Schelling Redux: An Evolutionary Dynamic Model of Residential Segregation

Emin Dokumaci† and William H. Sandholm‡

January 18, 2014

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’

†We thank many seminar audiences, Marzena Rostek, Larry Samuelson, and especially Daisuke Oyama for helpful comments. Financial support from NSF Grants SES-0617753 and SES-1155135 is gratefully acknowledged.
‡Wisconsin Center for Education Research, University of Wisconsin, 1025 West Johnson Street, Madison, WI 53706, USA. e-mail: edokumaci@wisc.edu; website: http://www.ssc.wisc.edu/~edokumaci.
§Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA. e-mail: whs@ssc.wisc.edu; website: http://www.ssc.wisc.edu/~whs.
preferences about the racial composition of their neighborhoods play an important role.\footnote{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. A basic reference on U.S. residential segregation is Massey and Denton (1993); see also Charles (2003).} 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.\footnote{See the previous references, as well as Clark (1991), Bobo and Zubrinsky (1996), and Farley et al. (1997).} 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.\footnote{Schelling further develops this insight in a broad array of contexts in his 1978 book, \textit{Micromotives and Macrobehavior}, which contains an abridged version of his 1971 paper.}

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 preferring 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.\footnote{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-dimensional 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).}

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.\footnote{Schelling refers to this model as the “bounded neighborhood” model.} Schelling focuses on how the distributions of tolerances in the two populations influence the set of equilibrium outcomes and the nature of dise-
equilibrium 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 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 \( \mu^w \) and \( \mu^b \), are measures on \([0, \infty)\) with total masses \( m^w \) and \( m^b \), respectively. Both of these measures are assumed to have bounded support and 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.

It will be convenient to introduce the decumulative distributions \( G^w \) and \( G^b \) corresponding to \( \mu^w \) and \( \mu^b \). Thus for \( \theta^w \in [0, \infty) \), \( G^w(\theta^w) = \mu^w([\theta^w, \infty)) \) is the mass of agents in the white population with tolerances of at least \( \theta^w \). \( G^b(\theta^b) = \mu^b([\theta^b, \infty)) \) is interpreted similarly. It is convenient to include the point at infinity in the domains of \( G^w \) and \( G^b \), and to define \( G^w(\infty) = G^b(\infty) = 0 \): in Schelling’s model, no one is infinitely tolerant.

Social state \( x = (x^w, x^b) \neq (0, 0) \) is an equilibrium if

\[
(1a) \quad G^w(\frac{x^b}{x^w}) = x^w \quad \text{and} \\
(1b) \quad G^b(\frac{x^w}{x^b}) = x^b.
\]

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. Condition (1a) says that the mass of white agents whose tolerance is at least the current white/black ratio is equal to the mass of white agents in the neighborhood. If there are only black agents in the neighborhood (\( x^w = 0 \)), then \( x^b/x^w = \infty \), and hence \( G^w(x^b/x^w) = 0 \), so the equilibrium condition (1a) is satisfied. (1b), the equilibrium condition for the black agents, is interpreted similarly. To avoid the expression 0/0, we rule out \( x^w = x^b = 0 \) by fiat.\(^7\)

\(^6\)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).

\(^7\)When we introduce our best response dynamics in Section 3, we take a less crude approach to avoiding
We illustrate these definitions using Schelling’s (1971) first example (p. 168–171). Here 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 \([0, 2]\), so that \(G^w(\theta^w) = 100 - 50\theta^w\) and \(G^b(\theta^b) = 50 - 25\theta^b\) for \(\theta^w, \theta^b \in [0, 2]\). Figure 1 describes the states where each equilibrium condition binds.\(^8\) Condition (1a) binds on curve \(W\) and on the vertical axis (since \(G^w(x^b/0) = 0\)); condition (1b) binds on the curve \(B\) and on the horizontal axis (since \(G^b(x^w/0) = 0\)). Finding the intersection points (and excluding the origin), we see that there are segregated equilibria at \((100, 0)\) and \((0, 50)\), and an integrated equilibrium at \(x_* = (x^w_*, x^b_*) \approx (21.7401, 34.0276)\).

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)

---

\(^8\)This graph, Figure 9 of Schelling (1978), is a cleaner version of Figure 18 of Schelling (1971).
We can represent Schelling’s dynamics using using ordinary differential equations on $X = X^w \times X^b$. At a minimum, these equations should satisfy

\begin{align}
\text{(2a)} & \quad \text{sgn}(\dot{x}^w) = \text{sgn}(G^w(x^b_x - x^w)), \\
\text{(2b)} & \quad \text{sgn}(\dot{x}^b) = \text{sgn}(G^b(x^w_x - x^b)).
\end{align}

The effect of dynamics from this class on the state $(x^w, x^b)$ is represented by the arrows in Figure 1. When the state is below curve W in the figure, $\dot{x}^w$ is positive,\(^9\) so the number of whites in the neighborhood increases. When the state is above curve W, $x^w$ falls, and when the state is on the curve, $x^w$ 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 curve B.

While one would need to specify the dynamics more precisely than in (2) to obtain exact solutions, the solutions’ qualitative features are 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 Disequilibrium Sorting

In order to formulate his dynamics on the set of social states $X$, Schelling must impose a 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 strong assumption to make about behavior during disequilibrium adjustment. We refer to it as the *disequilibrium sorting assumption*.

Without this sorting assumption, it is not immediately clear that the requirements (2) on the evolution of behavior need hold. To see why not, consider a state at which $G^w(x^b/x^w) > x^w$. Since the number of whites who prefer to be in the neighborhood exceeds the number who are in the neighborhood, the sorting assumption ensures that all discontented whites are outside the neighborhood, and thus system (2) says that the number of whites in the neighborhood should increase. But without this assumption, some whites in the neighborhood may be quite intolerant, and thus prefer to move out. If they do so more quickly than tolerant whites move in, the number of white agents in the neighborhood could fall.

---

\(^9\)To see this, observe that when $x^w \in (0, m^w)$ and $x^b = 0$, we have $\dot{x}^w = G^w(0) - x^w = m^w - x^w > 0$. 

---
3. Bayesian and Aggregate Best Response Dynamics

In order to construct dynamics that do not rely on disequilibrium sorting, one must track the behaviors of agents with different tolerances separately. We do so using the Bayesian best response dynamics, introduced in an abstract setting by Ely and Sandholm (2005). While these dynamics are infinite-dimensional, Ely and Sandholm (2005) prove that most important questions about these dynamics can be answered by examining finite-dimensional dynamics obtained by aggregation. In the present context, these aggregate best response dynamics dynamics satisfy condition (2) above.

Section 3.1 describes the Bayesian best response dynamics and aggregation an informal way, concluding with the definition of the aggregate best response dynamics 3 studied in the remainder of the paper. Stepping back, Section 3.2 states assumptions needed for the dynamics to be well defined and have basic regularity properties. With this groundwork in place, Section 3.3 presents the aggregate best response dynamic 3 for two of Schelling’s examples. A formal treatment of the Bayesian and aggregate best response dynamics is provided in Appendix A.

3.1 Informal Derivation and Definition of the Aggregate Dynamics

As before, we suppose that the distributions of types in the two populations are described by measures $\mu^w$ and $\mu^b$ on the extended interval $\Theta^w = \Theta^b = [0, \infty]$. The points in this interval are called tolerances, or, in Bayesian game language, types. Agents of type $\infty$, called committed types, prefer to live in the neighborhood under all circumstances; the role of these types will be explained below.

