S = k ln(B(W)): Boltzmann entropy, the Second Law, and the Architecture of Hell

TIM MAUDLIN, NYU & JBI

WARSAW SPACETIME COLLOQUIUM,

MAY 7, 2021

JOHN BELL INSTITUTE FOR THE FOUNDATIONS OF PHYSICS

”Definition” of Theoretical Terms: Conceptual Issues

- A significant part of any scientific account of the world is the development of an appropriate terminology and definitions.
- The question of what constitutes a “correct” definition becomes even more acute when the theoretical vocabulary in one system is being “reduced to” or “defined in terms of” some other system.
- The concept of “entropy” was originally introduced in classical thermodynamics, in a setting which was agnostic about the atomic theory of matter and the kinetic theory of heat.
- Acceptance of the atomic theory and kinetic theory invited the “reduction” of classical thermodynamical concepts to statistical mechanical ones. But what is the criterion of success?

Crude Functionalism

- One thought—loosely associated with functionalism and particularly David Lewis—is that one identifies the “role” that the concept plays in the original theory and then seeks to define something that “plays the same role” or at least nearly the same role in the underlying theory.
- In the case of entropy, there are fundamental reasons why *nothing* definable in statistical mechanical terms can *exactly* “play the role” of entropy in classical thermodynamics. So that presents one problem.
- But there is another problem: what if *many* things definable in the underlying theory can equally well “play the same role”?

A Simple Example

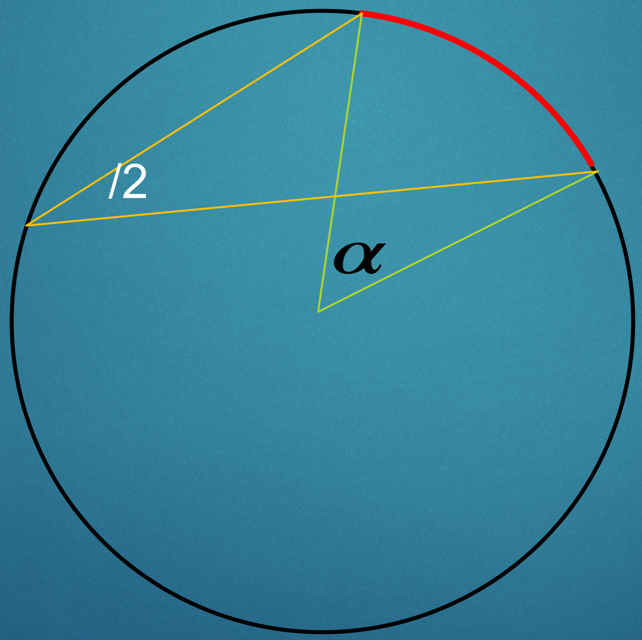
- What is the “correct” definition of a *circle* in Euclidean geometry?
- In one sense of “play a role” any two definitions that are materially equivalent—that pick out all and and only the circles—are functionally equivalent to each other and “play the same role”: neither could be more correctly called the “right” or “true” or “real” or “correct” definition if they are necessarily coextensive.
- But this seems like the wrong result: some definitions “capture the essence” of a circle better than other, materially equivalent ones.
- Definition 1: A circle is the locus of all points equidistant from a given point (the ‘center’).
- Definition 2: A circle is a closed curve with the largest area for a given perimeter.

