What Is Black Hole Entropy?

Erik Curiel

Munich Center For Mathematical Philosophy
Ludwig-Maximilians-Universität

Black Hole Initiative
Harvard University

Smithsonian Astrophysical Observatory
Radio and Geoastronomy Division

email: erik@strangebeautiful.com

$$\frac{A}{4}$$

possibly plus corrections of order $\hbar$

the philosopher versus the physicist

Recently I was at a talk in which a philosopher argued that we do not understand the nature of black-hole entropy, that it poses problems unrelated to the standard puzzles about the nature of entropy in thermodynamics and statistical mechanics, and that we need to try to address these problems if we are to confidently conclude that black holes are real thermodynamical objects. Five minutes into the talk she was interrupted by an eminent theoretical physicist, who claimed (at some length) that her project is otiose. Physics, he said—by which he meant string theory—has in the past twenty years already provided us with a sound understanding of the quantum nature of the statistical underpinning of the Bekenstein entropy, *i.e.*, the area of a black hole as a measure of its physical entropy. Any remaining puzzles about black-hole entropy are not peculiar to black holes, but are the same in kind and character as the standard puzzles about entropy as a physical quantity associated with any type of physical system.

I agree with the philosopher,
not the physicist

Outline

classical black holes (crudely speaking)

- regions of "no escape": would have to travel faster than light to leave (boundary is event horizon, need notion of "infinity")
- no ordinary matter present: **not an ordinary physical system**
- **"global"** objects: defined by entire future behavior
- "No Hair": **characterized entirely** by mass, angular momentum, electric charge

(*pace* Daniel, not all black hole spacetimes arise from Hamiltonian formulation—Kerr, evaporating black hole, . . .)

Event Horizon Telescope: M87

why black holes matter

1. they're incredibly cool
2. they're incredibly weird: can teach us much about conceptual structure of general relativity, what is (in some sense) physically possible in the world
3. black hole thermodynamics: one of most important, central, and fruitful fields of study in theoretical physics, closely connecting disciplines once seen as largely independent: cosmology, general relativity, quantum field theory, thermodynamics, statistical mechanics, particle physics, fluid dynamics, condensed matter, quantum optics, . . .

why black holes matter (cont.)

4. widely considered best source of clues to a more fundamental theory of quantum gravity
   a. simplest "purely gravitational" system to try to quantize
   b. their laws mix dependence on and independence from EFE in funny way, so get both pure kinematics and dynamics
   c. indirect access to micro-degrees of freedom and their dynamics by treating thermodynamics as arising from statistical mechanics

Field	Core Concepts
astrophysics	• compact object • region of no escape • engine for enormous power output
classical relativity	• causal boundary of the past of future null infinity (event horizon) • apparent horizon (all outgoing light rays "get turned around") • quasi-local horizon
mathematical relativity	• apparent horizon • singularity

Table: the core concepts common to different fields for characterizing black holes

(Curiel 2019, "The Many Definitions of a Black Hole", *Nature Astronomy*, 3:27–34)

Field	Core Concepts
semi-classical gravity	• same as classical relativity • thermodynamical system of maximal entropy
quantum gravity	• particular excitation of quantum field • ensemble or mixed state of maximal entropy • no good definition to be had
analogue gravity	• region of no escape for finite time (“long” compared to characteristic time) • same for low energy modes (“low” compared to characteristic energies)

Table: the core concepts common to different fields for characterizing black holes, cont.

(Curiel 2019, “The Many Definitions of a Black Hole”, *Nature Astronomy*, 3:27–34) serious methodological and epistemological (and ontological?) problems:

1. event horizon is "global": makes implicit reference to "all future time"
2. that, and fact that nothing locally distinguishable about event horizon ⇒ *no local measurements can ever determine its location, much less whether there is one*
3. the definitions used in classical general relativity do not match those in astrophysics
4. in particular, our universe is not "asymptotically flat", so it cannot have anything like an event horizon
5. *all* properties of black holes, all theorems, relied on by astrophysics assume event horizon—still applicable in real world?
6. how is SgrA* similar to and how different from a Kerr black hole?

different properties one may demand of a “black hole”

- possesses a horizon that satisfies the four laws of black hole mechanics
- possesses a locally determinable horizon
- possesses a horizon that is, in a suitable sense, vacuum
- is vacuum with a suitable set of symmetries
- defines a region of no escape, in some suitable sense, for some minimum period of time
- defines a region of no escape for all time
- is embedded in an asymptotically flat spacetime
- is embedded in a topologically simple spacetime
- encompasses a singularity
- satisfies the No-Hair Theorem

different properties one may demand of a "black hole" (cont.)

