Japanese Journal of Political Science 16 (4), 553–573
doi:10.1017/S1468109915000304

From KISS to TASS Modeling: A Preliminary Analysis of the Segregation Model Incorporated with Spatial Data on Chicago

GAKU ITO
Department of Advanced Social and International Studies, the Graduate School of Arts and Sciences, the University of Tokyo
gakuatgu@gmail.com

SUSUMU YAMAKAGE
Professor of International Relations, Aoyama Gakuin University, Tokyo
yamakage@sipeb.aoyama.ac.jp

Abstract

The 'keep it simple, stupid' slogan, or the KISS principle has been the basic guideline in agent-based modeling (ABM). While the KISS principle or parsimony is vital in modeling attempts, conventional agent-based models remain abstract and are rarely incorporated or validated with empirical data, leaving the links between theoretical models and empirical phenomena rather loose. This article reexamines the KISS principle and discusses the recent modeling attempts that incorporate and validate agent-based models with spatial (geo-referenced) data, moving beyond the KISS principle. This article also provides a working example of such time and space specified (TASS) agent-based models that incorporates Schelling's (1971) classic model of residential segregation with detailed geo-referenced demographic data on the city of Chicago derived from the US Census 2010.

1 Introduction

As agent-based modeling (ABM) has been increasingly applied to the study of social phenomena, there has been more empirical approaches than ever. During early days of ABM, exploratory attempts were extremely abstract with rich interpretations. It may be useful to make a list, which will surely become very lengthy, of the application of the Ising model, for instance, to various types of phenomena over recent decades (McCoy and Wu 2014). Without doubt, a simple model may be applicable to, and able to explain, a large number of phenomena that seem unrelated to one another. The making of simple models was not only advised but also accepted in order to extract the essence of various research activities. The most famous maxim in the field must be KISS ('Keep It Simple, Stupid') that was first introduced by Robert Axelrod (1997a: 4–5) to the field.

The current trend of ABM looks at a departure from the KISS principle. As such, KISS may be replaced by 'TASS', which stands for 'Time And Space Specified'. In either theory-oriented or history-oriented studies, the empirical approach requires some kind of empirical data that specifically represent a certain point or range of both time and space. In this sense, the introduction of empirical data into ABM may be characterized as TASS modeling.

This article aims to demonstrate the promising possibilities of the change of emphasis from KISS to TASS in ABM applications in social sciences. Namely, this article adopts the empirical approach employing ABM that used to be built based on KISS. Although the ways to specify time and space are numerous, this article focuses on the spatial dimension and attempts to analyze a KISS model with geographical data that are constructed from geographic information systems (GIS). While spatial analysis is widely attempted in ABM, because of the common setting of interaction among agents on a two-dimensional plane, it is difficult (actually time-consuming and tedious) at best to import spatial data into the computational model, even if such data are available from GIS. Some tools that incorporate ABM with spatial data are needed in order to build a TASS agent-based model. As for the tool for ABM, it is customary for the authors to use the ABM-specific simulator named artisoc developed by the Kozo Keikaku Engineering (KKE). Recently, KKE developed the GIS data converter for artisoc, which converts various types of spatial data into the spatial variables compatible with artisoc, which allows for automatic data transformation and easier use of spatial data in ABM. The empirical analyses in this article are made possible thanks to it.[1]

The KISS model that is used here is the segregation model developed by Thomas Schelling nearly half a century ago (Schelling, 1969, 1971, 1978). A pioneering and monumental attempt at ABM, the model is so well-known that it must be unnecessary to describe it here (see Clark and Fossett, 2008; Hatna and Benenson, 2014; Vinkovic and Kirman, 2006, for extensive reviews). Because the spatial structure of the segregation model is so simple that only the neighborhood is defined, it may be necessary to add some spatial variables in order to apply the model to the empirical analysis. The element of TASS in this article is geo-referenced (spatial) census data of the metropolitan Chicago area. Because Chicago is one of the cities in which residential segregation is clearly observed, it is an appropriate case to explicitly incorporate the KISS-type Schelling's segregation model with empirical data. Although analyses in this article are preliminary, they hopefully demonstrate the promising possibility of TASS modeling.

In the next section, existing applications of the ABM technique and their empirical turn are surveyed. Then, the classic Schelling's model of residential segregation is

[1] KKE GIS data converter for artisoc is currently in beta. We thank KKE for providing unlimited access to their software and answering numerous questions.

incorporated with newly available fine-grained demographic data to demonstrate the methodological utility of the TASS-type ABM. Preliminary analyses is followed by discussion and conclusions concerning the application of this empirically explicit computational modeling approach to social sciences.

2 From the KISS to the TASS principle

The micro–macro link, or the relationship between individual behavior at the micro level and collective outcomes at the macro level sits at the center of numerous scholarly debates in social science (Coleman, 1990). Theories of social science are often considered to be more informative if they account for how and why individual actions and interactions add up to collective behavior, or the causal mechanisms/relationships between 'micromotives' and 'macrobehavior' (Bruch and Mare, 2006: 667–668). As Schelling (1978) argues:

> Sometimes they are not easily guessed. Sometimes the analysis is difficult. Sometimes it is inconclusive. But even inconclusive analysis can warn against jumping to conclusions about individual intentions from observations of aggregates, or jumping to conclusions about the behavior of aggregates from what one knows or can guess about individual intentions. (p. 14)

Agent-based modeling (ABM) has been used to simulate and specify the relationship between autonomous actions and interactions of individuals and the dynamics and behavior of the whole system. Cederman (2005: 873) defines ABM as a 'computational methodology that allows scientists to create, analyze, and experiment with artificial worlds populated by agents that interact in nontrivial ways and that constitute their own environment'. In ABM, researchers code the actions and interactions of constituent elements of a system, called agents, into a program and examine what systemic or collective outcomes emerge through computational simulations. As such, ABM enables us to systematically explore whether, how, and why some specific micro-level mechanisms generate a class of macro-level outcomes not reducible to properties of the constituent agents (Axelrod, 1997a; Cederman, 2005; Epstein and Axtell, 1996; Macy and Willer, 2002).

2.1 Agent-based modeling and the KISS principle

A model serves as a purified experiment, and ABM technique provides a 'computer laboratory for virtual experimentation' (Crooks *et al.*, 2008: 418). Among the pioneering works, Axelrod (1997a: 4–5) argues:

> Although agent-based modeling employs simulation, it does not necessarily aim to provide an accurate representation of a particular empirical application. Instead, the goal of agent-based modeling is to enrich our understanding of fundamental processes that may appear in a variety of applications. This requires adhering to the KISS principle, which stands for the army slogan 'keep it simple, stupid'.

Most existing ABM applications adhere to the KISS principle and have specified the micro-mechanisms sufficient to generate macro-outcomes such as residential segregation, evolution of cooperation in the multi-player Iterated Prisoners Dilemma game, and dissemination of culture (Axelrod, 1984, 1997a, b; Schelling, 1969, 1971, 1978). In a sense, the concepts of ABM in social science date back to the 1970s, and a pioneering contribution must be the work of Thomas Schelling (1969, 1971, 1978) who developed a model of residential segregation by moving pennies and dimes on a chessboard. Assuming that coins represent householders of different race seeking to reside amongst one's own kind, he demonstrated that marked segregation within neighborhoods could emerge from seemingly mild, not strong, individual preferences (see Sakoda, 1971, for similar games). Having yielded the counterintuitive insight, Schelling's parsimonious model is now regarded as a 'theoretical basis' for the scholarly debate on the causes of residential segregation (Clark and Fossett, 2008: 4109).

The last few decades have witnessed an explosion of ABM applications. Political scientists and sociologists have examined and extended existing theories through ABM (e.g., Axelrod, 1997a; Bhavnani *et al.*, 2009; Bremer and Mihalka, 1977; Cederman, 1997, 2002, 2005; Cusack and Stoll, 1990; Johnson *et al.*, 2011; Laver and Sergenti, 2012; Lustick *et al.*, 2004; Schrodt, 1981; Siegel, 2009, 2011). Physicists have developed agent-based models to analyze social phenomena (see Castellano *et al.*, 2009; Helbing *et al.*, 2015, for comprehensive reviews). Because of the increasing number of modeling attempts, there is now more confidence that ABM is a 'valid technical methodology' for scientific inquiry (Janssen and Ostrom, 2006).