A Relevant Comment by Aristotle

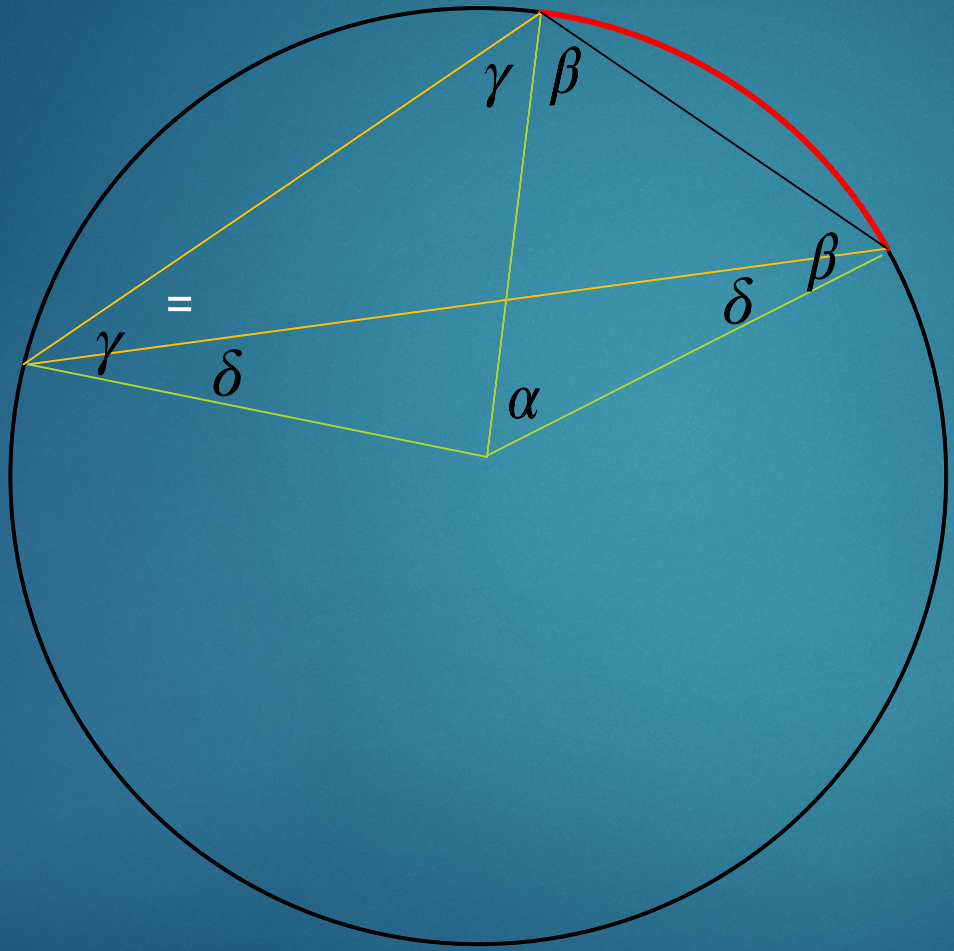
- “It seems not only useful for the discovery of the causes of the derived properties of substances to be acquainted with the essential nature of those substances (as in mathematics it is useful for the understanding of the property of the equality of the interior angles of a triangle to two right angles to know the essential nature of the straight and the curved or of the line and the plane) but also conversely, for the knowledge of the essential nature of a substance is largely promoted by an acquaintance with its properties: for, when we are able to give an account conformable to experience of all or most of the properties of a substance, we shall be in the most favorable position to say something worth saying about the essential nature of that subject; in all demonstration a definition of the essence is required as a starting-point, so that definitions which do not enable us to discover the derived properties, or which fail to facilitate even a conjecture about them, must obviously, one and all, be dialectical and futile.” de Anima 402b 15

A Simple Example

- Here is a fact about circles: the angle subtended by an arc at the center is twice the angle subtended by the arc at any other point on the circle, e.g.:



Proof



180° =
180° =
2 =
180°
So =
And hence
/2 =

Which of These is Not Like the Other?: Entropy Edition

- A) $S = -H =$ Boltzmann
- B) $S = k \ln(W)$ Planck
- C) $S =$ Gibbs
- D) $S =$ von Neumann
- F) $S =$ Shannon

The Reversibility Objection

- 1) Reversibility of the Microdynamics, i.e.
- There is an operator * such that if $S_i \dashrightarrow S_f$ then $S_f^* \dashrightarrow S_i^*$.
- 2) The entropy of S = the entropy of S*
- 3) Entropy can increase but cannot decrease.
- 1), 2) and 3) are jointly incompatible: they cannot all be true.
- The solution is to abandon 3): it must be *physically possible* for entropy to decrease.

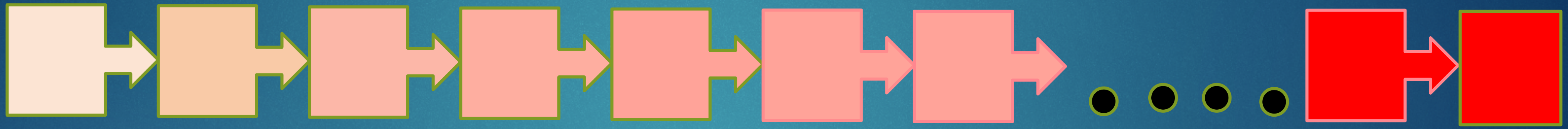
An Observation

- From a logical point of view, admitting merely the physical *possibility* entropy decrease solves the problem.
- Surprisingly, that retreat from the absolute exceptionlessness of the Second Law can be as tiny, as mild, as you like.
- In particular, 1) and 2) can be made compatible with the claim that entropy decrease is as unlikely, as improbable, as you like (short of being absolutely impossible).
- Furthermore, *there is no other reversibility problem that needs to be solved, and in particular none that requires the Past Hypothesis as a fundamental postulate.*

