Schelling Redux: An Evolutionary Dynamic Model of Residential Segregation*†

Emin Dokumacı and William H. Sandholm‡

June 19, 2007

Abstract

Schelling (1971) introduces a seminal model of the dynamics of residential segregation in an isolated neighborhood. His model combines agent heterogeneity with explicit behavior dynamics; as such it is presented informally, and with the use of "semi-equilibrium" restrictions on out-of-equilibrium play. In this paper, we use recent techniques from evolutionary game theory to introduce a formal version of Schelling's model, one that dispenses with equilibrium restrictions on the adjustment process. We show that key properties of the resulting infinite-dimensional dynamic can be derived using a simple finite-dimensional dynamic that captures aggregate behavior. We determine conditions for the stability of integrated equilibria, and we derive a strong restriction on out-of-equilibrium dynamics that implies global convergence to equilibrium: along any solution trajectory, one population's aggregate behavior adjusts monotonically, while the other's changes direction at most once. We present a variety of examples, and we show how extensions of the basic model can be used to study both alternative specifications of agents' preferences and policies to promote integration.

1. Introduction

The high degree of residential segregation in U.S. cities is well documented. While many factors contribute to this state of affairs, it is widely acknowledged that individuals'

*This paper contains color figures. A copy of the paper with high quality figures can be downloaded from http://www.ssc.wisc.edu/~edokumac/research/schelling.pdf.

†We thank Marzena Rostek and Larry Samuelson for helpful comments. Financial support from NSF Grants SES-0092145 and SES-0617753 is gratefully acknowledged.

‡Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA. e-mail: edokumaci@wisc.edu, whs@ssc.wisc.edu; websites: http://www.ssc.wisc.edu/~edokumac, http://www.ssc.wisc.edu/~whs.

preferences about the racial composition of their neighborhoods play an important role.[1] Still, the link between individual preferences and observed segregation is not as direct as one might expect. Indeed, surveys show that most individuals prefer neighborhoods that are more diverse than the neighborhoods we see.[2] This discrepancy suggests a variety of questions of clear practical import: Why do observed neighborhood racial compositions differ from what individuals would prefer? How do these compositions change over time? Under what circumstances are integrated neighborhoods most likely to survive? When is public policy likely to be helpful in sustaining integration?

The seminal work on the link between individual preferences and residential segregation is Thomas Schelling's 1971 article entitled "Dynamic Models of Segregation". This article considers residential location choices made by individuals from two groups, focusing on how these decisions can result in higher levels of segregation than nearly any individual agent would prefer.[3]

Schelling (1971) presents two distinct models of residential location dynamics. Both models feature agents whose preferences are described by scalar "tolerances"; these indicate the proportion of unlike residents in a neighborhood that an agent will accept before prefering to move to a new location. The two models differ in how physical space, and hence agents' choice sets, are described. In Schelling's "spatial proximity model", locations are arrayed in a discrete grid, and an agent's neighborhood is defined to be the set of locations that are adjacent to his own; when an opportunity to move arises, a dissatisfied agent moves to an empty square whose neighborhood he prefers.[4]

In contrast, Schelling's "isolated neighborhood" model describes decisions made by heterogeneous agents about whether to live in a certain mixed neighborhood or in some homogeneous outside locales.[5] Schelling focuses on how the distributions of tolerances

[1]Unless proper controls are put in place when preferences are elicited, racial composition can end up serving as a proxy for preferences about other neighborhood characteristics, in particular income levels. Nevertheless, the sociological consensus holds that preferences about neighborhood racial composition remain an important contributor to observed segregation even when other factors are held fixed. The basic reference on U.S. residential segregation is Massey and Denton (1993). For a more recent survey of work on this question, see Charles (2003). Additional references are discussed below.

[2]See the previous references, as well as Clark (1991), Bobo and Zubrinsky (1996), and Farley et al. (1997).

[3]Schelling further develops this insight in a broad array of contexts in his 1978 book, *Micromotives and Macrobehavior*, which contains an abridged version of his 1971 paper.

[4]Actually, Schelling presents two spatial proximity models that differ in important respects. The better known of these, the so-called "checkerboard model", describes locations as a discrete two-dimensinal grid, and allows agents who are not content to move to any unoccupied square in the grid. Before presenting this model, Schelling considers a one-dimensional model in which agents are able to insert themselves between pairs of opponents. See Pancs and Vriend (2007) for further discussion of these models. Variants of Schelling's spatial proximity models have been analyzed using techniques from stochastic evolutionary game theory: see Young (1998, 2001), Zhang (2004a,b), Möbius (2000), and Bøg (2006).

[5]Schelling refers to this model as the "bounded neighborhood" model.

in the two populations influence the set of equilibrium outcomes and the nature of disequilibrium dynamics. He then uses this model to consider the effects of various policy interventions, and, most famously, to address the phenomenon of "neighborhood tipping". While this model has been quite influential, it is not completely satisfying in all respects: it requires strong out-of-equilibrium sorting assumptions, and its disequilibrium dynamics are not specified explicitly. Still, Schelling's approach provides a workable model of disequilibrium behavior adjustment in an environment with heterogeneous agents—an environment that without strong simplifying assumptions appears impervious to a tractable dynamic analysis.

In this paper, we show that despite the complications generated by agent heterogeneity, Schelling's isolated neighborhood model can be placed on a firm theoretical footing. Using new tools from evolutionary game theory—namely, the Bayesian best response dynamic of Ely and Sandholm (2005)—we construct an explicit model of location choice dynamics. This model allows the behaviors of agents with different preferences to adjust separately, and so avoids "semi-equilibrium" restrictions on disequilibrium neighborhood compositions. While the model takes the form of an infinite-dimensional dynamical system, we adapt Ely and Sandholm's (2005) results to show that for many purposes, the analysis of this system can be reduced to that of an appropriate two-dimensional dynamical system, one that keeps track of aggregate behavior within each population. Thus, our approach to the dynamics of residential segregation allows us to model individual behavior in a satisfying way while still retaining Schelling's original insights.

Indeed, by specifying an explicit model of behavior dynamics, we are able to obtain a number of new qualitative results. For instance, we derive necessary and sufficient conditions for the stability of integrated equilibria. Surprisingly, the key requirement for stability is that in at least one of the populations, the number of agents who find themselves indifferent between staying in the neighborhood and moving out must not be too large.

More novelly, we are able to obtain restrictions on the nature of the adjustment process itself. Using techniques from the theory of competitive differential equations (see, e.g., Smith (1995)), we are able to prove that any solution trajectory of our model must obey strong monotonicity requirements: the state variable describing aggregate behavior in one of the two groups must change monotonically over time, while the state variable describing behavior in the other group switches directions at most once. It follows immediately that every solution trajectory of the dynamic converges to equilibrium. Restrictions on the nature of disequilibrium adjustment are not common in economics, but in studying issues like residential segregation, where dynamics are understood to be of central import,

implications of this sort seem likely to play a crucial role.

While we only explore one model in detail, our modeling technique is quite flexible, and can accommodate a wide array of specifications of policy instruments, individual preferences, and location choice alternatives. After presenting the analytical results for our basic model, we offer a variety of examples that reveal the complex interplay between population sizes and preference distributions that determine the nature and the stability of equilibrium behavior. We show that a distaste for overly homogeneous neighborhoods can lead to the formation of sparsely populated segregated districts, but can also allow an integrated equilibrium to be a global attractor. We address policies to sustain integration, and argue that the success of these policies can depend on the fine details of the instruments employed. Lastly, we describe in brief some further extensions of our model, including the replacement of the homogeneous outside options by additional neighborhoods subject to settlement by both groups, and the introduction of heterogeneity in income levels and in preferences for public goods.

The remainder of the paper is organized as follows. In Section 2, we describe Schelling's original isolated neighborhood model. In Section 3, we show how this model can be formalized using Bayesian population games and Bayesian best response dynamics, and introduce the aggregate dynamic that makes our qualitative analysis possible. Section 4 derives necessary and sufficient conditions for stability of equilibrium, establishes monotonicity properties of disequilibrium behavior trajectories, and demonstrates the implications of these results through some examples. Section 5 describes extensions of the model, and Section 6 offers concluding remarks. All proofs are relegated to an appendix.

2. Schelling's Isolated Neighborhood Model

2.1 Colors and Tolerances

Here is the basic idea behind Schelling's (1971) model, in his own words:

> In this model there is one particular bounded area that everybody, black or white, prefers to its alternatives. He will reside in it unless the percentage of residents of the opposite color exceeds some limit. Each person, black or white, has his own limit. ('Tolerance', we shall occasionally call it.) If a person's limit is exceeded in this area he will go someplace else—a place, presumably, where his own color predominates or where color does not matter. (p. 167)

Thus, in Schelling's model, there are two populations of agents, one of whites and one of blacks. Agents in each group choose between residing in a (possibly) mixed neighborhood and residing at an alternate location inhabited solely by members of their

own group. Each agent's preferences are characterized by a *tolerance*. If the ratio of other-group residents to own-group residents in the neighborhood is below the agent's tolerance, he prefers to live in the neighborhood; if this ratio is below the agent's tolerance, he prefers to live at the alternate location.

Schelling presents his model without using any notation: he employs only verbal descriptions and diagrammatic analysis. Nevertheless, introducing some notation here will enable us to explain his work in an efficient fashion. We let m^w and m^b denote the sizes of the white and black populations. The distributions of tolerances in the white and black populations, which we denote by μ^w and μ^b, are measures on $[0, \infty)$ with total masses m^w and m^b, respectively. Both of these measures are assumed to be absolutely continuous (i.e., to admit density functions).

2.2 Equilibrium

The natural definition of equilibrium in this setting requires that no agent be able to benefit from switching locations.[6] To describe equilibrium formally, let $x^w \in X^w \equiv [0, m^w]$ and $x^b \in X^b \equiv [0, m^b]$ denote the numbers of whites and blacks in the neighborhood. Elements of $X \equiv X^w \times X^b$ are called *social states*.

Let the functions $t^w : X^w \to [0, \infty)$ and $t^b : X^b \to [0, \infty)$ be defined implicitly as follows:

$$\mu^w([t^w(x^w), \infty)) = x^w;$$
$$\mu^b([t^b(x^b), \infty)) = x^b.$$

Thus, $t^w(x^w)$ is the (x^w)th highest tolerance in the white population (where $t^w(0)$ is the very highest tolerance and $t^w(m^w)$ is the very lowest), and $t^b(x^b)$ is the (x^b)th highest tolerance in the black population. To ensure that t^w and t^b are well-defined, we assume in the remainder of this section that the supports of μ^w and μ^b are bounded intervals. In this case, t^w and t^b are the inverses of the decumulative distribution functions of μ^w and μ^b.[7]

A white agent will be indifferent between living inside or outside the neighborhood when the ratio of blacks to whites in the neighborhood is equal to his tolerance. In particular, if x^b blacks and the x^w most tolerant whites live in the neighborhood, then the

[6] When Schelling (1971) introduces his notion of "static viability", he does not consider whether agents outside the neighborhood would prefer to move in; however, he introduces this possibility when considering location choice dynamics (p. 170).

[7] Without this assumption, the dynamics described in equations (2a) and (2b) are not well defined. Schelling (1971) uses distributions with both nonconvex supports and mass points in some of his examples (see, e.g., his p. 176-178), but nothing about these examples changes if the distributions are modified so as to have convex supports. In our model, we will allow nonconvex supports, but mass points at finite tolerances will be forbidden.

marginal white agent is indifferent if the ratio x^b/x^w equals his tolerance $t^w(x^w)$. With this in mind, we define

$$T^w(x^w) = x^w\, t^w(x^w)$$

to be the number of blacks that must be in the neighborhood for this marginal white agent to be indifferent. Similarly, if the x^b most tolerant blacks are in the neighborhood, then

$$T^b(x^b) = x^b\, t^b(x^b)$$

is the number of whites who must be in the neighborhood for the marginal black agent to be indifferent.