2.2 Time and space specified (TASS) modeling

The KISS principle stresses simplicity rather than realism. Adhering to the KISS principle, most conventional ABM applications remain abstract and are rarely tested against empirical data (Janssen and Ostrom, 2006; Lustick and Miodownik, 2009). However, overly abstract models may lead to arbitrary interpretations and thus bring into question the appropriateness of ABM in explaining known empirical phenomena. For example, Axelrod (1997b) uses the simulation outcomes of a single agent-based model named dissemination of culture to explain various phenomena at different analytical levels such as the process of national integration, the survival of a heterogeneous minority surrounded by overwhelming majority, and regional integration seen in Europe. Researchers skeptical of ABM often acutely criticize this problem (Mitsutsuji and Yamakage, 2013). While parsimony is vital in modeling attempts, models must also have explicit connections with empirical observations to enrich our understanding of generating processes underlying social phenomena and address such criticism. Otherwise, it may be extremely difficult to test, evaluate, or validate agent-based models using empirical data. Consequently, a natural question emerges: Whether and to what extent can the model predictions explain empirical phenomena?

Such a speculation leads to the idea of incorporating agent-based models with empirical data to tighten the connections between models and empirical phenomena, or to make models time and space specified (TASS). Rather than yielding purely theoretical insights, TASS modeling may identify micro-mechanisms sufficient to generate macro-outcomes consistent with observed ones. While no computational models can perfectly 're-run' history, such models can serve as heuristic devices that enhance our understanding of generating processes underlying empirical phenomena (Janssen and Ostrom, 2006; Lustick and Miodownik, 2009; Weidmann and Salehyan, 2013).

Although the ways to specify time and space are numerous, increasingly available spatial data and the recent developments of geographic information systems (GIS) provide an opportunity to tighten the connection between purified models and empirical phenomena through the use of geo-referenced empirical data (Crooks *et al.*, 2008; Lustick and Miodownik, 2009; O'Sullivan, 2008; Torrens and Benenson, 2005). Indeed, some researchers have started to use increasingly available spatial data to test, evaluate, and validate their models with empirical observations and thereby have moved beyond the KISS principle. Initial contributions by sociologists include models that incorporate Schelling's segregation model with geo-referenced demographic and dwelling price data (e.g., Benenson *et al.*, 2002, 2009; Bruch, 2014; Crooks, 2008; Yin, 2009). For example, Benenson *et al.* (2002, 2009) and Yin (2009) show that modified or generalized versions of Schelling's segregation model were applicable to and explainable of segregation patterns observed in real cities using detailed spatial data on the Tel Aviv and Buffalo metropolitan areas respectively. These results illustrate that that observed patterns of residential segregation could plausibly arise from mild racial and economic preferences.

Political scientists and physicists have also developed another class of TASS models to examine the determinants of violence in civil war (e.g., Bhavnani and Choi, 2012; Bhavnani *et al.*, 2011, 2014; Lim *et al.*, 2007; Sakamoto, 2013; Weidmann and Salehyan, 2013). Lim *et al.* (2007) and Weidmann and Salehyan (2013) propose agent-based models similar to Schelling's model incorporating detailed spatial data of ethnic geography and analyzing the patterns of violence and segregation in the former Yugoslavia and in Iraq. They demonstrate that a simple mechanism of ethnically and/or security motivated migration and subsequent violence accounts for the spatial distributions of violence in actual conflicts and successfully predicts them. Such empirically plausible models could also serve as heuristic devices to explore the possible effects of proposed policy efforts. Having optimized model parameters such that the patterns of violence from the simulation closely fit the actual distributions, Bhavnani *et al.* (2014) use an agent-based model to assess how different levels of Israeli–Palestinian segregation would shape future violence in Jerusalem.

These modeling attempts have successfully specified the micro-mechanisms sufficient to generate empirically observed macro-outcomes and demonstrated that empirically grounded agent-based models are extremely useful in identifying and testing causal mechanism. Incorporating agent-based models with spatial data seems to be a promising way to go beyond the KISS principle toward the TASS principle.

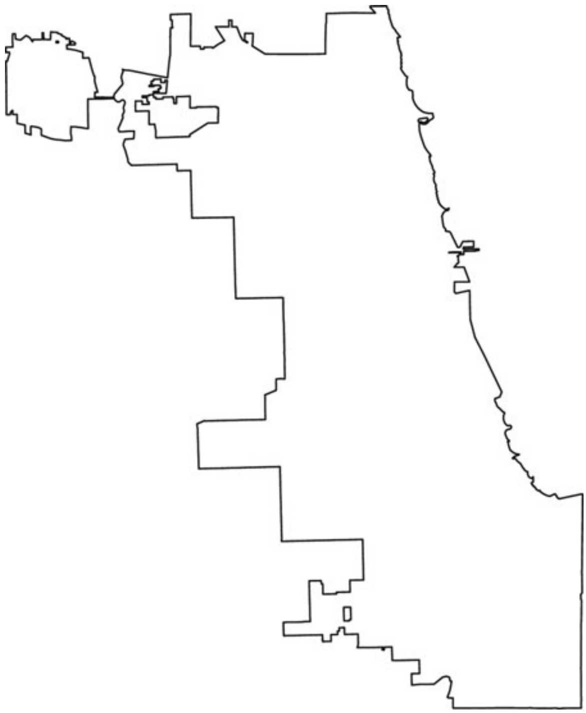
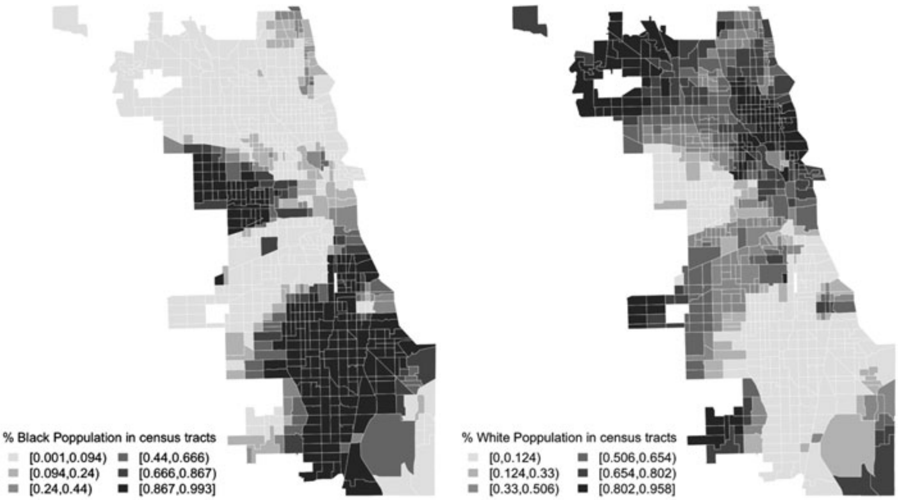


Figure 1. City-level administrative boundary of Chicago

3 A working example: applying Schelling's model of segregation to Chicago

This section provides a working example of TASS modeling incorporating Schelling's (1969, 1971, 1978) classic model of segregation with geo-referenced demographic data on Chicago derived from the US Census 2010. The city of Chicago is not only one of the most diverse cities, but also one of the most racially segregated cities in the United States. Figures 1 and 2 represent the city boundary and the population distribution among the 2010 Census tracts of Chicago, respectively. As Charles (2003) and Logan *et al.* (2011) argue, many of the large and ethnically diverse US cities remain highly, though not completely, segregated. Chicago contains not only highly homogeneous segregated clusters but also heterogeneous integrated mosaics as shown in Figure 2.