A Parable

- What Zeus wants:
- 1) There should be 100,000,000,000 rooms, each at a different level of unpleasantness, starting at neutral (O).
- 2) Zeus should be able to place a soul initially at any level of unpleasantness up to 99,999,000,000.
- 3) Each midnight, the soul will be required to exit the room it is in through a door.
- 4) For 1,000,000 days in succession, the soul will visit the next worse room.

The First Design



Where the arrows represent subway turnstiles.

Zeus's New Demand

- Zeus has already purchased a large supply of swinging doors. By their design, they allow motion in both directions.
- Furthermore, Zeus demands that the choice of which door to go through must be completely free and unconstrained. The soul can choose the door on any basis and by by whatever method it likes.

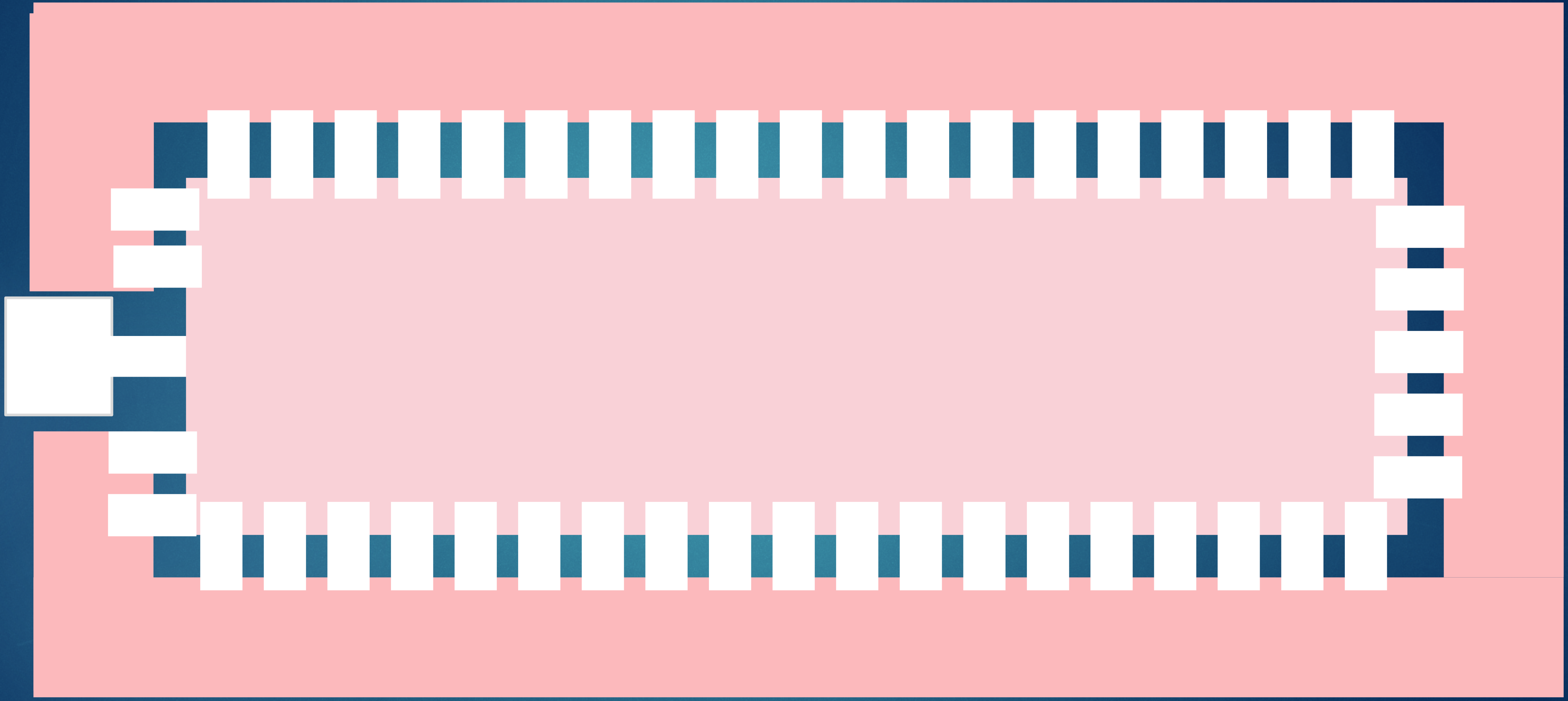
The Reversibility Objection

- You now complain that Zeus has set you a logically impossible task.
- If a soul on one night goes through a swinging door from room N to room N +1, then it must be possible on the next night for it to go back from room N + 1 to room N.
- Zeus concedes that given the swinging doors the demands are logically impossible to satisfy.

The Offer

- However, you ask Zeus to allow you a little bit of wiggle room, indeed as little as Zeus likes.
- You cannot absolutely guarantee that each soul will always move to the next worse room, but you can make it *overwhelmingly likely* that that will happen. Zeus can be *morally certain* that every soul will do exactly what he wants it to do.
- Zeus agrees that, given the logical impossibility of absolute certainty, this is good enough.

The Second Design



In Detail

- Room 0 has 1 door, that goes into Room 1.
- Room 1 has 1,000,001 doors, 1 to Room 0 and 1,000,000 to Room 2.
- Room 2 has 1,000,001,000,000 doors, 1,000,000 to Room 1 and 1,000,000,000,000 to Room 3….
- Etc., Etc.
- The ordering of the "badness of the rooms" can be read off the number of doors in the room. For convenience, instead of recording the number, record the log of the number in base 1,000,001.

One More Detail

- The choice that the soul makes must be free. But of course, if a soul remembers which door it came in, then it can just hang around and go back out that door. So you ask Zeus to require that every day the soul drink a draught from the river Lethe, which causes it to forget how it came into the room. But the choice of door (by color, location, shape, size, etc.) is made completely free.

- In the design, the locations, shapes, sizes, etc. are randomized, so there is no correlation between these features and whether the door leads to Room N + 1 or Room N – 1.

A Typicality Argument

- In every room, *overwhelmingly most* of the doors lead to the next higher room. So no matter what the soul's method of choice is, we can be *morally certain* that a door to the next higher room will be chosen. If we want to be even more certain, just increase the number of doors. Or Zeus can relax the criterion of success from *always* going to the next worst room to *almost always* going to the next worse room.

