Dynamic models of residential segregation: An analytical solution

Sébastian Grauwin [a,b,c], Florence Goffette-Nagot [a,d,*], Pablo Jensen [a,b,c,e]

[a] Université de Lyon, Lyon, F-69007, France
[b] Institut rhônalpin des systèmes complexes, IXXI, Lyon, F-69007, France
[c] ENS de Lyon and CNRS, Laboratoire de Physique, UMR 5672, Lyon, F-69007, France
[d] CNRS, GATE Lyon-St Etienne, UMR 5824, Ecully, F-69130, France
[e] CNRS, Laboratoire d'Économie des Transports (LET), UMR 5593, Lyon, F-69363, France

ARTICLE INFO

ABSTRACT

We propose an analytical solution to a Schelling segregation model for a relatively broad range of utility functions. Using evolutionary game theory, we provide existence conditions for a potential function, which characterizes the global configuration of the city and is maximized in the stationary state. We use this potential function to analyze the outcome of the model for three utility functions corresponding to different degrees of preference for mixed neighborhoods: (i) we show that linear utility functions is the only case where the potential function is proportional to collective utility, the latter being therefore maximized in stationary configurations; (ii) Schelling's original utility function is shown to drive segregation at the expense of collective utility; (iii) if agents have a strict preference for mixed neighborhoods but also prefer to be in the majority versus the minority, the model converges to perfectly segregated configurations, which clearly diverge from the social optimum. Departing from the existing literature, these conclusions are based on analytical results which open the way to analysis of many preference structures. Since our model is based on bounded rather than continuous neighborhoods as in Schelling's original model, we discuss the differences generated by the bounded- and continuous-neighborhood definitions and show that, in the case of the continuous neighborhood, a potential function exists if and only if the utility functions are linear. A side result is that our analysis builds a bridge between Schelling's model and the Duncan and Duncan segregation index.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Ethnic and immigrant residential segregation is a striking feature of most Western cities. Extensive views of segregation patterns in the U.S. are provided in Cutler et al. (2008), Iceland and Scopilliti (2008) and Reardon et al. (2008). Cutler et al. (2008) examine a range of potential determinants of immigrant segregation, including their cultural traits and nationalist sentiments among U.S. natives. Card et al. (2008) findings for racial segregation in the period 1970–2000 show evidence of tipping-like behaviors: an increase in the minority share in a neighborhood above a certain threshold leads to a further decrease in the native (in their case white) population. This analysis is one of the first to provide clear empirical evidence of non-linear dynamic aggregate behaviors predicted by social interaction models. According to these results, the utility of the white population in a mixed neighborhood seems to exhibit a sharp decrease beyond a certain minority share. The authors also provide evidence

of a direct link between white's attitudes toward minority members and aggregate configurations, as measured by the location of the tipping point. However, the theoretical relationship between individual preferences and aggregate configurations has not been fully explored.

Schelling provided an early contribution, proposing a model to formalize the aggregate consequences of individual preferences related to the social environment (Schelling, 1969, 1971, 1978). The two basic components of Schelling's "Dynamic models of segregation" (1971) are an individual utility function that entirely determines the level of satisfaction enjoyed by an agent in a location, and a dynamical rule that drives agents' location changes and, therefore, the way that the configuration of the city evolves. Using an inductive approach, Schelling shows that if the preferences considered are such that an environment of more than 50% of own-group agents is highly preferred to a less than 50% of own-group environment, then the equilibrium configuration exhibits high levels of segregation, although there is no preference for segregation *per se*. Schelling's (1971) paper is a seminal work in relation to his seemingly paradoxical finding that mild individual preferences for own-group neighbors lead to complete segregation at the global level. However, on reflection, we realize that, given the highly asymmetrical utility function used in this model, it could hardly lead to an integrated environment. Nevertheless, later

* Corresponding author at: GATE, 93 chemin des Mouilles, F-69130 Ecully, France. Tel.: +33 472866078; fax: +33 472866090.
E-mail addresses: sebastian.grauwin@ens-lyon.fr (S. Grauwin), goffette-nagot@gate.cnrs.fr (F. Goffette-Nagot), Pablo.Jensen@ens-lyon.fr (P. Jensen).

research shows that even a peaked utility function, that is, a function achieving its maximum for a perfectly mixed environment, can lead to a fully segregated equilibrium as soon as this function is asymmetric (Zhang, 2004b; Pancs and Vriend, 2007; Barr and Tassier, 2008).

The realism of Schelling's model can be criticized on the grounds that it ignores institutional causes of segregation, income effects, and the social structure of the city. Nevertheless, the model is popular for modeling social systems to show the unintended macro-level consequences of individual behavior, and Schelling's 1971 paper has become the most widely cited of his contributions (more than 460 citations to it on June 10, 2010). For many years there were relatively few citations to it, but since 2003 there have been around 40 citations per year, demonstrating the renewed interest in Schelling's model. It is interesting also that it is cited in widely different fields: economics and sociology are responsible for the highest number (40%) followed by computer science, mathematics and physics which together account for 24% of the citations to this paper. This substantial scientific activity has led to new insights: interpretations of the emergence of segregation patterns as the result of a coordination problem (Zhang, 2004a,b); a physical analog of Schelling's model (Vinkovic and Kirman, 2006); evidence of the robustness of Schelling's results with respect to different definitions of individual utilities and/or environments (Pancs and Vriend, 2007; Fagiolo et al., 2007); the impact of heterogeneous agents and public policies (O'Sullivan, 2009); the exploration of tipping behaviors (Zhang, 2011).[1]

Attempts to solve Schelling's model analytically include Dokumaci and Sandholm (2007); Mobius and Rosenblat (2000); Pollicott and Weiss (2001); Pancs and Vriend (2007); Zhang (2004a,b, 2011). Zhang (2004a,b, 2011)'s contributions are the most successful of these so far. Zhang proposes variations to Schelling's model which he analyzes formally using the concept of potential function developed in evolutionary game theory. Zhang (2004a) considers a model with vacant cells and linear utility functions. Zhang (2004b, 2011) use an asymmetric peaked utility function in a model with no vacant cells. The latter raises the issue of individual rationality, since it is assumed that, for individual moves to occur, two agents must coordinate and agree to exchange locations. This analysis, then, departs from Schelling's original framework. Also, the three contributions referred to above cover only two specific utility functions.

In the present paper, we build on Zhang (2004a,b, 2011). We transpose Schelling's model to an evolutionary game theory context with the aim of characterizing the equilibrium segregation level by means of a potential function. In contrast to Zhang (2004b, 2011), we consider bounded neighborhoods, i.e. blocks where all agents share the same neighbors. In this context, we show that a potential function of the model exists if and only if the utility functions are such that the externalities generated by one type of agent are symmetric to those generated by the other type of agent or in the limit case where there are no vacant cells. Under this condition, we can identify a general form of the potential function and can predict the global pattern emerging from different utility functions, which, to the best of our knowledge, has not previously been done. The main property of this potential function is that it reflects both the macro and micro scale. On the one hand, this aggregate function depends only on the number of agents of each type in each block. On the other hand, it keeps tracks of the individual level since it corresponds to the sum of the individual utility changes generated by individual moves.

We use this potential function to characterize the segregation level of the stationary configurations of the model for different utility functions, representing different degrees of preference for mixed environments. We examine successively: (i) linear utility functions, with a continuous preference for segregated environments; (ii) Schelling's original utility function in which there is a mild preference for a

mixed environment; and (iii) asymmetric peaked utility functions, according to which agents clearly exhibit a preference for a mixed environment. We show that there is no divergence between individual moves and social welfare with increasing linear utility functions, although segregation prevails in stationary configurations. We show also that even with the strongest preference for a mixed environment – in the asymmetric peaked utility function case – the model yields segregated stationary configurations. In this case, too, the divergence between the stationary segregation level and the optimal segregation level is highest. These results complement those obtained by Zhang (2004b, 2011) and Pancs and Vriend (2007).

In summary, our work provides a relatively general solution to Schelling's model with bounded neighborhoods, that encompasses previous work on this model and opens the way to analysis of many structures of preferences, for instance, those based on empirical findings concerning racial preferences. We show a few simulation results by way of illustration.

This paper is organized as follows. The model features are presented in Section 2. Section 3 defines the potential function concept and states our main result. The potential function is applied in Section 4 to study the stationary configurations obtained for three different utility functions. In one of these cases, we are able to build a connection with a commonly used segregation measure. In Section 5, we demonstrate the supplementary result that a potential function exists with continuous neighborhoods if and only if the utility functions are linear, and we discuss the differences between the bounded and continuous neighborhood cases. We also consider agents' preferences for local amenities and analyze a case where a tax on segregative behaviors is introduced.

2. A general dynamic model of segregation

2.1. The city and the agents

Our artificial city is a two-dimensional $N \times N$ square lattice with periodic boundary conditions, i.e. a torus containing N^2 cells. Each cell corresponds to a dwelling (residential unit) and all are of equal quality. We suppose that a certain characteristic divides the population of this city into two groups of households, which we will refer to as red and green agents. Each location may be occupied by a red agent, a green agent, or be vacant. We denote the number of vacant cells as N_V, and the number of respectively red and green agents as N_R and N_G. Thus the parameter N controls the size of the city, the parameter $v = N_V/N^2$ its vacancy rate, and the fraction $n_R = N_R/(N_R + N_G)$ its composition.

We define a *state x* of the city as a N^2-vector, each element of this vector labeling a cell of the $N \times N$ lattice. Each state x thus represents a specific configuration of the city. X is the set of all possible configurations, the demographic parameters (N, v, n_R) being fixed.

2.2. Neighborhoods

Since Schelling (1969)'s work, two ways of conceiving the neighborhood of an agent have been developed and used in analytical and simulation models.

Bounded neighborhood models (Fig. 1a) describe cities divided into geographical units within which all agents are connected. The neighborhood of an agent therefore is composed entirely and exclusively of the locations in the same geographical unit as his own. In the following, when we refer to a bounded neighborhood model, we assume implicitly that the city is divided into a set $\mathcal{Q}$ of blocks, each of which contains $H + 1$ locations, where H is a fixed integer that corresponds to the number of locations in an agent's neighborhood (hence, the relation $|\mathcal{Q}|(H + 1) = N^2$ must hold). Obviously, the description of the city as a lattice with periodic boundary conditions is unnecessary in this case. Note that since some locations remain empty, the size H of the neighborhood of an agent can also be interpreted as the maximum number of neighbors an agent can have. For a given configuration $x \in X$ of the city, we use $R_q(x)$ and $G_q(x)$

---

[1] See Clark and Fossett (2008) for a literature review.
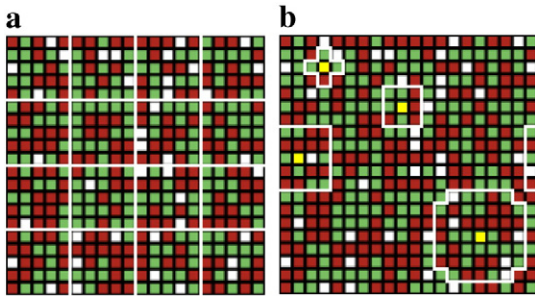
Fig. 1. Different forms of neighborhood. Red, green and white squares denote respectively red agents, green agents and vacant cells. a. Example of a bounded neighborhood in which the city is divided in square blocks containing $H+1=25$ cells/locations; b. In the case of a continuous neighborhood description, the neighborhood of an agent corresponds to his H nearest cells/locations. Around the agents marked in yellow, we enlightened by the white frontiers a $H=4$, a $H=8$, a $H=24$ and a $H=44$ continuous neighborhood. [If you printed this document in black and white, the red and green squares should appear respectively in dark gray and soft gray.]

respectively to denote the number of red and green agents that live inside the block $q \in \mathcal{Q}$. Taking into account that some locations of each block may remain empty, the $\{R_q\}$ and the $\{G_q\}$ must thus verify:

$$\sum_q R_q = N_R \tag{2.1}$$

$$\sum_q G_q = N_G \tag{2.2}$$

$$\left(R_q, G_q\right) \in E_{H+1} \equiv \{(R,G), 0 \leq R+G \leq H+1\}. \tag{2.3}$$

Continuous neighborhood models (Fig. 1b) describe cities where the neighborhoods do not correspond to zoning at the city level, but are centered on the local perception of each agent. In a continuous neighborhood description, we assume that the neighborhood of an agent is composed of the H nearest locations surrounding him/her. The $H=4$ "Von Neumann neighborhood" and the $H=8$ "Moore neighborhood", displayed among other examples in Fig. 1b are the most commonly used in agent-based computational models.

As we demonstrate in Section 5.1, a potential function exists in the continuous neighborhood case only with very specific utility functions. In the following, unless otherwise stated, we consider a bounded neighborhood setup. There are actually no arguments in favor of bounded or continuous neighborhoods as far as the realism of the assumptions is concerned. Bounded neighborhoods can be thought of as reproducing the effects of the administrative divisions of real cities such as census areas or school districts. Nevertheless, bounded neighborhoods have the drawback that the spatial arrangement of the neighborhoods is not taken into account: each neighborhood is considered independently without any connection to other neighborhoods. In Section 5.1 we present a formal analysis showing how the connections between neighborhoods, according to which agents close to neighborhood boundaries consider the composition of the contiguous neighborhood, can be introduced in our framework. Section 5.2 presents some simulation results illustrating the impact of the neighborhood description on the forms of segregation at city level.

2.3. Agent's utility function

Each agent has a utility level which depends only on his neighborhood composition. Let us consider an agent whose neighborhood is composed of R red agents, G green agents and V vacant cells. Since $R+G+V=H$, we need two independent parameters to describe the composition of the agent's neighborhood. In general, we can write

the utility of an agent, for example, as a function of R and G. Like most models in the literature, for simplicity we assume that agents in a same group share the same utility function.[2]

Without any loss of generality, we can write the utility of an agent as:

$u = u_R(R,G)$ for a red agent with R red and G green neighbors,
$u = u_G(R,G)$ for a green agent with R red and G green neighbors.

The utility of an agent is thus a function of $E_H \to R$. More specifically, in the case of a bounded neighborhood description, we have:

$u = u_R\left(R_q-1, G_q\right)$ for a red agent living in block q,
$u = u_G\left(R_q, G_q-1\right)$ for a green agent living in block q.

In order to facilitate the comparison of different utility functions, utility in the examples presented below is such that a zero utility level denotes complete dissatisfaction of the agent and a utility of 1 denotes complete satisfaction.