One might think that individual residents must have strong racial preferences for this segregation pattern to emerge. However, this intuition does not always hold. Survey evidence suggests that while individuals do favor residing with neighbors of the same race, such preferences are rather modest (e.g., Bruch and Mare, 2006; Clark, 1991; Clark and Fossett, 2008; Emerson *et al.*, 2001; Farley *et al.*, 1978; Krysan *et al.*, 2009; Xie and Zhou, 2012). This is a gap in search of a theory. As discussed in detail



(A) % Black population within tracts (B) % White population within tracts

Figure 2. Spatial distribution of Black/White population in Chicago
Notes: Data are derived from the US Census 2010. Note that the outer limit of census tracts is slightly different from the city-level administration boundary (Figure 1). Darker shade represents a higher proportion of Black/White population (P0030002/P0030003) over the total population (P0030001) within the tract, and solid lines represent census tract boundaries. Census tracts containing NAs for any of total (P0030001), White (P0030002), or Black population (P0030003) are dropped.

below, Schelling's segregation model provides a possible explanation to this seemingly puzzling phenomenon.[2]

3.1 Technical challenges

While the recent empirical turn in ABM is a welcome trend, it currently takes in-depth ABM and GIS skills to be able to exploit opportunities offered by TASS modeling. Nonetheless, the recent developments in software provide a solution to this technical challenge, at least partly. Our model is written in artisoc, a user-friendly agent-based toolkit developed by the Kozo Keikaku Enterprise (KKE). The KKE GIS data converter for artisoc is used to couple our agent-based model with spatial data on Chicago. The software allows for easier building of agent-based models and handling of spatial data in ABM. In the current context, high-resolution geo-referenced demographic data provided by the US Census 2010 are extremely useful and freely available from the US National Historical Geographic Information System (NHGIS, https://www.nhgis.org/).

[2] The scholarly debate on the causes of urban residential segregation mainly focuses on three themes: racial discrimination, own-race preference, and economic differences (Yin, 2009: 2749–50). Adhering to the original Schelling's model, we focus on the second cause. See Benenson *et al.* (2009); Yin (2009); Zhang (2004) for models incorporating economic factors.

While combining and converting spatial data may be another technical challenge, a series of open-source packages named UScensus-suite (Almquist, 2010) for R allows for convenient handling of the US Census data. As R is one of the most popular statistical software packages among students of social science, Almquist (2010) is a welcome addition to the spatial econometrics/statistics packages released on CRAN (http://cran.r-project.org/; see Bivand *et al.*, 2013 and CRAN Task View: Analysis of Spatial Data at http://cran.r-project.org/web/views/Spatial.html, for further details). Our empirical data are derived from the US Census 2010 and handled using R and GIS data converter for artisoc.

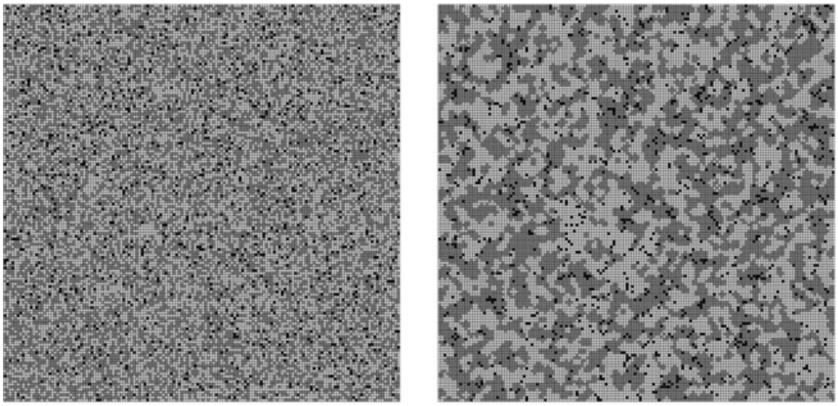
Offering a comprehensive explanation for residential segregation in Chicago is beyond the scope of this article. Instead, we aim to demonstrate how TASS modeling helps to tighten the connection between theoretical models and empirical phenomena and allows for easier empirical evaluation of models. We present a working example in the following way. First, we replicate Schelling's segregation model using a hypothetical square lattice and examine simulation outcomes. Second, we replace the hypothetical lattice with more a specified one representing the 2010 Census tracts within the Chicago city boundary. This modification enables us to explore whether and how well the classic segregation model is able to generate spatial segregation patterns consistent with the observed one as Schelling (1969: 488) himself once implied.

3.2 Replicating Schelling's model of segregation

Schelling's (1969, 1971, 1978) model considers spatial distribution of racial groups as an outcome of repeated interactions between individual householders. Our baseline agent-based model replicates the original model on a two-dimensional 150 × 150 square lattice; that is, a grid with 225,000 cells as shown in Figure 3(A). The lattice is populated with 'penny' (dark gray) and 'dime' (light gray) agents that represent householders of different racial groups. A small proportion of grid-cells are left unpopulated. Each agent can occupy one grid-cell, or a 'housing unit', at a discrete moment of time step and move to another unpopulated cell when she/he is not satisfied with the current position. Because agents are assumed to have mild preferences for living amongst like agents, their decisions whether or not to move to another cell depend on the number of like and unlike agents among neighborhoods.

To describe the model more formally, let **Ag** denote the set of agents belonging to one of two mutually avoiding groups $g \in \{\text{penny}, \text{dime}\}$. Each agent $Ag_i \in \mathbf{Ag}$ is randomly placed on a different cell at the beginning of a simulation run. A model parameter π determines the proportion of populated cells. Once the lattice has been populated with agents, the model proceeds in a series of time steps t.

For simplicity, we assume that (1) there are 50% penny and 50% dime agents on the lattice, (2) the number of agents is held for the entire model period, (3) preferences of agents are held for the entire model period, (4) agents' groups are observable and thus Ag_i can perfectly distinguish neighbor agents belonging to g from those not belonging to g, and (5) agents possess knowledge about all the neighbors surrounding every



(A) $t = 1$ (initial state) (B) $t = 48$ (equilibrium state)

Figure 3. Schelling's model of segregation on a two-dimensional 150 × 150 lattice
Notes: Panel (A) shows the initial spatial configuration of 'penny' and 'dime' agents. The lattice becomes highly segregated at time step 48 as panel (B) shows. Dark gray and light gray cells represent penny and dime agents respectively. Black cells are currently unoccupied ($\pi = 0.95$).

agent.[3] Following Benenson *et al.* (2009, 468–470), the agents' behavior algorithm is defined as follows:

(1) At every time step t, each agent Ag_i belonging to a group g observes the set of neighborhood cells $\mathcal{N}_j$ around the currently located cell j and computes the fraction s_{jt} of the cells occupied by agents belonging to g among $\mathcal{N}_j$.[4]

(2) If $s_{jt} <$ predetermined threshold (common racial preference) θ, then Ag_i decides to relocate; otherwise, it decides to stay.

(3) If Ag_i decides to relocate, it randomly chooses a currently unoccupied cell k and runs rule (1) to compute s_{kt}.

(4) If $s_{kt} > s_{jt}$, Ag_i moves to cell k; otherwise, it stays.

(5) Previously occupied cell j becomes unoccupied and available to other agents.

Once all agents have executed the behavior rule, time step count increases to $t + 1$. A simulation run continues until the spatial distribution of agents reaches either (1) an equilibrium where all agents are satisfied with their current locations or (2) a dynamic equilibrium where agents may still be dissatisfied and moving but the composition of neighborhoods is no longer changing, or (3) time step count t reaches to T. We set $T = 10,000$ and the proportion of populated cells $\pi = 0.95$ for the following simulation

[3] These assumptions and parameter choices can easily be relaxed or modified. See, for example, Benenson *et al.* (2009); Bruch and Mare (2006); Clark and Fossett (2008); Crooks (2008); Fossett and Waren (2005); Hatna and Benenson (2014); Laurie and Jaggi (2003); Vinkovic and Kirman (2006); Zhang (2004).

[4] Formally, $s_{jt} = \frac{\text{\#agents belonging to } g \text{ within } \mathcal{N}_j}{\text{\#agents belonging to } g \text{ within } \mathcal{N}_j + \text{\#agents not belonging to } g \text{ within } \mathcal{N}_j}$. Note that there is no contribution from empty cells.

