



Sociohydrodynamics: Data-driven modeling of social behavior

Daniel S. Seara^a , Jonathan Colen^{a,b,c} , Michel Fruchart^{a,b,d}, Yael Avni^a, David G. Martin^{e,f} , and Vincenzo Vitelli^{a,b,e,1}

Edited by Jean-Philippe Bouchaud, Académie des Sciences, Paris, France; received April 14, 2025; accepted July 20, 2025 by Editorial Board Member Paul Chaikin

Living systems display complex behaviors driven by physical forces as well as decision-making. Hydrodynamic theories hold promise for simplified universal descriptions of socially generated collective behaviors. However, the construction of such theories is often divorced from the data they should describe. Here, we develop and apply a data-driven pipeline that links micromotives to macrobehavior by augmenting hydrodynamics with individual preferences that guide motion. We illustrate this pipeline on a case study of residential dynamics in the United States, for which census and sociological data are available. Guided by Census data, sociological surveys, and neural network analysis, we systematically assess standard hydrodynamic assumptions to construct a sociohydrodynamic model. Solving our minimal hydrodynamic model, calibrated using statistical inference, qualitatively captures key features of residential dynamics at the level of individual US counties. We highlight that a social memory, akin to hysteresis in magnets, emerges in the segregation–integration transition even with memory-less agents. While residential segregation is a multifactorial phenomenon, this physics analogy suggests a simple mechanistic explanation for the phenomenon of neighborhood tipping, whereby a small change in a neighborhood's population leads to a rapid demographic shift. Beyond residential segregation, our work paves the way for systematic investigations of decision-guided motility in real space, from micro-organisms to humans, as well as fitness-mediated motion in more abstract spaces.

hydrodynamics | economics | sociology | active matter | machine learning

Individual social organisms, from bacteria to ants to humans, display complex behaviors shaped by their interactions with each other and their environment. Groups of such organisms often form large, coherent patterns across space and time (1–5). This regularity suggests that aspects of social behavior may be captured using generalized hydrodynamic theories that account for individual choices. Hydrodynamic theories are mathematical descriptions of the time evolution of spatially extended systems that involve only a small number of slowly varying fields (6–9). This approach, originating in fluid mechanics, has since been applied to living systems in the field of active matter, ranging from microbial suspensions (10–15) and cellular tissues (16–23) to insect swarms (24, 25) and human crowds (26–28). In these systems, active mechanical (29, 30) or “social” forces (31–33) between individuals drive the dynamics of the hydrodynamic variables, such as density and polarization. However, we lack a principled way to incorporate cognitive decision-making into hydrodynamic models.

Here, we develop a data-driven pipeline to capture the physical manifestations of nonmechanical choices within a hydrodynamic theory. We take inspiration from microeconomics to codify individual preferences (micromotives) into utility functions (34, 35), and then we incorporate them into a “sociohydrodynamic” theory that can account for collective behavior (macrobehavior). We illustrate our approach on the case study of human residential dynamics, focusing on segregation between non-Hispanic White and non-Hispanic Black residents in the United States, for which both research (36–45) and data (46) from the field of sociology are available. Theoretical explorations have examined the connection between micromotives and macrobehavior (47–54), including recently proposed hydrodynamic theories (55–57). In addition, recent work suggests that statistical methods can forecast local trends in segregation observed in US census data (58–60). Our analysis combines these two perspectives to forecast demographic distributions using a hydrodynamic theory constructed directly from data.

We demonstrate that, for a period of four decades, both local and global aspects of the dynamics of US population distributions can be partly described by an intelligible, analytical model constructed from data. Our model's segregation–integration transition is history-dependent, suggesting a possible mechanism for a phenomenon dubbed “neighborhood tipping,” whereby a small change in a neighborhood's population leads to a rapid demographic shift (61–63).

Significance

In communities of living organisms, dynamic behavior is often driven by social interactions. Hydrodynamics simplifies the description of systems composed of many interacting constituents. While hydrodynamic models incorporating social interactions have been proposed, it is unclear the extent to which they describe real data. To address this issue, we combine tools from economics and statistical physics to construct and validate a hydrodynamic theory of social behavior. We show how to incorporate individual preferences within a hydrodynamic framework, verify our underlying hypotheses using neural networks, and apply the model to predict human residential dynamics across the United States.

Author affiliations: ^aJames Franck Institute, University of Chicago, Chicago, IL 60637; ^bDepartment of Physics, University of Chicago, Chicago, IL 60637; ^cJoint Institute on Advanced Computing for Environmental Studies, Old Dominion University, Norfolk, VA 23529; ^dGulliver, CNRS, École Supérieure de Physique et Chimie Industrielles, Université Paris Sciences & Lettres, Paris 75005, France; ^eLeinweber Institute for Theoretical Physics, University of Chicago, Chicago, IL 60637; and ^fEnrico Fermi Institute, University of Chicago, Chicago, IL 60637

Author contributions: D.S.S. and V.V. designed research; D.S.S., J.C., M.F., Y.A., and D.G.M. performed research; D.S.S. and J.C. analyzed data; and D.S.S., J.C., M.F., Y.A., D.G.M., and V.V. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. J.-P.B. is a guest editor invited by the Editorial Board.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: vitelli@uchicago.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2508692122/-DCSupplemental>.

Published August 29, 2025.

1. The Sociohydrodynamic Pipeline

Fig. 1 summarizes our sociohydrodynamic pipeline. First, we identify candidate hydrodynamic variables that evolve slowly both in time and space (Fig. 1A). To be useful, these variables must contain enough information to forecast their own future values—they must be self-predictive. To test this self-predictability in a model-agnostic way, we turn to machine learning and train a neural network to forecast the evolution of the candidate hydrodynamic variables and examine whether the forecasting is possible. Examining further the trained neural network allows us to assess whether the dynamics are local, i.e. whether hydrodynamic variables are only instantaneously influenced by their immediate surroundings (Fig. 1B). This locality, when it holds, considerably simplifies the models we consider next. We then construct an analytical, phenomenological model that relates micromotives to macrobehavior using a combination of physics and economic theory (Fig. 1C). Finally, we apply this model to predict real data by first inferring the equation’s coefficients and then checking their numerical solution against experimentally measured trajectories (Fig. 1D).

In the sections below, we detail how we apply this pipeline to the case of human residential dynamics. There, we make precise the variables we use, how we check predictability and locality, and the specific model we build to describe the data.

1.1. Identifying Hydrodynamic Variables in Social Behavior. Before building a hydrodynamic theory, we must first find suitable hydrodynamic variables. When the collective variables are not easy to guess, data-driven techniques can be used to propose candidates (64–66). In our example of residential dynamics, a reasonable guess for collective variables are the populations of different groups. Here, we focus on non-Hispanic White and non-Hispanic Black residents. Data from the decennial US Census (46) give us access to population densities of each group a at position $\mathbf{r}$ and time t , $\rho^a(\mathbf{r}, t)$. From these densities, we define a proxy for the local housing capacity $h(\mathbf{r})$ by finding the highest density of people who have lived at each location within a county over time. In this work, we take fill fractions as our hydrodynamic variables, defined as the population densities normalized by the maximum housing availability in the county— $\phi^a(\mathbf{r}, t) = \rho^a(\mathbf{r}, t)/\max(h)$ (*Materials and Methods*).

In order to apply the usual tools of hydrodynamics, we must check that the collective variables are slowly varying both in space and time, compared to the space and time scales we are able to resolve. Qualitatively, we observe that the main features of the population distributions appear unaffected by spatial smoothing over the length-scale of single counties. Quantitatively, fill fractions are correlated over distances four to seven times larger than the typical census tract ℓ in the region (*Materials and Methods*). In the following analysis, we smooth the fill fractions using a Gaussian filter with an SD of 3 km (*SI Appendix, Fig. S1*).

We also require hydrodynamic variables to vary slowly in time. We assess the temporal change in populations via measurements of segregation using the entropy index, which measures how local demographic distribution differs from the global composition (67–69) (*Materials and Methods*). Qualitatively, maps of the entropy index appear similar over a span of 40 y. Quantitatively, its overall magnitude decreases slowly

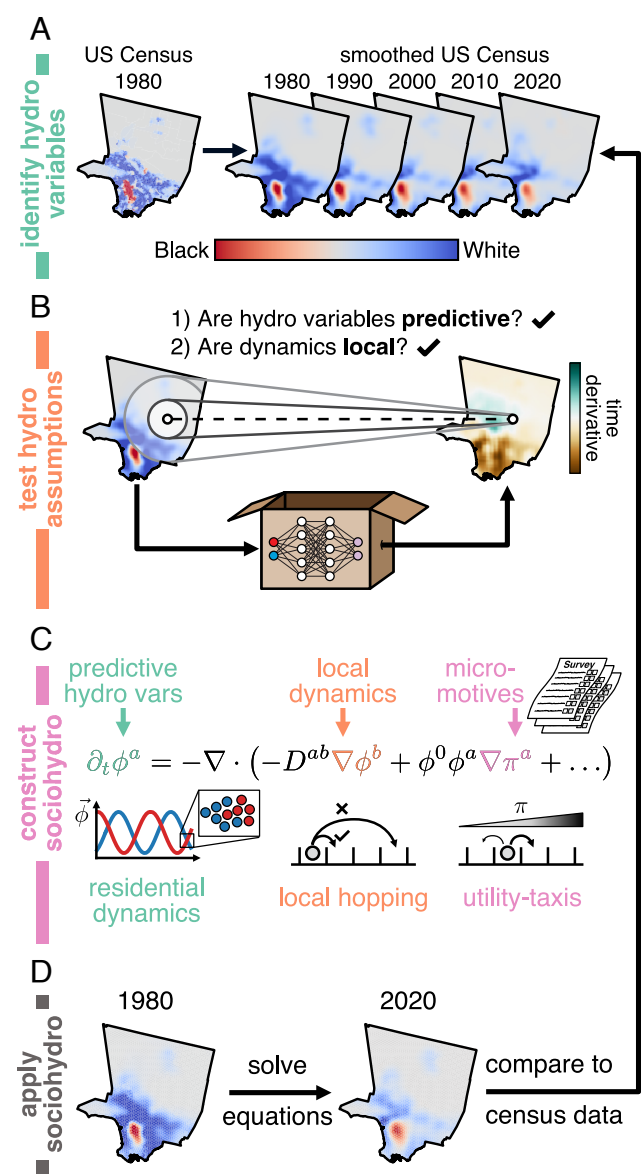


Fig. 1. How to construct a sociohydrodynamic model from data. (A) 1980 Census data of Los Angeles County, California, colored by the difference in relative population density, $\phi^W - \phi^B$, where $\phi^a = \rho^a / \rho_{\max}$ (Left). Right shows the result of smoothing the census data on a regular square lattice for decennial Census data from 1980, 1990, 2000, 2010, and 2020. (B) We test the assumptions made in simple hydrodynamic theories using a model-free, data-driven approach. Namely, we test whether population distributions are predictive of their own future, and if they evolve locally. The prior is determined by the accuracy of the neural network predictions, and the latter is determined by the network’s saliency (see Section 1.2 and *Materials and Methods*). (C) Constructing our sociohydrodynamic model based on dynamics of $\tilde{\phi}$ (green), with local movements (orange) that are biased according to spatial gradients of utility (purple). (D) We verify our sociohydrodynamic equations by solving them on individual US counties. The coefficients are learned from Census data for each county, and then simulated with the initial condition set as the 1980 US Census data. The results in 2020 are then compared to the 2020 US Census.