We introduce a notation in order to characterize the level of utility on the global (city) scale:

$$U(x) = \sum_k u_k \tag{2.4}$$

where u_k is the utility of agent k and $U(x)$ denotes the collective utility of a configuration x.

2.4. A behavioral rule: the logit dynamical rule

The core of dynamic segregation models is that agents are given opportunities to move in order to increase their individual utility. Once the static description of the model is specified, we need to add a dynamic rule that governs these moves. In the following, the city configuration evolves according to an iterative process. In the first iteration, an initial configuration is randomly chosen. In each successive iteration, one agent and one vacant cell are chosen at random. The agent chosen then chooses to move to that vacant cell with a probability $Pr\{move\}$ that depends on the utility gain Δu he would achieve if he was to move, as follows[3,4]:

$$Pr\{move\} = \frac{1}{1 + e^{-\Delta u/T}} \tag{2.5}$$

where $T>0$ is a fixed parameter.

Eq. (2.5) represents a logit choice function as developed in (McFadden, 1974) as the outcome of a random-utility model. In such a model, it may be that an agent makes a utility-decreasing move, either because she is making a mistake or because of lack of information. The scalar T is used to determine the relative importance of the random part with respect to the deterministic part of the random utility function. The probability of an agent to make a utility-decreasing move decreases as $T \to 0$ and the described rule thus converges to the non-strict best response rule. For any finite $T>0$, agents choose non-best replies with a non-zero probability, but actions that yield smaller payoffs are chosen with smaller probability. This kind of perturbed best-response dynamics is developed in Young (1998) in the context of evolutionary games.

---

[2] See O'Sullivan (2009) for a treatment with heterogeneous agents.
[3] Instead of assuming, as Schelling did, that the agents move to the nearest satisfactory position (the idea being that the cost of moving increases with distance), we suppose here that the distance between the current and envisaged locations of an agent does not matter in his decision about whether to move or not.
[4] This rule is also adopted in Zhang (2004a). Another choice would be to choose an agent at random and to compute her probability to move based on the utility of all vacant cells and her current cell as in Pancs and Vriend (2007). Both formulations give the same results.

Here, the neighborhood composition is supposed to be the main determinant of agents' actions: we restrict our analysis to the case of low values of T.

Compared to a best-response behavior rule (where the agents move if and only if they strictly improve their utility), the logit rule allows for some fluidity in the model in the sense that blocked states are avoided (Vinkovic and Kirman, 2006). Also, it provides a strong analytical framework for the Schelling model. Obviously, it implies that the probability that the state at the tth iteration x^t is equal to a given state x depends only on the state at the previous iteration x^{t-1}:

$$Pr\left(x^t = x \middle| x^{t-1}, \dots, x^1, x^0\right) = Pr\left(x^t = x \middle| x^{t-1}\right). \tag{2.6}$$

The dynamic rule thus yields a finite Markov process.

It is then straightforward to figure out that the Markov chain describing our system is irreducible (since $T>0$ each imaginable move has a non-zero probability of occurring and it is thus possible to get to any state, from any state), aperiodic (given any state x and any integer k, there is a non-zero probability that we return to state x in a multiple of k iterations) and recurrent (given that we start in state x, the probability that we will never return to x is 0). These three properties ensure that the probability to observe any state x after t iterations starting from a state y converges toward a fixed limit independent of the starting state y as $t \to \infty$.

In other words, for each set of parameters and dynamic rule, there exists a stationary distribution

$$\Pi : x \in X \to \Pi(x) \in [0, 1], \sum_{x \in X} \Pi(x) = 1 \tag{2.7}$$

which gives the probability with which each state x will be observed in the long run.

Clearly, for $T \to \infty$, the randomness introduced in the dynamical rule prevails and the stationary distribution is just a constant. Similarly, for any finite $T>0$, our dynamic system (the city) evolves toward an attractor composed of a subset A of X. It follows that any measure $\mathcal{M}$ — such as the global utility U — performed on the state space X will in the long run fluctuate around a mean value $\mathcal{M}_\infty = \Sigma_{x \in A} \Pi(x) \mathcal{M}(x)$. These mean values may depend on the intensity of the noise T, but the amplitude of the fluctuations decreases as $T \to 0$.

In the following, we refer to two states x and y as *immediately communicating states* (ICS) if we can switch from state x to state y by moving one single agent. We denote by $\Delta_{xy}u$ the variation of utility of this agent induced by this particular move and P^T_{xy} the probability to be in state y at a given iteration if the system was in state x at the previous iteration.

According to the dynamic rule presented above, we have:

$$P^T_{xy} = \gamma\left(1 + e^{-\Delta_{xy}u/T}\right)^{-1} \text{ if } x \text{ and } y \text{ are ICS} \tag{2.8}$$

$$P^T_{xy} = 0 \text{ if } x \text{ and } y \text{ are not ICS} \tag{2.9}$$

where the parameter $\gamma = 1/(N_V(N_R + N_G)) = 1/(v(1-v)N^4)$ takes account of the probability to pick the right agent and the right vacant cell for the system to pass from x to y. P^T thus corresponds to the probability transition matrix for a fixed T and the stationary distribution Π by definition is the unique normalized function defined on X verifying for all $x \in X$:

$$\sum_y P^T_{yx} \Pi(y) = \Pi(x). \tag{2.10}$$

3. Model solving with a potential function

3.1. Definitions and properties

Following Zhang (2004a,b), we place our model in an evolutionary game theory context and solve it using the concept of potential function. The concept of potential function in game theory was proposed by Monderer and Shapley (1996). A game is said to be a potential game if the incentive of all players to choose a strategy can be expressed in one global function, the potential function. In our context, the definition of a potential function takes the rather simple form:

Definition 1. Let $\mathcal{F} : x \in X \to \mathcal{F}(x) \in \mathbb{R}$ be an aggregate function describing each of the potential configurations. By definition, $\mathcal{F}$ will be a (cardinal) potential function of our model if and only if each gain in utility Δu of a moving agent is equal to the variation $\Delta \mathcal{F}$ that is induced on the global level by the move of this agent.[5] A cardinal potential function will thus verify: $\mathcal{F}(y) - \mathcal{F}(x) = \Delta_{xy}u$ with $\Delta_{xy}u$ previously defined (Section 2.4).

The main property of a potential function is to link the variation of a purely individual function (the utility of the moving agent) to the variation of a global function defined on the space X of all possible configurations and characterizing the city configuration. The ensuing lemma highlights the value of the potential function as an analytical tool.

Lemma 1. If $\mathcal{F}$ is a potential function of the system, then the stationary distribution Π is such that for any configuration x:

$$\Pi(x) = \frac{e^{\mathcal{F}(x)/T}}{\sum_{z \in X} e^{\mathcal{F}(z)/T}} \tag{3.1}$$

It follows that for $T \to 0$, the stationary configurations are those that maximize $\mathcal{F}$.

Proof. The following proof follows the classical argument presented in Young (1998).

Let π be the function defined as $\pi : X \to [0, 1]; x \to \pi(x) = e^{\mathcal{F}(x)/T} / \sum_z e^{\mathcal{F}(z)/T}$. The first step in the proof consists of checking that π satisfies the *detailed balance condition*:

$$\pi(x)P^T_{xy} = \pi(y)P^T_{yx} \tag{3.2}$$

If x and y are two different and not communicating states, equality 3.2 is trivially satisfied since in this case $P^T_{xy} = P^T_{yx} = 0$. If $x = y$, the detailed balance condition is also trivially verified. In the case where $x \neq y$ and x and y are two communicating states, we have:

$$\begin{aligned} \pi(x)P^T_{xy} &= \pi(x)\gamma \frac{1}{1 + e^{-\Delta_{xy}u/T}} = \pi(x)\gamma \frac{1}{1 + e^{-(\mathcal{F}(y)-\mathcal{F}(x))/T}} \\ &= \pi(x)\gamma \frac{e^{\mathcal{F}(y)/T}}{e^{\mathcal{F}(x)/T} + e^{\mathcal{F}(y)/T}} = \pi(y)\gamma \frac{e^{\mathcal{F}(x)/T}}{e^{\mathcal{F}(x)/T} + e^{\mathcal{F}(y)/T}} \\ &= \pi(y)\gamma \frac{1}{1 + e^{-(\mathcal{F}(x)-\mathcal{F}(y))/T}} = \pi(y)\gamma \frac{1}{1 + e^{-\Delta_{yx}u/T}} = \pi(y)P^T_{yx} \end{aligned}$$

recalling that $\gamma = 1/(N_V(N_R + N_G)) = 1/(v(1-v)N^4)$.

Hence the detailed balance condition is always verified and

$$\sum_{x \in X} \pi(x)P^T_{xy} = \sum_{x \in X} \pi(y)P^T_{yx} = \pi(y) \sum_{x \in X} P^T_{yx} = \pi(y) \cdot 1 = \pi(y), \tag{3.3}$$

---

[5] Games can be either ordinal or cardinal potential games. In cardinal games, the difference in individual payoffs for each player from individually choosing his or her strategy *ceteris paribus* has to be of the same value as the corresponding difference in value for the potential function. In ordinal games, only the signs of the differences have to be the same.

which defines π as a stationary distribution of the process. Because the Markov chain is finite and irreducible, it has a unique stationary distribution. Hence, for each state x, $\Pi(x) = \pi(x) = e^{\mathcal{F}(x)/T}/\Sigma_z e^{\mathcal{F}(z)/T}$.

Define then X_F as the subset of X of the states that maximize the potential function $\mathcal{F}$:

$$X_F = \{y, \forall x \in X \, \mathcal{F}(y) \geq \mathcal{F}(x)\}. \tag{3.4}$$

The second part of the lemma can now be proved as follows: for two states x and y of X_F, we will have $\mathcal{F}(x) = \mathcal{F}(y)$ and therefore $\Pi(x)/\Pi(y) = e^{[\mathcal{F}(x)-\mathcal{F}(y)]/T} = 1$, which means that two states that strictly maximize $\mathcal{F}$ are observed with the same probability in the long run; for two states $x \in X \backslash X_F$ and $y \in X_F$, we have $\mathcal{F}(x) - \mathcal{F}(y) \leq 0$ and therefore $\Pi(x)/\Pi(y) = e^{[\mathcal{F}(x)-\mathcal{F}(y)]/T} \to 0$ as $T \to 0$. This means that for $T \to 0$, the probability to observe a state that does not maximize the potential function $\mathcal{F}$ becomes in the long run infinitesimally small. □

The potential function is hence a very powerful analytical tool. First, it establishes a relation between individual changes in utility and a global characteristic of the city configuration. Second, because stationary configurations can be defined as those maximizing the potential function for low noise levels, the existence of a potential function allows analytical qualification of the stationary configurations. The fact that knowing Δu is sufficient to say something on the global level is highly non-trivial since, in particular, there is no way to determine the externalities produced by the move of an agent – i.e. the variation of the utility of the agent's former and new neighbors – based only on knowing the utility variation of the moving agent. Note that a low level of T is required for the maximum of the potential function to be achieved at stationary configurations.[6] However, T has to remain strictly positive to avoid blocked states.

3.2. Main result: existence of a potential function

It is possible, using the potential function, to examine analytically the outcome of the model for different utility functions, representing different degrees of preference for mixed environments. To do so, we need to address two questions. First: given any pair of utility functions (u_R, u_G), does a potential function exist and can we compute it? And, reciprocally, given a potential function, can we find a pair of utility functions (u_R, u_G) that can be translated into this specific potential function?

We show in the following that in the context of bounded neighborhoods, one can achieve an analytical solution of the model under a rather mild condition. Let us begin with some definitions.

Definition 2. Let $\mathbb{U}$ be the set of pairs of utility functions (u_R, u_G) that verify, for all $(R,G) \in E_H$, the following condition:

$$u_R(R, G) - u_R(R, G+1) = u_G(R, G) - u_G(R+1, G). \tag{3.5}$$

Condition 3.5 imposes that if a block contains $R+1$ red agents and $G+1$ green agents, the utility gain a red agent would achieve if a green agent left must be the same as the utility gain a green agent would achieve if a red agent left. The results in the following apply to pairs of utility functions verifying this condition. The range of application of this condition is specified in Section 3.3.

Definition 3. Let $\mathbb{F}$ be the set of aggregate functions of the form $\mathcal{F}(x) = \Sigma_{q \in \mathcal{Q}} F(R_q, G_q)$, where F is an intermediate function defined on the set E_{H+1} of all possible numbers of red and green agents that can be present in a block.

The main result of this article is presented in the following proposition:

Proposition 1. Each aggregate function $\mathcal{F} \in \mathbb{F}: x \to \mathcal{F}(x) = \Sigma_{q \in \mathcal{Q}} F(R_q, G_q)$ is a potential function which corresponds to at least[7] one pair (u_R, u_G) of utility functions of $\mathbb{U}$ that can be expressed as:

$$\begin{cases} u_R(R, G) = F(R+1, G) - F(R, G) \\ u_G(R, G) = F(R, G+1) - F(R, G) \end{cases}. \tag{3.6}$$

Reciprocally, for each pair of utility functions (u_R, u_G) of $\mathbb{U}$, there exists one corresponding potential function $\mathcal{F}_{[u_R, uG]} \in \mathbb{F}$. This function can be expressed through the functional $F_{u_R, u_G}: E_{H+1} \to \mathbb{R}$ – such that $\mathcal{F}_{[u_R, u_G]}(x) = \Sigma_{q \in \mathcal{Q}} F_{[u_R, u_G]}(R_q, G_q)$ – which is defined for all $(R,G) \in E_{H+1}$ by [8]:

$$F_{[u_R, u_G]}(R, G) = \sum_{r=1}^{R} u_R(r-1, 0) + \sum_{g=1}^{G} u_G(R, g-1) \tag{3.7}$$

$$= \sum_{r=1}^{R} u_R(r-1, G) + \sum_{g=1}^{G} u_G(0, g-1). \tag{3.8}$$

Proof. See Appendix A. □

Proposition 1 states that it is always possible to define a function F at the neighborhood level corresponding to the variation of utility of a moving agent, but only a pair of utility functions verifying condition 3.5 allows this function to be path-independent and therefore uniquely defined for any given configuration. Reciprocally, for any pair of utility functions verifying condition 3.5, the game has a potential function that is maximized at stationary configurations, and this function is the sum of neighborhood-level intermediate components. As Eq. (3.7) shows, the intermediate component of the potential function corresponds to the sum of the utilities of the agents, arriving in succession, in the block. This sum is calculated starting from an empty block, agents being introduced one by one, first the red ones and then the green ones. As Eq. (3.8) shows, the same sum is obtained if green agents are introduced first and red agents after. Before giving a more general interpretation in Section 3.3, it is useful to provide the following corollary, aimed at showing that utility functions verifying condition 3.5 can be given a convenient formulation in which interactions between the two groups are expressed through the same function in the two utility functions, thus giving a more general formulation for the potential function.

