black or white agents, with the rest vacant. Neighboring vertices are defined as in the Moore neighborhood. We add all diagonal edges so that all neighboring vertices are connected. We refer to an edge by the two vertices it connects. For example, we refer to an edge as $\{b, w\}$ if it connects two vertices occupied by a black agent and a white agent. The order doesn't matter here.

We will proceed with the following notations:

n–total number of vertices.
n_b–the number of vertices occupied by black agents.
n_w–the number of vertices occupied by white agents.
n_v–the number of vacant vertices.
E_{ij}–total number of $\{i, j\}$ edges, $i, j \in \{b, w, v\}$.

Given the utility functions of the two groups of agents, we know $SW = nY - 2E_{bb} - 2E_{bw} + 2\theta E_{ww}$. If our goal is to maximize E_{ww} only, then we need to have white agents clustered together. This can be proved by negation.

Each vertex of the graph has degree 8, i.e., it is connected with eight neighboring vertices. If we count all the edges connecting a white vertex with a black or a vacant vertex, and double count all the edges connecting two white vertices, then we get the total degree of all white vertices. In fact, we have three such relations:

$$2E_{ww} + E_{wb} + E_{wv} = 8n_w \tag{i}$$
$$2E_{bb} + E_{wb} + E_{bv} = 8n_b \tag{ii}$$
$$2E_{vv} + E_{wv} + E_{wv} = 8n_v \tag{iii}$$

Suppose for the time being there are no vacant locations, then by equation (i) and (ii) we have:

$$(2E_{ww} + E_{wb} + 0) + (2E_{bb} + E_{wb} + 0) = 8(n_w + n_b).$$
$$\Rightarrow E_{ww} + E_{bb} + E_{wb} = 4(n_w + n_b) = \textit{constant}.$$

Cleary, to maximize $SW = nY - 2E_{bb} - 2E_{bw} + 2\theta E_{ww} = nY - 8(n_w + n_b) + 2(\theta + 1)E_{ww}$, we need to maximize E_{ww} given that $\theta = \pi - 1 > -1$. That is, we must have all white agents clustered together.

Now, we add some vacant locations. Let's first have the predetermined number of whites on the landscape and fill the other locations with black agents. At this moment, to achieve the maximum social welfare, we need to maximize E_{ww}. Then, to get the right number of black agents, we will take n_v blacks away from the landscape. Let's first take 1 black agent away; it must be true that E_{ww} remains unchanged and $(E_{bb}+E_{wb})$ decreases by eight. Notice that eight is the maximum number of those types of edges that can be reduced by removing one black agent. Suppose n_v is small enough so that we can remove n_v black agents, one after another, with each having eight black or white neighbors. In other words, after n_v black agents are removed, we still have $E_{vv}=0$. (This can be achieved as long as $n_v \leq n_b/4$. Interested readers can try this out.) By this construction, we guarantee that social welfare is at its maximum at each state. At the end, when all n_v blacks are removed, social welfare is still at its maximum. At the same time, E_{ww} is also maximized because we never moved any white agents.

We know any rearrangement of the configuration cannot increase social welfare. By equation (i) and (ii), we have:

$$(2E_{ww} + E_{wb} + E_{wv}) + (2E_{bb} + E_{wb} + E_{bv}) = 8(n_w + n_b).$$
$$\Rightarrow 2(E_{ww} + E_{bb} + E_{wb}) + (E_{wv} + E_{bv}) = 8(n_w + n_b). \tag{iv}$$

Since $E_{vv}=0$, equation (iii) reduces to $E_{wv}+E_{bv}=8n_v$. By equation (iv),

$$2(E_{ww} + E_{bb} + E_{wb}) = 8(n_w + n_b) - 8n_v.$$

$$E_{ww} + E_{bb} + E_{wb} = 4(n_w + n_b - n_v) = \mathit{constant}.$$

Therefore, $SW = nY - 2E_{bb} - 2E_{wb} + 2\theta E_{ww} = nY - 8(n_w + n_b - n_v) + 2(\theta + 1)E_{ww}$ can achieve its maximum only by maximizing E_{ww} and hence minimizing $(E_{bb}+E_{wb})$ at the same time. So whites are clustered together.

Proof of Proposition 2

Proof: It trivially follows from Theorem 1 and Claim 2.

REFERENCES

Blanchflower, D. & Oswald, A. (1994). *The Wage Curve*. Cambridge, MA: MIT Press.

Blank, D. M. & Winnick, L. (1953). "The Structure of the Housing Market." *Quarterly Journal of Economics*, *67*, 181–208.

