A MATHEMATICAL DEFINITION OF COMPLEX ADAPTIVE SYSTEM AS INTERACTION SPACE

PAOLO GIORDANO

Abstract. We define a mathematical notion of complex adaptive system by following the original intuition of G.K. Zipf about the *principle of least effort*, an intuitive idea which is nowadays informally widespread in complex systems modeling: adaptation means minimizing suitable costs and, at the same time, effectively distribute available resources. Generalizing Mandelbrot's ideas, we also understand when a large class of these systems satisfy a power law, and we give details in case of evolving languages. We also illustrate this notion with theorems describing Von Thünen-like models. These describe the appearance of emergent patterns in a large class of complex systems, and we detail the application in land use theory as well as in several other systems. We use the language of interaction spaces theory so that, thanks to its universality, this definition of complex adaptive system and emergent patter is formulated in very general terms, with a precise mathematical language, coupled to a clear intuitive description, but also exemplified using real-world systems like in natural selection, urban growth, text mining, microeconomics, biology, etc.

Contents

This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/PAT9221023.

1. Introduction: Complex adaptive systems and emergent patterns following Zipf's idea

It is widely recognized that complex adaptive systems (CAS) are among the most interesting types of complex systems (CS) (see, e.g., [28, 29]), and there are several attempts to provide a precise definition of CAS, see e.g. [30, 4, 20, 44].

Due to its unifying properties, see [18], interaction spaces (IS) theory can provide the appropriate context for a *general mathematical definition* of CAS, and the understanding of general dynamical laws governing these systems. In the present paper, we use all the notions of IS theory introduced in [18].

The main idea is to try a mathematical formalization of the original ideas expressed by G.K. Zipf in [45], because they seem intuitively clear, meaningful and applicable to a large class of CAS, where some kind of optimization and, at the same time, of information sharing between interacting entities in a CS is nowadays informally frequently used, see e.g. [3, 12, 16, 20, 25, 28, 22, 30, 38] and references therein.

In [45], the appearing of a power law is connected to what we think is an adapting behavior of the system. Zipf's *principle of least effort* intuitively explains this adaptation as a result of *two* opposing processes: *unification* and *diversification*. We will try to intuitively explain this principle through several examples listed below. Clearly, both these terms in [45] are sufficiently imprecise and could hence be misinterpreted. The mathematical formalization we present here also aims to gain a more clear understanding of these processes.

In our insight, unification processes are related to interactions that try to *decrease* convenient *costs* (possibly meant in an abstract way: e.g. loss of profits, loss of common goods, probability to get hungry, probability to loose reproductive possibilities, hormones such as cortisol or dopamine, etc.), whereas diversification processes are linked to *long term changes* of suitable interactions, i.e. to the increasing of the *causal information shared through the goods* exchanged between agents and patients of these interactions, see [18] for an explanation of these terms of IS theory. It is the implementation of these interactions and the most diversified exchange of fluxes of goods that enable the adaptive population to be resilient *and* keep a low value of costs.

To start a first understanding and a preliminary intuitive validation of the subsequent mathematical definition, we can keep in mind the following examples:

1) In a natural language, unification processes drift to shorten most frequently used words (or better: frequently used sounds, see [8]); diversification ones make evolve the language towards longer and specialized words, [13, 45].
2) In cities and their markets development, unification tends to bring near people so as to decrease suitable costs of living; diversification tends to use all the possible living locations so as to approach the appropriate rent costs, [42, 16].
3) In natural selection, unification forces push giraffes to search for eatable trees (see e.g. [9] and references therein); diversification can select all the best genetic codes that allow for a longer neck, [14]. We recall that a costs decreasing process (unification) can cause an evolution into different phenotypes (diversification), see [21]. More generally, natural selection seems to result by the dynamics of two processes: representation of biological information as chemical properties and

control flow (diversification), and the energetic constraints limiting the maintenance of that information (unification), [35, 36, 37]. See [34] for an alternative evolutionary explanation of long neck in giraffes.

4) Determining the direction to navigate to a safe place, such as a home or nest, is a fundamental behavior for all complex animals and a crucial first step in navigation. We can say that unification mental processes evaluate costs related to the achievement of these goals, whereas diversification ones are related to judgments based on a previously learned or planned behavior, [10].
5) Companies with a longer life span are able not only to decrease costs and increase profits (unification), but are also able to adapt to their complex environment by implementing long-term robustness. The latter are often realized through diversification processes such as: maintain heterogeneity of people, ideas, and endeavors, and preserve redundancy among components, [33].
6) In autism (but also in schizophrenia), a necessary decreasing in high costs related to social interactions (unification) is sometimes compensated by higher abilities (diversification) in very specialized or creative activities, [31, 11].
7) The ability to manage costs related to large varieties of goods (unification) is related to the ability to implement the same stable economic decisions (interactions) applicable to different goods (diversification), [43].
8) Phyllotaxis, the regular arrangement of leaves or flowers around a plant stem, is an example of developmental pattern formation. Phyllotaxis is characterized by the divergence angles between the organs, the most common angle being 137.5°, the golden angle. Different approaches and hypotheses has been used to model this formation mechanism, see e.g. [23]. In this process, we can see unification forces related to energy exploitation by each primordium, and diversification forces that tend to uniformly distribute these energy sources between old and new primordia.
9) An example of **non**-adaptive but still complex system is traders payments of Wall Street employees. It is well known that this payment has a base salary and a bonus, which is usually a percentage of trader's profit. If traders lose, they still get their base, and only if their loss is great enough, they are fired. However, they never have to return the money lost by the company due to their wrong trading. This is a clear financial incentive to be reckless because it rewards short-term gains (costs which identify unification forces) without regard to long-term consequences (diversification forces), [1].
10) The efficiency of a parliament (unification processes interpreted as decreasing of suitable costs) can be improved by inserting randomly selected legislators (increasing of diversification among legislators), see [32].
11) Whereas classical economic theories prescribe specialization of countries industrial production, inspection of the country databases of exported products shows that this is not the case: successful countries are extremely diversified, in analogy with biosystems evolving in a competitive dynamical environment. In fact, together with classical and necessary costs reduction (unification), diversification represents the hidden potential for development and growth, [39].
12) The evolutionary emergence of an egalitarian attitude in a population can be explained by using an evolutionary model of group-living individuals competing for resources and reproductive success (i.e. unification as costs to decrease), see [17]. Although the differences in fighting abilities lead to the emergence of

hierarchies where stronger individuals take away resources from weaker individuals, the logic of within-group competition implies that each individual benefits if the transfer of resources from a weaker group member to a stronger one is prevented. This model shows that this effect can result in the evolution of a particular behavior causing individuals to interfere in a bully–victim conflict on the side of the victim. A necessary condition in this process is a high efficiency of coalitions in conflicts against the bullies. The egalitarian drive leads to a dramatic reduction in within-group inequality. Simultaneously, it creates the conditions for the emergence of inequity aversion via the internalization of these adapted behavioral rules. All these interactions are general because they can be applied to different situations. They hence represent a long-term and diversified improving of society.

13) The present network of financial exposures among institutions (e.g. banks, companies) shows that this system is **not** well adapted. The centrality of certain institutions does not allow the system to be resilient to financial fails of few institutions, see [7]. As we will see later, this is the same mechanism showed by monopolies and the corresponding lack of diversity in the system. We could say that the system is flawed because it lacks in interactions that act when a node of the financial network fail: a more diversified network would show a greater resilience because of a lower centrality of these nodes. It is also clear that a globally directed taxation system can prevent the formation of such a centralized nodes. Even if these global taxes do not allow a maximization of profits (which could be clearly higher without this tax), they could be rightly justified as a measure to prevent global irreparable problems to the entire system, and hence to the institutions themselves. See also [41].

To underline that we are going to give our interpretation of Zipf's principle of least effort, we prefer to name our mathematical version *generalized evolution principle* (GEP). This name has also the merit to link this adapting dynamics to evolutionary theories that, in our opinion, are well inscribed into it.

Before starting an intuitive description of the GEP, we have first to specify what is a global state of an IS.

2. Global states of an interaction space

There are five sources of (possible) randomness acting in an IS $\mathfrak{I}$. One of them derives from the (possible) stochastic evolution equation, see [18, (EE)]. We recall that this also includes the stochastic behaviour of activation state $\mathrm{ac}_i^p(t) \in [0,1]$ and goods $\gamma_i(t) \in R_i$, see [18, Rem. 1 (e)]. We have this type of stochastic behavior because we do not want to model some details of the evolution of the interaction i, or because it is not possible at all: it can happens, e.g., that a pedestrian randomly chooses between two exists which are equally located with respect to its present position: a deterministic model could not be reasonable for a non-trivial large class of pedestrians.

Three other sources of randomness are the occurrence times $t_i^{\mathrm{s}}(t)$, $t_i^{\mathrm{a}}(t)$ and $t_i^{\mathrm{o}}(t)$, see [18, Sec. 3.5.1].

The last source is due to neighborhoods $\mathcal{N}_i(t) \subseteq E$ of interactions $i \in I$, see [18, (NE)].

Therefore, in principle, it is possible to define a new probability space that combines all these (possible) random sources. Elementary events of this space are of

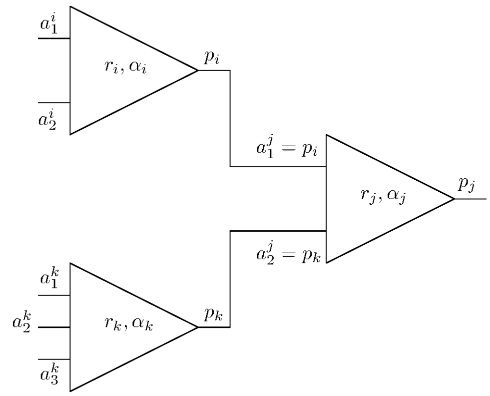

FIGURE 2.1. An example of interactions in cause-effect cascade.

the form $(t_s, t_a, t_o, \mathcal{N}, \omega)$, where $(t_s, t_a, t_o) \in [t_{st}, t_{end}]^3$ are e.g. distributed as explained in [18, Sec. 3.5.1], $\mathcal{N} \subseteq E$ are all possible neighbourhoods, probabilistically distributed depending on modeling choices, and $\omega \in \Omega_p$ is distributed as P_p for all patients p according to the evolution equation, see [18, (EE)].

The joint probability that we have to settle on this space of elementary events is certainly not easy to set. On the one hand, the universal properties of IS allow us to state that the aforementioned ones are all the possible sources of randomness that we have to take into account in several type of models. On the other hand, the causal graph defined by the activation function, [18, Sec. 2], can be of great help in finding this probability: we can say that the interaction i is a cause of the interaction j if along any possible solution of the IS the interaction i always precedes j, and if i activates for j an agent of j (see e.g. Fig. 2.1).

If the resulting directed cause-effect graph is acyclic, we can interpret it as a beliefs network and apply the methods of Bayesian networks to define the joint probability, see e.g. [15].

A *global state space $\bar{M}_{\mathcal{P},J,t}$ of the population $\mathcal{P} \subseteq E$ and the interactions $J \subseteq I$ up to time $t \in [t_{st}, t_{end}]$* is given by three components, each being a subset of all the possible *paths* (states, times, neighbourhoods) of our IS (recall that Y^X is the set of all the functions $f : X \longrightarrow Y$ and, to understand, that if the index set $J = \{j_1, \dots, j_n\}$ is finite, then the product of sets is $\prod_{j\in J} S_j = S_{j_1} \times \dots \times S_{j_n}$):

$$\bar{M}_{\mathcal{P},J,t} := \bar{M}^{\mathrm{s}}_{\mathcal{P},t} \times M^{\mathrm{t}}_{J,t} \times M^{\mathrm{n}}_{J,t} \tag{2.1}$$

$$\bar{M}^{\mathrm{s}}_{\mathcal{P},t} \subseteq \left(\prod_{e\in\mathcal{P}} \bar{S}_e\right)^{[t_{st},t]} =: \bar{S}_{\mathcal{P}}^{[t_{st},t]} \tag{2.2}$$

$$M^{\mathrm{t}}_{J,t} \subseteq \left([t_{st}, t_{end}]^3\right)^{J\times[t_{st},t]}$$

$$M^{\mathrm{n}}_{J,t} \subseteq \{\mathcal{N} \mid \mathcal{N} \subseteq E\}^{J\times[t_{st},t]}$$

$$\bar{M}_{\mathcal{P},J} := \bigcup_{t\in[t_{st},t_{end}]} \bar{M}_{\mathcal{P},J,t}. \tag{2.3}$$

The space $\bar{M}_{\mathcal{P},J}$ is called *global state space of $\mathcal{P}$ and J.*

Recall that the state space $\bar{S}_e$ also contains activation and states of goods, whereas the proper state space S_e is a subspace of $\bar{S}_e$ that includes all the other proper state values, see [18, (ST)]. The optimization performed by a CAS occurs in the space $\bar{M}_{\mathcal{P},J,t}$ and hence it also depends on the choice of such a global state space, see also Sec. 5.1.

A conceptual explanation of supersets appearing in (2.2) is as follows: for each interacting entity $e \in \mathcal{P}$ of the given population, in $\bar{M}^{\mathrm{s}}_{\mathcal{P},t}$ we have all the state *time functions* $x_e : [t_{\mathrm{st}}, t] \longrightarrow \bar{S}_e$ of our IS; in $M^{\mathrm{t}}_{J,t}$ we can consider all the occurrence times $t^{\mathrm{s}}_j(-)$, $t^{\mathrm{a}}_j(-)$, $t^{\mathrm{o}}_j(-) : [t_{\mathrm{st}}, t] \longrightarrow [t_{\mathrm{st}}, t_{\mathrm{end}}]$ of each interaction $j \in J$; finally, in $M^{\mathrm{n}}_{J,t}$ we have the neighbourhood $\mathcal{N}_j(-) : [t_{\mathrm{st}}, t] \longrightarrow \{\mathcal{N} \mid \mathcal{N} \subseteq E\}$ of $j \in J$ as function of time up to t.

Even if it is natural to consider in the global state space all the possible paths of independent variables we can consider in our IS model, in the mathematical definition of the GEP Def. 2 we will see why this is important to achieve a greater generality and flexibility in defining from what the adaptation property must depend on. The intuitive idea preliminary expressed in the global space $\bar{M}_{\mathcal{P},J,t}$ is that a population $\mathcal{P}$ can adapt by changing in time its proper state variables, or its activations, exchanged goods (and hence its cause-effect relations), occurrence times or even neighbourhoods where interactions $j \in J$ take information they need to run. We simply use the same symbols but omitting the subscripts $\mathcal{P}$ and J (e.g. $\bar{M}_t$, $\bar{S}$, $\bar{M}^{\mathrm{s}}_t$, etc.) if $\mathcal{P} = E$ and $J = I$.