Social state $x = (x^w, x^b)$ is an *equilibrium* if the following conditions hold:

$$\text{(1a)} \qquad \begin{cases} x^w = 0 & \Rightarrow \quad T^w(0) \leq x^b; \\ x^w \in (0, m^w) & \Rightarrow \quad T^w(x^w) = x^b; \\ x^w = m^w & \Rightarrow \quad T^w(m^w) \geq x^b. \end{cases}$$

$$\text{(1b)} \qquad \begin{cases} x^b = 0 & \Rightarrow \quad T^b(0) \leq x^w; \\ x^b \in (0, m^b) & \Rightarrow \quad T^b(x^b) = x^w; \\ x^b = m^b & \Rightarrow \quad T^b(m^b) \geq x^w. \end{cases}$$

In equilibrium, it must be that the x^w most tolerant white agents and the x^b most tolerant black agents live in the neighborhood. Conditions (1a) and (1b) tell us that at an interior equilibrium, the marginal agent in each population is indifferent between living inside or outside the neighborhood. If all agents of a certain color are in the neighborhood, the least tolerant agent of that color must weakly prefer to reside there; if no agents of this color are in the neighborhood, the most tolerant agent must weakly prefer not to reside there.

Let us illustrate these definitions by presenting Schelling's (1971) first example (p. 168-171). In this example, Schelling assumes that the white population is twice as large as the black population ($m^w = 100, m^b = 50$), and that the distribution of tolerances in each population is uniformly distributed on $[0, 2]$. It is easy to verify that this specification of μ^w and μ^b generates the functions $T^w(x^w) = 2x^w - \frac{1}{50}(x^w)^2$ and $T^b(x^b) = 2x^b - \frac{1}{25}(x^b)^2$. Schelling's diagram of these two functions is presented in Figure 1.[8] Examining this figure, we see that there are segregated equilibria at (100, 0) and (0, 50), and an integrated equilibrium at $x_\star = (x^w_\star, x^b_\star) \approx (21.7401, 34.0276)$.

[8]This graph, Figure 9 of Schelling (1978), is a cleaner version of Figure 18 of Schelling (1971).

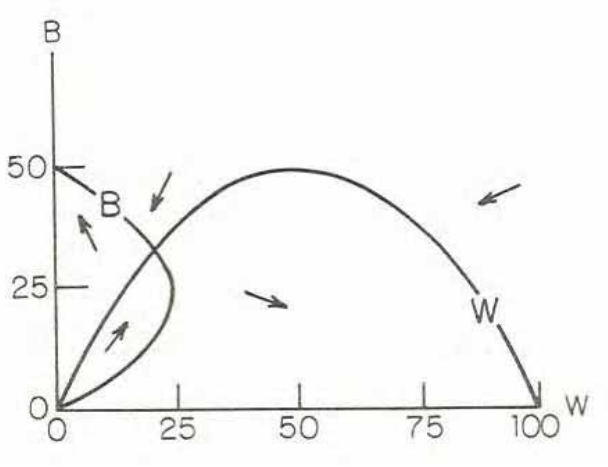


Figure 1: Schelling's first example.

It is worth noting one technical difficulty here. If one population, say the whites, is absent from the neighborhood ($x^w = 0$), then the ratio x^b/x^w used to define white agents' preferences is undefined. There is no difficulty in setting this ratio to ∞ when x^b is positive, and our definitions behave appropriately in this case. But if both x^w and x^b are 0, it is unclear how preferences ought to be defined, and it is clearly problematic to specify preferences in a neighborhood of this point in a continuous way. [9]

2.3 Schelling's Dynamics

Let us quote Schelling's (1971) description of his dynamics:

> It is the dynamics of motion, though, that determine what color mix will ultimately occupy the area. The simplest dynamics are as follows: if all whites present in the area are content, and some outside would be content if they were inside, the former will stay and the latter will enter; and whites will continue to enter so long as all those present are content and some would be content if present. If not all whites present are content, some will leave; they will leave in the order of their discontent, so that those remaining are the most tolerant; and when their number in relation to the number of blacks is such that the whites remaining are all content, no more of them leave. A similar rule governs entry and departure of the blacks. (p. 170)

In our notation, this description corresponds to the following family of dynamics on $X = X^w \times X^b$:

(2a) $\qquad \operatorname{sgn}(\dot{x}^w) = \operatorname{sgn}(T^w(x^w) - x^b);$

[9] For the record, we note that since $T^w(0) = T^b(0) = 0$, state $(0, 0)$ is always an equilibrium according to conditions (1a) and (1b).

(2b) $\qquad \operatorname{sgn}(\dot{x}^b) = \operatorname{sgn}(T^b(x^b) - x^w).$

The effect of these dynamics on the state (x^w, x^b) is represented by the arrows in Figure 1. When the state is below the graph of T^w (labeled W in the figure), so that $T^w(x^w) > x^b$, $\dot{x}^w$ is positive, so the number of whites in the neighborhood increases. When the state is above the graph of T^w, the number of whites in the neighborhood falls, and when the state is on the graph of T^w, the number of whites is momentarily fixed. Similarly, the number of blacks in the neighborhood rises, falls, or stays fixed according to whether the state is to the left, to the right, or directly on the graph of T^b.

While one would need to specify the dynamics more precisely than in (2) to obtain exact solutions, the qualitative features of such solutions is apparent from Figure 1: the integrated equilibrium $x_\star$ is a saddle, and hence unstable, and so almost all solution trajectories head toward one of the stable, segregated equilibria at states (100, 0) and (0, 50).

2.4 The Semi-Equilibrium Assumption

In order to formulate his dynamics on the set of social states X, Schelling must impose a major simplifying assumption: at any point in time, the agents who are in the neighborhood are always the ones whose tolerances are highest. While this property must hold in equilibrium, it is a very strong assumption to make about behavior during disequilibrium adjustment. We refer to Schelling's assumption here as the *semi-equilibrium assumption*, and describe system (2) as *semi-equilibrium dynamics*.

It is important to note that without the semi-equilibrium assumption, it is no longer clear that system (2) provides a reasonable description of the evolution of behavior. To see why not, suppose that at some moment in time, the x^w whites in the neighborhood are not the x^w most tolerant whites in the population. In this event, the least tolerant white agent in the neighborhood has a tolerance less than $T^w(x^w)$. Even if $T^w(x^w) > x^b$, if it is the case the least tolerant whites in the neighborhood adjust their behavior before the discontented whites outside the neighborhood, the number of white agents in the neighborhood could fall. This disagrees with equation (2a).

In order to construct dynamics that do not rely on an implicit equilibrium assumption, one must track the behaviors of agents with different tolerances separately. We accomplish this by employing tools introduced in an abstract context by Ely and Sandholm (2005): we express Schelling's model as a Bayesian population game, and describe the evolution of behavior using Bayesian best response dynamics.

3. Bayesian Population Games and Best Response Dynamics

3.1 Schelling's Model as a Bayesian Population Game

3.1.1 Types and Type Distributions

In our Bayesian population game, each agent is a member of one of two populations, w or b. Each agent in population $p \in \{w, b\}$ is endowed with a *type* (i.e., a *tolerance*) θ^p from the set $\Theta^p \equiv [0, \infty]$. In what follows, we denote the type of a white agent by θ^w and the type of a black agent by θ^b; as before, the distributions of blacks' and whites' types are denoted μ^w and μ^b. Notice that we have introduced agents of type ∞; these *committed types* prefer to live in the neighborhood under all circumstances.

We make the following assumptions about distributions μ^w and μ^b:

(A1) $m^w_\infty \equiv \mu^w(\{\infty\}) > 0$ and $m^b_\infty \equiv \mu^b(\{\infty\}) > 0$.

(A2) $\mu^w|_{[0,\infty)}$ and $\mu^b|_{[0,\infty)}$ admit bounded density functions.

We sometimes find it convenient to write $m^p = m^p_f + m^p_\infty$, where $m^p_f = \mu^p([0, \infty))$ denotes the mass of finite type agents in population $p \in \{w, b\}$. The roles of assumptions (A1) and (A2) will be made clear below.

3.1.2 Payoffs

The payoff functions for the Bayesian game are of the form $U^p_s : X \times \Theta^p \to (-\infty, \infty]$, where $p \in \{w, b\}$ is a population and $s \in \{in, out\}$ is a strategy. Recall that in Schelling's (1971) model, an agent prefers to live in the neighborhood if and only if the ratio of other-group residents to own-group residents in the neighborhood is below the agent's tolerance. Therefore, if we normalize the payoff to the outside option *out* to 0 (i.e., if $U^w_{out} \equiv U^b_{out} \equiv 0$), then the payoff to choosing *in* must satisfy

(3a) $\operatorname{sgn}\left(U^w_{in}(x; \theta^w)\right) = \operatorname{sgn}\left(\theta^w - \frac{x^b}{x^w}\right)$ when $\theta^w \neq \infty$, and

(3b) $\operatorname{sgn}\left(U^b_{in}(x; \theta^b)\right) = \operatorname{sgn}\left(\theta^b - \frac{x^w}{x^b}\right)$ when $\theta^b \neq \infty$.

Since the committed types always prefer to live in the neighborhood, we set

$$U^w_{in}(\cdot\,; \infty) \equiv U^b_{in}(\cdot\,; \infty) \equiv \infty.$$

For conditions (3a) and (3b) to be well-defined, x^w and x^b must be positive. We could skirt this issue temporarily by multiplying through by x^w and x^b, but as we saw at the end

of Section 2.2, doing so still leaves us with a severe discontinuity at the origin. This issue is addressed in the next two sections.

3.1.3 Bayesian Strategies

For analytical convenience, we assume that in each population p, there is a continuum of agents of each type $\theta^p \in \text{support}(\mu^p)$. The proportion of agents in subpopulation θ^p who choose *in* is denoted $\sigma^p(\theta^p) \in [0,1]$. Thus, the behavior of all types of agents in population p is described by the *Bayesian strategy* $\sigma^p : \Theta^p \to [0,1]$; the set of all such Bayesian strategies is denoted Σ^p. We evaluate distances between points in Σ^p using the L^1 norm,

$$\|\sigma^p - \hat{\sigma}^p\| = \int_{\Theta^p} \left|\sigma^p(\theta^p) - \hat{\sigma}^p(\theta^p)\right| \mathrm{d}\mu^p, \tag{4}$$

so that the distance between Bayesian strategies σ^p and $\hat{\sigma}^p$ is just the (μ^p-)average distance between subpopulations' behaviors under σ^p and $\hat{\sigma}^p$. We call σ^p and $\hat{\sigma}^p$ equivalent if $\sigma^p(\theta^p) = \hat{\sigma}^p(\theta^p)$ for (μ^p-)almost every $\theta^p \in \Theta^p$. In other words, we do not distinguish between Bayesian strategies that indicate the same action distribution in almost every subpopulation.

A complete description of behavior in both populations is provided by $\sigma = (\sigma^w, \sigma^b) \in \Sigma \equiv \Sigma^w \times \Sigma^b$. We refer to $\sigma \in \Sigma$ as a *Bayesian strategy profile*, or as a *profile* for short.

As we have seen, agents care about the proportions of blacks and whites in the neighborhood, but they do not care directly about opponents' types. For this reason, it is useful to introduce notation for the aggregate behavior generated by a Bayesian strategy profile. The masses of white and black agents who choose *in* under profile $\sigma = (\sigma^w, \sigma^b)$ is obtained by applying the *aggregation operator* A:

$$A\sigma = (A\sigma^w, A\sigma^b) = \left(\int_{\Theta^w} \sigma^w(\theta^w)\,\mathrm{d}\mu^w, \int_{\Theta^b} \sigma^b(\theta^b)\,\mathrm{d}\mu^b\right).$$

Like the expectation operator E, the aggregation operator A always integrates with respect to the measure or measures appropriate for its argument.

In Section 3.1.2, we saw that that preferences are not well-defined at social states x at which x^w or x^b equals 0. To begin to address this issue, recall that committed agents in both populations find *in* strictly dominant. Bearing this in mind, we let

$$\Sigma^p_\circ = \{\sigma^p \in \Sigma^p : \sigma^p(\infty) = 1\}$$

denote the subset of Σ^p on which committed agents in population p choose *in*. By defini-

tion, the total number of *in* players at any $\sigma^p \in \Sigma_\circ^p$ is at least m_∞^p; we thus have

$$A(\Sigma_\circ^p) = X_\circ^p \equiv [m_\infty^p, m^p].$$

For future convenience, we let $\Sigma_\circ = \Sigma_\circ^w \times \Sigma_\circ^b$ and $X_\circ = X_\circ^w \times X_\circ^b$, so that $A(\Sigma_\circ) = X_\circ$.