Each population $p \in \{w, b\}$, there is a continuum of agents of each type $\theta^p \in \Theta^p$. At a given point in time, behavior in population $p$ is described by a map $\sigma^p : \Theta^p \to [0, 1]$, where $\sigma^p(\theta^p) \in [0, 1]$ represents the proportion of agents in subpopulation $\theta^p$ who choose $\text{in}$. We call $\sigma^p$ a Bayesian population state. Being a map from the set of types $\Theta^p$ to distributions on $\{\text{in}, \text{out}\}$, $\sigma^p$ is the population analogue of a randomized strategy (more precisely, a behavior strategy) in a Bayesian game. We sometimes write $\sigma = (\sigma^w, \sigma^b)$, and call this object a Bayesian social state.

In our context, the Bayesian best response dynamic is a law of motion for the Bayesian social state $\sigma$. This dynamic describes how $\sigma$ changes if each agent occasionally receives opportunities to switch strategies, switching to his current best response when these opportunities arise.

Although the Bayesian best response dynamic is defined on a function space, it has one property that greatly eases its analysis: at each point in time, the target state toward which
\(\sigma\) moves—the so-called Bayesian best response to \(\sigma\)—only depends on the masses \(x^w\) and \(x^b\) of agents in each population who choose in, not on which types those agents are. This suggests the possibility of capturing certain aspects of the evolution of the Bayesian social state \(\sigma\) by examining an aggregate dynamic, a dynamic whose state variable \(x = (x^w, x^b)\) suppresses information about types.

Carrying out this plan, Ely and Sandholm (2005) define an aggregate best response dynamic with state variable \(x\). It is not difficult to show (see Appendix A.3) that there is a many-to-one map from solutions of the Bayesian dynamic to solutions of the aggregate dynamic. Because this map is many-to-one, it is not obvious whether stability results for the aggregate dynamic imply stability results for the Bayesian dynamic. Nevertheless, Ely and Sandholm (2005) prove that if a state is stable under the aggregate dynamic, then so is its counterpart under the Bayesian dynamic, where “stable” can refer to Lyapunov, asymptotic, or global asymptotic stability. They also show that instability under the aggregate dynamic implies instability of under the Bayesian dynamic. Thus, while we are directly concerned with the behavior of an infinite-dimensional dynamic on the set of Bayesian strategies, most of our questions about this dynamic can be addressed by studying a finite-dimensional aggregate dynamic.

We show in Appendix A that in the present setting, the aggregate best response dynamic takes the form

\[
\begin{align*}
\dot{x}^w &= G^w(\frac{x^b}{x^w}) - x^w, \\
\dot{x}^b &= G^b(\frac{x^w}{x^b}) - x^b,
\end{align*}
\]

where \(G^p(\theta^p) = \mu^p([\theta^p, \infty])\) is the mass of agents in population \(p\) whose type is at least \(\theta^p\). This is the simplest instance from the class of dynamics (2) introduced in our discussion of Schelling (1971). But while Schelling defines his dynamics by imposing a disequilibrium sorting assumption, the aggregate dynamics (3) do not require this; instead, they are obtained by aggregation from a full model of the evolution of the Bayesian social state.

One way to understand the difference Schelling’s dynamics and (3) is to consider what restrictions one would add to those in (2) to obtain dynamics that fit Schelling’s story more closely. A natural possibility is to link \(\dot{x}^w\) and \(\dot{x}^b\) to the number of “marginal agents” in each population—that is, the most tolerant agents who are out of the neighborhood, or the least tolerant agents who are in—since these are the agents who are switching locations.\(^{10}\)

\(^{10}\)A possible (though discontinuous) specification for such dynamics is

\[
\begin{align*}
\dot{x}^w &= f^w(t^w(x^w)) \text{sgn}(G^w(\frac{x^b}{x^w}) - x^w), \\
\dot{x}^b &= f^b(t^b(x^b)) \text{sgn}(G^b(\frac{x^w}{x^b}) - x^b),
\end{align*}
\]
By design, the dynamic (3) does not reflect this sort of restriction. In fact, since away from equilibrium it is common for each population to have discontented members both in the neighborhood and out of it, the time derivatives $\dot{x}^w$ and $\dot{x}^b$ under the aggregate dynamic (3) represent net rates of entry rather than exclusive entry or exit.

3.2 Additional Assumptions

Carrying out the constructions described above requires two new assumptions on the type distributions $\mu^w$ and $\mu^b$:

(D1) $\mu^w_{|[0,\infty)}$ and $\mu^b_{|[0,\infty)}$ admit bounded density functions.

(D2) $m_w^\infty \equiv \mu^w(\{\infty\}) > 0$ and $m_b^\infty \equiv \mu^b(\{\infty\}) > 0$.

These assumptions are used to ensure that the dynamics we study are well-defined, and admit unique solutions from every initial condition. Assumption (D1) is a smoothness condition on the distribution of finite types. Assumption (D2) requires there to be a positive mass of committed agents in each population; this assumption seems plausible from a descriptive point of view, and plays an important technical role in what follows. It will be useful 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 $p \in \{w, b\}$.

Finally, to avoid the technical problems noted in Section 2, we focus on the behavior of the Bayesian best response dynamic on the set of Bayesian social states at which all committed agents choose in. Because in is a dominant strategy for these agents, this set of states is forward invariant under the Bayesian best response dynamic, and so it is legitimate to focus on the behavior of the dynamics on this set.

Each Bayesian social state $\sigma = (\sigma^w, \sigma^b)$ in which committed agents choose in corresponds to an aggregate state $x = (x^w, x^b)$ with $x^w \geq m^w_\infty$ and $x^b \geq m^b_\infty$. Thus our subsequent analysis will focus on the behavior of the aggregate dynamic (3) on the set $X_\sigma = X^w_\sigma \times X^b_\sigma$, where $X^w_\sigma \equiv [m^w_\infty, m^w_\infty]$ and $X^b_\sigma \equiv [m^b_\infty, m^b_\infty]$. Since the ratios $x^b/x^w$ and $x^w/x^b$ are bounded on $X_\sigma$, the dynamic (3) is well-defined on this set. Assumption (D1) ensures further that (3) is Lipschitz continuous on $X_\sigma$, and so admits unique solutions from every initial condition in $X_\sigma$ (see Appendix A.3).

3.3 Revisiting Two of Schelling’s Examples

With the groundwork complete, we revisit two examples from Schelling (1971).

where $f^p$ is the type density function for population $p$ and $t^p(x^p) \in (G^p)^{−1}(x^p)$, so that mass $x^p$ agents in population $p$ have types above $t^p(x^p)$. (If $(G^p)^{−1}(x^p)$ is not a singleton, then the decumulative distribution has has a horizontal segment at height $x^p$, so any choice of $t^p(x^p)$ yields $f^p(t^p(x^p)) = 0$.)
Example 3.1. Schelling’s first example, described in Section 2, features 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 \(\infty\) agents to each population, so that the total population masses become \(m^w = m^w_f + m^w_{\infty} = 100.01\) and \(m^b = m^b_f + m^b_{\infty} = 50.01\).

We can express the aggregate best response dynamic (3) for this environment as

\[
\begin{align*}
\dot{x}^w &= \left(\max\left\{100 - 50\frac{x^b}{x^w}, 0\right\} + .01\right) - x^w, \\
\dot{x}^b &= \left(\max\left\{50 - 25\frac{x^w}{x^b}, 0\right\} + .01\right) - x^b.
\end{align*}
\]

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, the blacks’ and whites’ nullclines, contain those states at which the rate of entry of each group is 0, and correspond to curves B and W in Figure 1.

