

RESEARCH ARTICLE

A constructive approach to the epistemological problem of emergence in complex systems

Alberto Pascual-García ^{1,2*}

1 Centro de Biología Molecular “Severo Ochoa” CSIC-Universidad Autónoma de Madrid, Madrid, Spain, **2** Department of Life Sciences Imperial College London, Silwood Park, Ascot, United Kingdom

✉ Current address: Institute of Integrative Biology, ETH-Zürich, Zürich, Switzerland

* alberto.pascual.garcia@gmail.com, alberto.pascual@env.ethz.ch



 OPEN ACCESS

Citation: Pascual-García A (2018) A constructive approach to the epistemological problem of emergence in complex systems. PLoS ONE 13 (10): e0206489. <https://doi.org/10.1371/journal.pone.0206489>

Editor: Lawrence M Ward, University of British Columbia, CANADA

Received: September 4, 2017

Accepted: October 15, 2018

Published: October 30, 2018

Copyright: © 2018 Alberto Pascual-García. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: APG’s work was partially supported by fellowship BFU2012-40020 of the Ministerio de Ciencia e Innovación (Spanish Government) (<http://www.ciencia.gob.es/portal/site/MICINN/>) and by grant 542381 from the Simons Foundation (www.simonsfoundation.org), awarded to Sebastian Bonhoeffer. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis,

Abstract

Emergent patterns in complex systems are related with many intriguing phenomena in modern science. One question that has sparked vigorous debates is if difficulties in the modeling of emergent behaviours are a consequence of ontological or epistemological limitations. To elucidate this question, we propose a novel approximation through constructive logic. Under this framework, experimental measurements will be considered conceptual building blocks from which we aim to achieve a description of the microstates ensemble mapping the macroscopic emergent observation. This procedure allow us to have full control of any information loss, thus making the analysis of different systems fairly comparable. In particular, we aim to look for compact descriptions of the constraints underlying a dynamical system, as a necessary *a priori* step to develop explanatory (mechanistic) models. We apply our proposal to a synthetic system to show that the number and scope of the system’s constraints hinder our ability to build compact descriptions, being those systems under global constraints a limiting case in which such a description is unreachable. This result clearly links the epistemological limits of the framework selected with an ontological feature of the system, leading us to propose a definition of emergence strength which we make compatible with the scientific method through the active intervention of the observer on the system, following the spirit of Granger causality. We think that our approximation clarifies previous discrepancies found in the literature, reconciles distinct attempts to classify emergent processes, and paves the way to understand other challenging concepts such as downward causation.

Introduction

“What urges you on and arouses your, you wisest of men, do you call it will to truth? Will to the conceivability of all being: that is what I call your will! You first want to make all being conceivable: for, with a healthy mistrust, you doubt whether it is in fact conceivable. But it

decision to publish, or preparation of the manuscript.

Competing interests: The author has declared that no competing interests exist.

must bend and accommodate itself to you! Thus will your will have it. It must become smooth and subject to the mind as the mind's mirror and reflection."

Friedrich Nietzsche [1]

Scientific modelling is one of the best examples of a human activity fitting the words of Zarathustra: it requires the generation of conceptual representations for processes which frequently depend on uncomfortable features such as measurement inaccuracy, constituents interdependence or complex dynamics. We attempt to incorporate these representations within a mathematical or computational framework, which is nothing but a comfortable place where we reaffirm our confidence in the acquired knowledge. Building a formal framework provides a favourable environment to reach new analytical and computational results, thus accelerating the outcome of new predictions that can be firmly settled within the scientific knowledge after hypothesis testing.

Many interesting challenges in scientific modelling come from complex systems, which are commonly defined as systems composed of a large number of entities driven by non-linear interactions between their components and with the environment. In particular, complex systems may lead to a controversial phenomenon in modern science: the observation of emergent behaviours. Among this kind of collective behaviours we find phenomena such as magnetism, patterns observed in dissipative systems like hurricanes or convection cells or, in biological systems, patterns on animal skins or flocking behaviour. Looking at these examples it seems that the difficulty arises from an apparent discontinuity between the emergent macroscopic properties and their microscopic description. Since a basic tenet in the scientific method is that macroscopic properties are the consequence of the lower level constituents, a critical question arises here [2]: How is possible to obtain a satisfactory conceptual representation of emergent macroscopic behaviours when the definition of emergence apparently implies a discontinuity between the microscopic and the macroscopic representation?

Explaining the origin of this discontinuity has led to the famous controversy between the vitalist and reductionist approaches [3, 4]. Our first aim is not to discuss what an emergent property *is* or if it is possible to *generate* an emergent property—e.g. through a simulation or experiment. We rather wonder which are the features that a complex system exhibiting an emergent behaviour may have for being more or less accessible to an explanation of that behaviour by applying the scientific method. Nevertheless, the term *explanation* is central in the above discrepancy, as it is clearly apparent in this phrase summarizing the so called *emergentist* position, seen by Kim as a meeting point in the controversy: Emergent properties are identified when we observe that a complex system “begins to exhibit genuinely novel properties that *at a first sight* are irreducible to, and neither predictable nor explainable in terms of, the properties of their constituents” [5]. Although this statement brings a good intuition on when an emergent behaviour is identified, we added the words “at a first sight” to emphasize our discrepancy: although we agree in that emergent properties are not just a simple extrapolation of lower level properties –thus rejecting a simplistic compositional physicalism– this does not necessarily mean that emergent behaviours cannot be *explained* from a microscopic description and, why not, eventually *predicted*. At least as scientists, unless strong experimental and theoretical support reject this possibility, we prefer to remain agnostic. Following the view of Francis Crick, a natural starting point for a scientist is the decomposition of the system and we expect that “. . . its behaviour can, at least in principle, be understood from the nature and behaviour of its parts plus the knowledge of how all these parts interact” (reproduced from [4] referring to [6] p.11), i.e. this perspective admits an explanatory physicalism, and in the following we will try to precisely define what is understood here as *explanatory*.

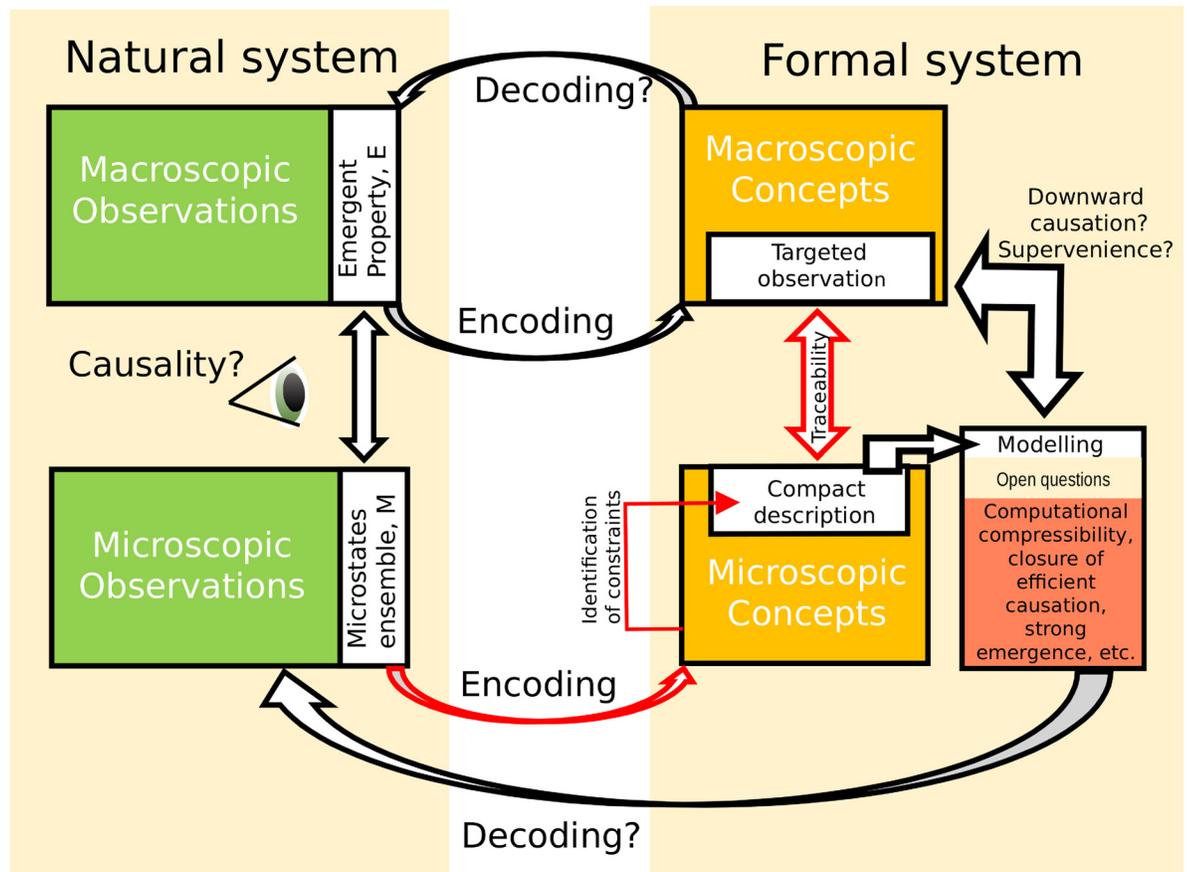


Fig 1. Scheme of the epistemological approximation of an observer in the analysis of emergent properties. Research starts with the observation of an emergent macroscopic property E and the associated ensemble of microstates M , both characterized through measurements. These observations are then encoded to build a formal framework. In this work, we are interested in finding a *minimal epistemological map* between the formal macroscopic and microscopic descriptions (red arrows). This map consists of finding a microscopic compact description of the ensemble, which is achieved through the identification of microscopic constraints in the dynamics of the system, in which case we will say that the macroscopic property is traceable. This is a necessary *a priori* step before a model is built (box modelling) that will allow the researcher to address other questions (see questions marks) such as supervenience of upper and lower level constituents or how to decode the formal framework under the scientific method through experiments. Interestingly, although we focus in a very specific epistemological process, and we do not explicitly address questions related with modelling or the ontology of emergent properties, the analysis of microscopic constraints sheds light over these questions.

<https://doi.org/10.1371/journal.pone.0206489.g001>

In the above emergentist definition one should distinguish between *theoretical prediction* of the features of an emergent property E and its *inductive prediction* [5]. Inductive prediction occurs if, every time that the property E is observed, a particular ensemble of microstates M (see Fig 1) is also observed. The fact that this situation can be considered a phenomenological law justifies us to say that understanding what comprises the interrelations of the components of the system and the influence of the environment that leads to the observation of M , already constitutes an explanation of E . Still we may argue that this explanation is incomplete if we do not understand how visiting M translates into the appearance of E . Or if we are not able to *theoretically* predict the emergence of E when we operate over the system to make it run into M , formally through simulations or, in the natural side, decoding the formalism with experiments such that we intend to control the system to generate M , and hence to observe E . But we will defend here that it is the first step that should be tackled in the search for an explanatory framework of E under the scientific method.

Following this attitude, we wonder which are the properties that a complex system (and, in particular, those exhibiting emergent behaviours) may have for being more or less accessible to our knowledge by applying the scientific method, and we address this question by focusing on the analysis of the encoded description of M . In this sense, we align with the proposal of de Haan [7], who highlights the necessity of a general epistemological framework in which emergence can be addressed, and we need to incorporate in this framework arguments to determine what should be understood as epistemologically accessible.

An interesting condition for epistemological accessibility was proposed by Bedau when he coined the term *weak emergence* for those emergent processes that are epistemologically accessible only by simulation [8]. The idea is that a simulation would demonstrate the *supervenience* of upper level properties from lower level constituents, even if the mechanistic process leading to the observed pattern is not completely understood, i.e. it is not possible to compress the simulation into a compact set of rules explaining how the outcome is determined. Therefore, it provides an objective definition of emergence based on computational incompressibility, that has been explored by different models such as cellular automata [9, 10] or genetic algorithms [11], approximations that were later called *computational emergence* [12].

Nevertheless, Bedau also pointed out that if it is not possible from the simulations to recover regularities that may lead us to describe the system under compact laws, the above computational approximations should be understood as non-epistemic, hence being useless in deciphering the principles governing the emergence of a property [8]. Furthermore, Huneman rightly emphasized that this is an important question to understand the relationship between computational models and processes observed in nature [13] and, for those computational models depicting regularities in their global behaviour, coined the term *robustly emergent* models.

There is a last notion of emergence we would like to discuss that has been considered fundamental –as opposed to epistemological–, which is called *strong emergence* [14]. In physics it is frequently assumed that knowing the positions and velocities of particles is sufficient to determine the pairwise interactions. This assumption is frequently found in physics-inspired models of collective behaviour, where individual motion results from averaging the pairwise responses to each neighbour. Bar-Yam reasoned, however, that this assertion would not hold if the system is embedded in responsive media –such as the motions of impurities embedded in a solid– or in any process where global optimization (instead of local) is involved. In this way, if there is a constraint in the system acting on every component simultaneously and it is strong enough –i.e. a strong global constraint– it is not possible to determine the state of the system considering only pairwise interactions. In a sense, the parts are determined *downwards* from the state of the whole, with consciousness being a paradigmatic example suggested for strong emergence [15].