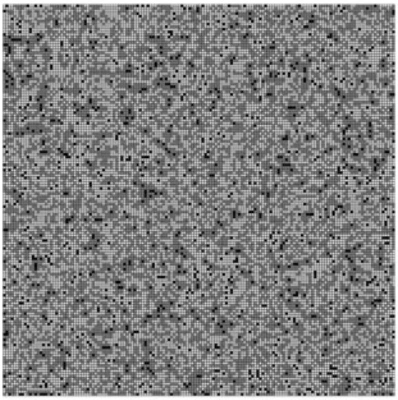
runs. The above representation is slightly modified and generalized compared with the original model (Benenson *et al.*, 2009: 468–70). The original model assumes that preference threshold $\theta = 1/3$ and the size and the form of local neighborhoods each agent cares about $\mathcal{N}$ is a 3×3 Moore neighborhood (eight nearest cells). While further generalization or modification is possible, it is known that simulation outcomes critically depend on these two parameters (e.g., Bruch and Mare, 2006; Clark and Fossett, 2008; Crooks, 2008; Hatna and Benenson, 2014; Laurie and Jaggi, 2003; Vinkovic and Kirman, 2006). Note that the baseline value of θ is quite small, meaning that both penny and dime agents have mild preferences and thus tolerate a large proportion of unlike neighbors among their neighborhoods.

Figure 3 (B) shows a typical simulation outcome with θ and $\mathcal{N}$ set as above. The figure illustrates that a highly segregated pattern at the macro-level, a phenomenon not directly foreseen from the individual motives, could emerge from mild segregation preferences at the micro-level (Schelling, 1978; 148–55). Fraction s_{jt} defined in rule (1) serves as a measure of segregation for a given agent at time step t, and the measure of global segregation level at t, S_t, can be given by calculating the mean fraction value. Because agents are randomly placed at the beginning of a simulation run, the initial segregation level S_1 roughly equals 0.5. In the illustrative run shown in Figure 3, S_{48} increases to 0.763 at time step 48 when all agents are satisfied, suggesting that mild individual preferences to segregate are sufficient for global segregation.

In Figures 4 and 5, we vary two main model parameters, the preference threshold θ and the definition of neighborhoods $\mathcal{N}$, to explore whether and how agents' segregation preferences and neighborhood scopes influence the spatial patterns of segregation. Three results emerge. First, as easily expected, the square-lattice becomes more segregated as θ increases, meaning that stronger segregation preferences at the micro level lead to higher segregation at the macro level as shown in Figure 4.[5] Second, the resulting segregation level is not very sensitive to the definition of $\mathcal{N}$. The global segregation level at time step T, S_T, remains 0.761 and 0.729 for panels (A) and (B) of Figure 5 respectively. Table 1 reports the mean resulting segregation level$\langle S_T \rangle$ over 300 replications for each parameter setting with different random seeds, which confirms that the results shown in Figure 5 are not accidental.[6] Finally and qualitatively, as the size of $\mathcal{N}$ increases so do the sizes of homogeneous clusters (of agents belonging to the same group) *and* heterogeneous areas. While a simulation run with a small $\mathcal{N}$ converges to segregation in a long run (Figures 3 and 4), a simulation run with a

[5] More generally, it is known that Schelling's segregation model exhibits a tipping point behavior: The spatial configuration of agents remain integrated for $\theta < \theta_{critical}$, while it converges to segregation for $\theta \geq \theta_{critical}$, with $\theta_{critical} \sim 1/3$ (Hatna and Benenson, 2014, 2).

[6] Another tendency apparent from Table 1 is that the standard deviations of S_T positively correlate with the neighborhood size $\mathcal{N}$, meaning that the levels of segregation tend to vary more in simulation runs with greater $\mathcal{N}$.


(A) $\theta = 1/4$

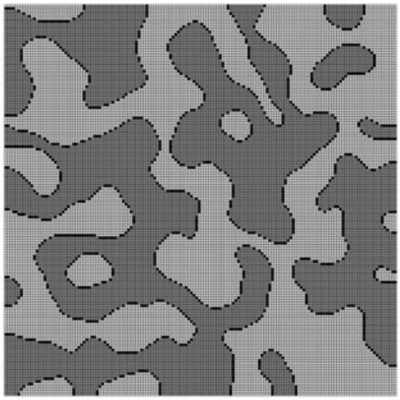

(B) $\theta = 3/4$

Figure 4. Representative cases with different values of racial preference threshold θ
Notes: Panel (A) shows a typical simulation outcome with highly tolerant agents and panel (B) represents another outcome with highly intolerant agents. A greater θ means a stronger preference for neighbors of the same kind. The two panels demonstrate that both resulting segregation level and the sizes of homogeneous clusters are positively correlated with preference threshold θ.

Table 1. *300 replications of simulation runs for each parameter setting*

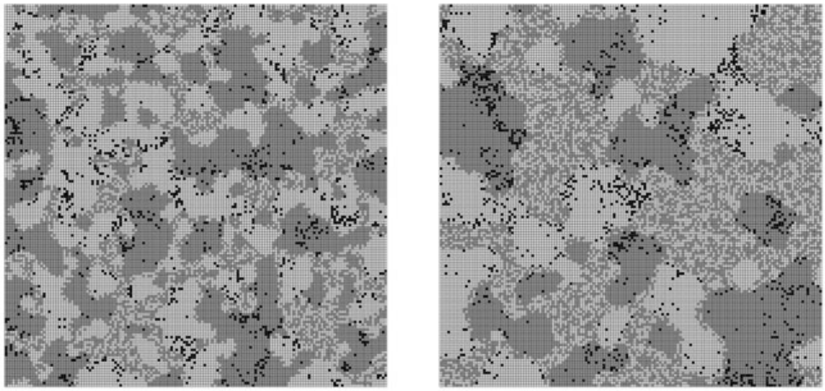
	Parameter values				
	$\theta = 1/3$,	$\theta = 1/4$,	$\theta = 3/4$,	$\theta = 1/3$,	$\theta = 1/3$,
	$\mathcal{N} = 3 \times 3^a$	$\mathcal{N} = 3 \times 3$	$\mathcal{N} = 3 \times 3$	$\mathcal{N} = 5 \times 5$	$\mathcal{N} = 7 \times 7$
$\langle S_T \rangle$	0.757(0.004)	0.566(0.003)	0.980(0.001)	0.752(0.014)	0.741(0.088)

Note: Standard deviations in parentheses.
[a]Baseline setting.

greater $\mathcal{N}$ yields a mixed spatial pattern that contains both segregated patches for each group as well as integrated patches where both groups coexist (Figure 5).[7]

These results suggest that the *levels* of segregation mainly depends on the preference threshold θ while the *forms* of segregation is highly sensitive to the neighborhood definition $\mathcal{N}$.[8] Because real cities are rarely completely segregated but rather exhibit mixed patterns that consist of both homogeneous clusters and heterogeneous mosaics, the simple exercise indicates that a segregation model with a greater $\mathcal{N}$ is likely to be a better heuristic tool to enrich our understanding of empirical phenomena than the original model.

[7] See Fossett and Waren (2005) and Laurie and Jaggi (2003), for more comprehensive examinations on the influence of neighborhood definitions.
[8] See Crooks (2008) for a similar experiment using a vector (continuous) space.



(A) $\mathcal{N}$ = a 5×5 Moore neighborhood (B) $\mathcal{N}$ = a 7×7 Moore neighborhood

Figure 5. Representative cases with different sizes of neighborhoods, $\mathcal{N}$
Notes: Two panels show typical simulation outcomes with $\mathcal{N}$ a 5×5 Moore neighborhood (24 nearest cells) and $\mathcal{N}$ = a 7×7 Moore neighborhood (48 nearest cells) respectively, with $\theta = 1/3$. The number of neighborhood cells agents observe increases as the size $\mathcal{N}$ increases. The sizes of both homogeneous (segregated) clusters and heterogeneous (integrated) mosaics increase as the size of $\mathcal{N}$ increases

3.3 Incorporating the segregation model with spatial data

Although the spatial data shown in Figures 1 and 2 are available in vector polygon form, we converted (rasterized) the corresponding census-tract data into numbers of grid-cells in order to specify the model's context because artisoc currently does not work with polygon data.[9] The edge length of each grid-cell roughly corresponds to 150 meters, and the whole Chicago-like lattice consists of 26,507 grid-cells as shown in Figure 6. This spatial modification allows for comparing simulation outcomes with empirical observations. For simplicity, we only consider Black-White neighborhood segregation and agents in our model represent householders of each racial group.[10] As in the simulation runs above, 95% of grid-cells are populated with agents, and the same number (50%) of penny and dime agents are randomly created. We set preference threshold $\theta = 1/3$ and $\mathcal{N}$ = a 7×7 Moore neighborhood. Model parameters π and T are also set as in the previous subsection, i.e., $\pi = 0.95$ and $T = 10,000$.

