Journal of Economic Behavior & Organization
Vol. 54 (2004) 533–550

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Residential segregation in an all-integrationist world

Junfu Zhang*

Public Policy Institute of California, 500 Washington Street, Suite 800, San Francisco, CA 94111, USA

Received 20 September 2001; accepted 3 March 2003

Available online 28 January 2004

Abstract

This paper presents a variation of the Schelling [J. Math. Sociol. 1 (1971) 143; T.C. Schelling, Micromotives and Macrobehavior, Norton, New York, 1978] model to show that segregation emerges and persists even if every person in the society prefers to live in a half-black, half-white neighborhood. In contrast to Schelling's inductive approach, we formulate neighborhood transition as a spatial game played on a lattice graph. The model is rigorously analyzed using techniques recently developed in stochastic evolutionary game theory. We derive our primary results mathematically and use agent-based simulations to explore the dynamics of segregation.

JEL classification: C73

Keywords: Residential segregation; Potential function; Stochastic stability; Agent-based simulation

1. Introduction

In 1982, the General Social Survey asked all black respondents this question: "If you could find the housing that you would want and like, would you rather live in a neighborhood that is all black; mostly black; half black, half white; or mostly white?" It turned out that 55.3% preferred to live in a half-black, half-white neighborhood. In 1990 and 1996, the General Social Survey asked all respondents (most of whom were white) a similar question regarding their attitude towards living in a neighborhood where half of their neighbors were blacks. Over 60 percent of the respondents answered "neither favor nor oppose," "favor," or "strongly favor" (Davis and Smith, 1999).

The survey results sharply contrast with the realities of our society where residential segregation persists at high levels. Even if we refer to neighborhoods that are 40–60 percent

* Corresponding author. Tel.: +1-415-291-4454; fax: +1-415-291-4428.
E-mail address: zhang@ppic.org (J. Zhang).

doi:10.1016/j.jebo.2003.03.005

Table 1
Percentage of blacks who live in a half–half neighborhood

Metro area	Neighborhoods in sample	Half–half neighborhoods	Percentage (%)
Baltimore	581	25	3.80
Buffalo	289	12	2.49
Chicago	1885	71	2.91
Cleveland	829	34	2.97
Detroit	1273	40	2.56
Milwaukee	428	16	2.74
St. Louis	456	20	4.24
Washington	911	63	5.96

black as "half black, half white" neighborhoods, we find that a very few number of blacks actually live in these neighborhoods in major metropolitan areas (Table 1, data from 1990 Census).[1] How then should we interpret the survey data? If many whites and blacks desire integrated neighborhoods, then why do we not observe more of them in our society?

Social scientists frequently use survey data to study people's preference of neighborhood race and their attitude toward racial integration. Many authors repeatedly note that surveys reveal more tolerance for mixed race neighborhoods than is actually found in our highly segregated urban areas (Farley et al., 1978, 1997; Bobo and Zubrinsky, 1996). A classic study by Thomas Schelling sheds light on this disjunction between individual preferences and social outcomes. By moving dimes and pennies on a checkerboard, Schelling (1971) demonstrated that moderate preferences for like-color neighbors at the individual level can be amplified into high levels of segregation. Later on, Schelling (1978) used numerous examples to illustrate further that macro social patterns could deviate from micro motives substantially. Although presented informally, Schelling's findings have been influential across disciplines in social sciences. In the past three decades, his work inspired many "tipping" models and agent-based simulations (e.g., Epstein and Axtell, 1996). However, because proper analytical tools were unavailable, there was no important development following Schelling's original model. Young (1998) was the first to point out that techniques newly developed in the theory of stochastic dynamical systems can be used to analyze the Schelling model. This paper represents such an attempt.

In this paper, we will extend Schelling's original results and mathematically prove that segregation may emerge and persist even if everybody prefers integrated neighborhoods. In particular, we will construct an artificial world where all agents think integrated neighborhoods most desirable. However, as we shall show, even in such a world, residential segregation almost surely appears. Our purpose here is not to replicate the real world exactly. Instead, we will attempt to provide some insights that can help us better understand the real world phenomenon of segregation. The most important message we want to convey is that a phenomenon may prevail in a society despite what people want or feel is "good." It is quite possible that such phenomena persist only because the cost of deviation is too high. We think residential segregation at present might be such a phenomenon.

[1] In empirical studies, neighborhoods are usually approximated by census tracts. In general, census tracts have between 3000 and 8000 residents and boundaries that follow visible features such as rivers, highways, or major streets.

We want to point out that once blacks and whites are segregated, it is a long journey back to reintegration. It is particularly worth noting that the first step toward integration involves some blacks moving into predominantly white neighborhoods or some whites moving into predominantly black neighborhoods or both. This first step may never be achieved if people feel extremely uncomfortable when isolated in an opposite-color neighborhood, even if they like the idea of a 50–50 mixed neighborhood. If nobody wants to take the first step, then the society is stuck with segregation.

In contrast to Schelling's inductive approach, we shall derive our results mathematically. We formulate neighborhood transition as a spatial game played on a lattice graph. Agents decide to move in response to local racial composition, creating a game-theoretic situation with feedback loops. The dynamic game is analyzed using techniques recently developed for Markov process with random perturbations. We prove that the stationary distribution of the Markov chain concentrates almost exclusively on the states with the minimum number of black–white neighboring pairs. That is, in the long run, we rarely find blacks and whites living as neighbors. In addition, we show that the segregational pattern emerges regardless of the initial state. For the first time, we put a Schelling-type model on rigorous footing.

The rest of the paper is organized as follows. Section 2 presents the model and the main result. Section 3 explores the dynamics of segregation with agent-based simulation. Section 4 concludes with some remarks.