In this paper, we investigate which are the conditions that a natural system depicting an emergent property may possess to be compatible with the notion of robust emergence. To address this task, we will shift the attention from computational models to focus on the analysis of experimental data (see Fig 1). We will follow the view in which emergence is considered a “relation between descriptions of models of natural systems and not between properties of an objective reality in itself” [16] although, as we will see immediately, we will not renounce to talk about ontology. In particular, we aim to understand the relation between macroscopic descriptions of emergent behaviours and their microscopic counterparts, a map similar to the one proposed by Kim [5]. This map has been criticized by Mitchell saying that it is static, and that it cannot account for the dynamical nature of systems depicting emergent behaviours [17]. But here we show that the ability to find what we will call a *compact* description of the microstates’ ensemble depends indeed on the dynamical features of the system and, in

particular, on how the components are entangled through internal and external constraints. In this way, the minimal explanation we look for resides in the description of the microstates' constraints, and the epistemological accessibility will be related with our ability to find and compress such a description. Our strategy starts from experimentally characterized microscopic states associated to a macroscopic emergent observation, and investigates which kind of regularities found in these microstates are more difficult to compress and why. We believe that this is a necessary *a priori* step for any approach aiming to model the observed process, and thus scientifically sound notions of emergence should focus first on this process.

An immediate risk in this endeavour is that we must select a framework to work with and, by doing this, any result may depend on the framework selected. To circumvent this difficulty, we will apply constructive logic, which is both predicative and intuitionistic [18]. Since we aim to investigate the concept of emergence within a scientific setting, we will focus on how formal models are built starting from experimental measurements –that are considered basic characteristics– following a purely constructive attitude. This will allow us to monitor any loss of information [18], which we will see is particularly important to understand the difficulties in the formalization of emergence. In particular, the analysis of different synthetic examples with the rather limited repertoire of logical operations considered, will allow us to establish fair comparisons between systems, and to investigate further the ontological origins of these difficulties.

The article is articulated as follows. In Methods we introduce the formalism. We provide a careful definition of the system, and then introduce the different mathematical tools needed. Then we describe our epistemological approximation to the problem, explaining how a number of key concepts such as *constraint* or *model* should be understood within this framework. Then we explain the general procedure followed to analyse the systems, in which logical disjunction has a prominent role.

In Results, we analyse synthetic examples of 3-bit systems, to illustrate how the number and scope of the constraints hinder our ability to describe the systems. We next discuss the limitations of our approach, and how it is possible to elucidate which is the complexity of the constraints underlying the system under analysis despite these limitations, through the active *intervention* of the observer on the system. This procedure will lead us to propose a definition of emergence strength that we believe reconciles different conceptions of emergence found in the literature.

Methods

Main definitions and operations

System definition. We start by proposing a glossary of terms concerning the system definition, some of them close to those proposed by Ryan in [19]. We will call a set of basic quantities associated with a given entity an (object of) observation o_i . Each of these quantities is a function f of the Cartesian product of a finite collection of sets –at least one of which is determined by experimental measurement– into real numbers \mathbb{R} , i.e. $f : A \times B \times \dots \times P \rightarrow \mathbb{R}$.

The non-measurable sets may refer to a set of measurement units (e.g. grams, meters), to a set of reference frameworks, or to any other set necessary to determine the final quantity. For simplicity, we will consider that any variation in magnitude is a consequence of a variation in the outcome of a measurement and thus, in the following, we will refer to measurements when the quantities describing objects of observation are discussed.

Hence, a system is characterized by a collection of M quantitative and/or qualitative (i.e. binary) magnitudes $X = \{x_k, k = 1, \dots, M\}$. Given that we are interested in complex systems, we will consider that the system consists of a large number of entities, that we denote N . We will

call this selection of objects *scope*, whose size, N , implicitly determines the spatio-temporal boundaries of the system. Determining the scope is already a difficult task for large complex systems. Difficulties arise, on the one hand, from the identification of these entities because, when their number is large, a complete characterization may be unfeasible. On the other hand, it will also be difficult to define the separation between system and environment, as oftentimes this separation cannot be achieved using strictly objective arguments [20, 21].

The selected variables X are intended to be sufficient to answer the questions addressed. For simplicity, we start by considering an ideal scenario in which all of these variables can be quantified for any entity within the system, leading to $N \times M$ specific values. In order to discuss how the complexity of a system can be explored using the scientific method, we will propose a procedure to relax this assumption in section.

Every variable x_k has a *precision* r_k which is its finest interval of variation and may be determined by different means. For instance, the precision may be limited by the intrinsic measurement error, which would be an ontological limitation. Another possibility arises when the expected influence of a given variable on the system's description is small for a given shift in value, and a coarser discretization is then justified (a methodological choice). Calling I_k the interval of viable values of x_k , the number of possible values considered for this variable will be $\zeta_k = I_k/r_k$. Note that ζ_k can be seen as the resolution of the variable x_k , and thus we will call resolution R of the system the finest description that allows us to distinguish two of its states $R = \max_k(\{\zeta_k\})$ ($k = 1, \dots, M$).

This choice of variables together with the set of viable values will be called the *focus* F of the knowing subject, upper-bounded by $F \sim M \times R$. We finally call the set of specific values $\{N, M, R\}$ the *scale*. A factor multiplying any of these values represents a change in the scale of the scope (if N is modified), or the focus (M or R). Note that, according to this definition, the scale is an ontological attribute as determined by N , but it also depends on epistemological ones, determined by M and R . Therefore, the breadth of the focus is heavily influenced by epistemological choices. Interestingly, it has been claimed that emergent behaviours (emergent theories) are the consequence of a change in the scope [19] (and not in the focus [22]).

Measurable properties and the definition of concepts. Let us start introducing some definitions, most of them already provided and justified in [23], that we recover here for completeness. For the sake of simplicity we will start considering that our objects of observation $o \in O$ are the components of a complex system at a given time, i.e. we focus on a single microstate μ with N components, each of them described by M variables. For the moment, each of these components is what we consider an object of observation. We will move later towards a description where each object of observation is a microstate itself, becoming the whole space of objects the observed phase space. All the definitions considered in the following for a single microstate can be extended for other objects with a different scale.

Definition: We call a *basic concept or characteristic* $c_a = x_k^*$ the specific value x_k^* of a variable x_k , out of the ζ_k possible values, measured over an object of observation o . In this way, if we consider two different measurements for the same entity, each of them will constitute a different object of observation.

Definition: We call *focus* F the whole set of characteristics considered by the observer: $F = \{x_k^l; k = 1, \dots, M; l = 1, \dots, \zeta_k\} \equiv \{c_a; a = 1, \dots, \tilde{M}\}$, with $\tilde{M} = \sum_{k=1}^M \zeta_k$. We make explicit here the discrete nature of the conceptual setting and the relation between resolution and focus, which achieves a suitable description in terms of characteristics, leading to the definition of concepts. Discreteness is ultimately a consequence of the non-vanishing systematic error associated with any measurement, and any transition into a continuous description would be a formal abstraction made during the modelling process.

Definition: We call a *concept* v any non-empty finite subset of F : $v = \{c_1, \dots, c_p\}$, with $P \leq \bar{M}$. We defined the intension of concepts. Given that a concept may contain a single characteristic, $v = \{c\}$, any characteristic can be considered a (basic) concept as well. Characteristics are atomic concepts in the sense that, for research purposes, they are not further decomposable into other concepts. In this sense, they constitute a basis for any other concept. The distinction between concept and characteristic will be needed in certain circumstances, but given the simplicity of the examples we discuss below, we will rarely use this distinction and refer simply to concepts.

Logical operations. It follows from the previous definitions the operations below to build new concepts.

Definition: (Conjunction of concepts). Let $v_1 = \{c_1, \dots, c_p\}$ and $v_2 = \{d_1, \dots, d_Q\}$ be two concepts. Then, the conjunction of v_1 and v_2 is the concept

$$v_1 \wedge v_2 = \{c_1, \dots, c_p, d_1, \dots, d_Q\} \tag{1}$$

The conjunction of concepts is, in turn, a concept which consists of the set of all characteristics contained in concept v_1 AND v_2 . Although the new concept is the union of two sets of concepts, note the choice of the meet symbol \wedge made in [23] that we also follow here. With this choice it is stressed that there are *fewer* concepts enjoying all the characteristics in the union than concepts enjoying all the characteristics of the finite sets v_1 and v_2 separately (Silvio Valentini, personal communication). The situation is confirmed by the choice of the empty set as the greatest element among collections of characteristics, given that the characteristics it contains (namely, no characteristics) are enjoyed by all concepts.

Now we would like to know the supremum of two concepts, namely the subset of concepts U that allows us to speak of the concept v_1 OR v_2 .

Definition: (Disjunction of concepts). Let $v_1 = \{c_1, \dots, c_p\}$ and $v_2 = \{d_1, \dots, d_Q\}$ be two concepts. Then, the disjunction of v_1 and v_2 is the subset

$$U = v_1 \vee v_2 = \{v_1, v_2\} = \{\{c_1, \dots, c_p\}, \{d_1, \dots, d_Q\}\} \tag{2}$$

This operation is a sensible one because we will use it to describe subsets of observations such as $\{v_1, v_2\}$, and we would like to find a concept for this subset describing either the observation of v_1 OR the observation of v_2 OR the observation of v_1 AND v_2 . Unfortunately, this concept does not generally exist in a closed form, a fact that will be central to our investigation.

The space of concepts described so far, equipped with a join and a meet for every finite subset, fulfils the properties of a distributive lattice [24] (moreover, it is a topology [25], although we will not exploit this fact) and hence we can get some more intuition about disjunction through the following property.

(Distributivity of the disjunction of concepts). Let $v_1 = \{c_1, \dots, c_p, b_1, \dots, b_L\}$ and $v_2 = \{d_1, \dots, d_Q, b_1, \dots, b_L\}$ be two concepts. Then the subset of concepts $U = v_1 \vee v_2$ can be expressed as

$$U = v_1 \vee v_2 = \{b_1, \dots, b_L\} \wedge [\{c_1, \dots, c_p\} \vee \{d_1, \dots, d_Q\}]. \tag{3}$$

Therefore, disjunction of concepts leads to a subset of concepts containing the set of all characteristics in common to both concepts, plus those that are observed either in one or the other. As we already noted, we are unable to build a concept, let us say q , in terms of conjunctions of concepts to talk about U , but it is clear that the more concepts v_1 and v_2 have in common the closer we are to finding such a concept. Other basic properties of conjunction and

disjunction that will be used are the following [24]:

	\wedge	\vee
<i>Commutativity</i>	$v_1 \wedge v_2 = v_2 \wedge v_1$	$v_1 \vee v_2 = v_2 \vee v_1$
<i>Associativity</i>	$(v_1 \wedge v_2) \wedge v_3 = v_1 \wedge (v_2 \wedge v_3)$	$(v_1 \vee v_2) \vee v_3 = v_1 \vee (v_2 \vee v_3)$
<i>Unit laws</i>	$v_1 \wedge \text{true} = v_1$	$v_1 \vee \text{false} = v_1$
<i>Idempotence</i>	$v_1 \wedge v_1 = v_1$	$v_1 \vee v_1 = v_1$
<i>Absorption</i>	$v_1 \wedge (v_1 \vee v_2) = v_1$	$v_1 \vee (v_1 \wedge v_2) = v_1$

The space of objects of observation. We have introduced a *formal* conceptual apparatus that a knowledge subject may use to build from experimental measurements a model, but we still do not know what a model means. Before getting into this discussion, we should establish a map between the formal side in which we developed our framework and the *concrete* space, in which the objects of observation live, i.e. the objects that constitute the physical world. In particular, we are interested in understanding how the set of concepts we considered, which determines a partition of the focus, will induce, in turn, a partition in the set of objects of observation. In other words, we aim to determine a constitutive relationship between any single characteristic belonging to the focus F and the set of objects O . Understanding this map will allow us to work in the formal space while ignoring the concrete space, but being certain that any new concept arising in the formal side has a correspondence in the concrete side. The following constitutive relationship will express how the objects become cognitively significant by means of the characteristics measured and, in turn, by the concepts we build from them [23].

Definition: (Constitution relation). Let F be the focus over a set O of objects. Given $o \in O$ and $v \in F$, we introduce a binary relation, \Vdash , that we call *constitution relation*, such that by $o \Vdash v$ we mean that v is one of the concepts constituting o . With the constitution relation we determine how the objects of observation are expressed via the conceptual apparatus of the knowing subject. In addition, we would like to know which objects are constituted by a given concept.

Definition: (Extension of a concept). Let $v \in F$ be a concept. Then, the extension Ext of v is the subset of objects of O constituted by v , that is

$$\text{Ext}(v) = \{o \in O \mid o \Vdash v\} \tag{4}$$

We note here that an immediate consequence of Eq 4 is that any object of observation has necessarily associated with it a concept, i.e. it is just cognitively accessible by means of the conceptual apparatus of the knowing subject. This assertion, if accepted in general, leads to a Kantian epistemological positioning [25]. In our case, it is a consequence of the fact that our objects of observation are built from measurements of a reproducible experimental setting, and hence it is true by construction. Nevertheless, the opposite is not true, as we may deal with concepts for which no object is observed, i.e. $\text{Ext}(v) = \emptyset$. Such a situation would arise for example if we have *a priori* expectations of the viable values of the system. For instance, we know that a group of birds can fly in any direction even if we systematically observe them flying in a single direction. From the point of view of the scientific method, concepts built from *a priori* expectations are very important, as they may be used to propose null hypotheses which, in general, can be formulated as H_0 : " v observed". It is when we reject the hypothesis through experiments that we acquire a scientific knowledge of the process analysed, i.e. that v is not observed with some statistical significance.