- is the result of evolution from initial data satisfying an appropriate Hadamard condition (stability of evolution)
- allows one to predict that final, stable states upon settling down to equilibrium after a perturbation correspond, in some relevant sense, to the classical stationary black hole solutions (Schwarzschild, Kerr, Reissner-Nordström, Kerr-Newman)
- agrees with the classical stationary black hole solutions when evaluated in those spacetimes
- allows one to derive the existence of Hawking radiation from some set of independent principles of interest
- allows one to calculate in an appropriate limit, from some set of independent principles of interest, an entropy that accords with the Bekenstein entropy (*i.e.*, is proportional to the area of a relevant horizon, with corrections of the order of $\hbar$)
- possesses an entropy that is, in some relevant sense, maximal
- has a lower-bound on possible mass
- is relativistically compact.

different subsets of these properties are used in different contexts in different investigations, often in the same field

but they are jointly inconsistent

⇒ no definition can accommodate all actual uses of the concept in contemporary physics even more deep and interesting methodological and epistemological (and ontological?) problems, such as how empirical content of gets theories fixed, what an appropriate semantics of theoretical terms can be, whether there is a consistent ontology across theories, . . . , but, sadly, no time to go into it all now

for the purposes of the rest of this talk, we will assume the classical definition: a global event horizon in an asymptotically flat spacetime

first hint at black hole thermodynamics

No Hair $\Rightarrow$ black hole "macrostate" completely independent of how it formed, current constitution of interior (compatible with *any* "microstructure")!

just like ordinary thermodynamical systems

Zeroth Law

Thermodynamics

The temperature T is constant throughout a body in thermal equilibrium.

Black Holes

The surface gravity κ is constant over the event horizon of a stationary black hole.

First Law (Energy Conservation)

Thermodynamics

change in energy = (temperature × change in entropy) + work done

$$(\mathsf{d}E = T\mathsf{d}S + p\mathsf{d}V + \Omega\mathsf{d}J)$$

Black Holes

change in mass = (surface gravity × change in area) + "rotational work"

$$(\delta M = \frac{1}{8\pi}\kappa\delta A + \Omega_{\mathsf{H}}\delta J)$$

Second Law

Thermodynamics

Entropy never decreases ($\delta S \geq 0$) in any physical process.

Black Holes (The Area Theorem)

The area of the event horizon never decreases ($\delta A \geq 0$) in any physical process.

Third Law (Nernst Theorem)

Thermodynamics $T = 0$ is not achievable by any physical process

Black Holes $\kappa = 0$ is not achievable by any physical process

“Minus First Law” (Brown and Uffink)

Thermodynamics isolated thermodynamical systems tend to approach a unique equilibrium state

Black Holes isolated, non-stationary black holes tend to settle down to a unique stationary state (Kerr-Newman spacetime)

a formal or physical analogy?

- classical black holes are **perfect absorbers**, emitting nothing $\Rightarrow$ temperature **absolute zero**
- Geroch's infamous thought-experiment $\Rightarrow$ temperature **absolute zero**
- area **nothing like entropy**
- surface gravity **nothing like temperature**

$\Longrightarrow$ **analogy is purely formal**

(everything and its mother has EOM of SHO, but an alternating-current circuit is not physically a pendulum)

black hole thermodynamics, based on the Hawking effect, arises from the combination of our 3 best physical theories: general relativity, thermodynamics, and quantum field theory

general relativity

- best theory of space, time and gravity ("dynamical spacetime geometry"), classically conceived
- regime of applicability: intermediate energy and large spatiotemporal scales, super-dense matter to vacuum
- **no quantum effects**
- **extreme causal weirdness**, seemingly inconsistent with quantum mechanics

quantum field theory

- best theory of fundamental properties and behavior of matter (ignoring gravity)
- regime of applicability: high (or *extremely* low) energy and small spatiotemporal scales
- fixed, flat, static spacetime geometry
- **extreme quantum weirdness** (superposition, entanglement, uncertainty principle, non-locality), seemingly inconsistent with general relativity

thermodynamics

- study of relationship between work and heat for ordinary matter classically conceived, and how that constrains gross properties and behavior
- regime of applicability: all matter at low to intermediate energy and intermediate spatiotemporal scales
- **no gravity, no quantum effects**
- **temporal asymmetry**, seemingly inconsistent with general relativity and quantum field theory

in general, we do not know how they fit together
(or even whether they do)

with one major exception...

