Assumptions of Physics blueprint: counting evolutions as a foundation for entropy and system/state definition

Gabriele Carcassi[*] and Christine Aidala

Physics Department, University of Michigan, Ann Arbor, MI 48109

(Dated: June 8, 2021)

This paper presents ideas and progress towards a more general reconceptualization of entropy that can be used to recover and extend thermodynamics and statistical mechanics. This would address foundational questions such as: why does entropy posses both objective qualities (i.e. it is a measurable property) and in (i.e. it is an ensemble property)? Is there a general notion of equilibrium that includes both dynamical and thermodynamical equilibria in a way that is relativistically invariant? What is entropy maximization, why is so successful in many fields and how is it linked to state definition? The working hypothesis is that entropy is the logarithm of the count of the possible evolutions of a system within a specific process. That is, given the description of system at the desired granularity level, we count all the way the system can evolve in time that are compatible with that description. This switch from counting states to counting evolutions makes the concept more relativistically sound and allows to properly keep track of how the information about the system is transported back and forward in time by the process, which is essential to apply the concept in a more general setting. Defining states requires, at least in some setting, the independence of the system, which indirectly leads to entropy maximization. The objective is to formulate a precise mathematical framework that makes this intuition precise and recovers the usual entropies in specific settings.

This work is part of Assumptions of Physics (https://assumptionsofphysics.org), a project that aims to identify a handful of physical principles from which the basic laws can be rigorously derived.

1. INTRODUCTION

This paper presents the goal and the current status of our research. It is intended to gather early feedback, including pointers to other relevant work. Feedback is welcome and encouraged.

The overall goal of Assumptions of Physics is to identify a handful of physical principles from which the basic laws can be rigorously derived. Part of this work is establishing a formal framework for states and processes, of which the definition of entropy is a crucial aspect. We will not assume specific knowledge of the rest of our framework[1, 2], though ultimately the goal is to integrate this work with the rest. As usual, a formal construction from scratch will force us to make explicit the physical requirements that are baked into the use of the common concepts and provide us with a framework that is, hopefully, still valid when those assumptions fail.

Previous work focused on showing how requiring experimental verifiability of a theory leads to topological and σ-algebraic structures within that theory. Another effort is exploring how the ability to compare the level of granularity within the theory is connected to geometric and measure theoretic structures, giving us a way to quantify the information contained in each statement of the theory. We now want to characterize the relationship between the granularity (i.e the information content) of the descriptions at different times. **The working hypothesis is that entropy captures how coarse or fine grained is the description of a particular system within a particular process. It does so by measuring the logarithm of the count of the possible evolutions that are compatible.** One way to look at it is that as the Boltzmann entropy fails to take into account interactions and correlations between molecules, the count of the microstates and the Shannon/Gibbs entropy fail to take into account system/environment correlations. Since these correlations may change in time, the same description of the same system will corresponds to different entropy in time, and this is what the process entropy is designed to solve. The central problem of this paper is therefore to understand what are the appropriate basic definitions, how do they map to the standard ones and how system independence, as characterized by entropy, is required for the very definitions of systems and states.

To give a sense of how this would work, we start with defining a **process domain** $\bar{\mathcal{P}}$, which represents all meaningful statements, all descriptions at all times and at all scales, that apply to a particular process. Mathematically, it is a σ-complete Boolean algebra, which means it is closed under negation, countable disjunction (i.e. logical OR) and countable conjunction (i.e. logical AND). It will also come equipped with a partial order $\preccurlyeq$ such that $s_1 \preccurlyeq s_2$ if $s_1 \vee s_2 \equiv s_2$ that indicates whether one statement gives a **narrower**, more specific, description than the other. For example:

---

[*] carcassi@umich.edu

- *"The position of the object is between 0 and 1 meters" $\preccurlyeq$ "The position of the object is between 0 and 1 kilometers"*
- *"The fair die landed on 1" $\preccurlyeq$ "The fair die landed on 1 or 2"*
- *"The first bit is 0 and the second bit is 1" $\preccurlyeq$ "The first bit is 0"*

Additionally, we will have a way to compare any two statements and decide whether one provides a description at a finer level of granularity. Mathematically, we have an additional preorder $\leqq: \bar{\mathcal{D}} \times \bar{\mathcal{D}} \to \mathbb{B}$. Saying $\mathsf{s}_1 \leqq \mathsf{s}_2$ means that the description provided by s_1 is **finer**, gives more information, is more precise, than the description provided by s_2. We say that two descriptions are **equigranular**, noted $\mathsf{s}_1 \doteq \mathsf{s}_2$, if they provide the same level of detail. For example:

- *"The position of the object is between 0 and 1 meters" $\leqq$ "The position of the object is between 2 and 3 kilometers"*
- *"The fair die landed on 1" $\leqq$ "The fair die landed on 3 or 4"*
- *"The first bit is 0 and the second bit is 1" $\leqq$ "The third bit is 0"*

In these cases, the first statement may not be contained or overlap with the second. We will assume fineness have all the properties it needs such that if a statement $\mathsf{u} \in \bar{\mathcal{P}}$ is chosen, we can construct a measure μ_{u} that quantifies the granularity of a statement in terms of the unit. For example, if $\mathsf{u} =$ *"the position of the ball is within 0 and 1 meters"* and $\mathsf{s} =$ *"the position of the ball is within 3.5 and 5 meters"*, then $\mu_{\mathsf{u}}(\mathsf{s}) = 1.5$.

Each process domain will include a set of possibilities E: these statements are the narrowest possible statement and provide the full description the whole process at all times and at all level of detail. We call **evolutions** these statements as they effectively describe the evolution of the all systems within the process. We will assume that all evolutions are equigranular (i.e. $e_1 \doteq e_2$ for all $e_1, e_2 \in E$). The measure is therefore always uniform over the possibilities and therefore characterizing the granularity level of a description is effectively equivalent to counting the evolutions that are compatible with such description. The process entropy $S(\mathsf{s}) = \ln \mu(\mathsf{s})$ is defined to bethe logarithm of that count, and therefore the logarithm of the measure.

With these basic definitions we can transport statements back forth in time and see how the process entropy changes. More specifically, for each instant of time $t \in T$ we can define a respective subdomain $\bar{\mathcal{D}}_t \subseteq \bar{\mathcal{P}}$ that contains all the descriptions at that particular time. Given a statement s_0 at time t_0, we can define the best prediction s_1 at time $t_1 > t_0$ and find that its process entropy must be greater (i.e. $S(\mathsf{s}_1) \geq S(\mathsf{s}_0)$). We can also define the best reconstruction s_{-1} at time $t_{-1} < t_0$ and find that its process entropy also must be greater (i.e. $S(\mathsf{s}_{-1}) \geq S(\mathsf{s}_0)$). Intuitively, information about a system at a particular time can only get worse (or stay the same) as we transport that description forward or backward in time along the process. This relationship is symmetric with time.

Similarly, we can define a trajectory $x(t)$ of a particular system, where x is the full description of that system. A process is deterministic if knowing $x(t)$ means we can predict $x(t + \Delta t)$, meaning that all evolutions that start in the same state must also end in the same state: evolutions can merge and cannot diverge. Therefore the evolution count, and therefore the entropy, can never decrease: $S(x(t)) \leq S(x(t + \Delta))$. A process is reversible if knowing $x(t + \Delta t)$ means we can reconstruct $x(t)$, meaning that all evolutions that end in the same state must also start in the same state: evolutions can diverge and cannot merge. If a process is deterministic end reversible, then, the evolution count, and therefore the entropy, must be conserved $S(x(t)) = S(x(t + \Delta))$.