Finally, we aim to know what is the extension of a subset U of concepts $U = \{v_1, \dots, v_L\}$.

Lemma: Let U be a subset of the set F of concepts. Then, the extension of U is defined by setting

$$\text{Ext}(U) = \bigcup_{v \in U} \text{Ext}(v) \tag{5}$$

Hence, if we consider two concepts v_1 and v_2 , we should not confuse the extension of a concept built by conjunction of concepts $v = v_1 \wedge v_2 = \{c_1, \dots, c_P\} \wedge \{d_1, \dots, d_Q\} = \{c_1, \dots, c_P, d_1, \dots, d_Q\}$ with the extension of a subset of concepts built by disjunction $U = \{v_1, v_2\} = v_1 \vee v_2 = \{\{c_1, \dots, c_P\}, \{d_1, \dots, d_Q\}\}$. In the former case, we look for objects containing both concepts and, thus, the number of these objects is smaller than or equal to the number of objects described by v_1 or v_2 , $\text{Ext}(v) = \text{Ext}(v_1 \wedge v_2) = \text{Ext}(v_1) \cap \text{Ext}(v_2)$. On the other hand, the subset U extends over objects containing *any* of the concepts, since its extension is the union of the extension of both concepts, as shown in Eq 5.

Epistemological approximation

Microscopic and macroscopic descriptions. Following the definitions introduced, we would like now to differentiate between two types of variables providing a description of the system at different scales: microscopic and macroscopic variables. Note that by microscopic we do not mean “atomistic” but instead a significantly shorter spatio-temporal scale of observation of the system. A particular feature of the interplay between both scales is that, when a macroscopic property is observed during the dynamical evolution of a system, even if the microscopic variables are continuously changing, the macroscopic variable’s values remain the same.

In the following, we will call microstate μ a vector containing, at a given time, the values of a set of variables $\{x_k\}$ that fully determines the state of the microscopic objects, i.e. $\mu = \{x_k^*\}$, where x_k^* stands for a particular value of the variable x_k . Therefore, the basic objects of observation we are considering now are the microstates $o \equiv \mu$. When a coarse-graining of the microstates at the spatial, temporal, or spatiotemporal dimensions is performed, it may be possible to determine macroscopic variables y_k describing the state of the system at the new (coarser) scale. We will call macrostates the objects described at the macro scale $o \equiv \xi$. In some cases, the macroscopic variables y_k can be obtained by applying a surjective map f over the microscopic variables $f(x_k) \rightarrow y_k$. For instance, if we deal with an incomplete (statistical) microscopic description of an ensemble, $P(\mu)$, we can obtain a coarse determination of a macroscopic variable y_k averaging the correspondent microscopic variable x_k , weighted by the statistical probability of the microstates over the ensemble $\langle x_k(\mu)P(\mu) \rangle$. Nevertheless, in many other situations it is not possible to find such a map, and we argue that this fact underlies many problems surrounding the study of emergent properties: we observe a macroscopic property such as the collective behaviour of many interacting elements, and it does not seem possible to explain it from lower levels of description (for instance, from the properties of the entities themselves).

It is important to underline that macrostate and microstate definitions are relative to the scale of observation and they may change if we move from one scale of description to another. Consider a system described within a certain temporal scale by a set of microstates $\{\mu_i\}$ which are associated with the observation of a single macrostate ξ . Assume now that the system evolves along a sufficiently long path such that we observe different macrostates and we store T snapshots of these dynamics, leading to an ensemble of macrostates $\{\xi_u\}_{u=1}^T$. If at this larger scale new properties arise, we may be interested in considering that each of these macrostates is now a microstate $\hat{\mu}$ for a new system with a larger scope and lower resolution $\xi_u \rightarrow \hat{\mu}_i$.

Given that the scope of a macrostate will always be larger than that of a microstate ($N_\xi \geq N_\mu$), whereas the opposite is true for the resolution ($R_\xi \leq R_\mu$), in this exercise we have increased the scope and reduced the resolution. This is the reason why, the larger the scale, the more difficult it is to build a bottom-up explanatory framework.

We should also note that this change in scale requires an effort to reduce the system description, but this kind of reduction has been performed from the very first step: for the definition of scope, we have neglected entities; for the definition of focus, we have neglected variables and probably restricted their viable values assuming a lower resolution. Furthermore, any map between microstates and macrostates again considers a reduction in the information provided by the microstates. In general, for both very broad or very detailed questions the technical complexity increases and a reduction in description is unavoidable, and it is important to remark that this exercise is not linked with a *naive* reductionist positioning, in which it is accepted that any macroscopic description is a simple extrapolation of the properties of the microscopic description [26]. Instead, we accept that in complex systems there are discontinuities between the different levels of description and that, for each new level, new properties may arise. We are interested here in investigating when a microscopic description is a minimal representation of an emergent macroscopic observation.

Traceability, compact descriptions and models. Given the above considerations, we propose two formal definitions that will be helpful for understanding our rationale behind the further development of the paper. Before getting into the first definition, we should clarify what is understood as constraints of the system, which are considered here simply as restrictions in the viable values of the variables we handle [27]. Any system is constrained to some extent. But there are some constraints that belong to the definition of the system itself, that we will call *intrinsic*, and others that depend on particular conditions, that we will call *facultative*. Consider the structural differences between a protein and a heteropolymer whose sequence is the result of random shuffling of the protein sequence. Both chains of amino acids have the same number of intrinsic constraints (those derived from the existence of peptidic bonds), but a protein structure requires three additional constraint levels in the interactions between its amino acids that allow the polymer to behave as a protein: the first is needed for it to be kinetically foldable in a biologically relevant time, the second for making the fold thermodynamically stable under physiological conditions, and the third to perform its specific function (metal-binding, phosphorylation, etc.). Both chains have the same amino acids but evolution has selected for a specific order in the sequence that generates the constraints needed for the emergent property (the protein function) to arise. Quantitatively, the probability that these constraints should appear by chance is quite low: the number of possible heteropolymers of length N , considering an alphabet of 20 amino acids, is 20^N and, as a reference, the number of protein structures (of *any* length) deposited to date in the Protein Data Bank (although far from being complete) is only $\sim 1.2 \times 10^5$ (www.pdb.org).

Calling $\{\mu_E\}$ the set of microstates visited when an emergent behaviour is observed, $\{\mu_{NE}\}$ the set of microstates not visited, and Ω the whole phase space, we expect that $\#(\Omega) \approx \#(\{\mu_{NE}\}) \gg \#(\{\mu_E\})$; where $\#(\{\cdot\})$ is the cardinality of the set $\{\cdot\}$, i.e. the number of elements it contains. Therefore, we expect the volume of the region of the phase space where an emergent property arises to be much smaller than the whole phase space. This is probably why unpredictability or surprise are attributes frequently used to describe emergent properties.

Let us now define what we consider a macroscopic property, whose identification is typically the starting point of any research.

Definition: (Target macroscopic property). We will say that an observed macroscopic property is a target (of scientific research) if it is only observed when certain microscopic facultative constraints are present. With this definition we state a minimal condition for

considering that a macroscopic property is susceptible of being analysed through the scientific method, an assumption that also applies when an emergent property is interrogated. Therefore, although it is an epistemological condition, it implies an ontological assumption about emergent properties, namely that these are the consequence of a microscopic behaviour running under facultative constraints.

Hence, we consider situations in which the phase space of the system Ω is restricted to a smaller observed region $\Omega^O \subset \Omega$, and we say that there exists a (perhaps novel) macroscopic concept \hat{c} such that $\text{Ext}(\hat{c}) = \Omega^O$. This fact motivates the exploration of this region both in terms of macroscopic $\{\hat{c}\}$ and microscopic $\{c\}$ concepts (see Fig 1). We now introduce a condition that allows us to consider that a macroscopic description is in correspondence with a microscopic description.

Definition: (Traceability). Given a target macroscopic property \hat{c} and the observed phase space Ω^O associated with that property, i.e. $\text{Ext}(\hat{c}) = \Omega^O$, we will say that the macroscopic description obtained is traceable if we find an appropriate function or algorithm applied to microscopic properties $f: \{c\} \rightarrow q$ such that the new concept q derived compactly describes the ensemble of microstates, i.e. $\text{Ext}(q) = \text{Ext}(\hat{c}) = \Omega^O$. This definition paves the way for quantifying the correspondence between both descriptions within the proposed framework. Note it does not mean that we should be able to thoroughly *explain* the macroscopic properties from the microscopic properties. It is a weaker condition that only requires establishing a correspondence between microscopic and macroscopic variables describing the same region of the observed phase space Ω^O . In Fig 1 we make explicit this distinction showing how traceability appears as a relation between representations (double red arrows), while investigating causal relationships requires us to build models from the starting descriptions and to perform experiments (white arrow to modelling box). Therefore, traceability can be seen as a rather minimal epistemological condition that allows us to talk about emergent properties circumventing any discontinuity between both descriptions.

Let us think further about the existence of concepts q in the context of complex systems. After identifying a set of microstates μ visited when an emergent property is observed, we may be able to characterize this set as described above, through a set of concepts $\{e_i / \text{Ext}(e_i) = \mu_i, \forall i\}$. Given the intrinsic dynamical nature of these systems, a natural starting point in the search for a formal description will be given by $q = e_1 \vee \dots \vee e_L$, L being the number of microstates. Note, however, that the concept q simply presents the microstates but it does not provide any insight into the mechanistic processes underlying the observation of these microstates and no others. In the remainder of the article, we will show that to provide such mechanistic understanding it is necessary for q to constitute a description of the constraints of the system. Furthermore, the chances of building a mechanistic model will be related with our ability to build a compact description.

Definition: (Compact description) We say that a set of distinguishable microstates $\{\mu\}$, namely a set in which every microstate μ_i is unambiguously described by a concept $e_i \in F$, F being the microscopic focus, is compactly described by a concept q , if $\text{Ext}(q) = \{\mu\}$ and it *irreversibly* describes the constraints existing in the system. By irreversibility it is meant here that, even if q is analytically derived from the starting concepts e_i and it is still a definition of the system, and thus $\text{Ext}(q) = \text{Ext}(\{e_i\})$, we lose information in the derivation in such a way that it is impossible to reverse the operations and retrieve back the original states. To recover the full description of the system it will be needed to create a model.

Definition: (Model) A computable function or algorithm is a model Ψ if, given a compact description q of a set of microstates $\{\mu\}$, it interprets the constraints encoded in q and generates the description of the microstates from which the compact description was

derived, i.e. $\Psi(q) = \{e\}$. In summary, we are interested in investigating the traceability of complex systems through compact descriptions describing the constraints of the system, which may allow us to build mechanistic models to test hypotheses about the accuracy of the description found. Note that, if the description of the constraints is incomplete or incorrect, a model will generate a set of microstates that would not perfectly match the observed microstates (if it is able to match them at all), a fact that will be exploited in our proposal to quantify emergence.

The process we describe in the following is similar in spirit to Rosen’s attitude when he said that, in the modelling process, instead of throwing away the organization and keeping the matter (a compositional physicalism) one should throw away the matter and keep the organization (see Chap 5E in [28]). While looking for constraints we shed light on the entailment between the components of the system. Still, to prove that the constraints found correspond with the observed microscopic process, we should use a generative model to show that the original microstates are recovered. The condition of irreversibility tries to avoid an impredicative definition of the system’s constraints: we aim to avoid any definition of the system’s constraints that consists of the mere presentation of the system itself.

Results

Identification of constraints: Focusing on disjunction

As we anticipated, when an emergent property is observed the probability distribution of the values of one or several variables depart from the distribution observed when the system is free of constraints, thus losing ergodicity [29] (p. 186). Therefore, the existence of facultative external or internal constraints limit the behaviour of the system and, as we will attempt to clarify, a necessary condition to determine a microscopic property associated to every microstate visited requires the determination of the existing constraints. Indeed, it has been claimed that the reduction in the degrees of freedom of the system –a phenomenon for which it has been coined the term *dissolvence*– relies at the basis of the formation of complex emergent structures [30]. We will show here that the nature of the different constraints acting on the system determine its epistemological accessibility, and hence our ability to reach a satisfactory explanation of emergent behaviours.

Firstly, we show how the extension of those concepts built through binary operations over sets of concepts can be obtained. Given a new concept α built via conjunction of two concepts, v_1 and v_2 , i.e. $\alpha = v_1 \wedge v_2$, its extension is $Ext(\alpha) = Ext(v_1) \cap Ext(v_2)$. With conjunction we reduce the scope of the description but we refine it, being therefore the basic operation needed to build sharp descriptions of the observed objects.

Let us take as an example the description of polymers of amino-acids $\{o_\alpha\}$. Each amino-acid is described by two quantities that we will consider basic characteristics: the position in the sequence and its specific amino-acid identity. In our framework, an example of a concept describing an amino-acid is $v_i = \text{“cysteine in position } i\text{”}$ which, in turn, is built by conjunction of the characteristics *“cysteine”* and *“i”*. The whole polymer o_α will be described by a concept α , built by conjunction of a set of concepts v_i , i.e. $\alpha = (\lambda_1 \wedge \tau_2 \wedge \dots \wedge v_N)$ (see Fig 2). The sequence becomes uniquely determined with the concept α because its extension exactly maps the polymer under study: $Ext(\lambda_1 \wedge \tau_2 \wedge \dots \wedge v_N) = Ext(\alpha) = o_\alpha$. In summary, conjunction underlies bottom-up approximations, for which we look for sharp descriptions adding up concepts.

