# Exact solution for a metapopulation version of Schelling's model

**Richard Durrett**[1] **and Yuan Zhang**

Department of Mathematics, Duke University, Durham, NC 27708

Contributed by Richard Durrett, August 4, 2014 (sent for review July 23, 2014)

**In 1971, Schelling introduced a model in which families move if they have too many neighbors of the opposite type. In this paper, we will consider a metapopulation version of the model in which a city is divided into $N$ neighborhoods, each of which has $L$ houses. There are $\rho NL$ red families and $\rho NL$ blue families for some $\rho < 1/2$. Families are happy if there are $\leq \rho_c L$ families of the opposite type in their neighborhood and unhappy otherwise. Each family moves to each vacant house at rates that depend on their happiness at their current location and that of their destination. Our main result is that if neighborhoods are large, then there are critical values $\rho_b < \rho_d < \rho_c$, so that for $\rho < \rho_b$, the two types are distributed randomly in equilibrium. When $\rho > \rho_b$, a new segregated equilibrium appears; for $\rho_b < \rho < \rho_d$, there is bistability, but when $\rho$ increases past $\rho_d$ the random state is no longer stable. When $\rho_c$ is small enough, the random state will again be the stationary distribution when $\rho$ is close to 1/2. If so, this is preceded by a region of bistability.**

segregation | large deviations

In 1971, Schelling (1) introduced one of the first agent-based models in the social sciences. Families of two types inhabit cells in a finite square, with 25–30% of the squares vacant. Each family has a neighborhood that consists of a $5 \times 5$ square centered at their location. Schelling used a number of different rules for picking the next family to move, but the most sensible seems to be that we pick a family at random on each step. If the fraction of neighbors of the opposite type is too large, then they move to the closest location that satisfies their constraints. Schelling simulated this and many other variants of this model (using dice and checkers) to argue that if people have a preference for living with those of their own color, the movements of individual families invariably led to complete segregation (2).

As Clark and Fossett (3) explain "The Schelling model was mostly of theoretical interest and was rarely cited until a significant debate about the extent and explanations of residential segregation in US urban areas was engaged in the 1980s and 1990s. To that point, most social scientists offered an explanation that invoked housing discrimination, principally by whites." At this point Schelling's article has been cited more than 800 times. For a sampling of results from the social sciences literature, see Fossett's lengthy survey (4) or other more recent treatments (5–7). About 10 y ago, physicists discovered this model and analyzed a number of variants of it using techniques of statistical mechanics (8–14). However, to our knowledge, the only rigorous work is ref. 15, which studies the 1D model in which the threshold for happiness is $\rho_c = 0.5$ and two unhappy families within distance $w$ swap places at rate 1.

Here, we will consider a metapopulation version of Schelling's model in which there are $N$ neighborhoods that have $L$ houses, but we ignore spatial structure within the neighborhoods and their physical locations. We do this to make the model analytically tractable, but these assumptions are reasonable from a modeling point of view. Many cities in the United States are divided into neighborhoods that have their own identities. In Durham, these neighborhoods have names like Duke Park, Trinity Park, Watts-Hillendale, Duke Forest, Hope Valley, Colony Park, etc. They are often separated by busy roads and have identities that are reinforced by e-mail newsgroups that allow people to easily communicate with everyone in their neighborhood. Because of this, it is the overall composition of the neighborhood that is important not just the people who live next door. In addition, when a family decides to move they can easily relocate anywhere in the city.

Families, which we suppose are indivisible units, come in two types that we call red and blue. There are $\rho NL$ of each type, leaving $(1 - 2\rho)NL$ empty houses. This formulation was inspired by Grauwin et al. (16), who studied segregation in a model with one type of individual whose happiness is given by a piecewise linear unimodal function of the density of occupied sites in their neighborhood. To define the rules of movement, we introduce the threshold level $\rho_c$ such that a neighborhood is happy for a certain type of agent if the fraction of agents of the opposite type is $\leq \rho_c$. For each family and empty house, movements occur at rates that depend on the state of the source and destination houses:

| From/to | Happy | Unhappy |
|---|---|---|
| Happy | $r/(NL)$ | $\epsilon/(NL)$ |
| Unhappy | $1/(NL)$ | $q/(NL)$ |

where $q$, $r < 1$, and $\epsilon$ are small, e.g., 0.1 or smaller. Because there are $O(NL)$ vacant houses, dividing the rates by $NL$ makes each family moves at a rate $O(1)$. Because $\epsilon$ is small, happy families are very reluctant to move to a neighborhood in which they would be unhappy, whereas unhappy families move at rate 1 to neighborhoods that will make them happy. As we will see later, the equilibrium distribution does not depend on the values of the rates $q$ and $r$ for transitions that do not change a family's happiness. We do not have an intuitive explanation for this result.