The last insight is that states are description of the system and only of the system. This happens only when the system is independent: the evolution of the system tells us nothing about the evolution of other systems. Independence means both factorization of the measure (and linearity of entropy) and it's maximization: the lack of correlations means all evolutions are possible. States can then be characterized by entropy in a way that is process independent, and lead to the geometrical structures of thermodynamics, classical, quantum and statistical mechanics.

The strength of the approach is that these definitions will apply to processes independently of the discipline (i.e. physics, economics, ecology, ...) or the kind (i.e. deterministic, irreversible, stochastic, classical/quantum, ...). In particular, it will work in dissipative systems or dynamical systems with equilibria, where the standard notion of entropy decreases as it proceeds toward equilibrium. Additionally it reduces to the standard notions in the appropriate settings. For example, if we assume the system is characterized by a Hamiltonian system over time (i.e. the extended phase space charted by $\{q^i, p_i, t\}$ with Hamiltonian flow), the measure reduces to counting the states at each time. If the flow is not Hamiltonian, however, the phase space measure is not the correct measure at all times (i.e. the same precision over q and p at different times corresponds to a different amount of information). On the other hand, if we assume the system over a small Δt fluctuates randomly according to a distribution ρ, then the counting the possible evolutions means counting permutations. The process entropy, then, can be expressed as $\int \rho \ln \rho$, and the Shannon entropy is recovered.

2. OVERVIEW

We now give a more detailed overview of the goal and status. This will also cover the conceptual understanding and physical motivation, not just the mathematical aspects. Later sections will present both what has been established to

work and what not to work in full mathematical detail.

The overall goal is to describe a generic system as it evolves in time by defining a minimal set of concepts that such a setup must define, thus bringing to light the tacit assumptions underlying state spaces and evolution laws. We want to find notions that are equally valid and useful within different settings (e.g. thermodynamics, dynamical systems, classical mechanics, ...).

An excellent review of the history of thermodynamics and classical statistical mechanism, and therefore of the different concepts of entropy, has been compiled by Uffink (cite UFFINK). As for terminology, by Shannon entropy we will generally refer to the family of expressions based on the $p \log p$ formula. This will include the Shannon entropy, it's extension to the continuous, the Gibbs entropy and the Von Neumann entropy. By postulate entropy we will mean the one defined by the fundamental postulate of statistical mechanics, which is the logarithm of the count of microstates. We will not be discussing the Boltzmann entropy specifically, as this can be understood as a special case of the Gibbs entropy where all particles are identically and independently distributed. When correlations between particle/molecules are present, the Boltzmann entropy fails to correspond to the thermodynamic entropy.

2.1. Preamble

Before we start, we list of a set of problems this approach will need to solve. We also include common points of confusions on entropy, thermodynamics and statistical mechanics.

Success of entropy outside of physics. As entropy maximization is increasingly used outside of physics, we need a coherent explanation as to why that happens. The postulate entropy at this level is not insightful: it is unlikely that the primary explanation for ecological distributions is fluctuations of atoms. Therefore we need a principled account for the entropy maximization on more general grounds.

Entropy increases for dynamical equilibria. For a dynamical system, the presence of equilibria (or more in general attractors) leads to entropy decrease, which is the opposite of what one would expect. Take a dissipative system, like a damped harmonic oscillator: areas in phase space will shrink which leads to a lower count of states and therefore a lower postulate entropy; distributions will become more concentrated which leads a lower Shannon entropy. Therefore, as we approach equilibrium, the entropy decreases which is the opposite of what we would expect. This needs to be reconciled.

Ontological vs subjective view of entropy. On the nature of entropy, there are two positions that we find untenable. One position takes entropy to be an objective "ontological" quantity, ontological in the sense that, like charge or position, is an inherent property of the single instance of the system, such as total energy or total momentum. This does not work as the entropy is a property of the ensemble: for the Shannon entropy is a functional of the distribution ρ and for the postulate entropy is the volume of a phase space region. The other position takes entropy to be the subjective information one has about the system, which also does not work. The ensemble is connected to possible fluctuations of the system which are measurable and independent of what we may or may not believe about the system. The idea here is that the notion of equilibrium is process dependent: it depends on both the system and the environment. It depends what happens at the boundary. Even isolation is a type of boundary, which is also technically impossible (e.g. we cannot isolate from gravitational interactions). Therefore entropy should be epistemological but objective: it quantifies the level of granularity/information (epistemological) that is accessible through a particular process allows (objective).

Information content is time dependent. Both the postulate and the Shannon entropy are time independent. This presents a problem because the same statement at different times can provide different information. Take *"Mark is in bed"*. If it is 2 am, the statement is very likely to be true, therefore provides little information. If it is 1 pm, the statement is very unlikely to be true, and therefore provides more information. In the same vein, if we have a dissipative system, *"the energy of the system is less than 3 J"* provides less and less information as time goes on. It should be clear that a proper definition of entropy that would work far from equilibrium will need to take this into account.

Equilibria and relativity. Another problem related to time dependence is that the notion of theromdynamic equilibria is not, in general, relativistic. The notion of "equilibria as nothing is changing" is not invariant under boost. A volume of gas would be at equilibria only if at rest. For a boosted frame, something would be changing (i.e. the position). This also suggests an automatic interaction with gravitational forces: a volume of gas at rest would follow geodesics for other observers. Another failure of thermodynamic equilibria in a relativistic context is the following. Suppose a volume of gas is at rest for an observer reaches equilibrium at time t_0. This means there is a region of space Δx that after t_0 can be considered at equilibrium. But equal time for a boosted frame is different, so different observers will observe different parts reaching equilibrium at different times.

Thermodynamic reversibility vs dynamical reversibility. There are different notion of reversibility. Thermodynamic reversibility is the existence of a process that "undoes the change". This is problematic as this process

technically does not exist in many cases. Dynamical reversibility is the ability to reconstruct the initial state. This is also called by some retrodictability or reverse determinism. A dissipative system should be regarded as thermodynamically irreversible, yet is technically retrodictable.

Units for entropy. Units are often stripped out and not discussed, which leave the mathematics with imprecise physical meaning. If entropy is associated to phase space volume, we should note that the unit of phase space volume depends on the number of degrees of freedom of the system. This means that the entropy of two systems with different number of particles is not numerically comparable as the areas are not commensurable. We need a precise account for this in the math.

Entropy of a single microstate. The entropy of a single microstate is minus infinity, and not zero. A single microstate occupies zero area and the logarithm of zero is minus infinity. In terms of Gibbs/Shannon entropy, a single microstate would correspond to a delta function. The infinite density at that point contributes minus infinite entropy, and since the distribution is zero everywhere else, there are no other contributions. In quantum mechanics, all pure states have zero von Neumann/Shannon entropy. In a way, that is the problem that quantum mechanics fixes.

Entropy over continuous variables. The definitions of entropy are in general not invariant under coordinate transformations (i.e. change of continuous variables). It is the structure of phase space in classical mechanics and of the quantum Hilbert space that make entropy invariant. It can be shown that classical phase space (i.e. symplectic manifolds) is exactly the structure needed for a coordinate invariant (see [3]).