As shown in Figure 2, the Black-White segregation pattern observed in Chicago is best described as a mixture of highly segregated Black/White clusters and heterogeneous

[9] There are four different types of spatial data models: point, line, polygon, and grid. The first three are vector (continuous) data models and represent entities as exactly as possible. The final data model is a raster model and represents originally continuous surfaces by using a regular tessellation (Bivand *et al.*, 2013: 8–10). The grid-cell representation of the city boundary is directly comparable to the original Schelling's model. The 1990, 2000, and 2010 US Census datasets are available in a polygon form with corresponding demographic data (Almquist, 2010). See Benenson *et al.* (2002, 2009) and Crooks (2008) for polygon/vector-based Schelling's segregation model.

[10] Black and White population accounts for roughly 78% (2,120,659/2,718,590) of the total population of the city of Chicago in the 2010 Census.

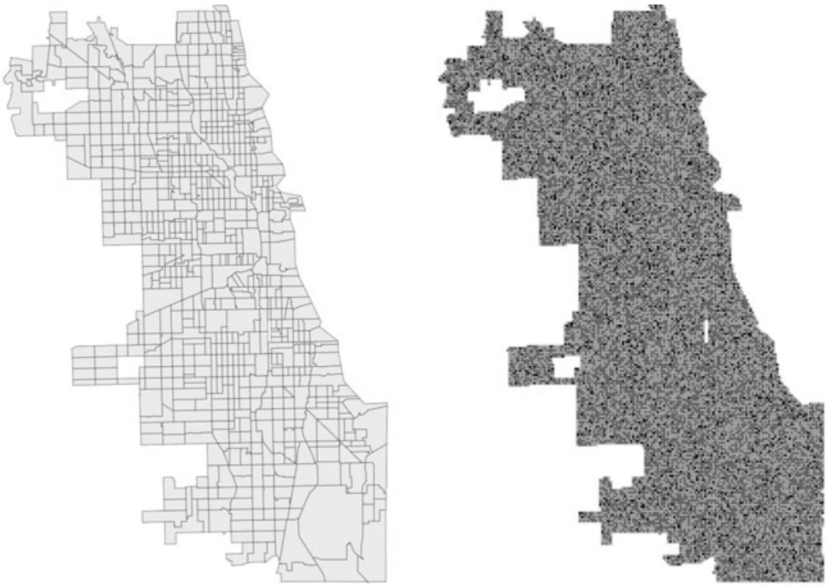


Figure 6. Grid-cell representation of the city of Chicago
Notes: Polygon-form spatial data (Panel (A)) was converted into grid-cell (raster) form (Panel (B)) using GIS data converter for artisoc. The edge length of each grid-cell roughly corresponds to 150 meters. Black cells represent empty cells. Dark gray and light gray cells represent penny and dime agents respectively. A geographically isolated tract (Federal Information Processing Standards (FIPS) code = 770602) is dropped for convenience.

mosaics where both groups coexist, or a spatial pattern of mixed segregation and integration. Figure 7 shows twelve illustrative simulation runs. As the simulation outcomes demonstrates, our specified model yields spatial segregation patterns that contain segregated patches for each agent group as well as patches where both groups coexist within the Chicago boundary. In this sense, the specified version of the segregation model seems to be able to generate the observed pattern of segregation.

Indeed, we need a more comprehensive replication and statistical tests on agreements between simulation outcomes and empirical data in order to formally test and validate these preliminary results. Nonetheless, these illustrative runs visually suggest that the combination of a mild level of segregation preferences and a specific neighborhood size may be sufficient to generate the residential segregation observed in Chicago. The original Schelling's model is explainable of the level of segregation but not the form of segregation empirically observed. A slightly modified model with a greater neighborhood size is explainable of both the level and the form of segregation, but it is not obvious whether the model is applicable to the real-world cities. Our specified model demonstrates that the modified model may be applicable to the residential segregation

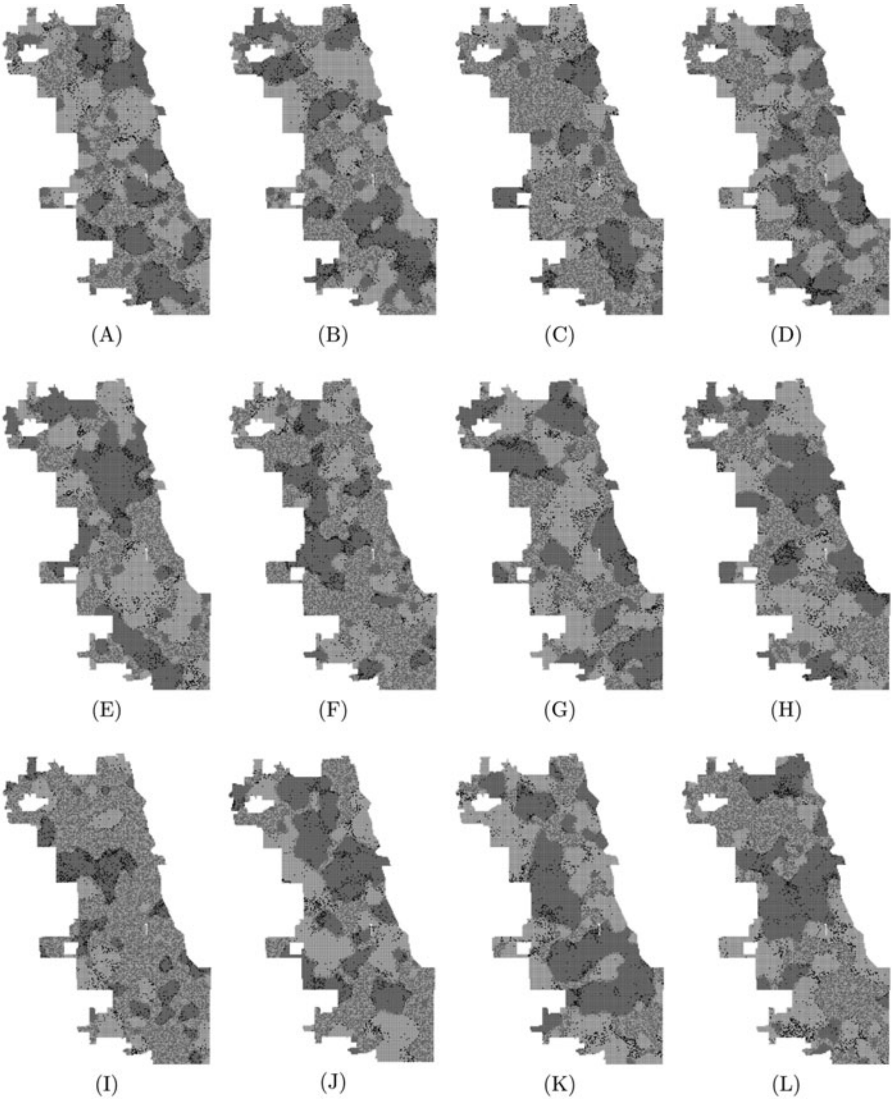


Figure 7. Twelve illustrative simulation runs of the specified segregation model.
Notes: Each panel shows the spatial distribution of agents when it reaches to an equilibrium where all agents are satisfied with their current positions.

observed in Chicago. KISS modeling enables the first two evaluations but not the last. Tightening the connection between the model and the empirical observation, TASS modeling makes such an evaluation possible.

One limitation of current model is that the spatial distribution of local clusters varies from one simulation run to another as shown in Figure 7. Because agents are randomly put at the beginning of each simulation run and the specific locations of clusters depend of the initial spatial configuration of agents, our model is not able to constantly generate a specific spatial distributions of agents similar to the observed segregation pattern. It seems that we need some deterministic and exogenous elements coupled with our model in order to constantly generate such a specific distribution of segregated clusters.

