 doi:10.20944/preprints202106.0019.v1

EMERGENCE AND ALGORITHMIC INFORMATION DYNAMICS OF SYSTEMS AND OBSERVERS

FELIPE S. ABRAHÃO AND HECTOR ZENIL

Abstract. Previous work has shown that perturbation analysis in algorithmic information dynamics can uncover generative causal processes of finite objects and quantify each of its element's information contribution to computably constructing the objects. One of the challenges for defining emergence is that the dependency on the observer's previous knowledge may cause a phenomenon to present itself as emergent for one observer at the same time that reducible for another observer. Thus, in order to quantify emergence of algorithmic information in computable generative processes, perturbation analyses may inherit such a problem of the dependency on the observer's previous formal knowledge. In this sense, by formalizing the act of observing as mutual perturbations, the emergence of algorithmic information becomes invariant, minimal, and robust to information costs and distortions, while it indeed depends on the observer. Then, we demonstrate that the unbounded increase of emergent algorithmic information implies asymptotically observer-independent emergence, which eventually overcomes any formal theory that any observer might devise. In addition, we discuss weak and strong emergence and analyze the concepts of observer-dependent emergence and asymptotically observer-independent emergence found in previous definitions and models in the literature of deterministic dynamical and computable systems.

1. Introduction

Perturbation (or intervention) analyses of changes in the algorithmic information necessary for computably constructing an object allow the investigation of underlying causal effectiveness of the parts (or elements) [47, 49] as well as the solution to the inverse problem of finding the best generative model [50]. This approach adopted by the framework of algorithmic information dynamics [46] is based on the universal optimality of algorithmic probability [17, 19, 26, 31] and stems from the

(Felipe S. Abrahão) National Laboratory for Scientific Computing (LNCC), 25651-075, Petropolis, RJ, Brazil

(Hector Zenil) Alan Turing Institute, British Library, 2QR, 96 Euston Rd, London NW1 2DB. Oxford Immune Algorithmics, RG1 3EU, Reading, U.K.. Algorithmic Dynamics Lab, Unit of Computational Medicine, Department of Medicine Solna, Center for Molecular Medicine, Karolinska Institute, SE-171 77, Stockholm, Sweden. Algorithmic Nature Group, Laboratoire de Recherche Scientifique (LABORES) for the Natural and Digital Sciences, 75005, Paris, France.

E-mail addresses: `fsa@lncc.br, hector.zenil@cs.ox.ac.uk`.

F. S. Abrahão acknowledges the partial support from CNPq (301429/2021-9), INCT in Data Science – INCT-CiD (CNPq 465.560/2014-8), and FAPERJ (E-26/203.046/2017).

We thank Virginia Chaitin, Michael Winter, Alyssa Adams, Hyunju Kim, Sara I. Walker, Paul Davies, and Mikhail Prokopenko for suggestions, directions, and rewarding discussion on related topics investigated in this article.

discovered high convergence rate of computable generative models to the algorithmic probability [44]. Under this framework, the present article tackles the problem of quantifying emergence of algorithmic information in discrete deterministic dynamical systems and computable systems.

The challenge of formalizing the notion of *emergence* usually resides in the definition of what the term ”reducibility” (”derivable” or ”predictable”) means when one says that a macro-level phenomenon is not reducible to its micro-level parts or the initial conditions. In order to eliminate the possibility of an observer classifying a phenomenon as emergent, while another observer classifies it as reducible to its isolated parts (to the parts at a lower level scale or to the initial conditions), one adopted approach is to define emergence as a relative property [7, 9, 14, 28]. That is, as a property or quantifiable value that can only be evaluated by comparing the behavior of the system functioning as a whole at the macro-state level with the behavior of the system's parts at the micro-state level functioning isolated from each other. The mathematical and empirical problem is to guarantee that such a dependency on the observer cannot occur even in the case of formalizing emergence as a relative property of the whole in comparison to the parts [23]. In particular, we show that mathematical measures of emergent behavior as a relative property nevertheless depend on the formal theories that the observer previously knows.

In this regard, one of the paradigm shifts brought by algorithmic information dynamics concerning previous methods based on computability and information theory is that perturbation analysis enables our results to hold even if one allows the very act of observing to substantially change (or introduce "noise" into) the observed system's behavior. In fact, the act of observing can be formally defined as an interaction in which the system being observed perturbs the observer and the observer perturbs the system being observed, where the observer is a particular type of system that can compute functions and is equipped with a formal theory.

We show that, in the context of finite discrete deterministic dynamical systems, or computable systems in general, emergence of algorithmic information is invariant (i.e., the values can only vary up to a constant that does not depend on the object or system) and minimal (i.e., it is a minimal quantity for any chosen method of measuring the irreducible information content), but it is still dependent on the formal theory held by the observer. In addition, it is robust to variations of the arbitrarily chosen method of measuring irreducible information content, to error (or distortions) in the very act of observing, and to variations of the algorithmic-informational cost to process the observed system in accordance with the observer's formal knowledge. This kind of emergence is called *observer-dependent emergence* (ODE). Then, we investigate some previous definitions and measures of emergence in the literature that fall under ODE.

On the other hand, we show that systems that display an unbounded increase of emergent algorithmic information overcome this dependency. This is called *asymptotically observer-independent emergence* (AOIE). Then, in Theorems 4.1 and 4.2, which are the main results of this article, we demonstrate the existence of mathematical models in the literature that display AOIE.

To achieve the results, this article introduces new definitions, lemmas, and theorems. We also apply our theoretical results in order to analyze other examples and previous work in the literature regarding emergence in discrete deterministic dynamical systems and computable systems and regarding definitions of weak and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

strong emergence. Some previous mathematical definitions of emergence suffer from the dependency on the formal theory that the observer knows at the moment. One aspect that is remarkable about the AOIE is the fact that this is a type of emergence for which emergent phenomena will eventually remain emergent for every formal theory that one might devise in the future.

In Section 2, we study how algorithmic information content can be quantified in finite discrete deterministic dynamical systems and computable systems. In Section 3, we introduce *algorithmic perturbations*, *formal observer systems*, and the minimum requirements (stated as the *observation principle*) for the observation to take place. In Section 4.1, we introduce ODE and analyze two previous works. In Section 4.2, we introduce AOIE, demonstrate that two previous mathematical models in the literature display AOIE, and analyze those models in comparison to other approaches to weak and strong emergence. Section 5 concludes the article.

2. Algorithmic information content

2.1. **Encoded objects.** *Algorithmic complexity* defines an *invariant* and *minimal* measure of the information necessary for constructing an encoded object x with Turing machines, computable processes, or computable functions [16]. The (prefix) *algorithmic complexity* $\mathbf{K}(x)$ is the length of the shortest prefix-free (or self-delimiting) program (which is denoted by x^*) that outputs the object x in a universal prefix Turing machine (UTM) $\mathbf{U}$, i.e., $\mathbf{U}(x^*) = x$ and the length of $|x^*| = \mathbf{K}(x)$ is minimum. In this way, we measure the *size* $\mathbf{I_{ac}}(x)$ *of the algorithmic information content* (a.i.c.) of x as the *equivalence class* of integer values $k \in \mathbf{I_{ac}}(x)$ in the interval

$$|\mathbf{K}(x) - k| \leq c_{\mathbf{I}} \ ,$$

where $c_{\mathbf{I}} \in \mathbb{N}$ is an arbitrary and sufficiently large *object-independent* constant. Besides fixed and independent of the objects x, the value of the constant $c_{\mathbf{I}}$ is sufficiently large to encompass the other object-independent constants that appear: in the algorithmic coding theorem [17, 19, 26, 31]; in $\mathbf{I_A}(x; x^*) = \mathbf{I_A}(x^*; x) \pm \mathbf{O}(1) = \mathbf{K}(x) \pm \mathbf{O}(1)$; in $\mathbf{I_K}(x : x) = \mathbf{K}(x) - \mathbf{O}(1)$; and in any difference $|\mathbf{I_{ac}}(x) - \mu(x)| \leq \mathbf{O}(1)$, where $\mu(\cdot)$ is an arbitrary (semi-)measure of information content that an observer might choose to employ such that μ is equivalent to $\mathbf{K}(\cdot)$.

Note that $\mathbf{I_A}(w\,; z) = \mathbf{K}(z) - \mathbf{K}(z\,|w^*)$ is the *mutual algorithmic information* between the arbitrary strings w and z and $\mathbf{I_K}(w : z) = \mathbf{K}(z) - \mathbf{K}(z\,|w)$ is the $\mathbf{K}$-complexity of information in w about z [31]. Without loss of generality, but with a little abuse of notation, one can employ hereafter $\mathbf{K}(x) \pm c_{\mathbf{I}}$ to denote the equivalence class $\mathbf{I_{ac}}(x)$ in this article. Also note that the above definition of a.i.c. applies analogously to the size $\mathbf{I_{ac}}(z\,|w)$ of the conditional a.i.c. of z given w, which is an equivalence class of values in the interval

$$c_{\mathbf{I}} - \mathbf{K}(z\,|w) \leq \mathbf{K}(z\,|w) \leq \mathbf{K}(z\,|w) + c_{\mathbf{I}} \ ,$$

where the *conditional* prefix algorithmic complexity of a binary string z given a binary string w, denoted by $\mathbf{K}(z\,|w)$, is the length of the shortest program z^*_w such that $\mathbf{U}(\langle w, z^*_w \rangle) = z$ and $\langle \cdot \, , \, \dots \, , \, \cdot \rangle$ denotes any arbitrarily chosen encoding of tuples [26, 31].

The known properties of the algorithmic coding theorem and mutual algorithmic information in AIT are important to note. This is because the fact that the a.i.c. measure is a constant-bounded equivalence class $\mathbf{I_{ac}}(\cdot)$ (instead of a fixed value

 doi:10.20944/preprints202106.0019.v1

given by the the algorithmic complexity $\mathbf{K}(\cdot)$) directly implies that the values of $\mathbf{I_{ac}}(\cdot)$ in Definitions 4.1 and 4.2 are *invariant* and *minimal* with respect to a particular observer. As the reader will notice in Section 4, although perceiving the behavior of a particular system as emergent depends on the observer's formal knowledge, the values of a.i.c. still are invariant and minimal with respect to this observer.

2.2. **Dynamical systems.** For systems composed or defined by stochastic processes, emergence of information has been studied in terms of statistical information (for example, those based on entropy) or other related measures [27, 40]. Also in the context of stochastic processes, the quantification of synergy in multivariate stochastic systems has been an active field [32]. However, it is already known that statistics faces insuperable distortions when trying to quantify irreducible information content of deterministic processes [44]. For example, this is seen in the existence of Borel-normal sequences that are in fact computable (and, therefore, logarithmically compressible) [13] and in the existence of low-algorithmic-complexity graphs in which the degree-sequence entropy is maximal, and also Borel-normal [48]. Thus, if one is interested in measuring irreducible information content (or measuring the emergence of new irreducible information) in deterministic systems, which are free of stochasticity, employing any computable method only capable to characterize statistical regularities (in order to approximate a compressed form) will inherit these distortions. For this reason, the present article only addresses discrete deterministic dynamical systems or computable systems in general.

When moving from the context of encoded objects to discrete dynamical systems, the above properties of algorithmic information in Section 2.1 are important to understand in which formal sense algorithmic information is a measure of irreducible information content for dynamical systems. Let $\mathcal{S} = (X_{\mathcal{S}}, f_{\mathcal{S}}, E_{\mathcal{S}}, T)$ be a *finite discrete deterministic dynamical system* (FDDDS) embedded in an environment $\mathcal{E}$ [28], where $X_{\mathcal{S}}$ is the state space of $\mathcal{S}$,

$$\begin{array}{cccc} f_{\mathcal{S}} & : & X_{\mathcal{S}} \times E_{\mathcal{S}} \times T & \to & X_{\mathcal{S}} \\ & & (\mathcal{S}_t, e_{\mathcal{S}_t}, t) & \mapsto & \mathcal{S}_{t+1} \end{array}$$

is the function that defines the *evolution rule* (e.r.) of $\mathcal{S}$, $E_{\mathcal{S}}$ is the space of all possible environmental surrounding states that belong to the boundary of $\mathcal{S}$, and T is the set of time instants. If the cardinality of the set $E_{\mathcal{S}}$ of a dynamical system $\mathcal{S}$ is finite, then the dynamical system is said to have a finite *boundary*. If both sets $X_{\mathcal{S}}$ and $E_{\mathcal{S}}$ are composed only of discrete finite states, the finite-boundary dynamical system is said to be *finite* and *discrete*. In this article we only deal with dynamical systems $\mathcal{S}$ that are finite, discrete, deterministic, and have finite boundaries. The environment $\mathcal{E} = (X_{\mathcal{E}}, r_{\mathcal{E}}, T)$ is a FDDDS into which the systems $\mathcal{S}$ and their environmental surroundings $E_{\mathcal{S}}$ are embedded, where $r_{\mathcal{E}} : X_{\mathcal{E}} \times T \to X_{\mathcal{E}}$ is the e.r. of $\mathcal{E}$. In the case the e.r. of a dynamical system is a computable function, or computable relation, then the dynamical system is said to be *computable*.