The idea to consider only the time interval $[t_{\mathrm{st}}, t]$ up to t, allows us to mathematically define that the system is better adapted at time t than at time s, and hence to distinguish between an improving dynamic and an emergent pattern, which is a best possible global state.

On the basis of what we said above, we can assume to have a probability space $(\Omega^{\mathrm{g}}, \mathcal{F}^{\mathrm{g}}, P^{\mathrm{g}})$ (the superscript "g" stands for *global*) and three random processes

$$\begin{aligned} &X : \Omega^{\mathrm{g}} \times [t_{\mathrm{st}}, t_{\mathrm{end}}] \longrightarrow \bar{S} \\ &T : \Omega^{\mathrm{g}} \times I \times [t_{\mathrm{st}}, t_{\mathrm{end}}] \longrightarrow [t_{\mathrm{st}}, t_{\mathrm{end}}]^3 \\ &N : \Omega^{\mathrm{g}} \times I \times [t_{\mathrm{st}}, t_{\mathrm{end}}] \longrightarrow \{\mathcal{N} \mid \mathcal{N} \subseteq E\} \end{aligned} \tag{2.4}$$

representing resp. the possible state maps of each interacting entity $e \in E$ (including activations, goods and proper state variables), the possible occurrence times, and the neighbourhood of each interaction $i \in I$. In other words, for any fixed elementary event $w \in \Omega^{\mathrm{g}}$, if we replace everywhere the state map $x_e(t) \in \bar{S}_e$, the occurrence times $t^{\mathrm{s}}_i(t)$, $t^{\mathrm{a}}_i(t)$, $t^{\mathrm{o}}_i(t) \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$, and the neighborhoods $\mathcal{N}_i(t) \subseteq E$ resp. with the time functions $X(w)(t)_e := X(\omega, t)_e \in \bar{S}_e$, $T^{\mathrm{s}}_i(\omega)(t) := T(\omega, i, t)_1$, $T^{\mathrm{a}}_i(\omega)(t) := T(\omega, i, t)_2$, $T^{\mathrm{o}}_i(\omega)(t) := T(\omega, i, t)_3$, and $N(\omega, i, t) \subseteq E$, then all the conditions in the definition of IS are satisfied. For example, $X(w)(t)_e$ satisfies the evolution equation whenever e is a patient fulfilling the assumptions of [18, (EE)].

The process $X = X(\omega, t)_e$ is formally a function of three variables $\omega \in \Omega^{\mathrm{g}}$, $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$ and $e \in E$, but we use flexible notations for the corresponding partial functions when one or more of these variables are fixed. For example, $X_e(t) : \omega \in \Omega^{\mathrm{g}} \mapsto X(\omega)(t)_e \in \bar{S}_e$ is the random state variable of the entity $e \in \mathcal{P}$ at time $t \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$. Similar notations are also used for the processes T and N.

Choosing a space $\bar{M}^{\mathrm{s}}_{\mathcal{P},t}$ which contains only constant state maps, we are equivalently considering the adaptation as occurring in a subspace (e.g. a manifold) of $\bar{S}_{\mathcal{P}} = \prod_{e \in \mathcal{P}} \bar{S}_e$; this is what we will examine both in Sec. 6 and in Sec. 7 below. On the contrary, an example of IS where the time state function $X(\omega)(-) \in \bar{M}^{\mathrm{s}}_{\mathcal{P},t_{\mathrm{end}}}$ is more important than the value itself $X(\omega)(t)_e \in \bar{S}_e$ is in the intelligent interpretation of a given text, let us say a clinical note, where we want to count how much frequently a given disease e is cited in the text: $X(\omega)(t_1)_e, \ldots, X(\omega)(t_N)_e$. In this example, the variable t represents the passing time while a given "user ω" is reading the text. If we consider only $t_1, \ldots, t_N \le t$, then $X(\omega)(-)_e|_{[t_{\mathrm{st}},t]}$ can be used to represent the amount of information the user ω is interpreting in the text up to the reading time t.

In the following, we also always assume that the global state space $\bar{M}_{\mathcal{P},J,t}$ is chosen so that

$$\begin{aligned} &\left(X_e(\omega)|_{[t_{\mathrm{st}},t]}\right)_{e\in\mathcal{P}} \in \bar{M}^{\mathrm{s}}_{\mathcal{P},t} \\ &\left(T_i(\omega)|_{[t_{\mathrm{st}},t]}\right)_{i\in J} \in M^{\mathrm{t}}_{J,t} \qquad \forall \omega \in \Omega^{\mathrm{g}}\, \forall t \in [t_{\mathrm{st}}, t_{\mathrm{end}}] \\ &\left(N_i(\omega)|_{[t_{\mathrm{st}},t]}\right)_{i\in J} \in M^{\mathrm{n}}_{J,t} \end{aligned} \tag{2.5}$$

for all populations $\mathcal{P} \subseteq E$ and all families of interactions $J \subseteq I$.

Finally, using $X(\omega)(t)_e \in \bar{S}_e$, we can also define the random processes of *activation state* and *goods* for each interaction $i \in I$:

$$\begin{aligned} \mathrm{AC}^e_i(\omega)(t) &:= X(\omega)(t)_{e,1,i} \\ \Gamma_i(\omega)(t) &:= X(\omega)(t)_{\mathrm{pr}(i),2,i}, \end{aligned}$$

so that $\mathrm{AC}^e_i(t) : \Omega^{\mathrm{g}} \longrightarrow [0,1]$ and $\Gamma_i(t) : \Omega^{\mathrm{g}} \longrightarrow R_i$.

3. Intuitive description of the generalized evolution principle

We start with an intuitive description of the GEP and with some thoughts to understand what we have to define in a precise mathematical language.

In an IS, we can have several populations of interacting entities. Since only some of these populations have to be described as adapting, we have to talk of the *adaptation of a given* subset (=: *population*) *of interacting entities*. This yields to a more general notion than simply talking of an entire "complex adaptive IS".

Therefore, in an adaptation process, we need to identify an adapting population $\mathcal{P}$ of interacting entities, $\mathcal{P} \subseteq E$, and a family of interactions $I_{\mathcal{P}} \subseteq I$ performed by the population $\mathcal{P}$ and representing how $\mathcal{P}$ is going to *decreases costs and stabilize the adaptation process through a suitable diversification of its goods*. Interactions in $I_{\mathcal{P}}$ are called *adaptive interactions*.
In the following definition, we specify what kind of interactions we have to consider:

Definition 1. Let $\mathcal{P} \subseteq E$, we say that $I_{\mathcal{P}} \subseteq I$ *is a family of interactions of (the population)* $\mathcal{P}$ if at least one agent of each interaction $i \in I_{\mathcal{P}}$ is a member of the population $\mathcal{P}$:

$$\forall i \in I_{\mathcal{P}}\, \exists a \in \mathrm{ag}(i) \cap \mathcal{P}.$$

We now continue the intuitive motivations and description of the GEP.

Note explicitly that not necessarily interactions of $I_{\mathcal{P}}$ act on entities of the population $\mathcal{P}$, i.e. not always $\mathrm{pa}(i) \in \mathcal{P}$ if $i \in I_{\mathcal{P}}$. This opens the possibility that the population adapts by acting on external entities and, only after a suitable chain of cause-effect related interactions, this causes a change in the state of $\mathcal{P}$.

As we will see more precisely below, the adaptation process changes functions state $X_{\mathcal{P}} := \left(X_e|_{[t_{\mathrm{st}},s]}\right)_{e\in\mathcal{P}}$ of the population $\mathcal{P}$, or occurrence times $T_{\mathcal{P}}(s) :=$

$\left(T_i|_{[t_{\rm st},s]}\right)_{i\in I_\mathcal{P}}$ or neighborhoods $N_\mathcal{P}(s) := \left(N_i|_{[t_{\rm st},s]}\right)_{i\in I_\mathcal{P}}$ of its adaptive interactions at time s to better values at time t, where unification and diversification are improved, i.e. costs are lowered and goods are resiliently distributed. We therefore precisely define when it happens that *at time t the population $\mathcal{P}$ is better adapted than at time s*. A particular case will be when at time t the population is at one of the *best* possible states, which usually corresponds to a steady or an equilibrium state and hence to an emergent pattern, even if, in general, not necessarily an emergent pattern can be really attained as actual state.
In the following, we use the notation for the global state

$$Y_\mathcal{P}(\omega)|_{[t_{\rm st},t]} := \left(\left(X_e(\omega)|_{[t_{\rm st},t]}\right)_{e\in\mathcal{P}}, \left(T_i(\omega)|_{[t_{\rm st},t]}\right)_{i\in I_\mathcal{P}}, \left(N_i(\omega)|_{[t_{\rm st},t]}\right)_{i\in I_\mathcal{P}}\right), \quad (3.1)$$

for $\omega\in\Omega^{\rm g}$ and $t\in[t_{\rm st},t_{\rm end}]$. Note that $Y_\mathcal{P}(\omega)|_{[t_{\rm st},t]}\in\bar{M}_{\mathcal{P},I_\mathcal{P},t}\subseteq\bar{M}_{\mathcal{P},I_\mathcal{P}}$ because of (2.5).

Together with the family of interactions $I_\mathcal{P}$, for each global state $y\in\bar{M}_{\mathcal{P},I_\mathcal{P}}$, we have to identify a function

$$C_y : \mathcal{P} \longrightarrow \mathbb{R}^k_{\ge 0} \qquad \forall y \in: \bar{M}_{\mathcal{P},I_\mathcal{P}} \quad (3.2)$$

defined on the population $\mathcal{P}$ and called *unification costs causing $I_\mathcal{P}$ at the state y*: an adaptive population $\mathcal{P}$ reacts with the interactions $I_\mathcal{P}$ to an increase in these costs $C_y = (C_y^1,\dots,C_y^k)\in\mathbb{R}^k_{\ge0}$ by changing the global state y and trying to decrease each cost C_y^j, $j=1,\dots,k$. The interactions $i\in I_\mathcal{P}$ that allow a decreasing of each component of the cost function represent the unification processes. Therefore, we are going to define when $\mathcal{P}$ is adaptive *with respect to a given cost function C and a family of interactions $I_\mathcal{P}$*. Of course, a population $\mathcal{P}$ can be adaptive with respect to several cost functions and families of interactions. Note that since the cost function C_y depends on the global state space $y\in\bar{M}_{\mathcal{P},I_\mathcal{P},t}$ (the union in (2.3) is actually a disjoint one), we can also consider costs depending on the activations or proper state variables of $e\in\mathcal{P}$, or on goods, occurrence times or neighbourhoods of interactions $i\in I_\mathcal{P}$, or even on time t.

In general, the adaptation of the population $\mathcal{P}$ does not occur by changing the state of a single entity $e\in\mathcal{P}$ but of several of them. For this reason, in most cases C_y is actually evaluated by averaging costs of the same type undergone by each interacting entity. The probability we use to average these costs is defined on the population $\mathcal{P}$, so that we can talk of costs paid *by the population $\mathcal{P}$* with respect to this probability. In general, this probability also depends on the global state $y\in\bar{M}_{\mathcal{P},I_\mathcal{P}}$: think e.g. at the following examples:

1) $y=(y^1,\dots,y^n)\in\mathbb{R}^n$ are the frequencies $y^k\in\mathbb{R}$ of each of n words in an evolving language. If the language is a CAS, these frequencies are distributed following a power law, and the average values of the costs in using this language depends on these frequencies, giving higher weights to more used words, see Sec. 7.
2) $y=(x,r)\in\mathbb{R}^{2n}\times\mathbb{R}^n_{\ge}$ are locations $x^j\in\mathbb{R}^2$ and rents $r^j\in\mathbb{R}_{\ge0}$ of $j=1,\dots,n$ companies producing suitable commodities, and the costs of these companies are averaged depending on both locations and rents, e.g. giving higher weights to more central and expansive locations, see Sec. 7.
3) y is the degree of attention (i.e. the activation in the language of IS) that each lecturer ω is giving to each part of speech in a given clinical note during its reading. The most adapted readings we are interested to are those where we

have a probability which are proportional to the degree of attention given to medical terms, and hence to a lower value of the related average costs.

Therefore, before talking of an adaptation process, for each global state $y \in \bar{M}_{\mathcal{P},I_{\mathcal{P}}}$ and for each cost C^j_y, we also need to identify a probability P^j_y on $\mathcal{P}$:

$$P^j_y \text{ is a probability on } \mathcal{P} \qquad \forall y \in \bar{M}_{\mathcal{P},I_{\mathcal{P}}}\, \forall j = 1, \dots, k.$$

It is with respect to P^j_y that we have to evaluate the expected value of the cost function C^j_y, and it is this expected value, and not the single costs experienced by each $e \in \mathcal{P}$, that have to be decreased in unification interactions. The probability P^j_y is called *probability to average the unification cost* C^j_y.