Although there are numerous possible extensions such as introduction of economic motives, geographic/physical barriers, 'quarter' agents representing Hispanics or Asians, and variation in the agents' tolerance thresholds, an arbitrary 'tuning' of the initial spatial configuration of agents is sufficient to generate clusters in a specific part of the city.[11] Specifically, we set some 'core communities' of agents belonging to each group in the initial state. The agents constituting these communities behave as do other agents – the only difference is their initial spatial locations. We select roughly 10% (80/801) of census tracts dominated by either Blacks or Whites, and fill the corresponding grid-cells with penny (representing Blacks) or dime agents (representing Whites) respectively, as Figure 8 shows.[12]

Figure 9 shows 12 illustrative simulation runs of the specified model with predetermined core communities. As shown in each panel, simulation outcomes have successfully generated segregation patterns with clusters located at specific locations. Gravitating toward the preexisting core communities, segregated clusters populated penny (dark gray) agents tend to form at the bottom and the center-left part of the lattice, and dime (light gray) clusters tend to form at the upper-left corner and the center of the lattice. This spatial distribution of segregated clusters, together with the general spatial segregation pattern of mixed segregation and integration, bear some resemblance to the observed spatial pattern.

Indeed, such an arbitrary tuning itself tells us little about the underlying process without a reasonable interpretation of the core communities and their initial locations. Incorporating the segregation model with additional spatial data of geographycal/physical barriers, dwelling prices, or empirical evidence on segregation preferences of different racial groups might yield more theoretically and empirically relevant insights. However, providing such a comprehensive examination of the causes

[11] See Benenson *et al.* (2002, 2009); Clark and Fossett (2008); Crooks (2008); Yin (2009); Zhang (2004), for modified segregation models coupled with economic motives and geographic/physical barriers. See Bruch and Mare (2006); Clark (1991); Clark and Fossett (2008); Emerson *et al.* (2001); Farley *et al.* (1978); Krysan *et al.* (2009); Xie and Zhou (2012), for empirical evidence on segregation preferences. Because it is known that variation in the (distribution of) agents' thresholds influences simulation outcomes significantly (e.g., Hatna and Benenson, 2014), incorporating Schelling's segregation model with survey data on segregation preferences, in addition to spatial data as presented in this article, is another promising way to specify the model's temporal and spatial context (e.g., Clark and Fossett, 2008).

[12] To implement this modification, we select 40 Black-dominated and 40 White-dominated census tracts with > 98% Black/White population share out of 801 tracts. The corresponding grid-cells are filled with either penny or dime agents at the beginning of a simulation run.

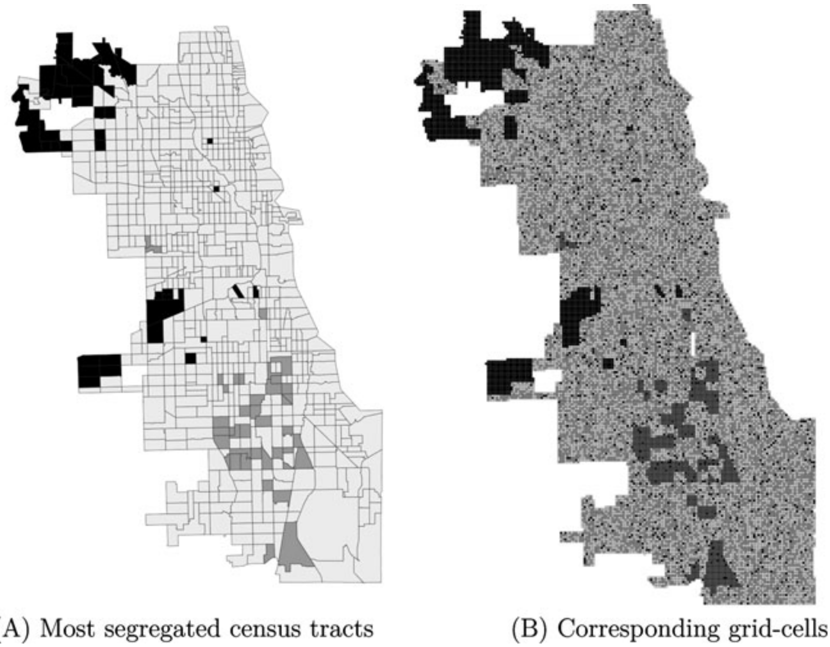


Figure 8. Distribution of the 10% most segregated census tracts and the corresponding grid-cells ('core communities').

Notes: Panel (A) shows the spatial distribution of 80/801 most segregated census tracts dominated by either Blacks (dark gray) or Whites (black). Panel (B) shows the corresponding grid-cell representation and the spatial distribution of penny agents representing Blacks (dark gray) and dime agents representing Whites (black) at the beginning of a simulation run.

of residential segregation in Chicago is beyond the scope of this article. Rather, our preliminary exercise aims to demonstrate how simple, but empirically explicit agent-based models, can serve as a heuristic that help researchers examine potential inferential leverage of abstract model on specific empirical cases. Indeed, the exercise here implies that the slightly modified version of Schelling's segregation model serves as a useful heuristic to specify micro-level mechanisms *sufficient*, if not necessary, to generate the observed pattern of residential segregation in the city of Chicago. Extending the model such that it incorporates the model with individual-level census data or other fine-grained empirical data is likely to help researchers further clarify and test the micro-level mechanisms underlying the macro-level patterns of residential segregation as some pioneering works have already demonstrated (e.g., Benenson *et al.*, 2002, 2009; Bruch, 2014; Crooks, 2008; Yin, 2009). As such, direct validation of hypothesized models is extremely difficult with conventional KISS modeling and other purely theoretical modeling techniques, the methodological utility of TASS-type ABM and the technical solution offered by GIS converter for artisoc contribute to the longstanding interests in examination of the micro-macro relationships in social sciences.

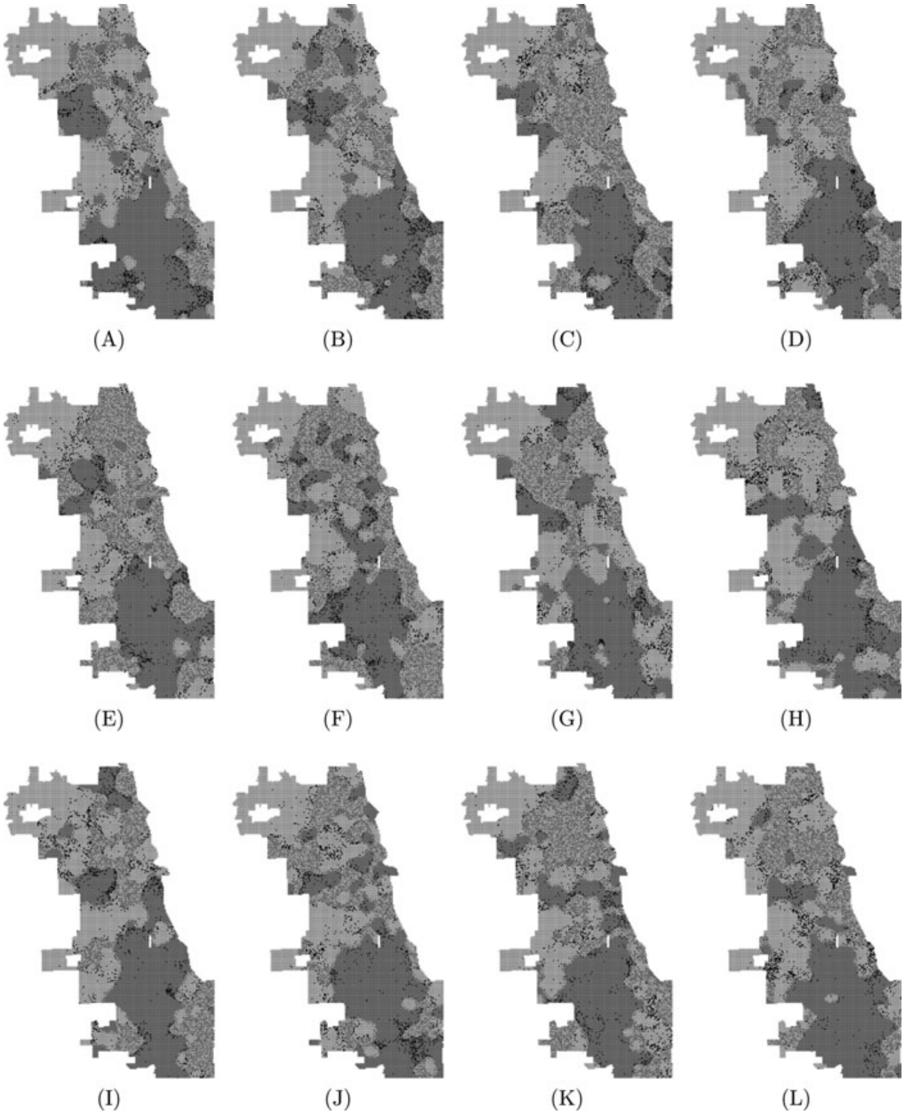


Figure 9. Twelve illustrative simulation runs of the specified segregation model with 'core communities'