between the years of 1980–2020, as measured previously (41) (Fig. 6 C and D).

Together, these two results indicate that human residential dynamics may indeed be described by hydrodynamic variables that evolve slowly over the scale of individual counties and over a time-scale of decades.

1.2. Testing Hydrodynamic Assumptions with Neural Networks.

1.2.1. Self-predictability. We next test how well the candidate hydrodynamic variables $\vec{\phi}$ identified in the previous section can predict their own future states (Fig. 1B). Formally, we check whether $\partial_t \phi^a(\mathbf{r}, t)$ depends only on $\vec{\phi}(\mathbf{r}', t)$ (for all $\mathbf{r}'$). Testing the self-predictability of collective variables requires a reliable rule for dynamics, which we do not have a priori. To resolve this, we train a neural network on US Census data to map the fields ϕ^a to the time derivatives of the fields $\partial_t \phi^a$ (*Materials and Methods*). Our dynamical system

$$\partial_t \phi^a(\mathbf{r}, t) = f_{\vec{\phi}}^a[\vec{\phi}; h](\mathbf{r}), \quad [1]$$

uses a machine-learned dynamical rule $\vec{f}_{\vec{\phi}}$ for $\vec{\phi}$ that depends on both $\vec{\phi}$ and the local housing approximation h introduced earlier (Fig. 2A). We train on 34 US counties, each with at least 10^6 inhabitants, and test on the three largest US counties—Los Angeles CA, Cook IL, and Harris TX, with the addition of Fulton GA to better represent geographically disparate parts of the United States. Besides their large populations, all these counties have simple boundaries (convenient for solving our hydrodynamic equations later), see *SI Appendix* for further discussion.

The predicted population dynamics, obtained by integrating Eq. 1, capture qualitative large-scale features of the population distributions of US counties. Fig. 2B shows an example of the predicted White population in Los Angeles County, CA. Similar results are seen for both White and Black populations in 3 additional counties dispersed throughout the United States (*SI Appendix*, Fig. S2). In *SI Appendix*, we also quantify the prediction accuracy and show that the neural network out-performs several alternative dynamical rules, including no dynamics, linear growth, and exponential growth (*SI Appendix*, Fig. S3).

1.2.2. Locality. Then, we evaluate whether the dynamics are generated locally (Fig. 1B). Locality allows us to approximate $\partial_t \phi^a(\mathbf{r}, t)$ as a function of $\vec{\phi}$ and its gradients, $f^a[\vec{\phi}] = f^a(\vec{\phi}, \nabla \vec{\phi}, \dots)$ evaluated at $\mathbf{r}$. To evaluate locality, we consider the saliency

$$\mathcal{K}^{ab}(r) = \left\langle \frac{\partial f_{\vec{\phi}}^a(\mathbf{r})}{\partial \phi^b(\mathbf{r}')} \right\rangle_{|\mathbf{r}-\mathbf{r}'|=r} \quad [2]$$

of the machine-learned dynamics. In essence, $\mathcal{K}^{ab}(r)$ measures how strongly the predicted dynamics of ϕ^a at a point $\mathbf{r}$ depends on the value of the field ϕ^b at another point a distance r away (Fig. 2C; see *Materials and Methods*) (64, 70). A $\mathcal{K}^{ab}(r)$ that decays rapidly to zero indicates that $\vec{f}_{\vec{\phi}}$ uses information within a very narrow region to generate its dynamics. Note that correlation functions would not be sufficient to assess whether the dynamics is local as they do not directly address the evolution rules of the system.

Indeed, the trained neural network identifies a local rule (Fig. 2D). $\mathcal{K}^{ab}$ exhibits a similar structure for all predicted regions and is narrowly peaked at $r < 5\ell$, where ℓ is the typical size of a census tract within that county (Fig. 2D). Furthermore, the cross-salencies ($a \neq b$) may echo the preferences underlying residential decision-making. Below, we codify these preferences in “utility functions.” We find that salencies extracted from a neural network trained on agent-based simulations based on the Schelling model (47) support this claim. Namely, the sign of the salencies at $r = 0$ directly reflect the slope of the utility functions input to the simulations (*SI Appendix*, Fig. S4).

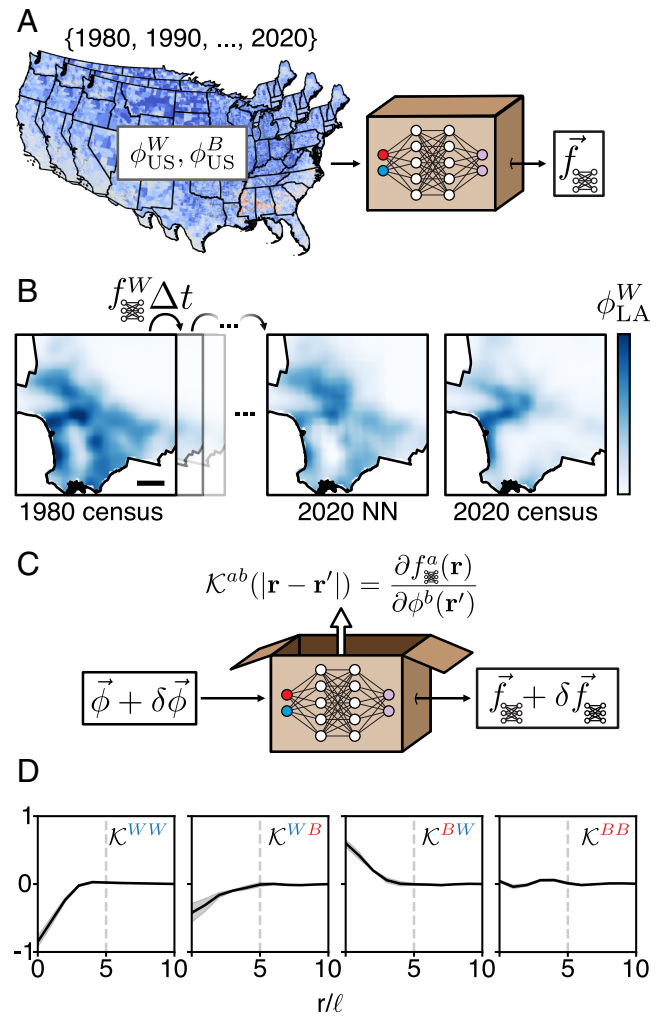


Fig. 2. Human residential dynamics are predictable and local. (A) Illustration of neural network training procedure. A convolutional neural network takes processed US Census data, $\vec{\phi}$ as input, and produces a functional $\vec{f}_{\vec{\phi}}$. This functional is optimized to provide an accurate prediction of 2020 census data when $\vec{\phi}(1980)$ is evolved by $\vec{f}_{\vec{\phi}}$. (B) Example of evolution by neural network for the White population in Los Angeles County. Starting with the 1980 census as an initial condition, populations within individual counties are evolved forward in time using $\vec{f}_{\vec{\phi}}$ to 2020, and then compared with the 2020 census. (Scale bar indicates a length of $5\ell = 11.5$ km.) (C) Schematic illustration of saliency $\mathcal{K}^{ab}$ as the linear response of the neural network. (D) Saliency measured radially for four regions, Cook County IL, Fulton County GA, Harris County TX, and Los Angeles County CA. Mean $\pm$ SD are shown by black line and gray shading, respectively.

Microscopically, nonlocality can both arise from the fact that agents are influenced by what happens away from them and from the fact that they can move far away. However, microscopic sources of nonlocality do not necessarily lead to a nonlocal effective description. For example, a variant of the Ising model with local interactions but infinite-range displacements (71) is described by a hydrodynamic theory with local motility at a sufficiently coarse-grained level (*SI Appendix*, Fig. S5). Indeed, human residential dynamics feature nonlocal displacements at the individual level: According to the US Census Bureau, 53% of moves were within the same county in 2022, down from 64% in 2012 (72).

1.3. Constructing a Hydrodynamic Theory. Using the insights and elements extracted from data in the initial steps of our pipeline, we now develop a hydrodynamic model designed to

capture universal features of residential dynamics found across US cities.

1.3.1. General theory. Having identified the collective variables $\phi^a(\mathbf{r}, t)$ and verified that they are predictive of their own future, we write a general equation of motion for these variables

$$\partial_t \phi^a(\mathbf{r}, t) = -\nabla \cdot \mathbf{J}^a + S^a \quad [3a]$$

in which we have separated the dynamics into two parts: the divergence of a flux $\mathbf{J}^a$ that redistributes ϕ^a in space and a source term S^a that changes the total populations. The self-predictability and locality of the dynamics for $\vec{\phi}$ (Fig. 2) motivates us to write $\mathbf{J}^a = \mathbf{J}^a(\vec{\phi})$ and $S^a = S^a(\vec{\phi})$.