2. The model

2.1. An artificial world

Our model is a variation of Schelling's famous "checkerboard model." One difference is that we do not have vacant spaces in this model and hence individuals move by switching residential locations. It is as if there exists a centralized agency that processes all the information about who wants to move and which two agents may want to switch.

Consider an $N \times N$ lattice graph embedded on a torus. V is the set of vertices. Each vertex in V is occupied either by a black agent or by a white agent. Consider any type of neighborhood defined locally. In particular, we assume that any agent considers $2n$ agents around her as her neighbors. Let E be the collection of all unordered pairs (i, j), where agents i and j are neighboring agents: $E = \{(i, j)|i \text{ and } j \text{ are neighboring agents}\}$.

An agent's payoff (utility) has two parts: a deterministic term u that depends on how many like-color neighbors she has in the local neighborhood, and a random term ϵ that captures the value of other relevant characteristics of the neighborhood. ϵ is assumed to be independent both across agents and across residential locations because different agents value different characteristics and different neighborhoods have different idiosyncratic traits. An agent's payoff is interpreted as how much he or she likes to pay for a residential location in a neighborhood. The deterministic term of every agent's payoff function is assumed to be a kinked curve, depicted in Fig. 1. The function peaks at n, which is half of her total number of neighbors. On the left side of n, the function is linearly increasing; on the right side of n, it is linearly decreasing. It is relatively steeper on the left side. Linearity is assumed for the

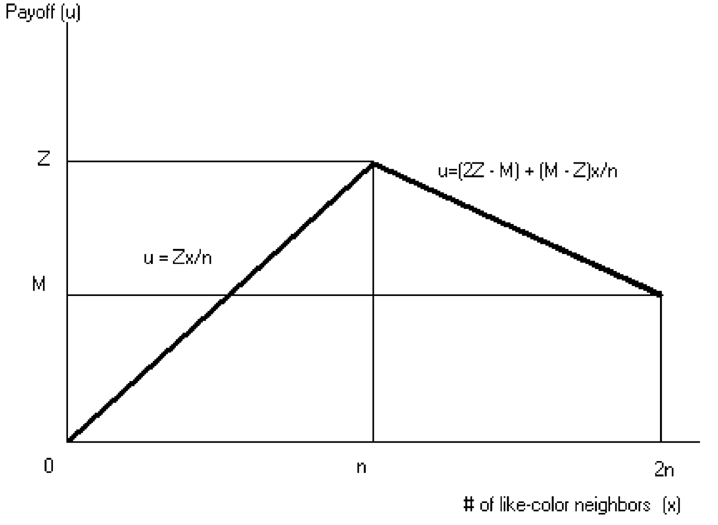

Fig. 1. Payoff profile.

simplicity of analysis. By letting x be the number of like-color neighbors one has, we can write the deterministic part of an agent's payoff function as

$$u = \begin{cases} \dfrac{Zx}{n} & \text{if } x \leq n, \\ (2Z - M) + \dfrac{(M - Z)x}{n} & \text{otherwise,} \end{cases} \qquad Z > M > 0. \tag{1}$$

The shape of the payoff function means that every agent wants 50–50 mixed neighborhoods most. Although people do not prefer segregated neighborhoods, they feel better if they belong to the majority group rather than the minority group. That is, for a white agent, a 30–70 black–white ratio is better than a 70–30 black–white ratio; the opposite is true for a black agent. Empirical evidence suggests that this is a plausible assumption. Multiple reasons could explain the inclination towards one's own race including cultural concerns, fear of potential hostility from the other group, or fear of isolation.

An agent's total payoff is written as

$$\beta u(\cdot) + \epsilon,$$

where β is a positive constant. Parameter β determines the relative importance of the random term. If β is close to zero, the random term is very important, and the racial composition of the neighborhood plays a minor role in an agent's decision; if β is close to infinity, the random term is unimportant, and only the racial composition matters. Every agent has the same β.

Agents may exchange residential locations. In each period of time, a pair of agents is randomly chosen from two different neighborhoods. The chosen agents are allowed to consider switching residential locations according to their own interests. Utility is transferable.

We assume that if one agent gains more in a switch than the other agent loses, then they will negotiate over a proper amount of transfer, and the switch is likely to happen. This allows us to focus on the sum of two chosen agents' payoffs because they always attempt to maximize it by a joint decision.

If the two picked agents do not switch residential locations, the sum of their payoffs is

$$\{\beta u_1(\cdot|\text{not switch}) + \epsilon_1\} + \{\beta u_2(\cdot|\text{not switch}) + \epsilon_2\}$$
$$= \beta\{u_1(\cdot|\text{not switch}) + u_2(\cdot|\text{not switch})\} + \{\epsilon_1 + \epsilon_2\} = \beta U + \eta.$$

If the two picked agents do switch residential locations, the sum of their payoffs is

$$\{\beta u_1(\cdot|\text{switch}) + \lambda_1\} + \{\beta u_2(\cdot|\text{switch}) + \lambda_2\}$$
$$= \beta\{u_1(\cdot|\text{switch}) + u_2(\cdot|\text{switch})\} + \{\lambda_1 + \lambda_2\} = \beta V + \xi.$$