Corollary 1. Any pair of utility functions (u_R, u_G) belonging to $\mathbb{U}$ can be written:

$$u_R(R, G) = \xi_R(R) + \sum_{g=0}^{G-1} \xi(R, g) \tag{3.9}$$

$$u_G(R, G) = \xi_G(G) + \sum_{r=0}^{R-1} \xi(r, G) \tag{3.10}$$

where ξ_R and ξ_G are arbitrary functions of $\{0, 1, .., H\} \to \mathbb{R}$ and ξ is an arbitrary function of $E_H \to \mathbb{R}$.

For each pair of utility functions (u_R, u_G) verifying 3.9 and 3.10, thanks to Eq. (3.7), we can rewrite the general form of the potential function $\mathcal{F}_{[u_R, u_G]}$ as:

$$\mathcal{F}(x) = const + \sum_q \left(\sum_{r=0}^{R_q-1} \xi_R(r) + \sum_{g=0}^{G_q-1} \xi_G(g) + \sum_{r=0}^{R_q-1} \sum_{g=0}^{G_q-1} \xi(r,g) \right) \tag{3.11}$$

---

[6] In the case of finite values of the noise level ($T>0$), it can be demonstrated using standard statistical physics tools, that the states that are the more likely to occur are those that maximize $F(x) + TS(x)$ where $S(x)$ is an entropy-like global function that takes account of the number of ways of locating R_q red agents and G_q green agents in each block q of the city (Grauwin et al., 2009a).

[7] Since the definition of the potential function is based only on its variation, $\mathcal{F}$ can be defined up to an additive constant. Similarly, a utility function can also be defined up to a constant. All the formulae in this insert are written according to the convention u $(0,0) = F(0,0) = F(0,1) = F(1,0) = 0$. For more details, see the proof in Appendix A.

[8] Our notation assumes that a sum is null whenever its upper bound of summation is inferior to its lower bound of summation.

Proof. For any pairs of utility functions (u_R, u_G), we can define ξ_R and ξ_G, two functions of $\{0, 1, ..., H\} \to R$ and ξ_{RG} and ξ_{GR}, two functions of $E_H \to \mathbb{R}$ by

$$\begin{cases} \xi_R(r) = u_R(r, 0) \\ \xi_G(g) = u_G(0, g) \end{cases} \begin{cases} \xi_{RG}(r, g) = u_R(r, g+1) - u_R(r, g) \\ \xi_{GR}(r, g) = u_G(r+1, g) - u_G(r, g) \end{cases}$$

for all $0 \le r \le H$ and $0 \le g \le H$. By definition, we can then write the utility functions as

$$u_R(R, G) = \xi_R(R) + \sum_{g=0}^{G-1} \xi_{RG}(R, g) \tag{3.12}$$

$$u_G(R, G) = \xi_G(G) + \sum_{r=0}^{R-1} \xi_{GR}(r, G) \tag{3.13}$$

for all $(R, G) \in E_H$.

The condition given by Eq. (3.5) is obviously equivalent to $\xi_{RG} = \xi_{RG} \equiv \xi$, which proves Corollary 1. $\square$

3.3. Interpretation

Here, we propose first an interpretation of condition 3.5 and then an interpretation of the form of the potential function. Proposition 1 ensures that a potential function exists for any pair of utility functions verifying condition 3.5. In its original form, this condition says that there is a symmetry in the externalities generated by green agents on red agents, and by red agents on green agents: starting from a given neighborhood composition, the variation in utility produced by the departure of an agent of the other type must be the same for both categories. This can be seen as rather limiting, as some real world situations do not conform to this condition. For instance, well-known surveys on white and black individuals' valuation of their preferred residential environment show that blacks favor integrated neighborhoods, while whites favor all-white neighborhoods (Farley et al., 1978; see Farley et al., 1997 for recent figures).

However, condition 3.5 covers more general types of preferences when the rate of vacant cells is low. Namely, at the limit of a very low vacancy rate, that is there are no vacant cells in most of the blocks, i.e. in these blocks the relation $R_q + G_q = H + 1$ holds. Hence, only one parameter is needed among (R_q, G_q, V_q) to define a utility function and all possible cases can be described if the agent's utility depends only on his number of similar neighbors. This can be done by taking $\xi \equiv 0$ in Eqs. (3.9) and (3.10), while keeping the functions ξ_R and ξ_G independent and free. In other words, in the limit of no vacant cells, each agent arriving in a neighborhood receives a utility that is fully determined by the number of like-neighbors. Therefore, the order in which the agents settle in the neighborhood does not matter and the condition for having a potential function holds. The set U hence describes all possible pairs of utility functions in the limit $v \to 0$. It follows also that condition 3.5 holds for all pairs of utility functions in situations where vacant cells are considered in the same way as unlike-color neighbors. Note also that condition 3.5 applies to the utility functions considered in Zhang (2004a), where preferences for neighborhoods are determined by the number of like-agents only and the symmetric effect of unlike-color neighbors on each type of agent emerges as the result of the determination of housing prices by density.

In its original form, the potential function $\mathcal{F}$ can be interpreted as the sum of the incentives of the agents (when they settled) to move into the neighborhood where they located. Indeed, if $x(t)$ denotes the state of the city at iteration t, then the potential can be rewritten as

$$\mathcal{F}(x(t)) - \mathcal{F}(x(0)) = \sum_{t'=1}^{t} \Delta_{x(t'-1)x(t')} u \tag{3.14}$$

where $\Delta_{x(t'-1)x(t')} u = 0$ by definition if no move happens at iteration t' and where we can take $\mathcal{F}(x(0)) = 0$ since the potential is defined up to a constant. Conversely, the potential function $\mathcal{F}$ can also be viewed as the minimum utility level each agent would require to agree to leave his/her neighborhood. As such, it represents, in the case $T \to 0$, the stability of the configuration x: the higher the potential function, the smaller the incentive for an agent to move.

To interpret the potential function further, note that condition 3.5 can also be written as:

$$u_R(R, G) + u_G(R+1, G) = u_G(R, G) + u_R(R, G+1) \tag{3.15}$$

which means that starting from any initial composition of a block, the sum of the utilities of a red agent and a green agent entering this block successively is the same whatever the order in which they enter. This expression stresses that, under condition 3.5, the value of function F in a given neighborhood q does not depend on the particular path of events that lead to the composition of this neighborhood. This is also particularly clear in the form of condition 3.5 given in corollary 1. It follows therefore that the potential function $\mathcal{F}$, which is the sum of the F intermediate functions, is independent of the particular order in which the agents arrived in the neighborhoods.

Hence, it is also possible to define $\mathcal{F}$ as the average over all the possible ways of ordering the agents, which is shown formally in Grauwin et al. (2010).[9]

4. Segregation for different levels of preferences for mixed environments

The main property of the potential function obtained in the previous section is that it reflects both the macro and micro scales. On the one hand, $\mathcal{F}$ is an aggregate function, defined at the city level, which depends only on the number R_q and G_q of red and green agents in each block. On the other hand, $\mathcal{F}$ also keeps tracks of the individual level since it corresponds to the sum of the utility differences generated by individual moves. When the stationary states are reached in the case $T \to 0$, $\mathcal{F}$ is maximized, which means that no agent can strictly improve its utility by moving. The potential function can now be used to assess the outcomes of our location model for different utility functions, representing different degrees of preference for mixed environments, as far as these functions verify condition 3.5. We examine successively (i) linear utility functions, with a continuous preference for segregated environments; (ii) Schelling's original utility function in which there is a mild preference for a mixed environment; and (iii) an asymmetric peaked utility function, according to which agents exhibit a strict preference for a mixed environment.[10]

4.1. Linear utility functions

We consider here utility functions that exhibit a monotonic effect of the number of same-color neighbors on utility, through linear utility functions. Zhang (2004a) proposes an analytical solution to a dynamic model of segregation with a linear utility function, and shows that the halved sum of individual utilities is a potential function of the game thus defined. In this section, we show that Proposition 1 produces similar results for all linear utility functions verifying condition 3.5 in the context of bounded neighborhoods.

---

[9] It is also shown in Grauwin et al. (2010) that the potential function of our model can be interpreted in terms of the Shapley value of a coalitional game.
[10] Refer to Grauwin et al. (2009b) for the study of other utility functions.

Suppose that u_R and u_G are expressed as:

$$u_R(R, G) = aR + bG$$
$$u_G(R, G) = bR + dG \qquad (4.1)$$

where a, b, d are constant parameters.[11]

It is easy to verify that this particular pair of utility functions verifies condition 3.5 and compute the corresponding potential function:

$$\mathcal{F}(x) = \frac{1}{2} \sum_q \left(aR_q(R_q - 1) + dG_q(G_q - 1) + 2bR_qG_q \right). \qquad (4.2)$$

We can rewrite this potential function as:

$$\mathcal{F}(x) = \left(b - \frac{a + d}{2} \right) \rho_{RG}(x) - \frac{a}{2} \rho_{RV}(x) - \frac{d}{2} \rho_{GV}(x) \qquad (4.3)$$

with:

$\rho_{RG} = \sum_q R_q G_q$ the number of red–green pairs of neighbors,

$\rho_{RV} = \sum_q R_q(H + 1 - R_q - G_q)$ the number of red–vacant pairs of neighbors and

$\rho_{GV} = \sum_q G_q(H + 1 - R_q - G_q)$ the number of green–vacant pairs neighbors.

This last form provides a convenient interpretation of the potential function. Putting aside at this point the last two terms, $\mathcal{F}$ is proportional to ρ_{RG}, this gives a measure of the relative contact between the two groups. Hence, the sign of the prefactor $b - (a + d)/2$ indicates whether mixed states (when positive) or segregated states (when negative) are obtained at the global level. Notice that the two groups do not need to both have strong preferences for like neighbors, for segregation to emerge. It is the average preference over the two groups that determines the level of segregation.

The terms proportional to ρ_{RV} and ρ_{GV} show that agents avoid the proximity of vacant cells when $a > 0$ and $d > 0$. All these insights from the study of the potential function can be checked by means of simulations (Fig. 2).

Turning now to the link between segregation of the stationary configurations and collective utility, it is useful, using Proposition 1, to write the potential function as follows[12]:

$$\mathcal{F}(x) = \frac{1}{2} \sum_q \left(aR_q(R_q - 1) + dG_q(G_q - 1) + 2bR_qG_q \right)$$
$$= \frac{1}{2} \sum_q \left(R_q u_R(R_q - 1, G_q) + G_q u_G(R_q, G_q - 1) \right) = \frac{1}{2} U(x).$$

With this choice of utility functions, the potential function is thus proportional to collective utility and therefore Lemma 1 ensures that, for low values of T, the stationary configuration maximizes collective utility. Reciprocally, we can verify (see proof in Appendix B) that if we want the potential function to be proportional to the collective utility, so that states that maximize the potential function also maximize collective utility, then the constant of proportionality is necessarily 0.5 and the pair of utility functions must take the form displayed in Eq. (4.1) (up to a constant).

To sum up, the linear utility functions as defined in Eq. (4.1) lead to segregated or mixed states depending on the values of the parameters. However, in all cases, there is no divergence between stationary

configurations and the optimum: these utility functions are such that utility-improving moves also improve collective utility.

4.2. Schelling utility function

Suppose that the agents compute their utility with a Schelling utility function, which is equal to 1 if the fraction of similar neighbors is superior or equal to 0.5, and equal to 0 otherwise. This utility function can be expressed in terms of the number of red and green neighbors as follows:

$$u_R(R, G) = \Theta(R - G) = \frac{1}{2}(1 + |R + 1 - G| - |R - G|) \qquad (4.4)$$

$$u_G(R, G) = \Theta(G - R) = \frac{1}{2}(1 + |R - 1 - G| - |R - G|)$$

where Θ is the Heaviside function defined by: $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ if $x \geq 0$.

It is easy to figure out that this particular pair of utility functions respects condition 3.5, and is therefore in the set $\mathbb{U}$. Indeed, the form of the symmetric externality produced by a new unlike-color neighbor is as follows:

$$u_R(R, G + 1) - u_R(R, G) = |R - G| - \frac{1}{2}(|R - G - 1| + |R - G + 1|) \qquad (4.5)$$

$$= u_G(R + 1, G) - u_G(R, G). \qquad (4.6)$$

It is possible to compute the potential function directly thanks to its interpretation. It is indeed the sum of the utilities of the agents introduced one by one in the city, and this sum is independent of the precise order in which agents are introduced. To do this for a given configuration $x \equiv \{R_q, G_q\}$, let us consider that in each block, we introduce first the agents in the majority (i.e. the red ones if $R_q > G_q$, the green ones if $G_q > R_q$, either the red or the green ones if $R_q = G_q$) and second the agents in the minority. Each agent of the majority has a utility of 1 as he/she settles in the city while each of the other minority agents has zero utility.[13] Hence it is straightforward to write the potential function as[14]:

$$\mathcal{F}(x) = const + \sum_{q \in \mathcal{Q}} \max(R_q, G_q)$$
$$= const + \sum_{q \in \mathcal{Q}} \frac{1}{2} \left(R_q + G_q + |R_q - G_q| \right)$$
$$= const' + \frac{1}{2} \sum_{q \in \mathcal{Q}} |R_q - G_q|.$$

The same expression can be found using the relation 3.7 (see Appendix C), the computation in this case is more formal than is presented here.

The reader recognizes a well-known expression, widely used in the literature on residential segregation. This potential function is a linear form of the Duncan and Duncan dissimilarity index, which in the case that the total numbers of red and green agents in the city are equal ($N_R = N_G = N$), is written as $D(x) = \frac{1}{2} \sum_q |R_q/N_R - G_q/N_G| = \frac{1}{2N} \sum_q |R_q - G_q|$ (Duncan and Duncan, 1955). To the best of our knowledge, an analytical connection between the "historical" works of Schelling and Duncan and Duncan on segregation has not been made before.

---

[11] Zhang (2004a)'s utility function corresponds to $b = d = -1$ and $a \geq -1$, the utility also includes a fixed income term which makes it positive. In Zhang (2004a)'s framework, the impact of unlike neighbors is not due to preferences, but to the impact of density on housing prices.

[12] Note that this result is similar to the one obtained in the continuous neighborhood case; see Section 5.1. See also Grauwin et al. (2009b).