## Significance

**More than 40 y ago, Schelling introduced one of the first agent-based models in the social sciences. The model showed that even if people only have a mild preference for living with neighbors of the same color, complete segregation will occur. This model has been much discussed by social scientists and analyzed by physicists using analogies with spin-1 Ising models and other systems. Here, we study the metapopulation version of the model, which mimics the division of a city into neighborhoods, and we present the first analysis to our knowledge that gives detailed information about the structure of equilibria and explicit formulas for their densities.**

Author contributions: R.D. and Y.Z. performed research and wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: rtd@math.duke.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1414915111/-/DCSupplemental.

## Convergence to a Deterministic Limit

To describe the dynamics more precisely, let $n_{i,j}(t)$, be the number of neighborhoods with $i$ red and $j$ blue families for $(i, j) \in \Omega = \{(i, j)\colon i, j \geq 0, i + j \leq L\}$. The configuration of the system at time $t$ can be fully described by the numbers $\nu_t^N(i,j) = n_{ij}(t)/N$. If one computes infinitesimal means and variances, it is natural to guess (and not hard to prove) that if we keep $L$ fixed and let $N \to \infty$, then $\nu_t^N$ converges to a deterministic limit.

Motivated by individual-based models in finance, Remenik (17) proved a general result that takes care of our example. To describe the limit, we need some notation. Let $\lambda(a_1, b_1; a_2, b_2)$ be $N$ times the total rate of movement from one $(a_1, b_1)$ neighborhood to one $(a_2, b_2)$ neighborhood. Let $b(\omega_1, \omega_2; \omega_1', \omega_2')$ be $N$ times the rate at which a movement from one $\omega_1 = (a_1, b_1)$ neighborhood to one $\omega_2 = (a_2, b_2)$ neighborhood turns the pair $\omega_1$, $\omega_2$ into $\omega_1', \omega_2'$. The exact formulas for these quantities are not important, so they are shown in *SI Text*.

**Theorem 1.** *As $N \to \infty$ the $\nu_t^N(i,j)$ converge in probability to the solution of the ordinary differential equation:*

$$\frac{d\nu_t(i,j)}{dt} = -\nu_t(i,j)\sum_{\omega\in\Omega}[\lambda(i,j;\omega)+\lambda(\omega;i,j)]\nu_t(\omega) + \sum_{\omega_1,\omega_2\in\Omega}\{b[\omega_1,\omega_2;(i,j),\omega'] + b[\omega_1,\omega_2;\omega',(i,j)]\} \times \nu_t(\omega_1)\nu_t(\omega_2). \quad [1]$$

We do not sum over $\omega'$ because its value is determined by $\omega_1$ and $\omega_2$. The first term comes from the fact that a migration from $(i, j) \to \omega$ or $\omega \to (i, j)$ destroys an $(i, j)$ neighborhood, whereas the second reflects the fact that a migration $\omega \to \omega'$ may create an $(i, j)$ neighborhood at the source or at the destination.

## Special Case $L = 2$

To illustrate the use of Theorem 1, we consider the case $L = 2$, and let $\ell_c = [\rho_c L]$ be the largest number of neighbors of the opposite type that allows a family to be happy. Here, $[x]$ is the largest integer $\leq x$. When $L = 2$, a neighborhood with both types of families must be (1, 1), so the situation in which $\ell_c \geq 1$ is trivial because there are never any unhappy families. In the case $L = 2$ and $\ell_c = 0$, it is easy to find the equilibrium because there is detailed balance, i.e., the rate of each transition is exactly balanced by the one in the opposite direction.

$$\begin{aligned}
r\nu_{1,0}^2 &= 4r\nu_{0,0}\nu_{2,0} && (1,0)(1,0) \rightleftharpoons (0,0)(2,0)\\
r\nu_{0,1}^2 &= 4r\nu_{0,0}\nu_{0,2} && (0,1)(0,1) \rightleftharpoons (0,0)(0,2)\\
2\nu_{0,0}\nu_{1,1} &= \epsilon\nu_{1,0}\nu_{0,1} && (1,1)(0,0) \rightleftharpoons (1,0)(0,1)\\
\nu_{1,0}\nu_{1,1} &= 2\epsilon\nu_{2,0}\nu_{0,1} && (1,1)(1,0) \rightleftharpoons (0,1)(2,0)\\
\nu_{0,1}\nu_{1,1} &= 2\epsilon\nu_{0,2}\nu_{1,0} && (1,0)(1,1) \rightleftharpoons (1,0)(0,2).
\end{aligned}$$

After a little algebra (*SI Text*), we find that this holds if and only if

$$\nu_{2,0} = \nu_{0,2} = x \quad \nu_{1,1} = 2\epsilon x \quad \nu_{1,0} = \nu_{0,1} = y \quad \nu_{0,0} = y^2/4x.$$

At first, it may be surprising that the rate $r$ has nothing to do with the fixed point, but if you look at the first two equations you see that the $r$ appears on both sides. The parameter $q$ does not appear either, but when $L = 2$, it is for the trivial reason that transitions $(1, 1)(1, 0) \to (1, 0)(1, 1)$, which occur at rate $q$, do not change the state of the system.