We define the measure of the *size of the a.i.c. of a FDDDS* $\mathcal{S} = (X_{\mathcal{S}}, f_{\mathcal{S}}, E_{\mathcal{S}}, T)$ until time instant t by $\mathbf{I_{ac}}(\mathcal{S} \upharpoonright_0^t)$, where $\mathcal{S} \upharpoonright_0^t$ is just a notation for an arbitrary encoding of the sequence $(\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_t)$ of states (i.e., a state space trajectory of $\mathcal{S}$ until $t \in T$). It is straightforward to show that, with respect to the encoded object $\mathcal{S} \upharpoonright_0^t$, the equivalence class $\mathbf{I_{ac}}(\mathcal{S} \upharpoonright_0^t)$ inherits the same properties discussed in Section 2.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

It is known that TMs can be simulated by computable FDDDSs. For example, one can construct an *elementary cellular automata* (ECA) employing Rule 110 that simulates a TM [43]. In addition, the decision problem of one is reducible to a decision problem of the other and the time complexity of the TM simulation by ECAs can be improved to a polynomial time overhead [35].

Lemma 2.1. *Let $\mathcal{S}$ be a* computable *FDDDS such that* $\mathbf{U}\left(\langle \mathcal{S} \restriction_0^t, p\rangle\right) = y$ *and, for every* $t' < t$*, there is* z *such that* $\mathbf{U}\left(\left\langle \mathcal{S} \restriction_0^{t'}, p\right\rangle\right) = z$ *and* $z \neq y$ *. Then,*

$$\left|\mathbf{K}\left(y\right) - \mathbf{K}\left(\mathcal{S} \restriction_0^t\right)\right| \leq \mathbf{K}\left(p\right) + \mathbf{O}\left(1\right) \ .$$

Proof. From the minimality of $\mathbf{K}\left(\cdot\right)$, we have that $\mathbf{K}\left(y\right) \leq \mathbf{K}\left(\mathcal{S} \restriction_0^t\right) + \mathbf{K}\left(p\right) + \mathbf{O}\left(1\right)$. Since $\mathcal{S}$ is computable and $\mathbf{U}\left(\left\langle \mathcal{S} \restriction_0^{t'}, p\right\rangle\right) = z \neq y$ for every $t' < t$, then we also have that $\mathbf{K}\left(\mathcal{S} \restriction_0^t\right) \leq \mathbf{K}\left(y\right) + \mathbf{K}\left(p\right) + \mathbf{O}\left(1\right)$. □

Moreover, in the particular case the system $\mathcal{S}$ is simulating an arbitrary Turing machine w and the decision problem of $\mathcal{S}$ until t is *Turing equivalent* to the decision problem of $\mathbf{U}\left(w\right)$, Lemma 2.1 implies that one can *equivalently measure the a.i.c.* of $\mathcal{S}$ by $\mathbf{I_{ac}}\left(y\right)$ instead of $\mathbf{I_{ac}}\left(\mathcal{S} \restriction_0^t\right)$, where $\mathbf{U}\left(w\right) = y$. And the conditional case $\mathbf{I_{ac}}\left(\cdot|\cdot\right)$ applies analogously.

We can now obtain from Lemma 2.2 another property related to the predictability or computability of FDDDSs that will be important to the results in Section 4. As usual, let $f(x) = \mathbf{o}(g(x))$ denotes function g *strongly* dominating function f asymptotically.

Lemma 2.2. *Let $\mathcal{S}$ be a FDDDS for which*

$$\mathbf{K}\left(A\right) + \mathbf{K}\left(B\right) + \log\left(t\right) = \mathbf{o}\left(\mathbf{K}\left(\mathcal{S} \restriction_0^t\right)\right) \ ,$$

where $\mathbf{K}\left(A\right)$ and $\mathbf{K}\left(B\right)$ are the algorithmic complexity of encoded finite subsets $A \subset X_{\mathcal{S}}$ and $B \subset E_{\mathcal{S}}$, respectively. Then, for every program p and for every formal axiomatic theory $\mathbf{F}$, there is t' such that, for every $t \geq t'$,

$$\mathbf{U}\left(\langle t, \mathbf{F}, A, B, p\rangle\right) \neq \mathcal{S} \restriction_0^t \ .$$

Proof. The proof directly follows by contradiction from the fact that, for any fixed p and $\mathbf{F}$, if $\mathbf{U}\left(\langle |s|, \mathbf{F}, A, B, p\rangle\right) = s$, then $\mathbf{K}\left(s\right) \leq \mathbf{O}\left(\log\left(|s|\right)\right) + \mathbf{K}\left(A\right) + \mathbf{K}\left(B\right) + \mathbf{O}(1)$. □

In this regard, Hernandez-Orozco et al [28] achieves a more general result. The notion of a system getting adapted to the environment at infinitely many time instants is formalized in [28] as *weak convergence* of $\mathcal{S}$ toward $\mathcal{E}$, i.e., for an $\epsilon > 0$, there is an infinite set $T' \subset T$ such that, for every $t \in T'$, $\mathbf{K}\left(\mathcal{E}_{t+1} | \mathcal{S}_{t+1}\right) \leq \epsilon$. If $\mathcal{S}$ displays *strong Open-Endedness*, Hernandez-Orozco et al [28] demonstrates that one cannot even decide at which time instants the system $\mathcal{S}$ is getting adapted to the environment.

3. Observers and algorithmic perturbations

3.1. **Algorithmic perturbations.** Intuitively, an observation of $\mathcal{S}$ by the observer should be realized when the interaction between them somehow sends sufficient information about $\mathcal{S}$ to the observer. To formally tackle this problem within the scope of AID, any interaction between systems is reduced to a number of algorithmic perturbations. Note that the following Definition 3.1, with no loss of generalization

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

it general applies to finite discrete dynamical systems. FDDDSs are particular cases of finite discrete dynamical systems and computable FDDDSs are particular cases of FDDDSs.

Definition 3.1. *An* algorithmic perturbation $\mathcal{P}$ *is a perturbation of the state* A_t *of a finite discrete dynamical system* A *at some time instant* t *that replaces the next state* A_{t+1} *with another state* A'_{t+1}*, where* $\mathcal{P}$ *is a program and* $\mathbf{U}\left(\langle A_t, \mathcal{P} \rangle\right) = A'_{t+1}$.

Any finite state change in a finite discrete dynamical system A caused (or determined) by another finite state and e.r. of another finite discrete dynamical system B can be reduced to an equivalent algorithmic perturbation (AP) from B into A (see also Section 4.1.1).

Other particular examples of APs are stochastically random (destructive or constructive) perturbations on strings, networks, encoded objects, or systems in general. We employ here the term ‘stochastic randomness’ to distinguish it from algorithmic randomness. A stochastic random event is the one produced by a stochastic processes. Thus, a stochastically random perturbation is a perturbation that changes (i.e., deletes or inserts) the elements of a system according to a probability distribution [46, 47, 49]. Hence, it follows that stochastically random perturbations are particular types of algorithmic perturbations because, for every change in the elements of an object or system, there is an AP $\mathcal{P}$ that executes the same exact changes. In the case of networks, we have that stochastic-randomly deleting (or inserting) $|F|$ edges in a network G, which results in a new network G', is equivalent to applying an AP $\mathcal{P}_F$ on G such that $\mathbf{K}\left(\mathcal{P}_F\right) \leq 2\,|F|\log_2\left(N\right) + \mathbf{O}\left(\log_2\left(|F|\right)\right) + \mathbf{O}(\log_2\left(\log_2(N)\right))$ and $\mathbf{U}\left(\langle G, \mathcal{P}_G \rangle\right) = G'$ [45, 46], where F is the subset of edges that were perturbed and N is the number of vertices. As one of the important properties implied by this equivalence in algorithmic information dynamics, a stochastically random perturbation on a single edge can only change the final algorithmic complexity of the network by $\mathbf{O}(\log_2(N))$ bits, which gives rise to a thermodynamic-like behavior of the reprogrammability of networks when these are subjected to stochastically random single-edge perturbations [49].

3.2. **Formal observer systems.** The act of observing is an act of a system (the observer) perturbing the object (another system), while being perturbed by the object. Thus, following the framework in Section 2.2, we define a *formal observer system* (FOS) as a particular type of FDDDS $\mathcal{S}$, which we will denote by $\mathcal{O}$ in order to avoid ambiguity with the general FDDDSs denoted by $\mathcal{S}$. The mathematical challenge is then to establish the necessary conditions for this mutual perturbation to result in sufficient knowledge acquisition by the observer. In this sense, we first introduce a variation of Turing machines (TMs), called *observer Turing machine*, that are going to be simulated by the FOSs. Then, we present the *observation principles.*

An *observer Turing machine* (OTM) O is a slight variation of the usual 2-tape Turing machine [30, 41], where $O(x_1, x_2)$ denotes the output of O with x_1 in the first tape and x_2 in the second tape. The *first tape* works as the first tape of the 2-tape Turing machine, but the *second tape* is filled by an encoded form of a *formal axiomatic theory* (FAT) $\mathbf{F}$ before the OTM starts to run. In other words, the OTM is a 2-tape Turing machine with access to a finite-size oracle, and this oracle is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

precisely the encoding of a FAT, where $O(x, \mathbf{F})$ denotes the output of O with input x when the formal theory $\mathbf{F}$ is previously known.

Since $\mathbf{F}$ is always finite, it is straightforward to show that, for every OTM, there is an equivalent single-tape Turing machine that simulates the OTM. Let M be a *non-halting* TM that repeatedly simulates an OTM such that: M does not halt, if the OTM does not halt; and M indefinitely repeats the simulation of the OTM from the beginning, if the OTM halts. Moreover, we know that one can construct a FDDDS that simulates M (see Section 2.2). This universal FDDDS can be constructed so that the decision problem of the OTM is Turing equivalent to the decision problem of the universal FDDDS. As a consequence, one can define a formal observer system as a FDDDS that simulates M in which the set T is infinite:

Definition 3.2. *A* formal observer system *(FOS) $\mathcal{O}$ is defined as a FDDDS that simulates the non-halting TM M which repeatedly simulates an OTM.*

Thus, $\mathcal{O}$ is a FDDDS that extends indefinitely in time either because the OTM does not halt or because it is simulating a halting OTM in an infinite number of repeated *simulation cycles*. In particular, if the OTM halts, then $\mathcal{O}$ is a FDDDS whose *recurrence time* [9] (see also Section 4.1.1) corresponds to the time steps necessary for the simulation of the OTM by the FDDDS. In other words, if the OTM halts, simulation cycles correspond to contiguous recurrent state space trajectories.

The construction of the FOSs allows us to formalize in Definition 3.3 the notion of a function being computed by a state space trajectory, which will be proved to be satisfiable in Lemma 3.1.

Definition 3.3. *Let $\mathcal{O}$ be a FOS and let f be a computable function. We say* $(\mathcal{O}_0, \dots, \mathcal{O}_t, \dots, \mathcal{O}_{t'}, \dots)$ computes *the value of $f(x)$ (or* decide *a problem) between time instants t and t' iff the state space trajectory $(\mathcal{O}_t, \dots, \mathcal{O}_{t'})$ corresponds to the simulation cycle of the OTM that computes the function f with input x (or decide the problem).*

This way, the (time or space) *computational class* of $\mathcal{O}$ becomes the (time or space) complexity class of the OTM.