[13] This example shows that to compute the potential function corresponding to a given pair (u_R, u_G) of utility functions, it is worth deciding in advance on a practical order for introducing the agents. The computation of $\mathcal{F}$ is indeed easier and more meaningful when agents are appropriately ordered.

[14] Note that in this particular example, we do not use the convention $u(0,0) = 0$. See Appendix C for details.
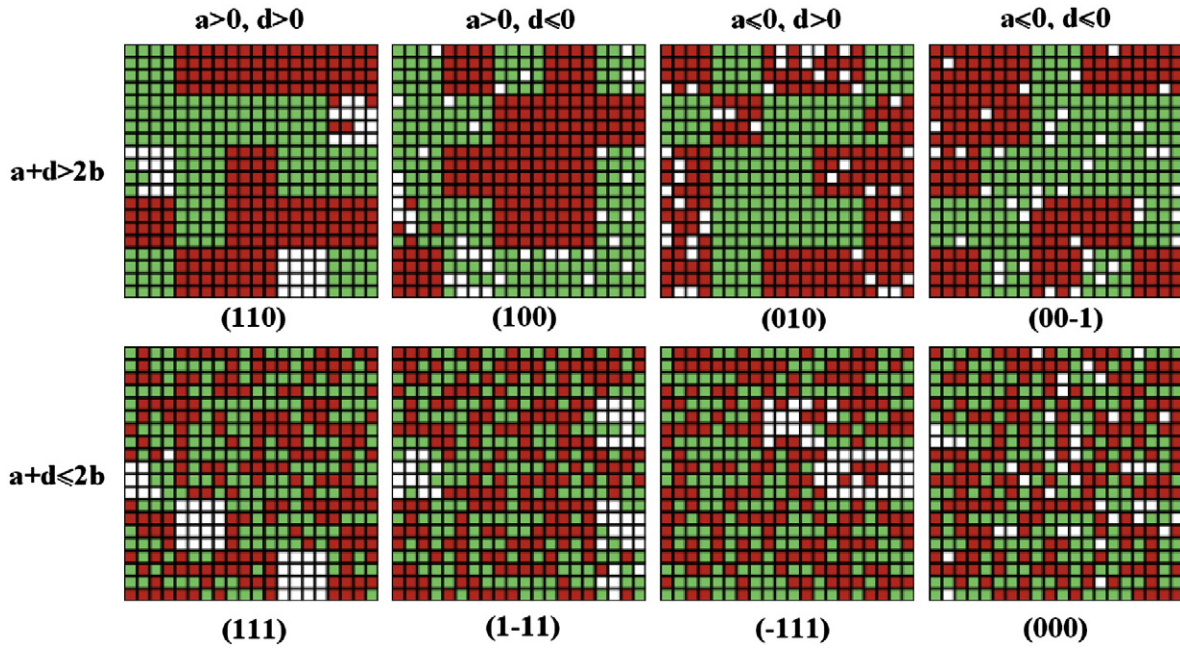
Fig. 2. Typical stationary configurations obtained by simulations for different values of (adb). Top panel: for $2b - (a+d) < 0$, the system evolves towards segregated configurations where red and green agents tends to live in different blocks. Bottom panel: for $2b - (a+d) > 0$, the system evolves towards mixed configurations where the number of red–green pairs of neighbors is maximized. From left to right: the sign of a and d controls the tendency of red and green agent to prefer to live in dense or uncrowded areas. The demographic parameters are ($N = 20, v = 10\%, n_R = 0.5$). Neighborhood size is fixed to $H + 1 = 16$ and the level of noise is $T = 0.1$.

It is worth investigating the link between the potential function and collective utility. The collective utility in a neighborhood q is:

$$
\begin{cases}
U_q &= \frac{1}{2}\left(R_q + G_q + \left|R_q - G_q\right|\right) \quad \text{if } R_q \neq G_q \\
U_q &= R_q + G_q \quad\quad\quad\quad\quad\quad\quad \text{if } R_q = G_q
\end{cases}
\tag{4.7}
$$

The potential function can therefore be written as:

$$
\mathcal{F}(x) = const + U(x) - \sum_{q \mid R_q = G_q} \frac{1}{2}\left(R_q + G_q\right)
\tag{4.8}
$$

This expression shows that the configuration that maximizes $\mathcal{F}$ does not correspond to the maximum collective utility. The divergence is due to the existence of perfectly mixed neighborhoods. More specifically, we compare two configurations differing in terms of the existence in configuration x_1, of two perfectly mixed neighborhoods with $K < (H+1)/2$ agents of each color, that are changed to segregated ones in configuration x_2 due to the exchange of two agents of different color. The difference in the potential function between the two configurations is due only to the change affecting these two neighborhoods. It is written as:

$$
\Delta\mathcal{F} = \mathcal{F}(x_2) - \mathcal{F}(x_1) = 2
\tag{4.9}
$$

because each of the two neighborhoods gained one agent of one color and lost one agent of the other color. The difference in collective utility consists of the loss of utility of the agents who are now in minority in their neighborhood, that is:

$$
\Delta U = U(x_2) - U(x_1) = -2(K-1) = 2 - 2K
\tag{4.10}
$$

Comparing the difference in the potential function and in the collective utility between these two configurations, we observe that decreasing the number of perfectly mixed neighborhoods decreases collective utility (due to the loss of those who are in the minority in the new configuration) while increasing the value of the potential

function. This is because the utility of the two moving agents is still 1 in their new neighborhood, while they clearly exert negative externalities on the agents of the group which is now in minority in this neighborhood. Stationary configurations will tend therefore to exhibit few perfectly mixed neighborhoods, at the expense of collective utility.

Our analysis based on Schelling's original utility function shows thus two points.[15] First, it provides an analytical demonstration of Schelling's result, that this pair of utility functions leads to segregated stationary configurations. Second, it sheds light on the source of the discrepancy between the collective utility of stationary configurations and the maximum utility that could be attained with a perfectly mixed environment.

4.3. Asymmetric peaked utility functions

In this section, we apply our analytical framework to asymmetric peaked utility functions displayed in Fig. 3, which are studied in Zhang (2004b) and Pancs and Vriend (2007). These utility functions are particularly appealing to demonstrate Schelling's intuition, that the aggregate outcome of the game can run against individual preferences. Indeed, these functions consider a case where agents strictly prefer perfectly mixed neighborhoods versus any level of segregation. We will show that our potential function provides a criterion for global segregation or for integration.

In the following, we assume the case $\xi \equiv 0$ where the utility of each type of agents can be described entirely by the number of like-color

---

[15] In Schelling's original model, the agents are satisfied if more than a fraction f of their neighbors is similar to them. Here, we only investigate the case $f = 1/2$. In fact, it can be shown that in the case $f \neq 1/2$ our results do not hold because condition 15 is not always fulfilled and no potential function can be derived. This case is a counter-example showing that our solution is not completely general. However, Schelling's original utility functions can be approximated by slightly different utility functions verifying condition 15. For example, we could consider that instead of requiring a minimum fraction of similar neighbors (i.e. a utility depending on $R/(R+G)$ for a red agent), the agents require a minimum density of same-color agents in their blocks (i.e. a utility depending on $R/H = R/(R+G+V)$ for a red agent).

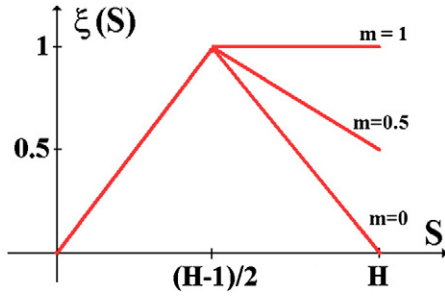**Fig. 3.** Asymmetric peaked function for some values of *m*.

agents. This choice permits an easier writing of the analytical results. It also can be understood as the limit case $v \to 0$ For simplicity, we suppose that the number H of possible neighbors of an agent is even. The asymmetric peaked utility functions can then be written as:

$$\begin{cases} u_R(R,G) = \xi_{ap}(R) \\ u_G(R,G) = \xi_{ap}(G) \end{cases}$$

with:

$$\begin{cases} \xi_{ap}(s) &= 2s/H & \text{if } s \leq H/2 \\ \xi_{ap}(s) &= 2-m-2(1-m)s/H & \text{if } s > H/2 \end{cases}$$

with m a fixed parameter (see Fig. 3).

Using Θ the Heaviside function defined by: $\Theta(x) = 0$ if $x<0$ and $\Theta(x) = 1$ if $x \geq 0$, this utility function can also be written as:

$$\xi_{ap}(s) = 2\frac{s}{H} - (2-m)\frac{2}{H}\left(s - \frac{H}{2}\right)\Theta\left(s - \frac{H}{2}\right).$$

For a given state x of the city, simple calculations show that the collective utility can be written as $U(x) = \sum_q (\tilde{U}(R_q) + \tilde{U}(G_q))$, with

$$\tilde{U}(S) = 2\frac{S(S-1)}{H} - (2-m)\frac{2S}{H}\left(S-1-\frac{H}{2}\right)\Theta\left(S-1-\frac{H}{2}\right), \quad \forall 0 \leq S \leq H. \tag{4.11}$$

Likewise, the corresponding potential function is $\mathcal{F}(x) = const + \sum_q F(R_q, G_q) = const + \sum_q \left(\tilde{F}(R_q) + \tilde{F}(G_q)\right)$, where

$$\tilde{F}(S) = \sum_{s=0}^{S-1} \xi_{ap}(s)$$
$$= \frac{(S-1)S}{H} - \frac{2-m}{H}\left(S-1-\frac{H}{2}\right)\left(S-\frac{H}{2}\right)\Theta\left(S-\frac{H}{2}-1\right) \tag{4.12}$$
$$= \frac{1}{2}\left[\tilde{U}(S) + (2-m)\left(S-1-\frac{H}{2}\right)\Theta\left(S-1-\frac{H}{2}\right)\right].$$

Proof. See Appendix D. □

This expression again implies that the potential $\mathcal{F}$ and the collective utility U are linearly related when the individual utility is linear (case $m=2$). The lower m, the less linear the individual utility and the greater the divergence from the $\mathcal{F} = const + U/2$ relation. Thus, relation 4.12 highlights the crucial role of the asymmetric parameter m which is the driver of the moves that produce externalities. We thus provide an analytical result that complements that of Pancs and Vriend (2007) based on simulations.

To be more specific, let us compare two configurations x_1 and x_2, which differ only in the repartition of $H+1$ red and $H+1$ green agents in two neighborhoods. In configuration x_1, the repartition is rather homogeneous, with $H/2+1$ red and $H/2$ green agents in the first neighborhood and $H/2$ red and $H/2+1$ green agents in the second one. In configuration x_2, the repartition is more segregated, with $H/2+$

$1+K$ red and $H/2-K$ green agents in the first neighborhood and $H/2-K$ red and $H/2+1+K$ green agents in the second one, $K \in \{0,1,...,H/2\}$ being an integer determining the level of segregation. The difference in the potential function between the two configurations is due only to the change affecting these two neighborhoods. It is written:

$$\Delta\mathcal{F} = 2\tilde{F}(H/2+1+K) + 2\tilde{F}(H/2-K) - 2\tilde{F}(H/2+1) - 2\tilde{F}(H/2)$$
$$= \frac{2m}{H}K(K+1). \tag{4.13}$$

Proof. See Appendix D. □

As might be expected, $\Delta\mathcal{F}$ increases with m, which means that the segregated configuration is more probable and stable than the mixed one since the asymmetry toward like-agents is stronger. It also increases with K, which means that for a given m a highly segregated block is more probable than a slightly segregated one. More importantly, a perfectly segregated block will be more probable than a perfectly mixed one if and only if relation 4.13 is positive, i.e. if and only if $m>0$.

The corresponding difference in collective utility consists of the loss of all the agents. It can be written:

$$\Delta U = 2\tilde{U}(H/2+1+K) + 2\tilde{U}(H/2-K) - 2\tilde{U}(H/2+1) - 2\tilde{U}(H/2)$$
$$= 4K\left(m\left(\frac{1}{2}+\frac{K+1}{H}\right)-1\right). \tag{4.14}$$

Proof. See Appendix D. □

It is obvious that increasing segregation also increases collective utility as soon as $m \geq 1$. It is straightforward to verify based on Eq. (4.14) that $\Delta U \geq 0 \Leftrightarrow m \geq m^* = H (H+1)^{-1}$.[16]
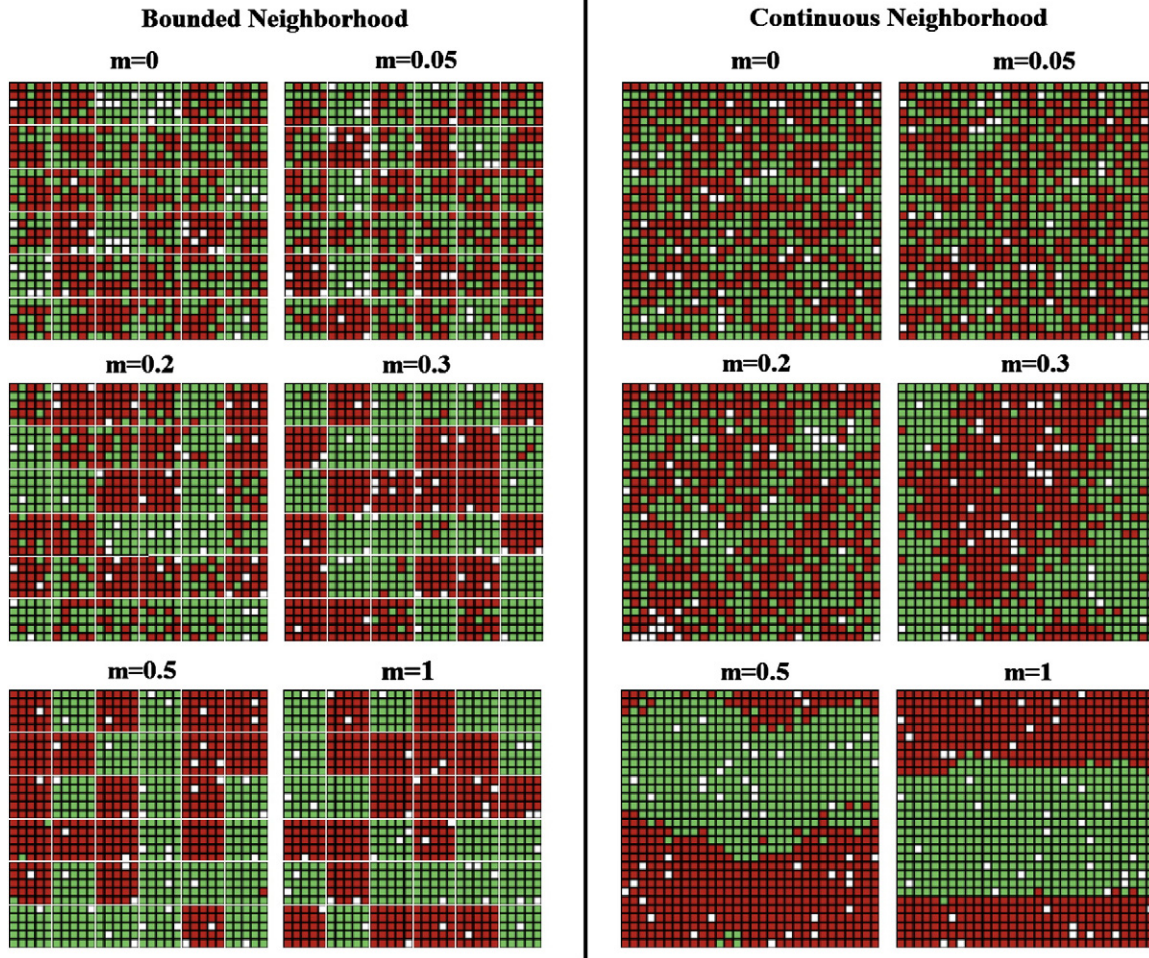
Our analysis provides a microscopic criterion allowing us to predict a global outcome. For $0<m<m^* \simeq 1$, complete segregated configurations will be obtained at the expense of the collective utility and for $m<0$, perfectly mixed configuration will be obtained. These results hold of course in the limit of a low noise level $(T \to 0)$.