Let’s now consider disjunction. Given two polymer sequences α_1 and α_2 (see Fig 2), described as $\alpha_1 = (\lambda_1 \wedge \delta_2 \wedge \delta_3 \wedge v_4 \wedge \tau_5)$ and $\alpha_2 = (\tau_1 \wedge \lambda_2 \wedge \delta_3 \wedge v_4 \wedge v_5)$, we would like to look for a concept α_{12} that sharply describes both sequences and no others. Intuitively, such a concept might be found if we can build a concept through conjunction of those concepts

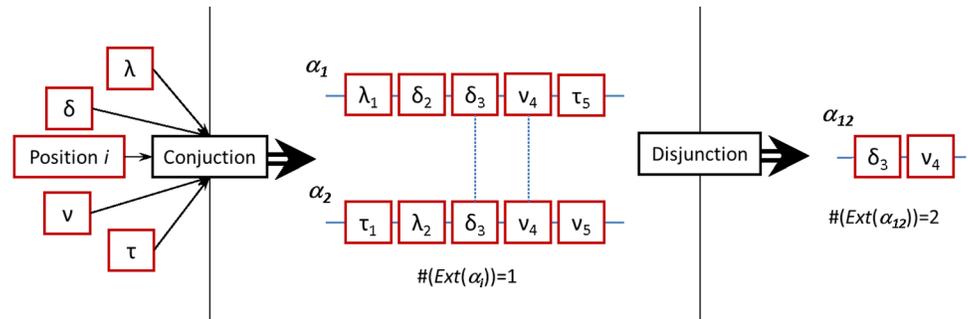


Fig 2. Illustration of conjunction and disjunction of concepts. Starting from the knowing subject’s conceptual apparatus (Greek letters, left), two sequences α_1 and α_2 are built through conjunction of the basic concepts, being themselves concepts (center). These sequences uniquely determine single objects, for instance protein sequences, and thus $\#(\text{Ext}(\alpha_i)) = 1$. Comparing both sequences we observe two common concepts (linked by dotted lines) that we identify through binary disjunction. If these two concepts are only found at these sequences, we can say that a new concept α_{12} sharply describes these sequences (right). This concept contains less basic concepts, but the extension is larger than the original sequences, i.e. $\#(\text{Ext}(\alpha_{12})) = 2$.

<https://doi.org/10.1371/journal.pone.0206489.g002>

shared by both sequences. Following the example, by using Eq 3 we obtain:

$$\alpha_1 \vee \alpha_2 = (\delta_3 \wedge v_4) \wedge [(\lambda_1 \wedge \delta_2 \wedge \tau_5) \vee (\tau_1 \wedge \lambda_2 \wedge v_5)], \tag{6}$$

what it immediately highlights is that there is a candidate for a common concept describing both sequences, $\alpha_{12} = \delta_3 \wedge v_4$ (see Fig 2). However, finding a concept describing common information is a necessary but not sufficient condition, because we should also guarantee that α_{12} is only found in these sequences in order to safely separate them from any other, what is not true in general.

In which situations can be guaranteed that a concept describing a set of objects exists? We can get some intuition with a simple example. Let’s consider two objects o_1 and o_2 sharply described by the concepts $\alpha_1 = v_1 \wedge v_2$ and $\alpha_2 = v_1 \wedge v_3$, meaning that $\text{Ext}(\alpha_1) = o_1$ and $\text{Ext}(\alpha_2) = o_2$. Distributivity of disjunction leads to

$$\alpha_1 \vee \alpha_2 = v_1 \wedge (v_2 \vee v_3). \tag{7}$$

which tell us that, to separate these objects from any other, a limited number of situations, illustrated in Fig 3, can occur:

1. Case 1. If $\text{Ext}(\alpha_1) \subseteq \text{Ext}(v_2)$ and $\text{Ext}(\alpha_2) \subseteq \text{Ext}(v_3)$, then, necessarily $\text{Ext}(v_1) = \text{Ext}(\alpha_1 \vee \alpha_2)$ and $\alpha_{12} = v_1$ constitutes a sharp description of both concepts.
2. Case 2. If $\text{Ext}(\alpha_1 \vee \alpha_2) \subset \text{Ext}(v_1)$ then $\text{Ext}(\alpha_1) = \text{Ext}(v_2)$ and $\text{Ext}(\alpha_2) = \text{Ext}(v_3)$ must hold, what leads to define the concept $\alpha_{12} = v_2 \vee v_3$.
3. Case 3. If $\text{Ext}(\alpha_1) \neq \text{Ext}(v_2)$ and $\text{Ext}(\alpha_2) \neq \text{Ext}(v_3)$, and $\text{Ext}(\alpha_1 \vee \alpha_2) \subset \text{Ext}(v_1)$ and we can only consider the possibility that $\alpha_{12} = v_1 \wedge (v_2 \vee v_3)$ to be a sharp description.

Independently of the situation found, with disjunction we are able to observe the constraints existing in the system. In the first two examples we can talk about the two sequences saying that are those with v_1 in the first position (first case) or with either v_2 or v_3 in the second position (second case), highlighting the restrictions to other values. The third example, which is a combination of both, requires more information to be expressed.

We argue that, in complex systems, the third situation is rather the rule because variables are interlinked through internal and external constraints. And we could expect that the larger

Sequence	Case 1		Case 2		Case 3	
...	Grey	Cyan	Red	Grey	Grey	Cyan
α_i	Grey	Grey	Red	Grey	Red	Grey
α_1	Red	Cyan	Red	Cyan	Red	Cyan
α_2	Red	Green	Red	Green	Red	Green
α_j	Grey	Green	Red	Grey	Grey	Green
...	Grey	Grey	Red	Grey	Red	Grey

Fig 3. Illustration of disjunction of two concepts. We consider two concepts α_1 and α_2 , sharply defined by conjunction of red (v_1 in the Main Text) and cyan (v_2) for α_1 , and by conjunction of red and green (v_3) for α_2 . The grey concepts denote any other colour needed to sharply determine the other α_i concepts. Since these three colours suffice to sharply describe α_1 and α_2 , their extensions over the different objects of observation must follow one of the three general cases shown in the figure and described in the Main Text.

<https://doi.org/10.1371/journal.pone.0206489.g003>

the number of objects and properties are, the more complex the expressions are. Under some circumstances, too complex expressions might reflect an imperfect state of knowledge that do not allow us to find more compact descriptions of the system. To avoid this situation, we will deal with a system in which the whole phase space is known in the next sections, and we will show in this way that it is the number and scope of the constraints what makes the description of the system's constraints more or less complex.

A synthetic example: The three bits system

We are already equipped with the necessary tools to analyse a synthetic example in detail. The toy model we consider consists of a system of three entities whose physical states are described by a single binary variable, i.e. a system modelled with three bits. Examples of this kind of systems may be sets of genes that are expressed (silent) when the amount of the correspondent protein is above (below) a certain threshold, species that are observed (absent) in certain environmental sample or the attractor of a boolean network. Each measurement performed over these entities will be considered an observation, taking a value of one or zero. Note that variable distinguishability is relevant, but the order in which the information is collected is unimportant. The reason is that these are not time-ordered strings but rather states of a system, and we assume that a state does not change within the scale of time needed to characterize it through measurements.

For a system composed by three binary entities the focus is

- $c_1 = 'ON \text{ at object } 1'; d_1 = 'OFF \text{ at object } 1';$
- $c_2 = 'ON \text{ at object } 2'; d_2 = 'OFF \text{ at object } 2';$
- $c_3 = 'ON \text{ at object } 3'; d_3 = 'OFF \text{ at object } 3';$

and, with this focus, we can potentially observe $2^3 = 8$ microstates $\mu_k = (x_1, x_2, x_3)$ (with $k = 1, \dots, 8$; and $x_i = \{0, 1\}$; see Table 1).

Each microstate is defined in terms of this focus through concepts e_k ($k = 1, \dots, 8$) built by conjunction of characteristics. For instance, the microstate $\mu_7 = (1, 1, 0)$ is defined in terms of the basic characteristics as $e_7 = c_1 \wedge c_2 \wedge d_3$, in turn being a concept. We can define for the

Table 1. Microstates of a 3-bit system.

Microstate	
$\mu_1 = (1, 1, 1)$	$\mu_5 = (0, 1, 1)$
$\mu_2 = (1, 0, 0)$	$\mu_6 = (1, 0, 1)$
$\mu_3 = (0, 1, 0)$	$\mu_7 = (1, 1, 0)$
$\mu_4 = (0, 0, 1)$	$\mu_8 = (0, 0, 0)$

<https://doi.org/10.1371/journal.pone.0206489.t001>

3-bit system $\binom{N}{n_{ext}}$ possible combinations of constraints involving n_{ext} variables, and thus the number of final microstates will depend on the number of constraints and on their scope, i.e. the number of elements influenced by the constraint. In the following, we consider examples with a different number and type of constraints, all resulting in the same number of microstates (four out of the eight viable states). Hence, these constraints are codified in one bit of information, but we will see that the number of concepts needed to express these constraints can change from system to system.

The simplest macroscopic description associated to the observed ensemble arises if we consider a coarse graining of the microscopic properties such that there is a surjective map between microstates and macrostates: a macroscopic variable takes the value 'ON' if these microstates are visited and 'OFF' otherwise. In this way, only if there is a statistically significant bias towards these microstates we can say that a novel target emergent macroscopic property is observed.

Taking these considerations in mind, we aim to disentangle the microscopic constraints in the system following the present formalism. Given that we build our conceptual setting starting from the basic characteristics (obtained from measurements) and then performing binary logical operations, we expect that the results obtained for the different systems are fairly comparable.

System with a single constraint of scope one (S1). The rationale is the same for the three systems introduced. We consider that there is an observed macroscopic emergent observation \hat{c}_E , and we know the microstates belonging to the associated region of the phase space $\{\mu_E\}$. Then, we analyze the set of microstates looking for its constraints.

The first system we consider is a system where the first bit is constrained to a fixed value (c_1), leading to the observations $\{\mu_1, \mu_2, \mu_6, \mu_7\}$ that we explicitly show in Table 2.

In order to find the system's constraints we start presenting the ensemble through disjunction of the concepts e_i , which sharply describe the different microstates, and then we investigate if it is possible to find a compact description. We first expand the definition of the system:

$$\Omega = e_1 \vee e_2 \vee e_6 \vee e_7 = (c_1 \wedge c_2 \wedge c_3) \vee (c_1 \wedge d_2 \wedge d_3) \vee (c_1 \wedge d_2 \wedge c_3) \vee (c_1 \wedge c_2 \wedge d_3)$$

Table 2. 3-bit microstates of a system with a single constraint of scope one.

$\{\mu_E\}$
$\mu_1 = (1, 1, 1)$
$\mu_2 = (1, 0, 0)$
$\mu_6 = (1, 0, 1)$
$\mu_7 = (1, 1, 0)$

<https://doi.org/10.1371/journal.pone.0206489.t002>

For clarity, we start looking for a compact description analysing the first two terms:

$$\begin{aligned}
 e_1 \vee e_2 &= (c_1 \wedge c_2 \wedge c_3) \vee (c_1 \wedge d_2 \wedge d_3) = c_1 \wedge (c_1 \vee d_2) \wedge (c_1 \vee d_3) \\
 &\quad \wedge (c_2 \vee c_1) \wedge (c_2 \vee d_2) \wedge (c_2 \vee d_3) \\
 &\quad \wedge (c_3 \vee c_1) \wedge (c_3 \vee d_2) \wedge (c_3 \vee d_3) \\
 &= c_1 \wedge (c_3 \vee c_2) \wedge (d_2 \vee d_3) \\
 &= c_1.
 \end{aligned}$$

In the simplification, c_1 absorbed several terms and expressions of the type $(c_i \vee d_i)$ or $(c_3 \vee c_2) \wedge (d_2 \vee d_3)$ are mutually exclusive, hence vanishing. We proceed now similarly to simplify $e_6 \vee e_7$:

$$\begin{aligned}
 e_6 \vee e_7 &= (c_1 \wedge d_2 \wedge c_3) \vee (c_1 \wedge c_2 \wedge d_3) = c_1 \wedge (c_1 \vee c_2) \wedge (c_1 \vee d_3) \\
 &\quad \wedge (d_2 \vee c_1) \wedge (d_2 \vee c_2) \wedge (d_2 \vee d_3) \\
 &\quad \wedge (c_3 \vee c_1) \wedge (c_3 \vee c_2) \wedge (c_3 \vee d_3) \\
 &= c_1 \wedge (c_2 \vee c_3) \wedge (d_2 \vee d_3) \\
 &= c_1.
 \end{aligned}$$

Combining both expressions yields

$$\Omega = e_{v_1} \vee e_2 \vee e_6 \vee e_7 = c_1$$

The result highlights that the system has a single constraint over the first variable, that does not allow us to observe the value d_1 , keeping the other two variables free. And the cost of this reduction is that we have irreversibly lost all the information needed to analytically recover the original concepts describing the system: a model would be needed to interpret this constraint and generate back these states.