Using now the fact that the equilibrium must preserve the red and blue densities, we can solve for $x$ and $y$ to conclude that

$$y = \frac{1-\sqrt{8(1-\epsilon)\rho^2 - 4(1-\epsilon)\rho + 1}}{1-\epsilon},$$

$$x = \frac{(2-2\epsilon)\rho - 1 + \sqrt{8(1-\epsilon)\rho^2 - 4(1-\epsilon)\rho + 1}}{2(1+\epsilon)(1-\epsilon)}.$$

The argument given here shows that this is the only fixed point that satisfies detailed balance. We prove in *SI Text* that it is the only fixed point. Because the formulas, which result from solving a quadratic equation, are somewhat complicated, Fig. 1 shows how the equilibrium probabilities $\nu_{i,j}$ vary as a function of $\rho$.

Unfortunately, when $L \geq 3$, there is no stationary distribution that satisfies detailed balance. One can, of course, solve for the stationary distribution numerically. Fig. 2 shows the limit behavior of the system with $L = 20$ and $\rho_c = 0.3$, i.e., $\ell_c = 6$ for initial densities $\rho = 0.1$, 0.2, 0.25, and 0.35. In the first two cases, most of the families are happy. In the third situation, the threshold is $\ell_c = 6$, whereas the average number of reds and blues per neighborhood is five, but because fluctuations in the makeup of neighborhoods can lead to unhappiness, there is a tendency toward segregation. In the fourth case, segregation is almost complete, with most neighborhoods having 0 or 1 of the minority type.

## Outline of Our Solution

Finding the stationary distribution requires solving roughly $L^2/2$ equations. To be precise, there are 231 equations when $L = 20$ and 5,151 when $L = 100$. In this section, we will adopt a different approach: we concentrate on the evolution of neighborhood 1 and consider the other $N - 1$ neighborhoods to be its environment, which can be summarized by the following four parameters: $(i)$ the average number of happy red and blue families per neighborhood, $h_R^1$ and $h_B^1$, and $(ii)$ the average number of vacant sites happy for red or blue, $h_R^0$ and $h_B^0$, again per neighborhood.

The first in our solution is to identify the stationary distribution of the neighborhood–environment chain that are self-consistent. That is, if we calculate the expected values of $h_R^1$, $h_R^0$, $h_B^1$, and $h_B^0$ in equilibrium in neighborhood, they agree with the original parameters. To begin to do this, we divide the state space $\Omega$ into four quadrants based on red and blue happiness. Writing 0 for $H$ and 1 for $U$, we have

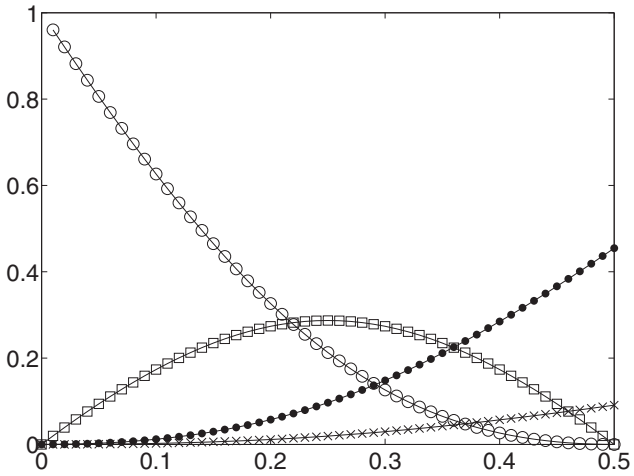


**Fig. 1.** Equilibrium probabilities $\nu_{i,j}$ for the case $L = 2$ plotted against $\rho$.

| | | |
|---|---|---|
| $j > \ell_c$ | $Q_{1,0}$ | $Q_{1,1}$ |
| $j \le \ell_c$ | $Q_{0,0}$ | $Q_{0,1}$ |
| | $i \le \ell_c$ | $i > \ell_c$ |

If we let $Tri(p_R, p_B)$ be the trinomial distribution

$$\frac{L!}{i!j!(L-i-j)!} p_R^i p_B^j (1-p_R-p_B)^{L-i-j}, \qquad [2]$$

then inside $Q_{k,\ell}$, the detailed balance condition is satisfied by $Tri(p_R, p_B)$ where

$$p_R = \frac{\alpha_{k,\ell}}{1+\alpha_{k,\ell}+\beta_{k,\ell}}, \quad p_B = \frac{\beta_{k,\ell}}{1+\alpha_{k,\ell}+\beta_{k,\ell}},$$

and formulas for $\alpha_{k,\ell}$ and $\beta_{k,\ell}$ are given in *SI Text*.