A switch will happen if and only if $\beta U + \eta < \beta V + \xi$. Following McFadden (1973), we assume that η and ξ are independent and follow identical extreme value distribution whose cumulative distribution function and probability distribution function are

$$F(x) = \exp(-\mathrm{e}^{-x}), \qquad f(x) = \exp(-x - \mathrm{e}^{-x}).$$

Then,

$$\begin{aligned}
\Pr\{\text{switch}\} &= \Pr\{\beta U + \eta < \beta V + \xi\} = \Pr\{\eta < \beta V - \beta U + \xi\} \\
&= \int_{-\infty}^{\infty} F(\beta V - \beta U + \xi) \cdot f(\xi)\,\mathrm{d}\xi \\
&= \int_{-\infty}^{\infty} \exp(-\mathrm{e}^{-\beta V + \beta U - \xi}) \cdot \exp(-\xi - \mathrm{e}^{-\xi})\,\mathrm{d}\xi \\
&= \int_{-\infty}^{\infty} \exp\left[-\xi - \mathrm{e}^{-\xi}\left(\frac{\mathrm{e}^{\beta U} + \mathrm{e}^{\beta V}}{\mathrm{e}^{\beta V}}\right)\right] \mathrm{d}\xi \\
&= \int_{-\infty}^{\infty} \exp[-\xi - \mathrm{e}^{-\xi}\mathrm{e}^{\phi}]\,\mathrm{d}\xi \\
&= \exp(-\phi) \cdot \int_{-\infty}^{\infty} \exp[-(\xi - \phi) - \mathrm{e}^{-(\xi-\phi)}]\,\mathrm{d}(\xi - \phi) \\
&= \exp(-\phi) \cdot \int_{-\infty}^{\infty} f(\xi - \phi)\,\mathrm{d}(\xi - \phi) \\
&= \exp(-\phi) \cdot 1 = \frac{\mathrm{e}^{\beta V}}{\mathrm{e}^{\beta U} + \mathrm{e}^{\beta V}}, \quad \text{where } \phi \equiv \ln\left(\frac{\mathrm{e}^{\beta U} + \mathrm{e}^{\beta V}}{\mathrm{e}^{\beta V}}\right).
\end{aligned}$$

Therefore,

$$\Pr\{\text{switch}\} = \frac{\mathrm{e}^{\beta[u_1(\cdot|\text{switch}) + u_2(\cdot|\text{switch})]}}{\mathrm{e}^{\beta[u_1(\cdot|\text{not switch}) + u_2(\cdot|\text{not switch})]} + \mathrm{e}^{\beta[u_1(\cdot|\text{switch}) + u_2(\cdot|\text{switch})]}}, \quad \beta > 0.$$

We call this a log-linear switch rule, which is derived using the method developed by McFadden. This kind of behavioral rule is commonly used in the literature (Blume, 1997; Brock and Durlauf, 2001; Young, 1998). It simply says that if a switch increases the agents' deterministic utilities, they are more likely to do it. Notice, even if a switch decreases the two agents' deterministic utilities, it is still possible that they do it. The possibility of such a "mistake" depends on β. A big β implies that "mistakes" are rarely made. In particular, as β approaches infinity, the probability of a switch approaches 1 if the switch gives higher utilities. In that case, the model reduces to one with agents playing best-reply to their environments.

By integrating out random utilities, we now have a behavioral rule that only involves deterministic utilities. Random utilities are unobservable. Thus, from now on, we will ignore random utilities by working with the behavioral rule.

2.2. A potential function

We define a set ED as the set of all edges that connect two agents of different colors: $\mathrm{ED} = \{(i, j) \in E | i \text{ and } j \text{ have different colors}\}$. A function ρ is then defined as the cardinality of the set ED : $\rho = |\mathrm{ED}|$.

In empirical analysis, many indices have been developed to measure residential segregation. As Massey and Denton (1988) point out, different indices measure different dimensions of segregation such as evenness, exposure, clustering, and concentration, although all are highly correlated. In the artificial world, the function ρ (after being properly normalized) serves as a natural index of segregation. In particular, it measures the degree of exposure (degree of potential contact) between the members of the two ethnic groups. It indicates the extent to which blacks and whites physically confront one another by virtue of sharing a common residential area.[2]

The function ρ has another attractive property. It can serve as a potential function for this spatial game.[3] When two like-color agents exchange their residential locations, the residential pattern is not affected, and the value of function ρ does not vary either. Thus we can focus our attention on the cases in which a black and a white agent are chosen.

Figs. 2 and 3 illustrate two types of black–white switch. As we shall show below, the changes in ρ are always proportional to the changes in the moving agents' payoffs.

2.2.1. Type I

Consider two agents 1 and 2. Agent 1 is at point A in her payoff profile, and 2 is at point B. Suppose agents 1 and 2 decide to switch residential locations. After the switch, 1 ends

[2] In the empirical literature, two related measures of exposure, isolation and interaction are commonly used. The isolation index measures the degree to which minority members are exposed only to one another, which is computed as the minority-weighted average of the minority proportion of the population in each neighborhood. The interaction index measures the exposure of minority members to majority members, which is computed as the minority-weighted average of the majority proportion in each neighborhood. When there are only two groups, the two indexes sum to 1. High values of isolation and low values of interaction indicate high levels of segregation. See Massey and Denton for more detailed discussion.

[3] A game is a potential game if the changes in every player's payoff can be characterized by the first difference of a function. The function is then called the potential function of the game (Monderer and Shapley, 1996).