A direct consequence of Definitions 3.1, 3.2, and 3.3 is that they enable an AP to change the machine that $\mathcal{O}$ is simulating. If an AP at time instant t replaces $\mathcal{O}_{t+1}$ with another state $\mathcal{O}'_{t+1}$ such that $\mathcal{O}'_{t+1}$ is the initial state of the universal FDDDS that simulates the same OTM of $\mathcal{O}$ but with the first tape containing w, then the state space trajectory $\left(\mathcal{O}'_{t+1}, \mathcal{O}'_{t+2}, \dots\right)$ simulates the OTM with input w in the first tape. If the OTM with input w computes a function $f(w)$, then $\left(\mathcal{O}'_{t+1}, \mathcal{O}'_{t+2}, \dots\right)$ also computes a function $f(w)$ after t time steps. For short, we refer to the first tape and the second tape of the OTM that $\mathcal{O}$ is simulating as the *first tape* and *second tape* of $\mathcal{O}$, respectively.

All of these above results and remarks lead us to the following Lemma 3.1 about FOSs that can receive external information (i.e., being algorithmically perturbed) over time. Lemma 3.1 also demonstrates that Definition 3.3 is satisfiable.

Lemma 3.1. *For every w_1, w_2, p with $\mathbf{U}(\langle w_1, p\rangle) = y_1$ and $\mathbf{U}(\langle w_2, p\rangle) = y_2$, there is a FOS $\mathcal{O}$, an AP $\mathcal{P}$, and a time instant $t \in T$ such that $\left(\mathcal{O}_0, \dots, \mathcal{O}_t, \mathcal{O}'_{t+1}, \mathcal{O}'_{t+2}, \dots\right)$ computes y_1 until time instant t and computes y_2 after time instant t.*

Proof. Let A be a universal FDDDS that simulates any TM p. Let M' be a single-tape TM that simulates a 2-tape TM with x_1 in the first tape and x_2 in the second tape. Let M be the TM that receives any TM p', simulates p' and: if p' reaches a halting state, writes p' in the tape again with the head on the first cell at the same initial state of M; otherwise, the simulation of p' continues indefinitely. Therefore, we have that $M(p')$ is undefined for every p'. Now, let $\mathcal{O}$ be a FDDDS that has the same e.r. of A and the initial state $\mathcal{O}_0$ corresponds to the initial configuration of A simulating $M(M'(w_1, p))$. Then, $\mathbf{U}(\langle w_1, p\rangle) = y_1$ implies that there are $t' \geq 0$ and $k \geq 1$ such that $(\mathcal{O}_0, \dots, \mathcal{O}_{t'})$ computes y_1 and $\mathcal{O}_{t'+k} = \mathcal{O}_0$. Let $\mathcal{O}'_0$ be the initial state that corresponds to the initial configuration of A simulating $M(M'(w_2, p))$. Since both $\mathcal{O}'_0$ and $\mathcal{O}_{t'+k}$ are finite states, we have that there is an AP $\mathcal{P}$ at time instant $t = t' + k - 1$ such that $\mathbf{U}(\langle \mathcal{O}_t, \mathcal{P}\rangle) = \mathcal{O}'_{t'+k}$. Therefore, the resulting state space trajectory $\left(\mathcal{O}_0, \dots, \mathcal{O}_t, \mathcal{O}'_{t+1}, \mathcal{O}'_{t+2}, \dots\right)$ computes y_1 until time instant t and computes y_2 after time instant t. □

3.3. **Observation principle.** The main idea of the *Observation Principle* 1 is that a proper observation occurs when both $\mathcal{O}$ and $\mathcal{S}$ are perturbing each other in such a way that the perturbation on the state $\mathcal{O}_t$ introduces sufficient information about $(\mathcal{S}_{t-k}, \dots, \mathcal{S}_t)$ and $\mathcal{S}'_{t+1}$ in the first tape of $\mathcal{O}$. The condition that the AP needs to give information about the post-perturbation future state $\mathcal{S}'_{t+1}$ is necessary. This is because, otherwise, $\mathcal{O}$ could have acquired sufficient algorithmic information about $(\mathcal{S}_{t-k}, \dots, \mathcal{S}_t)$ to compute $(\mathcal{S}_{t-k}, \dots, \mathcal{S}_t, \mathcal{S}_{t+1})$, while not acquired sufficient to compute $\mathcal{S}'_{t+1}$. In other words, the APs in the act of observing can be informative about the past of the observed system at the same time that it is not sufficient to determine what changes the very act of observing caused on the observed system. Informally, this means that (after the algorithmic perturbation in the act of observing) the system being observed would actually become substantially distinct from what the observer "thinks" the system should be and, therefore, the act of observing would not have given sufficient information about the observed system to the observer.

Analogous to $\mathcal{S} \restriction_0^t$ in Section 2.2, let $\mathcal{S} \restriction_t^{t'}$ denotes an arbitrarily chosen encoding of the state space trajectory $(\mathcal{S}_t, \dots, \mathcal{S}_{t'})$ of $\mathcal{S}$ between time instants t and t'. For short, let $\left(\mathcal{S} \restriction_{t-k}^t, \mathcal{S}'_{t+1}\right)$ denote the state space trajectory $\left(\mathcal{S}_{t-k}, \dots, \mathcal{S}_t, \mathcal{S}'_{t+1}\right)$ and $\left\langle \mathcal{S} \restriction_{t-k}^t, \mathcal{S}'_{t+1}\right\rangle$ denote any arbitrarily chosen encoding of the sequence $\left(\mathcal{S} \restriction_{t-k}^t, \mathcal{S}'_{t+1}\right)$.

Observation principle 1. *Let O be the OTM that $\mathcal{O}$ is simulating. Let $c_{\mathcal{O}} \in \mathbb{N}$ be a constant that only depends on $\mathcal{O}$ and does not depend on $\mathcal{S}$. Let $\mathcal{P}_{(\mathcal{O},\mathcal{P},t)}$ be an AP from $\mathcal{O}$ into $\mathcal{S}$ at time instant t. Let $\mathcal{P}_{(\mathcal{P},\mathcal{O},t)}$ be an AP from $\mathcal{S}$ into $\mathcal{O}$ at time instant t. We say $\mathcal{O}$* observes *the past $k+1$ states of $\mathcal{S}$ at time instant t if $\mathcal{P}_{(\mathcal{P},\mathcal{O},t)}$ causes a bit string w to be encoded into the first tape of O such that*

$$\mathbf{K}\left(\left\langle \mathcal{S} \restriction_{t-k}^t, \mathcal{S}'_{t+1}\right\rangle \middle| \left\langle w, \mathcal{O} \restriction_0^t\right\rangle\right) \leq c_{\mathcal{O}} \ , \tag{1}$$

where $\mathcal{S}'_{t+1} = \mathbf{U}\left(\left\langle \mathcal{S}_t, \mathcal{P}_{(\mathcal{O},\mathcal{P},t)}\right\rangle\right)$.

It follows directly from the basic properties in algorithmic information theory that the constant $c_{\mathcal{O}}$ sets the error margin for the extent in which the mutual algorithmic information is preserved between the system being observed and the information obtained by the observer during the observation. The smaller the value of $c_{\mathcal{O}}$, the more *mutual algorithmic information* (and also $\mathbf{K}$-complexity of information) is preserved. Formally, if the Observation Principle 1 is satisfied, then

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

we have that

$$\begin{aligned}\mathbf{K}\left(\left\langle \mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}'_{t+1} \right\rangle\right) - c_{\mathcal{O}} &\leq \mathbf{I_K}\left(\left\langle w, \mathcal{O} \restriction_{0}^{t} \right\rangle : \left\langle \mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}'_{t+1} \right\rangle\right) \leq \\ &\leq \mathbf{I_A}\left(\left\langle w, \mathcal{O} \restriction_{0}^{t} \right\rangle ; \left\langle \mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}'_{t+1} \right\rangle\right) + \mathbf{O}(1) \leq \\ &\leq \mathbf{K}\left(\left\langle \mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}'_{t+1} \right\rangle\right) + \mathbf{O}(1) \ . \end{aligned} \tag{2}$$

Thus, the Observation Principle 1 is general enough to encompass the case in which observation takes place, but it is defective. That is, when $\mathcal{O}$ observes $\mathcal{S}$ at time instant t and gets only partial information about $\mathcal{S}$. In other words, this defective information about $\mathcal{S}$ can differ from the actual information about $\mathcal{S}$, but only up to a bounded error margin (which is given by the constant $c_{\mathcal{O}}$).

4. Emergence of algorithmic information

4.1. **Observer-dependent emergence.** Intuitively, emergence of algorithmic information occurs when the formal theory known by the observer is not sufficient for computing, predicting, or completely explaining the object's future behavior from its constituent parts or previous conditions. In the process of trying to explain or predict the behavior of an observed system (or object), the observer employs the resources available, its own previously known formal knowledge, and the information it could gather from the observation.

The *main idea* of Definition 4.1 is that, even taking into account the equivalent ways to measure the irreducible information content (by the presence of the constant $c_{\mathbf{I}}$), the error margin of defective information in the observation (by the presence of the constant $c_{\mathcal{O}}$), and the algorithmic-informational cost of processing all the previous information that the observer has (by the presence of the constant c_e), there is an insufficient amount of algorithmic information that is still necessary to compute the future behavior of the observed system.

For short, extending the same notation we employed in Section 3.3, let $\left(\mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}' \restriction_{t+1}^{t'}\right)$ denote the state space trajectory $(\mathcal{S}_{t-k}, \dots, \mathcal{S}_t, \mathcal{S}'_{t+1}, \dots, \mathcal{S}'_{t'})$ and $\left\langle \mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}' \restriction_{t+1}^{t'}\right\rangle$ denote any arbitrarily chosen encoding of the sequence $\left(\mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}' \restriction_{t+1}^{t'}\right)$. Also consistently with our notation, we employ $(\mathcal{S} \restriction_{t}^{t})$ to denote the single state $\mathcal{S}_t$ at time instant t, i.e., $(\mathcal{S} \restriction_{t}^{t}) = \mathcal{S}_t$ for every t.

Definition 4.1 (Observer-dependent emergence). *Let $t' \geq t+m$, where $m \geq 1$. Let O be the OTM that $\mathcal{O}$ is simulating. Let $c_e > 0$ be a constant that may depend on $\mathcal{O}$, but does not depend on $\mathcal{S}$. A state space trajectory $\left(\mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}' \restriction_{t+1}^{t'}\right)$ is* emergent with respect to the observer $\mathcal{O}$ *after the observation of the $k+1$ past states of $\mathcal{S}$ at time instant t if*

$$\mathbf{I_{ac}}\left(\left\langle \mathcal{S} \restriction_{t-k}^{t}, \mathcal{S}' \restriction_{t+1}^{t'} \right\rangle \,\middle|\, \left\langle w, \left(\mathcal{O} \restriction_{0}^{t}, \mathcal{O}' \restriction_{t+1}^{t+m}\right)\right\rangle\right) > c_{\mathbf{I}} + c_{\mathcal{O}} + c_e \ , \tag{3}$$

where w is the bit string in the first tape of O that satisfies the Observation Principle 1 at time instant t.

The constant c_e can be arbitrarily large with the purpose of allowing high-algorithmic-complexity programs to process the previous information and try to compute the future behavior. However, once fixed its value, it does not depend on the choice of the observed system (or object). It can only depend on the observer.

 doi:10.20944/preprints202106.0019.v1

This way, since the three constants $c_{\mathbf{I}}$, $c_{\mathcal{O}}$, and c_e are already counted for a fixed observer, the *invariance* of $\mathbf{I_{ac}}\left(\left\langle \mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'} \right\rangle \middle| \left\langle w, \left(\mathcal{O} \upharpoonright_{0}^{t}, \mathcal{O}' \upharpoonright_{t+1}^{t+m}\right)\right\rangle\right)$ follows directly from the invariance mentioned in Section 2.1. In the same manner, the *incompressibility* or *irreducibility* of $\mathbf{I_{ac}}\left(\left\langle \mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'} \right\rangle \middle| \left\langle w, \left(\mathcal{O} \upharpoonright_{0}^{t}, \mathcal{O}' \upharpoonright_{t+1}^{t+m}\right)\right\rangle\right)$ follows directly from the minimality mentioned in Section 2.1.