2.2. Descriptions and logical relationships

The first step is to give the most basic mathematical structure we must have to be able to describe a process at all levels of granularity. This consists of a purely logical structure that imposes logical relationships between all possible descriptions.

Process domain. Our starting point is the **process domain** $\bar{\mathcal{P}}$, which is the set that contains all possible descriptions of the system at all levels of detail at all possible times. Depending on the process, it will contain statements like *"the average volume of the system between 1 and 2 seconds is between 3 and 4 liters"* or *"the trajectory of the particle is $y = 10\,m - t^2 \cdot 9.80665\,m/s^2$"*. It does not matter at this point what the statements themselves are, except they must be, at least in line of principle, defined from experimentally well-defined starting points. The translation of these physical requirements into a formal structure has already been carried out in previous work (see [1]). Mathematically, $\bar{\mathcal{P}}$ is a σ-complete Boolean algebra[1] and it is the closure of the set of experimentally verifiable statements $\mathcal{P} \subseteq \bar{\mathcal{P}}$.[2]

Logical relationships. As $\bar{\mathcal{P}}$ contains different levels of description, some statements will be more or less specific. For example, we say that *"the horizontal position is between 2.5 and 3 meters"* is **narrower** than *"the horizontal position is between 2 and 3.5 meters"*. Note that whenever the first statement is true, the second one must be true as well. Given two statements $s_1, s_2 \in \bar{\mathcal{P}}$, we formally capture this relationship by noting $s_1 \preccurlyeq s_2$ (read "is narrower than"). Narrowness can describe relationships at different scales, on different quantities (e.g. *"the horizontal position is between 2.5 and 3 meters and the vertical position is between 1 and 1.5 meters"* $\preccurlyeq$ *"the horizontal position is between 2.5 and 3 meters"*), constraints between variables at the same time (e.g. *"the temperature of the water in the glass is 3.98 °C"* $\preccurlyeq$ *"the density of the water in the glass is 1 g/cm^3"*) or at different times (e.g. *"at time 0 s the position is 1 m and the velocity is 1 m/s"* $\preccurlyeq$ *"at time 1 s the position is 2 m"*). It is a general tool to capture relationships within the theory which intuitively can also be thought of as implication.[3] Mathematically, a Boolean algebra is also a lattice (i.e. a partially ordered set with supremums and infimums) and $\preccurlyeq$ is the ordering relationship for $\bar{\mathcal{P}}$.

We can define other two logical relationships, which will be useful later. For example, we want to say that *"the horizontal position is between 2.5 and 3.5 meters"* is **compatible** with *"the horizontal position is between 2 and 3 meters"*, meaning that they can be true at the same time. Given two statements $s_1, s_2 \in \bar{\mathcal{P}}$ we denote $s_1 \asymp s_2$ (read "is compatible with") if it is possible for them to be both true. Finally, we want to capture whether the truth of one or more statements influences the truth of others. For example, we want to say that *"the temperature of the water in the glass is 3.98 °C"* is **independent** from *"the volume of the water in the glass is 10 cL"* but is not independent from *"the density of the water in the glass is 1 g/cm^3"*. We note $s_1 \perp\!\!\!\perp s_2$ (read "is independent of") when two statements are independent.

Possible evolutions, topology and σ-algebra. Within our process domain $\bar{\mathcal{P}}$ we can define the set E of the narrowest possible statements. Its elements give a complete description of our system at all times and therefore we

---

[1] A set of statements that is closed under negation (NOT), countable disjunction (OR) and countable conjunction (AND).

[2] The set $\mathcal{P}$ will form a Heyting algebra which, to be experimentally reachable, must allow a countable basis. The closure $\bar{\mathcal{P}}$ is in terms of negation and countable disjunction.

[3] Technically, implication in classical logic is a truth function as classical logic does not incorporate the idea of different "possible cases". What we have is more similar in spirit to semantic consequence in modal logic, though modal logic brings in a lot of undesirable elements we do not want.

call them **evolutions** while we call $E \subset \bar{\mathcal{P}}$ the set of all possible evolutions. In the Assumptions of Physics framework, E is the set of possibilities.

Every statement $\mathsf{s} \in \bar{\mathcal{P}}$ can now be characterized by the set $A(\mathsf{s}) \subseteq E$ of all evolutions compatible with s, that is all the evolutions for which s will be true. Mathematically, $A(\mathcal{P})$ is a **topology** over E while $A(\bar{\mathcal{P}})$ forms a σ-algebra over E, the **Borel algebra** of $A(\mathcal{P})$. Narrowness and compatibility become set relationships in the σ-algebra: $\mathsf{s}_1 \preccurlyeq \mathsf{s}_2$ if and only if $A(\mathsf{s}_1) \subseteq A(\mathsf{s}_2)$ (i.e. all evolutions for which s_1 is true are such that s_2 is also true) and $\mathsf{s}_1 \approx \mathsf{s}_2$ if and only if $A(\mathsf{s}_1) \cap A(\mathsf{s}_2) \neq \varnothing$ (i.e. two descriptions are compatible if there is an evolution that is compatible with both). This connection gives a direct physical meaning to these foundational set-based mathematical structures.

Symbol	Name	Meaning
$\mathcal{P}$	Process domain	The set of all possible descriptions at all times and at all levels of granularity
$E \subset \mathcal{P}$	Possible evolutions	The set of descriptions that give a full account of the process; they correspond to the narrowest statements within $\bar{\mathcal{P}}$
$A(\mathsf{s})$	Compatible evolutions	The set of evolutions that are compatible with the statement s
$\preccurlyeq$	narrower	A statement is narrower than another $\mathsf{s}_1 \preccurlyeq \mathsf{s}_2$ if the second one is true whenever the first one is
$\approx$	compatible	A statement is compatible with another $\mathsf{s}_1 \approx \mathsf{s}_2$ if they can be true at the same time
$\perp\!\!\!\perp$	independent	Two statements are independent $\mathsf{s}_1 \perp\!\!\!\perp \mathsf{s}_2$ if the truth of one does not influence the truth of the other

TABLE I. Process descriptions and their logical relationships

2.3. Granularity and quantifying evolutions

Now that we have a way to track all possible descriptions about a process and their logical relationship, we need to characterize their granularity. It turns out that this is equivalent to counting how many evolutions are compatible with each description: the more fine-grained the description, the fewer evolution (i.e. the fewer cases) it will identify.

Insufficiency of a single measure. The next step is to quantify the granularity of the description provided by each statement. Since statements at a finer level of description will put a greater constraint on the set of compatible evolutions, quantifying the granularity of the description of s means quantifying the size of $A(\mathsf{s})$. The statement *"at time 0 s the particle is between 0 m and 1 m"* is doubly more specific than *"at time 0 s the particle is between 0 m and 2 m"* precisely because it corresponds to half the evolutions of the second statement.