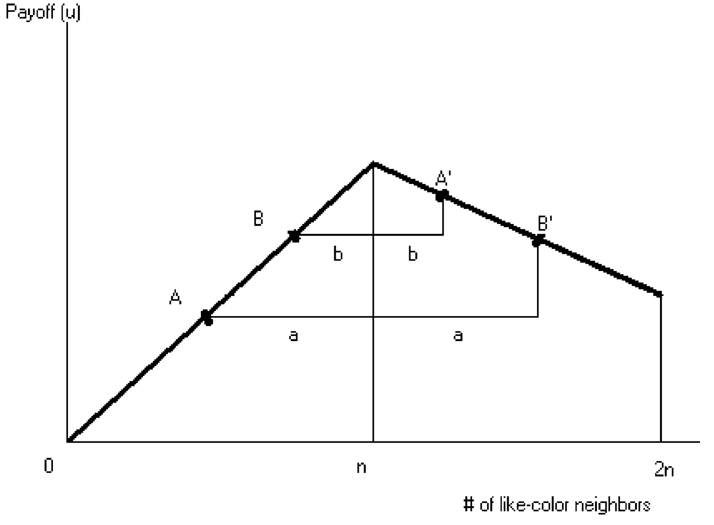

Fig. 2. Type I switch.

up with A′ and 2 ends up with B′ (Fig. 2). Let *a* be the horizontal distance from A to *n* and *b* be the horizontal distance from B to *n*. These two agents' payoffs before and after the switch are then summarized in Table 2.

We can now compute agent l's gain: $(2Z - M) + (M - Z)(n + b)/n - Z(n - a)/n = \{Mb + Z(a - b)\}/n$. Agent 2's gain is $(2Z - M) + (M - Z)(n + a)/n - Z(n - b)/n = \{Ma + Z(b - a)\}/n$. Their gains sum up to $(Ma + Mb)/n = 2(a + b) * M/(2n)$. Here $M/(2n)$

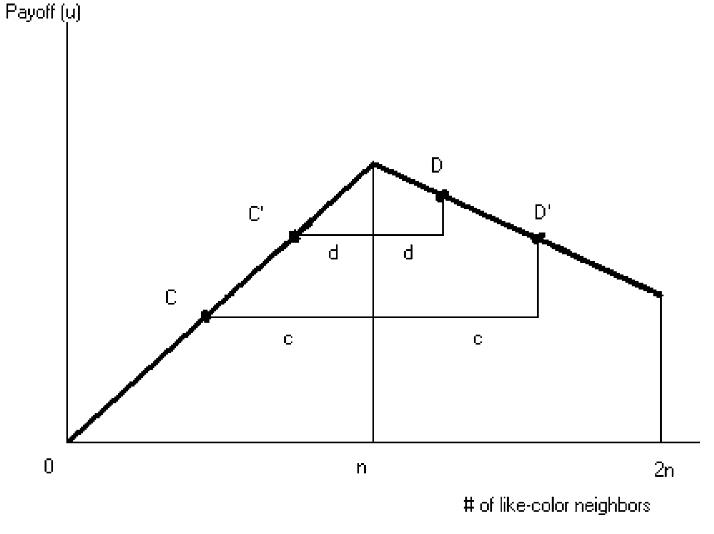

Fig. 3. Type II switch.

Table 2
Utility changes in a type I switch

Location	Utility
A	$\frac{Z(n-a)}{n}$
B	$\frac{Z(n-b)}{n}$
A′	$(2Z-M)+\frac{(M-Z)(n+b)}{n}$
B′	$(2Z-M)+\frac{(M-Z)(n+a)}{n}$

is a constant, and $2(a+b)$ is the change (decrease) in the total number of neighboring black–white pairs (i.e., $\Delta\rho$).

2.2.2. *Type II*

Consider two agents 3 and 4. Agent 3 is at point C in her payoff profile and 4 is at point D. Suppose 3 and 4 decide to switch and they end up with points C′ and D′, respectively (Fig. 3). Let c be the horizontal distance from C to n and d be the horizontal distance from D to n. These two agents' payoffs before and after the switch are summarized in Table 3.

Agent 3's gain is $Z(n-d)/n - Z(n-c)/n = Z(c-d)/n$. Agent 4's gain is $(2Z - M) + (M-Z)(n+c)/n - (2Z-M) - (M-Z)(n+d)/n = (M-Z)(c-d)/n$. The sum is $(Mc - Md)/n = 2(c-d) * M/(2n)$. Again, $M/(2n)$ is the same constant, and $2(c-d)$ is the change (decrease) in the total number of neighboring black–white pairs (i.e., $\Delta\rho$).

Types I and II and their reverse switches in fact exhaust all possible types of black–white switches. We can inflate the value of ρ by a constant $M/(2n)$. Therefore, in general, if a switch increases the moving agents' payoffs by ΔU, the value of function ρ will decrease by ΔU:

$$\{u_1(\cdot|\text{switch}) + u_2(\cdot|\text{switch})\} - \{u_1(\cdot|\text{not switch}) + u_2(\cdot|\text{not switch})\}$$
$$= -\{\rho(\cdot|\text{switch}) - \rho(\cdot|\text{not switch})\}.$$

Hence, $-\rho$ is a potential function of this game. That is, utility-improving switches always reduce the total number of black–white neighboring pairs. In other words, if a

Table 3
Utility changes in a type II switch

Location	Utility
C	$\frac{Z(n-c)}{n}$
C′	$\frac{Z(n-d)}{n}$
D	$(2Z-M)+\frac{(M-Z)(n+d)}{n}$
D′	$(2Z-M)+\frac{(M-Z)(n+c)}{n}$

black agent and a white agent gain from trading residential locations, they will necessarily have more like-color neighbors after the switch. This is easy to understand if before the trade the black agent lived in a predominantly white neighborhood while the white agent lived in a predominantly black neighborhood. In this case, both agents have more like-color neighbors after the switch and hence the total number of black–white pairs becomes smaller. From the payoff profile, we know that in general an agent could also improve her utility by moving into a neighborhood with fewer like-color neighbors. For example, a white agent will be happier if moving out of a 80% white neighborhood to a 60% white neighborhood. However, in that case, she can never find a black trading partner because the black agent has to move out of a 40% black neighborhood to a 20% black neighborhood. The white agent's gain is not enough to compensate her partner's loss.