To illustrate the relational rather than compositional nature of the approximation, we propose two representations that further allow us to see the interplay between the formal result found and the concrete representation of objects (see Fig 4). The first network represents the concrete side, where each microstate μ_i is linked with another μ_j if the same value is observed in the same variable. The constraints determine the microstates acting on the variables, so we should still move to the formal side to identify them. We create a second network considering now basic concepts as nodes, linking two concepts c_i or d_i if they extend over the same microstates (see Fig 4). More precisely, we link two concepts c_i and c_j with a directed edge if $\text{Ext}(c_i) \subseteq \text{Ext}(c_j)$, and with an undirected edge if $\text{Ext}(c_i) \cap \text{Ext}(c_j) \neq \emptyset$. In this way, we compactly represent all the dependencies present in the system with relationships of subordination (directed edges) cooccurrence (undirected) or exclusion (absent link) between the different values. This representation resembles the information that we would recover if we build a network of variables from a covariance matrix: positive (negative) correlation arises when similar (dissimilar) values are found between two objects.

From the formal network in Fig 4 it is easy to observe that one of the values of the first variable, d_1 , is never observed, a fact that we can express with the proposition:

$$\text{Ext}(d_1) = \emptyset.$$

What this proposition simply states is that, in order to identify that a given microstate belongs to this system, it is necessary to evaluate that the value measured at the first component of the system is different from zero. In other words, as soon as we reject the hypothesis

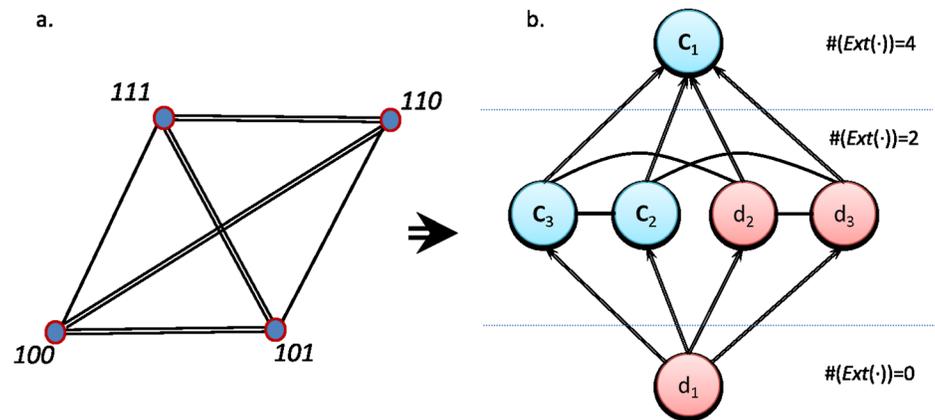


Fig 4. Representations of a three bits system with a single constraint of scope one. (Left) In the concrete network, each node represents a microstate and it is linked with another microstate if they share the same observation for any component, where the number of links represent the number of concepts shared. (Right) Formal network of concepts extracted from the analysis of the microstates. Two links c_i and c_j are linked with a directed edge if $Ext(c_i) \subseteq Ext(c_j)$ and with an undirected link if $Ext(c_i) \cap Ext(c_j) \neq \emptyset$. The concepts are hierarchically ordered according to the cardinality of their extension, i.e. the number of microstates they map. In this example, a single constraint on x_1 naturally arises, as one of its possible values maps the empty set.

<https://doi.org/10.1371/journal.pone.0206489.g004>

H_0 : " d_1 is observed", we will be confident of the fact that we are dealing with a microstate contained in S_1 . Similarly, to propose a generative model of the states, we just need to fix the first variable to one for every state and randomly generate values for the other two variables.

System with two constraints of scope two (S2). We now select four microstates that are the result of imposing two constraints over two pairs of variables, which leads to the set $\{\mu_1, \mu_4, \mu_5, \mu_8\}$ (see Table 3).

We look for a compact description following the same reasoning as for S_1 . The expanded presentation of the system reads

$$\Omega = \mu_1 \vee \mu_4 \vee \mu_5 \vee \mu_8 = (c_1 \wedge c_2 \wedge c_3) \vee (d_1 \wedge d_2 \wedge c_3) \vee (d_1 \wedge c_2 \wedge c_3) \vee (d_1 \wedge d_2 \wedge d_3).$$

We start by looking for a simplified expression for $\mu_4 \vee \mu_5$:

$$\begin{aligned} \mu_4 \vee \mu_5 &= (d_1 \wedge d_2 \wedge c_3) \vee (d_1 \wedge c_2 \wedge c_3) = d_1 \wedge (d_1 \vee c_2) \wedge (d_1 \vee c_3) \\ &\quad \wedge (d_2 \vee d_1) \wedge (d_2 \vee c_2) \wedge (d_2 \vee c_3) \\ &\quad \wedge (c_3 \vee d_1) \wedge (c_3 \vee c_2) \wedge c_3 \\ &= d_1 \wedge c_3 \end{aligned}$$

and the simplification is made after d_1 and c_3 absorbed several terms, and again neglecting those that are trivially true by construction. Next, we would like to compress states $e_1 \vee e_8$, but this is not possible: by expanding the expressions it is possible to rearrange them, but there is

Table 3. 3-bit microstates of a system with two constraints of scope two.

$\{\mu_E\}$
$\mu_1 = (1, 1, 1)$
$\mu_4 = (0, 0, 1)$
$\mu_5 = (0, 1, 1)$
$\mu_8 = (0, 0, 0)$

<https://doi.org/10.1371/journal.pone.0206489.t003>

no reduction in the end (not shown). The reason is that, if only these states would be observed, a global constraint should be acting on all three variables simultaneously, thus limiting their behaviour, and it is not possible to express with this formalism such a constraint except by presenting the states themselves. We will further explore this question in our last example below. For the present example, the description of the S2 system would be:

$$\Omega = \mu_1 \vee \mu_4 \vee \mu_5 \vee \mu_8 = (c_1 \wedge c_2 \wedge c_3) \vee (d_1 \wedge c_3) \vee (d_1 \wedge d_2 \wedge d_3)$$

The expression achieved is an unambiguous compact description of S2, and it would be interpreted in a model as follows (reading first the left parenthesis): either the subject observes all three states ON (1,1,1), OR (s)he observes OFF in the first position and OFF in the third one (0,?,1) OR all three are OFF (0,0,0). The same result would be obtained reading the expression in any other order: independently of the path followed, a decision tree representing the constraints of the system can be established. To recover all four states we can develop a generative model applying the rules proposed and drawing random values in the positions that, when the leaves of the decision tree are reached, remain undetermined (as it was for the second position in (0,?,1)). Note that we have lost again information in the compact representation achieved, although not as dramatically as in the previous example, because there are more constraints and they have a larger scope.

Now, looking into the concrete representation and, again following the procedure of the previous example (see Fig 5), we observe that the disconnected components in the formal network lead us to identify the constraints, that can be expressed with the propositions:

$$\text{Ext}(c_1 \wedge d_2) = \emptyset$$

$$\text{Ext}(c_2 \wedge d_3) = \emptyset$$

$$\text{Ext}(c_1 \wedge d_3) = \emptyset$$

It is easy to observe that one of these constraints is redundant. Given that c_2 and d_2 cannot be observed simultaneously, if c_1 is observed it means that c_2 is also observed and thus d_3

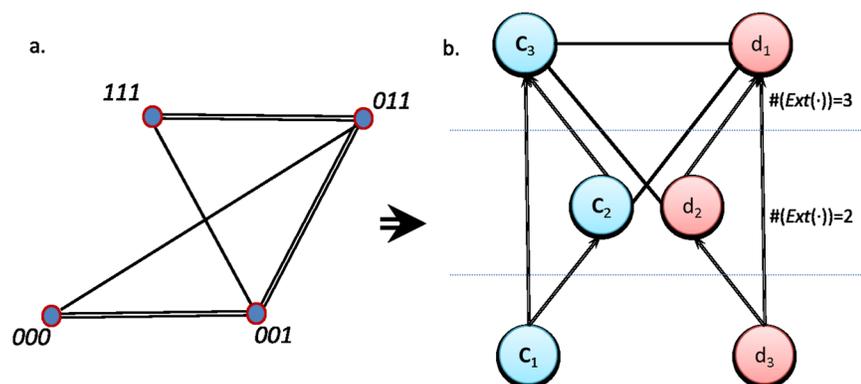


Fig 5. Representations of a three bits system with two constraints of scope two. (Left) In the concrete network, each node represents a microstate and it is linked with another microstate if they share the same observation for any component, where the number of links represent the number of concepts shared. (Right) Network of concepts extracted from the analysis of the microstates. Two links c_i and c_j are linked with a directed edge if $\text{Ext}(c_i) \subseteq \text{Ext}(c_j)$ and with an undirected link if $\text{Ext}(c_i) \cap \text{Ext}(c_j) \neq \emptyset$. The concepts are hierarchically ordered according to the cardinality of their extension, i.e. the number of microstates they map. In this example we identify the constraints observing those links that, despite of being viable, are absent. For instance, there is no link between d_3 and c_2 .

<https://doi.org/10.1371/journal.pone.0206489.g005>

cannot be observed. And the other way around, if d_3 is observed c_2 will not be observed and thus c_1 cannot be observed. Therefore, the third constraint $\text{Ext}(c_1 \wedge d_3) = \emptyset$, is a consequence of the other two, and we can just speak about the constraints of the system with two propositions.

System with a single constraint of scope three (S3; the parity bit system). Our last example is a set of microstates having an even number of ON bits, i.e. a single constraint involving all three components. This system has been previously introduced by Bar-Yam as a toy example of the particular type of emergent behaviour we introduced above called *strong emergence* [14]. For this system, given that we find two random values in two randomly selected bits, the third bit is constrained in such a way that the number of bits in the microstate is always odd. This rule is used in the control of message transmission, where the last bit (called the parity bit) is used to monitor the presence of errors in the chain transmitted. Note that we are not interested in understanding the system under this engineering perspective, as it provides already a rather *ad hoc* explanation on how the system is built [31] and, in this work, we assume no *a priori* knowledge of the underlying mechanisms generating the observation. It is just one possible observation that will be analysed as in the previous examples. The microstates we will consider are $\{\mu_1, \mu_2, \mu_3, \mu_4\}$, as explicitly shown in Table 4.

Again, we follow the same reasoning as in previous examples

$$\Omega = e_1 \vee e_2 \vee e_3 \vee e_4 = (c_1 \wedge c_2 \wedge c_3) \vee (c_1 \wedge d_2 \wedge d_3) \vee (d_1 \wedge c_2 \wedge d_3) \vee (d_1 \wedge d_2 \wedge c_3)$$

We look for a compact description of $e_1 \vee e_2$:

$$e_1 \vee e_2 = (c_1 \wedge c_2 \wedge c_3) \vee (c_1 \wedge d_2 \wedge d_3) = c_1 \wedge (c_2 \vee d_3) \wedge (c_3 \vee d_2).$$

Proceeding similarly with e_3 and e_4 yields

$$e_3 \vee e_4 = (d_1 \wedge c_2 \wedge d_3) \vee (d_1 \wedge d_2 \wedge c_3) = d_1 \wedge (c_2 \vee c_3) \wedge (d_2 \vee d_3),$$

and the description of the whole system would be

$$\Omega = e_1 \vee e_2 \vee e_3 \vee e_4 = [c_1 \wedge (c_2 \vee d_3) \wedge (c_3 \vee d_2)] \vee [d_1 \wedge (c_2 \vee c_3) \wedge (d_2 \vee d_3)].$$

The expression provides a kind of “bottom-up” description of the constraints of the system, that would allow us again to build a decision tree to propose a model. For instance, reading the expression from left to right we would say that if we observe ON in the first position (1,?,?) then, either we observe (1,1,?) and then (1,1,1), or (1,0,?) and then (1,0,0). The two remaining states would be obtained if OFF is observed in the first position. However, the expression achieved is not a compact description. The reason is that, even if in the search for a reduced description we absorbed a large number of terms, these were redundant, because if we expand the expression found we exactly recover the definition of the four microstates. For instance,

Table 4. 3-bit microstates of a system with a single constraint of scope three.

$\{\mu_E\}$
$\mu_1 = (1, 1, 1)$
$\mu_2 = (1, 0, 0)$
$\mu_3 = (0, 1, 0)$
$\mu_4 = (0, 0, 1)$

<https://doi.org/10.1371/journal.pone.0206489.t004>

expanding the expression between the first square brackets we get the first two states:

$$\begin{aligned} c_1 \wedge (c_2 \vee d_3) \wedge (c_3 \vee d_2) &= c_1 \wedge ((c_2 \wedge c_3) \vee (d_2 \wedge c_3)) \\ &= (c_1 \wedge c_2 \wedge c_3) \vee (c_1 \wedge d_2 \wedge d_3) = e_1 \vee e_2, \end{aligned}$$

and the other two are recovered from the second pair of brackets. Since a global constraint is acting on the system, we obtain that all the possible values of the variables are entangled. Therefore, the description of the constraint cannot be compressed, and hence it requires as much information as the presentation of the system itself, we are somehow “walking in circles” when we look for a description of the global constraint. We may say that we reached a more “expressive” description of the system, because we observe the relation between the values of the properties rather than the properties themselves but, since it is not compact, it may be also seen as an impredicative definition of the system’s constraints: we are defining the constraints with an expression that, rearranged, leads to the presentation of the system itself. This fact strongly resembles Rosen’s closure of efficient causation but, while the relational closure in his approximation focuses on the causal relationships between the concepts, we still have no clue about the nature of the constraints. However, it seems remarkable that global constraints seems to induce a closure from the point of view of the relation between the concepts what, in the jargon used by Rosen, would mean that the system is complex, as opposed to a mere complicated system.