3.2 Bayesian Best Responses, Equilibria, and Dynamics

3.2.1 Bayesian Best Responses

The *Bayesian best response correspondence* for population $p \in \{w, b\}$, denoted $B^p : X \rightrightarrows \Sigma^p$, is defined as follows:

$$B^p(x)(\theta^p) = \begin{cases} 1 & \text{if } U_{in}^p(x;\theta^p) > U_{out}^p(x;\theta^p), \\ [0,1] & \text{if } U_{in}^p(x;\theta^p) = U_{out}^p(x;\theta^p), \\ 0 & \text{if } U_{in}^p(x;\theta^p) < U_{out}^p(x;\theta^p). \end{cases}$$

Notice that $B^p(x) \in \Sigma^p$ is a Bayesian strategy; for each $\theta^p \in \Theta^p$, $B^p(x)(\theta^p) \in [0,1]$ is the set of mixed best responses to state x for agents of type θ^p.

All of the analysis to come hinges on the following fact about the restrictions of B^w and B^b to the set $X_\circ$, which we prove in the Appendix.

Lemma 3.1. *Under assumptions* (A1) *and* (A2)*, the maps $B^w : X_\circ \to \Sigma^w$ and $B^b : X_\circ \to \Sigma^b$ are single-valued and Lipschitz continuous.*

According to Lemma 3.1, restricting attention to Bayesian strategy profiles in $\Sigma_\circ$ dispels the discontinuity problem noted earlier: if x is restricted to lie in $X_\circ$, the maps $(x^w, x^b) \mapsto x^b/x^w$ and $(x^w, x^b) \mapsto x^w/x^b$ are continuous with bounded derivatives on this set. This observation is crucial to establishing the basic properties of the Bayesian best response dynamic.

3.2.2 Bayesian Equilibria

Bayesian strategy profile $\sigma \in \Sigma$ is a *Bayesian equilibrium* if

$$(\sigma^w, \sigma^b) = \Big(B^w(A\sigma), B^b(A\sigma)\Big).$$

In a Bayesian equilibrium, almost every agent in each population chooses a best response to the current aggregate behavior $A\sigma$. We denote the set of Bayesian equilibria by $\Sigma_\star$.

3.2.3 The Bayesian Best Response Dynamic

Under the Bayesian best response dynamic, behavior in the subpopulation corresponding to each type θ^p adjusts in the direction of that type's current best response. Formally, the *Bayesian best response dynamic* on $\Sigma_\circ$ is defined by the law of motion

$$\text{(B)} \qquad \begin{aligned} \dot{\sigma}^w &= B^w(A\sigma) - \sigma^w, \\ \dot{\sigma}^b &= B^b(A\sigma) - \sigma^b, \end{aligned}$$

where the time derivatives $\dot{\sigma}^w$ and $\dot{\sigma}^b$ are defined in terms of the L^1 norm (4). Evidently, the rest points of (B) are the Bayesian equilibria of the underlying Bayesian population game.

Since committed types always prefer *in*, the best response dynamic leaves the set $\Sigma_\circ$ forward invariant. Together, this observation, Lemma 3.1, and results in Ely and Sandholm (2005) imply that the dynamic (B) is well-behaved on $\Sigma_\circ$, in the sense that solutions from each initial condition in $\Sigma_\circ$ exist and are unique.

Theorem 3.2. *For each Bayesian strategy profile $\sigma \in \Sigma_\circ$, there exists a unique L^1 solution $\{\sigma_t\}_{t\geq 0}$ to the dynamic* (B) *with $\sigma_0 = \sigma$. This solution remains in $\Sigma_\circ$ at all positive times.*

3.3 Aggregation

The Bayesian best response dynamic (B) describes the evolution of behavior without recourse to a semi-equilibrium assumption. However, since the dynamic is defined on the L^1 space $\Sigma_\circ$, it is cumbersome to study directly. In this section, we appeal to results from Ely and Sandholm (2005) to show that most interesting properties of the dynamic (B) are captured by an aggregate dynamic defined directly on $X_\circ \subset \mathbf{R}^2$. This allows us to recover the simplicity of Schelling's (1971) analysis without assuming any undue coordination in agents' disequilibrium behavior.

Bayesian equilibria and the Bayesian best response dynamic are defined in terms of the composite map $B \circ A : \Sigma_\circ \to \Sigma_\circ$. Given a Bayesian strategy $\sigma \in \Sigma_\circ$, this map first aggregates to obtain social state $A\sigma \in X_\circ$, and from this computes the Bayesian best response profile $B(A\sigma) \equiv (B^w(A\sigma), B^b(A\sigma)) \in \Sigma_\circ$.

We want to work with equilibria and dynamics defined on the set $X_\circ$. To do so, we reverse the order of the operators in the composition $B \circ A$, and thus consider the map $A \circ B : X_\circ \to X_\circ$. Given a social state x, this map first computes the Bayesian best response profile $B(x) = (B^w(x), B^b(x)) \in \Sigma_\circ$, and then aggregates to obtain the new social state $A(B(x)) = (A(B^w(x)), A(B^b(x))) \in X_\circ$.

3.3.1 Aggregate Equilibria

We call social state $x_\star$ an *aggregate equilibrium*, denoted $x_\star \in X_\star$, if

$$x_\star = A(B(x_\star)).$$

To see the relevance of this definition, suppose that $\sigma_\star \in \Sigma_\star$ is a Bayesian equilibrium, so that $\sigma_\star = B(A(\sigma_\star))$. Then $A(B(A\sigma_\star))) = A\sigma_\star$, and so $A\sigma_\star$ is an aggregate equilibrium. Conversely, if $x_\star \in X_\star$ is an aggregate equilibrium, so that $x_\star = A(B(x_\star))$, then $B(A(B(x_\star))) = B(x_\star)$, and so $B(x_\star)$ is a Bayesian equilibrium. This demonstrates that there is a one-to-one correspondence between Bayesian equilibria and aggregate equilibria; in fact, one can further establish that the restricted map $A|_{\Sigma_\star} : \Sigma_\star \to X_\star$ is a homeomorphism whose inverse is $B|_{X_\star} : X_\star \to \Sigma_\star$.

3.3.2 The Aggregate Best Response Dynamic

The *aggregate best response dynamic* is described by the law of motion

$$\text{(A)} \qquad \begin{aligned} \dot{x}^w &= A(B^w(x)) - x^w \\ \dot{x}^b &= A(B^b(x)) - x^b \end{aligned}$$

on $X_\circ$. Evidently, the rest points of (A) are the aggregate equilibria of the underlying Bayesian population game.

To understand the importance of this dynamic, suppose that $\{\sigma_t\}_{t\geq 0}$ is a solution to the Bayesian best response dynamic (B), so that $\dot{\sigma}_t = B(A\sigma_t) - \sigma_t$ for all $t \geq 0$. Then

$$\frac{d}{dt} A\sigma_t = A\dot{\sigma}_t = A(B(A\sigma_t)) - A\sigma_t.$$

Thus, if the trajectory of Bayesian strategy profiles $\{\sigma_t\}_{t\geq 0}$ is a solution to (B), then the aggregate behavior trajectory $\{A\sigma_t\}_{t\geq 0}$ is a solution to (A). This argument shows that the aggregation operator A defines a many-to-one map from solutions of the Bayesian dynamic (B) to solutions of the aggregate dynamic (A). In other words, the evolution of aggregate behavior under (B) is completely determined by aggregate behavior at time 0: two Bayesian strategies that induce the same aggregate behavior induce the same aggregate behavior trajectories.

Because the map from solutions of (B) to solutions of (A) is many-to-one, it is not obvious whether stability results for rest points $x_\star$ of (A) imply stability results for the corresponding rest points $B(x_\star)$ of (B). Nevertheless, Ely and Sandholm (2005) prove that if

$x_\star$ is stable under (A), then $B(x_\star)$ is stable under (B), where "stable" can refer to Lyapunov, asymptotic, or global asymptotic stability. They also show that instability of $x_\star$ under the aggregate dynamic (A) implies instability of $B(x_\star)$ under the Bayesian dynamic (B). Thus, while we are directly concerned with the behavior of the infinite-dimensional dynamic (B) on the set of Bayesian strategies $\Sigma_\circ$, most of our questions about this dynamic can be addressed by studying the finite-dimensional dynamic (A) on the set of social states $X_\circ$. Therefore, the remainder of the paper focuses on the latter dynamic.

3.3.3 Alternate Expressions for the Aggregate Dynamic

Before turning to the analysis of the aggregate dynamic (A), it will prove useful to express it directly in terms of the primitives of the model. The compositions $A \circ B^w$ and $A \circ B^b$ can be written explicitly as

$$A(B^w(x)) = \int_{\Theta^w} B^w(x)(\theta^w)\,\mathrm{d}\mu^w = \mu^w\left(\left[\tfrac{x^b}{x^w}, \infty\right]\right) \quad \text{and}$$
$$A(B^b(x)) = \int_{\Theta^b} B^b(x)(\theta^b)\,\mathrm{d}\mu^b = \mu^b\left(\left[\tfrac{x^w}{x^b}, \infty\right]\right).$$

We can therefore write the dynamic (A) as

$$\text{(5a)} \qquad \dot{x}^w = \mu^w\left(\left[\tfrac{x^b}{x^w}, \infty\right]\right) - x^w,$$
$$\text{(5b)} \qquad \dot{x}^b = \mu^b\left(\left[\tfrac{x^w}{x^b}, \infty\right]\right) - x^b.$$

By setting the left hand sides of equations (5a) and (5b) equal to 0, we obtain explicit conditions for state (x^w, x^b) to be an aggregate equilibrium:

$$\text{(6a)} \qquad x^w = \mu^w\left(\left[\tfrac{x^b}{x^w}, \infty\right]\right),$$
$$\text{(6b)} \qquad x^b = \mu^b\left(\left[\tfrac{x^w}{x^b}, \infty\right]\right).$$

Thus, in an aggregate equilibrium, the mass of population p agents in the neighborhood equals the mass of such agents who prefer to reside in the neighborhood. Apart from the accounting for committed types, equations (6a) and (6b) are an alternate form of Schelling's equilibrium equations (1a) and (1b).

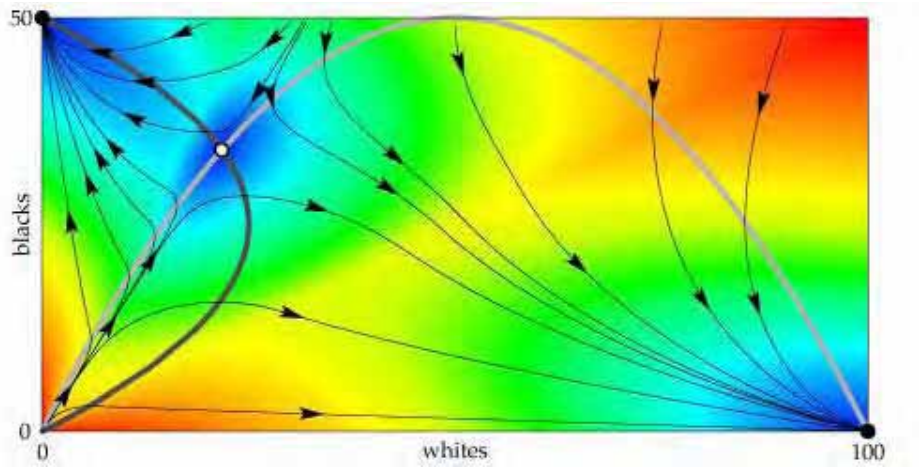


Figure 2: Schelling's first example revisited: no stable integration.

3.4 Revisiting Two of Schelling's Examples

Example 3.3. With the groundwork complete, let us revisit Schelling's first example, described in Sections 2.2 and 2.3 above. Recall that that example featured a white population twice as large as the black population ($m^w = 100, m^b = 50$), with distributions of tolerances in each population uniformly distributed on [0, 2].