It is important to recognize that, in contrast to the utilities that measure individual effects of the switch, the potential function $-\rho$ summarizes the social effects. Particularly, ρ covers the social externalities caused by the moving agents. For example, when a white agent in a predominately black neighborhood decides to trade with a black agent in a predominantly white neighborhood, their personal utilities may increase. However, this move causes both neighborhoods to deviate further from the 50–50 racial mixture. Therefore, all their neighbors in the two neighborhoods, either white or black, suffer from this switch. The moving agents have no reason to take into account these externalities. However function ρ decreases as the switch leaves fewer black–white neighboring pairs in the society, reflecting the loss of utilities in the whole population as the society moves a step further toward segregation.

2.3. The main result

We use Λ_N to denote the $N \times N$ lattice. A state is defined as a function $x : \Lambda_N \rightarrow \{\text{black}, \text{white}\}$, which labels each site with the color of its occupant. Letting x^t be the state at time t, we have a finite Markov process. Let P^β denote the Markov process (its transition probability matrix). We call it a perturbed process because agents do not always make "correct" decisions, depending on the value of β. Small values of β imply big perturbations; the perturbation vanishes as β approaches to 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,[4] P^β has a unique *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. (Foster and Young, 1990): A state $x \in X$ is *stochastically stable* relative to the perturbed process if $\lim_{\beta\rightarrow\infty} \mu^\beta(x) > 0$. The *stochastically stable set* is the smallest set that contains all the stochastically stable states.

4 Karlin and Taylor (1975) is a standard reference.

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 very likely that the system is found to be in the stochastically stable set.

Let X be the set of all states. Define S as a set of states that minimize the value of function $\rho : S = \{x|\rho\{x\} \leq \rho(y), \forall y \in X\}$. A finite N guarantees that S is non-empty.

Proposition 1. *In the artificial world we just described, S is stochastically stable, i.e.,* $\lim_{t\to\infty} \lim_{\beta\to\infty} \Pr\{x^t \in S\} = 1$..

Proposition 1 is a special case of Theorem 6.1 in Young (1998). In general, in any potential game with the log-linear revision rule, the set of all the states that maximize the potential function is stochastically stable. In our model, $-\rho$ is the potential function, so the states that maximize $-\rho$ (minimize ρ) are stochastically stable. That is, in the long run, we observe such states almost all the time.

We know ρ measures inter-racial contacts in our model. If a state x^* is in S, it minimizes inter-racial contacts. Hence $x^* \in S$ represents a segregated configuration. It then follows:

Proposition 2. *In the long run, if β is large, residential segregation is observed almost all the time.*

Note that how much people prefer 50–50 neighborhoods is irrelevant here. The world is locked in segregation primarily due to people's attitudes toward the two extremes: all-black and all-white neighborhoods. This model, in spirit, is similar to the insight recently discovered in evolutionary game theory that shows that an evolutionary process is more likely to select a risk-dominant equilibrium rather than a Pareto-dominant equilibrium in a 2×2 game (Kandori et al., 1993; Young, 1993).

3. Agent-based simulation

Proposition 2 is a limiting result for large β. In this section, we use agent-based simulation to study the dynamics of segregation with finite parameter values. Although our analytical results do not depend on any specific neighborhood definition, one naturally wonders how neighborhood size may affect the time to converge. This is also examined here using simulation.

3.1. Neighborhood definitions and payoff functions

Three simple definitions of neighborhood are commonly used in the agent-based simulation literature: the von Neumann neighborhood, the Moore neighborhood, and the $r(2)$ neighborhood (see Fig. 4). In a von Neumann neighborhood, an agent considers the four adjacent agents as neighbors; a Moore neighborhood includes the eight surrounding agents; and the $r(2)$ neighborhood covers 12 agents inside a "circle" with radius 2.

We arbitrarily pick $Z = 1$ and $M = 0.6$ for Eq. (1) and obtain a payoff function as Eq. (2). Corresponding to each of the three neighborhoods, Eq. (2) gives a specific

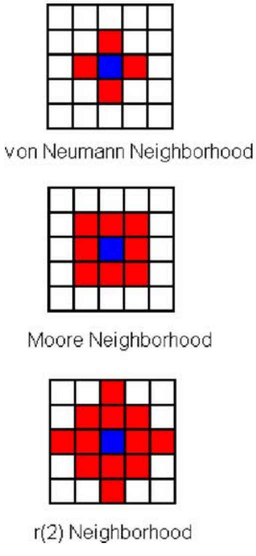

Fig. 4. Neighborhood definition.

payoff profile that yields value 0 for no like-color neighbor, peaks at value 1 for the half–half mixed neighborhood, and declines to 0.6 for 100 percent like-color neighbors (see Table 4).

$$u = \begin{cases} \dfrac{x}{n} & \text{if } x \leq n, \\ 1.4 - \dfrac{0.4x}{n} & \text{otherwise.} \end{cases} \quad (2)$$

Table 4
Payoff functions used in the simulation

x	$u^{VN}(x)\vert_{n=2}$	$u^{MN}(x)\vert_{n=4}$	$u^{r(2)}(x)\vert_{n=6}$
0	0	0	0
1	0.5	0.25	0.167
2	1	0.5	0.333
3	0.8	0.75	0.5
4	0.6	1	0.667
5	–	0.9	0.833
6	–	0.8	1
7	–	0.7	0.933
8	–	0.6	0.867
9	–	–	0.8
10	–	–	0.733
11	–	–	0.667
12	–	–	0.6

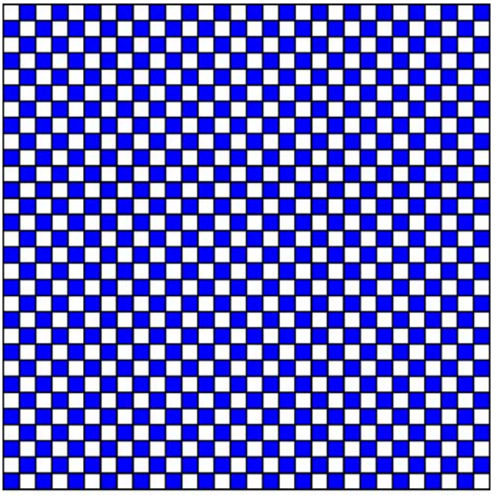

Fig. 5. A checkerboard initial state.