The source term S^a in Eq. 3a describes how ϕ^a changes locally due to, for example, reproduction or immigration (*SI Appendix*, Fig. S6). In addition to human populations, these processes play a crucial role in contexts such as microbiology and ecology (73, 74). Growth can be related to a local evolutionary “fitness function” that describes to what extent the environment promotes the growth of a certain group (75–78). Although fitness and utility functions share similarities, they have no reason to be identical (79).

Assuming isotropy, we write the flux $\mathbf{J}^a$ using a gradient expansion as

$$\mathbf{J}^a(\vec{\phi}) = \phi^a \mathbf{v}^a - \sum_b \left(D^{ab} \nabla \phi^b + \Gamma^{ab} \nabla^3 \phi^b \right). \quad [3b]$$

In the above, $D^{ab}[\vec{\phi}]$ and $\Gamma^{ab}[\vec{\phi}]$ accounts for diffusion of ϕ^a down the gradients of ϕ^b and gradients of the Laplacian $\nabla^2 \phi^b$, respectively. The diagonals of the latter, $\Gamma^{aa}[\vec{\phi}]$, parameterize a surface tension, penalizing spatial gradients in ϕ^a (80, 81). Higher-order terms in the gradients have been neglected. The first term in Eq. 3b describes advection of ϕ^a at a velocity $\mathbf{v}^a$. We assume that the velocity is proportional to the gradient of a utility function π^a ,

$$\mathbf{v}^a \propto \nabla \pi^a. \quad [3c]$$

The proportionality factor will be determined below based on a microscopic, agent-based model. The utility function quantifies the preference of an a individual for the location $\mathbf{r}$ at time t , providing a link between motility and socioeconomic incentives. The gradient reflects the propensity of individuals to move toward regions they prefer, i.e. up gradients in their utility. We call this behavior “utility-taxi,” in reference to other guided navigation strategies such as chemotaxis (82–84) or infotaxis (85).

1.3.2. Utility functions. The question remains—what is π^a ? Although many socioeconomic and personal factors may contribute (40, 41), here we focus on the impact of neighborhood demographic preferences in driving residential dynamics. In other words, we seek a utility function written as

$$\pi^a(\mathbf{r}, t) = \pi^a(\vec{\phi}(\mathbf{r}, t)). \quad [4]$$

This is the key feature of sociohydrodynamics: It establishes a feedback loop between the slowly evolving hydrodynamic variables and the decision-making processes that lead to motility in the first place.

To model these utility functions in the case at hand, we turn to the sociology literature measuring neighborhood demographic preferences (Fig. 3). Social scientists have found that residential preferences remained consistent across time over a span of 16 years between 1976 and 1992 in the Detroit metropolitan area (86),

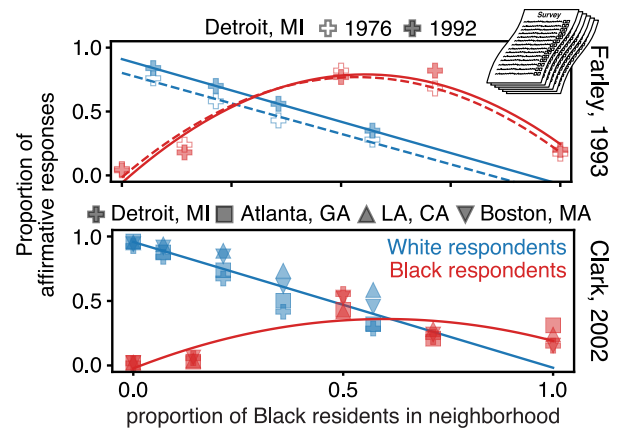


Fig. 3. Persistent residential preferences from sociological surveys. (Top) Reproduction of survey results shown in figures 8 & 10 in ref. 86. Survey's taken around Detroit, MI (+). Top plot shows proportion of White respondents “indicating they would be feel comfortable in the neighborhood” with the proportion of Black residents shown on the x-axis. Bottom shows proportion of Black residents ranking a neighborhood with the given proportion of Black residents as either their first or second choice among five options. Hollow circles show results from surveys taken in 1976, and filled circles show results from surveys taken in 1992. Dashed (1976) and solid (1992) lines are linear (Top) and quadratic (Bottom) fits to each set of data. (Bottom) Reproduction of figures 4 & 7 in ref. 87, which itself accumulates data from ref. 90. White respondents were asked whether they would move into a hypothetical neighborhood with the given proportion of Black residents. Black respondents, similarly to ref. 86, were asked to rank hypothetical neighborhoods and the results for the proportion of respondents ranking each neighborhood as their first choice is shown. Surveys were taken in Detroit, MI (+), Atlanta, GA (□), Los Angeles, CA (△), and Boston, MA (▽). Solid lines are linear (Top) and quadratic (Bottom) fits to each set of data. Precise wording of survey questions are reproduced from the respective references in *SI Appendix*.

and across space over several major US metropolitan areas (87) (Fig. 3). White residents show a monotonic decrease in their preference of neighborhoods with increasing proportion of Black residents. On the other hand, Black residents show a marked preference for mixed neighborhoods, which remains consistent between the two surveys. Qualitatively similar results were obtained in other US-based surveys (52, 86–89) (*SI Appendix*, Fig. S7). Thus, we assume that each group a has a distinct, time-invariant utility function written as a nonlinear function of $\vec{\phi}$.

1.3.3. Coarse-graining the Schelling model. To further constrain our equations of motion, we construct and coarse-grain an agent-based model for residential dynamics, based on the Schelling model (47–55, 61, 62). The same approach has been independently developed in refs. 55 and 56, resulting in hydrodynamic equations similar to ours. In short, it models agents on a lattice that randomly move to adjacent sites with a bias toward increasing their utility (*SI Appendix*). Within a mean-field approximation, coarse-graining this agent-based model leads to Eq. 3a with $S^a = 0$ and

$$\mathbf{v}^a(\vec{\phi}) = \phi^0 \nabla \pi^a(\vec{\phi}) \quad [5a]$$

$$D^{ab}(\vec{\phi}) = T^a (\phi^a + \delta^{ab} \phi^0) \quad [5b]$$

$$\Gamma^{ab}(\vec{\phi}) = \Gamma^a \delta^{ab} \phi^b \phi^0. \quad [5c]$$

In the above, $\phi^0 = 1 - \sum_b \phi^b$ is a vacancy fraction that arises because each lattice site has a maximum carrying capacity (91), mimicking the availability of housing. The parameters T^a and Γ^a control the rate of hopping and penalize spatial gradients of

$\vec{\phi}(x)$, respectively (see *SI Appendix* for discussion of the origin of the Γ term, *SI Appendix*, Fig. S8).

1.4. Applying the Sociohydrodynamic Model. To validate our model, we solve Eq. 3 and compare against the observed evolution of human populations in the United States (Fig. 4). To do so, we use linear regression to infer the coefficients of our model focusing on the same metropolitan areas as used to train the neural network in Section 1.2 (*SI Appendix*, Fig. S9).

In light of the responses to the social surveys in Fig. 3, we parameterize π^a as

$$\pi^a(\vec{\phi}) = \sum_{b,c} \kappa^{ab} \phi^b + \nu^{abc} \phi^b \phi^c. \tag{6}$$

This is a Taylor expansion of a nonconstant π^a around $\phi^a = 0$. Therefore, we fit 5 coefficients for the utility function of each field ($\kappa^{aW}, \kappa^{aB}, \nu^{aWW}, \nu^{aWB}, \nu^{aBB}$), as well as the diffusion constant T^a and the surface tension Γ^a .

The growth term S^a in Eq. 3 is fixed by fitting the total population change assuming an exponential growth model,

$S^a = r^a \phi^a$. This simple model captures the roughly exponential growth rates we empirically measure across US counties (*SI Appendix*, Fig. S6), although it neglects growth that may occur from immigration outside of the region. With this growth rate in hand, we fit the mass-conserving portion of Eq. 3, $-\nabla \cdot J^a = \partial_t \phi^a - S^a$.

We then numerically solve Eq. 3 using the learned coefficients. We take $\vec{\phi}(\mathbf{r}, 1980)$ as our initial condition and evolve in time until 2020. As shown in Fig. 4A, our simulations capture several qualitative features of observed population dynamics. For example, we predict the increase of the Black in the southwest of Harris county, as well as a decrease in the White population in the southeastern portion of Los Angeles county (see *SI Appendix*, Fig. S2 for results from other counties).

The fitted coefficients are given in Fig. 4B and C. Across multiple counties, we observed that the fitted diffusion constant of the Black population is larger than that of the White population, $T^B > T^W$ (Fig. 4B and *SI Appendix*, Fig. S3). The coefficients of the utility functions are also consistent across counties (Fig. 4B). The linear coefficients κ^{ab} have opposite signs for the two groups, signaling that White and Black residents have