Unfortunately, there is no stationary distribution that satisfies detailed balance on the entire state space. To verify this, we note that if the stationary distribution $\pi$ and the jump rates $q$ satisfy detailed balance $\pi(x)q(x, y) = \pi(y)q(y, x)$, then around any closed path $x_0, x_1, \ldots x_n = x_0$ with $q(x_{i-1}, x_i) > 0$ for $1 \le i \le n$, we must have

$$\prod_{i=1}^{n} \frac{q(x_{i-1}, x_i)}{q(x_{i-1}, x_i)} = 1,$$

but this is not satisfied around loops that visit two or more quadrants. To avoid this problem, we cut the connections between the different quadrants and identify the self-consistent distributions for the modified chain. At this point, we can only do this under

**Assumption 1.** *Stationary distributions are symmetric under interchange of red and blue.*

The answer given in the next section is a one-parameter family of stationary distributions indexed by $a \in [0, 1/2]$. There, and in the next two steps, we have a pair of results: one for $\rho < \rho_c$ and one for $\rho > \rho_c$.

In the second step, we investigate the flow of probability between quadrants when transitions between the quadrants are restored. The key idea is that the measures in each quadrant are trinomial, so the probabilities will decay exponentially away from the mean $(p_R L, p_B L)$. This observation implies that the flow between quadrants occurs at rate $\exp(-cL)$, which is much smaller than the time, $O(1)$, it takes the probability distributions to reach equilibrium. In words, the process comes to equilibrium on a fast time scale, whereas the parameters change on a much slower one. We will prove this separation of time scales in a version of the paper for a mathematical audience. Here, we will only give the answers that result under

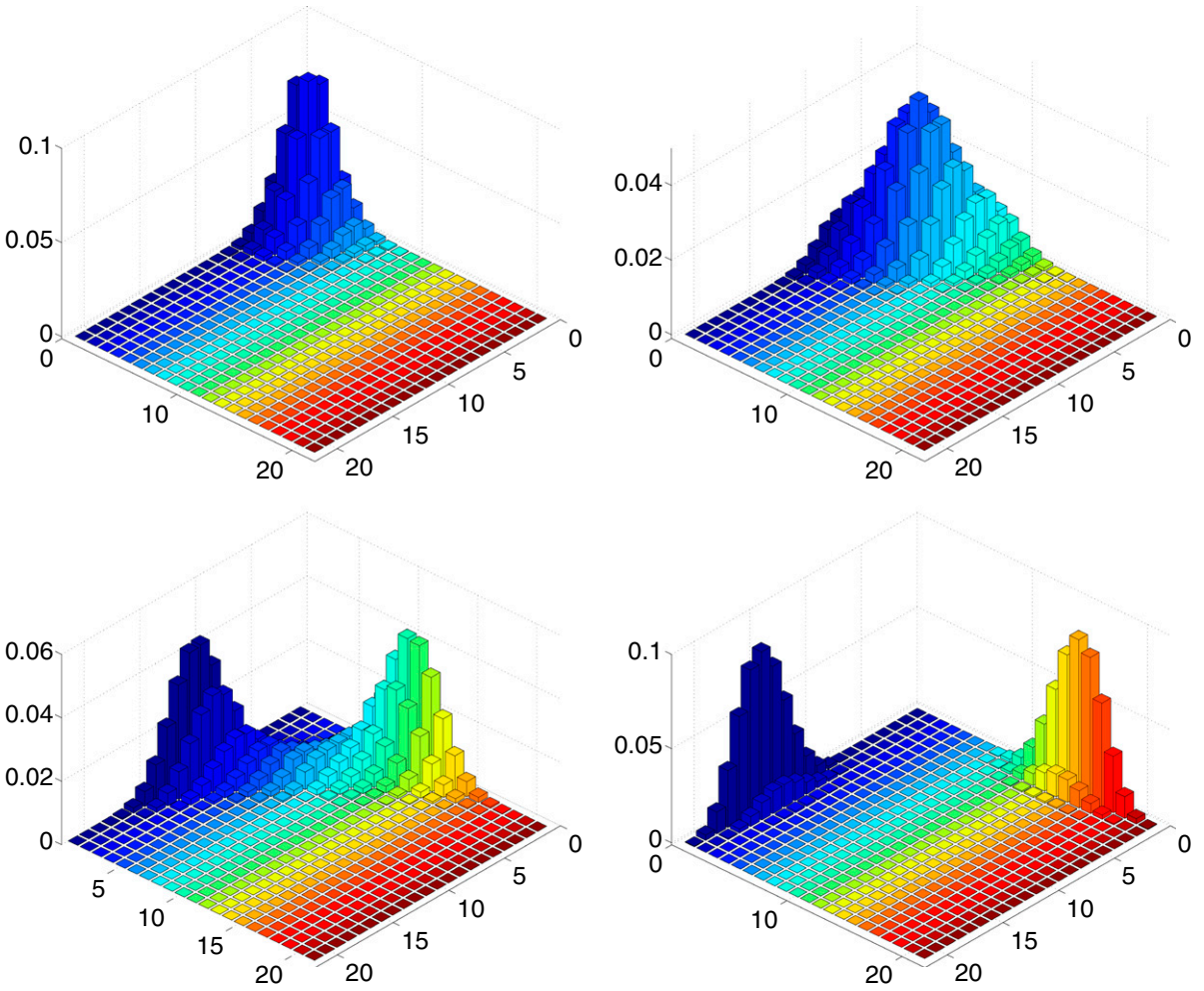


**Fig. 2.** Limiting behavior of limit differential equation, with $\rho_c = 0.3$, $\epsilon = 0.01$, and $\rho = 0.1$, 0.2, 0.25, and 0.35.