To ensure that the ratios $\frac{x^b}{x^w}$ and $\frac{x^w}{x^b}$ are always well-defined, we introduce small masses $m^w_\infty = m^b_\infty = .01$ of type ∞ agents to each population, so that the total population masses become $m^w = m^w_f + m^w_\infty = 100.01$ and $m^b = m^b_f + m^b_\infty = 50.01$. Applying equations (5a) and (5b), we express the aggregate best response dynamic for this game as

$$\dot{x}^w = \left(\max\left\{100 - 50\tfrac{x^b}{x^w},\ 0\right\} + .01\right) - x^w,$$
$$\dot{x}^b = \left(\max\left\{50 - 25\tfrac{x^w}{x^b},\ 0\right\} + .01\right) - x^b.$$

Figure 2 presents a phase diagram for this dynamic. In this diagram and those to follow, solution trajectories are lines marked with arrows representing the direction of motion. Colors represent the speed of motion, with red fastest and blue slowest. The dark and light gray curves represent the blacks' and whites' nullclines—that is, the states at which the rate of entry of each group is 0. Reviewing Figure 1 reveals that the graphs of the functions T^w and T^b from Schelling (1971) correspond to the nullclines in our figure.

Turning to the equilibrium states, we see that the white dot in Figure 2, located at $x_\star = (21.7401, 34.0276)$, represents an unstable integrated equilibrium, while the black dots at (100.0050, .01) and (.01, 50.0050) represent stable segregated equilibria. Evidently, solution trajectories from almost all initial conditions converge to a segregated equilibrium. §

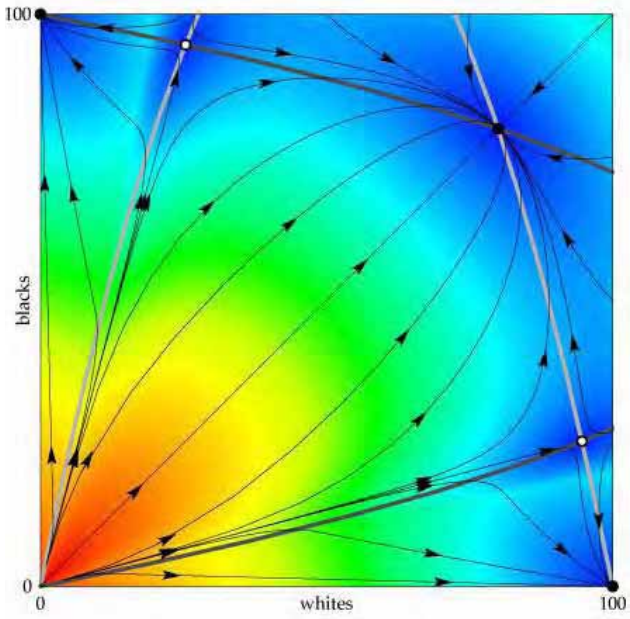


Figure 3: Schelling's second example: stable integration.

Example 3.4. While in the previous example integration is unstable, Schelling (1971) also offers an example in which integration is stable. Suppose that each population consists of a set of agents of mass $m_f^w = m_f^b = 100$ with tolerances distributed uniformly on $[0, 5]$, as well as a small mass $m_\infty^w = m_\infty^b = .01$ of committed types. The aggregate best response dynamic generated by equations (5a) and (5b), described by

$$\dot{x}^w = \left(\max\left\{100 - 20\tfrac{x^b}{x^w},\ 0\right\} + .01\right) - x^w,$$
$$\dot{x}^b = \left(\max\left\{100 - 20\tfrac{x^w}{x^b},\ 0\right\} + .01\right) - x^b,$$

is illustrated in Figure 3. States (100.0080, .01), and (.01, 100.0080) are stable segregated equilibria, state (80.01, 80.01) is a stable integrated equilibrium, and states (25.3590, 94.6410) and (94.6410, 25.3590) are unstable integrated equilibria. §

4. Analysis of the Aggregate Dynamic

The Bayesian best response dynamic (B) provides a satisfying but technically cumbersome model of the evolution of behavior. As we saw in Section 3.3, many important properties of this dynamic are captured by the aggregate best response dynamic (A).

To begin our analysis of the aggregate dynamic, we express it in a more convenient form. Let $f^w : [0, \infty) \to \mathbf{R}$ and $f^b : [0, \infty) \to \mathbf{R}$ denote the density functions for $\mu^w|_{[0,\infty)}$ and

$\mu^b|_{[0,\infty)}$. Then the dynamic (A) (cf equations (5a) and (5b)) becomes

$$\dot{x}^w = \int_{\frac{x^b}{x^w}}^{\infty} f^w(\theta^w)\,\mathrm{d}\theta^w + m_\infty^w - x^w, \tag{7a}$$

$$\dot{x}^b = \int_{\frac{x^w}{x^b}}^{\infty} f^b(\theta^b)\,\mathrm{d}\theta^b + m_\infty^b - x^b. \tag{7b}$$

Let us henceforth assume that the density functions f^w and f^b are continuous, so that the dynamic (7) is continuously differentable (C^1).[10] Writing the dynamic (7) as $\dot{x} = V(x)$, we observe that the derivative matrix for V at x is easily computed as

$$DV(x) = \begin{pmatrix} \dfrac{\partial V^w}{\partial x^w}(x) & \dfrac{\partial V^w}{\partial x^b}(x) \\ \dfrac{\partial V^b}{\partial x^w}(x) & \dfrac{\partial V^b}{\partial x^b}(x) \end{pmatrix} = \begin{pmatrix} \dfrac{f^w(r^{bw})\,x^b}{(x^w)^2} - 1 & -\dfrac{f^w(r^{bw})}{x^w} \\ -\dfrac{f^b(r^{wb})}{x^b} & \dfrac{f^b(r^{wb})\,x^w}{(x^b)^2} - 1 \end{pmatrix}, \tag{8}$$

where we let $r^{bw} \equiv x^b/x^w$ and $r^{wb} \equiv x^w/x^b$.

The following lemmas note two important consequences of this calculation. The first is the key to determining local stability of equilibria, while the second is the basis for our analysis of global behavior.

Lemma 4.1. *The eigenvalues of $DV(x)$ are $\lambda(x) = \dfrac{f^w(r^{bw})\,x^b}{(x^w)^2} + \dfrac{f^b(r^{wb})\,x^w}{(x^b)^2} - 1$ and -1.*

Lemma 4.2. $\dfrac{\partial V^w}{\partial x^b}(x)$ *and* $\dfrac{\partial V^b}{\partial x^w}(x)$ *are nonpositive.*

4.1 Local Stability of Equilibria

4.1.1 Stability of Segregated Equilibria

We call equilibrium $x_\star$ *segregated* if one group only has committed types residing in the neighborhood. A segregated equilibrium is *predominantly black* if $x_\star^w = m_\infty^w$ and $x_\star^b > m_\infty^b$, *predominantly white* if $x_\star^w > m_\infty^w$ and $x_\star^b = m_\infty^b$, and *empty* if $x_\star^w = m_\infty^w$ and $x_\star^b = m_\infty^b$.

Theorem 4.3 shows that under mild assumptions, there are unique predominantly black and predominantly white equilibria, and that both of these equilibria are locally stable. For brevity, the statement of the theorem focuses on predominantly black equilibria.

[10] For our local stability results to hold, it is enough that the densities f^w and f^b be continuous at the equilibrium ratios $x_\star^b/x_\star^w$ and $x_\star^w/x_\star^b$, respectively.

Theorem 4.3. *(i) Let* $\underline{x}^b = \mu^b([\frac{m^w_\infty}{m^b_\infty}, \infty]) > m^b_\infty$*, and suppose that* $\mu^w([\frac{\underline{x}^b}{m^w_\infty}, \infty]) = m^w_\infty$*. Then there is at least one predominantly black equilibrium* $x_\star = (m^w_\infty, x^b_\star)$*, and all such equilibria satisfy* $x^b_\star \geq \underline{x}^b$*.*

(ii) Suppose in addition that

$$f^b\left(\frac{m^w_\infty}{x^b}\right) < \frac{(x^b)^2}{m^w_\infty} \tag{9}$$

for all $x^b > \underline{x}^b$*. Then the predominantly black equilibrium is unique and asymptotically stable.*

(iii) More generally, if $x_\star = (m^w_\infty, x^b_\star)$ *is a predominantly black equilibrium, then* $x_\star$ *is asymptotically stable if it satisfies condition* (9)*, and it is unstable if it satisfies condition* (9) *with the inequality reversed.*

The proof of this result is presented in the Appendix.

Part (i) of Theorem 4.3 tells us that if some noncommitted blacks have at least moderate tolerances, and if no noncommitted whites have very high tolerances, then a predominantly black equilibrium exists, and that all such equilibria have a nonnegligible mass of blacks in the neighborhood. Part (ii) states that if there is no tolerance level at which the tolerance density f^b is exceptionally high, then the predominantly black equilibrium is unique and locally stable. Finally, part (iii) shows that even in the absence of uniqueness, the density condition (9) and its opposite still provide sufficient conditions for stability and instability of predominantly black equilibria.

In the next section, we will see that the link between tolerance densities and local stability persists for integrated equilibria.[11]

4.1.2 Stability and Instability of Integrated Equilibria

We call an equilibrium $x_\star$ *integrated* if there are noncommitted types from each population who reside in the neighborhood: that is, if $x^w_\star > m^w_\infty$ and $x^b_\star > m^b_\infty$. Theorem 4.4, a direct consequence of Lemma 4.1, characterizes the stability of integrated equilibria.

Theorem 4.4. *An integrated equilibrium* $x_\star$ *is a sink (and hence asymptotically stable) if*

$$f^w(r^{bw}_\star)\, r^{bw}_\star + f^b(r^{wb}_\star)\,(r^{wb}_\star)^2 < x^w_\star, \tag{10}$$

while $x_\star$ *is a saddle (and hence unstable) if this inequality is reversed.*

What makes an integrated equilibrium stable? According to Theorem 4.4, stability of equilibrium depends directly on the masses $x^w_\star$ and $x^b_\star$ of agents of each type in the

[11] For related results in the context of purified equilibria of normal form games, see Sandholm (2007).

neighborhood, as well as on the equilibrium tolerance densities $f^w(r_\star^{bw})$ and $f^b(r_\star^{wb})$. In particular, having small numbers of nearly indifferent agents leads to stability, while large numbers of indifferent agents leads to instability.

For intuition, consider an equilibrium $x_\star$ at which no whites are close to indifferent, and suppose that a shock causes the number of whites in the neighborhood to fall by a small amount. Since no whites were initially close to indifferent, the proportion of whites whose best response is *in* remains fixed, so the number of whites in the neighborhood rises back toward the equilibrium level $x_\star^w$. At the same time, the mass of blacks whose best response is *in* goes up, causing the number of blacks choosing *in* to increase from $x_\star^b$.

How evolution proceeds depends on the number of blacks who are initially nearly indifferent. If there are few, then the rise of x^w back toward $x_\star^w$ will proceed quickly relative to rise of x^b away from $x_\star^b$. When x^w comes close enough to $x_\star^w$, x^b starts falling back toward $x_\star^b$, and the equilibrium $x_\star$ is restored. On the other hand, if there are many blacks initially close to indifferent, then the disequilibrating change in x_b outpaces the equilibrating change in x_w; at some point, enough blacks enter the neighborhood that whites begin to leave, and the integrated equilibrium is destroyed.

Examining expression (10) more carefully, we find that if the equilibrium white/black ratio $r_\star^{wb}$ is, say, relatively large, then it is the density of indifferent blacks that is key to determining stability. The reason for this is not difficult to divine. When the ratio r^{wb} is large, changes in x^w and x^b have more dramatic effects on r^{wb} than on r^{bw}. Since the equilibrium ratios $r_\star^{wb}$ and $r_\star^{bw}$ are also the equilibrium tolerance levels of indifferent black and white agents respectively, the claim immediately follows.

To illustrate the behavior of his class of dynamics, Schelling (1971) draws the nullclines of x^w and x^b to represent his restrictions on feasible directions of motion. Theorem 4.5 shows that in the context of the Bayesian best response dynamic, these same nullclines are very useful for local stability analysis: except in nongeneric cases, local stability under this dynamic is completely determined by the slopes of the nullclines at the equilibrium.

Theorem 4.5. *Suppose that equilibrium $x_\star$ is hyperbolic (i.e., that $\lambda(x_\star) \neq 0$). Then $x_\star$ is asymptotically stable if*

(*i*) *both nullclines have negative or infinite slopes at $x_\star$, and*

(*ii*) *the whites' nullcline is steeper than the blacks' nullcline at $x_\star$*

Otherwise, $x_\star$ is unstable.