The naive approach would be to assign a measure over $A(\bar{\mathcal{P}})$: since this is a σ-algebra over E and a measure is exactly what one uses to give a size to each set, it would seem we are done. Unfortunately it is not that simple: a single measure can only compare objects of the same dimensionality. If we imagine all evolutions to be points on an n-dimensional manifold, if the measure gives finite values for k-dimensional regions it will necessarily give measure zero for all regions with lesser dimensionality and infinite measure for those with greater dimensionality. Coarse grained and subsystem descriptions will have different dimensionality as they describe different numbers of degrees of freedom, therefore we need a way to quantify objects at all levels.[4]

Fineness and equigranularity. Our starting point will be the ability to compare two statements and decide which one gives a finer description of the process. Formally, we have preorder $\leq$ on $\bar{\mathcal{P}}$ which, given two statements, tells us whether the description of one is **finer**, more refined, than the other. Two statements are **equigranualar**, noted $\mathsf{s}_1 \doteq \mathsf{s}_2$, if they provide the same level of description. For example, we can say that *"the trajectory of the particle is $y = t \cdot 1\,m/s$"* $\doteq$ *"the trajectory of the particle is $y = t \cdot 2\,m/s$"* and that *"the position at time 0 sec is between 0 and 1 m"* $\doteq$ *"the position at time 0 sec is between 1 and 2 m"*, even though the second pair is infinitely less discerning than the first. Note that if one statement is narrower than the other (i.e. $\mathsf{s}_1 \preccurlyeq \mathsf{s}_2$) then it is also finer than the other (i.e. $\mathsf{s}_1 \leq \mathsf{s}_2$) while the converse is not necessarily true.

---

[4] Typically one uses geometrical structures for his purpose. In phase space, the symplectic form allows us to define areas of two-dimensional surfaces, and use those to define volumes or even-dimensional areas. Note that it does not quantify the size of the points, which are implicitly assumed to be of equal size. This is a special case and cannot be the starting point.

We say two statements are **comparable** if one is finer than the other. The preorder $\leq$ is partial because not all statements are comparable to each other. For example, consider the pressure/volume state space for an ideal gas. Comparing *"the pressure is 1 kPa and the volume is between 1 and 2 liters"* and *"the pressure is between 1 and 2 kPa and the volume is 1 liter"* would mean saying whether one liter is bigger or smaller than one kPa.[5] The idea of physical dimension would therefore be something that is formally captured mathematically.

The idea is that geometrical, measure theoretic, information theoretic and probabilistic structures all descend from this more fundamental structure, the partial order that describes this "information granularity". We note that all these structures can be used to determine a size and therefore a partial order on the set. It would be possible, given sufficient conditions, to recover those structures from the order itself. The mathematical detail of how exactly this would work (i.e. what conditions are necessary and sufficient) is out of the scope of this work (see CITE other blueprint). Here we will simply assume that the process domain is such that the geometrical structures can be rederived.

Measures of granularity. To recover measures, we pick a statement $u \in \bar{\mathcal{P}}$, called unit, and construct a **measure** μ_u that is normalized (i.e. $\mu_u(u) = 1$), monotonic (i.e. $s_1 \leq s_2$ will mean $\mu_u(s_1) \leq \mu_u(s_2)$) and additive (i.e. $\mu_u(s_1 \vee s_2) = \mu_u(s_1) + \mu_u(s_2)$ if $s_1 \not\leq s_2$).[6] That is, we can quantify how much more or less precise a statement is compared to a fixed statement taken as a unit. If two statements are such that the measure of one with respect to the other is finite and non-zero, we say they are **finitely comparable**. We can show that if u_1 and u_2 are finitely comparable we have $\mu_{u_1}(s) = \mu_{u_1}(u_2)\mu_{u_2}(s)$. Furthermore, since all evolutions give a complete description of the whole process, we will assume they are all equigranular. Therefore if we pick $u \in E$, then $\mu_u(s) = \#(A(s))$ would be the counting measure of evolutions compatible with s.[7]

Independence. If we fix the evolution of one subsystem, it should be intuitive that the possible evolutions of another subsystem depends on the correlation between the two. The maximum number of joint evolutions corresponds to the case were the systems are independent: the choice of evolution for one does not contrain the choice of evolution for the other. In this case, we expect the measure to factorize. Mathematically, we take two subdomains $\bar{\mathcal{P}}_1 \subset \bar{\mathcal{P}}$ and $\bar{\mathcal{P}}_2 \subset \bar{\mathcal{P}}$. By subdomains we mean a set of descriptions that also form a σ-complete Boolean algebra. We say the two subdomains are **independent** if knowledge of one does not constrain knowledge of the other. Mathematically, any pair of statements $(s_1, s_2) \in \bar{\mathcal{P}}_1 \times \bar{\mathcal{P}}_2$ from the two domains is independent $s_1 \perp\!\!\!\perp s_2$. As we expect the count of evolution to factorize, we have $\mu_{u_1 \wedge u_2}(s_1 \wedge s_2) = \mu_{u_1}(s_1)\mu_{u_2}(s_2)$. In terms of the math, this can be imposed but it would be nice to show that this has to be the case. It can be shown for finite sets. It is not clear whether it must be for infinite ones. Also, note the units of the measures are different. As an analogy to classical phase-space, the area of each degree of freedom can be expressed in units of $\hbar$. The volume of two degrees of freedom, however, is in units of $\hbar^2$. This is an additional detail that needs to be fully understood mathematically in a way that is typically glossed over in statistical mechanics.

Process entropy. We are now ready to define the **process entropy** with respect to u as the quantity $S_u(s) = \log \mu_u(s)$. This can be thought of as the number of bits, yes/no questions, that separate the level of description of u from the one provided by s. Note that process entropy combines linearly for independent systems: $S_{u_1 \wedge u_2}(s_1 \wedge s_2) = \log \mu_{u_1 \wedge u_2}(s_1 \wedge s_2) = \log(\mu_{u_1}(s_1)\mu_{u_2}(s_2)) = \log \mu_{u_1}(s_1) + \log \mu_{u_2}(s_2) = S_{u_1}(s_1) + S_{u_2}(s_2)$. Also note that there is a unit dependence, so one can't simply compare the entropy of two different systems as they would be, most likely, expressed in different units. It typically does not pose a problem because we compare entropy of different description of the same object. This issue already exists with the other entropies, we are simply making in explicit so that we can study it.[8]

Recovering the postulate entropy. Entropy as the logarithm of the state count can be easily recovered as a special case. Suppose the process is deterministic and reversible in the sense that the state at one time identifies the whole evolution. Suppose that all state descriptions at one time are equigranular. In that case, the count of the evolution corresponds to the count of states and therefore the process entropy coincides with the postulate entropy.

Recovering the Shannon entropy. One way to recover the Shannon entropy (though probably not the only one) is to assume we have a process that is fluctuating according to a stable distribution $\rho(s)$. That is, our time resolution is fixed at a scale Δt, the state s keeps changing within that interval. The distribution $\rho(s)$ represents how often the system will visit the state s within that timeframe. To calculate the process entropy we need to calculate how many possible evolutions we would have within Δt that satisfy $\rho(s)$. An evolution is a sequence of states, one for each instant in time. If we assume the evolution is continuous, countably many states will suffice to identify one evolution. Therefore we are looking to count the possible permutations of a countably infinite sequence with recurrence set by $\rho(s)$. But this is exactly one way to recover the Shannon entropy. Therefore, in this case, the process entropy coincides with the Shannon entropy.

---

[5] In phase space, for example, we cannot in general compare q^i and p_i, yet we can always compare areas of phase space

[6] Differentiating between finite or countable additivity is beyond the scope.

[7] We still need to understand the necessary and sufficient conditions under which such measures would exist and be unique. For the present work, we assume it can be done.

[8] The phase space volume will be in units $\hbar^n$ where n is the number of degrees of freedom, typically $n = 3m$ where m are then number of molecules. While $3 < 4$, there is no sense in which $3\hbar^6$ is smaller than $4\hbar^{12}$ therefore comparing the entropy of gasses with different number of particles is not well defined.