Notes: Each panel shows the spatial distribution of agents when a simulation run reaches to an equilibrium where all agents are satisfied with their current positions. Note that the locations of segregated clusters become more stable than simulation runs shown in Figure 7 due to the introduction of 'core communities'.

4 Conclusion and outlook

In this article, we have scrutinized the recent development of ABM and its empirical turn. While the ABM technique becomes increasingly popular in social science research, conventional models adhere to the KISS principle and are rarely tested, evaluated, or validated using empirical data, leaving the links between models and empirical phenomena rather loose. While parsimony is vital in modeling attempts and the KISS principle cannot be disregarded, temporally and spatially specified modeling, or TASS modeling provides a promising way to fill this gap. While it currently takes in-depth ABM and GIS skills to exploit opportunities offered by TASS modeling, the most recent development of softwares presents a possible solution to such technical challenges. The working example that incorporates Schelling's segregation model with geo-referenced census data has showed that TASS modeling enables us to empirically test, evaluate, and validate agent-based models using empirical data. Although successfully providing a comprehensive explanation for the residential segregation in Chicago is likely to require additional modifications and systematic replications, our preliminary exercise has demonstrated the potential methodological utility of TASS modeling.

While our working example has focused on the sociological question of urban residential segregation, the possible methodological contribution of TASS modeling is not limited to the specific research agenda. Among others, the TASS-type, empirically explicit computational approach is likely to be extremely useful in civil war study as some pioneering studies have already demonstrated (e.g., Bhavnani and Choi, 2012; Bhavnani *et al.*, 2011, 2014; Lim *et al.*, 2007; Weidmann and Salehyan, 2013). The dynamics and spatial distributions of violence in civil war emerge from autonomous actions and interactions between armed groups and individuals. As hypotheses and theories of civil war mainly focus on the micro-level interactions, a methodological challenge arises: 'The mechanisms we postulate are located at the micro-level ... we can only observe the result of this process at the macro-level' (Weidmann and Salehyan, 2013: 55). TASS-type agent-based models incorporated with increasingly available detailed spatial data on violence taking place in the context of civil war, offer a methodological solution to explore the relationship between often unobserved or unobservable micro-mechanisms and observed macro-outcomes.

Indeed, TASS modeling is not a substitute but complement to existing methodologies. Nonetheless, incorporating parsimonious agent-based models with increasingly available spatial data offers a unique opportunity to systematically test and specify micro-level mechanisms underlying macro-level phenomena through computational experiments (Janssen and Ostrom, 2006; Lustick and Miodownik, 2009). As abstract KISS models can be extremely useful in developing theories, empirically grounded TASS models can help to yield empirically plausible insights. Empirically plausible TASS agent-based models could also serve as heuristic devices to assess the potential impacts of alternative policy options that would otherwise not be quantitatively comparable (e.g., Bhavnani *et al.*, 2014; Weidmann and Salehyan, 2013), which is likely to be of great interest to policy-makers and scholars alike.

There are also new methodological challenges for developing and testing TASS agent-based models. Among other things, effort is needed to develop commonly agreed and systematic standards for fitting models with empirical data and comparing alternative models. Currently, existing applications have used a variety of resolutions of empirical data and testing schemes, and there are no commonly applied standards. Developing such standards is likely to require accumulation of TASS modeling applications, but an effort itself may increase the acceptance of empirically grounded agent-based computational models in social sciences. We leave for future research this challenging task.

About the author

Gaku ITO is a Ph.D. student, the Department of Advanced Social and International Studies, the Graduate School of Arts and Sciences, the University of Tokyo (2012—) as well as Project Academic Support Specialist, the Department of Health Education and Sociology, the School of Public Health, the Graduate School of Medicine, the University of Tokyo (2015—). His research focuses on conflict dynamics, civil and international conflicts, and the integration of Agent-Based Modeling with Geographic Information System. He received MA in Arts and Sciences from the University of Tokyo in 2012.

Susumu YAMAKAGE is Professor of International Relations at Aoyama Gakuin University, Tokyo. Before taking the current position in 2012, he had taught International Relations at the Graduate School of Arts and Sciences, the University of Tokyo for three decades. His academic interests include the methodology of building and analyzing artificial (virtual) societies. He received Ph.D. in political science from Massachusetts Institute of Technology in 1982.

References

Almquist, Zack W. (2010), 'US Census Spatial and Demographic Data in R: The US Census 2000 Suite of packages', *Journal of Statistical Software*, 37(6): 1–31.

Axelrod, Robert (1984), *The Evolution of Cooperation*, New York, NY: Basic Books.

Axelrod, Robert (1997a), *The Complexity of Cooperation: Agent-Based Models of Competition and Col-Laboration*, Princeton, NJ: Princeton University Press.

Axelrod, Robert (1997b), 'The Dissemination of Culture: A Model with Local Convergence and Global Polarization', *Journal of Conflict Resolution*, 41(2): 203–26.

Benenson, Itzhak, Itzhak Omer, and Erez Hatna (2002), 'Entity-Based Modeling of Urban Residential Dynamics: The Case of Yaffo, Tel Aviv', *Environment and Planning B: Planning and Design*, 29(4): 491–512.

Benenson, Itzhak, Erez Hatna, and Ehud Or (2009), 'From Schelling to Spatially Explicit Modeling of Urban Ethnic and Economic Residential Dynamics', *Sociological Methods and Research*, 37(4): 463–97.

Bhavnani, Ravi and Hyun J. Choi (2012), 'Modeling Civil Violence in Afghanistan: Ethnic Geography, Control, and Collaboration', *Complexity*, 17(6): 42–51.

Bhavnani, Ravi, Michael G. Findley, and James H. Kuklinski (2009), 'Rumor Dynamics in Ethnic Violence', *Journal of Politics*, 71(3): 876–92.

Bhavnani, Ravi, Dan Miodownik, and Hyun Jin Choi (2011), 'Three Two Tango: Territorial Control and Selective Violence in Israel, the West Bank, and Gaza', *Journal of Conflict Resolution*, 55(1): 133–58.

Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing (2014), 'Group Segregation and Urban Violence', *American Journal of Political Science*, 58(1): 226–45.

Bivand, Roger S., Edzer Pebesma, and Virgilio Gómez-Rubio (2013), *Applied Spatial Data Analysis with R*, New York, NY: Springer.

Bremer, Stuart A. and Michael Mihalka (1977), 'Machiavelli in Machina: Or Politics among Hexagons', in Karl W. Deutsch, Bruno Fritsch, Helio Jaguaribe, and Andrei S. Markovits (eds.), *Problems of World Modeling: Political and Social Implications*, Cambridge, MA: Ballinger.

Bruch, Elizabeth E. (2014), 'How Population Structure Shapes Neighborhood Segregation', *American Journal of Sociology*, 119(5): 1221–78.

Bruch, Elizabeth E. and Robert D. Mare (2006), 'Neighborhood Choice and Neighborhood Change', *American Journal of Sociology*, 112(3): 667–709.

Castellano, Claudio, Santo Fortunato, and Vittorio Loreto (2009), 'Statistical Physics of Social Dynamics', *Reviews of Modern Physics*, 81(2): 591–646.