At first glance, since the constant c_e can be arbitrarily large, one might think that it is possible to cancel any presence of necessary extra algorithmic information to compute the future behavior of $\mathcal{S}$. This does not hold, i.e., no matter how large one chooses the value of c_e to be, there will be FDDDSs for which Equation 3 holds. One can always construct a state space trajectory of finite states whose global algorithmic information content is as larger as one wants. More formally, for every ϵ, there is $\left(\mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'}\right)$ such that $\mathbf{K}\left(\left\langle \mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'}\right\rangle\right) > \epsilon$.

For example, this can be achieved by ordering the finite-size states $\mathcal{S}_t$'s in such a way that the resulting sequence is as closer as one wants to be an algorithmically random sequence. Another way to obtain high-algorithmic-complexity state spaces trajectories is with FDDDSs whose states $\mathcal{S}_t$ themselves have a larger algorithmic information content. That is, although the sequence of states can be highly redundant (i.e., there is ϵ' such that $\mathbf{I_A}\left(\mathcal{S}_t; \mathcal{S}_{t'}\right) < \epsilon'$ for any t and t'), one can always construct a set $X_{\mathcal{S}}$ such that $\min\left\{\mathbf{K}\left(\mathcal{S}_t\right) \middle| \mathcal{S}_t \in X_{\mathcal{S}}\right\} > \epsilon + \mathbf{O}(1)$, which immediately implies that $\mathbf{K}\left(\left\langle \mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'}\right\rangle\right) > \epsilon$.

From any of these examples of constructing or obtaining a FDDDS with a larger global algorithmic information content, one directly employs Lemma 2.2 to show that, for any c_e, $\mathcal{O}$, $c_{\mathcal{O}}$ and $c_{\mathbf{I}}$, there is $\mathcal{S}$ and time instant t' for which the state space trajectory $\left(\mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'}\right)$ cannot be computed within the range $c_{\mathbf{I}} + c_{\mathcal{O}} + c_e$ of algorithmic information. Thus, for any observer, there are FDDDSs whose state space trajectories are ODE as in Definition 4.1.

Other important property is that the fact that $\left(\mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'}\right)$ satisfies Definition 4.1 for an observer $\mathcal{O}_1$ does not imply that it will always satisfy for another observer $\mathcal{O}_2$. In other words, a sequence $\left(\mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'}\right)$ being emergent (as in Definition 4.1) to an observer does not imply that it will be emergent to another observer. To demonstrate that there is such another observer $\mathcal{O}_2$ that can compute $\left(\mathcal{S} \upharpoonright_{t-k}^{t}, \mathcal{S}' \upharpoonright_{t+1}^{t'}\right)$, it suffices to extend the FAT $\mathbf{F}_1$ of $\mathcal{O}_1$ with an encoded form of the interpretation of the needed $c_{\mathbf{I}} + c_{\mathcal{O}} + c_e + \mathbf{O}(1)$ bits of algorithmic information in the language of $\mathbf{F}_1$, resulting in a FAT $\mathbf{F}_2$, and define $\mathcal{O}_2$ as simulating the same OTM of $\mathcal{O}_1$ but with $\mathbf{F}_2$ in its second tape. That is, information can always be converted into an extension of a FAT and, then, converted into new formal knowledge about what was supposed to be emergent. Whenever and wherever there is a sufficient finite amount of algorithmic information that can be employed to compute the behavior of a system, that emergence is in fact dependent on the observer's previous formal knowledge. Thus, for any emergence that results from a lack of finite algorithmic information—this way, satisfying Definition 4.1—, one has a type of *observer-dependent emergence* (ODE).

Following the equivalence of Turing machines and computable FDDDSs from Lemma 2.1 as discussed in Section 2.2, Definition 4.1 can assume the alternative

 doi:10.20944/preprints202106.0019.v1

form by only replacing Equation 3 with

$$\mathbf{I_{ac}}\left(\mathbf{U}\left(\langle t', p\rangle\right) \mid \langle w, t+m, O\rangle\right) > c_{\mathbf{I}} + c_{\mathcal{O}} + c_e \tag{4}$$

in the case p is an observed TM, O is an OTM, and w is the information received by O at time instant t.

4.1.1. *Models and examples of observer-dependent emergence.* In [14], a weakly emergent phenomenon is defined as the one for which the macro-level states of a system can only be derivable by simulating the very system. Later, Bedau [15] refines the notion of derivability in this definition with the notion of explanatory incompressibility. For example, in Conway's Game of Life one cannot in general decide whether or not a macrostate behavior will have a certain property from the initial configurations. Only by simulating the game it would be possible to gain sufficient irreducible information about whether or not the macrostate behavior has a property in general. One can easily demonstrate that such a "simulation irreducibility via explanatory incompressibility" in [14, 15] is implied by Definition 4.1. To this end, suppose there is a way for the FOS $\mathcal{O}$ to simulate and compute the state space trajectory $\left(\mathcal{S}\restriction_{t-k}^{t}, \mathcal{S}'\restriction_{t+1}^{t'}\right)$ of a FDDDS $\mathcal{S}$ in the $t+m$ time steps of $\mathcal{O}$. Then, we would have that there is a constant c_e such that

$$\mathbf{I_{ac}}\left(\left\langle \mathcal{S}\restriction_{t-k}^{t}, \mathcal{S}'\restriction_{t+1}^{t'}\right\rangle \middle| \left\langle w, \left(\mathcal{O}\restriction_{0}^{t}, \mathcal{O}'\restriction_{t+1}^{t+m}\right)\right\rangle\right) \leq c_{\mathbf{I}} + c_{\mathcal{O}} + c_e \ ,$$

which directly contradicts Definition 4.1 for this constant c_e. That is, the ODE in Definition 4.1 implies the weak emergence in [14, 15]. In the opposite direction, it is straightforward to show that the weak emergence in [14, 15] implies that there are $\mathcal{O}$, c_e and $t' > t+m$ for which Definition 4.1 is satisfied.

It is claimed in [14] that the simulation irreducibility is a property that is not dependent on the current limited knowledge of the observer. However, we can now employ Definition 4.1 to demonstrate that this claim is not precisely true. Let $\mathcal{O}_1$ and $\mathcal{S}$ be FDDDSs that, for infinitely many p, there are t and t' with

$$t' \geq \mathbf{K}\left(t'\right) > \left(t+m\right)^3 + \mathbf{K}\left(\left\langle w, \left(\mathcal{O}_1\restriction_{0}^{t}, \mathcal{O}_1'\restriction_{t+1}^{t+m}\right)\right\rangle\right) + c_{\mathbf{I}} + c_{\mathcal{O}} + |p| + \mathbf{O}(1) \tag{5}$$

and $m > 1$ such that

$$\mathbf{I_{ac}}\left(\left\langle \mathcal{S}\restriction_{t-k}^{t}, \mathcal{S}'\restriction_{t+1}^{t'}\right\rangle \middle| \left\langle w, \left(\mathcal{O}_1\restriction_{0}^{t}, \mathcal{O}_1'\restriction_{t+1}^{t+m}\right)\right\rangle\right) > c_{\mathbf{I}} + c_{\mathcal{O}} + |p| + \mathbf{O}(1) \ .$$

Clearly, $\mathcal{O}_1$ cannot always predict or compute the state space trajectory $\left(\mathcal{S}'\restriction_{t+1}^{t'}\right)$ during the $t+m$ time steps of $\mathcal{O}_1$. However, for any p, t, and t' for which this holds, we already showed in Section 4.1 how one can construct another $\mathcal{O}_2$ that can compute $\left(\mathcal{S}\restriction_{t-k}^{t}, \mathcal{S}'\restriction_{t+1}^{t'}\right)$ in $t+m$ time steps. In addition, since Equation 5 holds, then neither $\mathcal{O}_1$ nor $\mathcal{O}_2$ can be simulating $\left(\mathcal{S}\restriction_{t-k}^{t}, \mathcal{S}'\restriction_{t+1}^{t'}\right)$ during the $t+m$ time steps in first place.

Thus, since the "simulation irreducibility via explanatory incompressibility" in [14, 15] implies that there are conditions in which Definition 4.1 is satisfied, and vice-versa, we defend that the weak emergence described in [14, 15] can be understood as an informal alternative, but *equivalent*, to the ODE captured by Definition 4.1.

We will also demonstrate below that the ODE in Definition 4.1 implies *unbounded evolution* (UE) and *innovation* (INN) in [8, 9].

 doi:10.20944/preprints202106.0019.v1

If the interaction of a (finite) dynamical system A with its environment $\mathcal{E}$ (i.e., another finite dynamical system) gives rise to a recurrent state space trajectory whose length is larger than the length of all the other recurrent state space trajectories of any isolated system with the same size of A, the pair of systems A and $\mathcal{E}$ is said to exhibit UE. If the new emerging recurrent state space trajectory from the interaction between A and $\mathcal{E}$ is not contained in none of the other recurrent state space trajectories of any isolated system with the same size of A, the pair of systems A and $\mathcal{E}$ is said to exhibit INN. In addition, since the interaction with the environment can introduce state-dependent changes in the evolution rules of the system A, the increase of the recurrence time shown in the models investigated in [9] can be classified as an example of emergence through downward (or top-down) causation [9, 25, 42].

Note that, in the models of cellular automata (CA) in [9], the interaction with the environment can produce changes, i.e., perturbations, in the e.r. of system A. So, it differs from the AP in Definition 3.1 because the latter impacts the states of the system A instead of its e.r.. In fact, one can reduce each state-dependent rule perturbation on a CA A in [9] to an equivalent AP on an equivalent universal CA that emulates A. To this end, just note that every finite CA is computable by a TM and there are universal cellular automata (for example, ECA Rule 110) that can simulate any Turing machine. Thus, for each state-dependent rule perturbation in [9], one constructs the equivalent AP in the same manner as in the proof of Lemma 3.1.

Hence, in order to show that Definition 4.1 implies UE and INN, it suffices to choose a large enough constant c_e that depends on the observer $\mathcal{O}$ but does not depend on $\mathcal{S}$ nor $\mathcal{E}$, where both $\mathcal{S}$ and $\mathcal{E}$ are FDDDSs. Let $n_e = \max \{ |\mathcal{E}_t| \, | t \in T \}$ be the size of $\mathcal{E}$ and $n_s = \max \{ |\mathcal{S}_t| \, | t \in T \}$ be the size of $\mathcal{S}$. In the particular case of ECAs, then n_e and n_s become the respective widths of each ECA. Let $|X_{\mathcal{S}}|$ be the maximum number of distinct states of a system $\mathcal{S}$ and $|X_{\mathcal{E}}|$ be the maximum number of distinct states of the environment $\mathcal{E}$. In the case of ECAs, we would have $|X_{\mathcal{S}}| = 2^{n_s}$ and $|X_{\mathcal{E}}| = 2^{n_e}$. Thus, the maximum number of possible state space trajectories is upper bounded by $|X_{\mathcal{S}}|^{t_r - t + 1}$, where t is the time instant at which the observation of the initial conditions of isolated $\mathcal{S}$ occurred and t_r is the recurrence time of $\mathcal{S}$ when it is interacting with $\mathcal{E}$. Now, let t_p be the largest recurrence time of any isolated $\mathcal{S}$ with $t < t_p$. Finally, we choose a constant

$$c_e > |p'| + \mathbf{O}\left(\log_2\left(|X_{\mathcal{S}}|^{t_p - t + 1}\right)\right) + \mathbf{O}\left(\log_2\left(t_p\right)\right) + \mathbf{O}\left(\log_2\left(t_r\right)\right) \ , \tag{6}$$

where p' is the program that reads the index of a recurrent state space trajectory of length $t'' \leq t_p$ (index which is encoded in $\mathbf{O}\left(\log_2\left(|X_{\mathcal{S}}|^{t_p - t + 1}\right)\right)$ bits) and returns the respective state space trajectory by extending and repeating the t'' update steps (where t'' is an integer encoded in $\mathbf{O}\left(\log_2\left(t_p\right)\right)$ bits) until the t_r-length sequence of states is completed (where t_r is an integer encoded in $\mathbf{O}\left(\log_2\left(t_r\right)\right)$ bits). Clearly, once t_r can be arbitrarily larger than t_p, if a state space trajectory of $\mathcal{S}$ interacting with $\mathcal{E}$ satisfies Definition 4.1 with such a constant c_e, where $k = 0$ and $t' = t_r$, then it cannot be any recurrent state space trajectory of length $\leq t_p$ and, therefore, this state space trajectory satisfies both the definitions of UE and INN in [9].