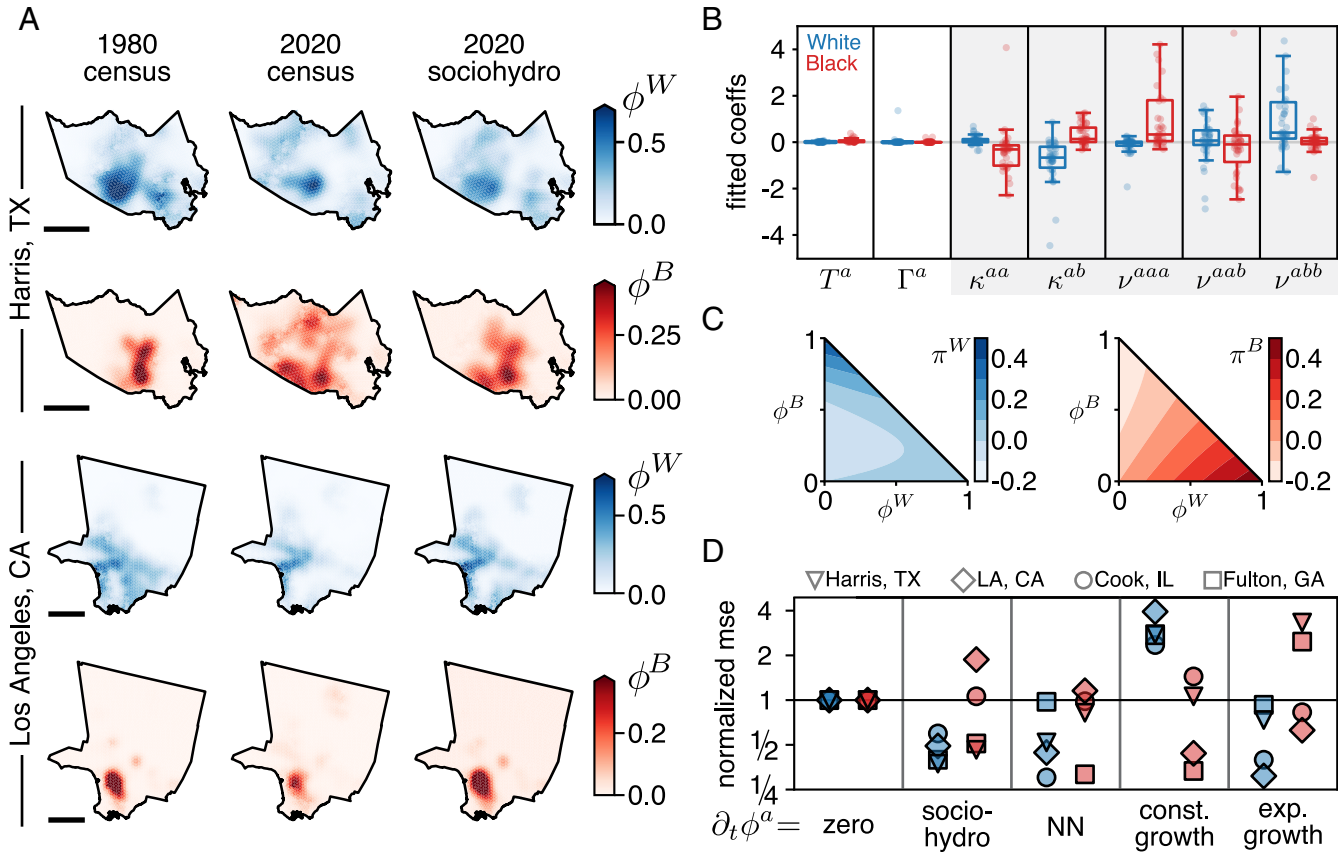


Fig. 4. Sociohydrodynamics predicts real population dynamics. (A) (Top) Sociohydrodynamic predictions for Harris County, TX. The *Left* column shows fill fractions, ϕ^W and ϕ^B the 1980 Census data for White and Black populations (first and third rows, respectively), that serve as initial conditions for the simulations. The second column shows the 2020 Census data for White and Black populations, in addition to the total change from the 1980 census, (second and fourth row). The third column shows the neural network's predictions for 2020 fill fractions starting from 1980. The fourth column shows the result from solving Eq. 3 with our learned coefficients starting from 1980 and continuing to 2020. (Bottom) Similar to *Left*, but for Fulton County, GA. (Scale bars indicate 25 km for each county.) See *SI Appendix* for results on other counties. (B) Learned coefficients resulting from linear regression on US Census data. Dots show results from fitting on the 38 counties considered. Box and whisker shows the quartiles and maxima and minima of the data, excluding outliers. Red and Blue symbols show results for Black and White populations, respectively. (C) Utility functions learned by the linear regression, highlighting the presence of nonreciprocity between the two groups. (D) Mean-square errors of various model predictions of human residential dynamics for the counties used to test our neural networks—Harris TX (∇), Los Angeles CA ($\diamond$), Cook IL ($\circ$), and Fulton GA ($\square$). All values are normalized to the error measured assuming that there is no change in the population from 1980 to 2020. The second column shows the results from numerically solving Eq. 3. The third column shows the errors from the neural network, whose results are shown in Fig. 2. The fourth column shows the results from assuming linear growth, defined as an extrapolation from 1990 to 2020 after a linear spline interpolation between 1980 and 1990. The final column shows the results from assuming only exponential growth.

incompatible residential preferences. More formally, the utilities π^W and π^B are incompatible in the sense that

$$\frac{\partial \pi^W}{\partial \phi^B} \neq \frac{\partial \pi^B}{\partial \phi^W}. \quad [7]$$

When this incompatibility condition is met, our equations of motion are nonequilibrium in the sense that they cannot be derived by optimizing a potential function (50, 92). These signs are consistent with survey results showing Black residents will move toward areas of higher White populations, while White residents will move away from areas of higher Black population (Fig. 3). However, the nonlinear terms for the utility of White residents v^{Wbc} additionally indicate a preference for neighborhoods with low White populations and high Black populations (Fig. 4C). Such behavior may originate from gentrification, where wealthy residents move into lower-income neighborhoods. Future work could test this hypothesis by supplementing our model with income or housing cost data.

Fig. 4D compares the sociohydrodynamic model's performance against other time-evolution models, including the neural networks in Fig. 2. Despite the simplicity of our sociohydrodynamic equations compared to f_{NN}^a , they achieve comparable

accuracy to the model-free estimate from the neural networks (see SI Appendix, Fig. S3 for results from other counties).

2. A Phase Diagram for Social Behavior

Inspired by the above results, we simplify our residential dynamics equations to a toy model in order to gain a better mathematical understanding of the possible sociohydrodynamic behaviors. We consider two groups $a = X, Y$ whose utility functions π^a are linear in the fields ϕ^a ($v^{abc} = 0$; see Fig. 5A and B).

The flux in Eq. 3b becomes

$$J^a(\vec{\phi}) = \sum_b (-D^{ab} + \phi^a \phi^0 \kappa^{ab}) \nabla \phi^b + \Gamma^{ab} \nabla^3 \phi^b.$$

This results in diffusion of ϕ^a either up or down gradients of ϕ^b , depending on the symmetries of the matrix κ^{ab} . Based on known results on the Schelling model (47–51, 61, 62), we expect to observe segregation (where the two groups concentrate at different places) and integration (where the two groups occupy mainly the same place), depending on the utility functions.

For every utility function, we solve Eq. 3 numerically, and then report the resulting dynamical phase of the system in