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. §

Example 3.2. 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 \(m^w_f = m^b_f = 100\) with tolerances distributed uniformly on \([0, 5]\), as well as a small mass \(m^w_{\infty} = m^b_{\infty} = .01\) of committed types. The aggregate best response
Figure 3: Schelling’s second example: stable integration.

The aggregate dynamic (3) for this environment is

\[
\dot{x}^w = \left(\max\left\{100 - 20x_b^w x^w, 0\right\} + .01\right) - x^w,
\]

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

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

To begin our analysis of the aggregate dynamic (3), we express it in a more convenient form. Letting \( f^p : [0, \infty) \to \mathbb{R} \) denote the density function for \( \mu^p|_{[0,\infty)} \), we can write the aggregate dynamic as

\[
\dot{x}^w = \int_{\frac{x^w}{\theta^w}}^{\infty} f^w(\theta^w) \, d\theta^w + m^w_\infty - x^w,
\]

\[
\dot{x}^b = \int_{\frac{x^b}{\theta^b}}^{\infty} f^b(\theta^b) \, d\theta^b + m^b_\infty - x^b.
\]
Let us henceforth assume that the density functions $f^w$ and $f^b$ are continuous, so that the dynamic (3) is continuously differentiable ($C^1$).\footnote{For our local stability results, it is enough that the densities $f^w$ and $f^b$ be continuous at the equilibrium ratios $x^w_*/x^b_*$ and $x^w_*/x^b_*$, respectively.} Writing the dynamic as $\dot{x} = V(x)$, we observe that the derivative matrix for $V$ at $x$ is easily computed as

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

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

The following lemmas note two basic consequences of this calculation.

**Lemma 4.1.** The aggregate best response dynamic (3) satisfies

$$
\frac{\partial V^w}{\partial x^b}(x) \leq 0, \quad \text{and} \quad \frac{\partial V^b}{\partial x^w}(x) \leq 0,
$$

$$
\frac{\partial V^w}{\partial x^b}(x) = 0 \quad \text{implies that} \quad \frac{\partial V^w}{\partial x^w}(x) < 0, \quad \text{and} \quad \frac{\partial V^w}{\partial x^b}(x) = 0 \quad \text{implies that} \quad \frac{\partial V^b}{\partial x^b}(x) < 0.
$$

Condition (5) says that the off-diagonal elements of the derivative matrix $DV(x)$ are always nonpositive: higher numbers of black agents in the neighborhood weakly reduce the net entry rate of white agents, and vice versa. Differential equations with this property are said to be competitive. Classical results from mathematical biology and dynamical systems theory show that two-dimensional competitive systems have very simple global behavior—see Hirsch and Smale (1974, Section 12.3) or Hofbauer and Sigmund (1998, Section 3.4).

Condition (6) holds force when a competitiveness condition is satisfied with equality. It states, for instance, that if increasing the number of black agents in the neighborhood has no effect on the net entry rate of white agents, then increasing the number of white agents must reduce their net entry rate. Under the aggregate dynamic (3), the antecedent holds when no white agents are indifferent between locations at the current state. In this case, moving some white agents into the neighborhood does not change any white agents’ best responses—that is, $f^w(r^{bw}) = 0$. But a tolerant agent who is moved in the neighborhood goes from being potential entrant to being content, and an intolerant agent who is moved into the neighborhood goes from being content to being a potential departee; both possibilities reduce the net entry rate, as stated in the consequent.

In Sections 4.1 and 4.2, we by establishing local and global stability results that hold
for any dynamic satisfying conditions (5) and (6), and so are quite robust to changes in the specification of the dynamic. In Section 4.3, we obtain further insights into local stability by making use of the functional form of the dynamic (3).

4.1 Monotonicity and Global Convergence

Theorem 4.2 presents the consequences of the competitiveness condition (5) for global behavior. 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 (3). We define the sets \( R_{++}, R_{-+}, \) and \( R_{--} \) analogously.

**Theorem 4.2.** Under any dynamic on \( X_\circ \) satisfying condition (5),

(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 \( \{x^w_t\}_{t \geq 0} \) or \( \{x^b_t\}_{t \geq 0} \) 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 4.2 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 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

\[ \frac{d}{dt} V^b(x) = \frac{\partial V^b(x)}{\partial x^w} \dot{x}^w + \frac{\partial V^b(x)}{\partial x^b} \dot{x}^b. \]

At the escape point, the second term on the right hand side of (7) 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 (7) 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 $X$ is compact, it follows immediately that the solution must converge to an equilibrium, proving part (iii) of the theorem.

4.2 Nullclines and Local Stability

To illustrate the behavior of his dynamics, Schelling (1971) draws the nullclines of $x^w$ and $x^b$ to represent his restrictions on feasible directions of motion. By doing so, he is able to distinguish between stable and unstable integrated equilibria in the examples he constructs. In the mathematical biology literature, it is known that in competitive systems in the plane, local stability of equilibrium can be assessed by examining the manner in which the nullclines cross—see Hirsch and Smale (1974, Section 12.3). We now show that under dynamics satisfying conditions (5) and (6), except in nongeneric cases, the local stability of equilibrium is determined by the slopes of the nullclines at the equilibrium.

The nullclines for the white and black populations are defined by

$$N^w = \{x \in X_0 : V^w(x) = 0\} \quad \text{and} \quad N^b = \{x \in X_0 : V^b(x) = 0\}.$$

To prove our characterization of local stability, we must first describe the behavior of the nullclines under conditions (5) and (6). Proposition B.1 in Appendix B.1 shows that each nullcline is the union of smooth curves through $X_0$, and establishes a variety of properties of these curves. This result allows us to define functions $s^w : N^w \to [-\infty, \infty)$ and $s^b : N^b \to (-\infty, \infty) \cup \{\pm \infty\}$ describing the nullclines’ slopes—see equations (14) and (15). Combining these functions with an exhaustive analysis of the possible eigenvalues of the derivative matrix $DV(x^*)$, we obtain the following result, which we prove in Appendix B.2.\footnote{Equilibrium $x^*$ is hyperbolic if $DV(x^*)$ has no eigenvalue with zero real part, so that its local stability properties are determined by its eigenvalues.}

--15--
Theorem 4.3. Consider a dynamic $\dot{x} = V(x)$ on $X$ satisfying condition (5) and (6). Suppose that equilibrium $x_*$ is hyperbolic. If $-\infty \leq s^w(x_*) < s^b(x_*) \leq 0$, then $x_*$ is asymptotically stable; otherwise, $x_*$ is unstable.

All of the examples in Section 3.3 and in Section 4.5 below illustrate Theorem 4.3: all stable equilibria satisfy conditions (i) and (ii) of the theorem, while all unstable equilibria fail one condition or the other. However, in nongeneric examples in which the nullclines overlap, one can have Lyapunov stable equilibria that are not asymptotically stable—see Appendix B.3 for an example.

4.3 Further Properties of Equilibrium

Here we establish results that depend on the specific form of the dynamic (3).

Define $r_{bw}$ and $r_{wb}$.

Lemma 4.4. The eigenvalues of $DV(x)$ are $\lambda(x) = \frac{f^w(r_{bw}) x^b}{(x^w)^2} + \frac{f^b(r_{wb}) x^w}{(x^b)^2} - 1$ and $-1$.

Implication: No sources.