The proof of this result can be found in the Appendix.

All of the examples in Section 3.4 and in Section 4.3 below illustrate Theorem 4.5: all stable equilibria satisfy conditions (i) and (ii) of the theorem, while all unstable equilibria fail one condition or the other. However, in nongeneric cases in which the nullclines

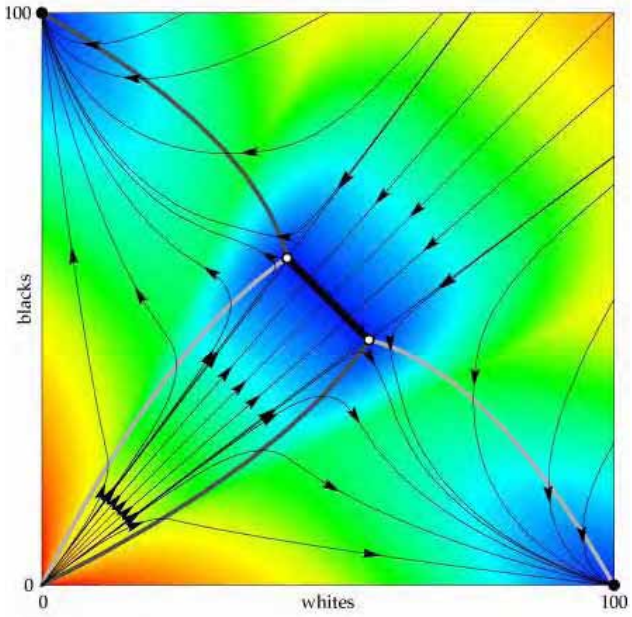


Figure 4: Lyapunov stable equilibria that are not asymptotically stable.

overlap, one can have Lyapunov stable equilibria that are not asymptotically stable. The next example illustrates this point.

Example 4.6. Let $m^p = m_f^p + m_\infty^p = 100 + .1$ for $p \in \{w, b\}$, and let the distributions of tolerances satisfy

$$f^p(\theta^p) = \begin{cases} 80 & \text{if } \theta^p \in [0, \frac{3}{4}], \\ \frac{100}{(\theta^p+1)^2} & \text{if } \theta^p \in (\frac{3}{4}, \frac{4}{3}], \\ \frac{450}{7} & \text{if } \theta^p \in (\frac{4}{3}, 2], \\ 0 & \text{otherwise.} \end{cases}$$

Figure 4 presents the phase diagram of the resulting aggregate best response dynamic. The relative interior of the thick line black consists of Lyapunov stable equilibria; white and black dots represent unstable and stable equilibria, respectively. §

4.2 Global Behavior of the Dynamic

We saw in Lemma 4.2 that the off-diagonal elements of the derivative matrix $DV(x)$ are always nonpositive: higher numbers of black agents in the neighborhood reduce the entry rate of white agents, and vice versa. Differential equations with this property are said to be *competitive*. Classical results from the dynamical systems literature show that

two-dimensional competitive systems have very simple global behavior—see Smith (1995, Theorem 3.2.2) or Hofbauer and Sigmund (1998, Section 3.4). Theorem 4.7 presents the consequences of these results in the current setting.

To state this theorem, we let

$$R_{+-} = \{x \in X_\circ : V^w(x) \geq 0 \text{ and } V^b(x) \leq 0\},$$

denote the set of states at which x^w is weakly increasing and x^b is weakly decreasing under the aggregate dynamic (A). We define the sets R_{++}, R_{--}, and R_{-+} analogously.

Theorem 4.7. *Under the aggregate dynamic* (A),

(*i*) *Sets R_{+-} and R_{-+} are forward invariant, and sets R_{++} and R_{--} are backward invariant.*

(*ii*) *Along each solution trajectory $\{x_t\}_{t\geq 0}$, either $\{x_t^w\}_{t\geq 0}$ or $\{x_t^b\}_{t\geq 0}$ is monotone, and the other changes direction at most once.*

(*iii*) *Every solution trajectory converges to an aggregate equilibrium $x_\star$.*

We begin our discussion of Theorem 4.7 with the intuition behind the invariance results in part (i). Consider, for instance, how a trajectory starting in region R_{+-} might escape into another region. Evidently, escape directly into region R_{-+} would require passing through a rest point of V, a contradiction. We therefore consider escape into a one of the two remaining regions, say R_{++}. This escape requires $V^b(x)$ to become positive; in particular, at the escape point it must be that $V^w(x) \geq 0$ and $V^b(x) = 0$.

Now, express the rate of change over time of $\dot{x}^b = V^b(x)$ as

$$\frac{d}{dt}V^b(x) = \frac{\partial V^b(x)}{\partial x^w}\dot{x}^w + \frac{\partial V^b(x)}{\partial x^b}\dot{x}^b. \tag{11}$$

At the escape point, the second term on the right hand side of (11) is zero, since $\dot{x}^b = 0$. Furthermore, since $\dot{x}^w \geq 0$ at the escape point, and since $\frac{\partial V^b(x)}{\partial x^w} \leq 0$ by the competitiveness of the dynamic, the first term on the right hand side of (11) is non-positive, implying that $\frac{d}{dt}V^b(x) \leq 0$. But since $V^b(x) = 0$ at the escape point, $V^b(x)$ cannot become positive, contradicting that an escape point has been reached. A similar argument establishes the forward invariance of R_{-+}. The backward invariance of R_{--} and R_{++} follows from the same argument, but with time running backwards.

Once part (i) is established, part (ii), and hence the absence of cycles, follows easily. The nullclines of the dynamic partition the state space into regions, each of which is contained in either R_{+-}, R_{-+}, R_{++}, or R_{--}. Clearly, forward invariance implies that solutions starting in R_{+-} or R_{-+} are monotonic in both components, with one component shrinking and the other growing. Furthermore, solutions starting in R_{++} or R_{--} are monotone in each

component while they remain in the region; these solutions either converge to a rest point or enter R_{+-} or R_{-+}, after which the previous analysis holds force. Either way, the solution obeys the restrictions stated in part (ii) of the theorem. Since the solution trajectory is eventually componentwise monotone, and since the state space $X_\circ$ is compact, it follows immediately that the solution must converge to an equilibrium, proving part (iii) of the theorem.[12]

4.3 Examples

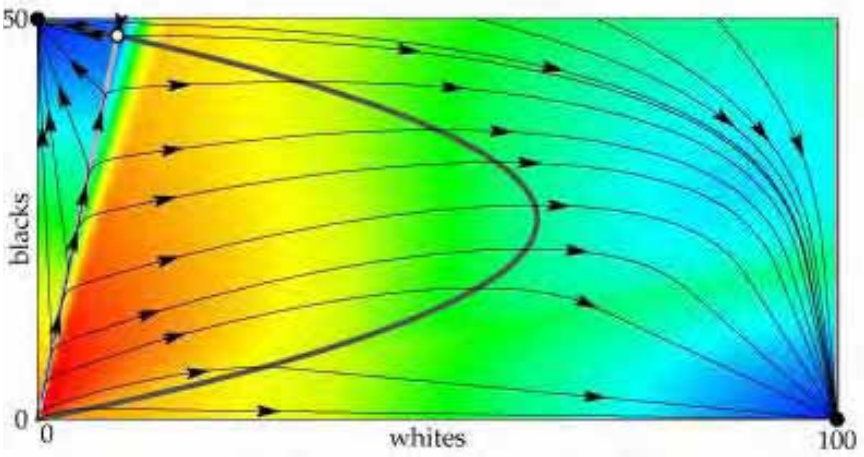
Example 4.8. Reducing tolerance to sustain integration. Figure 5 presents phase diagrams for three examples that differ only in the distribution of tolerances in the white population. In all three figures, we have $m^w = m_f^w + m_\infty^w = 100 + .1$ and $m^b = m_f^b + m_\infty^b = 50 + .1$, with the tolerances of noncommitted blacks uniformly distributed on [0, 5]. In Figure 5(i), all noncommitted whites have high tolerances: the full mass $m_f^w = 100$ of these agents have tolerances uniformly distributed on [3, 5]. In Figure 5(ii), we replace some high tolerance agents with low tolerance agents: mass 70 have tolerances uniformly distributed on [3, 5], and the remaining mass of 30 have tolerances uniformly distributed on [0, .5]. Figure 5(iii) takes this one step further, giving the *uniform*[3, 5] group mass 60, and the *uniform*[0, .5] group mass 40. Evidently, it is only in the final specification that a stable integrated equilibrium exists.

While at first glance this example seems counterintuitive, the logic behind it is simple. Since there are twice as many whites as blacks, entry by whites into an integrated neighborhood can leave it with a low percentage of blacks; this leads blacks with lower tolerances to start exiting the neighborhood, starting a feedback loop that leaves the neighborhood predominantly white. Replacing high tolerance whites with some low tolerance whites is tantamount to reducing the mass of the white population, making it less likely that entry by whites ultimately leads to an exodus of blacks.[13]

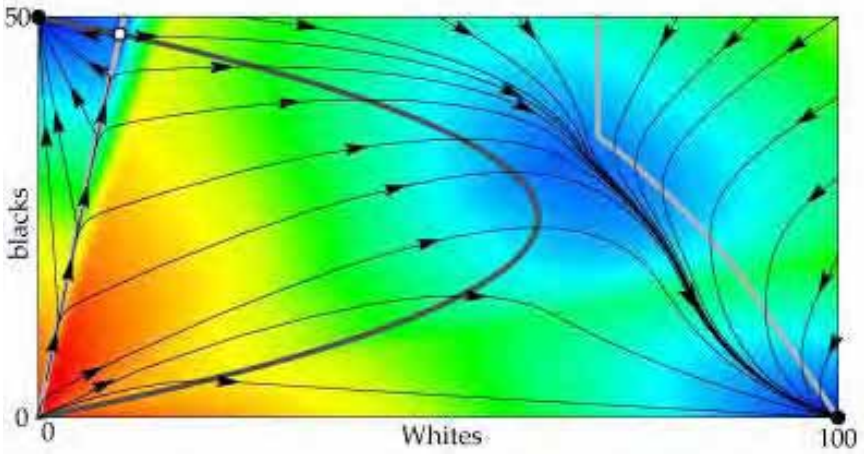
We can also interpret this result "locally", using the density condition from Theorem 4.4. In Figure 5(i), the unique integrated equilibrium, $x_\star \approx (10.0032, 48.0126)$, has a black/white ratio of $r_\star^{bw} \approx 4.7997$, and hence a white/black ratio of $r_\star^{wb} \approx .2083$, ratios that lie in the supports of f^w and f^b, respectively. As Theorem 4.4 indicates, the existence of relatively large numbers of indifferent agents causes adjustments away from equilibrium to be self-reinforcing, and so is a source of instability.

[12] In a competitive system, the nullcline condition in Theorem 4.5 is sufficient condition for local stability of equilibrium—see Hirsch and Smale (1974, p. 271) for an example. However, the necessity of the nullcline condition is not true for general competitive systems, but instead depends on the specific form of the dynamic (A).

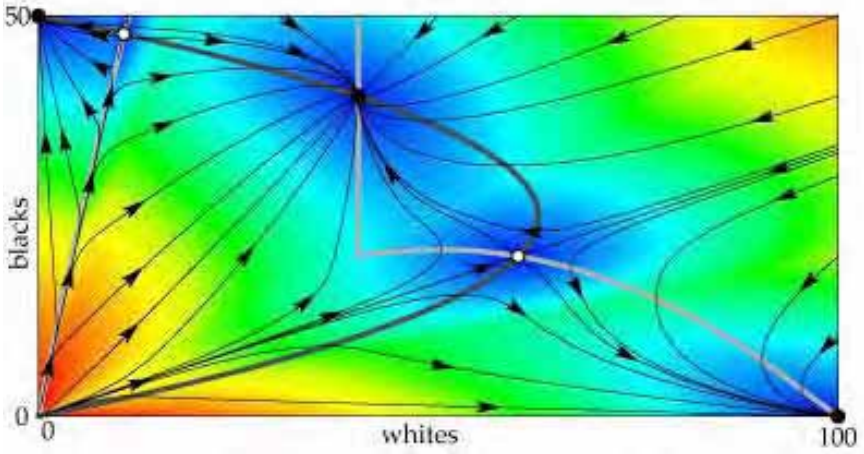
[13] Schelling (1971, p. 174-175) offers a similar example and discussion.