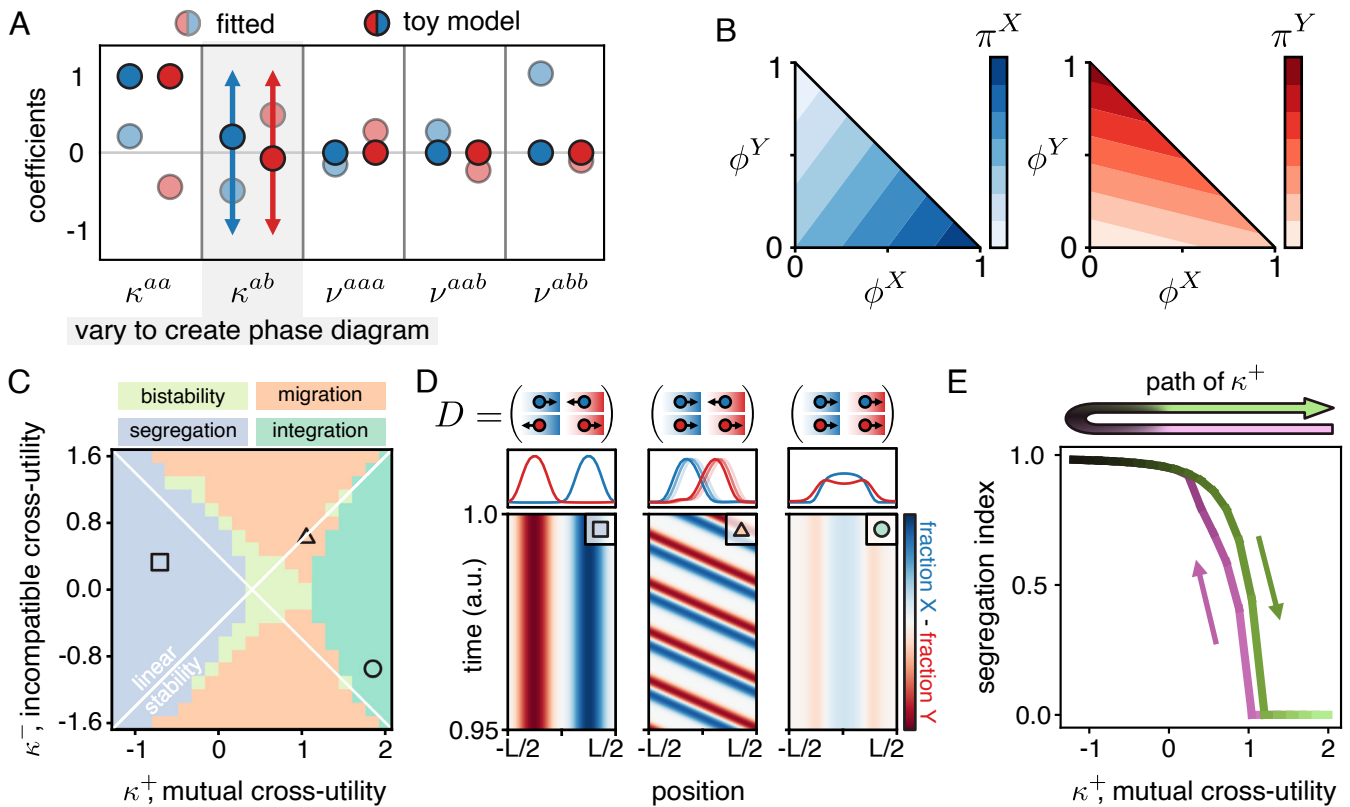


Fig. 5. Phase diagram for toy sociohydrodynamic model. (A) Toy model utility function coefficients (solid circles) compared to the learned coefficients from Fig. 4 (transparent circles). We consider $v^{abc} = 0$ for $\kappa^{aa} = 1 \forall a, b, c$, where $a \in \{X, Y\}$. We then vary the cross-utilities κ^{ab} to create our phase diagram. It will prove useful to parameterize our phase space in terms of $\kappa^\pm = \kappa^{XY} \pm \kappa^{YX}$. Other coefficients are $\tau^X = \tau^Y = 0.1$ and $\Gamma^X = \Gamma^Y = 1$. (B) Illustration of the linear utility functions for the two groups shown in the Left and Right subpanels, respectively. Plots show mutual cross-utility coefficient $\kappa^+ = \kappa^{XY} + \kappa^{YX} = -0.5$, incompatible cross-utility coefficient $\kappa^- = \kappa^{YX} - \kappa^{XY} = 1$. (C) Phase diagram showing different possible steady state dynamics depending on mutual and incompatible cross-utility coefficients, $\kappa^\pm$. We see 4 phases: segregation (purple), migration (orange), integration (green), and bistability of two phases (yellow). The white lines indicate results from a linear stability analysis for the onset of migratory states (details in SI Appendix). (D) Kymographs for steady state dynamics in the segregated phase, $(\kappa^+, \kappa^-) = (-0.07, 0.02)$ (Left), migrating phase, $(\kappa^+, \kappa^-) = (0.02, 0.04)$ (Middle), and integrated phase, $(\kappa^+, \kappa^-) = (0.09, -0.06)$ (Right). Plots on the Top show the state of the system at the final time. Schematics of the resulting diffusion matrix shown on Top of each phase. We set $T = 0.1$ and the average fill fraction of both types to be $\langle \phi^X \rangle = \langle \phi^Y \rangle = 0.25$. (E) The emergence of memory, in the form of hysteresis, in our simple model. Keeping $\kappa^- = 0$, simulations are run starting at $\kappa^+ = 2$ until a steady state is reached, and then κ^+ is slightly decreased. This process is repeated until $\kappa^+ = -1.2$, and then the process is reversed. The difference in the entropy index on the backward and forward pass of κ^+ .

Fig. 5C (see [SI Appendix](#) for criteria used in the categorization). We indeed observe the expected static states of segregation and integration (purple and green regions). In addition, we observe a time-dependent steady-state in which both groups continuously migrate (orange region; see also Fig. 5D), in agreement with the results of ref. 55. Finally, there are bistable regions (yellow) where multiple phases are observed depending on the initial state of the system ([SI Appendix](#), Fig. S10).

The mechanism leading to the migratory states in Fig. 5C and D stems from the incompatibility of the utilities π^X and π^Y in the sense of Eq. 7. For compatible utilities, meaning $\partial\pi^Y/\partial\phi^X = \partial\pi^X/\partial\phi^Y$, one can cast the dynamics as a gradient descent in a high-dimensional space ([SI Appendix](#) and refs. 50 and 92–94), thereby excluding time-dependent steady states like traveling waves. When the compatibility condition is violated, the corresponding interactions are nonreciprocal, a common ingredient to induce time-dependent steady-states (93–104). More precisely, Fig. 5B shows that the utility of X increases when ϕ^Y decreases, whereas the utility of Y increases when ϕ^X increases. In other words, Y tends to move toward X while X tends to move away from Y . When this tendency is strong enough, a time-dependent steady-state can emerge where the populations X and Y continuously chase or run away from each other.

To predict the behavior of our simplified model, we perform a linear stability analysis around an initially spatially uniform state for $\tilde{\phi}$. We find excellent agreement for the onset of pattern formation, when the uniform ϕ becomes unstable. The onset of the traveling states is not captured by the linear stability analysis due to the nonlinearities that play a role in their propagation. We see that no migration occurs when the overlap between the populations is below a threshold value, even when linear stability would predict migratory states ([SI Appendix](#), Fig. S11), in agreement with ref. 55.

The bistable regions in Fig. 5C illustrate that a single set of preferences can support multiple states of segregation. Which state is selected depends on the history of the system. To demonstrate this with our measured utility functions, we slowly change preferences of both groups over time. We implement a cycle in the mutual cross-utility $\kappa^+ = \kappa^{XY} + \kappa^{YX}$ (while fixing $\kappa^{XY} - \kappa^{YX}$), starting and ending with the same value. Fig. 5E shows the segregation indices measured in the resulting simulations. Some preferences (x-axis of the plot) result in two different segregation indices, depending on whether this preference evolved from a segregated or integrated state. This phenomenon is known as hysteresis: The state of the system depends on its past (105). This form of socioeconomic memory emerges at the community level, despite the fact that the individuals in our model have no memory. As a result, two communities with an identical set of preferences can be segregated or integrated depending on their histories, and abrupt changes can occur at the edge of the bistability region in phase space. This may shed light on a phenomenon known as “neighborhood tipping” in sociological research, where a threshold demographic distribution induces a rapid demographic shift in a neighborhood (61–63). We note that the region of phase space exhibiting this hysteresis can qualitatively change if utility functions are nonlinear ([SI Appendix](#), Fig. S12).

3. Conclusion and Outlook

To sum up, we introduced a data-driven pipeline to construct hydrodynamic descriptions of socially driven residential motility.

We illustrated this framework on the example of segregation in human residential dynamics. After testing the mathematical assumptions that enter a hydrodynamic theory, we validated our model by comparing its predictions to the evolution of demographic distributions in the US Census. We observed that our sociohydrodynamic model is sufficient to capture certain trends of residential dynamics in the United States, even though our model neglects many important personal, sociological, and geographical aspects of residential choices. We showed that segregation can persist even in the absence of external influence due to an emergent memory in the population of memoryless agents.

While limiting its accuracy, the simplicity of our model aids analysis, interpretation, and allows for several extensions. One could build a fluctuating hydrodynamic theory (106, 107), as modern models for growth (108) and human mobility (109) in cities are stochastic and long-ranged. One could increase the number of hydrodynamic fields, such as by including Hispanic citizens, the fastest growing demographic group in the United States. Finally, note that our theory is spatially homogeneous. Housing prices (57) and ideological indicators (110, 111) can provide spatial heterogeneity that can impact residential decisions (41).

More broadly, our work illustrates how to incorporate individual choices in a hydrodynamic theory to provide a precise mathematical mapping from micromotives to macrobehavior (48). Beyond humans, it could be applied to motile microorganisms (112–119) evolving (120–122) or adapting (123–125) in spatially extended time-varying environments. Further, our method may also be applied to motion through abstract rather than physical spaces, such as the chase and escape dynamics in antigenic space during hosts and pathogens coevolution (126, 127) or cell fate decisions during development (66, 128, 129).

4. Materials and Methods

4.1. US Census Data. Population data is collected from the decennial US census at the census tract level for decades 1980–2020, aggregated using the IPUMS database (46). Interpolation from census tracts to a square grid is performed using areal-weighted interpolation on population densities in units of $\#/(10\text{m})^2$ (130). Smoothing is then done using a Gaussian filter with width σ . GIS file information is provided in [SI Appendix](#).

4.2. Hydrodynamic Variables. We use US Census data at the census tract level to find population densities of each group $a \in \{W, B\}$ at position $\mathbf{r}$ at time t , denoted as $\rho^a(\mathbf{r}, t)$. We approximate the housing availability as $h(\mathbf{r}) = \max_t (\sum_a \rho^a(\mathbf{r}, t))$. Finally, we define the fill fraction as $\phi^a(\mathbf{r}, t) = \rho^a(\mathbf{r}, t)/h^*$, where $h^* = \max_{\mathbf{r}} (h(\mathbf{r}))$ acts as a carrying capacity. One could consider fill fractions that are normalized by a spatiotemporal capacity to capture how the dynamics of housing availability impacts residential dynamics (57). This is beyond the scope of this work.

4.3. Measuring Segregation. Let $i \in [1, 2, \dots, M]$ index local areas (e.g. census tracts) and $m \in [1, 2, \dots, M]$ index demographic groups. We define p_i^m to be the proportion of the population in the local area i that is composed of individuals from group m , and thereby define the local probability vector $\mathbf{p}_i = (p_i^1, p_i^2, \dots)$. Similarly, we define p^m to be the proportion of the population in the entire region (e.g. a county) that is composed of individuals from group m , and thereby define the regional probability vector $\mathbf{p} = (p^1, p^2, \dots)$. With these two quantities, we construct a local measure of segregation at location i , h_i , that is shown in Fig. 6C,

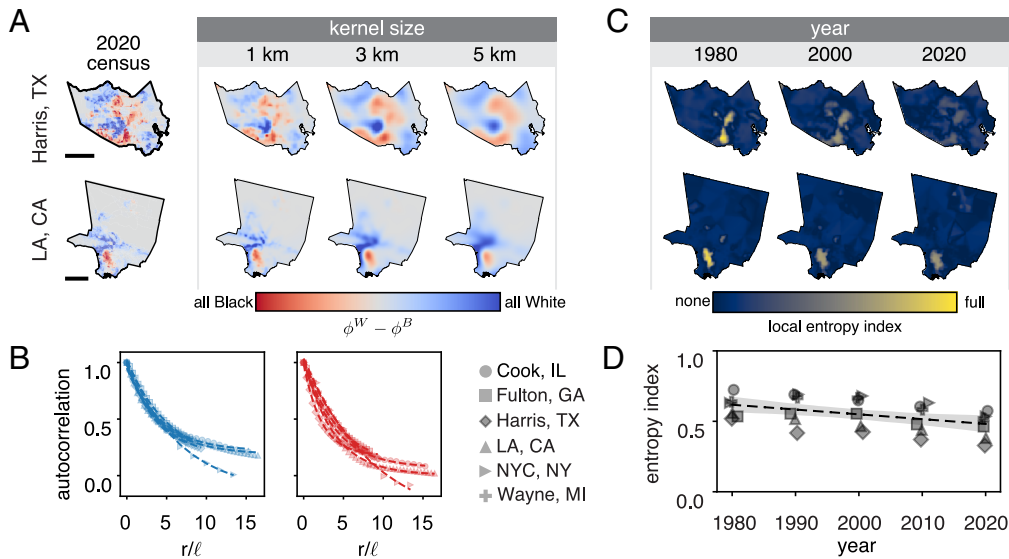


Fig. 6. Population distributions are slow variables across the United States. (A) Results of smoothing populations in various regions around the United States using increasing kernel sizes. From *Top* to *Bottom*, the regions are Harris County TX and Los Angeles County CA. Colors represent the proportion of local populations that are non-Hispanic White (blue) and non-Hispanic Black (red). (Scale bars are 25 km.) A Gaussian filter is used to smooth the populations. (B) Autocorrelation functions for White (blue) and Black (red) populations in six US regions (markers). Each set of data is fit to a decaying exponential (dashed line). The x-axis, measuring distance, is normalized by the median linear census tract size, ℓ of each region. Normalizing the correlation length by the ℓ accounts for differing densities, as each census tract is created to contain $\sim 10^3$ residents. The typical correlation lengths for White and Black populations in the regions we analyzed are $\langle \xi_W \rangle = 7.3 \pm 2.5 \ell$ and $\langle \xi_B \rangle = 4.3 \pm 1.0 \ell$. (C) Local segregation, h_i (Eq. 8a) for the two regions in A, shown from Census Data in 1980, 2000, and 2020. (D) Entropy index, S_H (Eq. 8b) measured as the weighted sum of the local segregation (*Materials and Methods*). Markers denote the same counties as B, and the dashed line is a linear fit to all data.