We therefore define the *unification "forces" at the state* y by

$$U^j_{\mathcal{P}}(y) := -E^j_y \left[C^j_y\right] \qquad \forall y \in \bar{M}_{\mathcal{P},I_{\mathcal{P}}}\, \forall j = 1, \dots, k, \tag{3.3}$$

where the expected value $E^j_y[-]$ is computed using P^j_y. The minus sign in (3.3) allows one to say that these forces are greater when the average costs are lower. For example, if $\mathcal{P} = \{e_1, \dots, e_N\}$ is finite, then

$$U^j_{\mathcal{P}}(y) = -\sum_{e \in \mathcal{P}} P^j_y(e) \cdot C^j_y(e). \tag{3.4}$$

Similarly, we can define the *unification forces for the event* $\omega \in \Omega^g$ *at time* $t \in [t_{\text{st}}, t_{\text{end}}]$ by

$$U^j_{\mathcal{P}}(\omega, t) := U^j_{\mathcal{P}}(Y_{\mathcal{P}}(\omega)|_{[t_{\text{st}},t]}). \tag{3.5}$$

Therefore, if the unification interactions allows the system to pass from the global state $Y_{\mathcal{P}}(\omega)|_{[t_{\text{st}},s]}$ to the better state $Y_{\mathcal{P}}(\omega)|_{[t_{\text{st}},t]}$, this can be expressed with

$$U^j_{\mathcal{P}}(\omega, t) \geq U^j_{\mathcal{P}}(\omega, s) \quad \forall j = 1, \dots, k. \tag{3.6}$$

In addition, the condition that $\bar{y} \in \bar{M}_{\mathcal{P},I_{\mathcal{P}}}$ is the best global state from the point of view of costs, can be expressed with

$$U^j_{\mathcal{P}}(\bar{y}) \geq U^j_{\mathcal{P}}(\omega, t) \quad \forall \omega \in \Omega^{\text{g}}\, \forall t \in [t_{\text{st}}, t_{\text{end}}]\, \forall j = 1, \dots, k. \tag{3.7}$$

The adaptive interactions $i \in I_{\mathcal{P}}$ also realize the diversification to a decreasing of C_y, and hence to stabilize this declining of costs. We measure the *diversification "forces"* $D_{I_{\mathcal{P}}}(\gamma)$ with the *information entropy of the fluxes of goods* $\gamma_i \in R_i$ *extracted by each population interaction* $i \in I_{\mathcal{P}}$ *from its resource space* R_i, *and exchanged from agents to patients through propagators*, see [18] and [12, 6] for a similar point of view. An intuitive way to motivate this idea is to say that the more the population is sharing its goods/resources/information the more it is adapting. See also Sec. 4 below for the relationships between GEP and different approaches in measuring information flows.

From what space these goods $\gamma = (\gamma_i)_{i \in I_{\mathcal{P}}}$ are taken from? If $x \in \bar{M}^{\text{s}}_{\mathcal{P},t}$ is a state map of our IS, then $\gamma_{x,i} := x(-)_{\text{pr}(i),i,2} : [t_{\text{st}}, t] \longrightarrow R_i$ are the corresponding goods of $i \in I$ as a function of time (up to t). It suffices to consider the family $(\gamma_{x,i})_{i \in I_{\mathcal{P}}}$; e.g. if $I_{\mathcal{P}} = \{i_1, \dots, i_d\}$ is finite, then $(\gamma_{x,i})_{i \in I_{\mathcal{P}}} = (\gamma_{x,i_1}, \dots, \gamma_{x,i_d})$ considers all the goods extracted by each interaction in the given order. The spaces we need to consider are therefore

$$R_{\mathcal{P},t}:=\left\{(\gamma_{x,i})_{i\in I_{\mathcal{P}}}\mid x\in\bar{M}^{\mathrm{s}}_{\mathcal{P},t}\right\}$$

$$R_{\mathcal{P}}:=\bigcup_{t\in[t_{\mathrm{st}},t_{\mathrm{end}}]}R_{\mathcal{P},t}. \tag{3.8}$$

We therefore need to finally identify a probability Q_γ defined on $I_{\mathcal{P}}$ to evaluate the corresponding forces of diversification $D_{I_{\mathcal{P}}}(\gamma)$ and $D_{I_{\mathcal{P}}}(\omega,t)$ as

$$Q_\gamma \text{ is a probability on } I_{\mathcal{P}} \qquad \forall\gamma\in R_{\mathcal{P}} \tag{3.9}$$

$$D_{I_{\mathcal{P}}}(\gamma):=\text{Entropy}\left(Q_\gamma\right) \tag{3.10}$$

$$D_{I_{\mathcal{P}}}(\omega,t):=D_{I_{\mathcal{P}}}\left(\left(\Gamma_i(\omega)|_{[t_{\mathrm{st}},t]}\right)_{i\in I_{\mathcal{P}}}\right) \qquad \forall\omega\in\Omega^{\mathrm{g}}\,\forall t\in[t_{\mathrm{st}},t_{\mathrm{end}}]. \tag{3.11}$$

The probability Q_γ is called *diversification probability*. For example, if $I_{\mathcal{P}}=\{i_1,\ldots,i_d\}$ is finite, then

$$D_{I_{\mathcal{P}}}(\gamma)=-\sum_{i\in I_{\mathcal{P}}}Q_\gamma(i)\cdot\log_2 Q_\gamma(i). \tag{3.12}$$

Note that, in general, the diversification forces depend on the whole history $\mathrm{n}_{\mathrm{pr}(i)}\,x_t$ of the state of the neighborhood of the propagator of $i\in I_{\mathcal{P}}$, because, e.g. for the case of $D_{I_{\mathcal{P}}}(\omega,t)$, the state of goods $\Gamma_i(t)(\omega)$ satisfy the evolution equation [18, (EE)], see also [18, Rem. 1.(e)]. Therefore, this neighborhood can include the influence of entities which are external to $\mathcal{P}$. Similarly, the propagator of i not necessarily belongs to $\mathcal{P}$, so that the diversification of resources can also involve external entities.

The second important condition of the GEP states that *in the sample $\omega\in\Omega^g$ diversification forces are greater at time t than at time s* if

$$D_{I_{\mathcal{P}}}(\omega,t)\geq D_{I_{\mathcal{P}}}(\omega,s). \tag{3.13}$$

As above, the condition that the global state of goods $\bar{\gamma}\in R_{\mathcal{P}}$ is the best from the point of view of diversification forces can be stated asking that

$$D_{I_{\mathcal{P}}}(\bar{\gamma})\geq D_{I_{\mathcal{P}}}(\omega,t) \qquad \forall\omega\in\Omega^{\mathrm{g}}\,\forall t\in[t_{\mathrm{st}},t_{\mathrm{end}}]. \tag{3.14}$$

For example, assume that we have two families of simultaneous *independent* interactions $J_{\mathcal{P}}$, $K_{\mathcal{P}}$ for the probability $Q(\gamma)$ (e.g. because there is no cause-effect relation between $j\in J_{\mathcal{P}}$ and $k\in K_{\mathcal{P}}$). Then the diversifications $D_{J_{\mathcal{P}}}(\omega,t):=D_{J_{\mathcal{P}}}\left(\left(\Gamma_i(\omega)|_{[t_{\mathrm{st}},t]}\right)_{i\in J_{\mathcal{P}}}\right)$ and $D_{K_{\mathcal{P}}}(\omega,t):=D_{K_{\mathcal{P}}}\left(\left(\Gamma_i(\omega)|_{[t_{\mathrm{st}},t]}\right)_{i\in K_{\mathcal{P}}}\right)$, thanks to the logarithm in (3.9), will contribute additively $D_{I_{\mathcal{P}}}(\omega,t)=D_{J_{\mathcal{P}}}(\omega,t)+D_{K_{\mathcal{P}}}(\omega,t)$ at increasing the diversification of the population $\mathcal{P}$.

In the GEP, we ask both (3.6) and (3.13) or both (3.7) and (3.14) (but where $\bar{\gamma}$ are the goods already included in the global state $\bar{y}$), see below Def. 2.

Note that not necessarily the maximization (5.4) of diversification forces $D_{I_{\mathcal{P}}}(\gamma)=\text{Entropy}\left(Q_\gamma\right)$ implies a uniform distribution for Q_γ: indeed, this holds only if the goods $\gamma\in R_{\mathcal{P}}$ allow for all possible probabilities on $I_{\mathcal{P}}$ (see also Thm. 3 below, and its assumption (6.2)).

More generally, one may be interested in locally maximizing some functional $F(U^1_{\mathcal{P}}(x),\ldots,U^k_{\mathcal{P}}(x),D_{I_{\mathcal{P}}},t)$ only, instead of the strong condition to maximize independently all the adaptation forces $U^1_{\mathcal{P}}(x),\ldots,U^k_{\mathcal{P}}(x),D_{I_{\mathcal{P}}}$. For example, this happens in Zipf-Mandelbrot law, where $F(U, D) = \frac{U}{D}$, see Thm. 3 below. In this case, we say that the GEP is satisfied *along the functional* F.

4. Generalized evolution principle, Shannon entropy, out of equilibrium systems and second law of thermodynamics

We recall that, in information theory, if two different messages are extracted from the same probability distribution, then they are undistinguishable from the point of view of information entropy. Actually, it is not correct to talk of Shannon entropy of single messages, because this notion can be applied only to probability distributions (and hence, e.g., to random variable). This is frequently summarized saying that entropy does not depend on the meaning of messages but only on their probability distribution. On the contrary, if we have two messages, and the former at time t (let us say *"to be or not to be"*) exchanges greater fluxes of goods with the population $\mathcal{P}$ than the latter (e.g. "*nttb e obt ooo e r*") at time s, then inequality (3.13) could be interpreted saying that the message exchanged at t is more meaningful for $\mathcal{P}$ than the message exchanged at time s. Therefore, the GEP has a different meaning than the simple Shannon entropy, exactly because it involves interactions and their propagators, and hence the cause-effect structure of the considered IS. This is ultimately due to the fact that we are not only talking of Shannon entropy of arbitrary random variables or time series, but of goods exchanged from agents to patients in a polyadic cause-effect relationships, see [18, Sec. 3.4.1]. In fact, in every IS, the cause-effect link is expressed by activation and occurrence times, see [18, (CE)] and [18, (SA)]. Moreover, propagators are themselves interacting entities, and hence goods can represent fluxes of any form between agents and patients, see the long list of examples in [18], examples in Sec. 1 above, and Sec. 6 below, where goods are probabilities, or Sec. 7.3.2 where goods are exchanged commodities. In other words, in order that the GEP holds in a *validated* IS we have to satisfy severe constraints, because we have to respect the idea of fluxes of goods exchanged in a polyadic relationships between agents and patients.

For example, the GEP can be used to distinguish the importance of different messages/states (thought of as interacting entities, and not as random variables) for a given population $\mathcal{P}$. After all, the intuitive difference between a meaningful/readable message and a completely random/unreadable one, is exactly that the former is able to send us signals (propagators) that we are able to interpret (goods), whereas random variables can be identified with their probability distribution, which are more unintelligible for our brain.

In our opinion, this gives elements to solve the critiques expressed by [22] in using transfer entropies in a mechanistic interpretation as information flow.

Similarly, it is also important to understand the differences between GEP and the decreasing of entropy for out of equilibrium systems. Indeed, it is well-known, see e.g. [30], that this type of systems constantly make efforts to expel waste having a high level of entropy, and this allows them to keep a *lower* level of *global* entropy. Therefore, this seems to contradict the GEP, where the diversification forces, related to the entropy of exchanged goods, have to be increased. However, these wastes have a great level of entropy but, exactly because they are expelled, do not have stable interactions relating their agents with patients remaining in the system. This decreasing of entropy is therefore completely different from the increasing of diversification: in the latter, the system adapts if it strives *to keep* these exchanged goods and the information/diversification they represent.

Finally, the GEP is also very different from the second principle of thermodynamics, because *we are not considering the whole entropy* of an isolated system but, on the contrary, only those related to the exchanged goods of the interactions $i \in I_{\mathcal{P}}$ that aim at decreasing the average cost function C. This costs essentially implies that the system is necessarily not isolated. Similar considerations, essentially based on the remark that in the GEP we do not calculate the entropy of the whole system, can be repeated for nonadditive entropic functionals, see e.g. [40] and references therein.

All these differences are even more meaningful if the GEP is satisfied only along some functional F.

5. Mathematical definition of the generalized evolution principle

The previous motivations justify the following

Definition 2 (Generalized evolution principle, CAS)**.** Let $\mathfrak{I}$ be an IS, s, $t \in [t_{\text{st}}, t_{\text{end}}]$ and $\mathcal{P} \subseteq E$. Then, we say that *in the sample $\omega \in \Omega^g$ at time t the population $\mathcal{P}$ is better adapted than at time s (with respect to $I_{\mathcal{P}}$, $\left(P_y^j\right)_{j,y}$, $\left(C_y^j\right)_{j,y}$, $(Q_\gamma)_\gamma$, in the global space $\bar{M}_{\mathcal{P},I_{\mathcal{P}}}$; briefly: $\mathcal{P}$ is a CAS* or that *$\mathcal{P}$ satisfies the GEP) if*

(i) $I_{\mathcal{P}}$ is a family of interactions of $\mathcal{P}$ called *adaptive interactions*.

(ii) P_y^j is a probability on the population $\mathcal{P}$ for each $j = 1, \ldots, k$ and each $y \in \bar{M}_{\mathcal{P},I_{\mathcal{P}}}$. The probability P_y^j is called *probability to average the $j-$th cost at the state y.*

(iii) $C_y^j : \mathcal{P} \longrightarrow \mathbb{R}_{\geq 0}$ is a measurable function with respect to P_y^j for each $y \in \bar{M}_{\mathcal{P},I_{\mathcal{P}}}$ and $j = 1, \ldots, k$. $C_y = (C_y^1, \ldots, C_y^k)$ is called *unification cost function at y.*

(iv) Q_γ is a probability on $I_{\mathcal{P}}$ for each $\gamma \in R_{\mathcal{P}}$ (see (3.8)) called *diversification probability.*

(v) We have

$$U_{\mathcal{P}}^j(\omega, t) \geq U_{\mathcal{P}}^j(\omega, s) \qquad \forall j = 1, \ldots, k, \tag{5.1}$$

$$D_{I_{\mathcal{P}}}(\omega, t) \geq D_{I_{\mathcal{P}}}(\omega, s), \tag{5.2}$$

where *unification forces* $U_{\mathcal{P}}^j(\omega, t)$ are defined by (3.5), and *diversification forces* $D_{I_{\mathcal{P}}}$ are defined by (3.9).

(vi) If $F : \mathbb{R}^{k+2} \longrightarrow \mathbb{R}$, we say that *along the functional F, in the sample $\omega \in \Omega^g$, at time t the population $\mathcal{P}$ is better adapted than at time s* if

$$F(U_{\mathcal{P}}^1(\omega, t), \ldots, U_{\mathcal{P}}^k(\omega, t), D_{I_{\mathcal{P}}}(\omega, t), t) \geq F(U_{\mathcal{P}}^1(\omega, s), \ldots, U_{\mathcal{P}}^k(\omega, s), D_{I_{\mathcal{P}}}(\omega, s), s).$$

We also say that the state $\bar{y} = (x, \bar{t}, N) \in \bar{M}_{\mathcal{P},I_{\mathcal{P}}}$ (see (2.1)) is an *emergent pattern for $\mathcal{P}$* if

$$U_{\mathcal{P}}^j(\bar{y}) \geq U_{\mathcal{P}}^j(\omega, s) \qquad \forall j = 1, \ldots, k, \tag{5.3}$$

$$D_{I_{\mathcal{P}}}\left((\gamma_{x,i})_{i \in I_{\mathcal{P}}}\right) \geq D_{I_{\mathcal{P}}}(\omega, s) \tag{5.4}$$

for all $\omega \in \Omega^{\mathrm{g}}$ and all times $s \in [t_{\mathrm{st}}, t_{\mathrm{end}}]$. We recall that $x \in \bar{M}^{\mathrm{s}}_{\mathcal{P},t}$ and hence $(\gamma_{x,i})_{i \in I_{\mathcal{P}}}$ are the goods included in the global state $\bar{y} = (x, \bar{t}, N) \in \bar{M}_{\mathcal{P}, I_{\mathcal{P}}}$. Similarly to (vi), we can define that $\bar{y}$ is an *emergent pattern for $\mathcal{P}$ along a functional F*.