- Zeus is satisfied. You get paid.

Prediction

- You are a soul in the afterlife. You have been provided with a complete blueprint of hell. You have drunk from Lethe, and don’t recall which door you came in through or indeed how long you have been in hell. Looking around, you see that you are in Room 100. Midnight is about to arrive, and you must choose a door to go through. Knowing your situation, you anticipate that you will go through a door into Room 101. And always (or almost always) you are right that you go through a door into the next worse room.
- Badness monotonically (or almost monotonically) increases.

Retrodiction

- You are a soul in the afterlife. You have been provided with a complete blueprint of hell. You have drunk from Lethe, and don’t recall which door you came in through or indeed how long you have been in hell. . Looking around, you see that you are in Room 100. You ask yourself where you likely were yesterday.
- Looking at the blueprint, you determine that either you were in Room 99 or in Room 101. You have no other evidence of which it was.
- What is reasonable to conclude?

The Reasoning

- You must choose between the hypothesis H99 that yesterday you were in Room 99 and the hypothesis H101 that yesterday you were in Room 101.
- The conditional probability that you would today be in Room 100 conditional on H99 is .999999.
- The conditional probability that you would today be in Room 100 conditional on H101 is .000001.
- Your priors, whatever they are, are not wildly imbalanced in favor of H101.
- So as a good Bayesian, you conclude that most likely you were in Room 99. And before that in 98. And before that in 97. And (at least almost always) you are right.

Note:

- So just by consideration of the relative numbers of doors in each room that lead up and lead down, you both predict that the “entropy” will continually monotonically increase and that in the past if *has been* continually and monotonically increasing, i.e. it was *lower* in the past.
- There is no disastrous retrodiction that has to be defeated or overridden by adding the Past Hypothesis.
- Indeed, the claim that the entropy was lower in the past should rightly be called the *Past Conclusion*.

Note:

- In the logic of the situation, it is immaterial whether there is a “top” room or not. The rooms could continue to grow *ad infinitum*, with no “largest” room or room with “the most doors”. That is neither here nor there.
- Of course if there *is* a largest room most of whose doors *lead back into the same room*, then once you get there you will (probably) stay there (for a very, very, very long time).
- In other words, the existence of an “equilibrium room” is neither here nor there.