We call equilibrium $x_*$ segregated if one group only has committed types residing in the neighborhood. A segregated equilibrium is predominantly black if $x^w_*$ $= m^w_\infty$ and $x^b_* > m^b_\infty$, predominantly white if $x^w_* > m^w_\infty$ and $x^b_* = m^b_\infty$, and empty if $x^w_* = m^w_\infty$ and $x^b_* = m^b_\infty$. We call an equilibrium $x_*$ integrated if there are noncommitted types from each population who reside in the neighborhood: that is, if $x^w_* > m^w_\infty$ and $x^b_* > m^b_\infty$.

Theorem 4.5. Suppose that $G^b(\frac{m^w_\infty}{x^b_*}) > x^b_*$ and $G^w(\frac{x^b_*}{m^b_\infty}) = m^w_\infty$ for some $x^b_* \in X^b$. Then there exists at least one predominantly black equilibrium $(m^w_\infty, x^b_*)$ with $x^b_* > x^b_1$. If all such equilibria are hyperbolic, then at least one is asymptotically stable.

See Appendix B.4 for a proof.

Theorem 4.6. Aggregate equilibrium $x_*$ is a stable node (and hence asymptotically stable) if

$$\frac{f^w(r^w_{\ast\ast}) x^b_\ast}{(x^w_\ast)^2} + \frac{f^b(r^b_{\ast\ast}) x^w_\ast}{(x^b_\ast)^2} < 1,$$

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

What makes an integrated equilibrium stable? According to Theorem 4.6, stability of equilibrium depends directly on the masses $x^w_\ast$ and $x^b_\ast$ of agents of each type in the neighborhood, as well as on the equilibrium tolerance densities $f^w(r^w_{\ast\ast})$ and $f^b(r^b_{\ast\ast})$. 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 \( \text{in} \) remains fixed, so the number of whites in the neighborhood rises back toward the equilibrium level \( x^w_\star \). At the same time, the mass of blacks whose best response is \( \text{in} \) goes up, causing the number of blacks choosing \( \text{in} \) to increase from \( x^b_\star \).

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 \( x^w_\star \) will proceed quickly relative to rise of \( x^b \) away from \( x^b_\star \). When \( x^w \) comes close enough to \( x^w_\star \), \( x^b \) starts falling back toward \( x^b_\star \), 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 (8) more carefully, we find that if the equilibrium white/black ratio \( r^\text{wb}_\star \) 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 \( r^\text{wb}_\star \) is large, changes in \( x^w \) and \( x^b \) have more dramatic effects on \( r^\text{wb} \) than on \( r^\text{bw} \). Since the equilibrium ratios \( r^\text{wb}_\star \) and \( r^\text{bw}_\star \) are also the equilibrium tolerance levels of indifferent black and white agents respectively, the claim immediately follows.

For related results in the context of purified equilibria of normal form games, see Sandholm (2007).
4.4 Dynamics of the Black/White Ratio

For the dynamic (3), we can learn a great deal by focusing on black/white ratios \( r = r^{bw} \). To start, we ask which ratios are possible in an aggregate equilibrium. If the current black/white ratio is \( r \), and the length of the vector \( x = (x^b, x^w) \) is \( ||x|| \), then simple geometry shows that the numbers of white and black agents in the neighborhood are

\[
x^w = ||x||/\sqrt{1 + r^2} \quad \text{and} \quad x^b = r ||x||/\sqrt{1 + r^2}.
\]

The number of white agents who prefer to be in the neighborhood at ratio \( r \) is \( G^w(r) \). For this to agree with the number who are in the neighborhood, it must be that

\[
||x|| = G^w(r) \sqrt{1 + r^2}.
\]

Similarly, at ratio \( r \), the number of black agents who prefer to be in the neighborhood equals the number who are in the neighborhood when

\[
||x|| = G^b(r^{-1}) \sqrt{1 + r^2}/r.
\]

Equating the right hand sides and rearranging shows that for \( r^* \) to be an equilibrium ratio—a black/white ratio consistent with aggregate equilibrium—it must be that

\[
g(r^*) \equiv \frac{G^b(r^{-1})}{G^w(r)} = r^*.
\]

Conversely, it is easy to verify that if \( g(r^*) = r^* \), then \( x^* = (x^w, x^b) = (G^w(r_*), G^b(r_*^{-1})) \) is an aggregate equilibrium with ratio \( r_* \).

The function \( g \) is nondecreasing, since

\[
g'(r) = \frac{G^w(r)r^{-2}f^b(r^{-1}) - G^b(r^{-1})f^w(r)}{G^w(r)^2} = \frac{f^b(r^{-1})}{r^2G^w(r)^2} + \frac{G^b(r^{-1})f^w(r)}{G^w(r)^2} \geq 0.
\]

And if \( r_* \) is an equilibrium ratio and \( x_* = (x^w_*, x^b_*) \) the corresponding aggregate equilibrium, then

\[
g'(r_*) = \frac{f^w(r_*)}{(x^w_*)^2} + \frac{f^b(r_*^{-1})}{(x^b_*)^2}.
\]

Comparing this with (8), we see that in hyperbolic cases, \( x_* \) is stable when \( g \) crosses the 45° line from above at \( r_* \), and \( x_* \) is unstable when \( g \) crosses the 45° line from below at \( r_* \).
If \( g(r) > r \), then the ratio of blacks who want to be in to whites who want to be in exceeds \( r \), so \( r \) is not an equilibrium ratio. An ratio \( r \) with \( g(r) < r \) is not an equilibrium for analogous reasons. See Figure 4 for pictures of the function \( g \) for the previous examples.

Finally, if \( r^* \) is a fixed point of \( g \) and \( x \) is any state satisfying \( x^b/x^w = r^* \), then

\[
\dot{x}^b = G^b(r^{-1} - 1) - x^b = r^* G^w(r^*) - r^* x^w = r^* x^w.
\]

So either \( x \) is a fixed point of (3), or \( x^b/x^w = r^* \). From this it follows that the ray from the origin through any aggregate equilibrium \( x^* \) is forward invariant under (3); so is the open region between such rays. In retrospect, this property is evident in Figures 2 and 3.

In fact,

\[
\frac{d}{dt} \left( \frac{x^b}{x^w} \right) = \frac{\frac{d^2}{dt^2} (x^b x^w - x^b x^w)}{(x^w)^2} = \frac{(G^b(r-1) - x^b) x^w - x^b (G^w(r) - x^w)}{(x^w)^2} = \frac{G^b(r^{-1}) x^w - x^b G^w(r)}{(x^w)^2}.
\]

So, after some rearranging,

\[
\text{sgn} \left( \frac{d}{dt} \left( \frac{x^b}{x^w} \right) \right) = \text{sgn} \left( \frac{G^b(r^{-1}) - x^b}{G^w(r)} \right) = \text{sgn}(g(r) - r).
\]

That is, the whether the black/white ratio is increasing or decreasing is determined by whether \( g \) is above or below the identity function.

Also, we have the following simple fact:

\[
\frac{d}{dt} (x^w + x^b) = (G^w(r) - x^w) + (G^b(r^{-1}) - x^b) = (G^w(r) + G^b(r^{-1})) - (x^w + x^b).
\]

That is, whether the total neighborhood population is growing or shrinking is determined by whether the number of agents who would prefer to be in the neighborhood given the current black/white ratio exceeds the number who are currently in the neighborhood. Put differently: along the ray from the origin with slope \( r \), the neighborhood population is increasing if it is less than \( G^w(r) + G^b(r^{-1}) \) and decreasing if it is above \( G^w(r) + G^b(r^{-1}) \). However, motion is only along the ray if \( r \) is a fixed point of \( g \), i.e. an equilibrium ratio.

4.5 Examples

In both examples, omit the first figure?