The simulations show the same results. The snapshots presented on the left panel of Fig. 4 are typical stationary configurations[17] obtained by simulating an artificial city where the agents' preferences are given by the asymmetric peaked utility function with bounded neighborhoods. These snapshots allow us to compare the analytical results obtained for $v \to 0$ and $T \to 0$ with a more realistic $v = 5\%$ and $T = 0.1$. We can see that for values of the asymmetry parameter m close to 0, the system converges toward randomly-organized mixed configurations, which also maximize the utility of most agents. On the contrary, for higher values of m, completely segregated configurations are obtained. The transition between these two extreme outcomes occurs for values of m included between 0.05 and 0.2. The relative smoothness of this transition is due to the non-zero values of the vacancy rate and of the level of noise.

The outcome for values of m higher than 0.2 illustrates the paradox in Schelling's model: large segregative patterns appear although they absolutely do not maximize the utility of most agents, as most of them are stuck inside an homogeneous area with a utility of 0.5. In this case, one of the key element driving segregation is the asymmetry of this utility function, that is, the fact that even if the agents have a strict preference for mixed environments, they still prefer to belong to the majority group instead of being in minority, as already shown in Pancs and Vriend (2007). In particular, with the asymmetric

---

[16] The limit value is not strictly equal to 1 because of the precise definition of our model: the argument of the utility function is the number of the agent's neighbors, which does not include the agent. In the limit $H \gg 1$, m^* converges toward 1.
[17] Refer to (Grauwin et al., 2010) for our precise definition of stationary configurations in the simulations.

Bounded Neighborhood

m=0   m=0.05

m=0.2   m=0.3

m=0.5   m=1

Continuous Neighborhood

m=0   m=0.05

m=0.2   m=0.3

m=0.5   m=1

Fig. 4. Typical stationary configurations obtained by simulations with the asymmetric peaked utility function. The demographic parameters are ($N = 30, v = 5\%, n_R = 0.5$). Neighborhood sizes are fixed to $H = 24$, and the level of noise is fixed to $T = 0.1$. Left: with a bounded neighborhood description. Right: with a continuous neighborhood description.

peaked utility function, a red (green) agent may move for example from a 49% red (green) neighborhood to a 51% red (green) neighborhood because it slightly increases the agent's utility. Meanwhile, this move is likely to decrease the utility of the previous and new neighbors and therefore decrease the level of collective utility. Both of these factors imply that a highly-segregated configuration is necessarily very stable. Indeed, once the city is divided into homogeneous areas, a red agent will have no incentive to go from the red area to the green one, because his/her utility would drop from 0.5 to 0.[18]

5. Discussion, limits and extensions

In this section, we justify the bounded neighborhood assumption with a supplementary result in the continuous neighborhood case and present an extension of our model in which the agents are affected by the composition of neighboring blocks. We then use simulations to compare the outcomes of a bounded- *versus* a continuous-neighborhood model. We also investigate some other extensions of the model, such as different agents' preferences and an analysis of taxation.

5.1. Bounded vs continuous neighborhoods: analytical results

Most papers based on simulations as well as those presenting analytical results for particular cases (Pancs and Vriend, 2007; Zhang, 2004a) deal with continuous neighborhoods. In this subsection, we first establish the important result that, in the context of continuous neighborhoods, a potential function only exists with linear utility functions. We then consider an extension of the analysis with bounded neighborhoods, in which a spatial connection between close neighborhoods is introduced, so that the main drawback of bounded neighborhoods, namely the fact that an agent is not affected by the composition of the blocks next to his/her own, is avoided.

Proposition 2. When using a continuous neighborhood description, a potential function exists if the agents' individual utility functions are bilinear functions of the form

$$u_R(R, G) = u_R(0, 0) + aR + bG$$
$$u_G(R, G) = u_G(0, 0) + bR + dG \qquad (5.1)$$

where a, b, d are three real constants. In this case, a potential function $\mathcal{F}$ can also be written as $\mathcal{F} : x \rightarrow \mathcal{F}(x) = const + U(x)/2$.

Reciprocally, no potential function exists for any other form of utility functions.

The first part of this claim – which is that a potential function exists if the utility functions are chosen as in Eq. (5.1) – is the main subject of Zhang (2004a). The second part of this claim — which is that no

---

[18] And even though a red agent from time to time goes into the green area by mistake, he/she will have a strong incentive to return to the red area because of the asymmetry in the utility function, and he/she will do so very likely before a second red agent enters the green area.

potential function exists for any other choice of utility functions (in the context of the model presented in this paper and with continuous neighborhoods) to our knowledge has not been proved elsewhere.[19] A sketch of the proof of Proposition 2 is given in Appendix E. The main reason why, in general, no potential function exists is that when a moving agent generates externalities on his/her (past and new) neighbors, these externalities depend on the type of his/her neighbors' own neighbors. A way to handle this issue in the continuous neighborhood case is to consider linear utility functions. For these, the generated externalities do not depend on the neighbors' neighbors, because linearity implies that the newcomer generates the same change in utility on his neighbors, independently of their initial situation; the same holds for his/her former neighbors. For non-linear utilities, a potential function exists only in the bounded neighborhood case.[20]

In the bounded neighborhood framework we use in our model, the main drawback compared to a continuous neighborhood framework is the fact that an agent is not affected by the composition of the blocks next to his/her own, which may seem unrealistic. In such a model, each neighborhood is considered as an island with no connection to other neighborhoods. In order to overcome this downside, we present here the sketch of a model which is an extension of our bounded-neighborhood model including transitivity between blocks while maintaining the existence of a potential function.

The bounded neighborhood formulation is based on a given partition (let us call it Q_1) of the city lattice into blocks, as shown on Fig. 1a. We can define a second partition Q_2 of the city lattice such that blocks of partition Q_2 overlap with blocks of partition Q_1. Fig. 5b shows an example where partitions are shifted from each other by half a block's diagonal. Let us assign to each red (resp. green) agent the two blocks $q_1 \in Q_1$ and $q_2 \in Q_2$ he is living in and the corresponding utilities $u_{R,i} = u_R(R_{q_i} - 1, G_{q_i})$ where $(i \in \{1,2\})$ (resp. $u_{Gi} = u_G(R_{q_i}, G_{q_i} - 1)$). Two potential functions $\mathcal{F}_1$ and $\mathcal{F}_2$ where $\mathcal{F}_i = \mathcal{F}(\{R_{q_i}, G_{q_i}\})$ can similarly be assigned to the two partitions.

Let us now define a final model encompassing both partitions. The utility u of an agent is taken as the sum $u_1 + u_2$ corresponding to his utility in both partitions. It is straightforward to verify that the analytical properties of our model hold in this larger description and that the states of the city can be characterized by the potential function $\mathcal{F}_{12} = \mathcal{F}_1 + \mathcal{F}_2$, whose maxima correspond to the stationary states of the city. Moreover, the fact that partitions overlap implies spatial transitivity between blocks of a given partition. Indeed, suppose that agents compute their utility with the asymmetric peaked utility function, as in Section 4.3. We showed earlier, in the model with one partition, that for $m > 0$, stationary states maximizing the potential function correspond to states where each block is either completely red or completely green, while mixed blocks are dynamically unstable. It is possible to show that in the two partitions model, the same result holds because the simultaneous maximization of the two potential functions can be achieved by minimizing the number of mixed blocks, which leads to a further concentration of blocks of the same color, as illustrated on Fig. 5b.

It is thus possible to extend our 'bounded neighborhood' framework where blocks are isolated from one another to a framework taking account of a contiguity between the blocks, while retaining the analytical solution, i.e. the existence of a potential function. We think that considering such an extension is a simple way to avoid the main limit of bounded neighborhood models. This extension is however not sufficient to bridge the gap between bounded- and
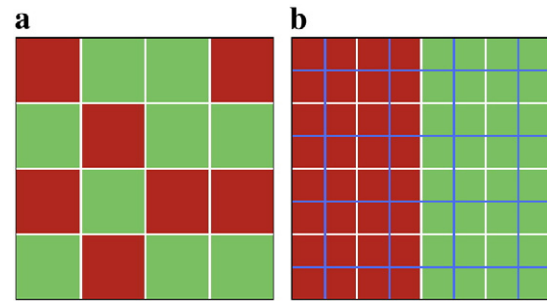
Fig. 5. Illustration of the expected effects of a "two partitions" model. a In the one partition model, homophilic preference leads to a pattern where blocks are exclusively populated by either red or green agents as depicted in Fig. 4. In this case however, agents are not aware of the composition of nearby blocks. b In the two partition framework (in each partition, blocks are delimited by either white or blue lines), nearby blocks of a given partition have a transitive influence on each other through the overlapping blocks of the other partition. The same segregation dynamics implying a minimization of the number of mixed blocks of either partition will lead to a gathering of same-color blocks.

continuous-neighborhood models and it remains true that, in general, no potential function can be derived in the continuous neighborhood case as shown in Proposition 2. Therefore, we use simulations in the following subsection to compare the segregation outcomes with bounded and continuous neighborhoods, in order to give an account of how the hypothesis of bounded neighborhoods affects our results.

5.2. Bounded vs continuous neighborhoods: segregation measures

We now investigate the differences between bounded *versus* continuous neighborhoods in the general case. We ran simulations using both neighborhood descriptions with the asymmetric peaked utility function for different values of the asymmetry parameter m. The neighborhood size is $H + 1 = 25$, the size of the city is $N^2 = 900$, the initial state is chosen at random and the simulation is stopped after each agent has been given on average 50 opportunities to move (which is more than twice the number needed to reach stationary configurations in the cases we studied[21]). For each set of parameters (neighborhood description and value of m), we ran 200 different simulations. The left and right panels of Fig. 4 present snapshots of stationary configurations obtained respectively with bounded- and continuous-neighborhood descriptions.

In order to compare the results of the two neighborhood definitions, a measure of segregation is calculated on the final configuration of each run of the simulations. We use an evenness index, that measures a weighted average of individual shares of same-color neighbors:

$$\sum_{i=1}^{N} s_i / \left(\sum_{i=1}^{N} s_i + \sum_{i=1}^{N} d_i \right) \qquad (5.2)$$

where s_i and d_i respectively stand for the number of similar and dissimilar neighbors of a given agent i in a local neighborhood defined as the H' closest cells around this agent. This segregation measure was used by Pancs and Vriend (2007). It accounts for the average share of same-color neighbors. Values of this index close to 0.5 correspond to mixed configurations while values close to 1 correspond to highly segregated patterns. The size of the neighborhood H' that is used here is *a priori* different from the size H considered in the utility function and we compute the index for different values of H' ($H' = 8$, $H' = 24$, $H' = 44$

---

[19] Note however that the non-existence of a potential function does not preclude the convergence of the game to a stationary state.

[20] Zhang (2004b) proposes a continuous neighborhood model without vacant cases which allows him to derive a potential function for utility functions defined by two linear pieces.

[21] See footnote 13.

Table 1
Share indices measured for different settings of the neighborhood description (either bounded or continuous) and of the asymmetry parameter m of the utility function. $N = 900$ and $H + 1 = 25$.

		Bounded neighborhood				Continuous neighborhood			
		$H' = 8$	$H' = 24$	$H' = 44$	$H' = 108$	$H' = 8$	$H' = 24$	$H' = 44$	$H' = 108$
$m = 0$	Average	0.561	0.565	0.547	0.558	0.494	0.516	0.509	0.503
	5%	0.542	0.548	0.533	0.514	0.481	0.511	0.506	0.501
	95%	0.579	0.578	0.558	0.532	0.507	0.523	0.514	0.506
$m = 0.05$	Average	0.582	0.582	0.561	0.528	0.506	0.519	0.51	0.503
	5%	0.563	0.567	0.545	0.514	0.492	0.511	0.505	0.503
	95%	0.601	0.598	0.577	0.543	0.518	0.528	0.516	0.507
$m = 0.2$	Average	0.679	0.654	0.617	0.555	0.558	0.544	0.525	0.509
	5%	0.653	0.63	0.595	0.529	0.539	0.531	0.514	0.503
	95%	0.705	0.677	0.64	0.583	0.576	0.558	0.536	0.517
$m = 0.3$	Average	0.775	0.723	0.673	0.582	0.726	0.704	0.681	0.645
	5%	0.742	0.689	0.636	0.545	0.688	0.666	0.638	0.608
	95%	0.81	0.759	0.711	0.624	0.767	0.75	0.728	0.69
	Average	0.846	0.777	0.716	0.602	0.915	0.889	0.864	0.807
$m = 0.5$	5%	0.824	0.746	0.678	0.561	0.903	0.877	0.852	0.797
	95%	0.872	0.809	0.757	0.648	0.926	0.899	0.873	0.816
$m = 1$	Average	0.85	0.778	0.715	0.598	0.942	0.913	0.885	0.824
	5%	0.822	0.743	0.673	0.551	0.939	0.909	0.881	0.82
	95%	0.875	0.813	0.759	0.659	0.945	0.916	0.889	0.829

and $H' = 108$). This allows us to give an account of the spatial scale of the observed segregation patterns.[22]

The average of the share measure in the final configuration of these 200 runs are displayed in Table 1, along with the corresponding 5th and 95th percentiles. When looking at local segregation patterns ($H' = 8$), we observe that mixed random configurations are obtained in both bounded and continuous neighborhood descriptions for low values of m. Segregation patterns appear in both descriptions for high values of m. However, the two neighborhood descriptions produce a different "transition range" between mixed and segregated configurations (roughly 0.1 ± 0.05 in the bounded neighborhood case *versus* 0.25 ± 0.05 in the continuous neighborhood case; see also Fig. 4).