$$h_i = \frac{\sum_m p_i^m \ln \left(\frac{p_i^m}{p^m} \right)}{-\sum_m p^m \ln p^m} = \frac{D_{\text{KL}}(\mathbf{p}_i || \mathbf{p})}{H(\mathbf{p})} \quad [8a]$$

where $D_{\text{KL}}(\mathbf{p}_i || \mathbf{p}) \geq 0$ is the Kullback-Leibler divergence between $\mathbf{p}_i$ and $\mathbf{p}$, and $H(\mathbf{p})$ is the Shannon entropy of $\mathbf{p}$. Taking a weighted sum of h_i gives the entropy index, S_H , a popular choice of measurement for racial segregation (60, 68). Specifically, weighting h_i by the fraction of the total population that lives in local area i , f_i , gives

$$S_H = \sum_i f_i h_i. \quad [8b]$$

4.4. Sociohydrodynamic Diffusion Matrix. We provide the full derivation of our sociohydrodynamic theory starting from an agent-based model in *SI Appendix*, including the effects of altruism. Here, we simply report the diffusion matrix in Eq. 3. Assuming that the utility does not have an explicit spatial dependence, i.e. $\pi^a = \pi^a(\phi)$, we have

$$J^a(x, t) = - \sum_b D^{ab} \partial_x \phi^b + \phi^a \phi^0 \partial_x \pi^a + \Gamma^{ab} \partial_x^3 \phi^b \quad [9]$$

$$= - \sum_b D^{ab} \partial_x \phi^b + \Gamma^{ab} \partial_x^3 \phi^b \quad [10]$$

where $\phi^0 = 1 - \sum_b \phi^b$ is the vacancy fraction and D^{ab} are elements of the matrix

$$D(\vec{\phi}) = -T \begin{pmatrix} 1 - \phi^Y & \phi^X \\ \phi^Y & 1 - \phi^X \end{pmatrix} + \phi^0 \begin{pmatrix} \phi^X \frac{\partial \pi^X}{\partial \phi^X} & \phi^X \frac{\partial \pi^X}{\partial \phi^Y} \\ \phi^Y \frac{\partial \pi^Y}{\partial \phi^X} & \phi^Y \frac{\partial \pi^Y}{\partial \phi^Y} \end{pmatrix} \quad [11]$$

The first term arises due to volume exclusion effects with finite carrying capacity of the lattice sites in the agent-based model. The second term arises from the utility functions. Similarly, we find

$$\Gamma^{ab} = \delta^{ab} \phi^b \phi^0, \quad [12]$$

where δ^{ab} is the Kronecker delta.

4.5. Numerical Methods. Parameter inference is done using linear regression on US Census data. We first interpolate the US Census data using a 3rd-order spline to estimate populations between decennial census years. We then estimate time derivatives and spatial gradients using finite differences. A feature matrix A is constructed using each term in Eq. 3 as a column, and a target vector b is created using the time derivatives. The coefficients in Eq. 3 are collected in a vector x , and we invert the equation $Ax = b$.

For the simulations shown in Fig. 4, we solve Eq. 3 over county boundaries using a finite-volume method. We construct a triangular mesh over the county geometry using GMSH, and use the Python package FiPy (131) to solve Eq. 3. More details can be found in *SI Appendix*.

For the simulations shown in Fig. 5, we simulate Eq. 3 in 1 spatial dimension using a finite difference method, with a 4th order discretization in space and 1st order discretization in time. Unless otherwise stated, we set $T = 0.1$ and $\Gamma = 1$, and use a time step of $\Delta t = 0.1$ and $\Delta x = 0.625$.

4.6. Linear Utility Functions. The linear utility functions used in Fig. 5 are given by

$$\pi^a(\vec{\phi}) = \sum_b \kappa^{ab} \phi^b \quad [13a]$$

$$\kappa = \begin{pmatrix} \kappa^{XX} & \kappa^{XY} \\ \kappa^{YX} & \kappa^{YY} \end{pmatrix} \quad [13b]$$

The matrix elements κ^{ab} quantify how the utility of a is affected by the presence of b . We call κ^{XX} and κ^{YY} the self-utility coefficients, and κ^{XY} and κ^{YX} the cross-utility coefficients. We find it convenient to define $\kappa^\pm = \kappa^{YX} \pm \kappa^{XY}$. We dub κ^+ the “mutual” cross-utility coefficient as it measures the degree of mutual (dis)like between the X and Y . Likewise, we call κ^- the “incompatible” cross-utility coefficient as it measures the incompatibility of the two utilities with each other (Fig. 5).

4.7. Assessing Locality via Neural Networks. We used a convolutional neural network to predict the dynamics $(\partial_t \phi^W, \partial_t \phi^B)$ from an initial condition (ϕ^W, ϕ^B) . Briefly, it contains two convolutional modules which computes latent features at the scale of the input data and uses a downsampled representation in order to aggregate spatial information over short or long distances. The network predicts the discrete time derivative and forecasts the next time step as

$$\phi^a(t + \Delta t) = \phi^a(t) + f_{\text{net}}^a[\vec{\phi}(t); h] \Delta t \quad [14]$$

Here, the population index $a \in (W, B)$ and the capacity h is taken as the maximum occupancy at each coordinate over the observed time range. The full network details are given in *SI Appendix*. During training, the network saw Census data from 1980 to 2020 for all counties except the four shown in Fig. 4D. In addition, the training set included samples from only the first decade of Census data (1980–1990) for these four counties. The results in Figs. 2B and 4D are extrapolations beyond the training regime for the remaining 30 y (1990–2020) in each test county. We used the trained network to predict population dynamics over the 40-y window spanned by decennial Census data.

To assess locality in the predicted dynamics, we computed the output saliencies for the trained network. Saliency is a measure of how much a model's predictions depend on its inputs and is here defined as (70)

$$K^{ab}(\mathbf{r}_i, \mathbf{r}_j) = \frac{\partial f_{\text{net}}^a[\vec{\phi}(\mathbf{r}_i)]}{\partial \phi^b(\mathbf{r}_j)} \quad [15]$$

Here a, b are population indices while i, j refer to spatial coordinates within each county. We compute $K^{ab}(\mathbf{r}_i, \mathbf{r}_j)$ from each input-output pair in each county dataset, for 100 randomly sampled output points $\mathbf{r}_i$ and all input points $\mathbf{r}_j$.

We azimuthally average this into a set of curves $K^{ab}(|\mathbf{r}_a - \mathbf{r}_b|)$ which are plotted in Fig. 2C.

Data, Materials, and Software Availability. Code and data have been deposited in Zenodo (132).

ACKNOWLEDGMENTS. We thank Michael Benzaquen, Luis Bettencourt, Jordan Kemp, and Ruben Zakine for fruitful conversations. D.S.S., M.F., and Y.A. acknowledge support from a MRSEC-funded Kadanoff-Rice fellowship and the University of Chicago Materials Research Science and Engineering Center, which is funded by the NSF under award no. DMR-2011854. Y.A. acknowledges support from the Zuckerman STEM Leadership Program. J.C. acknowledges support from the Hampton Roads Biomedical Research Consortium as part of the effort associated with the Old Dominion University-Thomas Jefferson National Accelerator Facility Joint Institute on Advanced Computing for Environmental Studies. D.M. acknowledges support from the Enrico Fermi Institute and the Kadanoff Center at UChicago. J.C. and M.F. acknowledge support from the NSF under grant DMR-2118415. M.F. and V.V. acknowledge partial support from the UChicago Materials Research Science and Engineering Center (NSF DMR-2011864). M.F. acknowledges support from the Simons Foundation. V.V. acknowledges support from the Army Research Office under grant W911NF-22-2-0109 and W911NF-23-1-0212 and the Theory in Biology program of the Chan Zuckerberg Initiative. This research was supported from the NSF through the Center for Living Systems (grant no. 2317138). This work was completed in part with resources provided by the University of Chicago's Research Computing Center. This work was supported in part by the Wahab Research Computing cluster at Old Dominion University (NSF CNS-1828593).

