From statistical physics to social sciences / *De la physique statistique aux sciences sociales*

Towards a statistical mechanics of cities

Vers une mécanique statistique des villes

Luís M.A. Bettencourt [a,b,c,*]

[a] *Mansueto Institute for Urban Innovation, Searle 222, 5735 S. Ellis Avenue, University of Chicago, Chicago, 60637, IL, USA*
[b] *Department of Ecology and Evolution, Department of Sociology, Earle 209, 1101 E 57th Street, University of Chicago, Chicago, 60637, IL, USA*
[c] *Santa Fe Institute, 1399 Hyde Pard Road, Santa Fe, 87501, NM, USA*

ARTICLE INFO

ABSTRACT

Cities are some of the most complex dynamical systems in human societies and in nature. There is growing interest in producing more comprehensive quantitative theory, capable of describing many of the features now observable in urban environments, especially those that show empirical regularities across cities of different sizes, geographies, and levels of development. The principal challenge of achieving such a goal is our ability to build frameworks that include realistic but simple accounts of agents' choices and strategic behavior, beyond current approaches in statistical physics or economics. Here, I propose a general framework that integrates agents' behavior with their resource and information management towards seizing opportunities in their environment. I show how this approach integrates urban scaling theory with a statistical mechanics of open-ended (economic) growth. The framework is general and, with appropriate modifications and elaborations, can account for the statistical dynamics of other complex systems.

© 2019 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

RÉSUMÉ

Les villes sont parmi les systèmes dynamiques les plus complexes des sociétés humaines et de la nature. On s'intéresse de plus en plus à la production d'une théorie quantitative plus complète, capable de décrire bon nombre de caractéristiques maintenant observables en milieu urbain, en particulier celles qui montrent des régularités empiriques dans des villes de tailles, de géographies et de niveaux de développement différents. Le principal défi pour atteindre un tel objectif est notre capacité à construire des cadres qui incluent des comptes rendus réalistes, mais simples, des choix et du comportement stratégique des agents, au-delà des approches actuelles en physique statistique ou en économie. Je propose ici un cadre général qui intègre le comportement des agents avec leur gestion des ressources et de l'information pour saisir les opportunités dans leur environnement. Je montre comment cette approche intègre la théorie de l'échelle urbaine à une mécanique statistique de la croissance (économique) ouverte. Le cadre est général et, avec des modifications et des

* Correspondence to: Mansueto Institute for Urban Innovation, Searle 222, 5735 S. Ellis Avenue, University of Chicago, Chicago, 60637, IL, USA.
  E-mail address: bettencourt@uchicago.edu.

précisions appropriées, il peut tenir compte de la dynamique statistique d'autres systèmes complexes.

1. Introduction

Cities, and the general phenomena associated with urbanization, are some of the most complex dynamical systems on Earth, shaping not only modern human societies, but also global resource flows and ecologies [1,2].

Recently, it has become apparent through the comparative analysis of urban systems across space, time, and levels of development that many quantities obey a number of simple statistical regularities, accounting for so-called scaling and agglomeration effects realized by cities [3–6]. This means that quantities – such as a city's built area or its Gross Domestic Product (GDP) – can be written as functions of time, t, and city size, $N_i(t)$ for city i, as

$$Y(N_i, t) = Y_0(t) \, N_i^{\beta}(t) \, e^{\xi_i(t)} \tag{1}$$

where $Y_0(t)$ is a time-dependent pre-factor, common to all cities, β is a scaling exponent also common to all cities, and the $\xi_i(t)$ are deviations from scaling. The scaling "law", $Y(N_i, t) = Y_0(t) N_i(t)^{\beta}$, represents average behavior across all cities in an urban system, while the deviations, $\xi(t)$, capture higher-order statistics [7].

This expression presents several targets for explanation, which have been partially tackled by different modeling approaches. The main goal of urban scaling theory has been the prediction of the exponents β for a number of different urban indicators. This theory expands on a long history of models in economic geography [8–10] and treats cities as self-consistent spatial equilibria, from which expressions for these exponents and the pre-factors, such as Y_0, are derived [6]. Predictions for many of these exponents stand in good agreement with observations from historical urban systems and cities in nations throughout the world today, from the Aztecs [11] to classical Rome [12] and from Brazil [13] to China [3,6], the US [3,5,6] to South Africa [13], from EU nations [4,14] to India [15] and beyond [3,5–7].

The statistical dynamics of these same quantities, however, have proven to be a harder problem because they are manifestly different from standard expectations in statistical physics [7,16]. There are two separate issues here: the first deals with predicting the (approximately) exponential behavior of the pre-factor in time, $Y_0(t) = Y_0(0) \, e^{\overline{\eta}t}$, which requires a theory for growth rates, $\overline{\eta}$. The second deals with the dynamical behavior of the "deviations" around scaling for each city, $\xi_i(t)$.

The first issue is necessarily entangled with theories of economic growth [17]. The present state of these theories in economics remains a work in progress, but a general insight is that "knowledge" is the primary ingredient of growth in incomes per capita [17,18]. The second issue has recently been developed in terms of the aggregation – from agents to cities – of multiplicative random growth models, which generate statistical behavior consistent with theory and data [16].

Here I take a few steps forward towards the integration of these developments into a common framework. This integration provides us with a more general, self-consistent statistical dynamics of cities and a number of formal connections to quantitative modeling in other complex systems, from population dynamics to financial markets.

2. Results

The complexity of cities is ultimately connected to the unique properties of people as social learning agents [8,19,20]. This means that it is the acquisition of resources and associated structuring of information – in a sense to be made specific below – that makes cities complex adaptive systems [6,8]. From this daily hustle at the individual level, immensely complex structures arise, including urban social networks, economic organizations with deep divisions of labor and knowledge, and the elaborate shapes of urban built environments. A statistical mechanics of cities must tell us how these structures are created and can change on average, connecting individual microscopic behavior with macroscopic consequences across many people and time.

The emphasis on choice, information and strongly heterogeneous interaction networks makes any statistical mechanics of cities very different from statistical physics [21]. The essential differences are illustrated schematically in Fig. 1. Any agent – a person, but also possibly an organization – can be described in terms of the dynamics of three general quantities: behavior, resources, and information.