These figures also show that the share index decreases with H', but more slowly in the continuous case than in the bounded case. In the bounded neighborhood case, even when all neighborhoods are mono-color (for high values of m), the absence of connections between contiguous neighborhoods produces a random spatial distribution of these green and red neighborhoods that is reflected by the low share index for high values of H'. This is the strongest difference between the two neighborhood definitions: as there are no connections between blocks in the bounded neighborhood case, the spatial arrangement of neighborhoods is purely random. Even if segregation patterns appear at the block level, red and green blocks can be observed side-by-side as well as in clusters of same-color blocks. In other words, the segregation patterns are local in the bounded neighborhood case whereas they are observed at the city level in the continuous neighborhood case.

5.3. Segregation by ethnic origin, income, and preferences for public amenities

So far, we have implicitly assumed that the sole characteristic on which agents base their evaluation of a location is the composition of its neighborhood. However, there are other determinants of residential location choice that are not necessarily correlated to neighbors' characteristics, such as local public goods. The red and green labeling of our two groups thus may correspond to two different ethnic

origins or two groups with different preferences for local public goods. Tiebout (1958)'s analysis of the sorting induced by local public goods is perhaps the main competitor to Schelling (1971) in terms of its influence on later work on neighborhood choice.

It is easy to write versions of our model which take account of agents' preferences for public goods while keeping the existence and properties of a potential function. For example if A is the set of all the public facilities and $d_{i,a}$ is the distance between an agent i and an amenity $a \in A$, the utility of an agent i can be rewritten in a general fashion as

$$u_i(R, G) \to u_i(R, G) + \tilde{u}_i\left(\left\{d_{i,a}\right\}_{a \in A}\right) \tag{5.3}$$

and one can easily derive the more general form of the potential function

$$\mathcal{F}(x) \to \mathcal{F}(x) + \sum_i \tilde{u}_i\left(\left\{d_{i,a}\right\}_{a \in A}\right). \tag{5.4}$$

This generalized approach provides a means to correct one of the biases of our analytical model, namely the lack of heterogeneity in locations. However, the extraction of the properties of the stationary state from this condensate global function becomes quite challenging, as the dimension of the state variable of the system increases with the number of added amenities.

5.4. Taxation

In order to illustrate the usefulness of the potential function, we further consider the introduction of a tax on segregation. The basic concept at the center of a model *à la* Schelling is that of an agent deciding where to move according solely to the utility gain she would achieve were she to move. Since her move will affect her former and new neighbors, an implicit consequence is that she could generate externalities that amount to $\Delta U - \Delta u$ when moving.

We consider now a benevolent planner who subsidizes positive externalities and taxes negative externalities. A way to model the action of that planner is to write the probability that a move happens as:

$$Pr\{move\} = \frac{1}{1 + e^{-(\Delta u + \alpha(\Delta U - \Delta u))/T}} \tag{5.5}$$

where $0 \le \alpha \le 1$ is a parameter controlling the tax level. The limit case $\alpha = 0$ corresponds to a standard Schelling model and the limit case

---

[22] More sophisticated measures of segregation exist, in which the spatial arrangement of neighborhoods intervenes (Morrill, 1991; Reardon and O'Sullivan, 2004; Wong, 2005). But they also require to define a spatial scale through contiguity or a distance-decay function. We opted here for a simple measure that can be computed for the two neighborhood definitions and that we can compute for different spatial scales.

$\alpha=1$ corresponds to the case where only the collective interest is taken into account.

Following the path of the proofs developed in Section 2.2, we can infer the stationary distribution in the bounded neighborhood framework:

$$\Pi(x) = \frac{e^{((1-\alpha)\mathcal{F}(x)+\alpha U(x))/T}}{\sum_z e^{((1-\alpha)\mathcal{F}(z)+\alpha U(z))/T}}.$$
(5.6)

The potential function can thus in this context be generalized to $(1-\alpha)\mathcal{F}(x) + \alpha U(x)$. We noted that the configurations maximizing $\mathcal{F}$ are not in general maximizing U and could even in certain cases (asymmetric peaked utility functions) be very unfavorable to U. We show here that the parameter $\alpha=1$ of the tax indeed allows the planner to obtain a stationary configuration in which the collective utility is maximized.[23]

Such a Pigouvian tax supposes that the central government has a precise knowledge of the neighborhood composition of each moving agent, which is a rather utopian assumption. We propose here a new variation of the Schelling model incorporating a different tax rule, based on more realistic assumptions regarding government's ability to intervene. We define a simple tax, which aims at preventing the emergence or maintenance of a dominant group in each block. Central government imposes on each agent a *tax* which is defined as:

$$r(S,\theta) = \begin{cases} \theta|S/H-0.5| & \text{if } S/H{>}0.5 \\ 0 & \text{otherwise} \end{cases}.$$
(5.7)

where S is the number of agents similar to the taxed agent within his block and θ is a fixed parameter controlling the tax level. Note that the tax is payable only by every agent belonging to the majority in their district and therefore does not penalize agents in the minority. Note also that this definition of the tax is defined at the level of block size H. This assumption implies direct competition between the segregation effect of a model *à la* Schelling, which favors values of s/H greater than 0.5, and the tax mechanism which favors values of S/H below 0.5.

The utility of red and green agents is then redefined, taking into account the penalty imposed by the tax, as:

$$\begin{cases} \hat{u}_R(R,G,\theta) = u_R(R,G) - r(R,\theta) \\ \hat{u}_G(R,G,\theta) = u_G(R,G) - r(G,\theta). \end{cases}$$
(5.8)

The tax is null for $\theta=0$. For $\theta{>}0$, one can expect that a rise in θ implies a rise in the probability of having non-segregated configurations in the stationary states.

From an analytical viewpoint, the properties of the model resulting from the Markov chain theory remain unchanged: there exists one unique stationary distribution and the final configurations do not depend on the initial ones. It is also straightforward to check that if the pair of utility functions (u_r, u_G) verifies condition 3.5 then so also does the pair $(\hat{u}_r, \hat{u}_G)$ and it is hence possible to derive a potential function incorporating the tax effects.

To be more specific, suppose that the utility of the agents is given by the asymmetric peaked function introduced in Section 4.3, which is a case where the stationary configuration clearly diverges from the collective optimum. An agent's utility $\hat{u}$ can be rewritten in this case as:

$$\begin{aligned} \hat{u}(S,\theta) &= u_{ap,m}(S) - r(S,\theta) \\ &= \begin{cases} 2S/H - 0 & \text{if } S/H{\leq}0.5 \\ 2-m-2(1-m)S/H-\theta|S/H-0.5| & \text{if } S/H{>}0.5 \end{cases} \\ &= \begin{cases} 2S/H & \text{if } s_i{\leq}0.5 \\ 2-(m-\theta/2)-2(1-(m-\theta/2))S/H & \text{if } S/H{>}0.5 \end{cases} \\ &= u_{ap,m-\theta/2}(S). \end{aligned}$$
(5.9)

---

[23] For development on the level of the tax α necessary or sufficient to break undesired stationary configurations, see Grauwin et al. (2009a).

Hence our tax mechanism can be interpreted in this case as direct control by central government over the asymmetry parameter m in the agents' utility function. According to the analytical results obtained in Section 4.3, the asymmetry parameter must be inferior to $m_c=0$ in order to avoid segregation. The equivalence stated in Eq. (5.9) thus allows us to predict that the minimal tax level necessary to break segregated patterns is:

$$\theta_c(m) = 2(m-m_c) = 2m$$
(5.10)

Our analytical framework thus allows us, very simply, to consider the consequences of different tax levels. Other public policies against segregation could also probably be analyzed.

6. Conclusion

In this paper, we used tools from evolutionary game theory to develop an analytical solution of Schelling segregation model for bounded neighborhoods and two homogeneous groups of agents. The analytical solution that we proposed exists if and only if the utility functions are such that there is a symmetry in the externalities generated by green agents on red agents and by red agents on green agents or with all possible pairs of utility functions in the limit case with no vacant cells. This represents a major step forward from work that is based on computer simulations or provides analytical results for specific models. We have shown that the stationary configurations for the selfish individual moves of agents maximize a potential function under relatively mild conditions on the agents' utility functions. This potential function can be interpreted as the sum of the agents' utilities as they move into their neighborhood, starting from a totally empty city. In other words, the potential function cumulates the incentives the agents had to move into the neighborhood where they are located.

This step forward was enabled by a partial reduction in the heterogeneity of agents' neighborhoods through the use of bounded neighborhoods. This allows us to track how each individual move affects the global configuration. When continuous neighborhoods are used, this information is lost because the way a moving agent affects his former and new neighbors depends on factors (the type of the neighbors' neighbors) that are not fully determined by the agent's decision. Therefore, it is generally impossible to know how an individual move affects a function of the global configuration unless the utility functions are linear.

We used the potential function to assess the outcomes of our location model for different utility functions, representing different degrees of preference for mixed environments. We examined successively linear utility functions, Schelling's original utility function and asymmetric peaked utility functions. The first two utility functions lead to segregated stationary configurations. In the linear utility case (and for meaningful values of the parameters), the segregated configurations, that maximize the potential function, also maximize collective utility. With Schelling's original utility function, a divergence between collective utility and the potential function appears. Asymmetric peaked utility functions lead to segregated configurations even for a slight asymmetry, because this asymmetry provokes moves to slightly segregated neighborhoods that will never be compensated by reverse moves.

Our analytical approach helps explain the factors that contribute to the paradoxical result that has generated interest in Schelling's model. Even if the dynamics is governed by agents moving to improve their own utility, the evolution leads to city configurations in which most of the agents are far from being satisfied. Our paper robustly reveals the two main components of segregation. First, the most important element driving segregation is the *asymmetry* of the utility function, as shown in Pancs and Vriend

(2007). In our framework, symmetric functions do not lead to segregation.[24] Once utility functions favor majority over minority status, segregation emerges, even if agents have a strict preference for mixed environments, as in the asymmetric peaked utility function. The second important element is the existence of externalities. As noted by Zhang (2004b) and Pancs and Vriend (2007), the existence of externalities explains why individual preferences for integrated environments may lead to segregated configurations. Indeed, the location choice of an agent is based only on his/her own utility level, even if it also affects the neighbors' utility levels. This makes mixed neighborhoods unstable, and segregated configurations very stable. The instability of mixed neighborhoods is particularly clear in the bounded neighborhood configuration for the asymmetric peaked function. Starting with the Nash equilibrium where $R_q = G_q = (H+1)/2$ and $T > 0$, the logit rule implies that there is a positive probability that an agent accepts a slight decrease of his utility, and leaves a block with composition $R_q = G_q = (H+1)/2$. The agents of the same color remaining in the focal agent's former block now have a lower utility and are even more likely to leave. This creates an avalanche effect which empties the block of agents of the same color, as each move away, further decreases the utility of the remaining agents. Conversely, highly-segregated configurations are very stable. Indeed, once the city is divided into homogeneous areas, a red agent will have no incentive to go from the red area to the green one, his utility dropping from m to 0.

The analytical tool we provide allows consideration of the outcomes of other types of utility functions, in particular those that emerge from empirical findings on social preferences, as soon as they verify the condition of asymmetric externalities or vacant cells can be ignored. It is conceivable then to analyze the theoretical outcomes of these preferences and possibly to test the effect of introducing public policy instruments aimed at decreasing segregation. It is also worth noting that with the original utility function suggested by Schelling, the potential function happens to be a linear form of the Duncan and Duncan segregation index. We thus bridge between theoretical models of segregation and residential segregation measures.

The kind of solution that is developed here has been used in physics in equilibrium statistical mechanics. Equilibrium statistical mechanics has developed powerful tools to link the microscopic and macroscopic levels. These tools are usually limited to physical systems, where the dynamics is governed not by a selfish criterion but by a global quantity such as the total energy. Blume (1993) bridges between statistical physics and a coordination game with local interactions. Here, by using the potential function, which is analogous to state functions in thermodynamics, we extend the analytical framework of statistical mechanics to Schelling's model. By doing so, our work opens the way to analytical treatments of a much wider class of social systems, where the dynamics is governed by individual strategies.

Acknowledgments

Appendix A. Proof of Proposition 1

We consider a bounded neighborhood model. Let us first prove the first part of Proposition 1, that is that any aggregate function $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$ is a potential function that corresponds to (at least) one pair of utility functions (u_R, u_G) of U.

Suppose that $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$ is a potential function of the game, where the intermediate function F is known. Let us assume that an agent is moving from a block 1, characterized by the numbers $(R_1, G_1) \in E_{H+1}$ of red and green agents who live in it, to a block 2 characterized similarly by the numbers $(R_2, G_2) \in E_H$ of red and green agents living in it (since there must be at least one vacant location in block 2 for an agent to move in it, we necessarily have $R_2 + G_2 < H + 1$). By definition, the utility variation of a moving agent must be equal to the variation of $\mathcal{F}$ it induces. Hence:

- to cover the cases when the moving agent is a red one: for all $(R_1, G_1) \in E_{H+1}$ with $R_1 \geq 1$,

$$u_R(R_2, G_2) - u_R(R_1 - 1, G_1) = \\ F(R_2 + 1, G_2) + F(R_1 - 1, G_1) - F(R_2, G_2) - F(R_1, G_1) \quad \text{(A.1)}$$

- to cover the cases when the moving agent is a green one: for all $(R_1, G_1) \in E_{H+1}$ with $G_1 \geq 1$,

$$u_G(R_2, G_2) - u_G(R_1, G_1 - 1) = \\ F(R_2, G_2 + 1) + F(R_1, G_1 - 1) - F(R_2, G_2) - F(R_1, G_1) \quad \text{(A.2)}$$

Taking $R_2 = G_2 = 0$ in Eqs. (A.1) and (A.2), one finds that the utility functions u_R and u_G verify for all $(R, G) \in E_H$:

$$u_R(R, G) - u_R(0, 0) = F(R + 1, G) - F(R, G) - F(1, 0) + F(0, 0) \quad \text{(A.3)}$$

$$u_G(R, G) - u_G(0, 0) = F(R, G + 1) - F(R, G) - F(0, 1) + F(0, 0). \quad \text{(A.4)}$$

These relations define (up to a constant $u(0,0)$) the utility functions the agents necessarily have if $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$ is a potential function of the game. It still remains to prove that this pair of utility functions belongs to the set $\mathbb{U}$. According to relations 54 and 55, one has for all $(R, G) \in E_H$:

$$u_R(R, G) - u_R(R, G + 1) = (F(R + 1, G) - F(R, G)) - (F(R + 1, G + 1) - F(R, G + 1)) \\ = (F(R, G + 1) - F(R, G)) - (F(R + 1, G + 1) - F(R + 1, G)) \\ = u_G(R, G) - u_G(R + 1, G).$$

Hence relation 3.5 holds, which means by definition that the pairs of utility functions (u_R, u_G) defined by relations A.3 and A.4 belongs to $\mathbb{U}$. Notice that in our demonstration no particular constraint has to be assumed on the form of function F. As a consequence, any aggregate function $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q) \in \mathcal{F}$ is a potential function of the game as soon as the pair of agents' utility functions is chosen so that relations A.3 and A.4 hold.