Example 4.7. 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.
Figure 4: Log-log plots of the function $g(r) = G''(r^{-1})/G''(r)$ for the first two examples
In all three figures, we have \( m_w = m_f^w + m_w^\infty = 100 + .1 \) and \( m_b = m_f^b + m_b^\infty = 50 + .1 \), 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 \(\text{uniform}[3, 5]\) group mass 60, and the \(\text{uniform}[0, .5]\) 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 4.6. In Figure 5(i), the unique integrated equilibrium, \( x_\star \approx (10.0032, 48.0126) \), has a black/white ratio of \( r_{bw}^\star \approx 4.7997 \), and hence a white/black ratio of \( r_{wb}^\star \approx .2083 \), ratios that lie in the supports of \( f_w \) and \( f_b \), respectively. As Theorem 4.6 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.

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 density \( f_w \) 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. \(^{\S}\)

\(^{13}\)Schelling (1971, p. 174-175) offers a similar example and discussion.
Figure 5: Making the white population less tolerant can create a stable integrated equilibrium.
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.
Example 4.8. 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. A Dual-Threshold Model

We postulate a preference for mixed living and simply reinterpret the same schedules of tolerance to denote merely the upper limits to the ratios at which people’s preference for integrated residence is outweighed by their extreme minority status (or by their inadequately majority status).

The same model fits both interpretations. The results are as pertinent to the study of preferences for integration as to the study of preferences for separation. (The only asymmetry is that we did not postulate a lower limit to the acceptable proportion of the opposite color, i.e., an upper limit to the proportion of like color in the neighborhood.) (p. 180)

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 $\theta^u_p$, but must also exceed a lower bound $\theta^l_p$. Thus, agents still avoid neighborhoods in which their group is insufficiently represented, but they also avoid neighborhoods in which their group is overrepresented.\textsuperscript{14}

\textsuperscript{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 ?? below.
More specifically, we define the type distribution $\mu^p$ for population $p \in \{w, b\}$ to be a measure on the type space

$$\Theta^p = \{\theta^p = (\theta^p_\ell, \theta^p_u) \in [0, \infty) \times (0, \infty) : \theta^p_\ell < \theta^p_u\}.$$  

The inequality in this definition ensures that every type is prefers to be in the neighborhood for some neighborhood composition. Let $\mu^p_\ell$ and $\mu^p_u$ denote the marginal measures on lower thresholds $\theta^p_\ell$ and upper thresholds $\theta^p_u$.

(D1') $\mu^w_{\ell\to(0,\infty)}, \mu^w_{u\to(0,\infty)}, \mu^b_{\ell\to(0,\infty)}, \mu^b_{u\to(0,\infty)}$ admit bounded density functions.

(D2') $m^w_{0,\infty} \equiv \mu^w((0, \infty)) > 0$ and $m^b_{0,\infty} \equiv \mu^b((0, \infty)) > 0$.

These assumptions allow a positive mass of types $(\theta^p_\ell, \theta^p_u) \in [0] \times (0, \infty)$, whose lower bound is inconsequential; if all noncommitted agents are of this type, we recover our earlier model.

Again assuming that committed types $\theta^p = (0, \infty)$ are always in, we can define the Bayesian and aggregate best response dynamics for this model are analogously to in the one-threshold model; again, the Bayesian dynamics admit unique solutions from every initial condition with committed agents choosing in, and the aggregate dynamics summarize the behavior of the Bayesian dynamics. In the present case, letting $G^p_\ell$ and $G^p_u$ denote the decumulative distributions of $\theta^p_\ell$ and $\theta^p_u$, one can show that the aggregate best response dynamic here takes the form

(9a) \[
\dot{x}^w = G^w_u\left(\frac{x^w}{x^u}\right) - G^w_\ell\left(\frac{x^w}{x^u}\right) - x^w,
\]

(9b) \[
\dot{x}^b = G^b_u\left(\frac{x^b}{x^u}\right) - G^b_\ell\left(\frac{x^b}{x^u}\right) - x^b.
\]

Focusing on equation (9a), note that when $r^{bw} \in (0, \infty)$, $G^w_u(r) - G^w_\ell(r) = \mu^w(|\theta^w : r \in (\theta^p_\ell, \theta^p_u))|$ is the mass of white agents who prefer to be in when the black/white ratio in the neighborhood is $r^{bw}$.\footnote{This conclusion depends on the assumptions that all types $\theta^p$ satisfy $\theta^p_\ell < \theta^p_u$, and that the marginal distributions admit densities on $(0, \infty)$. Because committed types are assumed to choose in, the ratios $x^b/x^w$ and $x^w/x^b$ are always in this interval, allowing us to ignore the discontinuity of $G^p_\ell$ at 0.} Thus, as under (9), the mass in population $p$ choosing in increases under the aggregate dynamic (9) when the mass from that population currently choosing in is less than the mass for whom it is optimal to choose in.

Discuss DV(x): we replace the density $f^p(r)$ in (4) with the difference in densities $f^p_u(r) - f^p_\ell(r)$. Only a competitive system if $f^p_u(r) - f^p_\ell(r) \geq 0$ for all $r$. If most agents have $\theta^p_\ell = 0$ this condition may well hold, and versions of all of the results from Section 4 also hold for the dual threshold model. However, it is easy to imagine environments where $f^p_u(r) - f^p_\ell(r) < 0$ at small values of $r$. 


Global behavior: fewer \textit{in} in integrated equilibria.

The following two examples illustrate that introducing a distaste for homogeneity leads to qualitatively different behavioral dynamics.

\textit{Example 5.1.} Omit the second figure.

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 $\theta^p = (\theta^p_l, \theta^p_h)$ of the noncommitted agents is uniform on the line segment with endpoints $(0, 1)$ and $(1, 3)$; thus, $(\theta^p_l, \theta^p_h) = (0, 1)$ is the type vector of the least tolerant agent, while $(\theta^p_l, \theta^p_h) = (1, 3)$ 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 $m_{w,\infty} = m_{b,\infty} = 2$, while in Figure 7(ii) they are $m_{w,\infty} = m_{b,\infty} = 6$. 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
Figure 8: Dual thresholds, high tolerances. A stable integrated equilibrium exists; with enough committed types it is the unique equilibrium.
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. §

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. §

6. Conclusion

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.

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

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.

A. Bayesian and Aggregate Best Response Dynamics

In this appendix, we formally define the Bayesian best response dynamic introduced informally in Section 3, and then derive from this the aggregate best response dynamic (3). The roles of the assumptions introduced in Section (3.2) are also made clear here.
A.1 Bayesian and aggregate population states and social states

Recall from Section 3 that the distribution of types (tolerances) in population \( p \in \{w, b\} \) is given by a measure \( \mu^p \) on \( \Theta^p = [0, \infty] \) that satisfies conditions (D1) and (D2) from Section 3. A Bayesian population state is a map \( \sigma^p : \Theta^p \rightarrow [0, 1] \), where \( \sigma^p(\theta^p) \in [0, 1] \) represents the proportion of agents in subpopulation \( \theta^p \) who choose in. We evaluate distances between points in \( \Sigma^p \) using the \( L^1 \) norm

\[
\|\sigma^p - \hat{\sigma}^p\| = \int_{\Theta^p} |\sigma^p(\theta^p) - \hat{\sigma}^p(\theta^p)| \, d\mu^p,
\]

so that the distance between Bayesian population states \( \sigma^p \) and \( \hat{\sigma}^p \) is just the \( (\mu^p-) \)average distance between subpopulations’ behaviors under \( \sigma^p \) and \( \hat{\sigma}^p \). States \( \sigma^p \) and \( \hat{\sigma}^p \) are considered equivalent if \( \sigma^p(\theta^p) = \hat{\sigma}^p(\theta^p) \) for \( (\mu^p-) \)almost every \( \theta^p \in \Theta^p \).