The latter conditions (5.3) and (5.4) can hold even if $\bar{y}$ is a global state which is not reached by the system, i.e. the equality $\bar{y} = Y_{\mathcal{P}}(\bar{\omega})|_{[t_{\mathrm{st}},t]}$ never holds for any $\bar{\omega}$ e any t. In fact, we may have a step from a state $Y_{\mathcal{P}}(\omega)|_{[t_{\mathrm{st}},s]}$ to a better state $Y_{\mathcal{P}}(\omega)|_{[t_{\mathrm{st}},t]}$ even if there does not exist a best possible state, i.e. an emergent pattern. An intuitive example that seems to satisfy this property may come from Darwinian evolution. Think at giraffes and their elongation of neck: the cost are related to the probability of finding leaves to eat; we have at least two interacting entities: giraffes and trees, but only one is adapting with respect to the cost of being hungry; the population interactions of the force of diversification allowed some giraffes to have a genetic code that causes a longer neck. It seems that there is not a maximum length of neck minimizing *this* cost, even if such a maximum is reached due to the increasing of *other* costs.

It is clear that a very general case, even if conceptually it cannot exhaust all the possibilities, is when the cost function C depends only on the state $X_{\mathcal{P}}(t)$ and not on $(T_{\mathcal{P}}(t), N_{\mathcal{P}}(t))$. This is an implicit assumption we will consider in the rest of the paper.

If the cost function $C_y \equiv C_y(e)$ does not depend on the entity $e \in \mathcal{P}$ nor on the global state y, and $\bar{y}$ is an emergent pattern, the GEP reduces to the classical entropy maximization principle for the family of probabilities $(Q_y)_y$. On the other hand, if we trivialize all the interactions $i \in I_{\mathcal{P}}$ by choosing constant resources (deterministic extraction of goods), then the population is adapting at an emergent pattern $\bar{y}$ if it minimizes the expected value of the cost function $C_{\bar{y}}$. Therefore, the GEP includes both the entropy maximization principle and classical minimization problems of the cost function $y \in \bar{M}_{\mathcal{P}, I_{\mathcal{P}}} \mapsto C^j_y$ if C^j_y is deterministic with respect to the probability P^j_y.

Because the expected costs $E^j_y[C^j_y]$, $j = 1, \dots, k$, in (3.3) are calculated with a probability distribution over the population $\mathcal{P}$, we can interpret the unification forces $U^j_{\mathcal{P}}(y) = -E^j_y(C^j_y)$, $j = 1, \dots, k$, as proportional to a quantification of suitable *common goods for the population $\mathcal{P}$*. As a consequence of the maximization properties of Shannon's entropy, the diversification forces $D_{I_{\mathcal{P}}}(\gamma)$ can be interpreted as a gauge of *long-term changes* (with respect to the given cost function C).

Note that an emergent patter $\bar{y} = (x, \bar{t}, N)$ results into probabilities $P^j_{\bar{y}}$ on the population $\mathcal{P}$ and $Q_{\bar{\gamma}}$ on adaptive interactions $I_{\mathcal{P}}$, where $\bar{\gamma} = (\gamma_{x,i})_{i \in I_{\mathcal{P}}}$ are the goods included in the global state $\bar{y}$. Starting from these probability distributions, and depending on the problem, we can then consider particular realizations $e \in \mathcal{P}$ and $i \in I_{\mathcal{P}}$, like the mode (i.e. where the probabilities assume highest value) in frequently used AI algorithms, see e.g. [5]; in this case, e represents the interacting entity where the weight $P^j_{\bar{y}}(e)$ to evaluate the average unification cost $C^j_{\bar{y}}$ is highest, whereas i represent the adaptive interaction which contributes more to the diversification $D_{I_{\mathcal{P}}}(\bar{\gamma})$. We can clearly have more that one of these e and i in case of multimodal probabilities.

Is it a classical complicated system, such as a spring-driven clock, a CAS with respect to Def. 2? We can think this clock as a continuous dynamical system (maybe with suitable generalized functions to represent the discrete ticking of the hands of the clock), so that it is surely an IS, see [18, Sec. 4.1]. This IS can be considered having simple cause-effect interactions starting from the spring and arriving at clock's hands. The natural cost function is hence the distance between the perfect speed of each hand, e.g. 1 Hz, and its actual eventually lower speed. The natural main resource is the potential energy stored in the spring. Therefore, the cost are eventually increasing and the diversification forces always constant (in a model with deterministic use of resources) or decreasing (stochastic use of resources with a decreasing variance), and indeed also intuitively the system is not adapting. On the other hand, note that the system given by the spring clock + a winding person is a CAS.

Note that this mathematical formalization of the GEP has been possible only thanks to the language we introduced for IS in [18]: interactions as cause-effect relation between agents and patients and carried by propagators, state map and state space, space of resources and goods; all these concepts are used in the previous definition.

We close this section by summarizing the steps we need to realize in order to model a CAS satisfying the GEP.

5.1. **General steps to model a complex adaptive system.** For the sake of clarity, we list here the general procedure we have to follow in order to model a system which obeys the GEP. The process presented here is clearly not linear and presents several cause-effect polyadic feedback interactions, i.e. it can be thought of as a kind of meta-IS.

1. We clearly have to start by defining an IS, or another type of model embedded as an IS. Therefore, in general we need (see [18, Tables 1, 2, 3]): interacting entities, interactions, activations, proper state, goods and resources, occurrence times, neighbourhoods and transition functions, even if in several particular models some of these notions are trivial.
2. Understand what are good global states $\bar{M}^{\mathrm{s}}_{\mathcal{P},t}$, $M^{\mathrm{t}}_{J,t}$, $M^{\mathrm{n}}_{J,t}$, see Sec. 6 and Sec. 7 for examples and the next point in this list.
3. We can try to define the global probability space Ω^{g} as explained in Sec. 2 above. However, an equivalent approach is to understand only the probability distributions of the processes X, T, N of (2.4) and set $\Omega^{\mathrm{g}} := \bar{M}_{\mathcal{P},I_{\mathcal{P}}}$ with probability P^{g} given by the joint probabilities of X, T, N. These processes X, T, N are then represented by the projections of each $\bar{M}_{\mathcal{P},I_{\mathcal{P}},t}$ resp. on $\bar{M}^{\mathrm{s}}_{\mathcal{P},t}$, $\bar{M}^{\mathrm{t}}_{I_{\mathcal{P}},t}$ and $\bar{M}^{\mathrm{n}}_{I_{\mathcal{P}},t}$. Note that knowing the distributions of X, T, N means to mathematically solves the considered IS, see [18, Sec. 2.8], by solving or simulating the evolution equations and understanding (if needed) the dynamics of occurrence times and neighbourhoods.
4. Understand what are the adaptive interactions $I_{\mathcal{P}}$, i.e. the interactions of the system that dynamically improve the state of the system by following the GEP.
5. Understand what are the unification costs $\left(C^j_y\right)_{j,y}$ causing unification forces. These are what the system tries to avoid as far as possible using adapting interactions $I_{\mathcal{P}}$.
6. Define probabilities on the population $\mathcal{P}$ to average unification costs $\left(P^j_y\right)_{j,y}$, e.g. giving a greater weight $P^j_y(e)$ at global states y and entities $e \in \mathcal{P}$ which are intuitively paying greater costs.

7. Define diversification probabilities $(Q_\gamma)_\gamma$ on the adaptive interactions $I_\mathcal{P}$; these are related to the goods exchanged during the dynamics in order to distribute resources and decrease the expected value of costs. Therefore, $Q_\gamma(i)$ will be higher for important adapting interactions $i \in I_\mathcal{P}$.
8. Assume to have an emergent pattern $\bar{y} \in \bar{M}_{\mathcal{P},I_\mathcal{P}}$, at least along a given functional F, and derive necessary conditions, or consider simulations where unification $U^j_\mathcal{P}(\omega,t)$ and diversification $D_{I_\mathcal{P}}(\omega,t)$ increases, see Sec. 6 and Sec. 7 for examples.
9. Validate the model by comparing, in the strongest possible way, emergent patterns predicted from your model with configurations of systems that are clearly independent from the model itself, e.g. real-world systems.

We now prove, in a more abstract setting and under mild conditions, that in every CAS the diversification probabilities Q_γ satisfy a power law when the population is at an emergent pattern along the functional $F(U,D) = \frac{U}{D}$ and if the partial derivative $\partial U_\mathcal{P}(x)/\partial x_k$ of the unification forces is of logarithmic type in k.

6. Power law distribution following Mandelbrot's idea

It is well known that power law distributions are frequently associated to CAS, and hence appear often both in nature and in social systems, see e.g. [38, 2, 16] and references therein.

Using the language of IS theory, this section follows and generalizes the classical ideas of B. Mandelbrot presented in [25, 26]; see also [27] and references therein for a deep analysis of this kind of model in linguistics.

We imagine to have an IS and a population $\mathcal{P}$ whose state is described by vectors $x \in \bar{S}_\mathcal{P} \subseteq \mathbb{R}^d$. The systems has to be thought of as a CAS that changes its state so as to decrease a suitable cost function $E_\mathcal{P} : \bar{S}_\mathcal{P} \longrightarrow \mathbb{R}_{>0}$ and, at the same time, to increase a corresponding information entropy:

$$D_{I_\mathcal{P}}(x) = -\sum_{k=1}^{d} x_k \cdot \log_2 x_k > 0 \qquad \forall x \in \bar{S}_\mathcal{P}. \tag{6.1}$$

For example, similarly to [25, 26, 27], we can think at d classes of words of an *evolving language*: each class $k = 1,\dots,d$ (sometimes called *rank*) representing all the words whose evolutionary state is described by comparable frequencies $x_k > 0$ of use, $\sum_{k=1}^d x_k = 1$. Using the language of IS theory, we can think to have interacting entities $E = \{1,\dots,d\} = \mathcal{P}$ and d interactions of the type $i_k : k \xrightarrow{(x'_k,x_k),\text{evo}} k$, for each $k \in E$ representing the evolution of the k-th interacting entities from a previous state x'_k to the new one x_k. For simplicity, we can directly say that $x_k = \gamma_{i_k}(t)$ are the goods of i_k, and all the other elements of the IS, e.g. occurrence times and transition functions f_k, are trivial. The adapting interactions are hence $I_\mathcal{P} = \{i_1,\dots,i_d\}$, and the diversification probabilities are $Q_x(i_k) = x_k$ for all $k = 1,\dots,d$ (compare (6.1) and (3.12)). We are only interested to the global state space $\bar{M}_{\mathcal{P},I_\mathcal{P}} = \bar{S}_\mathcal{P} \subseteq \mathbb{R}^d$ because all the other ones are trivial. In our opinion this type of models are highly idealized, and e.g. this stress on evolving IS exclude cases such as a given novel of a single author, or specialized subsets of a given language.

In general, i.e. not just thinking at the example of an evolving language, we assume that the global state $\bar{S}_\mathcal{P}$ is given by all the possible probabilities $x =$

$(x_1, \dots, x_d)$:

$$\bar{S}_{\mathcal{P}} = \left\{ x \in \mathbb{R}^d_{>0} \mid \sum_{k=1}^{d} x_k = 1 \right\}. \tag{6.2}$$

We consider several cases:

(i) The system adapts (e.g. it evolves) following the GEP with unification forces $U_{\mathcal{P}}(x) = -E_{\mathcal{P}}(x)$, diversification forces $D_{I_{\mathcal{P}}}$. The state $y \in \bar{S}_{\mathcal{P}}$ is an emergent pattern (not necessarily reached by the real dynamics of the system).

(ii) The system follows the GEP only along the functional $F(U, D) = \frac{U}{D}$, with $U_{\mathcal{P}}(x) = -E_{\mathcal{P}}(x)$. In other words, it evolves so as to minimize the ratio $\frac{E_{\mathcal{P}}}{D_{I_{\mathcal{P}}}}$ at the point $y \in \bar{S}_{\mathcal{P}}$, at least locally around y:

$$\forall x \in \bar{S}_{\mathcal{P}} \cap B_r(y) : \ 0 < \frac{E_{\mathcal{P}}(y)}{D_{I_{\mathcal{P}}}(y)} \le \frac{E_{\mathcal{P}}(x)}{D_{I_{\mathcal{P}}}(x)},$$

where $B_r(y) = \{x \in \mathbb{R}^d \mid |x - y| < r\}$ is the ball of radius $r \in \mathbb{R}_{>0}$.

(iii) In the previous case, we further assume that the partial derivative of the cost satisfies

$$\frac{\partial E_{\mathcal{P}}}{\partial x_k}(y) = \alpha(y) \cdot \log_2(k + k_0(y)) \tag{6.3}$$

for some functions α, $k_0 : \bar{S}_{\mathcal{P}} \longrightarrow \mathbb{R}$. This is the case considered in linguistic by Mandelbrot, [25, 26]. We do not need to assume that lower rank corresponds to higher frequency, i.e. $q_k(x) > q_{k+1}(x)$, but we prove it. Moreover, in this classical evolutionary interpretation, it is reasonable to assume that $k_0 = k_0(y)$ in (6.3) depends on the state y, so that the two parameters of Zipf-Mandelbrot law $k_0(y)$ and $-\alpha(y)\frac{D_{I_{\mathcal{P}}}(y)}{E_{\mathcal{P}}(y)}$ are functionally related through the adapted state y, as also confirmed by [27].

We have the following

Theorem 3. *Let $\bar{S}_{\mathcal{P}} \subseteq \mathbb{R}^d$ be the global state of an IS, and let $y \in \bar{S}_{\mathcal{P}}$. Assume that $\bar{S}_{\mathcal{P}}$ is given by (6.2). Consider the diversification force $D_{I_{\mathcal{P}}}$ given by (6.1). Let $U_{\mathcal{P}} = -E_{\mathcal{P}} \in \mathcal{C}^1(\bar{S}_{\mathcal{P}}, \mathbb{R}_{>0})$ be the unification force. Then, we have*

(i) If the system follows the GEP and y is an emergent pattern, then $y_k = \frac{1}{d}$ is always a uniform distribution.