Behavior can be modeled in terms of a trajectory (or "worldline" [6]), which is simply a temporal sequence of the agent's states in interaction with an external environment. This is represented here by $x[t]$, where the degrees of freedom x (typically a vector) could account, for example, for the position of an agent in an urban physical space. Executing any actual trajectory will incur dissipative costs and must therefore yield an income that, over time, compensates for the expenditure, so that the agent can live to fight another day.

The second quantity are *resources*, $r(t)$ [16]. In biological systems, these may be a storage of free energy. In a city, resources may account for more complex variables, including money. The variables $r(t)$ keep track of the compound resources available at each time to our agent, which result from its income minus costs.
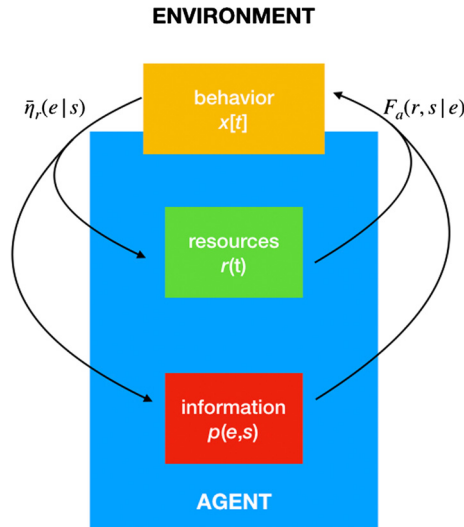
ENVIRONMENT

Fig. 1. A schematic representation of a strategic agent. The agent behaves in the external environment of the city and uses its resources and information to shape its behavior, in order to obtain more resources and information. This allows the agent to maintain its lifestyle in the city, and possibly also to grow its resources and knowledge.

The last quantity is *information*, $i(s, e)$, which is used to make probabilistic choices about the trajectory and associated resource expenditures with the forward-looking goal to extract new resources from the environment. I will assume that this information is personal (internal to the agent) and that information can grow through learning of the associations between signals, s, and states of the experienced environment, e.

An agent in statistical physics – say a Brownian particle – has neither internal free energy nor private information [21]. Thus, it cannot condition its future trajectory, or plan its "growth". As a result, it is at the mercy of entropy forces so that, for sufficient long times, its behavior will become maximally disordered, i.e. it will converge to thermal equilibrium (and "death").

The most essential difference between the statistical mechanics of physical objects and agents in complex systems – from the simplest cells to people in cities – is that the latter carry some resources and information that allows them – statistically, at least – to escape equilibration over long periods of time [22]. For dissipative systems, this can only happen if the agent spends some of its time and resources strategically, acting to harvest new free energy in their environment.

The city, from this perspective, is a general environment where people as learning agents can and must acquire new information, which can sustain them in material terms. This open-ended adaptive behavior, in turn, can lead over time to more complex social and spatial arrangements in populations and to the growth in knowledge for the society as a whole.

2.1. Behavior, resources, and information

The centerpiece of the statistical dynamics of cities sketched in Fig. 1 is the resources' equation [16] (also known in economic geography as the budget constraint [9,10]):

$$\Delta r(t) = y(x[t]) - c(x[t]) \tag{2}$$

This equation is pure accounting, analogous to the conservation of energy in physical systems. It keeps track of net resources (income, y, minus costs, c), accumulated by the agent over time. This appears simple, but it is rather complex because it is a functional of a field, in the form of each agent's trajectory [6]. This field is dynamical and depends on the available resources and information about environmental opportunities, costs and uncertainty. In cities, this involves dealing primarily with the behavior of other people, subject to transportation, land rents, and other living costs. The agent must also manage uncertainty arising from the variability in returns and types of interactions [6,8–10].

It turns out that the challenge of maximizing for resources in cities can be decomposed into two separable problems: one structural, applying at short time scales (days) and another, of growth and instability, applying on longer timescales (years to decades). Because these timescales are very different, these two problems have been approached separately by different disciplines.

2.2. Statistical mechanics of trajectories

The agent's trajectory can involve many different types of degrees of freedom and thus require different explicit dynamics, depending on context and level of detail. The reader may ponder how to model a tree and its leaves, a person in a city,

or a corporation. All of these are examples of agents interacting strategically with their external environments and using stored resources and information towards their maintenance and development.

I now introduce a simple model from statistical physics for the trajectory of an agent in space [23]. This can be used to model a person's mobility in urban built spaces. Consider then, x, as the physical coordinates of the agent in space. Then we can write conventional stochastic equations of motion for x as

$$\dot{x} = p/m, \qquad \dot{p} = -\gamma p - \frac{dV(x)}{dx} + \chi + F_\text{a} \tag{3}$$

Here, $V(x)$ is a standard potential, χ is an environmental thermalizing force, γ is the dissipation rate, and F_a is an additional "self-propelling" force, applied by the agent based on its known information and at the expenditure of resources, r. The corresponding energy, E, associated with the degrees of freedom x, obeys the familiar equation of motion, which can be integrated in time to give

$$\Delta E = -2(\gamma \Delta t)\langle \frac{p^2}{2m} \rangle + 2(\gamma \Delta t)k_\text{B}T + \oint F_\text{a} \cdot dx \tag{4}$$

If we consider that the trajectory works in *cycles*, returning to the initial state (home) after a period of time then $\oint E(t)\,dt = \Delta E = 0$, and we obtain the generalized Einstein relation,

$$\langle \frac{p^2}{2m} \rangle = k_\text{B}T + \frac{1}{2(\gamma \Delta t)}W_\text{a} \tag{5}$$

where I assumed that we are in two spatial dimensions. The average of the kinetic energy (momenta) is taken both over the environmental thermal fluctuations and the cycle. $W_\text{a} = \oint F_\text{a} \cdot dx$ is the work performed over the cycle by the agent's force. In statistical physics, there are typically no internal forces and consequently the trajectory thermalizes. This leads to the Maxwell–Boltzmann formulation of the corresponding velocities [23].