As described in Section 3, we restrict attention to Bayesian population states in which committed agents choose in. We denote the set of such states by

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

A complete description of behavior in both populations is provided by a Bayesian social state \( \sigma = (\sigma^w, \sigma^b) \in \Sigma = \Sigma^w \times \Sigma^b \). We let \( \Sigma_\circ = \Sigma^w_\circ \times \Sigma^b_\circ \) denote the set of social states in which committed agents choose in.

The mass of agents in population \( p \) who choose in at Bayesian population state \( \sigma^p \) is obtained by applying the aggregation operator \( A \):

\[
x^p = A\sigma^p = \int_{\Theta^p} \sigma^p(\theta^p) \, d\mu^p.
\]

This \( x^p \) is called an aggregate population state. Since \( \mu^p(\Theta^p) = m^p \) and \( \mu^p([\infty]) = m^p \), we have that \( A(\Sigma^p) = X^p \equiv [0, m^p] \) and \( A(\Sigma^p_\circ) = X^p_\circ \equiv [m^p_\circ, m^p] \). Similarly, \( x = A\sigma = (A\sigma^w, A\sigma^b) \) is an aggregate social state, and we write \( A(\Sigma) = X \equiv X^w \times X^b \) and \( A(\Sigma_\circ) = X_\circ \equiv X^w_\circ \times X^b_\circ \). (Like the expectation operator \( E \), the aggregation operator \( A \) integrates with respect to the measure or measures appropriate for its argument.)
A.2 Bayesian equilibrium and the Bayesian best response dynamic

To define the Bayesian best response dynamic, we first introduce the Bayesian best response correspondences \( B^w : X_0 \Rightarrow \Sigma^w_0 \) and \( B^b : X_0 \Rightarrow \Sigma^b_0 \):

\[
B^w(x)(\theta^w) = \begin{cases} 
1 & \text{if } \theta^w > x^b/x^w, \\
[0, 1] & \text{if } \theta^w = x^b/x^w, \\
0 & \text{if } \theta^w < x^b/x^w, 
\end{cases}
\]

\[
B^b(x)(\theta^b) = \begin{cases} 
1 & \text{if } \theta^b > x^w/x^b, \\
[0, 1] & \text{if } \theta^b = x^w/x^b, \\
0 & \text{if } \theta^b < x^w/x^b. 
\end{cases}
\]

Thus \( B^w(x) \in \Sigma^w_0 \) is a Bayesian population state, and \( B^w(x)(\theta^w) \) is the set of mixed best responses to aggregate social state \( x \) for agents of type \( \theta^w \). For instance, \( B^w(x)(\theta^w) \) equals \( \{1\} \), specifying that all agents of type \( \theta^w \) choose in, if tolerance \( \theta^w \) exceeds the black/white ratio at \( x \).

The following result is needed to establish the basic properties of the Bayesian best response dynamic.

**Lemma A.1.** Under assumptions (D1) and (D2), the maps \( B^w : X_0 \Rightarrow \Sigma^w \) and \( B^b : X_0 \Rightarrow \Sigma^b \) are single-valued and Lipschitz continuous.

**Proof.** We only consider the map \( B^w \). Given the definition of equivalence for Bayesian population states stated above, the single-valuedness of \( B^w \) 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 \( X_0 \), we find that

\[
\|B^w(x) - B^w(y)\| = \int_{\Theta^w} |B^w(x)(\theta^w) - B^w(y)(\theta^w)| \, d\mu^w \\
= \mu^w\left(\left[\frac{x^b}{x^w}, \frac{y^b}{y^w}\right]\right) \\
\leq M\left|\frac{x^b}{x^w} - \frac{y^b}{y^w}\right| \\
\leq MK\left|x - y\right|. \quad \blacksquare
\]

Bayesian social state \( \sigma \in \Sigma \) is a **Bayesian equilibrium** if \( (\sigma^w, \sigma^b) = (B^w(A\sigma), B^b(A\sigma)) \), or, more concisely, if \( \sigma = B(A\sigma) \). 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 \( \Sigma_* \subseteq \Sigma_0 \).

The **Bayesian best response dynamic** is defined by the law of motion

(11a) \( \dot{\sigma}^w = B^w(A\sigma) - \sigma^w \),

-31-
on the space on $\Sigma_\circ$. Under the Bayesian best response dynamic, behavior in the sub-population corresponding to each type $\theta^i$ adjusts in the direction of that type’s current best response. This dynamic, introduced by Ely and Sandholm (2005), extends the best response dynamic of Gilboa and Matsui (1991), defined for homogeneous populations, to settings where agents are heterogeneous with preferences drawn from continuous sets.

Since on $\Sigma_\circ$ is a function space, we must specify the norm with respect to which the time derivatives $\dot{\sigma}^w$ and $\dot{\sigma}^b$ are defined; we use the $L^1$ norm (10). Evidently, the rest points of (11) 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 A.1, and results in Ely and Sandholm (2005) imply that the dynamic (11) is well-behaved on $\Sigma_\circ$, in the sense that solutions from each initial condition in $\Sigma_\circ$ exist and are unique.

Theorem A.2. For each Bayesian social state $\sigma \in \Sigma_\circ$, there exists a unique $L^1$ solution $\{\sigma_t\}_{t \geq 0} \subset \Sigma_\circ$ to the dynamic (11) with $\sigma_0 = \sigma$.

A.3 Aggregate equilibrium and the aggregate best response dynamic

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

We call social state $x_* \in X_*$ an aggregate equilibrium, denoted $x_* \in X_*$, if $x_* = A(B(x_*))$. To see the point of this definition, suppose that $\sigma_* \in \Sigma_*$ is a Bayesian equilibrium, so that $\sigma_* = B(A\sigma_*)$. Then $A(B(A\sigma_*)) = A\sigma_*$, and so $A\sigma_*$ is an aggregate equilibrium. Conversely, if $x_* \in X_*$ is an aggregate equilibrium, so that $x_* = A(B(x_*))$, then $B(A(B(x_*))) = B(x_*)$, and so $B(x_*)$ 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 $A|_{X_*} : \Sigma_* \to X_*$ is a homeomorphism whose inverse is $B|_{X_*} : X_* \to \Sigma_*$.

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

\begin{equation}
\dot{x}^w = A(B^w(x)) - x^w
\end{equation}
(12b) \[ \dot{x}^b = A(B^b(x)) - x^b \]

on \( X \). Since the maps \( B: X \to \Sigma \) and \( A: \Sigma \to X \) are both Lipschitz continuous, so is their composition \( A \circ B: X \to X \), implying that from each initial condition \( x \in X \), the dynamic (12) admits a unique solution \( \{x_t\}_{t \geq 0} \subset X \) with \( x_0 = x \). And since

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

(12) is an alternate expression for the dynamic (3) from Section 3.1.

Evidently, the rest points of (12) are the aggregate equilibria of the underlying Bayesian population game. Moreover, if \( \{\sigma_t\}_{t \geq 0} \) is a solution to the Bayesian best response dynamic (11), 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 (11), then the aggregate behavior trajectory \( \{A\sigma_t\}_{t \geq 0} \) is a solution to (12). This argument shows that the aggregation operator \( A \) defines a many-to-one map from solutions of the Bayesian dynamic (11) to solutions of the aggregate dynamic (12). Furthermore, as described in Section 3.1, Ely and Sandholm (2005) show that stability and instability of Bayesian equilibrium under the Bayesian best response dynamic (11) are equivalent to stability and instability of the corresponding aggregate equilibrium under the aggregate best response dynamic (12). These results justify our conducting our analysis on the latter dynamic.