when the **effects of quantum fields** are taken into account, **black holes in general relativity**, even though they are nothing more than regions of empty spacetime, appear to become **true thermodynamical objects**, with an associated physical temperature and entropy

black holes and quantum fields

- 1974, Hawking demonstrates that black holes, in presence of quantum fields, are **not completely black** after all
- they emit radiation with a characteristically **thermal spectrum**—ordinary blackbody radiation—they **glow like lumps of hot iron**
- **temperature is surface gravity, entropy is area**
- $\Longrightarrow$ **physical equivalence**, not formal analogy: **black holes are thermodynamical systems!**
- $\Longrightarrow$ **laws of black holes are laws of thermodynamics** extended into new regime

$\Longrightarrow$ **black hole thermodynamics**, a deep and hitherto unsuspected connection among our three best, most fundamental theories

Black hole thermodynamics and results concerning quantum fields in the presence of strong gravitational fields more generally are without a doubt the most widely accepted, most deeply trusted set of conclusions in theoretical physics in which general relativity and quantum field theory work together in seemingly fruitful harmony.

This is especially remarkable when one reflects on the fact that we have absolutely no experimental or observational evidence for any of it, nor hope of gaining empirical access any time soon to the regimes where such effects may appreciably manifest themselves.

this all raises many deep philosophical problems and questions, most of which philosophers have not yet begun to address

traditional philosophical puzzles

general relativity

- the nature of spacetime (substantivalism versus relationalism)
- deterministic or indeterministic? (the Hole Argument, cosmic censorship)

quantum field theory

- the Measurement Problem
- non-locality (Bell's Theorem, "action at a distance")
- coherent ontology?

thermodynamics

- the status of the Second Law (empirical generalization? law of nature?)
- temporal asymmetry (arrow of time)
- reduction of thermodynamics to statistical mechanics

The Central Problem of Black Hole Thermodynamics

What does it mean to conceive of and treat black holes, in the presence of quantum fields, as thermodynamical systems?

It is no exaggeration to say that the Central Problem **affects essentially every traditional philosophical problem** of all three theories:

- restricting and refining how they can be **cogently formulated**
- changing the **criteria** for what may and may not count as **satisfying answers**
- suggesting **new avenues of attack**

AND it results in **several entirely new, deep problems**, independent of the traditional ones

importance of philosophical study

- without a doubt the **most widely accepted, most deeply trusted results** in theoretical physics in which those theories work together—but those theories **are *prima facie* inconsistent**
- AND **no empirical access** to those regimes
- $\Longrightarrow$ absolutely **no experimental or observational evidence** for any of it—why do we trust it?
- $\Longrightarrow$ investigations **necessarily speculative** in a way unusual even in theoretical physics
- $\Longrightarrow$ technically sophisticated physical questions **inextricable from subtle philosophical considerations** spanning ontology, epistemology, and methodology, again in a way unusual even in theoretical physics

how can an empty, locally undistinguished region of spacetime have thermodynamical properties?

- difficult to think of two more different quantities than entropy and spatial area...
- unless they be temperature and surface gravity
- how can these possibly be *identical*?
- $\Rightarrow$ *deep* problem for conceptual understanding of inter-theory relations: "same" quantity as represented in different theories

also:

- laws of ordinary thermodynamics are empirical generalizations
- laws of black hole mechanics are theorems of differential geometry
- $\Longrightarrow$ how can they possibly be "the same"?

why attribute entropy to black holes in the first place, besides formal analogies?

the most initially plausible way to explain what the entropy of a black hole measures, and why a black hole has such a property in the first place, is to point to the Hawking radiation it emits, and in particular the well defined temperature the radiation has

temperature and entropy go together like Wurst und Senf!

or Sturm und Drang?