In complex systems, though, the internal force dominates the dynamics, and we have instead:

$$\bar{E} = \langle \frac{p^2}{2m} \rangle \simeq \frac{1}{2(\gamma \Delta t)}W_\text{a} \equiv \bar{w}_\text{a} \tag{6}$$

Thus, the agent's "phase space" is now determined primarily by its use of stored resources and information, which drives the trajectory through F_a. The statistics of the agents' motion follow from the statistics of these internal forces, which must be generated in part to make the most of a stochastic environment. Knowing these forces in detail does determine the trajectory almost exactly (in an over-damped system) and therefore the trajectory's phase space. From the perspective of an outside observer, however, these forces are not knowable. Then we may ask for the statistics of an agent's motion, under the restriction that we only know their average resource use over a cycle (or perhaps their level of income, or even the average income over a population). In an analogy to the derivation of the Maxwell–Boltzmann distribution, this assumption predicts the statistics of an agent's momentum

$$P(p_1, p_2) = \frac{1}{Z} e^{-\frac{p_1^2 + p_2^2}{2m\bar{w}_\text{a}}} \tag{7}$$

with the normalization, $Z = 2\pi m \overline{w}_\text{a}$. Finally, we note that the entropy production from the agent's trajectory over a cycle is simply $\Delta S = W_\text{a}/T$, i.e. the work done by the internal forces in units of the environment's thermal energy exported to the environment as heat, $Q = T\Delta S = W_\text{a}$. $\bar{w}_\text{a}$ is, in general, a fluctuating quantity over cycles, inducing broad statistical variations on the agent's motion and energy dissipation on that timescale.

2.3. Short-term spatial equilibrium

The statistical mechanics described here that is specific to cities deals with the conditions that relate socioeconomic interactions between agents (and their net incomes) to space and associated transportation costs [6,9,10]. The singular characteristic of city life is that no agent can be self-sufficient: an agent's income must be obtained through social exchanges, making it a function of the agent's trajectory [6,9].

In this section, I derive the "mean field" behavior of urban indicators, meaning that $\xi_i(t) \to 0$. This simplification makes the derivations of this section apply to an idealized city where the behavior of individuals can be computed on average.

In this spirit, following [6], I start with income, which must derive from socioeconomic interactions, times factors g_k, which account for the average value of an interaction of type k in units of y. Then we can write the agent's income in the city, over some period of time Δt as

$$y_i = \int dt \sum_k \sum_{j=1}^{N} g_k F_{ij}^k(t) \equiv \sum_k g_k I_i^k \tag{8}$$

Here, $F_{ij}^k(t)$ is the instantaneous social network of the city, accounting for all connections between individuals i, j, at time t and of type k. Then, $I_i^k = \int dt \sum_{k=1}^{N} F_{ij}^k(t)$ is the total degree of individual i, of type k, over the time period. We can make this more explicit requiring that such interactions take place in the space of the city defined by a proximity kernel, K,

$$I_i^k = \int dt \sum_{j=1}^{N} F_{ij}^k(t) = \int dt \sum_{j=1}^{N} P(k|ij)K\left(x_i[t] - x_j[t]\right) \tag{9}$$

To obtain the average result, we can now replace the specific trajectories of each individual j by the background interacting population density, $\Gamma_n = \frac{N-1}{A_n}$, the ratio of the total population minus our agent, $N-1$, over the city's built area, A_n, denoted by the subscript n, for network. Taking the integral over the kernel, K, one obtains the interaction volume of the city swept by our agent, $a_0\ell$, where a_0, ℓ are the interaction cross section and length traveled, respectively. Then,

$$\overline{I}_{i,k} \simeq P(k) \int d^2x\, dt\, \Gamma(x,t) K\left(x - x_i[t]\right) \simeq P(k)\Gamma_n \int d^2x\, dt\, K\left(x - x_i(t)\right) \tag{10a}$$

$$= P(k)a_0\ell\, \Delta t\, \frac{N-1}{A_n} \simeq P(k)a_0\ell\, \Delta t\, \frac{N}{A_n} \tag{10b}$$

This allows us to write the average net income in a city of population N as

$$y \simeq G\Delta t \frac{N}{A_n} \tag{11}$$

where $G = a_0\ell\overline{g}$, with $\overline{g} = \int dk\, g_k P(k)$. The factor Δt is often implicit or omitted, so that y has units of a temporal rate.

Therefore, we obtain a very simple geometric result that, on average, income (and other urban products, proportional to social interactions) is simply proportional to the density of people in urban built spaces times the area swept by an agent. This a result recognizable from the problem of a scattering particle traveling through dense medium. Particular individuals will of course differ in many possible ways.

To bring in the concept of urban space self-consistently requires that we connect incomes to transportation costs, a condition central to all conceptualizations of the city as a spatial equilibrium [8–10], which is analogous to a bound-state problem in physics [6].

The spatial equilibrium condition sets a limit on the maximal area of the city, A, which can be reached using income, y, [6]. In its simplest version, this is written as

$$y = \epsilon A^{1/2} \tag{12}$$

where ϵ is a cost of travel per unit length (and time). When this cost represents energy, it will match the average force F_a, over the agent's spatial path, per unit time. This results in an average relationship between total area and population, $A(N) = aN^\alpha$, with $a = (G/\epsilon)^\alpha$ and $\alpha = 2/3$. This exponent value can be generalized if travel is not in a straight line, e.g., involving some fractal dimension [6]. Accounting for the built space of cities requires knowing how A is "filled" as a function of city size. This does vary somewhat, but empirical evidence and a space filling calculation using streets and lines with some additional observed characteristics [6] suggest the density-dependent average rule [6,24], $A_n \propto A^{1/2}N^{1/2} = a_n N^{1-\delta}$, $a_n \propto a^{1/2}$. This in turn leads to $y = Y_0 N^\delta$, with $Y_0 = G/a_n$, $\delta = \frac{1}{6}$. The finer features of the street network of cities can then also be used to compute average travel costs. This computation is a bit more involved and is given in Ref. [6], so I will not repeat it here. The result is that average travel costs per capita, c, obey a similar scaling law $c = c_0 N^\delta$, $c_0 \propto G^\alpha$. This means that both incomes and transportation costs scale on average in the same superlinear way with population size. However, because the prefactors of y and c have different parameter dependencies, the difference that defines the spatial equilibrium for individuals