(i) $m_f^w = 100$, distributed $uniform[3, 5]$



(ii) $m_f^w = 100$, distributed $(.7\ uniform[3, 5] + .3\ uniform\ [0, .5])$



(iii) $m_f^w = 100$, distributed $(.6\ uniform[3, 5] + .4\ uniform\ [0, .5])$

Figure 5: Making the white population less tolerant can create a stable integrated equilibrium.

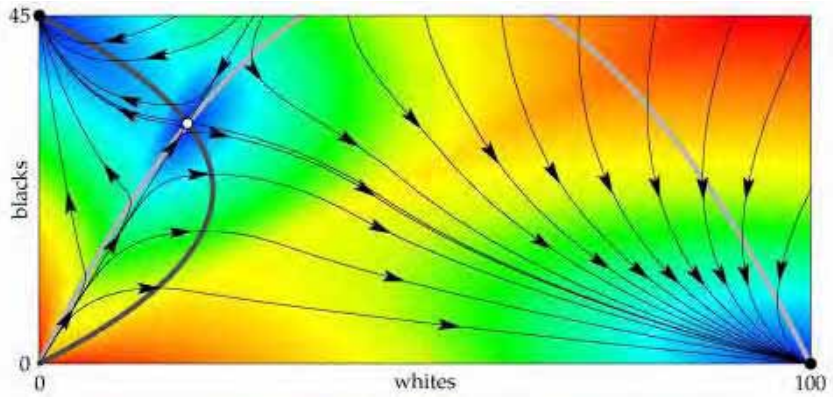
In contrast, Figure 5(iii) features an integrated equilibrium $x_\star \approx (40.04, 40.0933)$. Since the ratio of blacks to whites in this equilibrium is near unity, while the support of tolerance desnity f^w is $[0, .5] \cap [3, 5]$, there are no whites who are close to indifferent at this equilibrium. It follows that small changes in behavior do not affect any whites' best responses, with the consequence that whites' behavior tends to return to the integrated equilibrium level after any small disturbance. Consequently, even though a disturbance adding whites to the neighborhood (i.e., a disturbance sending the state to the right of the black dot representing $x_\star$) initially leads marginal blacks to leave, these blacks return to the neighborhood as x^w returns to $x^w_\star$.

Note, though, that Figure 5(iii) also shows two unstable integrated equilibria, at (60.0941, 20.0143) and (10.7189, 47.8557). In these equilibria, the black/white ratios lie in the support of f^w; the existence of many indifferent agents again generates instability. §

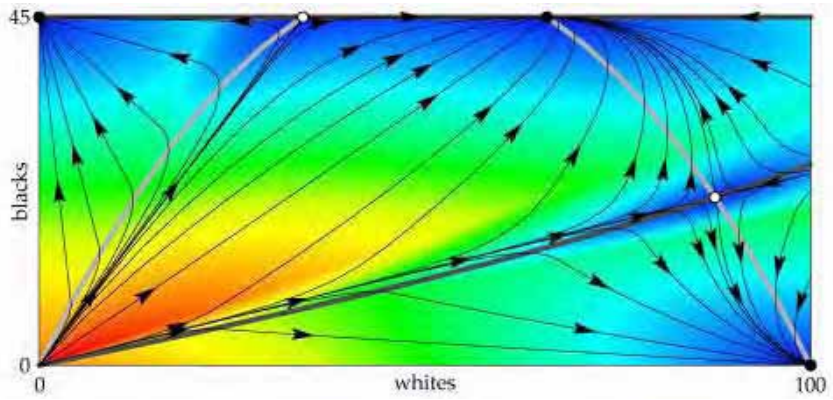
Example 4.9. Relative population sizes and stability of integration. In the examples pictured in Figure 6, the white population is of mass $m^w = m^w_f + m^w_\infty = 100 + .1$, and noncommitted types have tolerances that are uniformly distributed on $[0, 2]$. In Figure 6(i), the black population is of mass $m^b = m^b_f + m^b_\infty = 45 + .1$, and noncommitted types have tolerances distributed uniformly on $[0, 2]$; here the unique integrated equilibrium is unstable. In Figure 6(ii), we increase tolerances in the black population, distributing them uniformly on $[3, 5]$. Doing so creates two unstable integrated equilibria, as well as a stable integrated equilibrium at $x_\star \approx (65.7900, 45.1)$, at which all blacks and most whites reside in the neighborhood. By making tolerances high in the black population, we ensure that blacks are willing to reside in the neighborhood even when they are outnumbered by whites. Since the number of blacks is relatively small, the entry of all blacks does not cause the whites to leave. But if we increase the mass of the black population to $55 + .1$, as in 6(iii), then no positive number of noncommitted whites can coexist with all of the blacks in an integrated equilibrium. §

5. Extensions

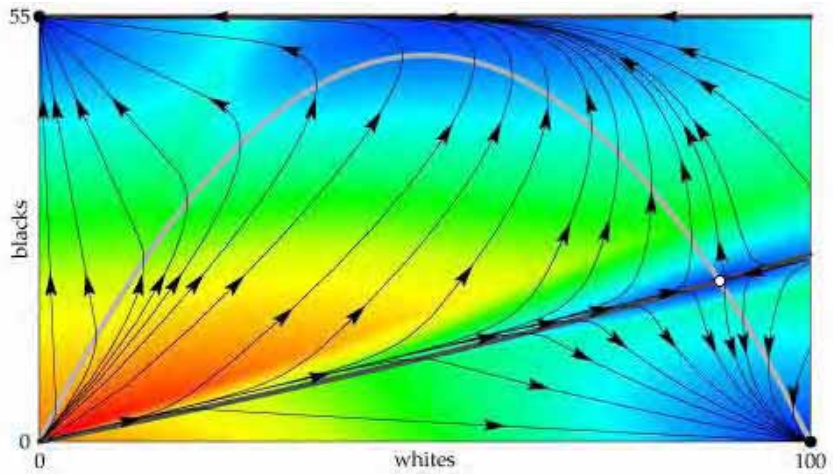
To this point, we have focused our attention on a formal version of Schelling's (1971) original segregation model. But beyond providing a platform from which to obtain analytical results for Schelling's model, the Bayesian population game framework allows us to consider a broader range of environments than would be possible using an informal approach. We now offer some preliminary examples in this vein.



(i) $m_f^b = 45$, distributed $uniform[0, 2]$



(ii) $m_f^b = 45$, distributed $uniform[3, 5]$



(iii) $m_f^b = 55$, distributed $uniform[3, 5]$

Figure 6: Increasing tolerances in the black population creates a stable integrated equilibrium. After this, increasing the mass of the black population destroys the equilibrium.

5.1 A Dual-Threshold Model

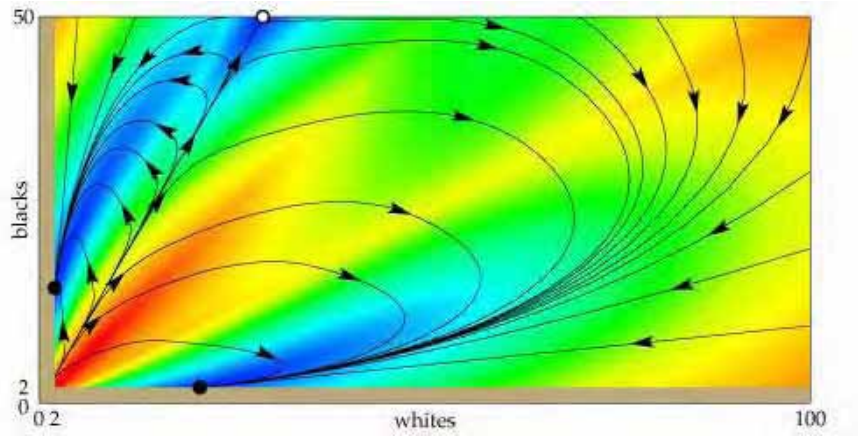
To this point, we have followed Schelling (1971) in assuming that agents' preferences are described by a single tolerance threshold. While this model is appealingly simple, is certainly not the only one worthy of study.

To take one simple variation, we can assume that agents' preferences are captured by two thresholds: for an agent to prefer to reside in the neighborhood, the ratio of other group to own group in the neighborhood must not only be less than an upper bound θ_h^p, but must also exceed a lower bound θ_l^p. Thus, agents still avoid neighborhoods in which their group is insufficiently represented, but they also avoid neighborhoods in which their group is overrepresented.[14]

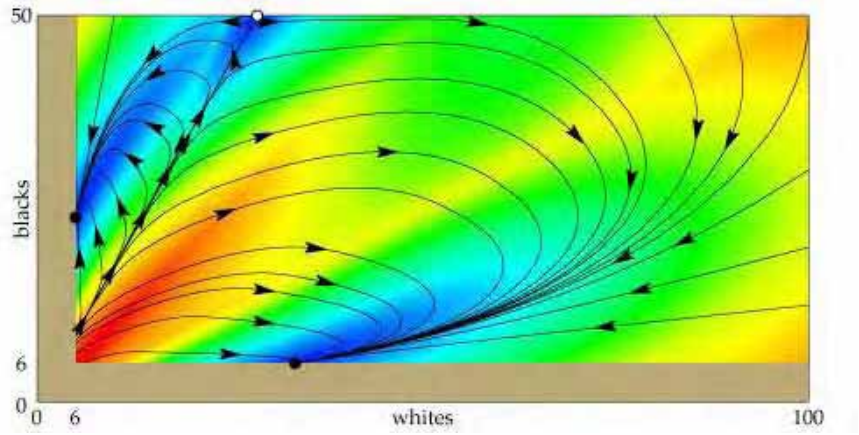
While a complete treatment of this model is beyond the scope of this paper, we assert that under mild smoothness conditions on the distributions of $\theta^w = (\theta_l^w, \theta_h^w)$ and $\theta^b = (\theta_l^b, \theta_h^b)$, the best response functions $B^w : X_\circ \Rightarrow \Sigma_\circ^w$ and $B^b : X_\circ \Rightarrow \Sigma_\circ^b$ are Lipschitz continuous. Thus, solutions to the Bayesian best response dynamic for this model exist, and satisfy appropriate analogues of the aggregation results from Section 3.3. The following two examples illustrate that introducing a distaste for homogeneity leads to qualitatively different behavioral dynamics.

Example 5.1. In the examples in Figure 7, the population's masses are $m^w = 100$ and $m^b = 50$. In each population, the distribution of the types $\theta^p = (\theta_l^p, \theta_h^p)$ of the noncommitted agents is uniform on the line segment with endpoints $(0, 1)$ and $(1, 3)$; thus, $(\theta_l^p, \theta_h^p) = (0, 1)$ is the type vector of the least tolerant agent, while $(\theta_l^p, \theta_h^p) = (1, 3)$ is the type vector of the most tolerant noncommitted agent. The two diagrams in the figure differ only in the masses of committed types: in Figure 7(i), the masses are $m_\infty^w = m_\infty^b = 2$, while in Figure 7(ii) they are $m_\infty^w = m_\infty^b = 6$. In both figures, the only stable outcomes are segregated, but in contrast to the segregated equilibria from the single threshold case, the segregated equilibria here are sparsely populated. The reason for this is easy to see. When, for example, there are very few whites in the neighborhood, the most tolerant whites will find the neighborhood too homogenous, and so will exit. This exodus is only halted by the presence of blacks who are committed to living in the neighborhood, which prevents the noncommitted whites with the lowest lower thresholds from exiting. Comparing Figures 7(i) and 7(ii), we see that increasing the number of committed blacks increases the number of whites residing in the neighborhood in the predominantly white equilibrium. §

[14]This model relies on an implicit assumption that the outside option offers some degree of group heterogeneity. Rather than specifying the racial compositions of outside locations exogenously, it would be preferable to determine the compositions of all locations endogenously—see Section 5.3 below.