BUT—no real connection with Hawking radiation/Hawking temperature:

1. some kinds of entropy (*e.g.*, Shannon/information, which many think important in this context) defined for systems without temperature
2. indeed, Hawking radiation is strictly kinematical: needs only a Lorentz metric and a horizon (analogue models), nothing to do with dynamics at all
3. but entropy is fundamentally dynamical—that we identify it with $A/4$ in general relativity depends on the form of the EFE

anyway, Hawking radiation is not blackbody radiation:

1. not generated by micro-dynamics of micro-degrees of freedom of event horizon (like electromagnetic blackbody radiation of hot iron is caused by jiggling of its atoms and electrons)
2. rather is feature of external, ambient quantum field
3. derived *even when no back-reaction*: *utterly no* connection with black hole dynamics/micro-states
4. how can it encode the temperature of the *black hole*?

but Hawking radiation is strongest argument for assigning a physical temperature to black holes!

Bekenstein's original motivation (pre-Hawking):

TO SAVE THE SECOND LAW!

- seems easy to violate standard Second Law when black holes are around:
  1. throw favorite highly entropic system into black hole
  2. the entropy of the world outside the event horizon—a causally isolated system—spontaneously decreases
- $\Rightarrow$ Bekenstein proposed Generalized Second Law: *total* entropy, black hole (area) + ordinary matter outside, never decreases
- many powerful (purely theoretical!) arguments supporting it (as Aron just showed)

for Bekenstein, black hole entropy was measure of information about black-hole interior inaccessible to an exterior observer

$\Longrightarrow$ extraordinary physical insight and understanding, great theoretical advance, based on terrible arguments

(not uncommon in the history of physics)

1. thermodynamical (phenomenological)?
2. Boltzmann?
3. Gibbs?
4. von Neumann (entanglement)?
5. Shannon?
6. something else entirely?

I claim: there had better be, at least, a good thermodynamical conception

- without an understanding of black-hole entropy as a truly thermodynamical entropy, in the sense of ordinary thermodynamics, we have no real evidence in the first place that black holes have a micro-structure appropriate for a statistical treatment of its dynamics that would yield a physically significant accounting of its entropy
- string theory, loop quantum gravity, . . . , can count all the "micro-states" they want, but until we have some indepedent reason to believe that what it's counting is relevant to a physically meaningful notion of entropy, it's nothing more than mathematics (and not even rigorous mathematics at that)

- classical GR alone cannot provide a statistical construction
- no way to describe a black hole as a system whose physical attributes arise as gross statistical measures over underlying, more finely grained quantities
- not even quantum field theory on curved spacetime (semi-classical gravity) can provide it
- still treat the black hole as an entity defined entirely by classical geometry of spacetime
- any statistical accounting, therefore, must come from a theory that attributes to the classical geometry itself a description based on an underlying, perhaps discrete collection of microstates, themselves describing the fine-grained dynamics of "entities"
- presumably quantum in nature, underlying the classical spacetime description of the black hole

a surprisingly common argument:

1. Planck length in combination with distinguished geometry provide natural coarse-graining: cover event horizon with Planck area tiles
2. the horizon carries some kind of information with density of approximately one bit (0/1) per unit area
3. total number of configurations of the order of $N \approx 2^A$ $\implies S := \log N \approx A \log 2$
4. *voilà!*

I think this is a crappy argument

- what is the yes-no question?
- either not counting micro-states relevant to dynamics in any straightforward way
- or else *strong* and *unwarranted* assumption that fundamental degrees of freedom are binary (or at least strictly and uniformly bounded by a very samll number)

I have similar problems with many "state-counting" arguments in weak-regime quantum gravity calculations

- Strominger and Vafa in string theory ("self-intersections of D-branes")
- Rovelli in loop quantum gravity ("ensemble/superposition of classical event horizon states")
- Sorkin from general perturbative quantum gravity plus discreteness assumptions
- ...