(ii) If the system follows the GEP along the functional $F(U, D) = \frac{U}{D}$ and y is an emergent pattern, then

$$y_k = \frac{1}{N(y)} 2^{-\frac{D_{I_{\mathcal{P}}}(y)}{E_{\mathcal{P}}(y)} \cdot \frac{\partial E_{\mathcal{P}}}{\partial x_k}(y)} \quad \forall k = 1, \dots, d, \tag{6.4}$$

where $N(y) := \sum_{j=1}^{d} 2^{-\frac{D_{I_{\mathcal{P}}}(y)}{E_{\mathcal{P}}(y)} \cdot \frac{\partial E_{\mathcal{P}}}{\partial x_j}(y)}$ is the normalization factor. In particular, this applies if the system follows the GEP, i.e. the cost $E_{\mathcal{P}}(y)$ attains its minimum and the diversification $D_{I_{\mathcal{P}}}(y)$ its maximum.

(iii) If the system follows the GEP along the functional $F(U, D) = \frac{U}{D}$, and (6.3) holds for some function $k_0 \in \mathcal{C}^1(\bar{S}_{\mathcal{P}}, \mathbb{R})$, then the Zipf-Mandelbrot law holds:

$$y_k = \frac{1}{N(y)} \cdot (k + k_0(y))^{-\alpha(y) \cdot \frac{D_{I_{\mathcal{P}}}(y)}{E_{\mathcal{P}}(y)}} \quad \forall k = 1, \dots, d,$$

and we have $N(y) = \sum_{j=1}^{d} (j + k_0(y))^{-\alpha(y) \cdot \frac{D_{I_{\mathcal{P}}}(y)}{E_{\mathcal{P}}(y)}}$.

Proof. We first note that the space $\bar{S}_{\mathcal{P}}$ is not an open set and hence, in each one of the cases (i)-(iii), to compute the constrained minimum state $y \in S_{\mathcal{P}}$, we have to use Lagrange multipliers. In case (i), the functions $E_{\mathcal{P}}$ and $D_{I_{\mathcal{P}}}$ attain resp. a minimum and a maximum at y, so that also the quotient $\frac{E_{\mathcal{P}}(y)}{D_{I_{\mathcal{P}}}(y)}$ is minimal. In all the cases, we therefore get the existence of $\lambda \in \mathbb{R}$ such that for all $k = 1, \ldots, d$

$$\partial_k \left(\frac{E_{\mathcal{P}}}{D_{I_{\mathcal{P}}}} \right)(y) = \lambda \partial_k \left(\sum_{j=1}^{d} y_j \right) = \lambda, \tag{6.5}$$

where $\partial_k := \frac{\partial}{\partial x_k}$. Recall that, by the method of Lagrange multipliers, λ does not depend on k. For simplicity, all the functions that will appear in the following are evaluated at the state y. From (6.5), we have

$$\partial_k E_{\mathcal{P}} \cdot D_{I_{\mathcal{P}}} + E_{\mathcal{P}} \cdot \sum_j \left(\partial_k y_j \cdot \log_2 y_j + y_j \frac{1}{y_j} \log_2 e \cdot \partial_k y_j \right) = \lambda D_{I_{\mathcal{P}}}^2.$$

Thereby

$$D_{I_{\mathcal{P}}} \cdot \partial_k E_{\mathcal{P}} + E_{\mathcal{P}} \left(\log_2 y_k + \log_2 e \right) = \lambda D_{I_{\mathcal{P}}}^2,$$

and hence

$$\begin{aligned} \log_2 y_k &= -\frac{D_{I_{\mathcal{P}}}}{E_{\mathcal{P}}} \partial_k E_{\mathcal{P}} - \log_2 e + \lambda \frac{D_{I_{\mathcal{P}}}^2}{E_{\mathcal{P}}}, \\ y_k &= 2^{-\frac{D_{I_{\mathcal{P}}}}{E_{\mathcal{P}}} \partial_k E_{\mathcal{P}}} \cdot \frac{1}{e} \cdot 2^{\lambda \frac{D_{I_{\mathcal{P}}}^2}{E_{\mathcal{P}}}}. \end{aligned} \tag{6.6}$$

We compute $\lambda \frac{D_{I_{\mathcal{P}}}^2}{E_{\mathcal{P}}}$ from (6.5):

$$\lambda \frac{D_{I_{\mathcal{P}}}^2}{E_{\mathcal{P}}} = \partial_k \left(\frac{E_{\mathcal{P}}}{D_{I_{\mathcal{P}}}} \right) \frac{D_{I_{\mathcal{P}}}^2}{E_{\mathcal{P}}} = \frac{D_{I_{\mathcal{P}}}}{E_{\mathcal{P}}} \cdot \partial_k E_{\mathcal{P}} + \log_2 y_k + \log_2 e,$$

so that

$$e \cdot 2^{-\lambda \frac{D_{I_{\mathcal{P}}}^2}{E_{\mathcal{P}}}} y_k = 2^{-\frac{D_{I_{\mathcal{P}}}}{E_{\mathcal{P}}} \cdot \partial_k E_{\mathcal{P}}}.$$

Taking the summation for $k = 1, \ldots, d$ (and considering that λ does not depend on k), we obtain

$$e \cdot 2^{-\lambda \frac{D_{I_{\mathcal{P}}}^2}{E_{\mathcal{P}}}} = \sum_{k=1}^{d} 2^{-\frac{D_{I_{\mathcal{P}}}}{E_{\mathcal{P}}} \cdot \partial_k E_{\mathcal{P}}} = N.$$

From this and (6.6) the claim (ii) follows.

In particular, if the system follows the GEP, then $E_{\mathcal{P}}$ attains a constraint minimum at $y \in \bar{S}_{\mathcal{P}}$. Once again using Lagrange multipliers, we get the existence of $\mu \in \mathbb{R}$ such that for all $k = 1, \ldots, d$

$$\partial_k E_{\mathcal{P}}(y) = \mu \partial_k \left(\sum_{j=1}^{d} y_j \right) = \mu.$$

Therefore, the right hand side of y_k in (6.4) does not depend on k and we hence have a uniform distribution $y_k = \frac{1}{d}$.

Finally, substituting (6.3) in (6.4) we get the final claim (iii). $\square$

Examples of cost functions satisfying (6.3) are given by the average value $E_{\mathcal{P}}(x) = \sum_{k=1}^{d} c_k(x) \cdot p_k(x)$, where $(p_1(x), \dots, p_d(x))$ is a probability, and $p_k(x) = x_k$. In the notations of the GEP, we have $U_{\mathcal{P}}(x) = -E_{\mathcal{P}}(x)$, the unification cost is $C_x(k) = c_k(x)$, and the probabilities to average the cost are $P_x(k) = p_k(x)$ for each interacting entity $k \in \mathcal{P} = \{1, \dots, d\}$, see (3.4).

The k-th component c_k of the cost is a modeling choice and has to satisfy (6.3). The classical example for languages is $c_k(x) = a \cdot \log_b(k + k_0)$, where a, $j_0 \in \mathbb{R}_{>0}$. Usually, $b > 1$ is the number of letters in an alphabet. Since $c_k(x)$ does not depend on x_k, we have $\partial_k E_{\mathcal{P}}(y) = c_k(y) = \frac{a}{\log_2 b} \log_2(k + k_0)$, and hence it suffices to set $\alpha := \frac{a}{\log_2 b}$ in order to apply (iii) of Thm. 3.

Note the importance of assuming that the function $\alpha(y)$ in (6.3) does not depend on k: otherwise, it would suffice to set $\alpha_k(x) := \partial_k E_{\mathcal{P}}(x) / \log_2(k + k_0(x))$ to trivialize this hypothesis.

Finally note that having more constraints in the global state, e.g.

$$\bar{S}_{\mathcal{P}} = \left\{ x \in \mathbb{R}^d_{>0} \mid \sum_{k=1}^{d} x_k = 1,\ c(x) = C \right\}$$

instead of (6.2), where $c(x) = c(x_1, \dots, x_d) = C \in \mathbb{R}^n$, does not allow to trivially repeat the proof, and this open the possibility to have different distributions. For example, in an intelligent interpretation of a clinical note, we can consider a reader looking for information about cancer first diagnosis. This gives a constraint on the words of the text, so that the reader would experience lower costs if the words is related to the information she/he is looking for.

7. Generalizing Von Thünen's model

Von Thünen's model, see [42], tries to answer the basic questions of location and land use theory: "where should a certain activity be located?" and "which activity should be chosen at a certain location?". Both questions address the principles underlying the spatial layout of an economy. Several original assumptions of Von Thünen's model can be weakened and generalized by simply using an appropriate mathematical notations and modeling; later, we will see them in details. Most important for us is that, in this model, costs and forces of diversification have simple properties that can be generalized to other CAS. This actually show another feature of having a common mathematical language for CS: describing a particular model with the language of IS theory, one can recognize that the obtained results can be actually generalized and hence potentially applied to several other CS.

In the present section, we explicitly use units of measurement, because this greatly helps the understanding of the different economic quantities that we are going to introduce. We will use notations such as $\left[\frac{€}{\mathrm{m}^2}\right]$ to denote the 1-dimensional (totally ordered real vector) space of quantities whose unit of measurement is € per square meter. Mathematically, it can be identified with the space of polynomials in the unknowns $€\cdot\mathrm{m}^{-2}$.

7.1. **Von Thünen's impedance zones.** We need to introduce several quantities and notations:

1. $B \in \mathbb{N}_{>0}$ the *number of commodities* produced by the considered economy. There are no a priory assumptions on the type of commodities (e.g. not necessarily of agricultural type).

2. For all $b = 1, \dots, B$, a unit of measurement u_b for the commodity b. For example u_b can be ton, or an integer number $n \in \mathbb{N}_{>0}$ so that $[u_b] = \mathbb{R}$ (dimensionless), or box, etc.
3. $x_m \in \mathbb{R}^2$ is the *location of the market* m.
4. $A \subseteq \mathbb{R}^2$ all possible *locations for the companies* that produce some commodity $b = 1, \dots, B$.
5. $y_b : A \longrightarrow \left[\frac{u_b}{\mathrm{m}^2}\right]_{>0}$, $y_b(x) > 0$ is the *yield of the commodity* b if the company is located at $x \in A$. The model is not stochastic, so quantities like these can denote average values. The use of a spatial unit of measurement like m^2 is only useful so as to not use too heavy notations. More appropriate units of production v_b, depending on the commodity b, could be introduced instead of m^2 (e.g. it could be $v_b = \mathrm{kwh}$, or $v_b = \frac{\mathrm{hour}}{\mathrm{man}}$).
6. $p_b(x_m) \in \left[\frac{€}{u_b}\right]_{>0}$ *price of the commodity* b in the market x_m per units of b.
7. $c_b : A \longrightarrow \left[\frac{€}{u_b}\right]_{>0}$, $c_b(x) > 0$ is the *production cost* of b at the location $x \in A$ per units of b.
8. $j(-, x_m) : A \longrightarrow [i]_{\geq 0}$, $j(x, x_m)$ is the *impedance* between the location $x \in A \subseteq \mathbb{R}^2$ and the market $x_m \in \mathbb{R}^2$ measuring the pure transportation costs to move from x to x_m (see also below); the impedance has unit of measurement i (e.g. i can be $i = \mathrm{km}$, $i = \mathrm{hour}$, $i = \frac{\mathrm{hour}}{\mathrm{man}}$, $i = €$, etc.). We do not need any assumption about existence or non existence of possible routes, neither on the nature of the transportation, nor on the linearity of j with respect to the distance between x and x_m along the shortest route, not even if we are actually moving goods from x to x_m or vice versa (this is only a useful intuitive interpretation in case of application in land use theory).
9. $F_b : [i]_{\geq 0} \longrightarrow \left[\frac{€}{u_b}\right]_{>0}$, $F_b(d)$ is the *transportation cost* of the commodity b per units of b and for any pair of points, $(x, x_m) \in \mathbb{R}^2$ having impedance $d \in [i]_{>0}$. For example, $F_b(d)$ can be lower if we assume the possibility to use refrigerators for the transportation of a dairy product b. This modeling assumption includes the particular case where $F_b(d) = \bar{F}_b \cdot d$, where $\bar{F}_b \in \left[\frac{€}{u_b \cdot i}\right]_{>0}$, i.e. the case where the transportation cost is proportional to the impedance d. Usually, one assumes that the transportation cost is increasing with the impedance d:

$$\forall d_1, d_2 \in [i]_{\geq 0} : \ d_1 \leq d_2 \Rightarrow F_b(d_1) \leq F_b(d_2). \tag{7.1}$$

   However, we will never use this assumption.
10. $k_b(-, x_m) : A \longrightarrow \left[\frac{€}{\mathrm{m}^2}\right]_{>0}$, $k_b(x, x_m)$ is the average *life cost* (everything but the rent of the company's location), per m^2, of the owner producing b at x and selling b in the market x_m.

We can now introduce the basic quantities of Von Thünen's model, using both a specific language of land use theory and a more general one:

Definition 4. The *(land) value* or *locational rent* at $x \in A$ with respect to the production of $b = 1, \dots, B$ and the market x_m is

$$L_b(x, x_m) := y_b(x) \cdot [p_b(x_m) - c_b(x) - F_b(j(x, x_m))] \in \left[\frac{€}{\mathrm{m}^2}\right]. \tag{7.2}$$

The *ideal value/rent* for a company located *at x with respect to the market located at x_m* is

$$R(x, x_m) := \max_{b=1,\dots,B} [L_b(x, x_m) - k_b(x, x_m)] \in \left[\frac{€}{\mathrm{m}^2}\right]. \tag{7.3}$$

From this, we deduce that the life cost must satisfy the constraint

$$k_b(x, x_m) < L_b(x, x_m) \quad \forall x \in A. \tag{7.4}$$

Moreover, we say that x *is a good (location) for (the production of) b* if

$$R(x, x_m) = L_b(x, x_m) - k_b(x, x_m), \tag{7.5}$$

i.e. if the ideal rent R equals the land value L_b minus the life cost k_b. Finally, the *impedance boundaries around the sink/market x_m* are given by

$$\underline{r}_b(x_m) := \inf \{ j(x, x_m) \mid x \in A,\ x \text{ is good for } b \} \in [i]_{\geq 0} \tag{7.6}$$

$$\overline{r}_b(x_m) := \sup \{ j(x, x_m) \mid x \in A,\ x \text{ is good for } b \} \in [i]_{\geq 0}. \tag{7.7}$$

Therefore, if a company is located at x, which is a good location for the production of b, we trivially have that $\underline{r}_b(x_m) \leq j(x, x_m) \leq \overline{r}_b(x_m)$, that is the company is located in the corresponding *impedance zone* bounded by $\underline{r}_b(x_m) \leq \overline{r}_b(x_m)$. Note that if at least two locations are good for b and have different impedance, the zone is non trivial, i.e. $\underline{r}_b(x_m) < \overline{r}_b(x_m)$.