**Assumption 2.** *The process is always in one of self-consistent stationary distributions, but the value of a changes over time.*

The third step is to use the stability results to show that the only possible stable equilibria are at the following end points: $a = 0$, which represents a random distribution; and $a = 1/2$, which represents a segregated state.

We will call these measures $\mu_r$ and $\mu_s$ when $\rho < \rho_c$ and $\hat{\mu}_r$ and $\hat{\mu}_s$ when $\rho > \rho_c$. Theorems 4A and 4B describe when they are stable fixed points.

## Self-Consistent Stationary Distributions

The results given here are proved in *SI Text*. The formulas, which again come from solving a quadratic equation, are not simple but they are explicit and easily evaluated.

**Theorem 2A.** *Suppose $\rho < \rho_c$. For $a \in (0, 1/2]$ let*

$$\rho_1(a,\rho) = \frac{-1 + (a+\rho)(1-\epsilon) + \Delta_1}{2a(1-\epsilon^2)}, \qquad [3]$$

*where*

$$\Delta_1 = \sqrt{[1-(a+\rho)(1-\epsilon)]^2 + 4a(1-\epsilon^2)\rho}.$$

Let $\rho_1(0, \rho) = \lim_{a\to 0} \rho_1(a, \rho) = \rho/[1 - \rho(1 - \epsilon)]$ and for $a \in [0, 1/2]$ let

$$\mu_a = (1-2a)\mathrm{Tri}(\rho_0,\rho_0) + a\mathrm{Tri}(\rho_1,\rho_2) + a\mathrm{Tri}(\rho_2,\rho_1).$$

*A symmetric distribution $\mu$ is self-consistent if and only if it has the form above with parameters $\rho_1 > \rho_c$, $\rho_2 = \epsilon\rho_1 < \rho_c$, and $\rho_0 = \rho_1/[1 + (1 - \epsilon)\rho_1] < \rho_c$.*

To clarify the last sentence: the definition of $\rho_1$ does not guarantee that the three conditions are satisfied for all values of $a \in [0, 1/2]$, so the inequalities are additional conditions. As shown in *SI Text*, $a \to \rho_1(a, \rho)$ is increasing, so the range of possible values of $\rho_1$ for a fixed value of $\rho$ is

$$[\rho_1(0,\rho), \rho_1(1/2,\rho)] = \left[\frac{\rho}{1-\rho(1-\epsilon)}, \frac{2\rho}{1+\epsilon}\right]. \qquad [4]$$

The possible self-consistent stationary distributions are similar in the second case but the formulas are different.

**Theorem 2B.** *Suppose $\rho \geq \rho_c$. For $a \in (0, 1/2]$ let*

$$\hat{\rho}_1(a,\rho) = \frac{\epsilon + (1-\epsilon)(a+\rho) - \Delta_2}{2a(1-\epsilon^2)}, \qquad [5]$$

*where*

$$\Delta_2 = \sqrt{[\epsilon + (1-\epsilon)(a+\rho)]^2 - 4a(1-\epsilon^2)\rho}.$$

Let $\hat{\rho}_1(0,\rho) = \lim_{a\to 0}\hat{\rho}_1(a,\rho) = \rho/[\epsilon + (1-\epsilon)\rho)]$, and for $a \in [0, 1/2]$ let

$$\hat{\mu}_a = a\mathrm{Tri}(\hat{\rho}_1,\hat{\rho}_2) + a\mathrm{Tri}(\hat{\rho}_2,\hat{\rho}_1) + (1-2a)\mathrm{Tri}(\hat{\rho}_3,\hat{\rho}_3).$$

*A symmetric distribution $\hat{\mu}$ is self-consistent if and only if it has the form above with parameters $\hat{\rho}_1 > \rho_c$, $\hat{\rho}_2 = \epsilon\hat{\rho}_1 < \rho_c$ and $\hat{\rho}_3 = \epsilon\hat{\rho}_1/[1-(1-\epsilon)\hat{\rho}_1] > \rho_c$.*

This time $a \to \rho_1(a, \rho)$ is decreasing, so the range of possible values of $\rho_1$ for a fixed value of $\rho$ is

$$[\hat{\rho}_1(1/2,\rho), \hat{\rho}_1(0,\rho)] = \left[\frac{2\rho}{1+\epsilon}, \frac{\rho}{\epsilon+(1-\epsilon)\rho}\right], \qquad [6]$$

i.e., the old upper bound on the range of $\rho_1$ in **4** has become the lower bound. See Fig. 3 for a picture.