$$\Delta r = y - c = f(G)N^\delta \tag{13}$$

can be optimized in G by balancing the design and efficiency of transportation networks to income productivity per interaction [6]. At the maximum, $G = G^*$, $c^* = \frac{1-\alpha}{\alpha}y^*$

$$\Delta r = y - c \xrightarrow{G=G^*} \frac{2\alpha - 1}{\alpha}y^* = \frac{2\alpha - 1}{1-\alpha}c^* = \eta_r r \tag{14}$$

with $\eta_r \equiv \frac{2\alpha-1}{\alpha}b = \frac{2\alpha-1}{1-\alpha}a$ (so that $a = \frac{1-\alpha}{\alpha}b$) the stochastic growth rate for resources. The stochastic variables $a = c/r$, $b = y/r$ are the relative costs and the return rate for investing resources in an uncertain environment. The last step shows that, by maximizing G, which appears in both incomes, and costs, we can make net incomes, $y - c$, simply proportional to income (costs reduce incomes by half, with $\alpha = 2/3$) and thus to the stochastic quantity $b = y/r$, which we will argue below is a constant, or at least a very slow function of time. We will also see next that this is a stochastic quantity, whose average reflects the information the agent has on states of the environment that can generate income.
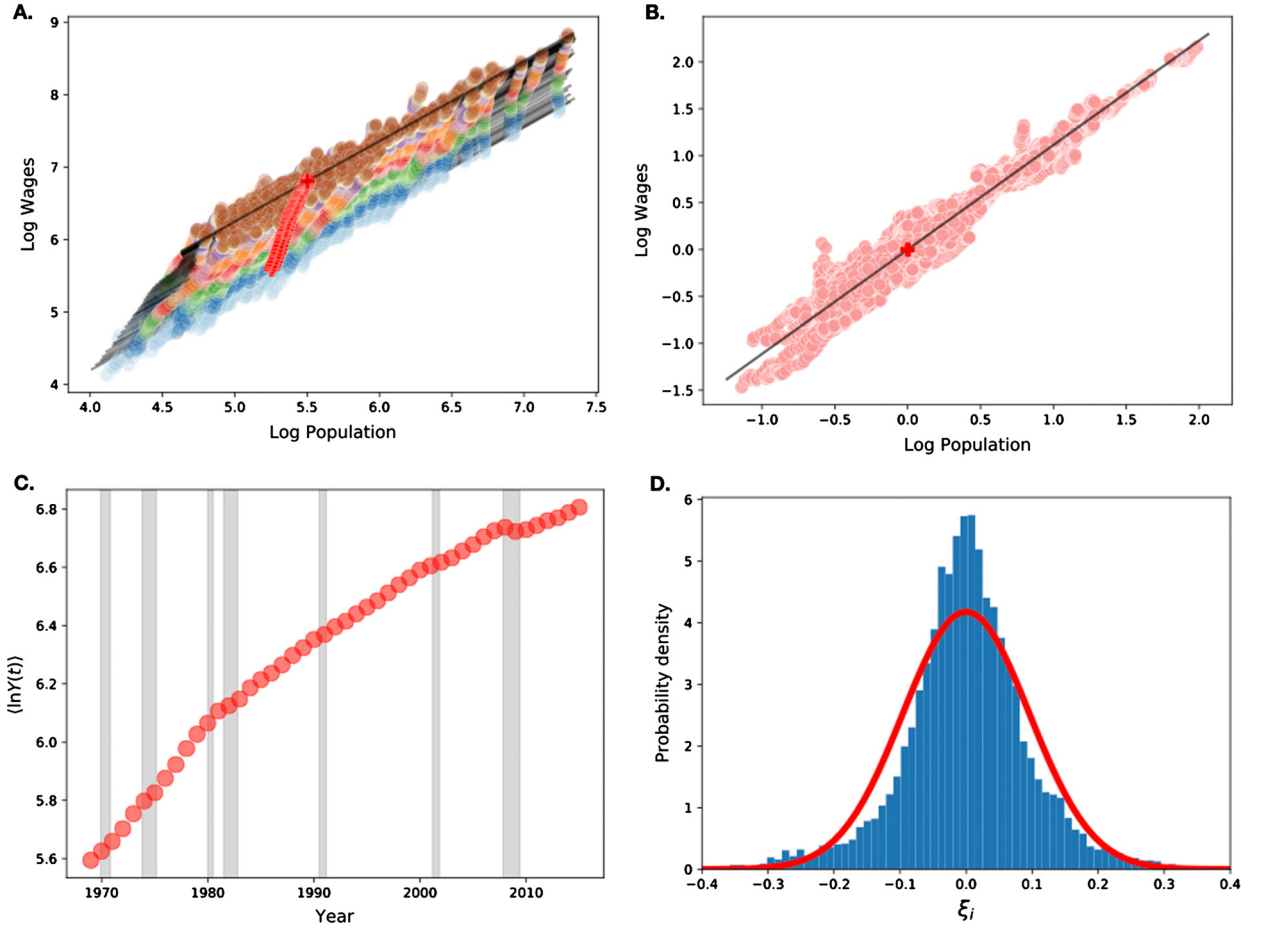
Fig. 2. Scaling, growth and statistics for wages in US Metropolitan Areas. A. The scaling of wages (1969–2016, blue to brown) with data for each year represented by a different color. The solid black line shows the scaling-law Eq. (1), which reproduced the same pattern every year ($\beta \simeq 1.12$) up to the motion of the centers ($\langle \ln N(t) \rangle, \langle \ln Y(t) \rangle$) shown as red crosses. B. The centered data, obtained by translating each data cloud so that its center (red cross) is at the origin (0,0). C. The motion of the average log wage over the system of cities, $\langle \ln Y(t) \rangle$, is approximately exponential in time with constant average rate over long periods of time. Shaded areas show periods of economic recession, at which growth rates tend to suffer a delay or change slope. D. The histogram of deviations from scaling shows that the fluctuations, ξ_i, are well localized and have mean zero. They are well described by a mixture of Gaussians for flow quantities such as wages, following Eq. (18). This is narrower than a single Gaussian (red line), emphasizing an abundance of small fluctuations (near the origin) at short times. The empirical density (blue bars) is well fit by two Gaussians, one narrow and another broad [16].