Note:

- Although the *volume* of the successive rooms gets larger and larger, the volume is *playing no explanatory role* in the behavior. What is important is not the structure of the volumes of the rooms, but the structure of their *boundaries*. What is playing the explanatory role is the fact that *overwhelmingly most of the boundary of every room leads to a room with a larger boundary*. Because what we want to explain is the nature of the *transitions* from one room to another, and that is determined by *which part of the boundary* the soul exits through, and *which other room* that boundary is shared with.

Note:

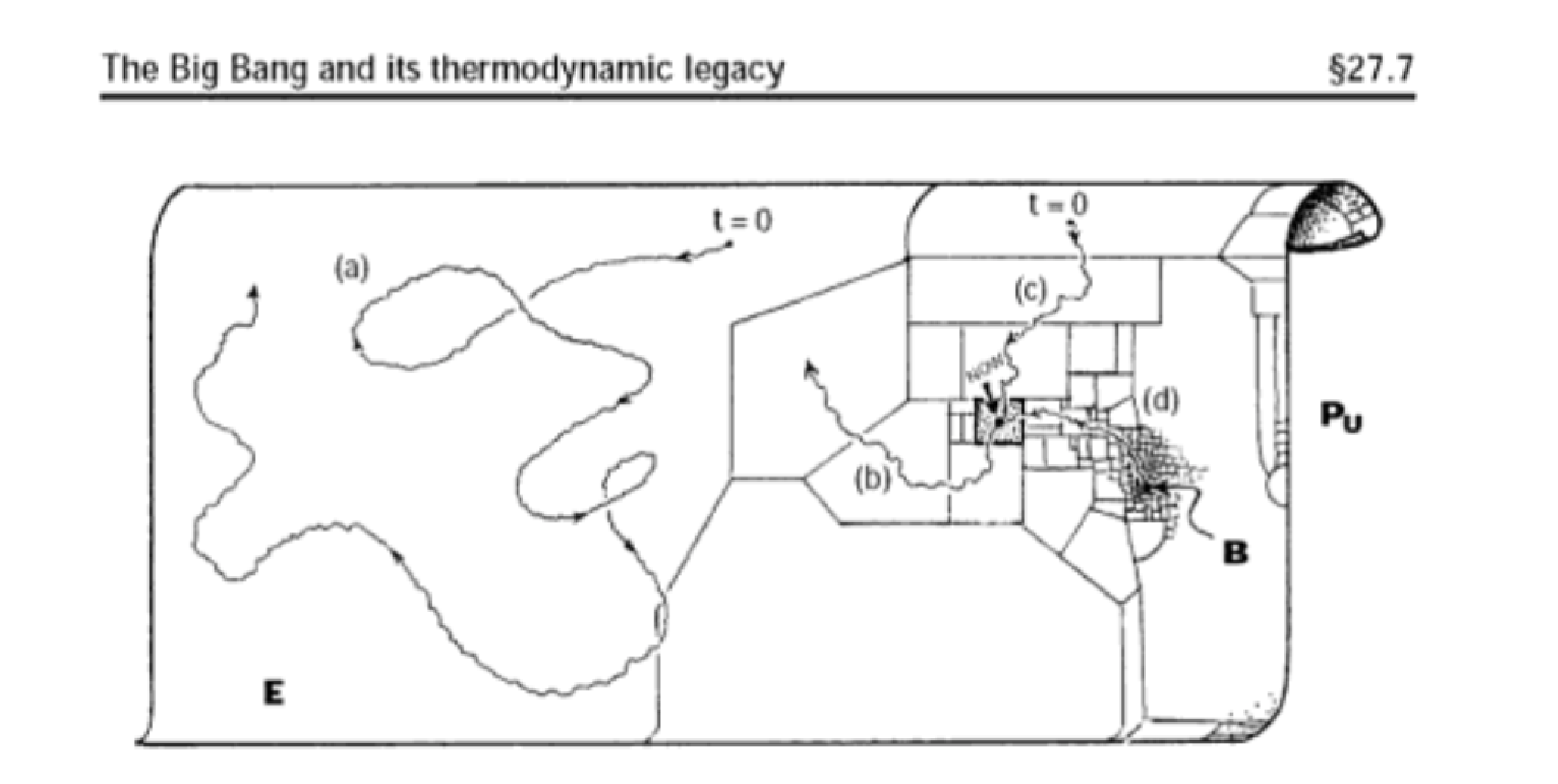
- If the rooms are more-or-less the same shape, and more-or-less convex, and if we choose a characteristic sort of average dimension D, and if the rooms are in an N-dimensional space, then the measure of the *volumes* of the rooms will be proportional to D^N and the measures of the *boundaries* of the rooms will be proportional to D^{N-1}. So the log of the volume will go as N ln(D) and the log of the boundary as (N-1) ln(D). So the *ratios of these logs will be the same whether one measures the volume or the boundary*.
- In particular, if the quantity k ln(volume) behaves the way we want the entropy to, then the quantity k' ln(boundary) will do just as well.