Let us now prove the second part of Proposition 1, which is that to any pair of utility functions (u_R, u_G) of $\mathbb{U}$ there is a corresponding potential function of the form $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$.

Let $(u_R, u_G) \in \mathbb{U}$ be a pair of utility functions verifying condition 3.5. Suppose that $F(0, 0)$, $F(0, 1)$ and $F(1, 0)$ are given and let us define recursively the function F on E_{H+1} by the following equations, verified for all $(R, G) \in E_H$:

$$F(R + 1, G) - F(R, G) = F(1, 0) - F(0, 0) + u_R(R, G) - u_R(0, 0) \quad \text{(A.5)}$$

$$F(R, G + 1) - F(R, G) = F(0, 1) + F(0, 0) + u_G(R, G) - u_G(0, 0). \quad \text{(A.6)}$$

The most important thing to notice is that these two relations are consistent with each other, thanks to condition 3.5 which

links the two utility functions u_R and u_G. By summing Eq. (A.5) on R and then Eq. (A.6) on G, one finds the following expression for function F:

$$F(R,G) - F(0,0) = R(F(1,0) - F(0,0)) + \sum_{r=1}^{R} (u_R(r-1,0) - u_R(0,0))$$
$$+ G(F(0,1) - F(0,0)) + \sum_{g=1}^{G} (u_G(R,g-1) - u_G(0,0))$$
(A.7)

or conversely by summing Eq. (A.6) on G then Eq. (A.5) on R,

$$F(R,G) - F(0,0) = R(F(1,0) - F(0,0)) + \sum_{r=1}^{R} (u_R(r-1,G) - u_R(0,0))$$
$$+ G(F(0,1) - F(0,0)) + \sum_{g=1}^{G} (u_G(0,g-1) - u_G(0,0)).$$
(A.8)

Hence, since $\mathcal{F} = \sum_{q \in \mathcal{Q}} F(R_q, G_q)$ we can obtain, after rearranging the different terms, a symmetric expression of the potential:

$$\mathcal{F} = |\mathcal{Q}| F(0,0) + N_R(F(1,0) - F(0,0) - u_R(0,0)) + N_G(F(0,1) - F(0,0) - u_G(0,0))$$
$$+ \frac{1}{2} \left[\sum_{r=1}^{R} (u_R(r-1,0) + u_R(r-1,G)) + \sum_{g=1}^{G} (u_G(0,g-1) + u_G(R,g-1)) \right].$$
(A.9)

Since the potential can be chosen up to a constant, it is clear from the previous expression that the choice of $F(0,0)$, $F(0,1)$, $F(1,0)$, $u_R(0,0)$ and $u_G(0,0)$ does not really matter. This justifies our choice to set them at zero to simplify the generic expressions of the potential given in Eqs. (3.7) and (3.8).

According to Eq. (A.7), $F(R,G)$ can be interpreted as the sum of the settling utility of R red and G green agents, considering that these agent settle one by one in an initially empty block, the red agents first and then the green ones. The same goes for Eq. (A.8) but the green agents settle first, followed by the red agents. In fact, relation 3.5 ensures that the sum does not depend on the exact order in which the agents settle (see Section 3.3 for more precisions). Hence $F(R_q, G_q)$ can be written as the mean of the settling utilities of the agents over all possible orders of settlement:

$$F(R_q, G_q) = \sum_{\substack{0 \le r \le R_q, \\ 0 \le g \le G_q \\ (r,g) \ne (0,0)}} \left(\frac{\alpha_R(r,g)}{\alpha(R_q, G_q)} u_R(r-1,g) + \frac{\alpha_G(r,g)}{\alpha(R_q, G_q)} u_G(r,g-1) \right)$$

where $\alpha(R_q, G_q) = (R_q + G_q)!$ is the number of settling orders of the $R_q + G_q$ agent living in block q, $\alpha_R(r,g)$ is the number of such orders in which a red agent settle after $r-1$ red and g green, in which case his/her utility is $u_R(r-1,g)$, and similarly $\alpha_G(r,g)$ is the number of such orders in which a green agent settles after r red and $g-1$ green, in which case the utility of this agent is $u_R(r-1,g) \cdot \alpha_R(r,g)$ is the product of:

- R_q, the number of ways of choosing the rth red settling agent,
- $\binom{R_q-1}{r-1}\binom{G_q}{g}$, the number of ways of sharing out the $R_q - 1 + G_q$ other agents − either before or after this agent,
- $(r+g-1)!$, the number of ways of ordering the agents who settle before and
- $(R_q + G_q - r - g)!$, the number of ways of ordering the agents who settle after.

$\alpha_G(r,g)$ can be similarly computed. We thus end up with a new formula for $F(R_q, G_q)$:

$$F(R_q, G_q) = \sum_{\substack{0 \le r \le R_q, \, 0 \le g \le G_q \\ (r,g) \ne (0,0)}} \binom{R_q}{r}\binom{G_q}{g} \frac{(r+g-1)!(R_q + G_q - r - g)!}{(R_q + G_q)!} (r\, u_R(r-1,g)$$
$$+ g\, u_G(r,g-1)).$$
(A.10)

Appendix B. Relation between the potential function $\mathcal{F}$ and the collective utility U

Let us suppose that $(u_R, u_G) \in \mathbb{U}$, and that the potential function of the system can be expressed as a linear function of the collective utility, i.e. $\mathcal{F}(\{R_q, G_q\}) = \lambda U(\{R_q, G_q\}) + \mu$. Since the potential function can be defined up to constant, we can take $\mu = 0$. Writing the utility functions in the form

$$u_R(R,G) = \xi_R(R) + \sum_{g=0}^{G-1} \xi(R,g)$$
$$u_G(R,G) = \xi_G(G) + \sum_{r=0}^{R-1} \xi(r,G)$$

introduced in Eqs. (3.9) and (3.10), the relation of proportionality between the potential and the collective utility can be written as

$$\sum_q \left(\sum_{r=0}^{R_q-1} \xi_R(r) + \sum_{g=0}^{G_q-1} \xi_G(g) + \sum_{r=0}^{R_q-1}\sum_{g=0}^{G_q-1} \xi(r,g) \right)$$
$$= \lambda \sum_q \left(R_q \xi_R(R_q - 1) + R_q \sum_{g=0}^{G_q-1} \xi(R_q-1,g) + G_q \xi_G(G_q - 1) \right.$$
$$\left. + G_q \sum_{r=0}^{R_q-1} \xi(r, G_q - 1) \right)$$

Since this relation must hold for all $\{R_q, G_q\}$, it follows that that for all $(R,G) \in E_H$, the following holds:

$$\sum_{r=0}^{R-1} \xi_R(r) + \sum_{g=0}^{G-1} \xi_G(g) + \sum_{r=0}^{R-1}\sum_{g=0}^{G-1} \xi(r,g)$$
$$= \lambda \left(R\xi_R(R-1) + R\sum_{g=0}^{G-1} \xi(R-1,g) + G\xi_G(G-1) + G\sum_{r=0}^{R-1} \xi(r, G-1) \right)$$
(B.1)

Taking successively $G=0$ and $R=0$ in that last equation provides three independent relations dissociating the three functions ξ_R, ξ_G and ξ:

$$\forall R>0, \quad \sum_{r=0}^{R-1} \xi_R(r) = \lambda R \xi_R(R-1)$$
(B.2)

$$\forall G>0, \quad \sum_{g=0}^{G-1} \xi_G(g) = \lambda G \xi_G(G-1)$$
(B.3)

$$\forall (R,G), \in E_H \sum_{r=0}^{R-1}\sum_{g=0}^{G-1} (\lambda \xi(R-1,g) + \lambda \xi(r,G-1) - \xi(r,g)) = 0.$$
(B.4)

Notice moreover that the convention $u(0,0) = 0$ implies $\xi_R(0) = \xi_G(0) = 0$. Let us also define $a = \xi_R(1)$, $d = \xi_G(1)$ and $b = \xi(0,0)$. Starting from Eqs. (B.2) to (B.4), it is straightforward to prove recursively that

$$\lambda = 1/2$$
$$\forall R>0, \xi_R(R) = aR$$
$$\forall G>0, \xi_G(G) = dG$$
$$\forall (R,G) \in E_H \xi(R,G) = b.$$

Hence the agents' utility functions corresponds exactly to those introduced in Eq. (4.1):

$$u_R(R,G) = aR + bG$$
$$u_G(R,G) = bR + dG.$$

The individual utilities are thus necessarily linear in the numbers of similar and dissimilar neighbors in the case that the potential function $\mathcal{F}$ is proportional to the collective utility U. □

Appendix C. Calculation of a potential function in Schelling case

Suppose that the agents compute their utility with the Schelling utility function (which is equal to 1 if their fraction of similar neighbors is superior or equal to 0.5, and equal to 0 otherwise). This utility function can be expressed in terms of the number of red and green neighbors as follows:

$$u_R(R,G) = \Theta(R-G) = \frac{1}{2}(1+|R+1-G|-|R-G|)$$

$$u_G(R,G) = \Theta(G-R) = \frac{1}{2}(1+|R-1-G|-|R-G|) \qquad \text{(C.1)}$$

where Θ is the Heaviside function defined by: $\Theta(x)=0$ if $x<0$ and $\Theta(x)=1$ if $x\geq0$. Notice that in this example (and only in this example) the convention $u(0,0)+0$ used in Proposition 1 is not respected. The form we choose to write the Schelling utility function imposes $u_R(0,0)=u_G(0,0)=1$. It is easy to figure out that this particular pair of utility functions respect condition 3.5, and is therefore in the set U. Indeed,

$$u_R(R,G)-u_R(R,G+1)$$
$$= \Theta(R-G)-\Theta(R-G-1) = \begin{cases} 0-0 &= 0 & \text{if } R\leq G-1 \\ 1-0 &= 1 & \text{if } R = G \\ 1-1 &= 0 & \text{if } R\geq G+1 \end{cases}$$

and

$$u_G(R,G)-u_G(R+1,G)$$
$$= \Theta(G-R)-\Theta(G-R-1) = \begin{cases} 1-1 &= 0 & \text{if } R\leq G-1 \\ 1-0 &= 1 & \text{if } R = G \\ 0-0 &= 0 & \text{if } R\geq G+1 \end{cases}.$$

Hence relation $u_R(R,G) - u_R(R,G+1) = u_G(R,G) - u_G(R+1,G)$ is always verified.

To compute a corresponding potential function, we can refer to the general form of Eq. (A.9) (since we do not use the convention $u(0,0)=0$ in this particular example) which can be written here as:

$$\mathcal{F} = const + \frac{1}{2}\sum_{q\in Q}\left[\sum_{r=0}^{R_q-1}\left(u_R(r,0)+u_R\left(r,G_q\right)\right) + \sum_{g=0}^{G_q-1}\left(u_G(0,g)+u_G\left(R_q,g\right)\right)\right]$$

$$= const + \frac{1}{4}\sum_{q\in Q}\left[\sum_{r=0}^{R_q-1}(3+|r+1-G_q|-|r-G_q|) + \sum_{g=0}^{G_q-1}(3+|R_q-1-g|-|R_q-g|)\right]$$

$$= const + \frac{1}{4}\sum_{q\in Q}\left[\left(3R_q+|R_q-G_q|-G_q\right) + \left(3G_q+|R_q-G_q|-R_q\right)\right]$$

$$= const + \frac{1}{2}\sum_{q\in Q}\left(R_q+G_q+|R_q-G_q|\right)$$

$$= const + \frac{1}{2}(N_R+N_G) + \frac{1}{2}\sum_{q\in Q}|R_q-G_q|$$

$$= const' + \frac{1}{2}\sum_{q\in Q}|R_q-G_q|.$$

Appendix D. Potential and collective utility in the case of the asymmetric peaked utility

Proof of Eq. (4.12)

We have to derive the expression $\tilde{F}(S) = \sum_{s=0}^{S-1}\xi_{ap}(s)$, where

$$\begin{cases} \xi_{ap}(s) &= 2s/H & \text{if } s\leq H/2 \\ \xi_{ap}(s) &= 2-m-2(1-m)s/H & \text{if } s>H/2 \end{cases}$$

For $S-1\leq H/2$, it is straightforward to write:

$$\tilde{F}(S) = \frac{2}{H}\sum_{s=0}^{S-1}s = \frac{(S-1)S}{H}. \qquad \text{(D.1)}$$

For $S-1>H/2$, one has

$$\tilde{F}(S) = \frac{2}{H}\sum_{s=0}^{H/2}s + (2-m)(S-1-H/2) - (1-m)\frac{2}{H}\sum_{s=H/2+1}^{S-1}s$$

$$= \frac{2}{H}(2-m)\sum_{s=0}^{H/2}s + (2-m)(S-1-H/2) - (1-m)\frac{2}{H}\sum_{s=0}^{S-1}s$$

$$= (2-m)(H/4+1/2+S-1-H/2) - (1-m)\frac{(S-1)S}{H}$$

$$= (2-m)\left[S-H/4-1/2-\frac{(S-1)S}{H}\right] + \frac{(S-1)S}{H}$$

$$= \frac{(S-1)S}{H} - (2-m)\frac{1}{H}\left[-\left(S-\frac{H}{2}\right)\left(\frac{H}{2}\right) - (S-1)\left(\frac{H}{2}\right) + (S-1)S\right]$$

$$= \frac{(S-1)S}{H} - \frac{2-m}{H}\left(S-\frac{H}{2}-1\right)\left(S-\frac{H}{2}\right). \qquad \text{(D.2)}$$