2.4. Emerging statistics and volatility management

Let us now revisit the resource balance equation taking its statistical significance more seriously [7,16]. This means that we now treat incomes and costs as statistical quantities with fluctuations in time and across individuals. It is well known that the rate equation, treated as a stochastic differential equation can be integrated (using Itō calculus) to give

$$\ln \frac{r(t)}{r(0)} = \left(\bar{\eta}_r - \frac{\sigma_r^2}{2} \right) t + \Theta(t) \tag{15}$$

where $\Theta(t) = \sum_{l=1}^{t} \epsilon_r(t_l)$, and $\epsilon_r(t_l) = \eta_r(t_l) - \bar{\eta}_r$. If the stochastic fluctuations obey the conditions of the central limit theorem, then we will obtain for long times that Θ approaches a normal variable with mean zero and variance $\sigma_\Theta^2(t) = \sigma_r^2 t$. This implies that, in this limit, $\ln r(t)$ becomes normal-distributed and that resources show lognormal statistics. I proposed recently in [16] that this is a consequence of the need for agents to dynamically control their resource growth (as we do by balancing a bank account), and thus avoid large fluctuations for the stochastic growth rates, which can be fatal. Geometric random growth models are also the starting point for most ideas of the dynamics of investment and financial markets [25, 26], and for the analysis of wealth inequality [25,27].

The average resource rates across cities require additional conditions to be satisfied in order to generate city wide lognormal statistics. Averaging Equations (2) and (14) over a population of size N results in effective growth rates and noise for the population mean $\bar{\eta}'_N, \epsilon'_N$ given by

$$\overline{\eta}'_N = \left[1 + \text{covar}_N(\frac{\overline{\eta}_i}{\overline{\eta}_N}, \frac{r_i}{r_N})\right]\overline{\eta}_N, \quad \epsilon'_N = \left[1 + \text{covar}_N(\frac{\epsilon_i}{\epsilon_N}, \frac{r_i}{r_N})\right]\epsilon_N \tag{16}$$

Here the quantities with subscript N without the prime, are standard population averages; those with an i refer to individuals in the population and covar_N is the covariance across the city's population [16]. When these effective fluctuations ϵ'_N obey the conditions of the standard limit theorem, the aggregate dynamics for the city will also be lognormal, which is observed approximately, at least for flow quantities [7,16]. Other more complicated regimes, typically involving power laws, can also emerge [25,28].

The population aggregate quantities are sensitive to resource inequality across the city through the covariance terms, which provide a selection effect familiar from the Price equation of evolutionary dynamics [29]: if richer individuals experience greater growth rates, then the aggregate growth rate in the population will be higher, and vice versa.

Another important effect is that, if the square volatility (or, less likely, the mean growth rate) depends on population size and time, e.g., as $\sigma^2_{\epsilon'_N} = \frac{\sigma^2}{t}\ln N$, then the mean-field scaling exponents will receive a 'running' correction due to fluctuations [16], which in this case is $\beta = \beta - \frac{\sigma^2}{2}$. For more complicated cases, the population size dependence of the effective growth rate may destroy strict scale-invariance (power law scaling), but this effect may be weak if volatilities are small. A statistical theory for the dynamics of the volatilities may then be necessary to account for any of these issues [26,30].

It also follows that incomes and costs inherit statistics from resources so that their amplitudes are related by the average factor, $\overline{b}$. Fluctuations in b contribute to the statistics of these flow quantities, so that for income, for example,

$$\xi_i(t) = \xi_i^r(t) + \ln b_i - \langle\ln b\rangle \tag{17}$$

resulting in [16]

$$d\xi_i(t) \simeq d\xi_i^r(t) + (\ln\overline{b}_i - \langle\ln\overline{b}\rangle)\,dt + \left[\frac{\Omega_i}{\overline{b}_i}\,dW_i(t) - \langle\frac{\Omega}{\overline{b}}\rangle\,dW(t)\right] \tag{18}$$

where Ω_i is the standard deviation of the fluctuations in returns b_i, and similarly for Ω for the returns averaged over the population. Using the spatial equilibrium conditions to relate the total area and, in turn, the networked (built) area of the city to incomes suggests that these spatial variables will also inherit their growth from resources, so that their statistics should also be approximately lognormal [16], but may also show short-term additional stochasticity, as in Fig. 2D. Thus, in the picture developed here, it is the dynamics of information and of associated volatility of growth that leads to systematic change in cities, both in terms of their economic and spatial expansion.

2.5. Growth and information

To complete the agent's dynamics, I want to illustrate the origins of growth rates for incomes and resources and their relation to knowledge of opportunities in a stochastic social environment. The problem was already set up in Section 2.3, by relating incomes to resources, $y = br$. We must now calculate the average returns, $\overline{b}$.

To proceed and establish the connection to information, I will assume that agents will seek to maximize their average resource growth rate over many cycles [31,32]. This is equivalent to the maximization of a logarithmic utility, but different from the behavior that would follow from other forms of utility optimized over the short term, as is often done in microeconomics [33].

I will consider here only a simple scenario, sufficient to motivate the introduction of individualized information for each agent. Consider the case in which the agent's resources are apportioned to different environmental stochastic events e_i, in terms of a fraction $f(e_i)$, $\sum_i f(e_i) = 1$, and that correctly predicting the outcome provides a payoff $o(e_i) \geq 1$. Thus, the agent's resources grow according to

$$r \to f(e_i)o(e_i)r \tag{19}$$

with the expectation that the payoff is large enough so that $f(e_i)\,o(e_i) > 1$; otherwise such event is not worthy of consideration. Note that the agent, in addition, will need to dedicate time and resources to other activities that support this investment: being productive at work requires rest, food, shelter, leisure time, etc. These resource investments are always necessary, but are in effect directed to alternative productive opportunities, e_i, through the differential allocation of the agent's time and effort.