Blume, L. E. (1993). "The Statistical Mechanics of Strategic Interaction." *Games and Economic Behavior*, *5*, 387–424.

Blume, L. E. (1997). "Population Games." In *The Economy as an Evolving Complex System II*. W. B. Arthur, S. N. Durlauf and D. A. Lane. (eds.) Reading, MA: Addison-Wesley.

Bobo, L. & Zubrinsky, C. L. (1996). "Attitudes on Residential Integration: Perceived Status Difference, Mere In-group Preference, or Racial Prejudice." *Social Forces*, *74*, 883–909.

Brock, W. A. & Durlauf, S. N. (1999). "Interaction Based Models." SSIR Working Paper 9910, University of Wisconsin.

Chambers, D. N. (1992). "The Racial Housing Price Differential and Racially Transitional Neighborhoods." *Journal of Urban Economics*, *32*, 214–32.

Clark, T. A. (1981). "National Trends in Black Suburbanization." In *The New Suburbanites: Race and Housing in the Suburbs*, R. W. Lake (ed). New Brunswick, N.J.: Center for Urban Policy Research, Rutgers University.

Connolly, H. X. (1974). "Black Movement into the Suburbs." In *Suburbanization Dynamics and the Future of the City*, J. W. Hughes (ed). New Brunswick, N.J.: Center for Urban Policy Research, Rutgers University.

Cutler, D. & Glaeser, E. (1997). "Are Ghettos Good or Bad?" *Quarterly Journal of Economics*, *112*(3), 827–72.

Cutler, D., Glaeser, E., & Vigdor, J. (1999). "The Rise and Decline of the American Ghettos." *Journal of Political Economy*, 107, 455–506.

Duncan, O. D. & Duncan, B. (1957). *The Negro Population of Chicago*. Chicago, IL: University of Chicago Press.

Epstein, J. M. & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, D.C.: Brookings Institution Press.

Farely, R., H. Schuman, Bianchi, S., Colasanto, D., & Hatchett, S. (1978). "'Chocolate City, Vanilla Suburbs': Will the Trend toward Racially Separate Communities Continue?" *Social Science Research*, *7*, 319–44.

Farley, R., Fielding E. L., & Krysian, M. (1997). "The Residential Preferenes of Blacks and Whites: A Four-Metropolis Analysis." *Housing Policy Debate*, *8*(4), 763–800.

Farley, R. & Frey, W. H. (1994). "Changes in the Segregation of Whites from Blacks During the 1980s: Small Step Toward a More Integrated Society." *American Sociological Review*, *59*, 23–45.

Foster, D. and Young, H. P. (1990). "Stochastic Evolutionary Game Dynamics." *Theoretical Population Biology*, *38*, 319–32.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*. New York, NY: Academic Press.

Kiel, K. A. & Zabel, J. E. (1996). "Housing Price Differentials in U.S. Cities: Household and Neighborhood Racial Effects." *Journal of Housing Economics*, *5*, 143–65.

King, A. T. & Mieszkowski, P. (1973). "Racial Discrimination, Segregation, and the Price of Housing." *Journal of Political Economy*, *81*, 590–606.

Massey, D. S. (1979). "Effects of Socioeconomic Factors on the Residential Segregation of Blacks and Spanish Americans in U.S. Urbanized Areas." *American Sociological Review*, *44*, 1015–22.

Massey, D. S. & Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

McFadden, D. (1973). "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarebka (ed). *Frontier in Econometrics*. New York, NY: Academic Press.

Monderer, D. and Shapley, L. S. (1996). "Potential Games." *Games and Economic Behavior*, *14*, 124–43

Reifel, J. W. (1994). "Black-White Housing Price Differentials: Recent Trends and Implications." *Review of Black Political Economy*, *23*, 67–93.

Rosen, K. T. and Smith, L. B. (1983). "The Price-Adjustment Process for Rentel Housing and the Natural Vacancy Rate." *American Economic Review*, *73*, 779–86.

Schelling, T. C. (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology*, *1*, 143–86.

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

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

Weaver, R. C. (1948). The Negro Ghetto. New York: Harcourt, Brace and Company.

Yinger, J. (1978). "The Black-White Price Differential in Housing: Some Further Evidence." *Land Economics*, *54*, 187–206.

Young, H. P. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

Zhang, J. (2001). *An Evolutionary Approach to Residential Segregation*. Unpublished dissertation. The Johns Hopkins University.

Zhang, J. (2004). "Residential Segregation in an All-Integrationist World." *Journal of Economic Behavior and Organization*, forthcoming.