Cederman, Lars-Erik (1997), *Emergent Actors in World Politics: How States and Nations Develop and Dissolve*, Princeton, NJ: Princeton University Press.

Cederman, Lars-Erik (2002), 'Endogenizing Geopolitical Boundaries with Agent-Based Modeling', *Proceedings of the National Academy of Sciences of the United States of America*, 99(3): 7296–303.

Cederman, Lars-Erik (2005), 'Computational Models of Social Forms: Advancing Generative Process Theory', *American Journal of Sociology*, 110(4): 864–93.

Charles, Camille Zubrinsky (2003), 'The Dynamics of Racial Residential Segregation', *Annual Review of Sociology*, 29(1): 167–207.

Clark, William A. V. (1991), 'Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model', *Demography*, 28(1): 1–19.

Clark, William A. V. and Mark Fossett (2008), 'Understanding the Social Context of the Schelling Segregation Model', *Proceedings of the National Academy of Sciences of the United States of America*, 105(11): 4109–14.

Coleman, James (1990), *Foundations of Social Theory*, Cambridge, MA: Belknap Press of Harvard University Press.

Crooks, Andrew T. (2008), 'Constructing and Implementing an Agent-Based Model of Residential Segregation through Vector GIS', *International Journal of Geographical Information Science*, 24(5): 661–75.

Crooks, Andrew T., Christian Castle, and Michael Batty (2008), 'Key Challenges in Agent-Based Modelling for Geo-Spatial Simulation', *Computers, Environment and Urban Systems*, 32(6): 417–30.

Cusack, Thomus R. and Richard J. Stoll (1990), *Exploring Realpolitik: Probing International Relations Theory with Computer Simulation*, Boulder, CO: Lynne Rienner Publishers.

Emerson, Michael O., George Yancey, and Karen J. Chai (2001), 'Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans', *American Sociological Review*, 66(6): 922–35.

Epstein, Joshua and Robert Axtell (1996), *Growing Artificial Societies: Social Science from the Bottom Up*, Cambridge, MA: MIT Press.

Farley, Reynolds, Howard Schuman, Suzanne Bianchi, Diane Colasanto, and Shirley Hatchett (1978), '"Chocolate City, Vanilla Suburbs": Will the Trend Toward Racially Separate Communities Continue?', *Social Science Research*, 7(4): 319–44.

Fossett, Mark and Warren Waren (2005), 'Overlooked Implications of Ethnic Preferences for Residential Segregation in Agent-Based Models', *Urban Studies*, 42(11): 1893–917.

Hatna, Erez and Itzhak Benenson (2014), 'Combining Segregation and Integration: Schelling Model Dynamics for Heterogeneous Population', http://arxiv.org/abs/1406.5215.

Helbing, Dirk, Dirk Brockmann, Thomas Chadefaux, Karsten Donnay, Olivia Woolley-Meza, Mehdi Moussaid, Anders Johansson, Jens Krause, Sebastian Schutte, and Matjaž Perc (2015), 'Saving Human Lives: What Complexity Science and Information Systems Can Contribute', *Journal of Statistical Physics*, 158(3): 735–81.

Janssen, Marco A. and Elinor Ostrom (2006), 'Empirically Based, Agent-Based Models', *Ecology and Society*, 11(2): article 37.

Johnson, Dominic D. P., Nils B. Weidmann, and Lars-Erik Cederman (2011), 'Fortune Favours the Bold: An Agent-Based Model Reveals Adaptive Advantages of Overconfidence in War', PLoS ONE, 6(6) e20851.

Krysan, Maria, Mick P. Couper, Reynolds Farley, and Tyrone A. Forman (2009), 'Does Race Matter in Neighborhood Preferences? Results from a Video Experiment', *American Journal of Sociology*, 115(2): 527–59.

Laurie, Alexander J. and Narendra K. Jaggi (2003), 'Role of "Vision" in Neighbourhood Racial Segregation: A Variant of the Schelling Segregation Model', *Urban Studies*, 40(13): 2687–704.

Laver, Michael and Ernst Sergenti (2012), *Party Competition: An Agent-Based Model*, Princeton, NJ: Princeton University Press.

Lim, May, Richard Metzler, and Yaneer Bar-Yam (2007), 'Global Pattern Formation and Ethnic/Cultural Violence', *Science*, 317(5844): 1540–4.

Logan, John R., Brian Stults, Advisory Board, Margo Anderson, Suzanne Bianchi, Barry Bluestone, Sheldon Danziger, Claude Fischer, Daniel Lichter, and Kenneth Prewitt (2011), 'The Persistence of Segregation in the Metropolis: New Findings from the 2010 Census', *Census Brief prepared for Project US2010*, http://www.s4.brown.edu/us2010 (accessed 19 August 2014).

Lustick, Ian S., Dan Miodownik, and Roy J. Eidelson (2004), 'Secessionism in Multicultural States: Does Sharing Power Prevent or Encourage It?', *American Political Science Review*, 98(2): 209–29.

Lustick, Ian S. and Dan Miodownik (2009), 'Abstractions, Ensembles, and Virtualizations: Simplicity and Complexity in Agent-Based Modeling', *Comparative Politics*, 41(2): 223–44.

Macy, Michael W. and Robert Willer (2002), 'From Factors to Actors: Computational Sociology and Agent-Based Modeling', *Annual Review of Sociology*, 28(1): 143–66.

McCoy, Barry M. and Tai Tsun Wu (2014), The Two-Dimensional Ising Model, second edition. Cambridge, MA: Harvard University Press.

Mitsutsuji, Katsuma and Susumu Yamakage (2013), 'Bringing Empirical Analysis in International Politics and Multi-Agent Simulation: Substitutiong Basic Norms in International Society', *Working Paprer Series: Study on Artificial Societies*, No. 43, University of Tokyo.

O'Sullivan, David O. (2008), 'Geographical Information Science: Agent-Based Models', *Progress in Human Geography*, 32(4): 541–50.

Sakamoto, Takuto (2013), 'Conflict Analysis in Virtual States (CAVS): A New Experimental Method Based on the Extensive Use of Multi-Agent Simulation (MAS) and Geographical Information System (GIS)', *JICA Research Institute Working Paper*, 56(3).

Sakoda, James M. (1971), 'The Checkerboard Model of Social Interaction', *Journal of Mathematical Sociology*, 1(1): 119–32.

Schelling, Thomas C. (1969), 'Models of Segregation', *American Economic Review*, 59(2): 488–93.

Schelling, Thomas C. (1971), 'Dynamic Models of Segregation', *Journal of Mathematical Sociology*, 1(2): 143–86.

Schelling, Thomas C. (1978), *Micromotives and Macrobehavior*, New York: W. W. Norton.

Schrodt, Philip A. (1981), 'Conflict as a Determinant of Territory', *Behavioral Science*, 26(1): 37–50.

Siegel, David A. (2009), 'Social Networks and Collective Action', *American Journal of Political Science*, 53(1): 122–38.

Siegel, David A. (2011), 'When Does Repression Work? Collective Action in Social Networks', *Journal of Politics*, 73(4): 993–1010.

Torrens, Paul M. and Itzhak Benenson (2005), 'Geographic Automata Systems', *International Journal of Geographical Information Science*, 19(4): 385–412.

Vinkovic, Dejan and Alan Kirman (2006), 'A Physical Analogue of the Schelling Model', *Proceedings of the National Academy of Sciences of the United States of America*, 103(51): 19261–5.

Weidmann, Nils B. and Idean Salehyan (2013), 'Violence and Ethnic Segregation: A Computational Model Applied to Baghdad', *International Studies Quarterly*, 57(1): 52–64.

Xie, Yu and Xiang Zhou (2012), 'Modeling Individual-Level Heterogeneity in Racial Residential Segregation', *Proceedings of the National Academy of Sciences of the United States of America*, 109(29): 11646–51.

Yin, Li (2009), 'The Dynamics of Residential Segregation in Buffalo: An Agent-Based Simulation', *Urban Studies*, 46(13): 2749–70.

Zhang, Junfu (2004), 'A Dynamic Model of Residential Segregation', *Journal of Mathematical Sociology*, 28(3): 147–70.