In this setting, only definition (7.2) is really specific of this model of land use. For an arbitrary IS, we can assume to have a *value function* $L_b : A \times A \longrightarrow \mathbb{R}$, a *cost function* $k_b : A \times A \longrightarrow \mathbb{R}$ satisfying (7.4) and an *impedance function* $j(-, x_m) : A \longrightarrow [i]_{\geq 0}$. With these, we can define the quantities $R(x, x_m)$ as in (7.3) (the *ideal value* for all the indexes $b = 1, \dots, B$), the property of being a *good x for b*, as in (7.5) and the impedance boundaries as in (7.6) and (7.7).

7.2. **Disjoint impedance zones.** We now want to see under what assumptions the impedance zones are disjoint, i.e. when $\underline{r}_\beta(x_m) \leq \overline{r}_\beta(x_m) \leq \underline{r}_b(x_m) \leq \overline{r}_b(x_m)$ if $b \neq \beta$ are two different indexes/commodities. We need to hypothesize that the cost of life k_b does not depend on the location x. We will use e.g. the notation $k_b \equiv k_b(-) \in \left[\frac{€}{\mathrm{m}^2}\right]$. In land use theory, this is clearly an assumption which holds only if $A \subseteq \mathbb{R}^2$ is not too large; e.g. it surely does not hold for locations situated in different countries, with different cost of labor, different climate conditions and different life costs.

Let b, $\beta = 1, \dots, B$, $b \neq \beta$, be two commodities. For simplicity, we omit the dependence by the market's location x_m. By contradiction, assume that

$$\overline{r}_\beta > \underline{r}_b \tag{7.8}$$

Definition (7.7) yields the existence of a location $x \in A$ which is good for β and such that $\underline{r}_b < j(x) \leq \overline{r}_\beta$. Since x is good for β, we have

$$R(x) = L_\beta(x) - k_\beta = \max_b [L_b(x) - k_b] \geq L_b(x) - k_b. \tag{7.9}$$

Analogously, from (7.6) and $\underline{r}_b < j(x)$ we get the existence of $y \in A$ which is good for b and such that

$$\underline{r}_b \leq j(y) < j(x) \tag{7.10}$$

$$R(y) = L_b(y) - k_b \geq L_\beta(y) - k_\beta. \tag{7.11}$$

How can it happen that x is a good location for β and y is not a good location for it, even if $j(y) < j(x)$? To understand this point, we compare the *gain (of net land value) passing from x to y*, for an arbitrary commodity β we have:

$$\Delta\eta_\beta(x,y) := [L_\beta(y) - k_\beta] - [L_\beta(x) - k_\beta] = L_\beta(y) - L_\beta(x) \tag{7.12}$$

If we assume $\Delta\eta_\beta(x,y) \ge \Delta\eta_b(x,y)$ and consider (7.12), we get

$$\begin{aligned} L_\beta(y) &\overset{(7.12)}{=} L_\beta(x) + \Delta\eta_\beta(x,y) \ge L_\beta(x) + \Delta\eta_b(x,y) \\ &\overset{(7.12)}{=} L_\beta(x) + L_b(y) - L_b(x) \\ &\ge L_\beta(x) + L_b(y) - L_b(x) - k_\beta + k_\beta \\ &\overset{(7.9)}{\ge} L_b(x) - k_b + L_b(y) - L_b(x) + k_\beta \\ &= -k_b + L_b(y) + k_\beta \end{aligned}$$

and hence $L_\beta(y) - k_\beta \ge L_b(y) - k_b$ so that $L_\beta(y) - k_\beta = L_b(y) - k_b$ by (7.11). Therefore, y is a good location both for β and b. We therefore proved that if $\Delta\eta_\beta(x,y) \ge \Delta\eta_b(x,y)$ and the impedance zones of b and β intersect, then $L_\beta(y) - k_\beta = L_b(y) - k_b$.

We can summarize this arguments with the following

Theorem 5 (J.H. von Thünen)**.** *Let us assume that k_b does not depend on the location $x \in A$. Let b, β be two commodities such that for all x, $y \in A$*

$$\Delta\eta_\beta(x,y) \ge \Delta\eta_b(x,y) \tag{7.13}$$

$$L_\beta(y) - k_\beta \ne L_b(y) - k_b. \tag{7.14}$$

Then impedance zones of b and β (around the market x_m) are disjoint, i.e. $\underline{r}_\beta \le \overline{r}_\beta \le \underline{r}_b \le \overline{r}_b$. Therefore, if any pair of different commodities always have different land values (7.14) and different variations of transportation costs (7.13) (for all x, $y \in A$), then we can order the commodities so that

$$\underline{r}_{b_1} \le \overline{r}_{b_1} \le \underline{r}_{b_2} \le \overline{r}_{b_2} \le \ldots \le \underline{r}_{b_B} \le \overline{r}_{b_B}.$$

The same result holds in any IS where we can define a value function $L_b : A \longrightarrow \mathbb{R}$, a cost $k_b \in \mathbb{R}$ and an impedance function $j : A \longrightarrow [i]_{\ge 0}$ for each interacting entities $b \in \mathcal{P}$ in a finite population, and for the quantities as defined in Def. 4. Note that the moving of commodities from $x \in A$ to the market x_m is only an intuitive interpretation in land use theory, but in more general IS we can also have the opposite movement.

All this serves to underscores that we can have disjoint impedance zones even if we do *not* have an optimized economy, i.e. a complex adaptive economy. Therefore, this is not a peculiarity of a CAS because we are not minimizing any cost nor maximizing any diversification force. On the contrary, in the next section, we will consider what happen when the subpopulation of companies producing the same commodity adapts following the GEP, i.e. evolves decreasing natural costs in the most diversified way.

7.3. **Von Thünen's model and the generalized evolution principle.** In this section, we assume that our IS satisfies the GEP, i.e. it is a CAS, and it evolved into a stationary emergent pattern configuration, where natural costs are minimum and diversification forces are maximum. For simplicity, we still continue to use a language of land use theory, for example assuming to have $n_b \in \mathbb{N}_{>0}$ companies producing the commodity $b = 1, \ldots, B$. However, it is clear that exactly the same deductions apply to any IS where the interacting entities are $E = \{(a, b) \mid b = 1, \ldots, B,\ a = 1, \ldots, n_b\} = \mathcal{P}$ and we have a value function $L_b : A \longrightarrow \mathbb{R}$ and a cost function $k_b : A \longrightarrow \mathbb{R}$ for all $b = 1, \ldots, B$. We do not actually even need an impedance function $j : A \longrightarrow [i]_{\geq 0}$ as above.

7.3.1. *Cost minimization.* In a non adapted economy, we do not necessarily have that rents coincide with their ideal values (7.3), e.g. because of ignorance of some agent with respect to the entire market configuration. For simplicity, in this section we assume to consider only one market x_m, so that all the prices, rents and values always refer to x_m and we will omit this variable. Then, the configuration space of a generic economy located around the market x_m is given by:

(i) A location $x_b^a \in A$ for each company $a = 1, \ldots, n_b$ producing the commodity $b = 1, \ldots, B$. We assume that $x_b^{a_1} \neq x_b^{a_2}$ if $a_1 \neq a_2$.
(ii) In each one of these locations, we have a rent $r_b^a \in \left[\frac{€}{\text{m}^2}\right]$ really applied to the company a producing b at x_b^a.

Using these notations, we can model configurations representing a positive cost for some agent acting as location renter or tenant owning a company. For example the inequalities

$$r_b^a < L_b(x_b^i) - k_b(x_b^a) < R(x_b^a)$$

imply

(i) $r_b^a < L_b(x_b^a) - k_b(x_b^a)$: the location x_b^a is rented at a price r_b^a which is less than the land value (it could be rented at a higher price).
(ii) $r_b^a < R(x_b^a)$: the rent r_b^a is less than the maximum rent that it would be possible to ask in the location x_b^a (at another company producing a different commodity).

Considering $<$, $=$ or $>$, all the possible inequalities are $3^3 = 27$, but several of them are mathematically impossible, repetitions, or impossible relations due to the definition of $R(x)$ (see (7.3)). Only the following possible inequalities remain:

Definition 6. The configuration space $M = M_{\mathcal{P}, I_{\mathcal{P}}}$ of the economies centered around the market x_m is defined by

$$(x_b^1, r_b^1, \ldots, x_b^{n_b}, r_b^{n_b})_{b=1,\ldots,B} \in M$$

if and only if for all $b = 1, \ldots, B$ and all $a = 1, \ldots, n_b$, we have $x_b^a \in A$, $r_b^a \in \left[\frac{€}{\text{m}^2}\right]$, $x_b^{a_1} \neq x_b^{a_2}$ if $a_1 \neq a_2$, and at least one of the following conditions is satisfied

$$\begin{aligned} r_b^a &= L_b(x_b^a) - k_b(x_b^a) = R(x_b^a) \\ r_b^a &= L_b(x_b^a) - k_b(x_b^a) < R(x_b^a) \\ r_b^a &< L_b(x_b^a) - k_b(x_b^a) = R(x_b^a) \\ r_b^a &< L_b(x_b^a) - k_b(x_b^a) < R(x_b^a) \\ r_b^a &> L_b(x_b^a) - k_b(x_b^a) = R(x_b^a) \end{aligned} \tag{7.15}$$

$$\begin{aligned} r_b^a &= R(x_b^a) > L_b(x_b^a) - k_b(x_b^a) \\ r_b^a &> R(x_b^a) > L_b(x_b^a) - k_b(x_b^a) \\ R(x_b^a) &> r_b^a > L_b(x_b^a) - k_b(x_b^a). \end{aligned} \tag{7.16}$$

We also use the simplified notation $(x,r) = (x_b^1, r_b^1, \ldots, x_b^{n_b}, r_b^{n_b})_{b=1,\ldots,B}$ to denote a configuration. We will think at (x_b^a, r_b^a) as two components of the proper state space of the interacting entity $(a,b) \in \mathcal{P}$, i.e. the company $a = 1, \ldots, n_b$ that produces the commodity $b = 1, \ldots, B$.

The first one (7.15) of these conditions is called *von Thünen configuration*. Each one of these, except the von Thünen one, corresponds to a possible configuration of an economy where at least one of its agents is paying a cost or is loosing a profit:

Definition 7. Let $b = 1, \ldots, B$, $a = 1, \ldots, n_b$ and $(x_b^1, r_b^1, \ldots, x_b^{n_b}, r_b^{n_b})_{b=1,\ldots,B} \in M$, then

(i) The cost paid by the tenant/company a, which produces the commodity b, is located in x_b^a and pays the rent r_b^a is

$$c_{\mathrm{t}}(x_b^a, r_b^a) := \begin{cases} r_b^a - [L_b(x_b^a) - k_b(x_b^a)] & \text{if } r_b^a > L_b(x_b^a) - k_b(x_b^a) \\ c_{0\mathrm{t}} & \text{otherwise.} \end{cases}$$

The quantity $c_{0\mathrm{t}} \geq 0$ represents a minimum non avoidable cost related to the rent of this location (e.g. administrative cost) and paid by the tenant. We assume $c_{0\mathrm{t}}$ sufficiently small, i.e. satisfying

$$r_b^a > L_b(x_b^a) - k_b(x_b^a) \;\Rightarrow\; c_{0\mathrm{t}} < c_{\mathrm{t}}(x_b^a, r_b^a) \tag{7.17}$$

for all a, b (note that this implies $c_{0\mathrm{t}} \leq c_{\mathrm{t}}(x_b^a, r_b^a)$, and if $c_{0\mathrm{t}} = c_{\mathrm{t}}(x_b^a, r_b^a)$, then necessarily $r_b^a \leq L_b(x_b^a) - k_b(x_b^a)$). Clearly, if $r_b^a > L_b(x_b^a) - k_b(x_b^a)$, the company a is paying a rent r_b^a higher than the company's gain $L_b(x_b^a) - k_b(x_b^a)$, and it is hence at a loss.

(ii) The loss of profit of the renter of the location x_b^a at rent r_b^a for the production of the commodity b is

$$l_{\mathrm{r}}^1(x_b^a, r_b^a) := \begin{cases} [L_b(x_b^a) - k_b(x_b^a)] - r_b^a & \text{if } r_b^a < L_b(x_b^a) - k_b(x_b^a) \\ l_{1\mathrm{r}} & \text{otherwise.} \end{cases}$$

The quantity $l_{1r} \geq 0$ represents a minimum non avoidable cost related to the rent of this location (e.g. some tax or the cost to know the land value $L_b(x_b^a)$ and the cost of life $k_b(x_b^a)$) and paid by the renter. We assume $l_{1\mathrm{r}}$ sufficiently small, i.e.

$$r_b^a < L_b(x_b^a) - k_b(x_b^a) \;\Rightarrow\; l_{1\mathrm{r}} < l_{\mathrm{r}}^1(x_b^a, r_b^a) \tag{7.18}$$

for all a, b (as above, this yields $l_{1\mathrm{r}} \leq l_{\mathrm{r}}^1(x_b^a, r_b^a)$, and if $l_{1\mathrm{r}} = l_{\mathrm{r}}^1(x_b^a, r_b^a)$, then $r_b^a \geq L_b(x_b^a) - k_b(x_b^a)$). If $r_b^a < L_b(x_b^a) - k_b(x_b^a)$, the renter is asking a lower rent r_b^a with respect to the better value $L_b(x_b^a) - k_b(x_b^a)$.