Also note that such a program p' can be employed to prove that UE and INN define a type of emergence that is dependent on the observer's previous knowledge. To this end, just note that, for any recurrent state space trajectory of length t_r,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

there is a constant $c_e \leq |p'| + \mathbf{O}\left(\log_2\left(|X_{\mathcal{S}}|^{t_r - t + 1}\right)\right) + \mathbf{O}\left(\log_2\left(t_r\right)\right)$ that negates Equation 3 (with $k = 0$ and $t' = t_r$).

Thus, the kind of emergence from UE and INN is dependent on the observer's formal knowledge and it is implied by the ODE in Definition 4.1. With the purpose of showing that both approaches are equivalent, it is important to remark that models displaying an empirical tendency of increase in the algorithmic complexity were investigated in [9]. In this line of research, the inverse problem (that is, to prove that a FDDDS displaying UE and INN *always* implies that there is at least one constant c_e for which the state space trajectory of Definition 4.1 is satisfied) is a necessary theoretical future research.

4.2. **Asymptotically observer-independent emergence.** The ODE studied in Section 4.1 establishes a formal mathematical characterization of emergence that takes into account observation errors or distortions, the arbitrarily chosen mathematical method of measuring information content, and any information cost to process the previous knowledge and the information gathered in the act of observing. The next natural question arises whether such an approach can be extended to formalize an emergent phenomenon that continues to be emergent for any observer. In this regard, the present section tackles the problem by presenting a precise definition and demonstrating that *asymptotically observer-independent emergence* (AOIE) of algorithmic information exists in mathematical models.

First, we naturally extend our notation so we can further explore the necessity of extra algorithmic information when a sufficiently large amount of time steps has passed. In order to avoid ambiguities, let $\mathbf{S} = \left(\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \dots, \mathcal{S}^{(k)}, \dots\right)$ denote a *discrete deterministic dynamical system* that is composed of a sequence of nested FDDDSs, where: the FDDDS $\mathcal{S}^{(i+1)}$ is an extension of the FDDDS $\mathcal{S}^{(i)}$ for every $i > 0$; and the transition from $\mathcal{S}^{(i)}$ to $\mathcal{S}^{(i+1)}$ occurs at time instant $t_{i+1} - 1$. Formally, the e.r. $f_{\mathbf{S}}$ of $\mathbf{S}$ is such that, for every $i > 0$ with $t_{i+1} - 1 > t \geq t_i$, one has

$$f_{\mathbf{S}}\left(\left(\mathcal{S}^{(i)} \upharpoonright_t^t\right), \left(e_{\mathcal{S}^{(i)}} \upharpoonright_t^t\right), t\right) = f_{\mathcal{S}^{(i)}}\left(\left(\mathcal{S}^{(i)} \upharpoonright_t^t\right), \left(e_{\mathcal{S}^{(i)}} \upharpoonright_t^t\right), t\right)$$

and

$$f_{\mathbf{S}}\left(\left(\mathcal{S}^{(i)} \upharpoonright_{t_{i+1}-1}^{t_{i+1}-1}\right), \left(e_{\mathcal{S}^{(i)}} \upharpoonright_{t_{i+1}-1}^{t_{i+1}-1}\right), t_{i+1} - 1\right) = \left(\mathcal{S}^{(i+1)} \upharpoonright_{t_{i+1}}^{t_{i+1}}\right) .$$

Consistently with the notation, note that a FDDDS $\mathcal{S}$ is a special case of $\mathbf{S}$ in which $\mathbf{S} = (\mathcal{S}, \mathcal{S}, \dots)$.

Now, we formalize emergence of algorithmic information that is able to eventually surpass any previous formal knowledge that the observer held or might come up with. The main idea of Definition 4.2 is that, for any equivalent way to measure the irreducible information content (by the presence of the constant $c_{\mathbf{I}}$), error margin of defective information in the observation (by the presence of the constant $c_{\mathcal{O}}$), and algorithmic-informational cost of processing all the previous information that the observer has (by the presence of the constant c_e), there is a certain stage of the system $\mathbf{S}$ from which a lacking amount of algorithmic information begins to be necessary to compute the future behavior of the observed system. In other words, for every FOS, there is a certain stage in which the subsequent behavior of $\mathbf{S}$ begins to display the ODE (as in Definition 4.1).

Definition 4.2 (Asymptotically observer-independent emergence)**.** *Let* $\mathbf{S}$ *be an arbitrary discrete deterministic dynamical system well defined for every time instant* *(i.e., $|T| = \infty$). A system* $\mathbf{S}$ *is* asymptotically observer-independently emergent *if, for every $\mathcal{O}$, for every time instant t, and for every $c_{\mathbf{I}}, c_{\mathcal{O}}, c_e > 0$, there is $t_e \in T$ such that, for every $t' \geq t_e$, one has that*

$$\mathbf{I_{ac}}\left(\left\langle \mathbf{S} \restriction_{t-k}^{t}, \mathbf{S}' \restriction_{t+1}^{t'} \right\rangle \,\middle|\, \left\langle w, \left(\mathcal{O} \restriction_{0}^{t}, \mathcal{O}' \restriction_{t+1}^{t+m}\right)\right\rangle\right) > c_{\mathbf{I}} + c_{\mathcal{O}} + c_e \ , \quad (7)$$

where $t' \geq t + m$, $m \geq 1$, and w is the bit string in the first tape of O that satisfies the Observation Principle 1 at time instant t.

Clearly, the AOIE in Definition 4.2 inherits the *invariance* and *minimality* of the ODE in Definition 4.1. The distinctive aspect now is the fact that, although there might be a FOS that can compute a finite-length state space trajectory of $\mathbf{S}$, the AOIE guarantees that this will eventually cease to happen. For this reason, the emergence of algorithmic information in Definition 4.2 is guaranteed to be *independent* of the observer, but only in the *asymptotic limit*. The existence of the time instant t_e assures that there is a phase transition for which, if the behavior of $\mathbf{S}$ is not emergent to a certain FOS, then it will start to be emergent after time instant t_e. Indeed, at each stage of $\mathbf{S}$, since it is a FDDDS, the strongest emergence that $\mathbf{S}$ can display is the ODE in Definition 4.1, but eventually any other FOS that might try to compute the behavior of $\mathbf{S}$ will be overcome after time instant t_e. We shall demonstrate in the next Sections 4.2.1 and 4.2.2 that there are indeed abstract mathematical models that satisfy Definition 4.2.

When dealing with programs or Turing machines instead of dynamical systems, Definition 4.2 can assume the alternative and more simplistic form. We say an infinite collection $\mathbf{P}$ of programs displays *asymptotically observer-independent emergence* if, for every O, for every time instant t, and for every $c_{\mathbf{I}}, c_{\mathcal{O}}, c_e > 0$, there are $p \in \mathbf{P}$ and $t_e \in T$ such that, for every $t' \geq t_e$, one has that

$$\mathbf{I_{ac}}\left(\mathbf{U}\left(\langle t', p\rangle\right) \mid \langle w, t+m, O\rangle\right) > c_{\mathbf{I}} + c_{\mathcal{O}} + c_e \ , \quad (8)$$

where $t' \geq t + m$, $m \geq 1$, and w is the bit string in the first tape of O that satisfies the Observation Principle 1 at time instant t.

4.2.1. *A model and example in evolutionary systems.* In [20, 21, 28], a theoretical model for the open-ended evolution of programs (or TMs) is presented with the purpose of obtaining a mathematical proof of darwinian evolution under the framework of algorithmic information theory (AIT). Inspired by (but not limited to) evolutionary biology, this field is called *metabiology* and in a general sense it constitutes a pursuit of mathematical proofs of meta-level fundamental properties and quintessential "laws" in evolutionary systems [22].

The cumulative evolution model in [20] is defined as a sequence of sole TMs that evolve over time due to the transformations performed by randomly generated algorithmic mutations: one starts with an arbitrary TM and then subsequentially applies randomly generated algorithmic mutations so that, if a mutation leads to a fitter TM than the previous TM, the new mutated TM replaces the old one, and so on. This way, Chaitin [20] demonstrates that n bits of algorithmic complexity is expected to be reached after $\mathbf{O}(n^2(\log(n))^2)$ successive randomly generated algorithmic mutations. This result is achieved by employing a theoretical analysis of the resulting algorithmic complexity from certain algorithmic mutations that are expected to occur over time. Abrahão [1, 2] demonstrates that the open-endedness obtained in the former cumulative evolution model trickles down to the more realistic resource-bounded case: n bits of *time-bounded* algorithmic complexity is expected to be reached after $\mathbf{O}(n^2(\log(n))^2)$ successive randomly generated *time-bounded* algorithmic mutations in the cumulative evolution of *time-bounded* TMs.

These abstract evolutionary models were then corroborated by empirical results in [29] not only showing that randomly generated algorithmic mutations produce a speed up in adaptation in comparison to the uniformly random point mutations[1], but also that it may be related to explanations of the occurrence of modularity, diversity explosion, and massive extinctions.

We remark that algorithmic mutations as in [20, 21] are exactly the APs we defined in Section 3.1, except that, in such particular evolutionary models, these APs are randomly generated following the usual i.i.d. probability distribution of prefix-free binary sequences.

We can now apply the result from [20] in order to demonstrate the existence of a system that displays *expected* AOIE. The main idea of Theorem 4.1 is that the cumulative evolution of sole TMs under successive perturbations performed by the randomly generated algorithmic mutations is able to guarantee (with probability as high as one wants) that the emergence of algorithmic information is larger than any FOS can keep up in the long run.

Theorem 4.1. *Let* $\mathbf{P}$ *be a sequence of TMs that results from the cumulative evolution model in [20] after* t *successive randomly generated algorithmic mutations. Then,* $\mathbf{P}$ *satisfies Definition 4.2 with probability arbitrarily close to* 1 *as* $t \to \infty$*.*

Proof. From [20], we know that, for any initial TM P_0 and sufficiently large t, the algorithmic complexity of the output of the t-th TM P_t in the sequence $\mathbf{P}$ is expected to be at least as larger as $\sqrt[3]{t}$, where t is the number of successive randomly generated algorithmic mutations. That is, $\mathbf{K}\left(\mathbf{U}\left(\langle t, P_t\rangle\right)\right) \geq \mathbf{\Omega}\left(\sqrt[3]{t}\right)$ holds with probability arbitrarily close to 1 as $t \to \infty$, where $\mathbf{\Omega}\left(\cdot\right)$ is the usual Big-**Omega** notation for asymptotic lower bounds. Suppose that there is at least one OTM O and there are $c_{\mathbf{I}}, c_{\mathcal{O}}$, and c_e such that, for every $P_{t'} \in \mathbf{P}$ and $t_e \in T$ with $t' \geq t_e$, one has that $\mathbf{I_{ac}}\left(\mathbf{U}\left(\langle t', P_{t'}\rangle\right) \mid \langle w, t_o + m, O\rangle\right) \leq c_{\mathbf{I}} + c_{\mathcal{O}} + c_e$, where $t' \geq t_o + m$ and w is the bit string in the first tape of O that satisfies the Observation Principle 1 at an arbitrarily fixed time instant t_o. Then, from basic inequalities in AIT and from our definition of $\mathbf{I_{ac}}$ in Section 2.1, we would have that

$$\begin{aligned}\mathbf{K}\left(\mathbf{U}\left(\langle t', P_{t'}\rangle\right)\right) &\leq \mathbf{K}\left(m\right) + \mathbf{O}(1) + 2\,c_{\mathbf{I}} + c_{\mathcal{O}} + c_e \\ &\leq \mathbf{O}\left(\log\left(t'\right)\right) + \mathbf{O}(1) + 2\,c_{\mathbf{I}} + c_{\mathcal{O}} + c_e\ .\end{aligned}$$

Therefore, for infinitely many t_e, we would have that $\mathbf{K}\left(\mathbf{U}\left(\langle t_e, P_{t_e}\rangle\right)\right) \leq \mathbf{O}\left(\log\left(t_e\right)\right)$ and $\mathbf{K}\left(\mathbf{U}\left(\langle t_e, P_{t_e}\rangle\right)\right) \geq \mathbf{\Omega}\left(\sqrt[3]{t_e}\right)$, which leads to a contradiction. □

The result in Theorem 4.1 is stated using the alternative form of Definition 4.2 for programs (or TMs) so it can be easily brought to the cumulative evolution model in [20]. However, it is easy see how Theorem 4.1 can be translated to a system $\mathbf{S}$ instead of a sequence $\mathbf{P}$. To this end, replace each program/organism P_t in [20] and in the proof of Theorem 4.1 with the Turing equivalent FDDDS $\mathcal{S}$, as done in Lemma 2.1 in Section 2.2. Then, to achieve the contradiction obtained in the proof of Theorem 4.1, we use the fact that Lemma 2.2 is eventually satisfied by those FDDDSs in the sequence $\mathbf{S}$ composed of sole FDDDSs that evolve over time due to the transformations performed by randomly generated APs.