Probability. Quantifying evolutions allows us to define probability. The idea is a process can be realized differently by each possible evolution. The more evolutions a statement is compatible with, the more it is likely to happen. For example, consider a process that implements a coin flip. If the coin (and the process flipping the coin) is fair, we must have that half the evolutions are compatible with *"the coin is head"* while the other half are compatible with *"the coin is tails"*. Therefore we define that the **probability** of s_2 given s_1 is

$$P(s_2|s_1) = \mu_{s_1}(s_1 \wedge s_2) = \frac{\mu_u(s_1 \wedge s_2)}{\mu_u(s_1)} \tag{1}$$

where $s_1 \wedge s_2$ is the logical conjunction (AND) between the two statements and u is finitely comparable to s_1. In other words, the probability quantifies the ratio of evolutions compatible with s_1 that are also compatible with s_2. This can be shown to satisfy the standard axioms of probability.

Symbol	Name	Meaning
$\lesssim$	finer	A statement is finer than another $s_1 \lesssim s_2$ if it provides a more refined description than the second
$\doteq$	equigranular	A statement is equigranular to another $s_1 \doteq s_2$ if they provide the same level of detail
$\mu_u(s)$	Measure	Quantifies the precision of a statement s in terms of a reference unit statement u; it quantifies how many evolutions are compatible with s compared to the evolutions compatible with u
$S_u(s)$	Evolution entropy	Defined as $\log \mu_u(s)$; represents the number of bits that separate the level of description of s from the one provided by the unit statement u; it quantifies the number of questions needed to go from the level of description of s to the level of description of u
$P(s_2|s_1)$	Probability	Defined as $\mu_{s_1}(s_1 \wedge s_2)$; quantifies the fraction of evolutions compatible with s_1 that are also compatible with s_2
	Comparable	Two statements are comparable if one is finer than the other
	Finitely comparable	Two statements are finitely comparable if the measure of one in terms of the other is finite and non-zero

TABLE II. Granularity, measures and probability

2.4. Time, predictions and reconstructions

Now that we have a way to keep the logical structure and granularity of all descriptions of a process, we want to understand how the information is transported back and forward in time by the process. That is, once a particular description is verified at a particular time, what can we say about past and future times.

Time domains. As a process extends over time and potentially describes multiple systems, we want to organize the statements into subdomains, one for each moment in time relative to a particular system. Given a **time parameter** $t \in T \subseteq \mathbb{R}$, we can imagine carving out a **system domain** $\bar{\mathcal{D}}_t$ that would correspond to all the statements about the system one can make at the given time. In particular, we define **one-step process** as one that only has an **initial domain** $\bar{\mathcal{D}}_{t_0}$ and a **final domain** $\bar{\mathcal{D}}_{t_1}$. Note that $\bar{\mathcal{D}}_t$ is a σ-complete Boolean sub-algebra of $\bar{\mathcal{P}}$ since any logical operation on statements at fixed t is also another statement at fixed t.

Temporal and scale coarse graining. Technically, the statements are not defined at an infinitesimal moment t, but rather within a finite interval $[t, t + \Delta t]$. This represents the time-scale at which the process will be described, meaning the description within $\bar{\mathcal{D}}_t$ should not be sensitive to what happens at faster scales. The quantities used to represent states at time t, then, should really be thought of as averages within Δt. In the same way, $\bar{\mathcal{D}}_t$ will in general represent the system at the chosen level of description, not the narrowest level possible. Statements about the environment or at a finer level of description (i.e. the positions of all molecules for a gas) will not be included.

Optimal coarse graining. In principle, this coarse graining is arbitrary and subjective in the sense that is completely at the mercy of the decision of the scientist. However, not all choices are equivalent. Ideally, we want the process at our level of description to not depend on the faster or smaller scale dynamics. If the dynamics at faster and smaller scale decouples from the one at slower and bigger scale, we can study the latter without worrying about the former. So, for example, we can study the trajectory of a cannonball without worrying about the motion of each individual molecule; we may not be able to predict the exact energy at a future time, as it fluctuates through interactions with the environment, but we may be able to predict it's average over a small Δt. We should stress here that **the process itself tells us whether and how the faster/smaller scale decouples decouple from the slower/bigger scale**. Therefore the choice of an **optimal** coarse graining is not arbitrary or subjective at all: it is determined by the process. The processes at our disposal guide us to carve domain at both temporal and size scale, and finding the optimal level of scale is part of the job of a physicist. At a theoretical level, we need to understand how a process guides us to that choice.

Predictions and reconstructions. A crucial step, then, is to characterize how the information at one time is carried to other times by the process. Suppose we found that the statement $s_0 \in \bar{\mathcal{D}}_{t_0}$ at time t_0 was true. What can we predict of the system at a future time $t \geq t_0$? Any future statement $s \in \bar{\mathcal{D}}_t$ such that $s_0 \preccurlyeq s$ will be true if s_0 is true. Therefore the best prediction will be the conjunction (the logical AND) of all such statement. In terms of order theory, it will be the narrowest broader statement. As all statements can be thought as sets of evolutions, it will be statement $s \in \bar{\mathcal{D}}_t$ such that $A(s)$ is the smallest set that contains $A(s_0)$.

In a similar way, we can ask what can be reconstructed about the system at a previous time $t \leq t_0$. Any past statement $s \in \bar{\mathcal{D}}_t$ such that $s_0 \preccurlyeq s$ will be true if s_0 is true. Therefore the best reconstruction will be the conjunction of all such statements, the narrowest broader statement, the smallest set $A(s)$ that contains $A(s_0)$.

Given that the formal definition for predictions and reconstructions are similar, we note $P_t(s_0)$ the prediction or recostribution for statement s_0 at time t.

Predictions and reconstructions always have higher entropy. Prediction and reconstruction are necessarily broader than the statement we start from and therefore they will necessarily be coarser: they cannot give us more information then what we started with. They may either give us the same information or less information. Mathematically, $s_0 \preccurlyeq P_t(s_0)$ for any s_0, t_0 and t. In this respect, there is no time asymmetry: information can only be diminished as we map it to past or future time.

The process dictates the relationship between the information at different moments in time. How much the information diminishes depends exclusively on the process. Suppose we take a deck of cards and memorize the order and leave it on the table. If sits undisturbed, full knowledge is maintained. If a person takes the top card and puts it at the bottom, full knowledge is still maintained. If a person shuffles the deck, the information about the order of the deck is lost. If a person cuts the deck, the information about the first card is lost, yet the information about the order is not.

Natural graining. As each process defines how information is transported over time, it will define optimal granularity levels for its description. We say $s_0 \in \bar{\mathcal{D}}_{t_0}$ **cuts the process along the grain**, or simply **is a natural cut**, if its prediction and reconstructions do not lose information. That is, $s_0 \doteq P_t(s_0)$ for all t. The certainty and the impossibility are both natural cuts of all process. More work needs to be done to understand the role of natural graining in a process, but the general sense is this would be a key element for a better definition of states and equilibria because of the following properties.

Natural graining is automatically relativistic. Consider a free particle, whose state at a given time x_0 is fully determined by position and momentum. This will give us predictions and reconstructions at the same level of granularity for all times, therefore x_0 cuts along the grain. It doesn't matter whether the particle is at rest or in motion. It doesn't matter whether "all properties" remain the same.