1. T. Reichenbach, M. Mobilia, E. Frey, Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. *Nature* **448**, 1046–1049 (2007).
2. I. D. Couzin, Collective cognition in animal groups. *Trends Cogn. Sci.* **13**, 36–43 (2009).
3. N. T. Ouellette, A physics perspective on collective animal behavior. *Phys. Biol.* **19**, 021004 (2022).
4. T. Vicsek, A. Zafeiris, Collective motion. *Phys. Rep.* **517**, 71–140 (2012).
5. L. M. A. Bettencourt, The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
6. L. P. Kadanoff, P. C. Martin, Hydrodynamic equations and correlation functions. *Ann. Phys.* **24**, 419–469 (1963).
7. L. D. Landau, E. M. Lifshitz, *Fluid Mechanics: Landau and Lifshitz* (Elsevier Science & Technology Books, 1987), p. 539.
8. P. W. Anderson, More is different. *Science* **177**, 393–396 (1972).
9. W. Van Saarloos, V. Vitelli, Z. Zeravcic, *Soft Matter: Concepts, Phenomena and Applications* (Princeton University Press, 2023).
10. A. Sokolov, I. S. Aranson, J. O. Kessler, R. E. Goldstein, Concentration dependence of the collective dynamics of swimming bacteria. *Phys. Rev. Lett.* **98**, 158102 (2007).
11. H. H. Wensink *et al.*, Meso-scale turbulence in living fluids. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14308–14313 (2012).
12. H. Wioland, F. G. Woodhouse, J. Dunkel, J. O. Kessler, R. E. Goldstein, Confinement stabilizes a bacterial suspension into a spiral vortex. *Phys. Rev. Lett.* **110**, 268102 (2013).
13. H. Li *et al.*, Data-driven quantitative modeling of bacterial active nematics. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 777–785 (2018).
14. K. Copenhagen, R. Alert, N. S. Wingreen, J. W. Shaevitz, Topological defects promote layer formation in *Myxococcus xanthus* colonies. *Nat. Phys.* **17**, 211–215 (2020).
15. A. I. Curatolo *et al.*, Cooperative pattern formation in multi-component bacterial systems through reciprocal motility regulation. *Nat. Phys.* **16**, 1152–1157 (2020).
16. A. F. Mertz *et al.*, Scaling of traction forces with the size of cohesive cell colonies. *Phys. Rev. Lett.* **108**, 198101 (2012).
17. T. B. Saw *et al.*, Topological defects in epithelia govern cell death and extrusion. *Nature* **544**, 212–216 (2017).
18. C. Pérez-González *et al.*, Active wetting of epithelial tissues. *Nat. Phys.* **15**, 79–88 (2018).
19. S. J. Streichan, M. F. Lefebvre, N. Noll, E. F. Wieschaus, B. I. Shraiman, Global morphogenetic flow is accurately predicted by the spatial distribution of myosin motors. *eLife* **7**, e27454 (2018).
20. R. Alert, C. Blanch-Mercader, J. Casademunt, Active fingering instability in tissue spreading. *Phys. Rev. Lett.* **122**, 088104 (2019).
21. D. Boockock, N. Hino, N. Ruzickova, T. Hirashima, E. Hannezo, Theory of mechanochemical patterning and optimal migration in cell monolayers. *Nat. Phys.* **17**, 267–274 (2020).
22. M. S. Yousafzai *et al.*, Cell-matrix elastocapillary interactions drive pressure-based wetting of cell aggregates. *Phys. Rev. X* **12**, 031027 (2022).
23. J. M. Armengol-Collado, L. N. Carenza, J. Eckert, D. Krommydas, L. Gioni, Epithelia are multiscale active liquid crystals. *Nat. Phys.* **19**, 1773–1779 (2023).
24. A. Cavagna *et al.*, Natural swarms in 3.99 dimensions. *Nat. Phys.* **19**, 1043–1049 (2023).
25. D. Gorboson *et al.*, An effective hydrodynamic description of marching locusts. *Phys. Biol.* **21**, 026004 (2024).
26. R. L. Hughes, The flow of human crowds. *Annu. Rev. Fluid Mech.* **35**, 169–182 (2003).
27. N. Bain, D. Bartolo, Dynamic response and hydrodynamics of polarized crowds. *Science* **363**, 46–49 (2019).
28. F. Gu, B. Guiselin, N. Bain, I. Zuriguel, D. Bartolo, Emergence of collective oscillations in massive human crowds. *Nature* **638**, 112–119 (2025).
29. M. C. Marchetti *et al.*, Hydrodynamics of soft active matter. *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
30. M. Vrugt, R. Wittkowski, Metareview: a survey of active matter reviews. *Eur. Phys. J. E* **48**, 12 (2025).
31. D. Helbing, P. Molnár, Social force model for pedestrian dynamics. *Phys. Rev. E* **51**, 4282–4286 (1995).
32. M. Ballerini *et al.*, Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1232–1237 (2008).
33. A. Corbetta, F. Toschi, Physics of human crowds. *Ann. Rev. Condens. Matter Phys.* **14**, 311–333 (2023).
34. J. V. Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, 2007), p. 776.
35. M. J. Osborne, A. Rubinstein, *A Course in Game Theory* (MIT Press, Cambridge, MA, 2006).
36. D. R. Williams, C. Collins, Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Rep.* **116**, 404–416 (2001).
37. D. Pager, H. Shepherd, The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Ann. Rev. Sociol.* **34**, 181–209 (2008).
38. S. F. Reardon, A. Owens, 60 years after brown: Trends and consequences of school segregation. *Ann. Rev. Sociol.* **40**, 199–218 (2014).
39. D. Alexander, J. Currie, Is it who you are or where you live? Residential segregation and racial gaps in childhood asthma. *J. Health Econ.* **55**, 186–200 (2017).
40. C. Z. Charles, The dynamics of racial residential segregation. *Ann. Rev. Sociol.* **29**, 167–207 (2003).
41. J. Hwang, T. W. McDaniel, Racialized reshuffling: Urban change and the persistence of segregation in the twenty-first century. *Ann. Rev. Sociol.* **48**, 397–419 (2022).
42. W. E. B. Du Bois, *The Souls of Black Folk*, *Penguin Classics* (Penguin Books, New York, 2018).
43. D. S. Massey, American apartheid: Segregation and the making of the underclass. *Am. J. Sociol.* **96**, 329–357 (1990).
44. P. A. Jargowsky, *Poverty and Place* (Russell Sage Foundation, New York, 1998).
45. K. E. Taeuber, A. F. Taeuber, *Residential Segregation and Neighborhood Change* (Transaction Publishers, 2008).
46. J. Schroeder *et al.*, IPUMS National Historical Geographic Information (Version 20.0) [dataset]. IPUMS (2025). <http://doi.org/10.18128/D050.V20.0>. Accessed 24 August 2023.
47. T. C. Schelling, Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186 (1971).
48. T. C. Schelling, *Micromotives and Macrobehaviour* (W. W. Norton, 1980), p. 256.
49. D. Vinković, A. Kirman, A physical analogue of the schelling model. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19261–19265 (2006).
50. S. Grauwlin, E. Bertin, R. Lemoy, P. Jensen, Competition between collective and individual dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20622–20626 (2009).
51. S. Grauwlin, F. Goffette-Nagot, P. Jensen, Dynamic models of residential segregation: An analytical solution. *J. Public Econ.* **96**, 124–141 (2012).