The general idea is that, in cities, there are large sets of events that are, in principle, worth dedicating time and effort to, in this sense. Then, after τ cycles, the agents resources will be

$$r(\tau) = \Pi_{t=1}^{\tau} f(e_i(t))o(e_i(t))\,r(0) \tag{20}$$

which depends on the specific sequence of the $e_i(t)$ observed over time. We can ask that the growth rate of resources is maximized on average. Assuming that the sum in Eq. (20) obeys the law of large numbers we can write the growth rate as

$$\overline{\eta}_r = \lim_{\tau \to \infty} \frac{1}{\tau} \ln \frac{r(\tau)}{r(0)} = v_\tau \sum_{i=1}^{l_s} P(e_i) \ln f(e_i) o(e_i) \tag{21}$$

where $P(e)$ is the probability distribution of environmental states, which I assume is approximately stationary. The size of the state space of the environment is given by l_s. The rate, $v_\tau = \tau/t$, is the *speed of allocations*, or the number investment moves per unit time. This quantity may vary over time reflecting ebbs and flows of optimism and opportunity. Below I set it to unity, for simplicity.

We can now ask for the assignment of resources that maximizes this rate. This is a constrained optimization problem, since the $\sum_j^{l_s} f_j = 1$, which results in $f(e_i) = p(e_i)$. In gambling, this is known as *proportional betting* [32]. The resulting maximal rate (denoted by the asterisk) is therefore

$$\overline{\eta}_r^* = \overline{\ln o} - H(E)$$

where $H(E) = -\sum_i P(e_i) \ln p(e_i) \geq 0$, is the Shannon entropy of the environment and $\overline{\ln o} = \sum_{i=1}^{l_s} P(e_i) \ln o(e_i)$. We see that, for the growth rate to be positive, the average rewards in the environment $\overline{\ln o}$ must be sufficiently large, i.e. there must be some "free lunch" (this is not so unusual, farmers receive a free lunch from the sun, social benefits and public goods provide some free resources to people in cities). On the other hand, a higher entropy (uncertainly) in the space of possible events, $H(E)$, hurts our agent proportionally [32].

We can now explore the meaning of this maximal growth rate. In gambling [31,32], the payoffs $o(e_i)$ are given by an aggregator, such as a bookie or the "house", who have access to the resource allocations of many agents. Prediction markets work in a similar way: prices provide estimates of the participants' average beliefs [33]. In the case of fair odds, the market estimate is $o(e_j) = 1/P_m(e_j)$, where $P_m(e_i)$ is the estimated frequency of events by the aggregator. Then the average growth rate can be written as

$$\overline{\eta}_r = \sum_j p(e_j) \ln \frac{f(e_j)}{P(e_j)} \frac{P(e_j)}{P_m(e_j)} = D_{KL}(P||P_m) - D_{KL}(P||f)$$

where D_{KL} is the relative entropy (or Kullback–Leibler divergence [32]) between the two probability estimates. This is a positive quantity that measures how different the two distributions are, vanishing when they coincide. This means that agents can benefit from the aggregator's wrong estimates, but that they also pay a price for their ignorance of the real event probabilities [31,32]. If markets are perfect estimators, the best someone can do is also to perfectly estimate the event probabilities [32,33]. Then the maximal rate is zero! This reflects the fact that the agent has no particular insight relative to the crowd and is, in the sense of information, redundant [34]. So, in this picture, unless there are free lunches, positive growth requires an information advantage [35].

This calls for a more elaborate picture of the growth rates, accounting for information specific to particular agents. It is well known that, if an agent has "inside knowledge" of the states of the environment, they can generate higher returns to their resources [31,32]. Let us call the information-bearing variables available to the agent "signals", s. Signals may come from the environment or from the agent's past experience (memory). Let us suppose that at each time for resource allocation, the agent can consult these signals. Then, the agent will apportion resources conditional on s. This now makes the estimation of the growth rate become

$$\overline{\eta}_r = \sum_{i,j} P(e_i, s_j) \ln f(e_i|s_j) o(e_i)$$

Maximizing this rate gives the allocation choice [31], $f(e_i|s_j) = P(e_i|s_j)$. For any payoff, the gain in average rate for using this additional signal is $\Delta \overline{\eta}_r = i(E, S) = \sum_{i,j} P(e_i, s_j) \ln \frac{P(e_i, s_j)}{P(e_i)P(s_j)}$, which is the definition of the *mutual information* between the signal and the states of the environment. If odds are fair [31,32], then

$$\overline{\eta}_r^* = i(E, S)$$

That is, the maximal average growth rate of resources is set by the special information that the agent has on the environment. Finally, it is worth noting that, in these circumstances, the agent must have an internal model of the probability of the signal given the state of the environment, $P(s_j|e_i)$. This model can be used in turn to learn the probability

$$f(e_i|s_j) = \frac{P(s_j|e_j)}{P(s_j)} f(e_i)$$

given that the trajectory over cycles results in the observed states of the environment and of the signal. In this way, the agent's trajectory is both a source of income and of evidence, which drive the accumulation of *both* resources and information. To the extent that the learning is imperfect, the growth rate will be diminished by a factor of $D_{KL}[P(e|s)||f(e|s)]$ [35].

This simple result can be made more complicated in a number of ways [32] briefly discussed, but not pursued, below. The point of its introduction here is that, though simple, it points to a number of features of urban life. First, competitive

learning and adaptation above and beyond all other agents is important: this means specialization so that different agents can have specific information on different aspects of their common environment [36,37]. Better signals and learning are key: making information as high as possible depends on the choices of signals and on learning their association with environmental states. Access to environmental variables with high payoffs is also essential, an issue that is often not open to all people due to inequality of opportunity [18,27,38].