Natural graining is compatible with both dynamic and thermodynamic equilibrium. Continuing with the card deck example, suppose a person takes the deck of card, cuts in half, shuffles the bottom half and orders the top part by suit and rank. This process creates an equilibrium that is, in a way, a mix of dynamic (the top part) and thermodynamic (the bottom part) equilibria. If we apply the process another time, in fact, the top part remains exactly in the same configuration; the bottom part, while it will most likely change, will retain its statistical description. Each equilibrium is fully identified by the cards in the top part. If we take the full description of the deck at the beginning, the prediction is coarser, therefore it is not a natural cut. The full description of the top half of the deck will also give a coarser prediction: the final order of the top half could have been reached by any initial order of the same cards in the top half. The description of what cards are in the top half, however, gives us exactly the information at equilibrium. Therefore it is a natural cut. This shows that the idea of natural cuts captures both types of behavior, even when they happen at the same time.

Natural graining is time independent. One of the problems with the notion of equilibria is defining when the equilibria is reached. Natural graining is a time independent concept, so we do not need to define when it is reached. If we consider a dynamical equilibrium, we have a basin of attraction for an equilibrium. All evolutions that

start in the basin of attraction will tend closer and closer to the equilibrium, without ever reaching it. The natural graining consists in the set of all evolutions taken together. In many cases, we are not interested in a specific evolution. For example, we have a cutoff at a certain time and we are only interested to note which equilibria the system was approaching. Mathematically, what needs to happen is that at the cutoff the description is a natural cut for both the previous and following process. At that point the behavior in the limit is not important, which makes the framework more compatible with realistic scenarios.

Natural graining decouples the internal dynamics. Consider a macroscopic object, like a balloon. Though we know it is formed by on the order of 10^23 molecules, we can still characterize its state with a few variables. Each molecule itself, maybe formed by a large number of fundamental constituents. At each level, the description is fairly independent of the description of the constituents. Natural graining picks out those descriptions that are effectively independent of the faster/smaller scale.

2.5. Snapshots, determinism and reversibility

In the previous section we explored how the information at a particular time moves across a process. Now we want to concentrate on the best possible descriptions of the system at each time. That is, how the descriptions with maximal information at each time relate to each other. This will give us straight forward definitions for determinism and reversibility.

Snapshots and trajectories. For each time domain $\bar{\mathcal{D}}_t$ we can find the set of possibilities X_t, which represent the statements that will give the most precise description at that time; we call these the possible **snapshots** of our system at that time. A **trajectory** is a sequence $x(t)$ of snapshots at each time. A particular evolution will identify a trajectory: that is, given $\mathsf{e} \in E$ we can find $x(t)$ such that $\mathsf{e} \preccurlyeq x(t_0) \wedge x(t_1) \wedge x(t_2) \wedge \ldots$. The converse is not necessarily true because $\bar{\mathcal{D}}_t$ will typically describe a particular system at a particular level of granularity and a particular time-scale. Therefore a single trajectory may correspond to multiple evolutions, since the evolutions potentially describe the same system at greater accuracy and/or other systems.

Snapshots are not states. One key insight is that the set of possible snapshots X_t may not be the same at all times and may not correspond to the state of the system. First of all X_t will only include those configurations that are compatible with at least one possible evolution (i.e. $A(x) \neq \varnothing$ for all $x \in X_t$), which may change in time. For example, the set of possible configurations for a system under dissipative processes will shrink as they converge to equilibrium. In quantum mechanics, the possible configurations after a measurement are restricted to the eigenstates, and a different choice of measurement will lead to a different process with different possible evolutions. Secondly, the best possible description may be broader than the full state of the system (e.g. when external interference forces us to give only a statistical account within our Δt) or narrower (e.g. when there are known correlations to the other degrees of freedom of the environment or of the system itself). In other words: the snapshots depend on the process and on other systems within the same process. These correlations prevents the snapshots to be proper states of a system, since they are not descriptions about and only about the system.

Determinism. We say a process is **deterministic** over the system domains $\{\bar{\mathcal{D}}_t\}_{t \in T}$ if given the snapshot at one time we can always predict the snapshot at all future times. Formally, let $t_1, t_2 \in T$ such that $t_1 \leq t_2$, for any $x_1 \in X_{t_1} \subset \bar{\mathcal{D}}_{t_1}$ we can find an $x_2 \in X_{t_2} \subset \bar{\mathcal{D}}_{t_2}$ such that $x_1 \preccurlyeq x_2$. Intuitively, all the evolutions that pass through x_1 will also pass through x_2. If $x(t)$ is the snapshot trajectory in time, we have $A(x(t)) \subseteq A(x(t + \Delta t))$: the set of compatible evolutions must stay the same or become larger. This also means that the process entropy associated to a snapshot cannot decrease during deterministic evolution:

$$S_{\mathsf{u}}(x(t)) \leq S_{\mathsf{u}}(x(t + \Delta t)) \tag{2}$$

since $x(t) \preccurlyeq x(t + \Delta t)$ means $x(t) \leq x(t + \Delta t)$ and $\mu_{\mathsf{u}}(x(t)) \leq \mu_{\mathsf{u}}(x(t + \Delta t))$.

Law of evolution. Deterministic processes are exactly the ones for which a law of evolution can be written. For each $x_1 \in X_1$, the final snapshot $x_2 \in X_2$ must be unique as all snapshots in X_2 are incompatible with each other. Therefore we can write a function $f : X_1 \to X_2$ such that $x_1 \preccurlyeq f(x_1) \equiv x_2$ that describes that particular step in the process. We call f the **law of evolution**. The law of evolution fully characterize the deterministic process.

Scientific privileging of deterministic processes. Deterministic processes play a special role in science because, insofar we want to have theories that make prediction, we are going necessarily focus on deterministic processes. This privileging is not due to nature per se: it is due to the practice of scientific investigation. We setup initial conditions and measure results, not the other way around. The time asymmetry is built into the practice of science as a fundamental trait. As such, it is likely to be taken as starting point, and cannot further be explained through scientific investigation itself.

Reversibility. Conversely, we say a process is **reversible** over the system domains $\{\bar{\mathcal{D}}_t\}_{t \in T}$ if given the snapshot at one time we can always reconstruct the snapshot at all past times. Formally, let $t_1, t_2 \in T$ such that $t_1 \leq t_2$, for any

$x_2 \in X_{t_2} \subset \bar{\mathcal{D}}_{t_2}$ we can find an $x_1 \in X_{t_1} \subset \bar{\mathcal{D}}_{t_1}$ such that $x_2 \preccurlyeq x_1$. In this case, all the evolutions that pass through x_2 must have passed through x_1. If $x(t)$ is the snapshot trajectory in time, we have $A(x(t)) \subseteq A(x(t - \Delta t))$: the set of compatible evolutions must stay the same or become smaller. This also means that the process entropy associated to a snapshot cannot increase during reversible evolution:

$$S_{\mathsf{u}}(x(t)) \leq S_{\mathsf{u}}(x(t - \Delta t)) \tag{3}$$

since $x(t) \preccurlyeq x(t - \Delta t)$ means $x(t) \leq x(t - \Delta t)$ and $\mu_{\mathsf{u}}(x(t)) \leq \mu_{\mathsf{u}}(x(t - \Delta t))$.