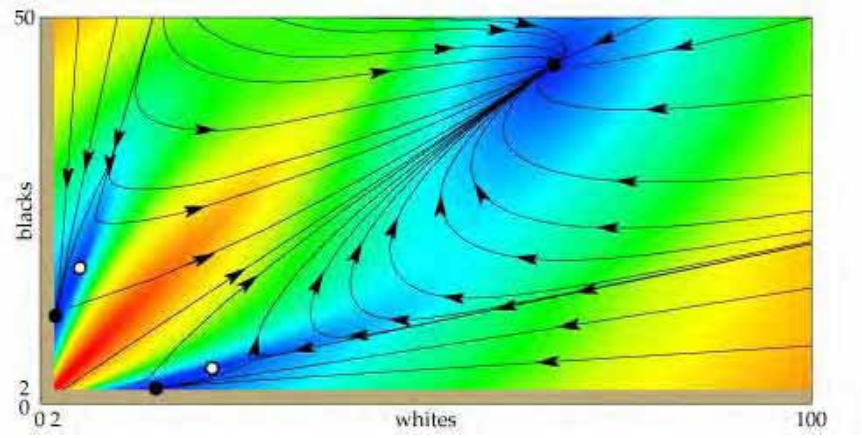


(i) $m_\infty^w = m_\infty^b = 2$

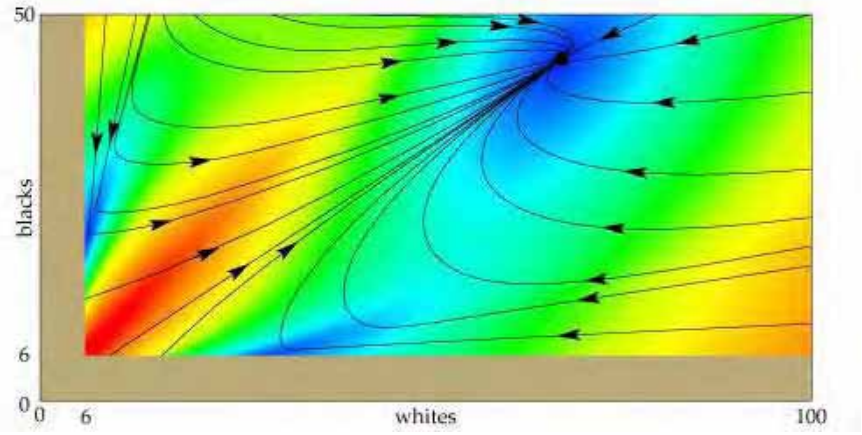


(ii) $m_\infty^w = m_\infty^b = 6$

Figure 7: Dual thresholds, low tolerances. Stable equilibria are segregated and sparsely populated.



(i) $m_\infty^w = m_\infty^b = 2$



(ii) $m_\infty^w = 6, m_\infty^b = 6$

Figure 8: Dual thresholds, high tolerances. A stable integrated equilibrium exists; with enough committed types it is the unique equilibrium.

Example 5.2. In the examples pictured in Figure 8, the population's masses are $m^w = 100$ and $m^b = 50$, with the masses of committed types equal to $m^w_\infty = m^b_\infty = 2$ in Figure 8(i), and equal to $m^w_\infty = m^b_\infty = 6$ in Figure 8(ii). Relative to those in the previous example, the agents here are more tolerant: in both cases, the types $\theta^w = (\theta^w_l, \theta^w_h)$ of noncommitted white agents are uniformly distributed on the segment with endpoints (0, 1) and (1, 3.5), while the types $\theta^b = (\theta^b_l, \theta^b_h)$ of noncommitted black agents are uniformly distributed on the segment with endpoints (0, 1) and (.9, 5.05). Evidently, making the agents more tolerant introduces the possibility of a stable integrated equilibrium.

Comparing Figures 8(i) and 8(ii), we see that increasing the masses of committed types has dramatic effects on the set of equilibria: when there are few committed types, the stable integrated equilibrium is supplemented by two stable segregated equilibria and two unstable integrated equilibria. With more committed types, these additional equilibria vanish, making the stable integrated equilibrium a global attractor. For intuition, bear in mind that in the dual threshold model, segregated equilibria, when they exist, tend to be sparsely populated. But if agents are relatively tolerant and the number of committed types is not too small, such equilibria cannot exist. For instance, if all committed agents and a relatively small number of noncommitted whites are in the neighborhood, the most tolerant blacks will prefer to enter. §

5.2 Taxation to Sustain Integration

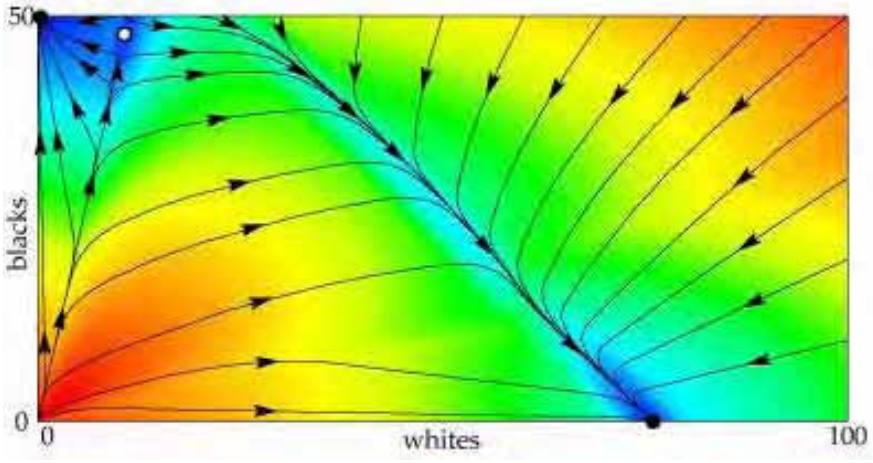
Since integration is generally viewed as a social goal, it is natural to consider policies whose aim is to promote this outcome. Here, we show how taxes can be used to sustain integration, and how the success of the policy can be sensitive to its fine details.

Example 5.3. Suppose as in Example 3.3 that the populations' masses are $m^w = m^w_f + m^w_\infty = 100 + .1$ and $m^b = m^b_f + m^b_\infty = 50 + .1$, and the tolerances of noncommitted agents in each population are distributed uniformly on [0, 5]. As we saw in Figure 2, the unique integrated equilibrium under this specification is unstable.

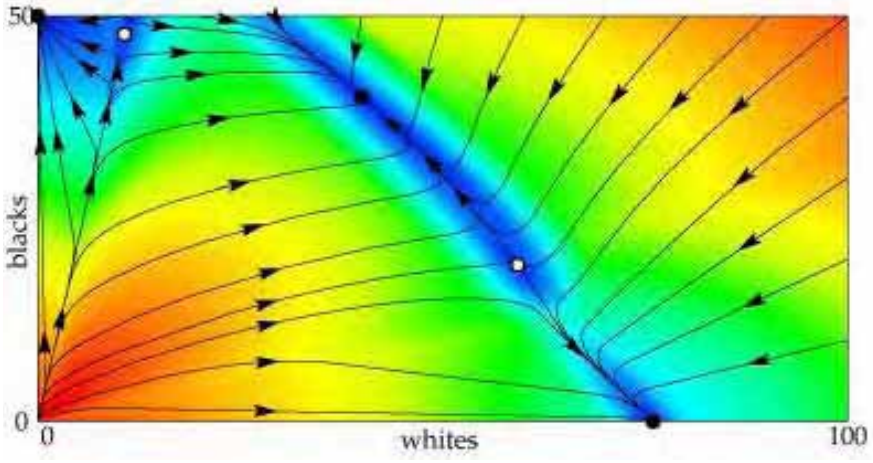
We now consider using taxes to sustain stable integration, while simultaneously keeping the total population of the neighborhood relatively small—in particular, below 80.[15] We suppose that tax rates can depend on the total mass $x^T = x_w + x_b$ of neighborhood residents, and can be set at different levels $\tau^w(x^T)$ and $\tau^b(x^T)$ for whites and blacks, as might be the case under various forms of affirmative action.

To introduce taxes, we must specify payoffs more precisely than we did in equation (3). We again normalize the payoff of *Out* to 0, but this time assume that the payoffs of

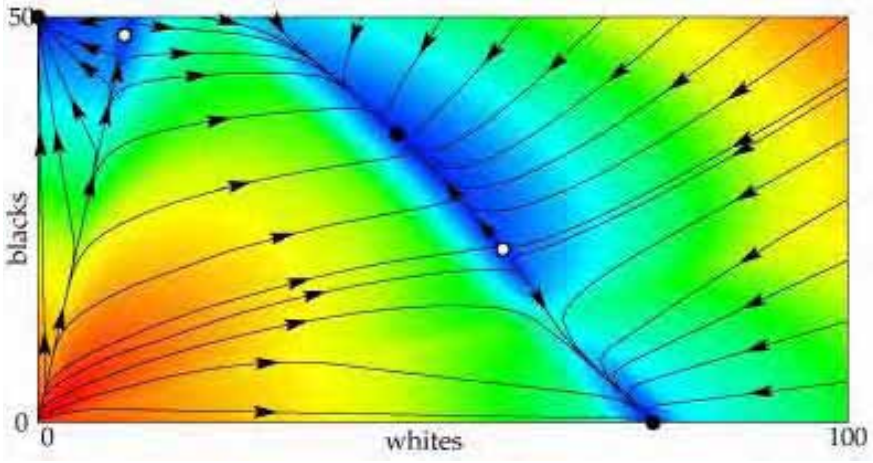
[15]Schelling (1971, p. 173-174) considers directly imposing a cap on the neighborhood's total population, but both the means of implementation and the consequences of such a cap are left vague.



(i) taxes on whites and blacks increase at rate $\frac{1}{5}$ for $x_T \in [70, 80]$



(ii) whites taxed as in (i), blacks' taxes increase at rate $\frac{1}{5}$ for $x_T \in [80, 90]$



(iii) whites taxed as in (i), blacks' taxes increase at rate $\frac{1}{50}$ for $x_T \in [70, 80]$

Figure 9: Taxation to sustain integration.

noncommitted agents to choosing *In* take the following functional form:

$$U^w_{In}(x, \theta^w) = \theta^w - \frac{x^b}{x^w} - \tau^w(x^T),$$
$$U^b_{In}(x, \theta^b) = \theta^b - \frac{x^w}{x^b} - \tau^b(x^T).$$

Thus, an agent's base payoff to residing in the neighborhood is the difference between his tolerance and the relevant neighborhood composition ratio; taxes enter the payoff function in a quasilinear fashion.[16]

In each diagram in Figure 9, we suppose that no taxes are imposed when the the neighborhood size x^T is below 70. To create Figure 9(i), we suppose that the tax on each resident increases linearly from 0 to 2 as the neighborhood size increases from 70 to 80, and remains fixed at 2 for higher levels of x^T. The taxes ensure that x^T cannot long remain above 80. Once x^T is approximately 76, it remains at this level. But the ratio of whites to blacks continues to change, with the state ultimately approaching the predominantly white equilibrium at (75.9908, .1).

In Figure 9(i), the region where $x^T \approx 76$ is blue, reflecting the fact that evolution proceeds slowly in this region. It stands to reason that small changes in tax policy could reverse the direction of motion of the dynamic in this region, thereby creating a stable integrated equilibrium. In Figure 9(ii), taxation of blacks is not initiated until $x^T = 80$; it increases linearly from 0 to 2 as x^T increases linearly from 80 to 90 and remains fixed at 2 thereafter. This policy generates a stable equilibrium at $x_\star \approx (40.0134, 40.1022)$. In Figure 9(iii), the tax rate for blacks increases less steeply than the tax rate for whites: in particular, we set $\tau^b(x^T) = \frac{1}{50}(x^T - 70)$ (versus $\tau^w(x^T) = \frac{1}{5}(x^T - 70)$) when $x^T \in [70, 80]$. This policy too generates a stable integrated equilibrium, this time at state $x_\star \approx (44.3790, 35.6387)$. §

5.3 Further Extensions

5.3.1 Multiple Unrestricted Neighborhoods

To this point, we have always assumed that each agents chooses between the neighborhood of interest and an outside location whose composition is fixed. Of course, a more complete model would account for the fact that when agents move to the outside location, they change the racial composition of that location, altering its appeal. Thus, a "general equilibrium" model in which the compositions of all locations are determined endogenously is of clear interest. Such a model is no more difficult to construct and analyze than

[16]This specification of payoffs is chosen for concreteness; the qualitative features of the example can be obtained using other functional forms.

the ones considered above, and we leave this task for future research.