(iii) The loss of profit of the renter of the location x_b^a at rent r_b^a with respect to all the possible commodities is

$$l_{\mathrm{r}}^2(x_b^a, r_b^a) := \begin{cases} R(x_b^a) - r_b^a & \text{if } r_b^a < R(x_b^a) \\ l_{2\mathrm{r}} & \text{otherwise} \end{cases}$$

The quantity $l_{2r} \geq 0$ represents a minimum non avoidable cost related to the rent of this location (e.g. some tax or the cost to know the ideal rent $R(x_b^b)$) and paid by the renter. We assume $l_{2\mathrm{r}}$ sufficiently small, i.e.

$$r_b^a < R(x_b^a) \;\Rightarrow\; l_{2\mathrm{r}} < l_{\mathrm{r}}^2(x_b^a, r_b^a) \tag{7.19}$$

for all a, b. Again: $l_{2\mathrm{r}} \le l^2_{\mathrm{r}}(x^a_b, r^a_b)$ and if $l_{2\mathrm{r}} = l^2_{\mathrm{r}}(x^a_b, r^a_b)$, then $r^a_b \ge R(x^a_b)$. If $r^a_b < R(x^a_b)$, the renter is asking a rent r^a_b lower that the best possible one $R(x^a_b)$.

To these costs and losses, we associate the following probabilities and hence expected costs and losses

Definition 8. Let $(x, r) \in M$, then

(i) The cost of the tenant C_{t} can be computed as

$$C_{\mathrm{t}}(x, r) := \sum_{b=1}^{B} \sum_{a=1}^{n_b} p_{\mathrm{t}}(x^a_b, r^a_b) \cdot c_{\mathrm{t}}(x^a_b, r^a_b), \tag{7.20}$$

where $\left(p_{\mathrm{t}}(x^a_b, r^a_b)\right)_{a,b}$ is *any* fixed *non-degenerate* probability distribution, i.e. $p_{\mathrm{t}}(x^a_b, r^a_b) > 0$ for all a, b. Note that $C_{\mathrm{t}}(x, r) \ge c_{0\mathrm{t}}$ because $c_{0\mathrm{t}} \le c_{\mathrm{t}}(x^a_b, r^a_b)$ for all a, b, and that the probability $\left(p_{\mathrm{t}}(x^a_b, r^a_b)\right)_{a,b}$ may depend on the cost C_{t}. Moreover, because of (7.17) and the non-degenerateness condition, if $C_{\mathrm{t}}(x, r) = c_{0\mathrm{t}}$, then $c_{\mathrm{t}}(x^a_b, r^a_b) = c_{0\mathrm{t}}$ for all a, b.

(ii) Let $j = 1, 2$, the losses of profits of the renter can be computed as

$$L^j_{\mathrm{r}}(x, r) := \sum_{b=1}^{B} \sum_{a=1}^{n_b} p^j_{\mathrm{r}}(x^a_b, r^a_b) \cdot l^j_{\mathrm{r}}(x^a_b, r^a_b), \tag{7.21}$$

where $\left(p^j_{\mathrm{r}}(x^a_b, r^a_b)\right)_{a,b}$, for $j = 1, 2$, are *any* fixed *non-degenerate* probability distribution, i.e. $p^j_{\mathrm{r}}(x^a_b, r^a_b) > 0$ for all a, b. As above $L^j_{\mathrm{r}}(x, r) \ge l_{j\mathrm{r}}$, and if $L^j_{\mathrm{r}}(x, r) = l_{j\mathrm{r}}$, then $l^j_{\mathrm{r}}(x^a_b, r^a_b) = l_{j\mathrm{r}}$ for all a, b, because of (7.18) and (7.19) and the non-degenerateness condition.

Compare (7.20), (7.21) with (3.4) to recognize that $P^{\mathrm{t}}_{(x,r)}(a, b) := p_{\mathrm{t}}(x^a_b, r^a_b)$ and $P^{j,\mathrm{r}}_{(x,r)}(a, b) := p^j_{\mathrm{r}}(x^a_b, r^a_b)$ are the probabilities to average the considered unification costs.

These expected costs and losses are minimum at a von Thünen configuration: $r^a_b = L_b(x^a_b) - k_b(x^a_b) = R(x^a_b)$ for all a, b. Vice versa, when a given configuration $(x^1_b, r^1_b, \dots, x^{n_b}_b, r^{n_b}_b)_{b=1,\dots,B} \in M$ minimize costs and losses, is it a von Thünen configuration? In general this is false: let $A = \{y_1, y_2\}$, $B = 2$, with

$$\begin{aligned} L_1(y_j) - k_1(y_j) &= 1\frac{€}{\mathrm{m}^2} \\ L_2(y_j) - k_2(y_j) &= 2\frac{€}{\mathrm{m}^2}. \end{aligned}$$

That is the commodity $b_1 = 1$ has a net land value equal to $1\frac{€}{\mathrm{m}^2}$ in both locations y_1, y_2, whereas the commodity $b_2 = 2$ has a double net land value in both locations. Assume, by contradiction, that $(x^1_1, r^1_1, x^2_2, r^2_2) \in M$ minimizes costs and losses. We would have

$$\begin{aligned} c_{\mathrm{t}}(x^1_1, r^1_1) = c_{0\mathrm{t}} &\iff r^1_1 \le L_1(x^1_1) - k_1(x^1_1) = 1\frac{€}{\mathrm{m}^2} \\ l^1_r(x^1_1, r^1_1) = l_{1\mathrm{r}} &\iff r^1_1 \ge L_1(x^1_1) - k_1(x^1_1) = 1\frac{€}{\mathrm{m}^2} \\ l^2_r(x^1_1, r^1_1) = l_{2\mathrm{r}} &\iff r^1_1 \ge R(x^1_1) = 2\frac{€}{\mathrm{m}^2}. \end{aligned} \tag{7.22}$$

This contradiction proves that, in this system, no configuration minimizes costs and losses to the minimum values $c_{0\mathrm{t}}$, $l_{1\mathrm{r}}$, $l_{2\mathrm{r}}$. In other words, this system does not allow for a von Thünen configuration. The problem here is that the system is lacking in possible configurations. This justifies the main assumption (7.23) of the following

Theorem 9. *If any company can be situated in a good location:*

$$\forall b = 1, \dots, B \, \forall a = 1, \dots, n_b \, \exists y \in A : \ y \text{ is good for } b, \tag{7.23}$$

then a given configuration $(x, r) \in M$ minimizes the average costs and losses

$$\begin{aligned} &C_t(x,r) \le C_t(y,s) \\ &L^1_r(x,r) \le L^1_r(y,s) \quad \forall (y,s) \in M \\ &L^2_r(x,r) \le L^2_r(y,s) \end{aligned} \tag{7.24}$$

if and only if (x, r) is a von Thünen configuration, i.e.

$$r^a_b = L_b(x^a_b) - k_b(x^a_b) = R(x^a_b) \quad \forall a, b.$$

The same result holds in any IS having a value function $L_b : A \longrightarrow \mathbb{R}$ and a cost function $k_b : A \longrightarrow \mathbb{R}$ for each interacting entities $(a, b) \in \mathcal{P}$ in a finite population, and for the quantities as defined in Def. 8.

Proof. If (x, r) is a von Thünen configuration, then $C_{\mathrm{t}}(x, r) = c_{0\mathrm{t}} \le C_t(y, s)$, $L^1_{\mathrm{r}}(x, r) = l_{1\mathrm{r}} \le L^1_{\mathrm{r}}(y, s)$, $L^2_{\mathrm{r}}(x, r) = l_{2\mathrm{r}} \le L^2_{\mathrm{r}}(y, s)$ by Def. 7, so that (7.24) hold. Vice versa, using assumption (7.23), we can construct a von Thünen configuration (y, s) by choosing a good location for every commodity:

$$\forall b, a \, \exists y^a_b \in A : \ R(y^a_b) = L_b(y^a_b) - k_b(y^a_b).$$

But (7.24) yields

$$\begin{aligned} c_{0\mathrm{t}} &\le C_{\mathrm{t}}(x,r) \le C_{\mathrm{t}}(y,s) = c_{0\mathrm{t}} \\ l_{1\mathrm{r}} &\le L^1_{\mathrm{r}}(x,r) \le L^1_{\mathrm{r}}(y,s) = l_{1\mathrm{r}} \\ l_{2\mathrm{r}} &\le L^2_{\mathrm{r}}(x,r) \le L^2_{\mathrm{r}}(y,s) = l_{2\mathrm{r}}. \end{aligned}$$

Therefore, $r^a_b = L_b(x^a_b) - k_b(x^a_b) \ge R(x^a_b)$ because of Def. 7 and Def. 8. Therefore, also (x, r) is necessarily a von Thünen configuration by (7.15). □

Clearly, assumption (7.23) is not completely realistic in real-world economies. Once again, this underscores that what cost functions to consider in this kind of economic models, is a modeling/philosophical/political choice. Depending on our political choices, other types of costs can be considered, such as: environmental costs, energy consumption, loss of profits, state's costs due to loss of job places because of transfer of branches, company's stock price, etc. One can also argue that also the maximization of profit is unrealistic because it implies the full knowledge of the entire system and unrealistic tendencies, such as that of trying the renting of centered locations to jewelries as soon as the demand of jewels increases.

In another political approach, we can denote by $\mathrm{tax}(x^a_b)$ the real estate tax that must be paid by the owner (renter) of the real estate property located (and rented) at x_b^a; then, we can consider the previous cost of the tenant C_t and, instead of the losses of profits L_r^j, and examine the cost

$$c_\mathrm{r}(x_b^a, r_b^a) := \begin{cases} \mathrm{tax}(x_b^a) - r_b^a & \text{if } r_b^a < \mathrm{tax}(x_b^a) \\ 0 & \text{otherwise,} \end{cases} \tag{7.25}$$

and the corresponding average cost C_r. In this model, any configuration where $r_b^a \le L_b(x_b^a) - k_b(x_b^a)$ and $r_b^a \ge \mathrm{tax}(x_b^a)$ minimizes the costs C_t and C_r. We have hence a more realistic model without any assumption of profits maximization or of full information about the market.

7.3.2. *Maximization of diversification forces.* A very natural flux of goods is the amount of commodity $b = 1, \ldots, B$ that each company $a = 1, \ldots, n_b$ is selling to the market x_m. We therefore want to see what the maximization of the diversification forces yields if we consider these interactions. We think at a generic IS where interacting entities are $E = \{(a, b) \mid b = 1, \ldots, B,\ a = 1, \ldots, n_b\} \cup \{x_m\}$, even if we still keep a language of land use theory. The population we want to consider is hence $\mathcal{P}_b = \{(a, b) \in E \mid a = 1, \ldots, n_b\}$ for each fixed $b = 1, \ldots, B$, i.e. the collection of the aforementioned companies, so that its cardinality is $|\mathcal{P}_b| = n_b$. We can therefore think at "selling" interactions $i_b^a : (a, b) \xrightarrow{r_b^a, \mathrm{s}} x_m$ having the company (a, b) as agent and the market x_m as patient, and hence we consider $I_{\mathcal{P}_b} := \{i_b^a \mid a = 1, \ldots, n_b\}$. The good of i_b^a is a quantity $\varphi_b^a \in (0, Q_b] =: R_{i_b^a}$ belonging to the space of resources $R_{i_b^a}$ of the interaction i_b^a. Using an example only to understand, in a simpler deterministic model of constant production, if s_b^a is the surface of the company a used for the production of 1 year of b (more generally, the amount of units of production v_b used in 1 year, see definition 5. in Sec. 7.1), then we have

$$\varphi_b^a = y_b(x_b^a) \cdot s_b^a \in \left[\frac{u_b}{\mathsf{year}}\right]$$
$$\Delta_b = \sum_{a=1}^{n_b} \varphi_b^a.$$

We can thus think at Δ_b as the amount of commodity b demanded by x_m in 1 year.

In general, i.e. independently from this particular example of land use theory, we consider a global state space $\bar{M}_{\mathcal{P}_b, I_{\mathcal{P}_b}}$ such that

$$R_{\mathcal{P}_b} = \left\{ (\varphi_b^1, \ldots, \varphi_b^{n_b}) \in (0, \Delta_b]^{n_b} \mid \sum_{a=1}^{n_b} \varphi_b^a = \Delta_b \right\}, \tag{7.26}$$

and the diversification probability corresponding to a Bernoulli process

$$Q_\gamma(i_b^a) := \frac{\gamma^a}{\Delta_b} = \frac{\varphi_b^a}{\Delta_b} =: q_b^a \quad \forall \gamma = (\varphi_b^1, \ldots, \varphi_b^{n_b}) \in R_{\mathcal{P}_b}. \tag{7.27}$$

In other words, the probability to extract a unit of commodity b produced by the company a, among all those flowing to the market x_m in one year, equals the fraction $\frac{\varphi_b^a}{\Delta_b}$ of goods φ_b^a produced by a over the demanded total Δ_b.