Law of inverse evolution. In a dual fashion, reversible processes allow us to write a law of inverse evolution. For each $x_2 \in X_2$, the initial snapshot $x_1 \in X_1$ must be unique as all snapshots in X_1 are incompatible with each other. Therefore we can write a function $g : X_2 \rightarrow X_1$ such that $x_2 \preccurlyeq g(x_2) \equiv x_1$ that describes that particular inverse step in the process. We call g the **law of inverse evolution**.

Determinism and reversibility. A deterministic and reversible process, then, will mean x_1 and x_2 are equivalent and that $A(x(t)) = A(x(t + \Delta t))$: the set of evolutions remains the same. This also means that the process entropy remains conserved:

$$S_{\mathsf{u}}(x(t)) = S_{\mathsf{u}}(x(t - \Delta t)) \tag{4}$$

The fact that past and future snapshots are equivalent $x(t) \equiv x(t + \Delta t)$ does not mean they are the same statement in terms of the description at each time: it means there is an if-and-only-if relationship between the descriptions at different times. For example, *"at time t the position is q and the velocity is v"* if and only if *"at time $t + \Delta t$ the position is $q + v\Delta t$ and the velocity is v"*. The differential equation $\dot{v} = 0$ is a short hand for all such relationships.

Entropy increase and irreversibility. According to these definitions, **entropy cannot decrease in deterministic processes**. Additionally, **reversible process (that are also deterministic) will conserve entropy**. For non-deterministic processes this may not be the case, therefore this is not an absolute law. However, insofar we want to write theories that give predictions, the result will apply. Therefore it is not that entropy never decreases, it never decreases for the processes we are interested in: the ones where we are able to properly define systems and make predictions. We can say that it is a **fundamental law of scientific theories but not a fundamental law of nature**.

Reversibility vs retrodictability. If the process is deterministic and reversible we have that the inverse of law of evolution $f^{-1} = g$ is equal to the law of inverse evolution. It is possible for f to be invertible even if the system is *not* reversible. In fact, we would have $x_1 \equiv f^{-1}(x_2) \preccurlyeq x_2$, which is not the same as $x_2 \preccurlyeq g(x_2) \equiv x_1$: narrowness is in the opposite direction. In other words, **an invertible deterministic system is not necessarily reversible**. For example, consider a damped harmonic oscillator: the dynamics is in principle invertible (we can reconstruct the trajectory) but it is not reversible (as the evolutions bunch together, finite precision knowledge of the initial conditions gives us more information than the same finite precision knowledge of the final conditions). These types of distinctions are crucial and they emerge naturally within the framework. It can be shows that this is possible only for systems identified by continuous variables, and not when all variables are discrete.

TODO: picture with deterministic evolution merging "streams" of evolutions, and detrev evolutions not merging

Symbol	Name	Meaning
$t \in T$	Time parameter	Time is treated as a real valued parameter within the range $T \subseteq \mathbb{R}$
$\mathcal{D}_t \subset \mathcal{P}$	System domain	The set of descriptions at a given time
$X_t \subset \mathcal{D}_t$	Possible snapshots	The set of descriptions that give a full account at time t; they correspond to the narrowest statements within $\bar{\mathcal{D}}_t$
$x(t) \preccurlyeq x(t + \Delta t)$	Determinism	Given the snapshot at one time we can predict the snapshot at future times
$x(t) \preccurlyeq x(t - \Delta t)$	Reversibility	Given the snapshot at one time we can reconstruct the snapshot at past times

TABLE III. System domain, snapshots, determinism and reversibility

2.6. System, states and state spaces

Systems. Lastly we need to answer the following question: how do we go from a set of descriptions to the definition of a system? Under what conditions do the snapshot actually represent states? And why do we do it? The general

idea is that, conceptually, we want to study the same object under different processes. That is, we want to say that the same object X can be subjected to different process at different intervals of time, and we are able to switch from one to the other. Note that this is not simply something we would like to do: it is a basic requirement to do science. Process A, in fact, can be the reliable preparation of a sample, process B can be some interaction we want to characterize, and process C can be the final measurement. What must happen, then, is that the description of the object at each junction is independent from the choice of the previous/future processes and from other objects. In other words, the description at the junction is about and only about the system at that time. This does not mean that it is the *full* description. It must be, then, that we have at our disposal a process under which the object is independent from the rest. We removed all correlations in time and with other systems. We call an object for which this is possible a **system**.

Independence as requirement. Therefore **independence is a requirement for the definition of a system**: we can talk about a system only insofar we have at our disposal a process under which the system is independent for the rest, such that the description of the system is truly a description of the system, with no correlations. We can speak of a chair as its own system because we have circumstances in which we can interact with the chair in a way that does not affect the other chairs or the tables. The definition of a system, then, is conditional to the existence of suitable processes.

What still needs to be clarified is the difference between the independence from other systems of the system coming in or of the process itself. Suppose we have system X and Y whose state is identified by a single continuous variable. If there is no correlation, all state combinations are possible, so the joint distribution is uniform over the whole plain. Now suppose they are correlated so that that $x = -y$. The joint distribution is uniform on the diagonal. If we have a process that acts and depends only on X, the two cases are indistinguishable since the marginal of x is uniformly distributed in both cases. The density of evolutions, however, is different, so there is a renormalization. But for the rest, the two cases are indistinguishable. So independence may not technically always map to factorization of the measure, but it may need to be indistinguishable from factorization.

Independence leads to maximization of entropy. Note that the presence of correlations between the system and the environment can only remove possible evolutions, not add them. Therefore independence is the case where the evolution count is maximal. In other word, requiring independence is equivalent to requiring maximization of entropy. Again note that this maximization of entropy is not due to some property of nature: it is due to the requirement of wanting to define a system, an objects for which we can mix and match processes. Of all processes nature presents us, we select those that decouple the world into separate systems, those that maximize entropy.

Relaxation processes and equilibria as independence. If we are to study systems, then, we must have at our disposal a process that takes an object and removes all possible correlations, rendering it independent. Moreover, this process must keep the system independent if it was already so. Therefore **independent descriptions must be equilibria** of such process. In general, the definition of the system requires the existence of such processes, which we call **relaxation** processes. States will be invariant, symmetries, of such processes.

System definition as object-environment relationship. Whether or not a particular set of attributes of an object can be seen as an independent system is not an a priori feature of the object, but rather a relationship between object and environment. This is not just because of independence, but also for the existence of the system itself. We can talk about a ball, with its position and velocity, on the surface of the earth at 20 Celsius. On the surface of the sun, we cannot talk about a ball. Therefore there must be a tight relationship between states, the processes defined on the system and the object-environment boundary. The process cannot alter the boundary, the state space must satisfy the constraints imposed by the boundary, the state space must be, as a whole, a symmetry of the processes.

System composition and its algebra. As we may want to describe systems, subsystems and their relationships, it useful to note that **the set of all possible systems is a Boolean algebra**. The ordering of the lattice is given by system inclusion. We say $X \in Y$ if X is a subsystem of Y. We can convince ourselves that this is a partial order (i.e. $X \in X$, if $X \in Y$ and $Y \in X$ then $X = Y$ and if $X \in Y$ and $Y \in Z$ then $X \in Z$). The join $X \vee Y$ is therefore the smallest system that contains both and represents system composition. The meet $X \wedge Y$ is the biggest system contained by both and represent the common subsystem. The 1 represents the system that contains everything. The 0 represent the null system, the one that contains nothing. The complement $\neg X$ represents everything except for the system.