5.3.2 Segregation by Race, Income, and Preferences for Public Goods

Throughout this paper, we have assumed that the sole characteristic that agents use to evaluate a neighborhood's desirability is its racial composition. While racial composition is an important determinant of residential location choice, it is hardly the only one. For instance, a neighborhood's average income level affects most people's assessments of its desirability. Indeed, attempts to elicit preferences for neighborhood racial composition is often subjected to the criticism that the preferences being elicited are those concerning income, for which race is serving as a proxy. By the same token, this paper has abstracted away from another key determinant of residential location choice, that of local public goods. Indeed, Tiebout's (1956) analysis of this issue is perhaps the main competitor of Schelling (1971) in terms of its influence on later work on neighborhood choice.

It is not too difficult to write down versions of our model that allow for heterogeneity in income or in public good preferences, and in which the existence and aggregation results from Section 3 continue to hold. However, determining the stability properties of the resulting aggregate dynamic, or even illustrating particular examples, becomes quite challenging, as the state variable for the aggregate dynamic is necessarily of higher dimension than 2. The construction and analysis of evolutionary models of segregation by race, income, and preferences for public goods is a challenging task for future research.

6. Concluding Remarks

In this paper, we use recent tools from evolutionary game theory to formalize, analyze, and develop extensions of the residential segregation model of Schelling (1971). Our approach captures the behavior dynamics of large, heterogeneous populations in a rigorous but tractable way. There many other economic issues whose modeling requires the introduction of behavior dynamics for heterogeneous populations. The present work demonstrates that the theory of Bayesian population games and Bayesian best response dynamics provides a powerful tool for analyzing such environments. We therefore have hope that this approach will prove fruitful as a foundation for models in other applied economic domains.

A. Appendix

The Proof of Lemma 3.1

We consider only the map B^w. The single-valuedness of this map is easily verified. To establish Lipschitz continuity, let x and y be social states, and assume without loss of generality that $x^b/x^w \leq y^b/y^w$. Letting M denote the bound on the density of $\mu^p|_{[0,\infty)}$ and K the Lipschitz coefficient for the map $x \mapsto x^b/x^w$ on domain $X_\circ$, we find that

$$\begin{aligned}\left\|B^w(x) - B^w(y)\right\| &= \int_{\Theta^w} \left|B^w(x)(\theta^w) - B^w(y)(\theta^w)\right| \mathrm{d}\mu^w \\ &= \mu^w\left(\left[\tfrac{x^b}{x^w}, \tfrac{y^b}{y^w}\right]\right) \\ &\leq M\left|\tfrac{x^b}{x^w} - \tfrac{y^b}{y^w}\right| \\ &\leq MK\left|x - y\right|.\end{aligned}$$

The Proof of Theorem 4.3

(i) By equilibrium conditions (6a) and (6b), a state $x_\star = (m^w_\infty, x^b_\star)$ with $x^b_\star > m^b_\infty$ is a predominantly black equilibrium if and only if

$$\text{(12a)} \qquad m^w_\infty = \mu^w\left(\left[\tfrac{x^b_\star}{m^w_\infty}, \infty\right]\right) \text{ and}$$

$$\text{(12b)} \qquad x^b_\star = \mu^b\left(\left[\tfrac{m^w_\infty}{x^b_\star}, \infty\right]\right).$$

Now as x^b increases from m^b_∞ to m^b, the map $x^b \mapsto \mu^b([\frac{m^w_\infty}{x^b}, \infty])$ is continuous and nondecreasing with $\mu^b([\frac{m^w_\infty}{m^b_\infty}, \infty]) = \underline{x}^b > m^b_\infty$. Thus, (12b) has a fixed point in $[\underline{x}^b, m^b]$, and all fixed points of (12b) lie in this interval. Since all fixed points of (12b) satisfy $x^b_\star \geq \underline{x}^b$, condition (12a) is satisfied at such points by assumption, proving the result.

(ii, iii) The slope of the map $x^b \mapsto \mu^b([\frac{m^w_\infty}{x^b}, \infty])$ is easily computed as $f^b(r^{wb})\, x^w/(x^b)^2$. Therefore, if inequality (9) holds this map can cross the 45° line only once, proving the uniqueness result in part (ii).

As for stability, since $x^b_\star \geq \underline{x}^b$ and since f^w is continuous, the fact that $\mu^w([\frac{x^b}{m^w_\infty}, \infty)) = 0$ implies that $f^w(r^{bw}_\star) = 0$. Thus, the eigenvalue $\lambda(x_\star)$ from Lemma 4.1 equals $f^b(r^{wb}_\star)\, m^w_\infty/(x^b_\star)^2 - 1$. Comparing this with condition (9) yields the stability result in part (ii), and part (iii) as well.

The Proof of Theorem 4.5

As in Section 2.2, let $t^w(x^w)$ denote the (x^w)th highest tolerance in the white population,

as implicitly defined by

$$x^w = \int_{t^w(x^w)}^{\infty} f^w(\theta^w)\,\mathrm{d}\theta^w + m_\infty^w.$$

There may be an interval of values of $t^w(x^w)$ that satisfy this inequality. However, if for some such $t^w(x^w)$ we have that $f^w(t^w(x^w)) > 0$, then $t^w(x^w)$ is uniquely defined. In this case, because of our maintained assumption that f^w is continuous, implicit differentiation reveals that

$$(t^w)'(x^w) = -\frac{1}{f^w(t^w(x^w))}.$$

Proceeding to follow Section 2.2, let $T^w(x^w) = x^w t^w(x^w)$, so that the whites' nullcline is the graph of T^w. If we suppose once more that $f^w(t^w(x^w)) > 0$, then the product rule tells us that the slope of the whites' nullcline is

$$(T^w)'(x^w) = t^w(x^w) - \frac{x^w}{f^w(t^w(x^w))}. \tag{13}$$

Now let $x_\star$ be an equilibrium. To defer special cases until the end of the proof, we suppose for now that

$$f_\star^w \equiv f^w(t^w(x_\star^w)) > 0,\ f_\star^b \equiv f^b(t^b(x_\star^b)) > 0,\ \text{and } f_\star^b \neq \frac{(x_\star^b)^2}{x_\star^w}. \tag{14}$$

Since $x_\star$ is an equilibrium, it lies on the whites' and blacks' nullclines. The former fact can be expressed as $T^w(x_\star^w) = x_\star^b$, and hence as $t^w(x_\star^w) = \frac{x_\star^b}{x_\star^w}$, so substituting into (13) reveals that

$$(T^w)'(x_\star^w) = \frac{x_\star^b}{x_\star^w} - \frac{x_\star^w}{f_\star^w}. \tag{15}$$

Repeating this entire argument for the blacks' nullcline shows that

$$(T^b)'(x_\star^b) = \frac{x_\star^w}{x_\star^b} - \frac{x_\star^b}{f_\star^b}. \tag{16}$$

These last two equalities imply that

$$(T^w)'(x_\star^w) \leq \frac{x_\star^b}{x_\star^w} \quad \text{and} \quad (T^b)'(x_\star^b) \leq \frac{x_\star^w}{x_\star^b},$$

from which it follows directly that

(17) if $(T^w)'(x^w)$ and $(T^b)'(x^b_\star)$ are nonnegative, then $(T^w)'(x^w_\star)\,(T^b)'(x^b_\star) < 1$.

Now compute as follows:

$$\begin{aligned} 1 - (T^w)'(x^w_\star)\,(T^b)'(x^b_\star) &= 1 - \left(1 - \frac{(x^b_\star)^2}{x^w_\star\, f^b_\star} - \frac{(x^w_\star)^2}{x^b_\star\, f^w_\star} + \frac{x^b_\star\, x^w_\star}{f^w_\star\, f^b_\star}\right) \\ &= \frac{x^b_\star\, x^w_\star}{f^w_\star\, f^b_\star}\left(\frac{x^b_\star\, f^w_\star}{(x^w_\star)^2} + \frac{x^w_\star\, f^b_\star}{(x^b_\star)^2} - 1\right) \\ &= \frac{x^b_\star\, x^w_\star}{f^w_\star\, f^b_\star}\,\lambda(x_\star). \end{aligned}$$

This computation shows that $1 - (T^w)'(x^w_\star)\,(T^b)'(x^b_\star))$ has the same sign as $\lambda(x_\star)$, the eigenvalue of $DV(x_\star)$ from Lemma 4.1; thus,

(18) $\lambda(x_\star) < 0$ if and only if $(T^w)'(x^w_\star)\,(T^b)'(x^b_\star) > 1$.

We now argue that under assumption (14), under which the whites' and blacks' nullclines have equilibrium slopes $(T^w)'(x^w_\star)$ and $((T^b)'(x^b_\star))^{-1}$, the theorem follows directly from equations (18) and (17). If the slopes of both nullclines are negative at $x_\star$, then (18) shows that $x_\star$ is stable if and only if the whites' nullcline is steeper than the blacks'. If one of the nullclines' slopes is negative and the other is not, then $(T^w)'(x^w_\star)\,(T^b)'(x^b_\star) \le 0$, so (18) shows that $x_\star$ is unstable. Finally, if both nullclines' slopes are nonnegative, then statement (17) shows that $(T^w)'(x^w_\star)\,(T^b)'(x^b_\star) < 1$, so (18) again shows that $x_\star$ is unstable.

We now address the cases in which assumption (14) does not hold. First, suppose that $f^w_\star = 0$, or, equivalently, that the slope of the whites' nullcline is $-\infty$. In this case, the blacks' nullcline is steeper than the whites' if and only if $((T^b)'(x^b_\star))^{-1}$ lies in in $(-\infty, 0)$, which is true if and only if $x^w_\star f^b_\star < (x^b_\star)^2$ by equation (16). But this inequality is exactly what is needed for $x_\star$ to be hyperbolically stable (i.e., to have $\lambda(x_\star) < 0$), proving the result in this case.

Next, suppose that $f^b_\star = 0$, or, equivalently, that the slope of the blacks' nullcline is 0. In this case, hyperbolic stability is equivalent to the requirement that $x^b_\star f^w_\star < (x^w_\star)^2$, which is in turn equivalent to the requirement that the slope of the white's nullcline lies in $[-\infty, 0)$ (cf equation (15)). This proves the theorem when $f^b_\star = 0$.

Finally, suppose that $f^b_\star = \frac{(x^b_\star)^2}{x^w_\star}$. Then the blacks' nullcline is vertical at $x_\star$, and so is at least as steep at the whites'; moreover, $\lambda(x_\star) \ge 0$, so $x_\star$ is not hyperbolically stable. This

completes the proof of the theorem.

References

Bobo, L. D. and Zubrinsky, C. L. (1996). Attitudes on residential segregation: Perceived status difference, mere in-group difference, or racial prejudice. *Social Forces*, 74:883–909.

Bøg, M. (2006). Is segregation robust? Unpublished manuscript, Stockholm School of Economics.

Charles, C. Z. (2003). The dynamics of racial residential segregation. *Annual Review of Sociology*, 29:167–207.

Clark, W. A. V. (1991). Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model. *Demography*, 28:1–19.

Ely, J. C. and Sandholm, W. H. (2005). Evolution in Bayesian games I: Theory. *Games and Economic Behavior*, 53:83–109.

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

Hirsch, M. W. and Smale, S. (1974). *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, San Diego.

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.

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

Möbius, M. M. (2000). The formation of ghettos as a local interaction phenomenon. Unpublished manuscript, MIT.

Pancs, R. and Vriend, N. J. (2007). Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics*, 91:1–24.

Sandholm, W. H. (2007). Evolution in Bayesian games II: Stability of purified equilibria. *Journal of Economic Theory*, forthcoming.

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.

Smith, H. L. (1995). *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems.* American Mathematical Society, Providence, RI.

Tiebout, C. M. (1956). A pure theory of local public expenditures. *Journal of Political Economy*, 64:416–424.

Young, H. P. (1998). *Individual Strategy and Social Structure*. Princeton University Press, Princeton.

Young, H. P. (2001). The dynamics of conformity. In Durlauf, S. N. and Young, H. P., editors, *Social Dynamics*, pages 133–153. Brookings Institution Press/MIT Press, Washington/Cambridge.

Zhang, J. (2004a). A dynamic model of residential segregation. *Journal of Mathematical Sociology*, 28:147–170.

Zhang, J. (2004b). Residential segregation in an all-integrationist world. *Journal of Economic Behavior and Organization*, 24:533–550.