3.2. Emergence of segregation

We first run a simulation on a 100×100 landscape to show how segregation emerges spontaneously. We use the Moore neighborhood and proceed with $\beta = 10$. The utility profile is specified in Table 4.

The simulation starts with a checker board configuration, where every agent is living in a 50–50 neighborhood (Fig. 5). In this "Pareto" state, everyone is getting the highest payoff and the residential pattern is socially optimal. However, the "Pareto" state cannot last long because the cost of deviation is very low. Once some agents accidentally change their residential locations, their neighbors become less satisfied than before, so when chances come, they will move too. It is not hard to imagine that the moves tend to lead to segregation. Consider two neighborhoods, one of which is predominantly black and the other predominantly white. If a white agent is selected from the predominantly black neighborhood and a black agent is selected from the other, it pays for them to switch because both will have higher utilities after they do so. This makes the predominantly black neighborhood "blacker" and the predominantly white neighborhood "whiter." On the contrary, if a black agent is picked from the predominantly black neighborhood and a white agent is picked from the other, they are not likely to switch because they will suffer a loss for doing so. This means that once segregation emerges, it tends to persist.

In our simulation, we see segregation start to emerge as time goes on (Fig. 6). If we wait long enough, we see that in Fig. 7, blacks and whites are almost completely separated. The boundary between them is reducing to its minimum, and blacks and whites rarely share same neighborhoods. As the simulation shows, once a clear color line is established, neighborhood transition becomes stagnant. Obviously, many agents are not happy with this

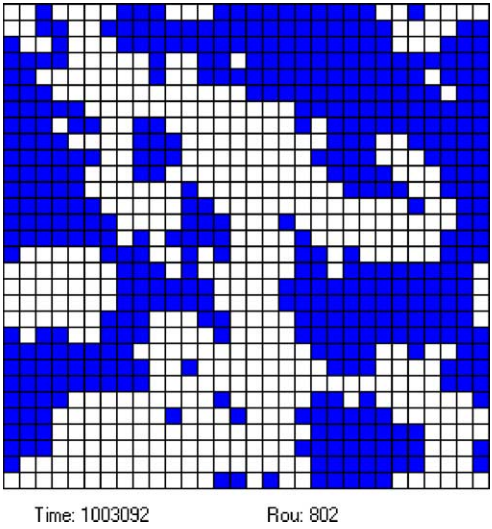

Fig. 6. A snapshot in the short run.

situation. However, they can do nothing about it individually. In a sense, this is a problem of coordination failure.

3.3. Neighborhood, β, and waiting time

Proposition 2 predicts that segregation emerges in the long run. Thus, it is important to know how long the long run really is. Although in theory the period of time is abstractly

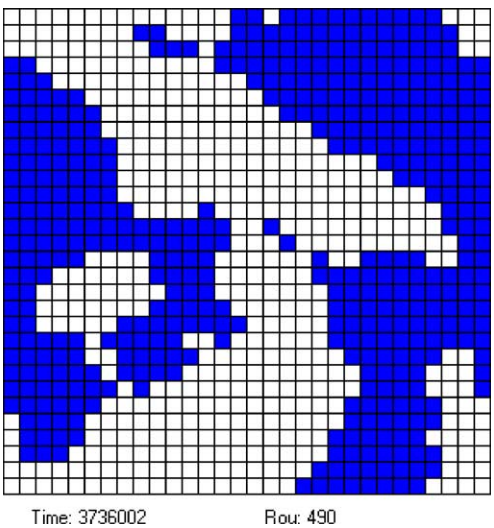

Fig. 7. A snapshot in the long run.

defined and thus hard to relate to real world time, it always makes sense to ask what affects the speed of convergence. Ellison (1993) first illustrated that local interaction has a big impact on transition time. Particularly, if agents play games with others in a smaller neighborhood, all else being equal, a shorter time is needed to reach a situation in which a large proportion of the population play an equilibrium strategy. Young (1998) further showed that the transition time is bounded from above regardless of the population size if the interaction structure satisfies certain conditions.

In our model, multiple factors determine the waiting time before a certain degree of segregation emerges. For example, coefficient β, neighborhood structure, initial state, population size, and the method used to pair trading partners are all relevant parameters. Since the effects of the last three are fairly transparent, we focus our attention on β and neighborhood definition. Two points are worth noting here. First, the value of β is meaningful only relative to the unit of the payoff function $u(x)$. One can always inflate the payoff values, reduce the value of β accordingly, and have exactly the same model. Second, the neighborhood in our model is a "reference" group instead of an "interaction" group as in many other spatial games (e.g., games analyzed by Ellison, 1993). In our model, an agent switches residential location with an agent in another neighborhood based on the racial composition of the two neighborhoods. In Ellison (1993), an agent plays the game only with other agents in the same neighborhood. Because of this difference, Ellison's results are not directly applicable here.

Again, we run the simulation on a 100×100 landscape. We use different neighborhood definitions and, under each definition, different β values. Simulations start with random initial states. We measure the waiting time before the absolute value of the potential function (ρ) falls below 600.[5] Payoff function are specified as in Table 4, and the results are plotted in Fig. 8.