## Stability Calculations

The results in this section are proved in *SI Text*. Using large deviations for the trinomial distribution, which in this case is just calculating probabilities using Stirling's formula, we conclude the following:

**Theorem 3A.** *Suppose $\rho < \rho_c$ and recall $\mu_a$ has no mass on $Q_{1,1}$. The flow into $Q_{0,0}$ from $Q_{0,1}$ and $Q_{1,0}$ is larger than the flow out if and only if*

$$\left(\frac{1-\epsilon\rho_1}{1-\rho_1}\right)^{1-\rho_c} < 1 + (1-\epsilon)\rho_1. \qquad [7]$$

**Theorem 3B.** *Suppose $\rho \geq \rho_c$ and recall $\hat{\mu}_a$ has no mass on $Q_{0,0}$. The flow out of $Q_{1,1}$ to $Q_{0,1}$ and $Q_{1,0}$ is larger than the flow in if and only if*

$$\left(\frac{\hat{\rho}_1}{1-\hat{\rho}_1}\right)^{1-\rho_c} < \left[1-(1-\epsilon)\hat{\rho}_1\right]^{-1}. \qquad [8]$$

## Phase Transition

Combining Theorems 2A and 3A, we can determine the behavior of the process for $\rho < \rho_c$. The set of possible values for $\rho_1(a, \rho)$ for a fixed $\rho$ is the interval $[\rho_1(0, \rho), \rho_1(1/2, \rho)]$ given in **4**. Because $0 \leq \rho \leq \rho_c$, we are looking for a solution to

$$\left(\frac{1-\epsilon x_0}{1-x_0}\right)^{1-\rho_c} = 1 + (1-\epsilon)x_0.$$

with $x_0 \in [0, 2\rho_c/(1+\epsilon))$. In *SI Text*, we show that $x_0$ exists and is unique. Here, we will concentrate on what happens in the example $\rho_c = 0.2$ and $\epsilon = 0.1$. When $\rho_c = 0.2$, the interval is [0, 0.4/1.1], and we have $x_0 = 0.2183$.

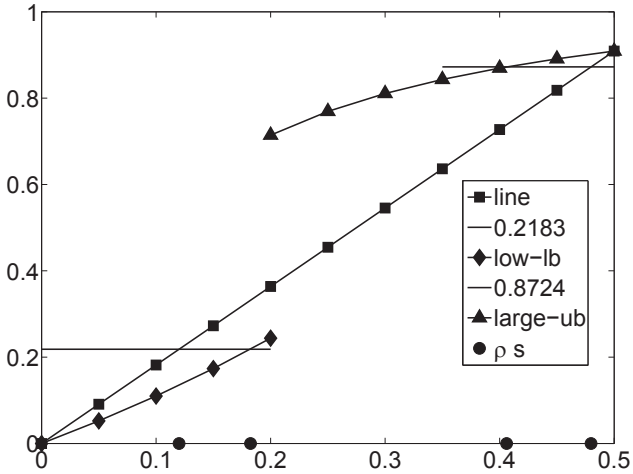


**Fig. 3.** Picture to explain calculation of the phase transition when $\rho_c = 0.2$ and $\epsilon = 0.1$. The $x$ axis gives the value of $\rho$. Dots on the axis are the locations of $\rho_b$, $\rho_d$, $\hat{\rho}_b$, and $\hat{\rho}_d$. The two curves are $\rho_1(0, \rho)$ for $\rho < \rho_c$ and $\hat{\rho}_1(0,\rho)$ for $\rho \geq \rho_c$, whereas the straight line is $\rho_1(1/2,\rho) = \hat{\rho}_1(1/2,\rho) = 2\rho/(1+\epsilon)$.

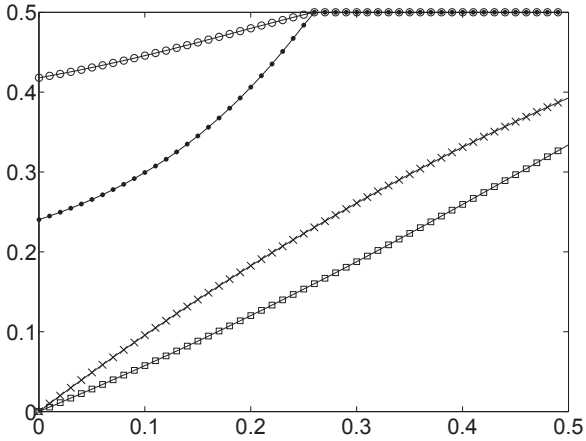


**Fig. 4.** $\rho_b < \rho_d < \hat{\rho}_b < \hat{\rho}_d$ as a function of $\rho_c$ when $\epsilon = 0.1$.

Let $\rho_b$ be chosen so that $x_0 = \rho_1(1/2, \rho_b)$ and $\rho_d$ be chosen so that $x_0 = \rho_1(0, \rho_d)$. See Fig. 3 for a picture. When a solution $x_0$ exists in the desired interval

$$\rho_b = \frac{(1+\epsilon)x_0}{2} \quad \text{and} \quad \rho_d = \frac{x_0}{1+x_0(1-\epsilon)}.$$

In our special case, $\rho_b = 0.1201$ and $\rho_d = 0.1825$.