[1] Which are the usual random mutations under consideration in mainstream models based on evolutionary modern synthesis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

Emergent phenomena usually can be divided into two kinds [12, 23]: a temporal version in which emergence occurs over time, as the system interacts with the environment (and, for this reason, it is called *diachronic* emergence [23, 37]); and a holistic (or synergistic) version in which emergence occurs as a distinctive feature of the "whole" in comparison to the parts [23, 37]. The kind of emergence shown in Theorem 4.1 falls under the diachronic type. In particular, the *open-endedness* proved in this theorem strictly refers to the unbounded increase of complexity over time, as the evolution goes by. For this reason, it is called *evolutionary open-endedness* [6, 7]. In this sense, we can adopt the convention of classifying the phenomenon in Theorem 4.1 as *asymptotically observer-independent diachronic open-endedness.*

4.2.2. *A model and example in networked systems.* The pervasiveness of non-homogeneous network topological properties has fostered the recent field of network science and showed its important role in complex systems science [11]. In this regard, motivated by the pursuit of a unified theory for complexity in network science and complex systems science [10, 33], the theory of *algorithmic networks* [3, 6, 7] allows the investigation of how network topological properties can trigger emergent behavior that is capable of irreducibly increasing the computational power of the whole network. An algorithmic (complex) network $\mathfrak{N}$ is a population of computable systems (or TMs) whose members can share information with each other according to a complex network topology. Each node of the network is a computable system (or a TM) and each (multidimensional) edge of the network is a communication channel.

In [7], it is shown that there are network topological properties, such as the small diameter, associated with a computationally cheap and simple communication protocol of plain diffusion that asymptotically trigger an unlimited increase of expected emergent algorithmic complexity of a networked node's final output as the number of nodes increases indefinitely. This unlimited increase of expected emergent algorithmic complexity is called *expected emergent open-endedness* (EEOE) [6, 7] and—by simplifying the notation from [6, 7] to meet our present purposes—it is mathematically defined by

$$\lim_{N \to \infty} \mathbf{E}_{\mathfrak{N}} \left(\overset{net}{\underset{iso}{\Delta}} \mathbf{K}(o_i, c) \right) = \infty \ , \tag{9}$$

where: $\mathbf{E}_{\mathfrak{N}}(\cdot)$ gives the average value over all possible randomly generated nodes in the algorithmic network $\mathfrak{N}$; $\overset{net}{\underset{iso}{\Delta}} \mathbf{K}(o_i, c)$ is the *emergent algorithmic complexity (EAC)* of a node o_i in c communication rounds; and N is the total number of nodes in the algorithmic network $\mathfrak{N}$.

$$\overset{net}{\underset{iso}{\Delta}} \mathbf{K}(o_i, c) = \mathbf{K}\left(P_{net}\left(o_i, c\right)\right) - \mathbf{K}\left(P_{iso}\left(o_i, c\right)\right)$$

denotes the difference between the algorithmic complexity of the node o_i in c communication rounds when running *networked* and the algorithmic complexity of the node o_i in c communication rounds when running *isolated* from any other node, respectively. More formally: $P_{net}\left(o_i, c\right)$ denotes the program that computes the sequence of all the outputs that are sent to o_i's neighbors at all the communication rounds until c communication rounds has passed; and $P_{iso}\left(o_i, c\right)$ denotes the program that computes all the computation steps of the isolated program o_i during the c communication rounds. Note that the limit of the EEOE eventually neutralizes

 doi:10.20944/preprints202106.0019.v1

any pair of constants $c_{\mathbf{I}}$ that one may subtract or add in this difference. Thus, one can equivalently define the EEOE as

$$\lim_{N \to \infty} \mathbf{E}_{\mathfrak{N}} \left(\overset{net}{\underset{iso}{\Delta}} \mathbf{I_{ac}}(o_i, c) \right) = \infty \ .$$

The proof of the occurrence of EEOE in the models studied in [7] is achieved by employing a theoretical analysis of the trade-off between the number of communication rounds and the average density of networked nodes with the maximum algorithmic complexity. There is an optimum balance between these two quantities in which, if a large enough average density of these nodes is achieved in a sufficiently small number of communication rounds, then EEOE is triggered.

Instead of the communication protocol of plain diffusion, Abrahão et al [6] shows that a susceptible-infected-susceptible (SIS) contagion scheme [38, 39] in algorithmic networks with a power-law degree distribution is also sufficient for triggering EEOE. In [4], it is shown that a slight modification in the communication protocol of plain diffusion from [7] is sufficient for enabling the whole algorithmic network to synergistically solve problems at a higher computational class than the computational class of its individual nodes.

As we mentioned in Section 4.2.1, it was shown in [1, 2] that the evolutionary open-endedness from [20, 21] also applies to the resource-bounded case. Regarding the EEOE from [6, 7], future research is still needed for establishing how a resource-bounded version of the EEOE in algorithmic networks mathematically unfolds.

Unlike Theorem 4.1, since converting the result from collections of TMs into the dynamical system version is not as straightforward as in Section 4.2.1—while the opposite direction is easy—, we choose to demonstrate Theorem 4.2 already in the dynamical system variant of AOIE, satisfying Definition 4.2. To this end, we slightly extend our notation to encompass the case of *macro-level* dynamical systems that are composed of other *micro-level* dynamical systems. Let $\mathfrak{S}$ denote a FDDDS from which each (macro-level) state $\mathfrak{S} \restriction_x^x$ at time instant x is a fixed arrangement of all the (micro-level) states $\mathbf{S}_i \restriction_x^x$ with $1 \leq i \leq N$, where x is an arbitrary time instant and N is the total number of dynamical systems in the form $\mathbf{S}_i$ that composes $\mathfrak{S}$. In this way, we can further explore the necessity of extra algorithmic information at a certain stage of a system when the macro-level dynamical system has reached a sufficiently large size.

The underlying *main idea* of Theorem 4.2 is that, for a sufficiently large population/network size, the state space trajectory of $\mathfrak{S}'$ after time instant t (which corresponds to the simulation of the population of nodes running networked) demands more algorithmic information to be computed on the average than the FOS is able to process after the observation of the state space trajectory of $\mathfrak{S}$ (which corresponds to the simulation of the population of isolated nodes) until time instant t. The key steps of the following proof are to convert each node's computation into the respectively equivalent FDDDSs, as we saw in Section 3.2. First, we construct the dynamical system $\mathbf{S}_i$ until time instant t, corresponding to the node o_i running isolated from each other. Secondly, we construct the dynamical system $\mathbf{S}'_i$ from time instant $t+1$ until t', corresponding to the node o_i running networked. Thus, a node o_i receiving information from its neighbor nodes (according to the network topology) is equivalent to the FDDDS $\mathbf{S}'_i$ being algorithmically perturbed by its neighbor FDDDSs (also according to the same network topology). Then, we combine those FDDDSs in order to form the macro-level dynamical system $\mathfrak{S}$ and $\mathfrak{S}'$,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

which refer to the isolated and networked case, respectively. In other words, $\mathfrak{S}'$ is a population of randomly generated FDDDSs that can perturb each other according to the network topology. On the other hand, although $\mathfrak{S}$ is composed of the same population, no FDDDS in $\mathfrak{S}$ can perturb other FDDDSs in $\mathfrak{S}$.

Theorem 4.2. *Let $\mathfrak{S}$ be the macro-level FDDDS whose decision problem is Turing equivalent to calculating the final output of every* isolated *node o_i, which belongs to a population of N isolated nodes as in [7], where $1 \leq i \leq N$. Let $\mathfrak{S}'$ be the macro-level FDDDS whose decision problem is Turing equivalent to calculating the final output of every* networked *node o_i, which belongs to an algorithmic network $\mathfrak{N}$ that displays EEOE as in [7] and the set of nodes of $\mathfrak{N}$ is the same population in the former isolated case. Then, for every $\mathcal{O}$, for every time instant t, and for every $c_{\mathbf{I}}, c_{\mathcal{O}}, c_e > 0$, there is $t_e \in T$ such that, for every $t' \geq t_e$, one has that*

$$\mathbf{E}_{i \leq N}\left(\mathbf{I_{ac}}\left(\left\langle \mathbf{S}_i \restriction_0^t, \mathbf{S}'_i \restriction_{t+1}^{t'} \right\rangle \middle| \left\langle w, \left(\mathcal{O} \restriction_0^t, \mathcal{O}' \restriction_{t+1}^{t+m}\right)\right\rangle\right)\right) > c_{\mathbf{I}} + c_{\mathcal{O}} + c_e$$

holds with probability arbitrarily close to 1 as $N \to \infty$, where: $t' \geq t+m$, $m \geq 1$, w is the bit string in the first tape of O that satisfies the Observation Principle 1 at time instant t; $\mathbf{S}_i$ and $\mathbf{S}'_i$ are the micro-level FDDDSs that simulate the node o_i in the isolated and the networked case, respectively; and the collection of all $\mathbf{S}_i$ and $\mathbf{S}'_i$, for every $i > 0$, forms the macro-level FDDDSs $\mathfrak{S}$ and $\mathfrak{S}'$, respectively.