B. Analysis of Nullclines and Local Stability

B.1 Behavior of the Nullclines

We begin with by establishing the properties of the nullclines under conditions (5) and (6) mentioned in the text. Let \( N^w = \{x \in X: V^w(x) = 0\} \) be the white’s nullcline, and let \( N^w(x^w) = \{x^b \in X^b: (x^w, x^b) \in N^w\} \) be its vertical section at \( x^w \).

**Proposition B.1.** Suppose conditions (5) and (6) hold. Then

(i) Each set \( N^w(x^w) \) is empty, a singleton, or a closed interval.

(ii) Suppose \( x \) satisfies \( x^b \in \text{int}(X^b) \) and \( \frac{\partial V^w}{\partial x^w}(x) \neq 0 \). Then in a neighborhood of \( x \), \( N^w \) is the
graph of a differentiable function from a neighborhood of \( x^w \) in \( X^w \) to \( X^b \) whose slope at \( x^w \) is \( -\frac{\partial V^w}{\partial x^w}(x) / \frac{\partial V^w}{\partial x^w}(x) \).

(iii) Suppose \( x \) satisfies \( x^b \in \text{int}(X^b) \) and \( \frac{\partial V^w}{\partial x^b}(x) = 0 \). Then in a neighborhood of \( x \), \( N^w \) is the graph of a differentiable function from a neighborhood of \( x^b \) in \( X^b \) to \( X^w \) whose slope at \( x^b \) is 0.

(iv) Suppose \( N^w(x^w) \) is nonempty with \( \frac{\partial V^w}{\partial x^w}(x^w, x^b) = 0 \) for all \( x^b \in N^w(x^w) \). Let \( \{(x_k^w, x_k^b)\} \) be a sequence with \( \lim x_k^w = x^w \) and \( x_k^b \in N^w(x_k^w) \). If \( \{x_k^w\} \) is increasing, then \( x_k^b > \max N^w(x^w) \) for large enough \( k \) and \( \lim x_k^b = \max N^w(x^w) \), and if \( \{x_k^w\} \) is decreasing, then \( x_k^b < \min N^w(x^w) \) for large enough \( k \) and \( \lim x_k^b = \min N^w(x^w) \).

(v) Suppose that \( N^w(x^w) \) is empty for all \( x^w \in (x_1^w, x_2^w) \subset X^w \), and that \( N^w(x_1^w) \) and \( N^w(x_2^w) \) are nonempty. Then either \( N^w(x_1^w) = \{m^b\} \) and \( \max N^w(x_2^w) = m^b \), or \( \min N^w(x_2^w) = m_{\infty}^b \) and \( N^w(x_2^w) = \{m^b\} \).

Part (i) says that each vertical section of the nullcline is a closed convex set. Together, parts (ii) and (iii) show that each component of the nullcline is a smooth parameterized curve through \( X_0 \). Part (iv) says that at if a vertical section of the nullcline is an interval, then it can only be approached from the left at its highest point and from the right at its lowest point. Roughly speaking, part (v) says that the nullcline bypasses an interval of empty sections from above or from below.

Proof. Part (i) follows from (5). Part (ii) follows from the implicit function theorem, and part (iii) follows from (6) and the implicit function theorem.

To prove part (iv), note first that since \( \frac{\partial V^w}{\partial x^w}(x^w, x^b) = 0 \) for every \( x^b \in N^w(x^w) \), it follows that \( \frac{\partial V^w}{\partial x^w}(x^w, x^b) < 0 \) for such \( x^b \) by condition (6). Thus, the continuity of \( V^w \) and the compactness of \( N^w(x^w) \) imply that for some \( x_0^w < x^w \), \( V^w(x) > 0 \) for all \( x \in (x_0^w, x^w) \times N^w(x^w) \). Moreover, since (5) implies that when \( x^b < \min N^w(x^w) \), \( V^w(x^w, x^b) > 0 \), the continuity of \( V \) and the previous conclusion imply that for some \( x_0^w < x^w \) and \( \varepsilon > 0 \),

\[
V^w(x) > 0 \quad \text{for all} \quad x \in (x_1^w, x_2^w) \times [\min N^w(x^w) - \varepsilon, \max N^w(x^w)].
\]

Now let \( \{(x_k^w, x_k^b)\} \) satisfy the stated assumptions with \( \{x_k^w\} \) increasing. Then by the continuity of \( V^w \), every convergent subsequence of \( \{x_k^w\} \) must converge to a point in \( N^w(x^w) \), but by (13), no sequence can converge to a point in \( N^w(x^w) \) other than \( \max N^w(x^w) \). Thus since \( X^b_0 \) is compact, \( \{x_k^b\} \) must converge to \( \max N^w(x^w) \), so (13) implies that that all points far enough along this sequence must exceed \( \max N^w(x^w) \). The proof of the second claim is similar.

To prove part (v), note first that together, parts (i)-(v) imply that \( N^w(x_1^w) \) must contain either \( m^b \) or \( m_{\infty}^b \). (In particular, if \( x^b = \min N^w(x_1^w) \neq m_{\infty}^b \) and \( \frac{\partial V^w}{\partial x^w}(x_1^w, x^b) = 0 \), then part (iii)
implies that $N^w$ must continue downward from $(x^w_\ell, x^b)$, and part (iv) implies that it can only do so to the right, contradicting that $N^w(x^w_\ell)$ is empty to the right of $x^w_\ell$.) Similarly, $N^w(x^w_r)$ must also contain either $m^b$ or $m^b_\infty$. If $N^w(x^w_\ell) = \{m^b\}$, then (5) implies that when $x^b < m^b$, $V^w(x^w_\ell, x^b) > 0$. Since $N^w$ and $(x^w_\ell, x^w_r) \times X^b_\circ$ are disjoint, it follows by continuity that $V^w(x^w_\ell, x^b) > 0$ throughout the latter set. If max $N^w(x^w_\ell)$ were less than $m^b$, then (5) would imply that $V^w(x^w_\ell, x^b) > 0$ for $x^b > \max N^w(x^w_\ell)$, which by continuity is a contradiction; thus max $N^w(x^w_\ell) = m^b$. If instead max $N^w(x^w_r) < m^b$, then $m^b_\infty$ must be in $N^w(x^w_r)$, so part (i) implies that it is the only point in $N^w(x^w_r)$; then a similar argument to the previous one shows that min $N^w(x^w_r) = m^b_\infty$. ■

In light of Proposition B.1, we define the function $s^w : N^w \rightarrow [-\infty, \infty)$ describing the slope of the white’s nullcline $N^w$ as follows:

$$s^w(x) = \begin{cases} -\frac{\partial V^w}{\partial x^w}(x) / \frac{\partial V^w}{\partial x^b}(x) & \text{if } \frac{\partial V^w}{\partial x^b}(x) \neq 0, \\ -\infty & \text{if } \frac{\partial V^w}{\partial x^b}(x) = 0. \end{cases}$$

The first case of this definition is justified by Proposition B.1(ii) when $x^b \in \text{int}(X^b_\circ)$, and by the continuity if the partial derivatives when it is not. Note that this defines $s^w(x)$ even when $x$ is an isolated point in $N^w$. The second case of this definition is justified by Proposition B.1(iii) and (iv); the former says that the slope in this case is infinite, and the second, which says that points from this case are passed from northwest to southeast, justifies defining the slope to be $-\infty$ at these points.