Assume now that the system is in an emergent pattern state $\gamma = (\varphi_b^1, \dots, \varphi_b^{n_b})$. We have that

$$D_{I_{\mathcal{P}_b}}(\gamma) =: D_{I_{\mathcal{P}_b}}(\varphi_b^1, \dots, \varphi_b^{n_b}) = -\sum_{i \in I_{\mathcal{P}_b}} Q_\gamma(i) \cdot \log_2 Q_\gamma(i) \tag{7.28}$$

$$= -\sum_{a=1}^{n_b} q_b^a \cdot \log_2 q_b^a \geq D_{I_{\mathcal{P}_b}}(s) \quad \forall s. \tag{7.29}$$

As we already have seen in Thm. 3, a global space of resources such as (7.26) allows the model to consider all the possible finite probabilities. As in Thm. 3, we can use Lagrange's multiplier to get

$$\exists \lambda \in \mathbb{R} : \left. \frac{\partial D_{I_{\mathcal{P}_b}}(\gamma^1, \dots, \gamma^{n_b})}{\partial \gamma^a} \right|_{\gamma^{(-)} = \varphi_b^{(-)}} =$$

$$= \frac{\partial}{\partial \gamma^a} \left. \left(-\sum_{a=1}^{n_b} \frac{\gamma^a}{\Delta_b} \cdot \log_2 \frac{\gamma^a}{\Delta_b} \right) \right|_{\gamma^{(-)} = \varphi_b^{(-)}} =$$

$$= \lambda \cdot \frac{\partial}{\partial \gamma^a} \left. \left(\sum_{a=1}^{n_b} \frac{\gamma^a}{\Delta_b} \right) \right|_{\gamma^{(-)} = \varphi_b^{(-)}}.$$

This gives $\varphi_b^a = \frac{2^{-\lambda}}{e \Delta_b}$, so that $\Delta_b = \sum_{a=1}^{n_b} \varphi_b^a = n_b \cdot \frac{2^{-\lambda}}{e \Delta_b}$ and

$$\varphi_b^a = \frac{\Delta_b}{n_b}. \tag{7.30}$$

We can state this result as a general

Theorem 10. *Let $\mathcal{I}$ be an IS with interacting entities $E = \{(a, b) \mid b = 1, \dots, B,\ a = 1, \dots, n_b\} \cup \{x_m\}$. Consider the population $\mathcal{P}_b = \{(a, b) \in E \mid a = 1, \dots, n_b\}$ for each fixed $b = 1, \dots, B$ and with adapting interactions $I_{\mathcal{P}_b} := \{i_b^a \mid a = 1, \dots, n_b\}$, where $i_b^a : (a, b) \xrightarrow{r_b^a, s} x_m$, with global space of resources given by (7.26) and diversification Bernoulli probabilities (7.27). Then $\gamma = (\varphi_b^1, \dots, \varphi_b^{n_b}) \in R_{\mathcal{P}_b}$ is a pattern with maximum diversification $D_{I_{\mathcal{P}_b}}$ if and only if $\varphi_b^a = \frac{\Delta_b}{n_b}$ holds for all $(a, b) \in E$.*

In land use theory, fluxes of commodities b are hence equally divided among all companies. We can interpret this property as a resilience characteristic of the adapted population $\mathcal{P}_b$, due to the absence of monopolies, maximal distribution of work, etc. We can hence say that in this case the GEP coincides with a well studied property of stable economies. In other words, if the population $\mathcal{P}_b$ follows the GEP, then the companies try to move in different configuration $(x, r) \in M$ so that to decrease the average costs of Def. 7 and, at the same time, thanks to some forces not explicitly represented in the model (e.g. a suitable taxation system or a clever use of resources or a population tendency to be resilient), they also evolve so that to satisfy (7.30), i.e. the absence of monopolies and a maximal *distribution* of work and resources. In case of a smaller space of resources $R_{\mathcal{P}_b}$(e.g. because some company is constrained to produce less than the best value), we do not have a uniform distribution.

Equation (7.30) can also be obtained from Thm. 3 and considering costs such as those of Def. 7 or (7.25) (in general any cost function $E_{\mathcal{P}}$ that does not depend on the probabilities, so that the corresponding derivative in (6.3) is zero). In Thm. 3, we can set $d = n_b$ and $x_j = \frac{\varphi_b^j}{\Delta_b}$, noting therefore that the costs do not depend on these probability but only on the configuration (x, r). Assuming (7.28) (or even only that the ratio $\frac{E_{\mathcal{P}_b}}{D_{I_{\mathcal{P}_b}}}$ is minimum) and setting $\alpha = 0$, we have $\sum_{k=1}^{n_b} k^{-\alpha_k \cdot \frac{D_{I_{\mathcal{P}_b}}}{E_{\mathcal{P}}}} = n_b$. Therefore, if at least Thm. 3 yields $q_k(y) = p_b^k(y) = p_b^1(y) = \frac{1}{n_b} = \frac{\varphi_b^a}{\Delta_b}$. This simple example shows how the general Thm. 3 can be applied to completely different settings.

It is clear that the result (7.30) does not correspond to a realistic behavior in real-world economies: However, it gives elements to start thinking at the GEP as a way to measure how far our economy is from a stable, adaptive, resilient system.

8. Conclusions and future developments

Even if it does not agree with a purely formalistic point of view in philosophy of mathematics, a mathematical definition must be validated as well. This validation ranges from useful and general theorems linked to this definition, to the inclusion of several interesting examples, in our case examples of CAS. It is therefore a very long process performed by the interested scientific community. On the one hand, we followed the ideas of G.K. Zipf, [45], which are nowadays informally frequently used in different modeling of CS, see e.g. [3, 12, 16, 20, 25, 28, 22, 30, 38] and references therein. The GEP can be explained and used even only at an intuitive level and the corresponding formalization is simple and corresponding to the intuition. This is already a form of validation. On the other hand, the present paper is only the first step in this validation process. For example, both the power law Thm. 3, or Thm. 9 and the results of Sec. 7.3.2 about von Thünen-like models of CS, viewed as sufficiently general mathematical results applicable to large classes of CS, move in this direction. Clearly, a better von Thünen's model (e.g. dynamical, stochastic also in the movement to different configurations (x, r), with explicit modeling of forces that allow the population to approach the ideal relation (7.30), with more than one markets, where the strong assumption (7.23) does not hold, etc.), or the applications of the related theorems to different CAS (e.g. phyllotaxis?) are only some of possible improvements.

As far as we know, the GEP is the first mathematical definition of CAS *with a proved universal applicability* supported by the mathematical embeddings of [18]. Having a clear mathematical notion could be of great advantage for the understanding and future development of the notion of CAS. For example, already the intuitive discussion to arrive at the GEP we had in Sec. 3, allowed us to recognize that the notion of CAS has to depend on validated costs, suitable probabilities to average these costs and measure the diversification forces, to identify a set of adaptive interactions, etc. Even the notion of emergent pattern can be mathematically defined, but can be reasonably considered less important than the general notion of GEP, where real-world CAS may only adapt evolving from a given state towards a better one.

The next important step in the theory of interaction spaces will be the definition of functors F preserving cause-effect relations between two interaction spaces $\mathcal{I}_1$ and $\mathcal{I}_2$, i.e. satisfying

$$a_1, \ldots, a_m \xrightarrow{\alpha, r} p \text{ in } \mathcal{I}_1 \quad \Rightarrow \quad F(a_1), \ldots, F(a_m) \xrightarrow{F(\alpha), F(r)} F(p) \text{ in } \mathcal{I}_2$$

In fact, this would allow us both to meaningfully define hierarchies of CS (each level of this hierarchy is related to another one by one of these cause-effect preserving functors). This notion includes a possible meaningful concept of *abstraction* if F is a forgetful functor, i.e. if it forgets some of the information in the states of the involved interacting entities. In this way, we also have a connection with the mathematical theory of multicategories, see [24]. Finally, if F is a forgetful functor, then a right adjoint functor $R : \mathcal{I}_2 \longrightarrow \mathcal{I}_1$ (in the opposite direction) formalizes the idea of *exploration of* $\mathcal{I}_1$ *from* $\mathcal{I}_2$; a left adjoint $L : \mathcal{I}_2 \longrightarrow \mathcal{I}_1$ formalizes the idea of *cause-effect simulation in* $\mathcal{I}_1$ *from* $\mathcal{I}_2$, see [19] for a first presentation of these ideas.

References

[1] Arnade C., Why It's Smart to Be Reckless on Wall Street, Scientific American Guest blog 2013, https://blogs.scientificamerican.com/guest-blog/why-its-smart-to-be-reckless-on-wall-street/

[2] Axtell, R.L., Zipf Distribution of U.S. Firm Sizes, Science 293, 1818 (2001). DOI: 10.1126/science.1062081

[3] Bak, P., *How nature works.* Oxford University Press, 1997.

[4] Bak P., Creutz M., Fractals and Self Organized Criticality. In: Bunde, A., Havlin, S. (eds) Fractals in Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-77953-4_2, 1994.

[5] Bahdanau, D., Cho, K., Bengio, J., Neural machine translation by jointly learning to align and translate. CoRR 2014, abs/1409.0473. https://api.semanticscholar.org/CorpusID:11212020. See also arXiv:1409.0473

[6] Barnett, N., Crutchfield, J.P., Computational Mechanics of Input–Output Processes: Structured Transformations and the ϵ-Transducer. J Stat Phys (2015) 161:404–451.

[7] Battiston, S., Puliga M., Kaushik R., Tasca P., Caldarelli G., DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk, Scientific Reports, 2:541, 2012, doi:10.1038/srep00541.

[8] Bouchard-Côté A., Hall D., Griffiths T.L., Klein D., Automated reconstruction of ancient languages using probabilistic models of sound change, PNAS vol. 110, no. 11, 2012, 4224–4229, doi: 10.1073/pnas.1204678110.

[9] Cameron E.Z., du Toit J.T., Winning by a Neck: Tall Giraffes Avoid Competing with Shorter Browsers, The American Naturalist 169, no. 1, 2007, 130-135, doi: 10.1086/509940

[10] Chadwick M.J., Jolly A.E.J., Amos D.P., Hassabis D., Spiers H.J., A Goal Direction Signal in the Human Entorhinal/Subicular Region, Current Biology 25, 87–92, January 5, 2015, http://dx.doi.org/10.1016/j.cub.2014.11.001

[11] Crespi B., Stead P., Elliot M., Comparative genomics of autism and schizophrenia, PNAS, vol. 107, suppl. 1, 2010, doi:10.1073/pnas.0906080106.

[12] Crutchfield, J.P., Between order and chaos, Nature Physics, Vol 8, January 2012, DOI: 10.1038/NPHYS2190

[13] B. Conrad, M. Mitzenmacher, Power Laws for Monkeys Typing Randomly: The Case of Unequal Probabilities, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 50, NO. 7, JULY 2004.

[14] Danowitz M., Vasilyev A., Kortlandt V., Solounias N., Fossil evidence and stages of elongation of the *Giraffa camelopardalis* neck, *R. Soc. opensci.* **2**:150393, 2015. http://dx.doi.org/10.1098/rsos.150393.

[15] Darwiche A., *Modeling and Reasoning with Bayesian Networks*, Cambridge University Press, 2009.

[16] Gabaix X., Zipf 's law for cities: An explanation. Quarterly Journal of Economics, CXIV:739–767, 1999.

[17] Gavrilets S., On the evolutionary origins of the egalitarian syndrome, PNAS vol. 109, no. 35, 2012, 14069–14074, doi: 10.1073/pnas.1201718109

[18] Giordano P., Interaction spaces: a universal mathematical theory of complex systems. See ttp://www.mat.univie.ac.at/~giordap7/IS.pdf

[19] Giordano P., Artificial general intelligence based on mathematical theory of complex systems. Slides of the seminar at SEAMS-UGM, 2023. http://www.mat.univie.ac.at/~giordap7/9th_SEAMS-UGM_2023.mp4.
[20] Haken H., *Synergetics. Introduction and Advanced Topics*, Springer Berlin, 2004.
[21] Herron M.D., Doebeli M., Parallel Evolutionary Dynamics of Adaptive Diversification in Escherichia coli. PLoS Biol 11(2): e1001490, 2013. doi:10.1371/ journal.pbio.1001490.
[22] James, R.G., Barnett, N., Crutchfield, J.P., Information Flows? A Critique of Transfer Entropies. Physical Review Letters 116, 238701 (2016). DOI: 10.1103/PhysRevLett.116.238701
[23] Kuhlemeier C., Phyllotaxis, Trends in Plant Science, Vol.12 No.4, 2007, doi:10.1016/j.tplants.2007.03.004.
[24] Leinster, T., *Higher Operads, Higher Categories*, Cambridge University Press, 2004.
[25] Mandelbrot B., An informational theory of the statistical structure of language, in W. Jackson (eds.), Communication Theory, Butterworths, London, 1953.
[26] Mandelbrot, B.,. Information theory and psycholinguistics: A theory of word frequencies. In P. F. Lazarsfield & N. W. Henry (Eds), Readings in Mathematical Social Sciences (pp. 350–368), 1966. Cambridge: MIT Press.
[27] Manin, D. Yu., Mandelbrot's Model for Zipf's Law: Can Mandelbrot's Model Explain Zipf's Law for Language?, Journal of Quantitative Linguistics, 16: 3, 274 - 285, 2009. doi: 10.1080/09296170902850358
[28] Miller J.H., Page S.E., *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press, 2007.
[29] Mitchell, M., Newman, M., Complex systems theory and evolution. In M. Pagel (editor), Encyclopedia of Evolution , New York: Oxford University Press, 2002.
[30] Nicolis G., Prigogine I., *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. Wiley, New York, 1977.
[31] Power R.A. et al., Polygenic risk scores for schizophrenia and bipolar disorder predict creativity, Nature neuroscience, Brief communications, 2015, doi:10.1038/nn.4040.
[32] Pluchino A., Garofalo C., Rapisarda A., Spagano S., Caserta M., Accidental politicians: How randomly selected legislators can improve parliament efficiency, Physica A: Statistical Mechanics and its Applications, Volume 390, Issues 21–22, 2011, 3944-3954, doi:10.1016/j.physa.2011.06.028.
[33] Reeves M., Levin S., Ueda D., The Biology of Corporate Survival. Natural ecosystems hold surprising lessons for business, Harvard Business Review, January-February 2016.
[34] Simmons R.E., Scheepers L., Winning by a Neck: Sexual Selection in the Evolution of Giraffe, The American Naturalist 148, no. 5, 1996, 771-786,doi:10.1086/285955.
[35] Smith E., Thermodynamics of natural selection I: Energy flow and the limits on organization, Journal of Theoretical Biology 252 (2008) 185–197.
[36] Smith E., Thermodynamics of natural selection II: Chemical Carnot cycles, Journal of Theoretical Biology 252 (2008) 198–212.
[37] Smith E., Thermodynamics of natural selection III: Landauer's principle in computation and chemistry, Journal of Theoretical Biology 252 (2008) 213–220.
[38] Sornette, D., (2006). *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-organization and Disorder: Concepts and Tools*. Springer Series in Synergetics (2nd ed.). Heidelberg: Springer.
[39] Tacchella A., Cristelli M., Caldarelli G., Gabrielli A., Pietronero, L. A New Metrics for Countries' Fitness and Products' Complexity. Sci. Rep. 2, 723, 2012; doi:10.1038/srep00723.
[40] Tsallis, C., When may a system be referred to as complex?—an entropic perspective. Frontiers in Complex Systems , 1, 2023. DOI:10.3389/fcpxs.2023.1284458
[41] Victor P., Questioning economic growth, Nature 468, 370–371, 2010, doi:10.1038/468370a.
[42] von Thünen J.H., *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalokonomie*, Perthes, Hamburg, 1826.
[43] Xie J., Padoa-Schioppa C., Neuronal remapping and circuit persistence in economic decisions, Nature neuroscience 2016, doi:10.1038/nn.4300.
[44] Yukalov V.I., Sornette D., Self-organization in complex systems as decision making, Advances in Complex Systems Vol. 17, No. 03n04, 2014.
[45] Zipf G.K., *Human Behavior and the Principle of Least Effort*, Addison Wesley, Cambridge, MA., 1949.

Paolo Giordano, Faculty of Mathematics, University of Vienna, AT.