**Theorem 4A.** *The stable stationary distributions for $\rho < \rho_c$ are*

$$\begin{array}{ccc} \mu_r & \text{for} & 0 \le \rho < \rho_b, \\ \mu_r \text{ and } \mu_s & \text{for} & \rho_b < \rho < \rho_d, \\ \mu_s & \text{for} & \rho_d < \rho < \rho_c. \end{array}$$

Why is this true? When $\rho < \rho_b$, $\rho_0 > \rho_1(1/2, \rho)$, the flow into $Q_{0,0}$ is always larger than the flow out, so $\mu_r$ is the stationary distribution. When $\rho_b < \rho < \rho_d$, there will be an $a_c \in (0, 1/2)$ so that $\rho_1(a_c, \rho) = \rho_0$. The flow into $Q_{0,0}$ is larger than the flow out when $a < a_c$ and the $a$ in the mixture will decrease, whereas for $a > a_c$, the flow out of $Q_{0,0}$ will be larger than the flow in and $a$ will increase. Thus, we have stable fixed points at 0 and 1/2. When $\rho_d < \rho < \rho_c$, $\rho < \rho_1(0, \rho)$, the flow out of $Q_{0,0}$ is always larger than the flow in, and the segregated fixed point with $a = 1/2$ is the stationary distribution.

Using Theorems 2B and 3B, we can determine the behavior of the process for $\rho \ge \rho_c$. The set of possible values for $\hat{\rho}_1(a, \rho)$ for a fixed $\rho$ is the interval $[\hat{\rho}_1(1/2, \rho), \hat{\rho}_1(0, \rho)]$ given in **6**. Because $\rho_c \le \rho \le 0.5$, we are looking for a solution to

$$\left(\frac{\hat{x}_0}{1-\hat{x}_0}\right)^{1-\rho_c} = \left[1-(1-\epsilon)\hat{x}_0\right]^{-1}.$$

with $\hat{x}_0 \in [2\rho_c/(1+\epsilon), 1/1+\epsilon]$. In *SI Text*, we show that there is a solution in the desired interval if and only if $\epsilon^{\rho_c} \ge (\epsilon+1)/2$, and it is unique. The condition comes from having $=$ in **8** at the right end point $1/(1+\epsilon)$. When $\epsilon = 0.1$, the condition is $\rho_c < 0.25964$.

When $\rho_c = 0.2$ and $\epsilon = 0.1$, this interval is [0.4/1.1, 1/1.1], and $\hat{x}_0 = 0.8724$. Let $\hat{\rho}_b$ be chosen so that $\hat{x}_0 = \hat{\rho}_1(0, \hat{\rho}_b)$ and $\hat{\rho}_d$ be chosen so that $\hat{x}_0 = \hat{\rho}_1(1/2, \rho_d)$. When a solution $\hat{x}_0$ exists in the desired interval, we have

$$\hat{\rho}_b = \frac{\epsilon\hat{x}_0}{1-\hat{x}_0(1-\epsilon)} \quad \text{and} \quad \hat{\rho}_d = \frac{(1+\epsilon)\hat{x}_0}{2}.$$

In our example, $\hat{\rho}_b = 0.4061$ and $\hat{\rho}_d = 0.4798$.

**Theorem 4B.** *The stable stationary distributions for $\rho \ge \rho_c$ are*

$$\begin{array}{ccc} \hat{\mu}_s & \text{for} & \rho_c \le \rho < \hat{\rho}_b, \\ \hat{\mu}_r \text{ and } \hat{\mu}_s & \text{for} & \hat{\rho}_b < \rho < \hat{\rho}_d, \\ \hat{\mu}_r & \text{for} & \hat{\rho}_d < \rho < 0.5. \end{array}$$

The reasoning behind this result is the same as for Theorem 4A. To show that these result can be used to explicitly describe the phase transition, Fig. 4 shows how the four critical values depend on $\rho_c$ when $\epsilon = 0.1$.