Several other issues are important in creating a more realistic theory of resource growth rates and establishing their relationship to information. One fundamental issue has to do with how information aggregates in populations, from an individual to an entire city. Information does not follow the familiar rules of conserved quantities – such as energy or time – and can be *redundant* or *synergistic* [34]. Redundant means that several agents have the same knowledge. This is typical of poor, subsistence societies where primary modes of production are replicated over the land. As a result, the total information in such a society is not much larger than that of a single household and it does not scale with total population size [37]. Such societies have typically very low rates of economic growth, but can be very stable. Conversely, information can be synergistic, meaning that different agents have different knowledge of the environment that, when taken together, has greater information than the sum of the parts. This occurs in interdependent "ecologies" and is typical of the deep divisions of knowledge characteristic of cities [37]. In this case, information will scale up with the addition of every agent, because each has some unique productive knowledge.

What then happens in cities? Redundancy or synergy? Necessarily there will be a bit of both. An argument for the value of the total information in the city can be given as follows. First consider different professions as a source of different productive knowledge. In US cities, the total number of different professions scales with city size approximately like $D(N) = D_0 N^{1-\delta}$ with $\delta \simeq 1/6$ [36]. This is a measure of specialization because it implies that the number of professional tasks *per person* is declining with city size as $\sim D_0 N^{-\delta}$. If these agents have a fixed "productive" time, t_T, on average, allocated to these tasks, they will spend more time per task, by a factor of about $t_T = t_0 N^\delta$. This implies a larger number of learning opportunities at which to get better at the task. Thus, we may estimate roughly that the average information per individual will scale like $\bar{i}(E, S) \simeq D_0 N^{-\delta} t_0 N^\delta \sim D_0 t_0$, independent of city size [34,37]. That is, even though in larger cities there are more professions, each with different specialized knowledge, there are fewer per capita. This is compensated for by greater depth of "learning by doing", gaining more information of the processes involved. As a consequence, in an urban system, the overall growth rate of incomes can be city size independent (Gibrat's law), even though it will depend on how many people can participate in the collective learning process at the national level.

2.6. Summary

In summary, in the simplest scenario we can write the statistics of total resources for all agents in a city (not per capita) over time as

$$R(t, N_i) = R_0(t) N_i(t)^\beta \, e^{\xi_i(t)}$$

where $R_0(t) = r_0(0) \, e^{(\overline{\eta}_r - \frac{\sigma_r^2}{2})t}$, $\overline{\eta}_r \to \bar{i}(E, S)$, $\beta = 1 + \delta \simeq 7/6$, [6] and where the fluctuations tend to Gaussian behavior at long times, with $\xi(t) = \xi(0) + \sigma_r t \, dW(t)$, with $dW(t)$ an approximately normal stochastic variable with unit variance. A statistical mechanics of cities allows us to compute each of these parameters beyond the simplest picture, including systematic corrections due to fluctuations and variations in the starting hypotheses. In this sense, scaling exponents may vary away from mean-field predictions and show some size and time dependence, fluctuations ξ_i may be non-Gaussian, especially for short times, and growth rates may depend on features of the environment and the agent's history, besides information.

However, if volatilities related to returns are sufficiently small and stationary, the statistical mechanics of cities will become much simpler, given by the mean-field scenario [6] with some calculable corrections. The regime of large volatilities remains unexplored, but clearly would lead in extreme cases to the demise of net growth and, in turn, of cities altogether [6, 7,39].

The statistical mechanics of cities that follows is not stationary. It depends on the self-consistent procurement of resources into the system and their partial dissipation as heat. Rates of change are in general exponential, set by the information about environmental affordances to agents in the system. A minimal amount of information (and average growth) is necessary, in this picture, to overcome the compounded volatility of multiplicative processes, without which resources will tend to zero as variance increases, leading to individual and systemic collapse.

3. Discussion and outlook

The development of a more comprehensive statistical mechanics of cities and other complex systems requires models of agents capable of active adaptive behavior, in ways that go beyond present models in statistical physics or economics [20, 39]. Here, I proposed a simple framework where agents modify their behavior as a function of their own extant resources and knowledge. While such a model can be as complex as one may want to make it, I showed that a simple solution can be obtained by separating fast degrees of freedom associated with cyclical structural behavior (budget constraint, spatial

"equilibrium") from slow degrees of freedom implicated in statistical variability, growth and learning, which become relevant only in the longer term, over many fast cycles.

In cities, and in human societies more generally, survival and cyclic behavior must be achieved on a daily timescale [6,9], while fast population and economic growth is optional and happens, at best, at a rate of a few percent per annum [18]. This is a daily rate of the order of 0.01% (for 3.5%/year), justifying this separation. Historically, growth rates have been much smaller.

Treating cities as large populations over long (annual) time scales then affords us a true statistical perspective, which unlike in physics is open-ended and can generate very broad (lognormal and power law) statistics [20,25,28], complexity and growth over the long term. Like in statistical physics, however, we are able to predict aggregate properties of cities and urban life.

A quantitative and predictive approach to these issues is just starting to become a reality [20] thanks to more disaggregated data and progress in multidisciplinary theory. This suggests a fertile ground for developments capable of tangible socioeconomic impact and for conceptual and technical progress on a more general statistical mechanics of human societies and other complex systems.