This is readily apparent when we look at the network representations (see Fig 6), because the formal network intuitively resembles a sphere in the sense that there are no “borders” –i.e. disconnected concepts from which propositions about the constraints can be derived by simple inspection, nor concepts mapping the empty set–. Thus, the identification of constraints is more difficult than in the previous examples also under these representations. Indeed, it seems possible to identify that there is a global constraint only because we already know the viable values. Considering all eight microstates of the free-of-constraints system highlights a lower cooccurrence of the different values of the three variables, but there will be no differences in the final network topology we obtain. This fact would be also observed in the marginal probabilities, as no bias will be observed for the free system nor for the parity bit system.

The comparison with the free system lead us to conclude again that in order to speak about the system we need to write down all the microstates that are not observed:

$$\begin{aligned} \text{Ext}(d_1 \wedge c_2 \wedge c_3) &= \emptyset \\ \text{Ext}(c_1 \wedge d_2 \wedge c_3) &= \emptyset \\ \text{Ext}(c_1 \wedge c_2 \wedge d_3) &= \emptyset \\ \text{Ext}(d_1 \wedge d_2 \wedge d_3) &= \emptyset \end{aligned}$$

which requires the same amount of information as presenting the system itself, as was pointed out above.

We conjecture that, for systems with global constraints, it is not possible to derive a compact description within this formalism. Of course, we may think that our formalism is simply insufficient and that there may be other more sophisticated formalisms that would be able to express the constraints in S3 under a compact description. For instance, there may be a function f which, given a description of a microstate e_i , returns a new *type* of concepts $\tilde{c} = f(e_i)$ that is able to differentiate the microstates of S3 from the rest. Let’s see this with an example.

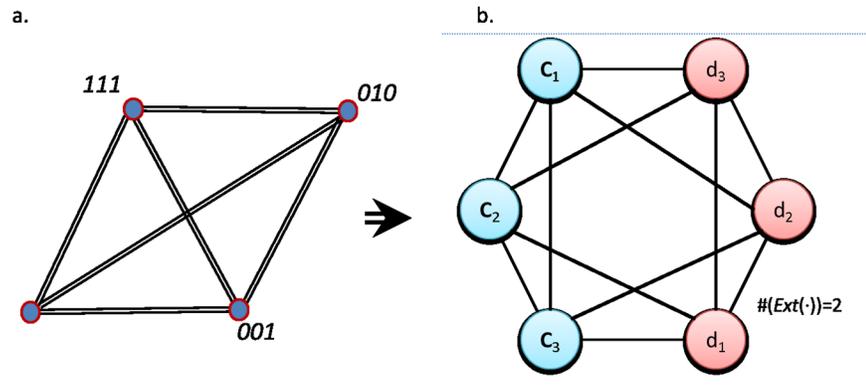


Fig 6. Representations of a three bits system with one constraint of scope three. (Left) In the concrete network, each node represents a microstate and is linked with another microstate if they share the same observation for any component, where the number of links represent the number of concepts shared. (Right) Formal network of concepts extracted from the analysis of the microstates. Two links c_i and c_j are linked with a directed edge if $Ext(c_i) \subseteq Ext(c_j)$ and with an undirected link if $Ext(c_i) \cap Ext(c_j) \neq \emptyset$. The graph of concepts is equivalent to the graph we would obtain for a free system, being just observed a reduction in the number of objects mapped by each concept (from $\#(Ext(\cdot)) = 4$ towards $\#(Ext(\cdot)) = 2$). It reflects the notion that the system has “no borders”.

<https://doi.org/10.1371/journal.pone.0206489.g006>

We can characterize the three above examples with their probability distributions:

$$P(\mu) = \delta(x_1, 1)/2^{n+1} \tag{S1}$$

$$P(\mu) = \delta(\delta(H_{12} + 1, 1), H_{23} + 1)/2^{n+1} \tag{S2}$$

$$P(\mu) = \delta(\text{mod}_2(\sum_i x_i), 1)/2^{n+1} \tag{S3}$$

where x_i is the value of the bit i , n is the number of bits, $\delta(a, b)$ is the Kronecker’s delta, $H_{ij} = H(x_i - x_j - 1)$ is the Heaviside function and $\text{mod}_2(\cdot)$ is the module two function. In principle, there is no reason to think that the probability distribution of S3 is more complex than those of S1 and S2. To say this there should be any objective reason showing, for instance, that the use of a summation operator and a module function is more complex than applying two Heaviside functions.

To circumvent these issues, measures like the algorithmic information complexity, also known as Kolmogorov complexity, have been proposed [32]. This is the spirit in which we developed the present proposal. We believe that the fact that the formalism we are using here has well defined limits (arising from the rather limited repertoire of operations allowed), is not a drawback but an advantage to fairly compare different systems. Furthermore, when we reach a limit, as it is the case for S3, we have a clear intuition of which are the ontological properties behind our epistemological limitation, in this case the existence of a global constraint. For instance, a desirable property of a framework should be that, if we increase the system’s size, our ability to describe the constraints of a larger system should change according with their number and scope. If we imagine how our examples will increase the number of microstates when the number of components N increases, both S1 and S3 will increase as 2^{N-1} while S2 increases as $N + 1$. According to our formalism, the length of the description of the system will change with increasing N . For instance, the number of propositions mapping the formal and concrete state of the type $Ext(c) = \emptyset$ we need in order to describe the constraints in the system remains equal to one for S1, increases as $N - 1$ for S2, and as 2^{N-1} for S3. We believe that this finding is remarkable and resembles the problems identified by Tsallis [33] with the extensivity of the entropy. In the following sections, we exploit it to provide a quantification of emergence compatible with the scientific method.

Quantification of emergence

According to the above results, the number and scope of the constraints are ontological properties that determine our epistemological ability to achieve a compact description, which we raised as a necessary condition to delineate a model aimed at explaining an emergent macroscopic property. If, as is the case with the formalism we selected, these constraints are fairly comparable among systems, then the amount of information that codifies its description seems to be a natural way to quantify emergence. It is, however, difficult to imagine that there exist a framework general enough to perform such quantification in real systems.

In this section we propose a series of concepts that we believe are compatible with computational modelling of complex systems, and that are likely independent of the system used and modelling approach followed. We start by supposing that we are dealing with a model of an emergent behaviour, and that our strategy consists of performing *interventions* to the system such that variables in the model are substituted for variables free of constraints. We then monitor what is the relative change in our ability to predict the system's behaviour after the intervention.

Coverage excess. Intervention is a basic strategy to link computational modelling with the scientific method [34]. When we neglect any variable describing a system, we reduce our predictive power if the variable is constrained with others, as the latter will lose specificity and so there will be more states of the system compatible with the remaining constraints. We can quantify the uncertainty we generate when we lose constraints; meaning that we can relate the causal effects a component has on the other components according to the notion of Granger causality [35].

We illustrate the proposed method in Fig 7 by considering the 3-bit synthetic examples. Starting from a 3-bit system, we neglect one of the variables, and then we explore which are the 2-bit states recovered. Returning to the system with the chosen variable neglected, but allowing it to now obtain any viable value (i.e. free of constraints), we will be able to build from these 2-bit states a number of 3-bit states. By doing this systematically for all the variables, we can infer what is the influence of the underlying constraints. For instance, for S1, if we neglect the first component x_1 , which is the source of the constraint, we obtain all possible states of a 2-bit system containing components x_2 and x_3 . This is because there is a single constraint in the system and it was deleted, hence we can recover the whole phase space Ω when we build up all the 3-bit states compatible with these 2-bit states. On the other hand, if we neglect any of the other two components, x_2 or x_3 , the constraint still remains in the first component x_1 . Therefore, if we look for the 3-bit states compatible with these 2-bit states, we will be constrained to come back to the original system S1. For the system S2, irrespective of the component neglected we will reach the same 2-bit states. From them, we obtain not only the original 3-bit states but two more, which means that we have lost one constraint every time, but the other one is still active. For S3, irrespective of the component neglected, we obtain all possible two bits states, and hence the whole phase space will always be recovered, meaning that the constraint is always completely lost, thus pointing towards the existence of a global constraint.

With these kinds of interventions we can quantify how the original system is *covered in excess* when an intervention takes place. Formally, let's call U the subset of concepts contained in the focus F describing a set of microstates $S = \{\mu\}$ containing N components in the phase space Ω , which we know is associated with a target macroscopic property. For simplicity, suppose that each component i is fully characterized by a single variable x_i . Next, remove any of the x_i variables and consider a new set of microstates S' of a system with $N - 1$ components. Let's then denote as $V(x_i)$ the subset of concepts describing the set of microstates S'' of a system

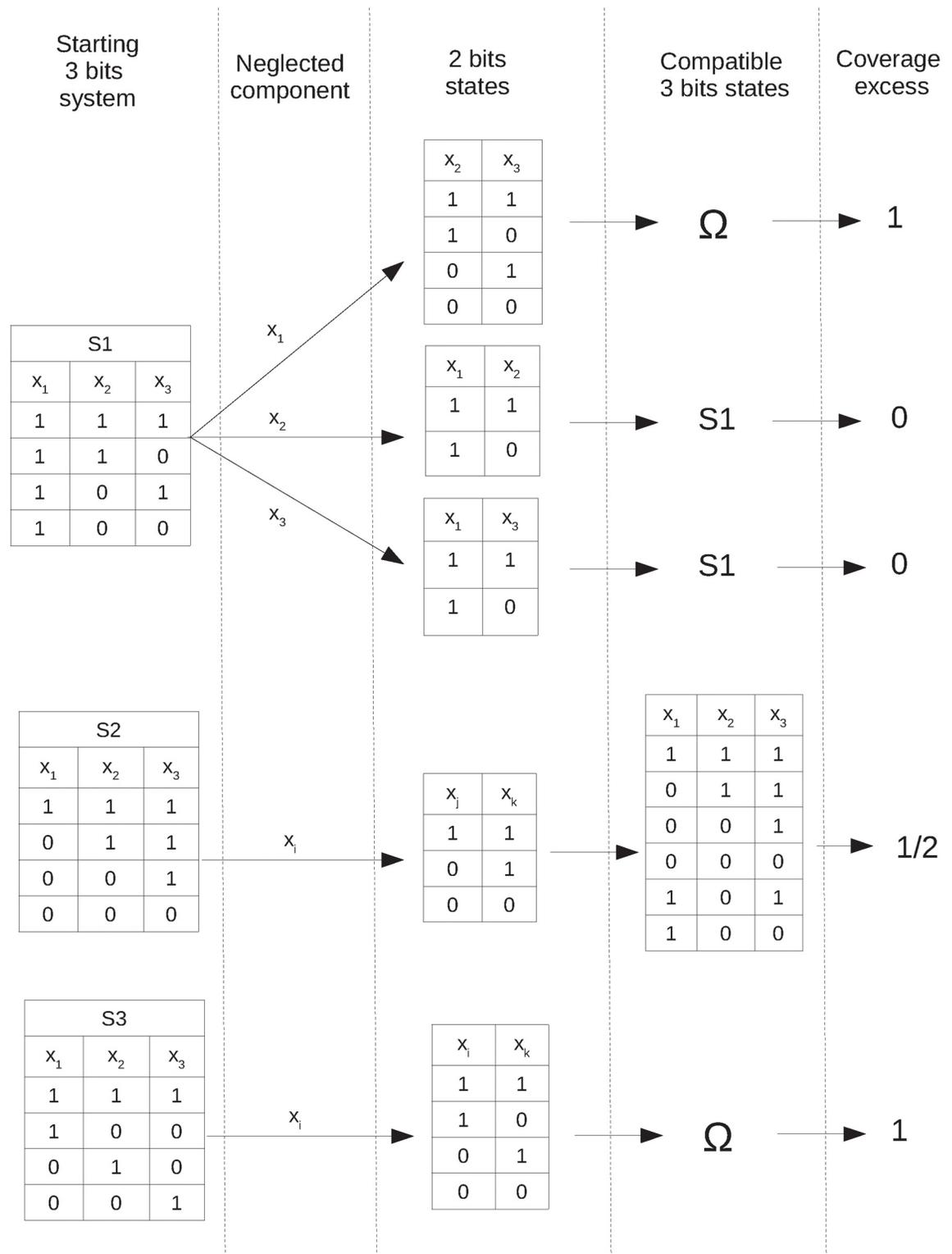


Fig 7. Scheme illustrating the definition of coverage excess. The scheme is divided in five columns (1-5) that we describe from left to right. (1) The 3-bit systems under analysis in the main text are shown. If we intervene in the systems neglecting one component (2) we will obtain a set of 2-bit states (3). For the system S1, removing x_1 lead to different states than if x_2 or x_3 are removed, while S2 and S3 lead to the same states independently of the component removed (see Main Text for details). From the 2-bit states, we recover the neglected component keeping it free of any constraint, which leads to a number of compatible 3-bit states (4). In the last column we

show the result for the coverage excess obtained from this procedure using Eq 8. The final value for the coverage excess of the system will be the average among the values obtained from the different interventions.