## Discussion

Here, we considered a metapopulation version of Schelling's model, which we believe is a better model for studying the dynamics of segregation in a city than a nearest neighborhood interaction on the 2D square lattice. Due to the simple structure of the model, we are able to describe the phase transition in great detail. For $\rho < \rho_b$, a random distribution of families $\mu_r = \text{Tri}(\rho, \rho)$ is the unique stationary distribution. As $\rho$ increases there is a discontinuous phase transition to a segregated state, $\mu_s$ at $\rho_d$ preceded by an interval $(\rho_b, \rho_d)$, in which both $\mu_r$ and $\mu_s$ are stable. Surprisingly the phase transition occurs to a segregated state occurs at a value $\rho_d < \rho_c$, i.e., at a point where in a random distribution most families are happy. This shift in behavior occurs because random fluctuations create segregated neighborhoods, which, as our analysis shows, are more stable than the random ones.

If $\rho_c$ is small enough, then as $\rho$ nears 1/2, there is another discontinuous transition at $\hat{\rho}_d$, which returns the equilibrium to the random state $\hat{\mu}_r = \text{Tri}(\rho, \rho)$, and this is preceded by an interval $(\hat{\rho}_b, \hat{\rho}_d)$ of bistability. To explain this, we note that when families are distributed randomly, everyone is unhappy and moves at rate 1, maintaining the random distribution. In our concrete example, $\rho_c = 0.2$ and $\epsilon = 0.1$, the fraction of vacant houses at $\hat{\rho}_d = 0.4798$ is only 4.04%, so it is very difficult to make segregated neighborhoods where one type is happy. Our stability analysis implies that these segregated neighborhoods are created at a slower rate than they are lost, so the random state prevails.

The results in this paper have been derived under two assumptions: (*i*) stationary distributions are invariant under interchange of red and blue and (*ii*) the process is always in one of a one-parameter family of self-consistent stationary distributions indexed by $a \in [0,1/2]$, but the value of $a$ changes over time. We are confident that *ii* can be rigorously proved. Removing *i* will be more difficult, because when symmetry is dropped, there is a two-parameter family of self-consistent distributions. A more interesting problem, which is important for applications to real cities, is to allow the initial densities of reds and blues and their threshold for happiness to differ. Although our solution is not yet complete, we believe it is an important first step in obtaining a detailed understanding of the equilibrium behavior of Schelling's model in a situation that is relevant for applications.

**ACKNOWLEDGMENTS.** We thank David Aldous, Nicholas Lanchier, Simon Levin, and Thomas Liggett for helpful comments. R.D. and Y.Z. were partially supported by Grants DMS 10-05470 and DMS13-05997 from the probability program at the National Science Foundation.

1. Schelling TC (1971) Dynamic models of segregation. *J Math Sociol* 1(2):143–186.
2. Schelling TC (1978) *Micromotives and Macrobehavior* (Norton, New York).
3. Clark WAV, Fossett M (2008) Understanding the social context of the Schelling segregation model. *Proc Natl Acad Sci USA* 105(11):4109–4114.
4. Fossett M (2006) Ethnic preferences, social science dynamics, and residential segregation: Theoretical explanations using simulation analysis. *J Math Sociol* 30(3-4):185–274.
5. Pancs R, Vriend NJ (2007) Schelling's spatial proximity model of segregation revisited. *J Public Econ* 91(1-2):1–24.
6. Kandler A, Perreault C, Steele J (2012) Cultural evolution in spatially structured populations: A review of alternative modeling frameworks. *Adv Complex Syst* 15:1203001.
7. Hatna E, Benenson I (2009) The Schelling model of ethnic residential dyanmics: Beyon the integrated-segregation dichotomy of patterns. *J Artif Soc Soc Simul* 15(1):6.

8. Vinkovic D, Kirman A (2006) A physical analogue of the Schelling model. *Proc Natl Acad Sci USA* 103(51):19261–19265.
9. Stauffer D, Solomon S (2007) Ising, Schelling and self-organizing segregation. *Eur Phys J B* 57(4):473–479.
10. Singh A, Vainchtein D, Weiss H (2007) Schelling's segregation model: Parameters, scaling, and aggregation. arXiv:0711.2212.
11. Dall'Asta L, Castellano C, Marsili M (2008) Statistical physics of the Schelling model of segregation. *J Stat Mechanics: Theory and Experiment* 7:L07002.
12. Gauvin L, Vannimenus J, Nadal J-P (2009) Phase diagram of a Schelling segregation model. *Eur Phys J B* 70(2):293–304.
13. Rogers T, McKane AJ (2011) A unified framework for Schelling's model of segregation. *J Stat Mechanics: Theory and Experiment* 7:P07006.
14. Domic NG, Goles E, Rica S (2011) Dynamics and complexity of the Schelling segregation model. *Phys Rev E* 83(5):056111.
15. Brandt C, Immorlica N, Kamath G, Kleinberg R (2012) An analysis of one-dimensional Schelling segregation. arXiv:1203.6346.
16. Grauwin S, Bertin E, Lemoy R, Jensen P (2009) Competition between collective and individual dynamics. *Proc Natl Acad Sci USA* 106(49):20622–20626.
17. Remenik D (2009) Limit theorems for individual-based models in economics and finance. *Stoch Proc Appl* 119(8):2401–2435.