52. W. A. V. Clark, M. Fossett, Understanding the social context of the schelling segregation model. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4109–4114 (2008).
53. M. Fossett, Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis. *J. Math. Sociol.* **30**, 185–273 (2006).
54. L. Gauvin, J. Vannimenus, J. P. Nadal, Phase diagram of a schelling segregation model. *Eur. Phys. J. B* **70**, 293–304 (2009).
55. R. Zakine, J. Garnier-Brun, A. C. Becharat, M. Benzaquen, Socioeconomic agents as active matter in nonequilibrium sakoda-schelling models. *Phys. Rev. E* **109**, 044310 (2024).
56. J. Garnier-Brun, R. Zakine, M. Benzaquen, From nonequilibrium to equilibrium: Insights from a two-population occupation model. *arXiv [Preprint]* (2024). <https://doi.org/10.48550/arXiv.2412.14996> (Accessed 15 April 2025).
57. A. C. Becharat, M. Benzaquen, J. P. Bouchaud, The diffusive nature of housing prices. *arXiv [Preprint]* (2024). <https://arxiv.org/pdf/2412.14624> (Accessed 27 April 2025).
58. Y. Chen *et al.*, Small-area population forecasting in a segregated city using density-functional fluctuation theory. *J. Comput. Soc. Sci.* **7**, 2255–2275 (2024).
59. Y. A. Kinkhabwala, B. Barron, M. Hall, T. A. Arias, I. Cohen, Forecasting racial dynamics at the neighborhood scale using density-functional fluctuation theory. *arXiv [Preprint]* (2021). <https://doi.org/10.48550/arXiv.2108.04084> (Accessed 10 September 2023).
60. B. Barron *et al.*, Extending the use of information theory in segregation analyses to construct comprehensive models of segregation. *arXiv [Preprint]* (2022). <https://doi.org/10.48550/arXiv.2212.06980> (Accessed 11 September 2023).
61. D. Card, A. Mas, J. Rothstein, Tipping and the dynamics of segregation. *Quart. J. Econ.* **123**, 177–218 (2008).
62. J. Zhang, Tipping and residential segregation: A unified schelling model. *J. Reg. Sci.* **51**, 167–193 (2011).
63. S. Gualdi, M. Tarzia, F. Zamponi, J. P. Bouchaud, Tipping points in macroeconomic agent-based models. *J. Econ. Dyn. Control* **50**, 29–61 (2015).
64. M. S. Schmitt *et al.*, Machine learning interpretable models of cell mechanics from protein images. *Cell* **187**, 481–494.e24 (2024).
65. M. S. Schmitt *et al.*, Information theory for data-driven model reduction in physics and biology. *arXiv [Preprint]* (2023). <https://doi.org/10.48550/arXiv.2312.06608> (Accessed 12 October 2024).
66. M. Lefebvre *et al.*, Learning a conserved mechanism for early neuroectoderm morphogenesis. *bioRxiv [Preprint]* (2023). <https://doi.org/10.1101/2023.12.22.573058> (Accessed 2 November 2024).
67. O. D. Duncan, B. Duncan, A methodological analysis of segregation indexes. *Am. Sociol. Rev.* **20**, 210 (1955).
68. S. F. Reardon, D. O'Sullivan, Measures of spatial segregation. *Sociol. Methodol.* **34**, 121–162 (2004).
69. E. Roberto, The divergence index: A decomposable measure of segregation and inequality. *arXiv [Preprint]* (2015). <https://doi.org/10.48550/arXiv.1508.01167> (Accessed 4 September 2024).
70. K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv [Preprint]* (2013). <https://doi.org/10.48550/arXiv.1312.6034> (Accessed 12 October 2023).
71. M. Conti, B. Meerson, A. Peleg, P. V. Sasorov, Phase ordering with a global conservation law: Ostwald ripening and coalescence. *Phys. Rev. E* **65**, 046117 (2002).
72. US Census Bureau, CPS historical migration/geographic mobility tables (2023). <https://www.census.gov/data/tables/time-series/demo/geographic-mobility/historic.html>. Accessed 22 July 2024.
73. R. A. Fisher, The wave of advance of advantageous genes. *Ann. Eugen.* **7**, 355–369 (1937).
74. O. Hallatschek *et al.*, Proliferating active matter. *Nat. Rev. Phys.* **5**, 407–419 (2023).
75. J. F. Crow, M. Kimura, *An Introduction to Population Genetics Theory* (Blackburn Press, 1970), p. 608.
76. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1983).
77. L. S. Tsimring, H. Levine, D. A. Kessler, RNA virus evolution via a fitness-space model. *Phys. Rev. Lett.* **76**, 4440–4443 (1996).
78. V. Mustonen, M. Lässig, From fitness landscapes to seascapes: Non-equilibrium dynamics of selection and adaptation. *Trends Genet.* **25**, 111–119 (2009).
79. J. Cremer *et al.*, Chemotaxis as a navigation strategy to boost range expansion. *Nature* **575**, 658–663 (2019).
80. J. W. Cahn, J. E. Hilliard, Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.* **28**, 258–267 (1958).
81. P. C. Hohenberg, B. I. Halperin, Theory of dynamic critical phenomena. *Rev. Mod. Phys.* **49**, 435–479 (1977).
82. E. F. Keller, L. A. Segel, Model for chemotaxis. *J. Theor. Biol.* **30**, 225–234 (1971).
83. J. Adler, Chemotaxis in bacteria. *Annu. Rev. Biochem.* **44**, 341–356 (1975).
84. H. C. Berg, Chemotaxis in bacteria. *Annu. Rev. Biophys. Bioeng.* **4**, 119–136 (1975).
85. M. Vergassola, E. Villermaux, B. I. Shraiman, 'Infotaxis' as a strategy for searching without gradients. *Nature* **445**, 406–409 (2007).
86. R. Farley, C. Steeh, T. Jackson, M. Krysan, K. Reeves, Continued racial residential segregation in detroit: "chocolate city, vanilla suburbs" revisited. *J. Hous. Res.* **4**, 1–38 (1993).
87. W. A. V. Clark, Ethnic preferences and ethnic perceptions in multi-ethnic settings. *Urban Geogr.* **23**, 237–256 (2002).
88. R. Farley, H. Schuman, S. Bianchi, D. Colasanto, S. Hatchett, "chocolate city, vanilla suburbs": Will the trend toward racially separate communities continue? *Soc. Sci. Res.* **7**, 319–344 (1978).
89. W. Clark, Residential preferences and neighborhood racial segregation: A test of the schelling segregation model. *Demography* **28**, 1–19 (1991).
90. L. Bobo, C. L. Zubrinsky, Attitudes on residential integration: Perceived status differences, mere in-group preference, or racial prejudice? *Soc. Forces* **74**, 883 (1996).
91. G. Giacomin, J. L. Lebowitz, Exact macroscopic description of phase segregation in model alloys with long range interactions. *Phys. Rev. Lett.* **76**, 1094–1097 (1996).
92. R. Lemoy, E. Bertin, P. Jensen, Socio-economic utility and chemical potential. *EPL* **93**, 38002 (2011).
93. M. Fruchart, R. Hanai, P. B. Littlewood, V. Vitelli, Non-reciprocal phase transitions. *Nature* **592**, 363–369 (2021).
94. A. Dinelli *et al.*, Non-reciprocity across scales in active mixtures. *Nat. Commun.* **14**, 7035 (2023).
95. Z. You, A. Baskaran, M. C. Marchetti, Nonreciprocity as a generic route to traveling states. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19767–19772 (2020).
96. S. Saha, J. Agudo-Canalejo, R. Golestanian, Scalar active mixtures: The nonreciprocal Cahn-Hilliard model. *Phys. Rev. X* **10**, 041009 (2020).
97. A. V. Ivlev *et al.*, Statistical mechanics where newton's third law is broken. *Phys. Rev. X* **5**, 011035 (2015).
98. C. Hargus, J. M. Epstein, K. K. Mandadapu, Odd diffusivity of chiral random motion. *Phys. Rev. Lett.* **127**, 178001 (2021).
99. Y. Baek, A. P. Solon, X. Xu, N. Nikola, Y. Kafri, Generic long-range interactions between passive bodies in an active fluid. *Phys. Rev. Lett.* **120**, 058002 (2018).
100. T. Frohoff-Hüsmann, J. Wrembel, U. Thiele, Suppression of coarsening and emergence of oscillatory behavior in a Cahn-Hilliard model with nonvariational coupling. *Phys. Rev. E* **103**, 042602 (2021).
101. F. Brauns, M. C. Marchetti, Nonreciprocal pattern formation of conserved fields. *Phys. Rev. X* **14**, 021014 (2024).
102. C. Scheibner *et al.*, Odd elasticity. *Nat. Phys.* **16**, 475–480 (2020).
103. M. Fruchart, C. Scheibner, V. Vitelli, Odd viscosity and odd elasticity. *Ann. Rev. Condens. Matter Phys.* **14**, 471–510 (2023).
104. J. P. Banerjee, R. Mandal, D. S. Banerjee, S. Thutupalli, M. Rao, Unjamming and emergent nonreciprocity in active ploughing through a compressible viscoelastic fluid. *Nat. Commun.* **13**, 4533 (2022).
105. N. C. Keim, J. D. Paulsen, Z. Zeravcic, S. Sastry, S. R. Nagel, Memory formation in matter. *Rev. Mod. Phys.* **91**, 035002 (2019).
106. J. M. O. de Zarate, *Hydrodynamic Fluctuations in Fluids and Fluid Mixtures*, J. V. Sengers, J. V. Sengers, Eds. (Elsevier Science and Technology, Oxford, ed. 1, 2006).
107. D. G. Martin, D. S. Seara, Y. Avni, M. Fruchart, V. Vitelli, The transition to collective motion in nonreciprocal active matter: Coarse graining agent-based models into fluctuating hydrodynamics. *arXiv [Preprint]* (2023). <https://doi.org/10.48550/arXiv.2307.08251> (Accessed 15 January 2025).
108. V. Verbatz, M. Barthelemy, The growth equation of cities. *Nature* **587**, 397–401 (2020).
109. M. Schlöpfer *et al.*, The universal visitation law of human mobility. *Nature* **593**, 522–527 (2021).
110. C. Borghesi, J. P. Bouchaud, Spatial correlations in vote statistics: A diffusive field model for decision-making. *Eur. Phys. J. B* **75**, 395–404 (2010).
111. A. J. Stier, S. Sajjadi, F. Karimi, L. M. A. Bettencourt, M. G. Berman, Implicit racial biases are lower in more populous more diverse and less segregated us cities. *Nat. Commun.* **15**, 961 (2024).
112. B. Kerr, M. A. Riley, M. W. Feldman, B. J. M. Bohannan, Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**, 171–174 (2002).
113. G. J. Velicer, Social strife in the microbial world. *Trends Microbiol.* **11**, 330–337 (2003).
114. M. E. Hibbing, C. Fuqua, M. R. Parsek, S. B. Peterson, Bacterial competition: Surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8**, 15–25 (2009).
115. E. D. Kelsic, J. Zhao, K. Vetsigian, R. Kishony, Counteraction of antibiotic production and degradation stabilizes microbial communities. *Nature* **521**, 516–519 (2015).
116. J. D. Van Dyken, M. J. Müller, K. M. Mack, M. M. Desai, Spatial population expansion promotes the evolution of cooperation in an experimental prisoner's dilemma. *Curr. Biol.* **23**, 919–923 (2013).
117. C. Ratzke, J. Barrere, J. Gore, Strength of species interactions determines biodiversity and stability in microbial communities. *Nat. Ecol. Evol.* **4**, 376–383 (2020).
118. M. B. Miller, B. L. Bassler, Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199 (2001).
119. C. M. Waters, B. L. Bassler, Quorum sensing: Cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346 (2005).
120. D. R. Mills, R. L. Peterson, S. Spiegelman, An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U.S.A.* **58**, 217–224 (1967).
121. G. P. Smith, Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317 (1985).
122. K. Chen, F. H. Arnold, Enzyme engineering for nonaqueous solvents: Random mutagenesis to enhance activity of subtilisin e in polar organic media. *BioTechnology* **9**, 1073–1077 (1991).
123. N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, S. Leibler, Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
124. E. Kussell, S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
125. A. Murugan *et al.*, Roadmap on biology in time varying environments. *Phys. Biol.* **18**, 041502 (2021).
126. J. Marchi, M. Lässig, A. M. Walczak, T. Mora, Antigenic waves of virus-immune coevolution. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103398118 (2021).
127. M. Lässig, V. Mustonen, A. M. Walczak, Predicting evolution. *Nat. Ecol. Evol.* **1**, 0077 (2017).
128. Alan Turing, The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **237**, 37–72 (1952).
129. D. Gilmour, M. Rembold, M. Leptin, From morphogen to morphogenesis and back. *Nature* **541**, 311–320 (2017).
130. N. S. N. Lam, Spatial interpolation methods: A review. *Am. Cartogr.* **10**, 129–150 (1983).
131. J. E. Guyer, D. Wheeler, J. A. Warren, Fipy: Partial differential equations with python. *Comput. Sci. Eng.* **11**, 6–15 (2009).
132. D. Seara *et al.*, Code and data accompanying "Sociohydrodynamics: data-driven modeling of social behavior". Zenodo. <https://doi.org/10.5281/zenodo.16809485>. Deposited 12 August 2025.