Thanks to the Heaviside function both results can then be written in the general form:

$$\tilde{F}(S) = \frac{(S-1)S}{H} - \frac{2-m}{H}\left(S-\frac{H}{2}-1\right)\left(S-\frac{H}{2}\right)\Theta\left(S-\frac{H}{2}-1\right). \qquad \text{(D.3)}$$

Proof of Eqs. (4.13) and (4.14)

The computation of $\Delta\mathcal{F}$ in relation 4.13 is based on the expression of $\tilde{F}(S)$ (Eq. D.3), which gives, with $K\in\{0,1,...,H/2\}$:

$$\Delta\mathcal{F} = 2\tilde{F}(H/2+1+K) + 2\tilde{F}(H/2-K) - 2\tilde{F}(H/2+1) - 2\tilde{F}(H/2)$$

$$= \frac{2}{H}\left[\left(\frac{H}{2}+K\right)\left(\frac{H}{2}+K+1\right) + \left(\frac{H}{2}-K-1\right)\left(\frac{H}{2}-K\right) - \left(\frac{H}{2}\right)\left(\frac{H}{2}+1\right)\right.$$

$$\left. - \left(\frac{H}{2}-1\right)\left(\frac{H}{2}\right)\right] - \frac{2(2-m)}{H}K(K+1)$$

$$= \frac{2m}{H}K(K+1).$$

In the same way,

$$\Delta U = 2\tilde{U}(H/2+1+K) + 2\tilde{U}(H/2-K) - 2\tilde{U}(H/2+1) - 2\tilde{U}(H/2)$$

$$= \frac{4}{H}\left[\left(\frac{H}{2}+K\right)\left(\frac{H}{2}+K+1\right) + \left(\frac{H}{2}-K-1\right)\left(\frac{H}{2}-K\right) - \left(\frac{H}{2}\right)\left(\frac{H}{2}+1\right)\right.$$

$$\left. - \left(\frac{H}{2}-1\right)\left(\frac{H}{2}\right)\right] - \frac{4(2-m)}{H}K(H/2+K+1)$$

$$= \frac{8}{H}K(K+1) - \frac{4(2-m)}{H}K(H/2+K+1)$$

$$= \frac{4K}{H}(m(H/2+K+1)-H)$$

$$= 2\Delta\mathcal{F} - 2(2-m)K. \qquad \text{(D.4)}$$

Appendix E. Proof of Proposition 2

In this section, we place ourselves in a continuous neighborhood description.

Constraint on the form of the utility functions

Suppose that there exists a potential function $\mathcal{F} : X \to \mathbb{R}$, and let $u_R : E_H \equiv \{(R,G), 0 \leq R+G \leq H\} \to \mathbb{R}$ and $u_G : E_H \to \mathbb{R}$ be the red and green agents' utility functions. We want here to investigate if the existence of the potential function imposes any constraint on the form of these utility functions. Suppose as shown on Fig. E.1 that A, B and C are three cells of the lattice such that cells B and C are in each other's neighborhood while cell A is neither in the neighborhood of cell B or C. Let x be a given state in which cells A and C are each occupied by a red agent while cell B is empty. We denote by $(R_A, G_A) \in E_H$, $(R_B, G_B) \in E_H$, $(R_C, G_C) \in E_H$ the number of red and green agents within the neighborhood of cells A, B and C in state x.[25]

Let y be the state obtained when the red agent on cell A moves on cell B and z the state obtained if next the red agent on cell C moves to the now empty cell A. The system goes back from state z to state x by a move of the red agent in cell B to cell C. The function $\mathcal{F}$ being a potential function, the following relations hold (taking into account that the move of the agents change the number of red agents in the neighborhoods of cells B and C):

$$\mathcal{F}(y) - \mathcal{F}(x) = \Delta_{xy} u = u_R(R_B, G_B) - u_R(R_A, G_A)$$
$$\mathcal{F}(z) - \mathcal{F}(y) = \Delta_{yz} u = u_R(R_A, G_A) - u_R(R_C + 1, G_C)$$
$$\mathcal{F}(x) - \mathcal{F}(z) = \Delta_{zx} u = u_R(R_C, G_C) - u_R(R_B - 1, G_B).$$

Summing all these relations yields:

$$u_R(R_B, G_B) - u_R(R_B - 1, G_B) = u_R(R_C + 1, G_C) - u_R(R_C, G_C).$$

If the values of (R_B, G_B) and (R_C, G_C) were independent, it would be straightforward to prove that for all values of $(R,G) \in E_H$ such that $R \geq 1$:

$$u_R(R, G) - u_R(R-1, G) = const.$$

We leave it to the reader to verify that the restrictions on the values of (R_B, G_B) and (R_C, G_C) due to the overlapping of the neighborhoods of cells B and C do not change this result. Hence $\forall (R,G) \in E_H$, $\exists a \in \mathbb{R}$

$$u_R(R, G) = u_R(0, G) + aR. \tag{E.1}$$

The same kind of reasoning proves that $\forall (R,G) \in E_H, \exists d \in \mathbb{R}$:

$$u_G(R, G) = u_G(R, 0) + dG. \tag{E.2}$$

Suppose now as shown on Fig. E.1 that A, B, C and D are four cells of the lattice such that cells B and C are in each other's neighborhoods while the neighborhoods of cells A and D do not overlap the neighborhoods of any of the other three cells. Let x be a given state in which cell A is occupied by a red agent while cell D is occupied by a green agent and cells B and C are empty. We denote by $(R_A, G_A) \in E_H$, $(R_B, G_B) \in E_H$, $(R_C, G_C) \in E_H$ and $(R_D, G_D) \in E_H$ the number of red and green agents within the neighborhoods of cells A, B, C and D in state x. Let y be the state obtained when the red agent in cell A moves to cell B, z the state obtained if next the green agent in cell D moves to cell C, w the state obtained if the red agent in cell B next goes back
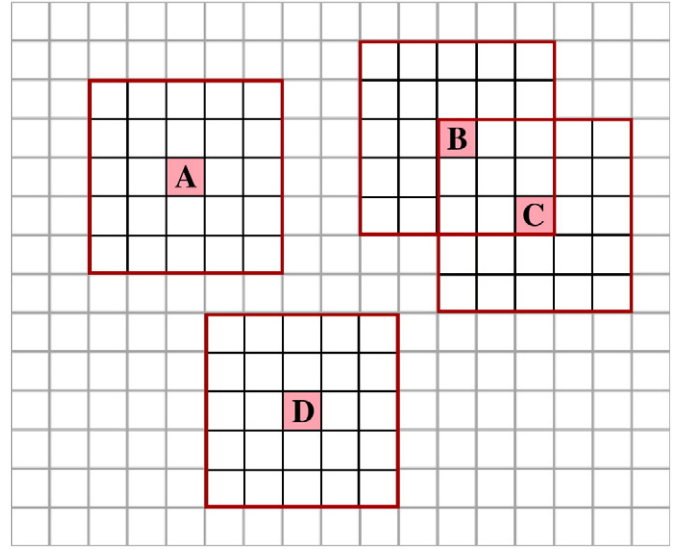

Fig. E.1. Example of configuration where the neighborhood of cells B and C overlap while those of cells A and D do not overlap with any of the other enlightened cells. The neighborhood of a cell is composed here by the $H=24$ nearest cells surrounding it.

to cell A. The system goes back from state w to state x by a move of the green agent in cell C to cell D. The function $\mathcal{F}$ being a potential function, the following relations hold (taking into account that the moves of the agents change the number of red agents in the neighborhoods of cells B and C):

$$\mathcal{F}(y) - \mathcal{F}(x) = \Delta_{xy} u = u_R(R_B, G_B) - u_R(R_A, G_A)$$
$$\mathcal{F}(z) - \mathcal{F}(y) = \Delta_{yz} u = u_G(R_C + 1, G_C) - u_G(R_D, G_D)$$
$$\mathcal{F}(x) - \mathcal{F}(z) = \Delta_{zx} u = u_R(R_A, G_A) - u_R(R_B, G_B + 1)$$
$$\mathcal{F}(x) - \mathcal{F}(z) = \Delta_{zx} u = u_G(R_D, G_D) - u_G(R_C, G_C).$$

Summing all these relations yields:

$$u_R(R_B, G_B + 1) - u_R(R_B, G_B) = u_G(R_C + 1, G_C) - u_G(R_C, G_C).$$

Once again, one can check that while the values of (R_B, G_B) and (R_C, G_C) are not independent, the intuitive results holds, which is $\forall (R,G) \in E_{H-1}, \exists b \in \mathbb{R}$:

$$u_R(R, G+1) - u_R(R, G) = u_G(R+1, G) - u_G(R, G) = b. \tag{E.3}$$

To summarize, the existence of the potential function imposes the existence of three real constants (a, b, d) such that the utility functions take the bilinear form:

$$u_R(R, G) = u_R(0, 0) + aR + bG$$
$$u_G(R, G) = u_G(0, 0) + bR + dG. \tag{E.4}$$

Reciprocal proof

Suppose that the utilities of the red and green agents are given by Eq. (E.4). The objective here is to demonstrate that these utility function being given, a potential function of the system exists.

Suppose that an agent moves from a neighborhood composed of R red and G green agents to a neighborhood composed of R' red and G' green agents. If the moving agent is red, then the variation of his utility can be written as

$$\Delta u = u_R(R', G') - u_R(R, G) = a(R' - R) + b(G' - G)$$

---

[25] Beware of the restrictions imposed on the values of (R_B, G_B) and (R_C, G_C) due to the overlapping of the neighborhoods of cells B and C.

while the variation of the collective utility is:

$$\Delta U = a(R'-R) + b(G'-G) \quad \text{which is the variation of the moving agent's utility}$$
$$+G'b - Gb \quad \text{which is the sum of the variation of the other green agents}$$
$$+R'a - Ra \quad \text{which is the sum of the variation of the other red agents}$$
$$= 2a\left(R'-R\right) + 2b\left(G'-G\right) = 2\Delta u.$$

Similarly, if the moving agent is green, the variation of his utility can be written as

$$\Delta u = u_G(R',G') - u_G(R,G) = b(R'-R) + d(G'-G)$$

while the variation of the collective utility is:

$$\Delta U = b(R'-R) + d(G'-G) \quad \text{which is the variation of the moving agent's utility}$$
$$+G'd - Gd \quad \text{which is the sum of the variation of the other green agents}$$
$$+R'b - Rb \quad \text{which is the sum of the variation of the other red agents}$$
$$= 2b\left(R'-R\right) + 2d\left(G'-G\right) = 2\Delta u.$$

Hence, whatever the starting configuration and whatever the move, the variation in utility of the moving agent is always half the variation in the collective utility, a global function depending on the configuration of the city. Hence, if the utility functions of the agents are given by Eq. (E.4), the function $\mathcal{F} : x \rightarrow U(x)/2$ is a potential function of the system.   □

References

Barr, J., Tassier, T., 2008. Segregation and strategic neighborhood interaction. Eastern Economic Journal 34, 480–503.

Blume, L.E., 1993. The statistical mechanics of best-response strategy revision. Games and Economic Behavior 11, 111–145.

Card, D., Mas, A., Rothstein, J., 2008. Tipping and the dynamics of segregation. Quarterly Journal of Economics 123, 177–218.

Clark, W.A.V., Fossett, M., 2008. Understanding the social context of the Schelling segregation model. Proceedings of the National Academy of Sciences 105, 4109–4114.

Cutler, D.M., Glaeser, E.L., Vigdor, J.L., 2008. Is the melting pot still hot? Explaining the resurgence of immigrant segregation. The Review of Economics and Statistics 90, 478–497.

Dokumaci, E., Sandholm, W.H., 2007. Schelling redux: an evolutionary dynamic model of residential segregation. Unpublished manuscript, University of Wisconsin-Madison.

Duncan, O.D., Duncan, B., 1955. A methodological analysis of segregation indexes. American Sociological Review 20, 210–217.

Fagiolo, G., Valente, M., Vriend, N.J., 2007. Segregation in networks. Journal of Economic Behavior and Organization 64, 316–336.

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

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

Grauwin, S., Bertin, E., Lemoy, R., Jensen, P., 2009a. Competition between collective and individual dynamics. Proceeding of the National Academy of Sciences USA 106, 20622–20626.

Grauwin, S., Goffette-Nagot, F., Jensen, P., 2009b. Dynamic models of residential segregation: brief review, analytical resolution and study of the introduction of coordination. GATE Working Paper. available at ftp://ftp.gate.cnrs.fr/RePEc/2009/0914.pdf.

Grauwin, S., Goffette-Nagot, F., Jensen, P., 2010. Dynamic models of residential segregation: an analytical solution. GATE Working Paper. available at ftp://ftp.gate.cnrs.fr/RePEc/2010/1017.pdf.

Iceland, J., Scopilliti, M., 2008. Immigrant residential segregation in US metropolitan areas. 1990–2000. Demography 45, 79–94.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics 76, 689–708.

Mobius, M.M., Rosenblat, T.S., 2000. The Formation of Ghettos as a Local Interaction Phenomenon. Unpublished manuscript, Harvard University.

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

Morrill, R.L., 1991. On the measure of spatial segregation. Geography Research Forum 11, 25–36.

O'Sullivan, A., 2009. Schelling's model revisited: residential sorting with competitive bidding for land. Regional Science and Urban Economics 39, 397–408.

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

Pollicott, M., Weiss, H., 2001. The dynamics of Schelling-type segregation models and a nonlinear graph Laplacian variational problem. Advances in Applied Mathematics 27, 17–40.

Reardon, S.F., O'Sullivan, D., 2004. Measures of spatial segregation. Sociological Methodology 34, 121–162.

Reardon, S.F., Matthews, S.A., O'Sullivan, D., Lee, B.A., Firebaugh, G., Farrell, C.R., Bischoff, K., 2008. The geographic scale of metropolitan racial segregation. Demography 45, 489–514.

Schelling, T.C., 1969. Models of segregation. American Economic Review, Papers and Proceedings 59, 488–493.

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.

Tiebout, C.M., 1958. A pure theory of local expenditures. The Journal of Political Economy 64, 416.

Vinkovic, D., Kirman, A., 2006. A physical analogue of the Schelling model. Proceedings of the National Academy of Sciences 103, 19261–19265.

Wong, D.W.S., 2005. Formulating a general spatial segregation measure. The Professional Geographer 57, 285–294.

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

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 54, 533–550.

Zhang, J., 2011. Tipping and residential segregation: a unified Schelling model. Journal of Regional Science 51, 167–193.