Acknowledgements

References

[1] M. Alberti, Advances in Urban Ecology: Integrating Humans and Ecological Processes in Urban Ecosystems, Springer, New York, 2008.
[2] C.M. Kennedy, The Evolution of Great World Cities: Urban Wealth and Economic Growth, University of Toronto Press, Toronto, Buffalo, 2011.
[3] L.M.A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G.B. West, Growth, innovation, scaling, and the pace of life in cities, Proc. Natl. Acad. Sci. USA 104 (17) (2007) 7301–7306, https://doi.org/10.1073/pnas.0610172104.
[4] D. Pumain, F. Paulus, C. Vacchiani-Marcuzzo, J. Lobo, An evolutionary theory for interpreting urban scaling laws, Cybergeo (2006), https://doi.org/10.4000/cybergeo.2519.
[5] M. Batty, The size, scale, and shape of cities, Science 319 (5864) (2008) 769–771, https://doi.org/10.1126/science.1151419.
[6] L.M.A. Bettencourt, The origins of scaling in cities, Science 340 (6139) (2013) 1438–1441, https://doi.org/10.1126/science.1235823.
[7] A. Gomez-Lievano, H. Youn, L.M.A. Bettencourt, The statistics of urban scaling and their connection to Zipf's law, PLoS ONE 7 (7) (2012) e40393, https://doi.org/10.1371/journal.pone.0040393.
[8] R.E. Park, E.W. Burgess, R.D. McKenzie, The City, University of Chicago Press, 1984.
[9] W. Alonso, Location and Land Use: Toward a General Theory of Land Rent, Harvard Univ. Press, Cambridge, MA, USA, 1977.
[10] E.L. Glaeser, Cities, Agglomeration, and Spatial Equilibrium, Oxford University Press, Oxford, UK, 2008.
[11] S.G. Ortman, A.H.F. Cabaniss, J.O. Sturm, L.M.A. Bettencourt, Settlement scaling and increasing returns in an ancient society, Sci. Adv. 1 (1) (2015) e1400066, https://doi.org/10.1126/sciadv.1400066.
[12] J.W. Hanson, S.G. Ortman, A systematic method for estimating the populations of Greek and Roman settlements, J. Roman Archaeol. 30 (2017) 301–324, https://doi.org/10.1017/S1047759400074134.
[13] C. Brelsford, J. Lobo, J. Hand, L.M.A. Bettencourt, Heterogeneity and scale of sustainable development in cities, Proc. Natl. Acad. Sci. USA 114 (34) (2017) 8963–8968, https://doi.org/10.1073/pnas.1606033114.
[14] L.M.A. Bettencourt, J. Lobo, Urban scaling in Europe, J. R. Soc. Interface 13 (116) (2016) 20160005, https://doi.org/10.1098/rsif.2016.0005.
[15] A. Sahasranaman, L.M.A. Bettencourt, Urban geography and scaling of contemporary Indian cities, J. R. Soc. Interface 16 (152) (2019) 20180758, https://doi.org/10.1098/rsif.2018.0758.
[16] L.M.A. Bettencourt, Urban Growth and the Emergence Statistics of Cities, Working Paper, Mansueto Institute, 2018.
[17] R.J. Barro, X.I.S. i Martin, Economic Growth, The MIT Press, Cambridge, MA, USA, 2003.
[18] C.I. Jones, P.M. Romer, The new Kaldor facts: ideas, institutions, population, and human capital, Am. Econ. J. Macroecon. 2 (1) (2010) 224–245, https://doi.org/10.1257/mac.2.1.224.
[19] R.E. Lucas, On the mechanics of economic development, J. Monet. Econ. 22 (1) (1988) 3–42, https://doi.org/10.1016/0304-3932(88)90168-7.
[20] J. Johnson, Non-Equilibrium Social Science and Policy: Introduction and Essays on New and Changing Paradigms in Socio-Economic Thinking, Springer, Berlin, Heidelberg, New York, 2016.
[21] S. keng Ma, Statistical Mechanics, World Scientific, Philadelphia, PA, USA, 1985.
[22] E. Schroedinger, What Is Life? The Physical Aspect of the Living Cell, Cambridge University Press, Cambridge, New York, 1992.
[23] J. Zinn-Justin, Phase Transitions and Renormalization Group, Oxford University Press, Oxford, New York, 2007.
[24] H. Samaniego, M.E. Moses, Cities as organisms: allometric scaling of urban road networks, J. Transp. Land Use 1 (1) (2008) 21–39, https://doi.org/10.5198/jtlu.v1i1.29.
[25] J.-P. Bouchaud, M. Mezard, Wealth condensation in a simple model of economy, Physica A 282 (3–4) (2000) 536–545, https://doi.org/10.1016/S0378-4371(00)00205-3.
[26] W. Paul, J. Baschnagel, Stochastic Processes: From Physics to Finance, Springer, Heidelberg, Germany, 2013.
[27] T. Piketty, G. Zucman, Capital is back: wealth–income ratios in rich countries 1700–2010, Q. J. Econ. 129 (3) (2014) 1255–1310, https://doi.org/10.1093/qje/qju018.
[28] S. Solomon, P. Richmond, Power laws of wealth, market order volumes and market returns, Physica A 299 (1–2) (2001) 188–197, https://doi.org/10.1016/S0378-4371(01)00295-3.
[29] S.A. Frank, Natural selection. IV. The Price equation*, J. Evol. Biol. 25 (6) (2012) 1002–1019, https://doi.org/10.1111/j.1420-9101.2012.02498.x.
[30] L. Bergomi, Stochastic Volatility Modeling, CRC Press, Boca Raton, FL, USA, 2016.
[31] J.L. Kelly, A new interpretation of information rate, IRE Trans. Inf. Theory 2 (3) (1956) 185–189, https://doi.org/10.1109/TIT.1956.1056803.
[32] T.M. Cover, J.A. Thomas, Information theory and statistics, in: Elements of Information Theory, 2nd edition, Wiley, 2006.
[33] J. Wolfers, E. Zitzewitz, Interpreting Prediction Market Prices as Probabilities, https://ssrn.com/abstract=898597, 2006.
[34] L.M.A. Bettencourt, The rules of information aggregation and emergence of collective intelligent behavior, Top. Cogn. Sci. 1 (4) (2009) 598–620, https://doi.org/10.1111/j.1756-8765.2009.01047.x.

[35] M. Hilbert, The more you know, the more you can grow: an information theoretic approach to growth in the information age, Entropy 19 (2) (2017) 82, https://doi.org/10.3390/e19020082.

[36] L.M.A. Bettencourt, H. Samaniego, H. Youn, Professional diversity and the productivity of cities, Sci. Rep. 4 (2014) 5393, https://doi.org/10.1038/srep05393.

[37] L.M.A. Bettencourt, Impact of changing technology on the evolution of complex informational networks, Proc. IEEE 102 (12) (2014) 1878–1891, https://doi.org/10.1109/JPROC.2014.2367132.

[38] J.R. Hipp, Income inequality, race, and place: does the distribution of race and class within neighborhoods affect crime rates?, Criminology 45 (3) (2007) 665–697, https://doi.org/10.1111/j.1745-9125.2007.00088.x.

[39] P.M. Allen, Cities and Regions as Self-organizing Systems: Models of Complexity, Vol. 1, Gordon and Breach, Luxembourg, 1997.