Note:

Neither the measure of the volume nor the measure of the boundary by itself is doing the explanatory work we need. If we take a space and *randomly* partition it into regions such that one region is overwhelmingly most of the volume, then typically the boundary of every other element of the partition will lead *directly* into that largest region. Interpreting the log of the volume (or boundary) as the "entropy", any system out of equilibrium will instantly "thermalize" to equilibrium.

So it isn't merely the relative *sizes* of the volumes or of the boundaries that is doing the explanatory work: it is also their precise geometrical *arrangement*.

A Typical Diagram



Definition, Explanation, and Reduction (or Emergence)

- Entropy, along with Volume, Pressure, Temperature, Moles of gas, is a concept in classical Thermodynamics.
- In order to “reduce” Thermodynamics to an underlying microphysical theory, these same terms must be somehow “defined” in the proprietary language of the microtheory.
- What are the criteria of adequacy or success of a proposal for such a definition?

Criteria of Success

- Insofar as the reduced theory has empirically adequate laws, we want the definition to provide a conceptual connection that accounts for that success.
- *In the first place*, we want the definiens to behave—at least to very good approximation—like the definienda in terms of the laws or generalizations they obey.
- *But more than that*, the definitions should reveal the features of the microtheory that *explains* or *accounts for* the behavior of the reduced theory.
- If k ln (V) k' ln(B(V)), then if one behaves the right way, so does the other. But still, it is the structure of the B(V)s, not of the Vs, that accounts for the behavior.

Example

- In Classical Thermodynamics, we have
- PV = nRT (Ideal gas Law)
- V is defined in the microtheory as V.
- P is defined in the microtheory as the average momentum flux per unit area. *At equilibrium*, that is the same everywhere and for all orientations of the unit area.
- T is defined as the average Kinetic Energy per particle or per degree of freedom.
- n is defined as the numbers of moles of the gas.
- k_B is defined as conversion constant.

Classical Thermodynamics

- In classical thermodynamics, P and T are only defined at equilibrium.
- The microphysical definitions allow for a more extensive definition, relative to a coarse-graining.
- Similarly, the quantity S in classical thermodynamics is only defined for a system at equilibrium.
- For them, differences in S are well-defined.
- We want the microphysical definition of S to recover the classical values for equilibrium, but also to be more broadly defined to cover non-equilibrium as well.

The Second Law

- Similarly, we want the microphysical definition to recover (at least to good approximation) the classical values for entropy of a system at equilibrium, but also to extend in a way that extends the Second Law.
- The classical Second Law (stated in terms of entropy) only pertains to comparison between equilibrium states, but once we fill in a definition for non-equilibrium we want to demand not merely that the entropy never go down in a closed system, but that if it rises it rises continuously, not by sudden jumps.
- The classical definition also pertains objectively to single systems, not only to large ensembles of systems. That is, we want an individualist definition of the entropy
- Either *S* = k ln(V) or *S* = k’ LN(B(V)) will work *under some more conditions.*

The Extra Conditions

- In order for the entropy to rise (or indeed fall) continuously, we require that the partitioning of the phase space into cells be correctly structured in term of which elements of the partition are adjacent to which.
- Adjacency is, by definition, sharing a boundary.
- So it is the boundary structure that accounts for the tendency of the so-defined entropy to behave as we want it to.
- So it is S = k In (B(V)) rather than S = k In(V) that provides the correct microscopic definition of the entropy of a system relative to a partition of the phase space.
- And as a bonus, the explanation of the Second Law is basically independent of the dynamics, unconcerned whether the dynamics is time-reversible, and never generates bad retrodictions that need to be mended by appeal to the Past Hypothesis.