Proof. Let $\mathfrak{N}$ be an algorithmic network studied in [7] that displays EEOE. Since every node o_i is a randomly generate program (or TM), we need to first construct its equivalent dynamical system that simulates the *networked* behavior of the node o_i, when the dynamical system is interacting with (i.e., perturbing and being perturbed by) its neighbors, and simulates the isolated behavior of the node o_i, when the dynamical system is *not* interacting with any of its neighbors. The isolated case is easily obtained by just replacing each node o_i with a FDDDS $\mathbf{S}_i$ that has the same e.r. of A such that the initial state $\left(\mathbf{S}_i \restriction_0^0\right)$ corresponds to the initial configuration of A simulating $M\left(M'\left(P_{iso}\left(o_i, c\right), \emptyset\right)\right)$, where the universal FDDDS A and TMs M and M' are exactly those defined in the proof of Lemma 3.1. Therefore, there is $h \geq 0$ such that, for every isolated o_i, we have that

$$\mathbf{K}\left(\mathbf{S}_i \restriction_0^t\right) = \mathbf{K}\left(\mathbf{U}\left(P_{iso}\left(o_i, c\right)\right)\right) \pm \mathbf{O}(1) \ , \tag{10}$$

where $t = c + h - 1$. In order to obtain the networked case, we then need to demonstrate that there is an equivalent AP $\mathcal{P}_{(o_i, c')}$ that transforms the FDDDS $\mathcal{S}_i'^{(c')}$ (which computes the output that is sent to o_i's neighbors at communication round c') into the FDDDS $\mathcal{S}_i'^{(c'+1)}$ (which computes the output that is sent to o_i's neighbors at communication round $c'+1$). To this end, by employing the universal FDDDS A and TMs M and M' from the proof of Lemma 3.1, one constructs a sequence $\mathbf{S}'_i = \left(\mathcal{S}_i'^{(1)}, \dots, \mathcal{S}_i'^{(c')}, \dots, \mathcal{S}_i'^{(c)}\right)$ of FDDDSs and the program $\mathcal{P}_{(o_i, c')}$ that returns the state $\left(\mathcal{S}_i'^{(c'+1)} \restriction_{t_{(c'+1)}}^{t_{(c'+1)}}\right)$ given the state $\left(\mathcal{S}_i'^{(c')} \restriction_{t_{(c'+1)}-1}^{t_{(c'+1)}-1}\right)$ as input, where: the state space trajectory $\left(\mathcal{S}'^{(1)} \restriction_{t+1}^{t_{(2)}-1}\right)$ computes $\mathbf{U}\left(P_{iso}\left(o_i, 1\right)\right)$; the state space trajectory $\left(\mathcal{S}_i'^{(c')} \restriction_{t_{(c')}}^{t_{(c'+1)}-1}\right)$ computes the output that is sent to o_i's neighbors at communication round c'; the state space trajectory $\left(\mathcal{S}_i'^{(c'+1)} \restriction_{t_{(c'+1)}}^{t_{(c'+2)}-1}\right)$ computes the output that is sent to o_i's neighbors at communication round $c'+1$; and so on.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

Note that c' is an arbitrary communication round with $1 \leq c' \leq c$. Therefore,

(11)

$$\mathbf{K}\left(\left\langle {\mathcal{S}'}_i^{(1)} \restriction_{t+1}^{t_{(2)}-1}, \dots, {\mathcal{S}'}_i^{(c')} \restriction_{t_{(c')}}^{t_{(c'+1)}-1}, \dots, {\mathcal{S}'}_i^{(c)} \restriction_{t_{(c)}}^{t'} \right\rangle\right) = \mathbf{K}\left(\mathbf{U}\left(P_{net}\left(o_i, c\right)\right)\right) \pm \mathbf{O}(1) \ .$$

Let $\mathfrak{S}$ and $\mathfrak{S}'$ be the FDDDSs composed of a fixed arrangement of all the (micro-level) states of $\mathbf{S}_i$ and $\mathbf{S}'_i$, respectively, where $1 \leq i \leq N$ and N is the total number of nodes. Let $\mathbf{E}_{i \leq N}\left(\cdot\right)$ denote the average over all the constituent systems $\mathbf{S}_i$ or $\mathbf{S}'_i$ that composes $\mathfrak{S}$ or $\mathfrak{S}'$, respectively. We know from [7] that Equation 9 holds with probability arbitrarily close to 1 as $N \to \infty$. Therefore, from basic inequalities in AIT and from the Observation Principle 1, one finally achieves the proof of Theorem 4.2 by combing Equation 9 with Equations 10 and 11. □

Extending the result of Theorem 4.2 to the algorithmic networks in [6] is straightforward from the method employed in the proof of Theorem 4.2. Thus, we leave up to the reader.

While the temporal (or diachronic) variant of AOIE presented in the previous Section 4.2.1 occurs over time due to the successive perturbations from the environment into the system $\mathbf{P}$ (or $\mathbf{S}$), the variant of AOIE presented in Theorem 4.2 occurs due to the interaction (in the form of perturbations) between the micro-level systems $\mathbf{S}_i$ as the number of these micro-level systems contained in the macro-level system $\mathfrak{S}'$ increases. Although there might be a FOS that can compute the expected behavior of an isolated micro-level system, the AOIE guarantees that this will eventually cease to happen as the size of the macro-level system becomes sufficiently large. The existence of the time instant t_e in Theorem 4.2 assures that, even if the FOS can computably predict the expected behavior of an isolated micro-level system that belongs to $\mathfrak{S}$, there is a phase transition for which, if the expected behavior of a micro-level system of $\mathfrak{S}'$ is not emergent to a certain FOS, then it will start to be emergent once the number of micro-level systems in $\mathfrak{S}'$ is sufficiently large. Thus, the sort of process that gives rise to the AOIE in Theorem 4.2 differs from the one in Theorem 4.1 in the same manner as the holistic variant of emergence differs from the temporal (or diachronic) one. For this reason, AOIE in Theorem 4.2 falls under the *holistic* variant of emergence. Hence, one can adopt the convention of calling the *emergent open-endedness* proved in Theorem 4.2 as *asymptotically observer-independent holistic open-endedness*.

Downward (or top-down) causation is usually described in the literature as a type of process in which the global (or macro-level) dynamics of the system as a "whole" gains causal efficacy over the micro-level systems (or parts) [23, 25, 37]. The holistic variant of AOIE in Theorem 4.2 demonstrates that, for sufficiently large $\mathfrak{S}'$, the expected behavior of a networked micro-level system $\mathbf{S}'_i$ is overruled by the algorithmic-informational dynamics of the algorithmic perturbations produced according to the network topology. This occurs because the algorithmic information of the dynamics of an isolated $\mathbf{S}_i$ is eventually not sufficient for computing the networked behavior of $\mathbf{S}'_i$, while the total algorithmic information shared through the network is. In this regard, Theorem 4.2 gives a proof of the existence of expected downward causation in FDDDSs (or in networked computable systems). It also offers the advantage of this expected downward causation being independent of the observer's formal knowledge in the asymptotic limit.

 doi:10.20944/preprints202106.0019.v1

4.2.3. *Weak, intermediate, or strong emergence.* In a broad sense, if weak emergent is characterized by phenomena that are *in principle* deducible or derivable from the simple initial or micro-level conditions, but that appear as unexpected at a higher coarse-grained level due to the lack of information, resources, or knowledge, one can classify the ODE in Section 4.1 as a weak emergence. This agrees with the approach to weak emergence as the unexpectedly complex behavior in [23], as explanatory incompressibility in [15], and as the type 0 and 1 weak emergence in [12]. Indeed, since there is always the possibility of another existing observer to which the phenomenon ceases to appear emergent, then the emergence in Definition 4.1 is, "*in principle*", deducible or derivable at the same time that there are observers for which the emergent behavior is "truly" incompressible and relatively uncomputable to these observers. The term "truly" is employed here in the precise sense that such an incompressibility or relative uncomputability do not depend on the chosen method to measure the information content, on the errors or distortions in the act of observing itself, nor on the algorithmic-informational cost to process the observed system in accordance with the observer's formal knowledge.

On the other hand, classifying the AOIE studied in Section 4.2 is not so easy. The crux of the matter is not quite in the notion of reducibility, derivability, or predictability (as in our case they have a formal unambiguous translation into sufficient algorithmic information), but in the term "in principle". If "in principle" means that the phenomenon should remain emergent for every formal observer system that belongs to the same computational class of the observed systems, then AOIE could be interpreted as a type of strong emergence. This is because, for every formal observer system at the same computational class (i.e., in the same Turing degree or in the same running time complexity class) of the observed systems, the behavior of an observed system that satisfies Definition 4.2 will eventually cease to be computable or predictable in the long run. In this sense, since AOIE implies ODE for infinitely many time steps in the future, the Church-Turing hypothesis entails that a system displaying AOIE (and, in this case, strong emergence) would always be understood as displaying ODE (and, therefore, the above weak emergence), while in fact never ceasing to display ODE (or weak emergence) for any possible observer. In other words, under the Church-Turing hypothesis, if AOIE is considered strong emergence, then this type of strong emergent phenomenon is a pseudoparadoxical type of emergent phenomenon that is always mathematically understood as weak emergence, while in fact being strong emergent if the observer could know the point of view of every observer. Another form of strong emergence has been described as the ultimate necessity of novel fundamental powers or laws to scientifically explain the macro-level behavior of a system [37]. In the context of FDDDSs or computable systems, AOIE offers a proof of this necessity, but now formally expressed as the never-ending necessity of new axioms (or new algorithmic information). Due to the presence of expected downward causation in the AOIE of Theorem 4.2, one can also successfully argue that the systems $\mathfrak{S}'$ satisfy the type 2 strong emergence in [12].

However, if "in principle" does not constrain the computational class of the observer, then AOIE can be brought back to the weak case. This is because, although no (finite) formal axiomatic theory held by the observer can compute the observed system in the long run, there might be *oracle* observers that can, if the e.r. of the observer itself belongs to higher Turing degrees. For example,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2021 doi:10.20944/preprints202106.0019.v1

it is true that both the sequence $\mathbf{P}$ of TMs in Theorem 4.1 and the macro-level FDDDS $\mathfrak{S}'$ in Theorem 4.2 cannot be computed by formal observer systems in the asymptotic limit, but both can be computed by an oracle machine of Turing degree $\mathbf{0}'$. In other words, if one allows observers to have access to an infinite source of algorithmic information, e.g., by filling out the *infinite* second tape of $\mathcal{O}$ with a halting probability (or Chaitin's Omega number) [17, 18], there are systems that satisfy Definition 4.2 at the same time that they are relatively computable by a *special* observer. Thus, in the case one believes in the existence of strong emergence that resists to ontological characterizations in terms of physical or informational causal efficacies, such as the emergence of qualia in the conscious mind [23, 37], it becomes consistent to classify AOIE as a type of weak emergence.

Although not constraining the computational class of the observer may seen reasonable, one is inherently assuming that there are "special" observers that belong to a higher computational class than that of all the other systems that can be observed by them, assumption which *per se* is just another type of constrain to be applied to Definition 4.2. One way to avoid this assumption, while still consistent with the fact that AOIE from Definition 4.2 is a stronger form of emergence than what is usually considered to be weak emergence (which generally falls under the ODE from Definition 4.1) is to classify AOIE as a type of *intermediate emergence* [23]. This kind of terminology has been proposed by Chalmers [23] to deal with a type of emergence that arises from a fundamental epistemological limitation given the known physical laws at the time of observation. In this sense, intermediate emergence is defined upon the existence of unbridgeable incompleteness of the observer's knowledge, so that even "in principle" one would not be able to deduce the macro-level complex behavior, which is still "in principle" determined by irreducible new laws that one always needs to devise or discover in the future. Regarding incompleteness, emergence as a consequence of uncomputability was also proposed by Cooper [24].

Once formal theories might be the best our scientific knowledge can grasp, classifying something as intermediate emergence entails that one is assuming the existence (or scientific pertinence) of a stronger form of emergence, which might be not the case. Nevertheless, we consider both hypotheses (i.e., with or without "special" observers) of mathematical and scientific relevance of pursuing. For the present purposes, since we have only dealt with formal axiomatic theories and not with physical, chemical, or biological theories in general, we hold on to what our theoretical results imply and, therefore, we avoid the claim of classifying AOIE as either weak, strong, or intermediate emergence. What we have shown is that, restricted to the context of algorithmic information dynamics, FDDDSs, computable systems, and FOSs, the AOIE is the *strongest* form of emergence that formal axiomatic theories can attain. Algorithmic information and algorithmic randomness have demonstrated and captured fundamental properties that underlie incompleteness of formal theories and the limits of mathematics [17, 19, 26]. Thus, within the scope of this article, it may be not a surprise that algorithmic information theory was the key to formalize emergence up to the limits that our formal mathematical knowledge can grasp.

5. Conclusion

Within the scope of algorithmic information dynamics, this article studies the fundamental role that algorithmic information plays in the act of observing and in the occurrence of emergent phenomena in discrete deterministic dynamical systems and computable systems.

We have formalized the act of observing a system as mutual perturbations occurring between the observer (which is a system) and the observed system. Formal observer systems are systems that previously know a formal axiomatic theory, which they can apply in order to compute the future behavior of the observed system. As a consequence, we demonstrate that a (finite discrete deterministic dynamical) system displaying emergent behavior with respect to an observer constitutes a type of emergence of algorithmic information that is invariant and minimal. Although it depends on the observer's formal knowledge, this emergence is robust to variations of the arbitrarily chosen method of measuring irreducible information content, to error (or distortions) in the very act of observing, and to variations of the algorithmic-informational cost to process the observed system in accordance with the observer's formal knowledge. Thus, this type of emergence is called *observer-dependent emergence* (ODE).