$\Longrightarrow$ begs the question by *assuming* that they are counting the dynamically relevant states, *and* that counting measure is the appropriate measure—but counting measure is almost never correct in statistical mechanics (always need some weighting) arguments for phenomenogical entropy

- based on Carnot-like cycle coupling ordinary matter to black holes as “heat sink”, and idea that gravitational radiation can carry “gravitational heat”, Curiel (2018) argues that there exists consistent thermodynamical theory of classical black holes (no one else believes it, including, from time to time, Curiel)
- based on Carnot-like cycle coupling Hawking radiation to black holes inside a box, Prunkl and Timpson (2019) argue that semi-classical black holes have consistent purely thermodynamical interpretation

plenty of grounds for questioning, criticizing both

interesting problem: Hawking radiation necessarily violates null energy condition (NEC)
BUT—*essentially all* GR black hole theorems used in BHT depend on NEC

1. No Hair theorem
2. the event horizon of a stationary black hole is a Killing horizon
3. Zeroth and Third Law of black hole mechanics
4. positivity of ADM and Bondi masses
5. if T_{ab} vanishes on a closed, achronal set, it vanishes in the domain of dependence of that set ("conservation of vacuum")
6. formation of trapped surface after gravitational collapse
7. black holes are (topologically) spherical
8. black holes don't bifurcate
9. apparent horizons hidden behind event horizons
10. domain of outer communication is topologically simple
11. Bousso's covariant entropy bound
12. asymptotically flat spacetimes without naked singularities are asymptotically predictable
13. many standard general forms of cosmic censorship
14. ...

where does black hole entropy "live"?

in particular, does black hole entropy have anything to do with the state or dynamics of anything in the interior?

standard answer: no

1. interior of black hole can be “arbitrarily large” (*e.g.*, glue an FLRW spacetime into the interior of Schwarzschild)—can have unbounded interior microstates
2. interior of black hole is wildly out of equilibrium, but black hole First Law is that appropriate for subsystem in equilibrium, and entropy counts microstates contributing to macroscopic equilibrium state—only event horizon is “in equilibrium”
3. exterior is causally isolated from interior, so why should interior be thermodynamically relevant to outside?
4. assume Hawking effect unitary (at least until close to complete evaporation); then Hawking radiation correlated with interior degrees of freedom; so number of interior states must remain large enough to store correlations; but entropy (proportional to area) is decreasing; so entropy better not count interior states (related to Page-Time Problem)

but this raises puzzles

1. black hole entropy is then extraordinarily smaller than an accounting of *all* microstates associated with the black hole—how is this statistical mechanics?
2. how can a derivation of the First Law ignore all that?
3. how does the event horizon "keep a record" of the entropy of all matter that passes through it?

- physical entropy has heretofore been attributed to *material* systems with non-trivial *dynamics*
- naively, spacetime geometry ("gravity") seems radically different from matter
- $\Longrightarrow$ does black hole thermodynamics, and in particular gravitational entropy, militate in favor of effacing this difference?

one fundamental and characteristic property of “matter”:

it possesses stress-energy
as represented by a T_{ab}

the thermodynamical fungibility of stress-energy

Ground of First Law of Thermodynamics: all forms of stress-energy are in principle ultimately fungible—any form of stress-energy can in principle be transformed into any other form

the family of all T_{ab} has a natural linear structure, and all stress-energy tensors must have the "physical dimension of stress-energy" $\implies$ the physical meaning of being able to add them together

gravity has no stress-energy tensor, so it can't be matter

but not so quick: sometimes possible to attribute *non-local* energy-like quantities, *i.e.*, not representable by a stress-energy tensor: gravitational radiation, ADM mass, various quasi-local masses, *etc.*—so this criterion is not so clear

anyway, other forms of energy in other theories are non-local (heat, work, Newtonian gravitational energy, . . .)

another try: EFE contains only contributions from "matter" stress-energy, so in general relativity another "obvious" answer is, matter is Ricci tensor and gravity is Weyl tensor

BUT:

- in classical general relativity, “matter can transform into curvature” (gravitational collapse into a singularity)
- in black hole thermodynamics, “curvature can transform into matter” (Hawking radiation)

$\Longrightarrow$ breakdown of distinction between “matter” and “geometry”? requires radical changes to picture of ontology of spacetime and matter?

intriguing speculation: matter and geometry not truly independent, but different “manifestations of underlying unified entity” (compare electric and magnetic fields in Maxwell theory, time and space in special relativity, ...)

(seems to be suggested by some programs of quantum gravity)