Remember, the value of parameter β determines the relative importance of the random utility. A β equal to zero means all agents move randomly without reference to racial composition. In that case, segregation is almost never achievable and the waiting time is close to infinity. In fact, for $\beta < 2$, our computer simulations never reach a state with a potential below 600. The waiting time declines sharply with β varying in a small range around 2: at $\beta = 4$, simulations always end fairly quickly. As β increases from 4 to around 40, the expected waiting time decreases at ever slower rates. When β goes beyond 40, the expected waiting time responds very little to β values (Fig. 8). Therefore, although the limiting result of Proposition 2 holds only as β approaches infinity, we could expect to see a segregational pattern fairly close to the limit with a finite β.

If every agent views the whole landscape as a single neighborhood, then moving does not alter the neighborhood's racial composition. In that case, moving becomes a random decision, and therefore the waiting time for segregation is infinitely long. Obviously, segregation will emerge and persist only if agents consider neighborhood a local concept. One naturally wonders, all else being equal, whether a smaller neighborhood always implies a shorter waiting time. This is a particularly important question because agent-based

[5] Notice, the potential function is defined with respect to neighborhood definition. In order to make results under different neighborhood structures comparable, here we always measure the potential with respect to the Moore neighborhood.

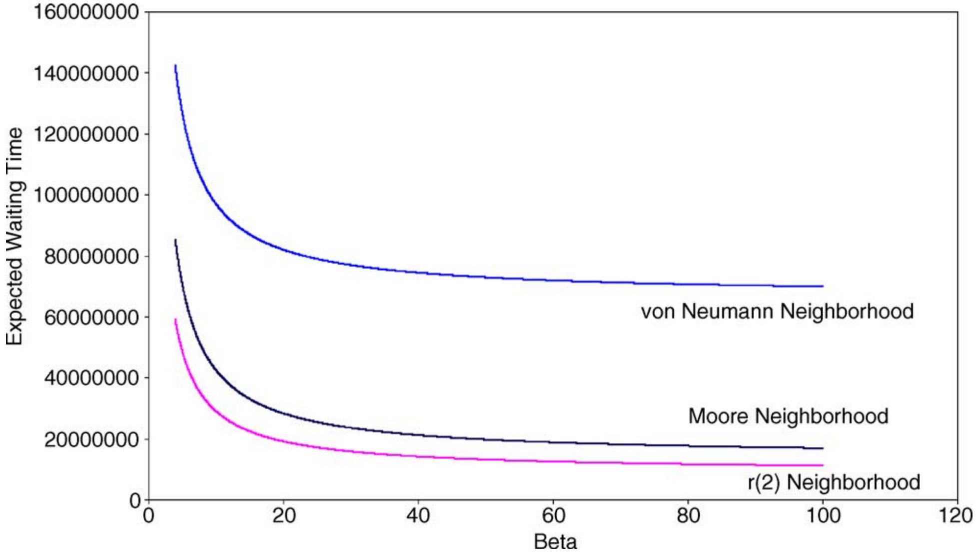

Fig. 8. Waiting time.

models almost always use small neighborhoods, much smaller than most real world notions of neighborhoods (see, e.g., Epstein and Axtell, 1996).

We run our simulation under different neighborhood definitions and compare the expected waiting time in Fig. 8. Interestingly, we find that a bigger neighborhood actually decreases the waiting time for segregation. Under the Moore neighborhood in which an agent has eight neighbors, the expected waiting time is less than half of that under the von Neumann neighborhood which only assigns four neighbors to each agent. The $r(2)$ neighborhood covers 12 neighbors for each agent, leading to even a shorter waiting time, although not much shorter than under the Moore neighborhood.

The relation between neighborhood size and waiting time becomes easier to understand if we consider different determinants of waiting time that are affected by neighborhood size. The "potential game" property of our model again helps clarify our thinking. The transition time to segregation is nothing but the time it takes for function ρ to decline from a high value to some cutoff value (in this case, 600). We call a switch *advantageous* if it could increase the two agents' utilities. Then, how soon ρ reaches the cutoff depends on the following factors:

(1) how quickly advantageous trading pairs are formed,
(2) how likely that a pair will switch residential locations,
(3) how much reduction of ρ results from each advantageous switch.

Of course, agents may mistakenly make disadvantageous switches. However, since such activities are rare given large β, we focus on factors (1)–(3). Suppose we shift from Moore neighborhoods to smaller von Neumann neighborhoods. When agents have fewer neighbors, they are less differentiated by neighborhood characteristics. That is, under the von

Neumann neighborhood, an agent finds many other people in the society living in the same kind of neighborhood as she does. As a consequence, it now takes longer to form an advantageous trading pair. On the other hand, once a pair is matched, the potential gain is higher, and hence the switch is more likely to happen, reducing ρ by a larger amount. When β is large, any positive gain will lead to a switch almost surely, thus factor (2) makes little difference. Factors (1) and (3) work in opposite directions. Our simulation results suggest that (1) dominates (3). That is, when neighborhoods are smaller, although each switch deducts a bigger value from ρ, it actually takes a lot more time to find such a trading pair. Therefore overall, smaller neighborhoods take longer to segregate. Notice that this is similar to the general idea that evolution tends to be more rapid when it may proceed via a sequence of smaller steps rather than requiring sudden large changes (Ellison, 2000).

This result should be understood with two caveats. First, larger neighborhoods may lead to shorter waiting time only when the neighborhood size is much smaller relative to the whole space. As neighborhood size approaches the whole area size, we know that segregation will never happen. Second, factor (1) works against small neighborhoods partly because of the way our model dynamics are defined. In particular, our model assumes that two agents are paired at random, regardless of the size of their potential gains. In reality, this may not be true since discontent agents may be motivated to find trading partners more quickly through information channels such as advertisements.