Then, we investigated the unbounded increase of emergent algorithmic information, which defines a type of emergence that we call *asymptotically observer-independent emergence* (AOIE). In addition to the above invariance, minimality, and robustness, any formal axiomatic theory that any formal observer system might devise will eventually fail to compute or predict the behavior of a system that displays AOIE. We demonstrated that there is an abstract evolutionary model that displays the temporal (or diachronic) variant of AOIE, which guarantees that no formal observer system is able to always compute the behavior of evolutionary software in the long run. We also demonstrated that there is an abstract model for networked systems that displays the holistic variant of the AOIE, which guarantees that no formal observer system is able to always compute the expected behavior of a micro-level subsystem as the size of the macro-level system becomes sufficiently large.

We also compared the ODE and AOIE studied in this article with other related work on weak and strong emergence in the literature. Depending on the interpretation of the term "in principle" in the usual definitions of weak and strong emergence, AOIE can be classified as weak, intermediate, or strong emergence. In any event, the results of the present article show that, within the context of finite discrete deterministic dynamical systems, or computable systems, AOIE is the strongest version of emergence that formal axiomatic theories can grasp. Whether this claim can be extended to other physical systems and physical theories is a problem that needs further discussion and future research. Nevertheless, given the relevance of formal axiomatic theories in mathematics and science in general, we consider the strength of AOIE demonstrated in this article to be remarkable.

References

[1] Abrahão FS. 2015 *Metabiologia, Subcomputação e Hipercomputação: em direção a uma teoria geral de evolução de sistemas*. Ph.D. thesis, Federal University of Rio de Janeiro (UFRJ), Brazil, Rio de Janeiro. Language: Portuguese. Available at http://objdig.ufrj.br/10/teses/832593.pdf

[2] Abrahão FS. 2016 The 'paradox' of computability and a recursive relative version of the Busy Beaver function. In *Information and Complexity* (eds C Calude, M Burgin), pp. 3–15. Singapure: World Scientific Publishing. (doi:10.1142/9789813109032_0001)

[3] Abrahão FS. 2016 Emergent algorithmic creativity on networked Turing machines. In *The 8th International Workshop on Guided Self-Organization at the Fifteenth International Conference on the Synthesis and Simulation of Living Systems* (ALIFE), Cancún.

[4] Abrahão FS, D'Ottaviano ÍML, Wehmuth K, Dória FA, Ziviani A. 2020 Learning the undecidable from networked systems. In *Unravelling Complexity* (eds S Wuppuluri, FA Doria), World Scientific Publishing. (doi:10.1142/9789811200076_0009)

[5] Abrahão FS, Wehmuth K, Zenil H, Ziviani A. 2021 An Algorithmic Information Distortion in Multidimensional Networks. In *Complex Networks and Their Applications IX* (eds RM Benito, C Cherifi, H Cherifi, E Moro, LM Rocha, M Sales-Pardo), pp. 520–531. Cham: Springer International Publishing. (doi:10.1007/978-3-030-65351-4_42)

[6] Abrahão FS, Wehmuth K, Ziviani A. 2018 Emergent Open-Endedness from Contagion of the Fittest. *Complex Systems* **27**. (doi:10.25088/ComplexSystems.27.4.369)

[7] Abrahão FS, Wehmuth K, Ziviani A. 2019 Algorithmic networks: Central time to trigger expected emergent open-endedness. *Theoretical Computer Science* **785**, 83–116. (doi:10.1016/j.tcs.2019.03.008)

[8] Adams AM, Berner A, Davies PCW, Walker SI. 2017 Physical Universality, State-Dependent Dynamical Laws and Open-Ended Novelty. *Entropy* **19**, 461. (doi:10.3390/e19090461)

[9] Adams A, Zenil H, Davies PCW, Walker SI. 2017 Formal Definitions of Unbounded Evolution and Innovation Reveal Universal Mechanisms for Open-Ended Evolution in Dynamical Systems. *Scientific Reports* **7**, 997. (doi:10.1038/s41598-017-00810-8)

[10] Barabási A-L. 2009 Scale-Free Networks: A Decade and Beyond. *Science* **325**, 412–413. (doi:10.1126/science.1173299)

[11] Barabási A-L. 2016 *Network Science*. 1st edn. USA: Cambridge University Press.

[12] Bar-Yam Y. 2004 A mathematical theory of strong emergence using multiscale variety. *Complexity* **9**, 15–24. (doi:10.1002/cplx.20029)

[13] Becher V, Figueira S. 2002 An example of a computable absolutely normal number. *Theoretical Computer Science* **270**, 947–958. (doi:10.1016/S0304-3975(01)00170-0)

[14] Bedau MA. 1997 Weak Emergence. *Philosophical Perspectives* **11**, 375–399. (doi:10.1111/0029-4624.31.s11.17)

[15] Bedau MA. 2010 Weak Emergence and Context-Sensitive Reduction. In *Emergence in Science and Philosophy* (eds A Corradini, T O'Connor), Routledge. (doi:10.4324/9780203849408)

[16] Burgin M. 2009 Theory of Information: Fundamentality, Diversity and Unification. World Scientific. (doi:10.1142/7048)

[17] Calude CS. 2002 *Information and Randomness: An algorithmmic perspective*. 2nd edn. Springer-Verlag. Available at

https://www.springer.com/br/book/9783540434665

[18] Chaitin GJ. 1975 A Theory of Program Size Formally Identical to Information Theory. *Journal of the ACM* **22**, 329–340. (doi:10.1145/321892.321894)

[19] Chaitin G. 2004 *Algorithmic Information Theory*. 3rd edn. Cambridge University Press. (doi:10.1017/CBO9780511608858)

[20] Chaitin G. 2012 Life as evolving software. In *A Computable Universe: Understanding and Exploring Nature as Computation* (ed H Zenil), pp. 1–23. Singapure: World Scientific Publishing. (doi:10.1142/9789814374309_0015)

[21] Chaitin GJ. 2012 *Proving Darwin: making biology mathematical*. New York: Pantheon Books.

[22] Chaitin VMFG, Chaitin GJ. 2018 A Philosophical Perspective on a Metatheory of Biological Evolution. In *The Map and the Territory: Exploring the Foundations of Science, Thought and Reality* (eds S Wuppuluri, FA Doria), pp. 513–532. Cham: Springer International Publishing. (doi:10.1007/978-3-319-72478-2_29)

[23] Chalmers DJ. 2008 Strong and Weak Emergence. In *The Re-Emergence of Emergence* (eds P Clayton, PCW Davies), Oxford: Oxford University Press. (doi:10.1093/acprof:oso/9780199544318.003.0011)

[24] Cooper SB. 2009 Emergence as a computability-theoretic phenomenon. *Applied Mathematics and Computation* **215**, 1351–1360. (doi:10.1016/j.amc.2009.04.050)

[25] Davies PCW. 2008 The Physics of Downward Causation. In The *Re-Emergence of Emergence* (eds P Clayton, PCW Davies), Oxford: Oxford University Press. (doi:10.1093/acprof:oso/9780199544318.003.0002)

[26] Downey RG, Hirschfeldt DR. 2010 Algorithmic Randomness and Complexity. New York, NY: Springer New York. (doi:10.1007/978-0-387-68441-3)

[27] Fernández N, Maldonado C, Gershenson C. 2014 Information Measures of Complexity, Emergence, Self-organization, Homeostasis, and Autopoiesis. In *Guided Self-Organization: Inception* (ed M Prokopenko), pp. 19–51. Springer, Berlin, Heidelberg. (doi:10.1007/978-3-642-53734-9_2)

[28] Hernández-Orozco S, Hernández-Quiroz F, Zenil H. 2018 Undecidability and Irreducibility Conditions for Open-Ended Evolution and Emergence. *Artificial Life* **24**, 56–70. (doi:10.1162/ARTL_a_00254)

[29] Hernández-Orozco S, Kiani NA, Zenil H. 2018 Algorithmically probable mutations reproduce aspects of evolution, such as convergence rate, genetic memory and modularity. *Royal Society Open Science* **5**, 180399. (doi:10.1098/rsos.180399)

[30] Lewis H, Papadimitriou CH. 1997 *Elements of the Theory of Computation*. 2nd edn. Prentice-Hall.

[31] Li M, Vitányi P. 2019 *An Introduction to Kolmogorov Complexity and Its Applications*. 4th edn. Cham: Springer, Cham. (doi:10.1007/978-3-030-11298-1)

[32] Lizier J, Bertschinger N, Jost J, Wibral M. 2018 Information Decomposition of Target Effects from Multi-Source Interactions: Perspectives on Previous, Current and Future Work. *Entropy* **20**, 307. (doi:10.3390/e20040307)

[33] Michail O, Spirakis PG. 2018 Elements of the theory of dynamic networks. *Communications of the ACM* **61**, 72–72. (doi:10.1145/3156693)

[34] Mitchell M. 2009 *Complexity: A Guided Tour*. Oxford University Press. Available at ISBN 9780195124415.

[35] Neary T, Woods D. 2006 P-completeness of Cellular Automaton Rule 110. In *Automata, Languages and Programming* (eds M Bugliesi, B Preneel, V Sassone, I Wegener), pp. 132–143. Berlin, Heidelberg: Springer Berlin Heidelberg. (doi:10.1007/11786986_13)
[36] Nicolis G, Nicolis C. 2009 *Foundations of complex systems*. World Scientific Publishing Co. Pte. Ltd. (doi:10.1017/S1062798709000738)
[37] O'Connor T. 2020 Emergent Properties. In *The Stanford Encyclopedia of Philosophy* (ed EN Zalta), Metaphysics Research Lab, Stanford University. Available at https://plato.stanford.edu/archives/fall2020/entries/properties-emergent/
[38] Pastor-Satorras R, Vespignani A. 2001 Epidemic spreading in scale-free networks. *Physical Review Letters* **86**, 3200–3203. (doi:10.1103/PhysRevLett.86.3200)
[39] Pastor-Satorras R, Vespignani A. 2002 Immunization of complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **65**, 036104. (doi:10.1103/PhysRevE.65.036104)
[40] Prokopenko M, Boschetti F, Ryan AJ. 2009 An information-theoretic primer on complexity, self-organization, and emergence. *Complexity* **15**, 11–28. (doi:10.1002/cplx.20249)
[41] Rich EA. 2007 *Automata Theory and Applications*. USA: Prentice-Hall, Inc.
[42] Walker SI, Davies PCW. 2012 The Algorithmic Origins of Life. *Journal of The Royal Society Interface* **10**, 20120869–20120869. (doi:10.1098/rsif.2012.0869)
[43] Wolfram S. 2002 *A new kind of science*. Champaign, IL: Wolfram Media.
[44] Zenil H. 2020 A Review of Methods for Estimating Algorithmic Complexity: Options, Challenges, and New Directions. *Entropy* **22**, 612. (doi:10.3390/e22060612)
[45] Zenil H, Kiani NA, Abrahão FS, Rueda-Toicen A, Zea AA, Tegnér J. 2020 Minimal Algorithmic Information Loss Methods for Dimension Reduction, Feature Selection and Network Sparsification. arXiv Preprints. Available at https://arxiv.org/abs/1802.05843
[46] Zenil H, Kiani N, Abrahão F, Tegnér J. 2020 Algorithmic Information Dynamics. *Scholarpedia Journal* **15**, 53143. (doi:10.4249/scholarpedia.53143)
[47] Zenil H, Kiani NA, Marabita F, Deng Y, Elias S, Schmidt A, Ball G, Tegnér J. 2019 An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems. *iScience* **19**, 1160–1172. (doi:10.1016/j.isci.2019.07.043)
[48] Zenil H, Kiani NA, Tegnér J. 2017 Low-algorithmic-complexity entropy-deceiving graphs. *Physical Review E* **96**, 012308. (doi:10.1103/PhysRevE.96.012308)
[49] Zenil H, Kiani NA, Tegnér J. 2019 The Thermodynamics of Network Coding, and an Algorithmic Refinement of the Principle of Maximum Entropy. *Entropy* **21**, 560. (doi:10.3390/e21060560)
[50] Zenil H, Kiani NA, Zea AA, Tegnér J. 2019 Causal deconvolution by algorithmic generative models. *Nature Machine Intelligence* **1**, 58–66. (doi:10.1038/s42256-018-0005-0)