<https://doi.org/10.1371/journal.pone.0206489.g007>

with N components obtained by adding a new unconstrained component x'_i to S' . We say that the coverage excess Ξ of S induced after neglecting the variable x_i and introducing back the component x'_i is the quantity

$$\Xi(x_i) = \frac{\#(\text{Ext}(V(x_i))) - \#(\text{Ext}(U))}{\#(\text{Ext}(F)) - \#(\text{Ext}(U))} = \frac{\#(S') - \#(S)}{\#(\Omega) - \#(S)} \tag{8}$$

where the function $\#(\cdot)$ returns the number of microstates contained in the set. This quantity takes a value of zero when the states recovered are the same as the original ones, and it is equal to one when the whole phase space is recovered. To consider a single value for the coverage excess we will average the result of the intervention over all variables:

$$\langle \Xi \rangle = \frac{1}{N} \sum_i \Xi(x_i) \tag{9}$$

If the system is very large an exhaustive computation may be unfeasible, and a random sampling of the different components or more complex interventions such as the removal of several variables would be needed. We indicate this average over interventions with the brackets $\langle \cdot \rangle$. The examples explored for the three bits system results in $\langle \Xi \rangle_{S1} = 1/3$, $\langle \Xi \rangle_{S2} = 1/2$ and $\langle \Xi \rangle_{S3} = 1$. The coverage excess reflects the vulnerability of the system to this intervention and, as the effects link the interventions with the number and scope of the constraints present in the system, it provides a mechanism to differentiate between upward and downward causation [36]. The system S1 is very vulnerable if x_1 is neglected, but it is not affected at all if any other variable is neglected, and thus there is upward causation from the first variable to the whole system. On the other hand, the system S3 is very vulnerable as the coverage excess is maximum irrespective of the variable over which we intervene, thus highlighting that there is a global constraint affecting the system downwards. We also note that these values will scale differently with the system's size. While for S1 $\langle \Xi \rangle_{S1} \rightarrow 0$ when $N \rightarrow \infty$, for S3 it will remain constant and equal to one. Finally, we consider the particular case in which we are already dealing with a subset of microscopic concepts U such that an emergent process described by the macroscopic concept \hat{c} is perfectly covered, i.e. such that $\text{Ext}(U) = \text{Ext}(\hat{c})$. In this case, we will say that the Eq 9 provides the coverage excess of the emergent property \hat{c} , that we denote as $\langle \hat{\Xi} \rangle$.

Loss of traceability and emergence strength. The sensitivity of the model when an intervention over the system takes place suggests that systems with higher coverage excess will be more difficult to analyse. This difficulty can be combined with the notion of traceability we proposed above. If we have a perfectly traceable system, we can quantify how much traceability we lose after intervention with the *loss of traceability*

$$\Upsilon = 1 - \langle \hat{\Xi} \rangle. \tag{10}$$

With this quantity, we can express those systems that are easily covered in excess have low traceability and the other way around. Therefore, the associated macroscopic property will be difficult to explain, from which the following definition for the *emergence strength* σ of the emergent property arises naturally

$$\sigma = -\log(\Upsilon). \tag{11}$$

We haven taken the logarithm of the traceability to represent the emergence strength as a dissimilarity between a state of perfect knowledge of the system (complete traceability) and the

state after intervention. If, after intervention, we still remain in a situation of perfect traceability, this dissimilarity will be zero. On the other hand, if we completely lose all the information about the system in a way that we recover an unconstrained phase space, this dissimilarity will be infinite, thus reflecting some epistemological gap for these systems.

With these definitions we expect to reconcile different positions on whether the origin of emergence is epistemological or ontological: even if we deal with a perfectly traceable system, which is therefore epistemologically accessible, we can still see that there are systems that are more inaccessible than others, and there are ontological reasons for that: the type of constraints involved in the system. For these systems, until perfect traceability is attained, we will probably be tempted to say that they are epistemologically inaccessible, and that there is a strictly ontological and not epistemological reason for that. But it is a combination of both: there is an ontological reason why their emergence strength is so high that hinders an epistemologically accessible (microscopic) compact description, but this doesn't mean that such a description cannot be achieved at some point. Of course, we should keep in mind that achieving a compact description does not mean that we have a full mechanistic understanding of the process: we still need to develop a model in the sense introduced here, that would interpret the constraints and generate the microscopic phase space. If the constraints are very complex the implementation of constraints may never be feasible. Furthermore, even if a satisfactory generative model is developed, it does not mean that we are able to generate the macroscopic emergent pattern with a model. This would require a final process to decode the formal model, *sensu* Rosen [28], which may further require a complex experimental setup (the decoding step shown in Fig 1).

Given that during the process of researching an emergent property there are different steps we could use to evaluate the accessibility of the underlying process, we propose to focus upon the analysis of the patterns obtained from the experimental data as the starting step. This may help to reconcile the definition of weak and strong emergent properties [14] using the emergent strength: if the emergence strength is infinite we deal with a strongly emergent pattern, whereas if the value is finite we deal with a weak emergent pattern with the associated strength as an indicator of how difficult it is to achieve traceability. We conjecture that further difficulties in computational or experimental modelling will likely go hand in hand with difficulties in the determination of the pattern's traceability.

Of course, we cannot discount the possibility that there exist systems which are epistemologically inaccessible. This may be the case for quantum or computational systems—for which some of the definitions of weak and strong emergence were originally proposed—but not for many systems of scientific interest, where we believe that the situation is the one that we examined here: these systems are very large and they are under constraints large in scope. Note that it may be argued that, with the above definitions, the emergence strength cannot be determined unless we achieve perfect traceability. However, this is only true for strongly emergent properties because Eq 10 can be easily modified to consider an intermediate (incomplete) model, using only the expression in Eq 9 (and not $\langle \hat{\Xi} \rangle$), which will give us an estimate of the emergence strength. We expect that, for natural systems, there is a complex structure of constraints with different scopes, and we will be able to progressively discover this structure—possibly from low to high scopes—and so provide an estimate at any time.

Discussion

In this article, we presented a novel approach to investigate the concept of emergence in complex systems. We tackled the problem through a constructive logical system that permits the investigation of the relationship between concepts and objects of observation [25]. In doing so,

we focused upon a particular kind of system, which we believe are of much interest to the current discussion of emergence. We start by supposing that we are analysing a naturally occurring macroscopic emergent property, and not a purely computational one. In addition, we neglect any vitalism, which means simply accepting explanatory physicalism; in the words of Mitchell, what else could there be? [17].

We then assume that we are able to describe microstates of the system through experimental measurements. This implicitly assumes that we are able to differentiate the system from its background [21] and to provide a bottom-up characterization in terms of concepts associated with the elements that constitute the system [16], thus justifying the constructive approach. Nevertheless, we allowed for the possibility that we have no clue about the mechanistic processes underlying these observations, as often happens when a research program is in its infancy.

This fact differentiates this work from other theoretical approximations aiming to understand which features belong to systems exhibiting emergent properties, but that already assume that sufficient knowledge about the system exists so as to test its computational compressibility (Bedau [8]), to postulate the existence of a closure of efficient causation (Rosen, which requires determining the causal relationships [28]) or to intervene over a system for which we already have a mechanistic model (see for instance Hoel et al. [37]). However, this is typically not the case when in the early stages of research, and this is why we believe our proposal may be helpful for a wide variety of scientists. Invoking a mild condition relating the macroscopic observation of an emergent property and the constrained walk of the system in a certain region of the phase space, we focus on systems from which we expect to find sufficient regularities in the analysis of their microstates so as to be able to build explanatory models, i.e. in potentially robust emergent systems, which we believe are of key interest to the scientific community [13].

We showed that building a microscopic model aimed to explain an emergent macroscopic observation requires identifying constraints in the viable values of the microscopic variables. Interestingly, we identified concept disjunction as the basic logic operation to find constraints. The search for similarity measures, dissimilarity measures or distances is an essential task in Biology and Ecology [38] aiming to understand, following a top-down approach, the information shared between the different observations. This probably explains the success of complex networks theory and its philosophical interest, or why methods comparing objects of observation, such as protein sequence alignments like BLAST [39], are among the most cited ever in the scientific literature [40]. In general, disjunction underlies dimensionality reduction techniques such as principal components analysis [41]. From the perspective of our framework, these are techniques aiming at obtaining a representation with the minimum number of concepts whose extension explains the full variability of the microstates. In this way, we are able to *talk* about the set of objects using a subset of concepts, which is essentially the task addressed by dimensionality reduction techniques, and that we defined here with the notion of compact description.

We applied these tools to three different ensembles of microstates of a 3-bit synthetic system. We observed that the scope of the constraints is the main difficulty in identifying them: the larger the scope of the constraint, the more difficult is to assess it. In particular, our method was unable to find a compact representation when the scope of the constraint has the same size as the system, which directly links the epistemological limitation of our framework with an ontological property of the system. We briefly considered other approximations, and were able to show that the number and type of constraints heavily influence the consequences that either an increase in system size or a loss of components may have upon our ability to identify them. This observation seems to be independent of the formalism used, and so will also be

independent of any subjectivity induced by the formalism we chose here. Notably, we were able to express this observation in the concrete space; thus further research would be needed to find equivalent definitions in the formal space.

We also proposed a procedure based on the intervention of the observer on the system, thus compatible with the scientific method, to compute the loss of information experienced when we neglect components in the system. Given that the loss of information depends on the type of constraints present, we can quantify how difficult it is to achieve traceability between the microscopic and macroscopic description. The loss of traceability was then used as a quantity to establish a distance between perfect traceability and our knowledge of the system after systematic interventions, which was what we called the emergence strength.

We believe that, for the kind of systems we are interested in, the emergence strength paves the way for us to reconcile different notions of emergence. For natural systems, we aim to develop computational models to reproduce experimentally measured data and to then simulate the emergent process, and thus it is compatible with weak emergence. Nevertheless, we propose to combine the ability to build a computational model with the identification of constraints from experimental data, since the identification of constraints is where we start learning about the natural process we face. In this way, we focus on disentangling the number and scope of the constraints, whose complexity will determine its emergence strength.

We conjecture that, for systems with different types of constraints, those with smaller scope are identified first. Accordingly, if a system has only constraints with a large scope or there is a big gap with respect to those with smaller scope, it may be simply impossible at a certain state of knowledge to assess them, an example of which may be our current knowledge of consciousness. For these processes, the emergence strength may be so high that we would be justified in calling them strongly emergent processes. For such processes, it may be the case that not only is it impossible to decipher the constraints, but even if the microscopic constraints are deciphered, it may still not be possible to build an experimental setup for a model to recreate the observed emergent pattern, given the complexity of the environment in which the system should be embedded to reproduce such a constraint. Our definition seems to also be compatible with the classification proposed by de Haan [7], as the existence of a microscopic emergent conjugated causally affecting the macroscopic pattern (in the strongest version, consciously), can be understood in terms of a global constraint (as he suggests in the relationship between this type of emergence and downward causation). This is the case for living systems, where we believe strong emergence may be pervasive.

Our findings might be criticized as saying that describing emergence in terms of constraints provides a static description for systems that are intrinsically dynamic; an approach which may be thought of as a kind of ontological reductionism [17]. Note, however, that constraints may themselves be dynamic and, either their variation occurs on a longer timescale, or the constraints dynamics is itself sufficiently constrained, e.g. periodic conditions such as day-night or seasonal temperatures. This would also address potential criticisms regarding multiple realizability: similar microscopic patterns can be found for systems under similar constraints even if the particular realizations of the microstates are substantially different for each system. Multiple examples of this can be found in the literature. For example, the evolutionary process allow us to classify protein structures in clusters if they have global structural similarity (that result from similar physico-chemical constraints) even if they perform different functions [42]. Similarly, ecological patterns such as the nestedness found in ecological networks describe complex constraints in the way in which species interact, even if they are found in different ecosystems, from plant-pollinators [43] to host-virus systems [44].

If patterns observed in organisms are the consequence of natural selection, and global constraints are acting on individuals in the selection process, natural selection itself can be thought

of as an expression of downward causation [45]. Interestingly, adaptation takes place when the organisms are able to predict and overcome environmental changes, and predictability is a consequence of the amount of structured information that exists in the environment [46]. As a corollary, the ability of organisms to modify the environmental constraints to make them more predictable, enhances the organisms' potential for adaptation. But, organisms share their environment with other organisms, and thus ecological interactions are fundamental to the adaptive process. This picture, in its stronger version, in which the influence of ecological interactions is so important that the notion of an individual as object of selection is challenged, becomes increasingly important in current research, particularly in the microbial world (see for instance [47]).