The simulation allows us to try more variations of the model. For example, we also learned that the linearity of the payoff function is not crucial for our qualitative results, although it simplifies our analysis substantially. What is crucial is that people feel happier in a neighborhood where their own color forms the majority than in a neighborhood where they are the minority. As survey evidence shows, this kind of residential preference is fairly common in reality.

4. Concluding remarks

Following Schelling (1971, 1978), we have constructed a dynamic model to understand the evolution of residential segregation. While Schelling's insights are enlightening, his original paper took an inductive approach and demonstrated all his results with simple examples. For more than three decades, the well-known Schelling model has never been mathematically analyzed. Our paper is a first attempt to put a Schelling-type model on rigorous footing. We formulate neighborhood transition as a spatial game played on a lattice graph. Segregation is characterized as a stochastically stable state that tends to emerge and persist in the long run regardless of the initial state.

Our mathematical approach has improved upon Schelling's analysis of segregation in at least three aspects. First, our model proposes an analytical measure of segregation. In Schelling's original model and in those variations analyzed by others using computer simulations, segregational patterns are always presented as a visual effect. That is, researchers simply look at simulation results to tell whether segregation emerges. Here we use the total number of black–white neighboring pairs to quantify segregation, enabling us to determine whether one configuration is more segregated than another. Conceptually, it is equivalent to the indices of inter-racial interaction widely used in empirical literature. Our measure is potentially useful in other agent-based models of segregation.

Second, our model explicitly defines agents' utilities, thus allowing us to make explicit welfare implications. In any free society, people move according to their own interests. When a person chooses a neighborhood, she helps redefine the neighborhood. Her entry may affect other neighbors' welfare, an aspect not taken into account when the person is making the decision. As a consequence, individually optimal actions may lead to a socially sub-optimal outcome. This kind of externality from social dynamics is made transparent in the potential function of our model that relates segregation, a non-optimum social outcome, to utility gains at the individual level.

Third, our model borrows the notion of stochastic stability from evolutionary game theory to characterize rigorously the phenomenon of residential segregation. For three decades, segregation in the Schelling model was known as an "emergent and persistent phenomenon." We translate this loose description into a rigorous equilibrium concept. The analysis of segregation is thus embedded in a game-theoretic framework in which a wide range of existing results and analytical tools are readily applicable.

The mathematical analysis is particularly insightful to highlight the stability property of segregation. In a dynamic model of segregation, there exist millions of Nash equilibria. However, to identify all Nash equilibria is not very useful. Remember, in a Nash equilibrium, no single agent has incentive to deviate, but if any agent makes a "stupid" move, the equilibrium may fall apart. Based on this idea, the stochastic stability concept asks which state is the most stable one. This state is the most difficult one to tip away and takes a large number of "stupid" moves to alter. By definition, the most stable state will prevail, regardless of its optimality. Scholars who study residential segregation with survey data only identify what residential pattern is most acceptable to people. However, without the knowledge of its stability, one has no reason to expect such a residential pattern to emerge and persist.

Our model has shown that segregation may prevail in an all-integrationist world because of its stability. Elimination of discrimination and fair housing legislation are not sufficient for residential desegregation, even if desegregation is an ideal of every individual in the society. If it is believed that mixed neighborhoods are socially optimal residential patterns, more specific public policies leading in that direction are warranted.

Acknowledgements

This work was completed at the Brookings Institution with the support of its Leo Model Research Fellowship. I would like to thank Professor Peyton Young for his guidance. I am also grateful to a coeditor and an anonymous referee for their invaluable comments and suggestions. Nikesh Patel's assistance is greatly appreciated.

References

Blume, L.E., 1997. Population games. In: Arthur, W.B., Durlauf, S.N., Lane, D.A. (Eds.), The Economy as an Evolving Complex System, vol. II. Addison-Wesley, Reading, MA, pp. 425–460.

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., 2001. Interaction-based models. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, Amsterdam, pp. 3297–3380.

Davis, J.A., Smith, T.W., 1999. General Social Surveys, 1972–1998: Cumulative Codebook. National Opinion Research Center, Chicago.

Ellison, G., 1993. Learning, local interaction, and coordination. Econometrica 61, 1047–1071.

Ellison, G., 2000. Basins of attraction, long run stochastic stability, and the speed of step-by-step evolution. Review Economic Studies 67, 17–45.

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

Farley, R., Schuman, H., Bianchi, S., Colasanto, D., Hatchett, S., 1978. Chocolate city, vanilla suburbs: will the trend toward racially separate communities continue? Social Science Research 7, 319–344.

Farley, R., Fielding, E.L., Krysan, M., 1997. The residential preferences of blacks and whites: a four-metropolis analysis. Housing Policy Debate 8, 763–800.

Foster, D., Young, H.P., 1990. Stochastic evolutionary game dynamics. Theoretical Population Biology 38, 319–332.

Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. Econometrica 61, 29–56.

Karlin, S., Taylor, H.M., 1975. A First Course in Stochastic Processes. Academic Press, New York.

Massey, D.S., Denton, N.A., 1988. The dimensions of residential segregation. Social Forces 67, 281–315.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontier in Econometrics. Academic Press, New York, pp. 105–142.

Monderer, D., Shapley, L.S., 1996. Potential games. Games Economic Behavior 14, 124–143.

Schelling, T.C., 1971. Dynamic models of segregation. Journal of Mathematical Sociology 1, 143–186.

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

Young, H.P., 1993. The evolution of conventions. Econometrica 61, 57–84.

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