Carving up the world. Philosophers talk about a description of reality that is joint-carving meaning that reality already contains divisions, prior to our investigation. This is an epistemological version of it: nature gives us access to processes for which some systems are independent from the rest. It is the accessibility of these processes that allows us to define systems, without them we cannot divide nature into parts. That is, if an object is really made of parts but the process driving their evolution is such that the correlations can never be eliminated or controlled, the parts cannot be defined as independent systems.

States and state spaces. A **state is a snapshot for which the system is independent**. That is, it is a description of the system and only of the system. A **state space** X is the set of all possible states of the system.

States as junction points. States will then act as the juncture point between processes. If we switch from process

A to process B at time t_0, given the requirement of independence, the description of the system at that time must be a state. If that's the case, both processes can be characterized independently from each other.

States as equilibria of faster/smaller scale processes. As noted above, the subdomain must choose both a physical scale and a time scale. The requirement of independence will mean that the description within that scale is not relevant to the processes at hand. This cutoff is, in effect, part of the boundary of the system: we are still defining what descriptions, which degrees of freedom, are under consideration and which are not. Note that this is consistency of the idea of relaxation process, which must therefore not only remove the correlations from other objects but also from internal degrees of freedom.

States as bundles of evolutions over short scale. In this light, is it better to understand states not as points at an instant of time, but rather as bundles of evolutions over short scale that we "connected" throughout the process.

State entropy. When we join two processes, we need to reconcile the evolution count at the junction. This means that the evolution count for states must be well defined and unique if we want to mix and match antecedent and subsequent processes. Note that independence is precisely the condition needed for the count of evolutions to factorize and process entropy to becomes additive. In this case, we can define a **state entropy** $S(x)$ which is the process entropy in condition of system independence. This means that **all state spaces must be endowed with a structure that define the state entropy**.

The entropic nature of geometric structures of state spaces. The claim is that all geometric structures that state spaces possess in different theories are precisely structures that keep track of the state entropy. This is obvious in thermodynamics, where the equation of state in entropric form precisely map state entropy. In classical mechanics, phase space is characterized by a symplectic structure, whose volume is used in statistical mechanics to compute the entropy. Moreover, orthogonality in the symplectic structure characterize the independence of degrees of freedom. In quantum mechanics, the inner product is connected to probabilities that, as we saw, are simply ratios of counts of evolutions. Moreover, those same probabilities are used in statistical mechanics for the computation of the entropy. Also note that these structure do not provide further characterization of the state space beyond the ability to compute the evolution count and the entropy. They provide entropic and only entropic information.

Zero state entropy and the null state. As we saw before, entropy is a relative concept and therefore needs a unit to be defined. What state should act as a reference and be assigned zero entropy? There is a natural choice. Compositing with the null system $X \vee 0 = X$ returns the original system, therefore the entropy must not change. Since the entropy is linear under system composition, the null system must have zero entropy. We should note that, in principle, all systems should have a null state, which is the state where the system is not there. This case is automatically independent and is therefore a state. This state should be given zero entropy and therefore can be taken as the unit of our measure for the state space.

Null state as the lowest entropy state. It should also be intuitive that this state has to have the lowest entropy of all. That is, saying the system is not there will give the finest level of description of the system. No other description can be finer (except for the impossibility). This means that any other state can give a description of the system that is, at best, as fine as the null state. This means that the **the state entropy can never be negative**. This is a more general form of the third law of thermodynamics, in the same way that 2 is a more general form of the second law.

State entropy and time dependence. We should stress that the state entropy fixes the relative process entropy between states at equal time in conditions of independence. In all other cases, or if these conditions changes throughout the process, the connection between state entropy and process entropy is more complicated: the same description for the same system will not correspond to the same count of evolutions. Consider a dynamical system with an attractor, such as a damped harmonic oscillator. As time proceeds, more and more evolutions will converge to the attractor. The same finite region around the attractor will correspond to more and more evolutions. In general, a process may introduce correlations between different variables which influences the value of the process entropy for a specific description.

State space as template. We should think of the state space as a template we can use to instantiate groups of statements at each time. We have a **state domain** $\bar{\mathcal{D}}_X$ which contains all statement fragments like *"the position of the ball is between 2 and 3 meters"* together with a surjective map $\iota : \bar{\mathcal{D}}_X \times T \twoheadrightarrow \bar{\mathcal{D}}_t$ that adds *"at time t"* for each possible time. The map ι must be a surjection that preserves the logical structure (i.e. logical operations, verifiability, narrowness, compatibility, ...) so that the basic logical relationships determined by the system itself are valid at all times (i.e. *"at time t the position of the ball is between 2 and 3 meters"* $\preccurlyeq$ *"at time t the position of the ball is between 0 and 300 meters"* for every t).

Note, though, that $\iota(\cdot, t_0)$ in general is not an isomorphism (i.e. a bijection that preserves the logical structure) for two reasons. The first is that not all states will be available at every moment in every process. As we said before, if the process is dissipative the set of possible states will become smaller. In that case, ι will map all the inaccessible states to $\perp$, as they will be impossible. Therefore ι is not in general a bijection. The second reason is that the statement fragments in $\bar{\mathcal{D}}_X$ represent only descriptions about the system itself and nothing else. In a particular process, however, knowing the state of one system might tell us something about other system as well. The presence

of correlations/coupling may mean that the precise knowledge of the degrees of freedom of the system may allow us to tell something about the environment or the internal structure (i.e. microstate). Mathematically, ι may map to statements that are narrower than the original fragment (e.g. we can have $\iota($ *"position of A is 1 meter"*, $t_0) \equiv$ *"at time t_0 the position of A is 1 meter and the position of B is -1 meter"*).

State spaces are constructed. Though state spaces are template, conceptually they are not the starting point. The state space of the system (i.e. the set of all possible complete descriptions of the system and only about that system, independently of the process at hand) is something we need to construct. We find processes for which the system decouples, and we elevate those descriptions to independent objects. This construction comes with a set of assumptions about the system which are at the foundations of thermodynamics specifically and physics more in general, and they need to be made explicit. The snapshots (i.e. best descriptions of a system at a time), instead, do not require any additional assumptions and therefore are more general concepts.

A rich mathematical structure To sum up, the definition of a system, then, must define several things.

- A boundary, that tells us where the system ends and the environment begins
- A characterization of the boundary, that defines in what cases the system remains well defined
- A set of relaxation processes, that tells us how the system can be decoupled from the rest
- A set of possible independent descriptions (i.e. states), that can be characterized about the system given the circumstances
- A set of possible evolutions for the system (i.e. the processes) under the given circumstances.

There is a lot more work to be done to fully characterize this structure mathematically, but the current physical theory can serve as guidance. The point is that this structure has to serve as a general template that all physical theories must follow. We need a general definition of states and processes of which classical mechanics, quantum mechanics, thermodynamics and statistical mechanics are a specialization. The fact that the theories are, on the surface, so different is actually a good thing as it makes it easier to understand what elements are specific to the theories and what are more general.

---

[1] G. Carcassi and C. A. Aidala, *Assumptions of Physics* (Michigan Publishing, 2021).
[2] G. Carcassi and C. A. Aidala, Assumptions of Physics overview: experimental verifiability and topology (2021), arXiv:2103.06053 [physics.gen-ph].
[3] G. Carcassi and C. A. Aidala, Hamiltonian mechanics is conservation of information entropy, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **71**, 60 (2020).