To illustrate this point, consider the following example proposed in [48], in which we consider one individual for which its fitness f_i can be decomposed into two components, where the first component reflects the fitness f_{ij}^{int} of the individual as a consequence of its ecological interactions with other species j , and the second its fitness $f_i^{\overline{\text{int}}}$ due to any other process, i.e. $f_i = f_{ij}^{\text{int}} + f_i^{\overline{\text{int}}}$. Now consider a particular example, in which two individuals belong to two different species, a and b , interacting mutualistically. The effect of the interaction on the fitness f_i would be positive through an increase reflected in the term f_{ij}^{int} . Finally, think of an evolutionary event which becomes fixed in the population of species a affecting its fitness, $f_a \rightarrow \hat{f}_a$, in such a way that the new fitness $\hat{f}_a < f_a$ and, in particular, $\hat{f}_a^{\overline{\text{int}}} = f_a^{\overline{\text{int}}}$ but $\hat{f}_{ab}^{\text{int}} < f_{ab}^{\text{int}}$. This means that the fitness of species b due to the interaction with species a will also be affected after the evolutionary event and thus there will be a change in the selection pressure on the regions of the genomes of both species codifying the traits needed for the interaction. Furthermore, if we consider an extreme scenario in which $f_{ab}^{\text{int}} \gg f_a^{\overline{\text{int}}}$ and $f_{ba}^{\text{int}} \gg f_b^{\overline{\text{int}}}$ —that may be the case for auxotrophs (see a synthetic ecological experiment in [49])—the relevance of these coevolving regions in the evolutionary process would be so important that the concept of an object of selection should be revisited [50]. In particular, it might be more appropriate to frame the evolution of both species by considering them as some sort of multicellular species. In this sense, even if the individual is still the main object of selection, it becomes entangled with an object of selection determined on a larger scale, which is the consortia of species. Furthermore, if this consortia acquires new functionality that make its members selectively advantageous in line with Kim [5], we may say that a new object of selection emerges. If that were the case, the single species' individuals would be constrained downwards by their fitness dependence with respect to the consortia they belong to, and the consortia is itself influenced upwards by the individual species. In the same way that we admit the existence of different levels of organization of living beings that depend on a hierarchy of constraints, we should also consider the possibility that individuals belong to different objects of selection influencing their fitness to different degrees. In this way, the term upwards (downwards) causation would be used in this context as the effect of constraints acting on a given level has ramifications for objects of selection on upper (lower) levels.

This perspective probably reflects the interest of modern science on mutualistic interactions and its relation with emergent processes [4], although mutualism may not be necessary to derive a measure of community-level fitness [51]. The determination of fitness above the individual level may be seen as a form of self-determination that would engage with the concept of closure of efficient causation. And in the same way that it is possible to argue that complete or perfect closure exists for any living being, it is also possible to argue for the concept of the object of selection to be a closed concept and reality. This is probably why it is aimed at extending the concept of closure to an ecological context [52]. A rigid definition of closure or of an

object of selection might be an acceptable concept in an ideal thermodynamical scenario of natural selection [53] –and thus theoretically interesting– but it may perhaps be a good idealization of the limits of biological organization, like simple cells or the whole biosphere. But these concepts become blurred in complex ecosystems in natural environments. We believe that there is no need to invoke any teleological principle [54]. Only in an ideal scenario in which life evolves spontaneously following thermodynamical principles may one envisage a teleological ideal in which biological organization maximizes any entropic or energetic principle [55], or arbitrary fitness. It can be argued that this scenario is true in general. A simple counterexample may be found in the way in which human beings not only do not optimize any spontaneous physico-chemical principle from which life may have emerged, but rather generate a process that may eventually lead to extinction.

In summary, we believe that the formalism introduced here improves our ability to synthetically understand complex systems. We believe that it could also be used to tackle other challenging questions, and thus we hope that our effort will stimulate both scientific and philosophical discussion. Looking for fresh formal approaches to talk about philosophical questions is particularly important because formally shaping our philosophical knowledge is a way to create new bridges between science and philosophy. This would be probably good news for science, as the benefits of philosophy seem to be, for current scientists, left behind.

Acknowledgments

I am particularly in debt with Silvio Valentini for his continuous and patient feedback. This work was partially developed in Ugo Bastolla's and Thomas Bell's labs, I acknowledge them their support. I also acknowledge to Andrés Moya's lab their hospitality during a short stay of research. I gratefully acknowledge discussions with Jorge Ibáñez-Gijón and Fernando Rosas, comments of Steven Vickers on an earlier version of the manuscript, and constructive criticisms of an anonymous referee. I finally thank to Peter Ashcroft, Judith Bouman, Arion Iffland-Stettner and David McLeod their help proofreading the paper.

Author Contributions

Conceptualization: Alberto Pascual-García.

Formal analysis: Alberto Pascual-García.

Investigation: Alberto Pascual-García.

Methodology: Alberto Pascual-García.

Validation: Alberto Pascual-García.

Visualization: Alberto Pascual-García.

Writing – original draft: Alberto Pascual-García.

Writing – review & editing: Alberto Pascual-García.

References

1. Nietzsche F. Thus spoke Zarathustra. Manis J, editor. Hazleton: Pennsylvania State University; 2013.
2. Brenner S. Sequences and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010; 365(1537):207–212. <https://doi.org/10.1098/rstb.2009.0221>
3. Emmeche C, Köppe S, Stjernfelt F. Explaining emergence: towards an ontology of levels. *Journal for General Philosophy of Science*. 1997; 28(1):83–117. <https://doi.org/10.1023/A:1008216127933>
4. Corning PA. The re-emergence of emergence, and the causal role of synergy in emergent evolution. *Synthese*. 2012; 185(2):295–317. <https://doi.org/10.1007/s11229-010-9726-2>

5. Kim J. Making sense of emergence. *Philosophical studies*. 1999; 95(1):3–36. <https://doi.org/10.1023/A:1004563122154>
6. Crick F, Clark J. The astonishing hypothesis. *Journal of Consciousness Studies*. 1994; 1(1):10–16.
7. de Haan J. How emergence arises. *Ecological Complexity*. 2006; 3(4):293–301. <https://doi.org/10.1016/j.ecocom.2007.02.003>
8. Bedau MA. Weak emergence. *Noûs*. 1997; 31(s11):375–399.
9. Crutchfield JP. When evolution is revolution—Origins of innovation. In: Crutchfield JP and Schuster P, editors. *Evolutionary dynamics: exploring the interplay of selection, neutrality, accident, and function*. Oxford: Oxford University Press; 2003; p. 101–134.
10. Bedau M. Downward causation and the autonomy of weak emergence. *Principia*. 2002; 6(1):5.
11. Holland JH. *Emergence: From chaos to order*. Redwood City, CA: Addison-Wesley. 1998
12. Huneman P, Humphreys P. Dynamical emergence and computation: an introduction. *Minds and Machines*. 2008; 18(4):425–430. <https://doi.org/10.1007/s11023-008-9124-4>
13. Huneman P. Determinism, predictability and open-ended evolution: lessons from computational emergence. *Synthese*. 2012; 185(2):195–214. <https://doi.org/10.1007/s11229-010-9721-7>
14. Bar-Yam Y. A mathematical theory of strong emergence using multiscale variety. *Complexity*. 2004; 9(6):15–24. <https://doi.org/10.1002/cplx.20029>
15. Chalmers DJ. Strong and weak emergence. In: Clayton P, and Davies P, editors. *The Re-Emergence of Emergence*, Oxford: Oxford University Press. 2006; p. 244–256.
16. Bich L. Complex emergence and the living organization: an epistemological framework for biology. *Synthese*. 2012; 185(2):215–232. <https://doi.org/10.1007/s11229-010-9722-6>
17. Mitchell SD. Emergence: logical, functional and dynamical. *Synthese*. 2012; 185(2):171–186. <https://doi.org/10.1007/s11229-010-9719-1>
18. Sambin G. Some points in formal topology. *Theoretical computer science*. 2003; 305(1):347–408. [https://doi.org/10.1016/S0304-3975\(02\)00704-1](https://doi.org/10.1016/S0304-3975(02)00704-1)
19. Ryan AJ. Emergence is coupled to scope, not level. *Complexity*. 2007; 13(2):67–77. <https://doi.org/10.1002/cplx.20203>
20. Alley TR. Organism-environment mutuality epistemics, and the concept of an ecological niche. *Synthese*. 1985; 65(3):411–444. <https://doi.org/10.1007/BF00869278>
21. Maturana HR, Varela FJ. *Autopoiesis and cognition: The realization of the living*. vol. 42. Dordrecht: Springer Science & Business Media; 1991.
22. Machta BB, Chachra R, Transtrum MK, Sethna JP. Parameter space compression underlies emergent theories and predictive models. *Science*. 2013; 342(6158):604–607. <https://doi.org/10.1126/science.1238723> PMID: 24179222
23. Boniolo G, Valentini S. Vagueness, Kant and topology: a study of formal epistemology. *Journal of Philosophical Logic*. 2008; 37(2):141–168. <https://doi.org/10.1007/s10992-007-9064-0>
24. Vickers S. *Topology via logic*. vol. 5. Cambridge: Cambridge University Press; 1996.
25. Boniolo G, Valentini S, et al. Objects: A study in Kantian formal epistemology. *Notre Dame Journal of Formal Logic*. 2012; 53(4):457–478. <https://doi.org/10.1215/00294527-1722701>
26. Anderson PW, et al. More is different. *Science*. 1972; 177(4047):393–396. <https://doi.org/10.1126/science.177.4047.393> PMID: 17796623
27. Hooker C. On the import of constraints in complex dynamical systems. *Foundations of Science*. 2013; 18(4):757–780. <https://doi.org/10.1007/s10699-012-9304-9>
28. Rosen R. *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia University Press; 1991.
29. Auyang SY. *Foundations of complex-system theories: in economics, evolutionary biology, and statistical physics*. Cambridge: Cambridge University Press; 1999.
30. Testa B, Kier LB. Emergence and dissolution in the self-organisation of complex systems. *Entropy*. 2000; 2(1):1–25. <https://doi.org/10.3390/e2010001>
31. Bersini H. Emergent phenomena belong only to biology. *Synthese*. 2012; 185(2):257–272. <https://doi.org/10.1007/s11229-010-9724-4>
32. Gell-Mann M, Lloyd S. Information measures, effective complexity, and total information. *Complexity*. 1996; 2(1):44–52.
33. Tsallis C, Mendes R, Plastino AR. The role of constraints within generalized nonextensive statistics. *Physica A: Statistical Mechanics and its Applications*. 1998; 261(3):534–554. [https://doi.org/10.1016/S0378-4371\(98\)00437-3](https://doi.org/10.1016/S0378-4371(98)00437-3)

34. Boschetti F. Causality, emergence, computation and unreasonable expectations. *Synthese*. 2011; 181(3):405–412. <https://doi.org/10.1007/s11229-010-9720-8>
35. Seth AK. Measuring autonomy and emergence via Granger causality. *Artificial Life*. 2010; 16(2):179–196. <https://doi.org/10.1162/artl.2010.16.2.16204> PMID: 20067405
36. Bitbol M. Downward causation without foundations. *Synthese*. 2012; 185(2):233–255. <https://doi.org/10.1007/s11229-010-9723-5>
37. Hoel EP, Albantakis L, Tononi G. Quantifying causal emergence shows that macro can beat micro. *PNAS*. 2013; 110(49):19790–19795. <https://doi.org/10.1073/pnas.1314922110> PMID: 24248356
38. Legendre P, Legendre L. *Numerical Ecology*. Oxford: Elsevier; 2012.
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990; 215(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
40. Van Noorden R, Maher B, Nuzzo R, et al. The top 100 papers. *Nature*. 2014; 514(7524):550–553. <https://doi.org/10.1038/514550a> PMID: 25355343
41. Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*. 1933; 24(6):417. <https://doi.org/10.1037/h0071325>
42. Pascual-García A, Abia D, Ortiz ÁR, Bastolla U. Cross-over between discrete and continuous protein structure space: Insights into automatic classification and networks of protein structures. *PLoS Computational Biology*. 2009; 5(3):e1000331. <https://doi.org/10.1371/journal.pcbi.1000331> PMID: 19325884
43. Bascompte J, Jordano P, Melián CJ, Olesen JM. The nested assembly of plant-animal mutualistic networks. *PNAS*. 2003; 100(16):9383–9387. <https://doi.org/10.1073/pnas.1633576100> PMID: 12881488
44. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host–phage interactions. *PNAS*. 2011; 108(28):E288–E297. <https://doi.org/10.1073/pnas.1101595108> PMID: 21709225
45. Bishop RC. Downward causation in fluid convection. *Synthese*. 2008; 160(2):229–248. <https://doi.org/10.1007/s11229-006-9112-2>
46. Zenil H, Gershenson C, Marshall JA, Rosenblueth DA. Life as thermodynamic evidence of algorithmic structure in natural environments. *Entropy*. 2012; 14(11):2173–2191. <https://doi.org/10.3390/e14112173>
47. Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology*. 2014; 12(4):263–273. <https://doi.org/10.1038/nrmicro3218> PMID: 24590245
48. Pascual-García A. Emergent patterns in protein, microbial and mutualistic systems. Ph.D. Thesis. Universidad Autónoma de Madrid; 2015. URI: hdl.handle.net/10486/668019
49. Mee MT, Collins JJ, Church GM, Wang HH. Syntrophic exchange in synthetic microbial communities. *PNAS*. 2014; 111(20):E2149–E2156. <https://doi.org/10.1073/pnas.1405641111> PMID: 24778240
50. Mayr E. The objects of selection. *PNAS*. 1997; 94(6):2091–2094. <https://doi.org/10.1073/pnas.94.6.2091> PMID: 9122151
51. Tikhonov M. Community-level cohesion without cooperation. *eLife*. 2016; 5:e15747. <https://doi.org/10.7554/eLife.15747> PMID: 27310530
52. Luz M, et al. Rosennean complexity and its relevance to ecology. *Ecological complexity*. 2017;.
53. Smith E. Thermodynamics of natural selection II: chemical Carnot cycles. *Journal of theoretical biology*. 2008; 252(2):198–212. <https://doi.org/10.1016/j.jtbi.2008.02.008> PMID: 18367209
54. Mossio M, Bich L. What makes biological organisation teleological? *Synthese*. 2017; 194(4):1089–1114. <https://doi.org/10.1007/s11229-014-0594-z>
55. Jorgensen SE, Svirezhev YM. *Towards a thermodynamic theory for ecological systems*. Oxford: Elsevier; 2004.