The analogue of Proposition B.1 (with the coordinates reversed) holds for the blacks’ nullcline $N^b$. It (and condition (5)) justifies describing its slope in the original $(x^w, x^b)$ coordinates by the function $s^b : N^b \rightarrow (-\infty, \infty) \cup \{\pm \infty\}$, where

$$s^b(x) = \begin{cases} -\frac{\partial V^b}{\partial x^w}(x) / \frac{\partial V^b}{\partial x^b}(x) & \text{if } \frac{\partial V^b}{\partial x^b}(x) < 0, \\ \pm \infty & \text{if } \frac{\partial V^b}{\partial x^b}(x) = 0. \end{cases}$$

We use $\pm \infty$ for the infinite slope in the second case because it is an instance of part (ii) of the proposition, not part (iii); in particular, no analogue of part (iv) of the proposition applies here.
B.2 Proof of Theorem 4.3

Let \( x_\star \in N^w \cap N^b \) be an equilibrium of (3), and write

\[
DV(x_\star) = \begin{pmatrix}
\frac{\partial V^w}{\partial x^w}(x_\star) & \frac{\partial V^w}{\partial x^b}(x_\star) \\
\frac{\partial V^b}{\partial x^w}(x_\star) & \frac{\partial V^b}{\partial x^b}(x_\star)
\end{pmatrix} = \begin{pmatrix} a & b \\
c & d \end{pmatrix}.
\]

By equations (14), (15), (5), and (6), \(-\infty \leq s^w(x_\star) < s^b(x_\star) \leq 0\) if and only if

\[
a < 0 \text{ and } d < 0 \text{ and } [-\frac{a}{b} < -\frac{c}{d} \text{ or } b = 0].
\]

To prove Theorem 4.3, we need to show if \( x_\star \) is a hyperbolic rest point of (3), then it is asymptotically stable if and only if (16) holds.

We evaluate the stability of \( x_\star \) under (3) by evaluating the of \( DV(x_\star) \). These can be expressed using the well-known formula

\[
\lambda_\pm(x_\star) = \frac{1}{2} \left( T \pm \sqrt{T^2 - 4D} \right),
\]

where \( T = \text{tr}(DV(x_\star)) \) and \( D = \text{det}(DV(x_\star)) \). Writing out (17) explicitly and rearranging, we obtain

\[
\lambda_\pm(x_\star) = \frac{1}{2} \left( (a + d) \pm \sqrt{(a + d)^2 - 4(ad - bc)} \right)
\]

\[
= \frac{1}{2} \left( (a + d) \pm \sqrt{(a - d)^2 + 4bc} \right).
\]

Since \( b \) and \( c \) are nonpositive by (5), the radicand in (18) is nonnegative, and so \( \lambda_\pm(x_\star) \) are real. It then follows from (17) that \( x_\star \) is hyperbolic if and only if \( D \neq 0 \). Considering the cases where \( T \geq 0 \) and \( T < 0 \) separately shows that \( x_\star \) is an unstable saddle when \( D < 0 \). Finally, if \( D > 0 \), then (5) implies that \( T \) cannot equal 0; \( x_\star \) is a source (and hence unstable) if \( T > 0 \), and a sink (and hence stable) if \( T < 0 \). Thus if \( x_\star \) is hyperbolic, it is stable if and only if \( D > 0 \) and \( T < 0 \).

In Table B.2, we partition all specifications of \( DV(x_\star) \) that are feasible under (5) and (6) into 16 cases. In each case, we determine the sign of the determinant \( D = \text{det}(DV(x_\star)) \), the trace \( T = \text{tr}(DV(x_\star)) \), or both. Evidently, the only cases consistent with stability are the four in the southwest quadrant. For three of these stability is automatic; for the fourth, in which \( a, b, c, d \in (\infty, 0) \), stability requires that \( D = ad - bc > 0 \), or equivalently \( -\frac{a}{b} < -\frac{c}{d} \). It is easy to verify that these requirements are equivalent to (16). This completes the proof of Theorem 4.3.
\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
$c < 0, d > 0$ & $D < 0, T \geq 0$ & $D < 0, T \geq 0$ & $D < 0, T > 0$ & $D \geq 0, T > 0$ \\
$(s^b \in (0, +\infty))$ & & & & \\
\hline
$c < 0, d = 0$ & $D = 0, T < 0$ & $D < 0, T < 0$ & $D < 0, T = 0$ & $D < 0, T > 0$ \\
$(s^b = \pm\infty)$ & & & & \\
\hline
$c < 0, d < 0$ & $D > 0, T < 0$ & $D \geq 0, T < 0$ & $D < 0, T < 0$ & $D < 0, T \geq 0$ \\
$(s^b \in (-\infty, 0))$ & & & & \\
\hline
$c = 0, d < 0$ & $D > 0, T < 0$ & $D > 0, T < 0$ & $D = 0, T < 0$ & $D < 0, T \geq 0$ \\
$(s^b = 0)$ & & & & \\
\hline
$\begin{array}{c}
b = 0, a < 0 \\
(s^w = -\infty)
\end{array}$ & $b < 0, a < 0$ & $b < 0, a = 0$ & $b < 0, a > 0$ & \\
$(s^w \in (-\infty, 0))$ & $(s^w = 0)$ & $(s^w \in (0, +\infty))$ & & \\
\hline
\end{tabular}
\caption{Evaluating stability of $x_*$ for all $DV(x_*)$ feasible under (5) and (6).}
\end{table}

B.3 A Nonhyperbolic Example

Example B.2. Let $m^p = m^p_0 + m^p_\infty = 100 + .1$ for $p \in \{w, b\}$, and let the distributions of tolerances satisfy

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

Figure 9 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. §

B.4 Proof of Theorem 4.5

Since $G^w(\frac{x^b}{m^w_\infty}) = m^w_\infty$, we know that

$$G^w(\frac{x^b}{m^w_\infty}) = m^w_\infty \text{ for all } x_b > x^b,$$

and since $G^b(\frac{m^w_\infty}{x^b}) > x^b$ and $G^b(\frac{m^w_\infty}{m^b}) \leq m^b$, the intermediate value theorem implies that there is an $x^*_b \in (\underline{x^b}, m^b]$ with $G^b(\frac{m^w_\infty}{x^*_b}) = x^b$. Thus $(m^w_\infty, x^*_b)$ is an aggregate equilibrium. For the
second claim, note that differentiating (19) yields $f^w(x^b/m^w_{\infty}) = 0$ for all $x_b > x^b$. Thus

$$
(20) \quad \lambda(m^w_{\infty}, x^b) = \frac{f^b(m^w_{\infty}) m^w_{\infty}}{(x^b)^2} - 1 = \frac{d}{dx^b} \left( G^b(m^w_{\infty}) - x^b \right) \quad \text{for all } x_b > x^b.
$$

By continuity, the function $x^b \mapsto G^b(m^w_{\infty})$ crosses the identity function from above at $x^b_\star = \min\{x^b > x^b : G^b(m^w_{\infty}) = x^b_\star\}$, so $\frac{d}{dx^b}(G^b(m^w_{\infty}) - x^b_\star) \leq 0$. Thus by equation (20) and the hyperbolicity assumption, $\lambda(m^w_{\infty}, x^b_\star) < 0$, implying that $(m^w_{\infty}, x^b_\star)$ is asymptotically stable.

References


