THE BLACK HOLE INFORMATION PARADOX

A RESOLUTION ON THE HORIZON

BY NETTA ENGELHARDT

The interior of a black hole is one of the most mysterious regions of the universe. By their very definition black holes cannot be directly observed, but a large body of indirect evidence strongly supports the existence of black holes in our own universe. Our job as physicists to describe the nature of our universe is incomplete without a full understanding of the interior of black holes.

General relativity, which governs the behavior of gravitating objects in the absence of quantum effects, predicts a singularity in the black hole interior. Such a singularity generally results in large curvatures on quantum scales; this is in contrast with our standard experiences in the lab, where quantum effects and gravity are happily more or less separate with little to no impact on one another. In the deep black hole interior, however, the strong interactions between gravity and quantum physics require a quantum theory of gravity.

Fortunately, though, we don't immediately lose control of black hole dynamics upon crossing the horizon: gravitational effects do not instantly become strong at the event horizon. Instead, the gravitational field strength increases gradually for an observer who falls into the black hole. The larger the black hole is, the longer it takes for gravitational effects to build up to a strength that requires a full quantum treatment. In particular, the gravitational effects at the putative event horizon of the M87 black hole are *weaker* than those at the surface of the sun! We certainly don't need quantum gravity for a good description of the surface of the sun. So it must be the case that semiclassical gravity—the approximately separate treatment of quantum effects and gravity—governs the physics of the black hole interior as well as it governs the physics of observers near the Sun (say, on Mercury). Which is to say, extraordinarily well! Put differently, if semiclassical gravity were to break down into strong quantum gravity effects at the event horizon of the M87 black hole, the same must also be true of the region between the Sun and Mercury. And we know that our current observational data of the Sun and Mercury can be well-described without any quantum gravity effects.

This means that a good chunk of the black hole interior fits within this approximate semiclassical gravity picture. This innocuous and straightforward conclusion, however, leads to one of the most longstanding problems in modern physics: the black hole information paradox.

THE PARADOX

The black hole information paradox is a conflict between two apparently incontrovertible facts: first, that semiclassical gravity is valid on scales where gravitational and quantum effects are more or less separate; second, that quantum mechanics is "*unitary*" and thus all quantum processes are in principle, though not necessarily in practice, reversible.

What precisely do we mean by "reversible"? How is a black hole different from a fire? Consider the following thought experiment: you write a message—classical information—on a notepad, which you then toss into a fire in some sealed chamber. Once the fire has consumed the notepad, the information appears to be destroyed: how can we possibly reverse the fire and read the message? Well, if we had arbitrarily powerful machinery that could track every molecule and collect all of the fine-grained information about the fire, and we knew the exact equations describing the behavior of every molecule as it interacts with other molecules, we could *in principle* recreate the message written on the notepad from the ash. This is the fundamental difference between a black hole and a fire.

A 1975 calculation by Stephen Hawking showed that if semiclassical gravity is approximately valid at the event horizon of a black hole, then black holes can evaporate. The black hole evaporation process appears to create

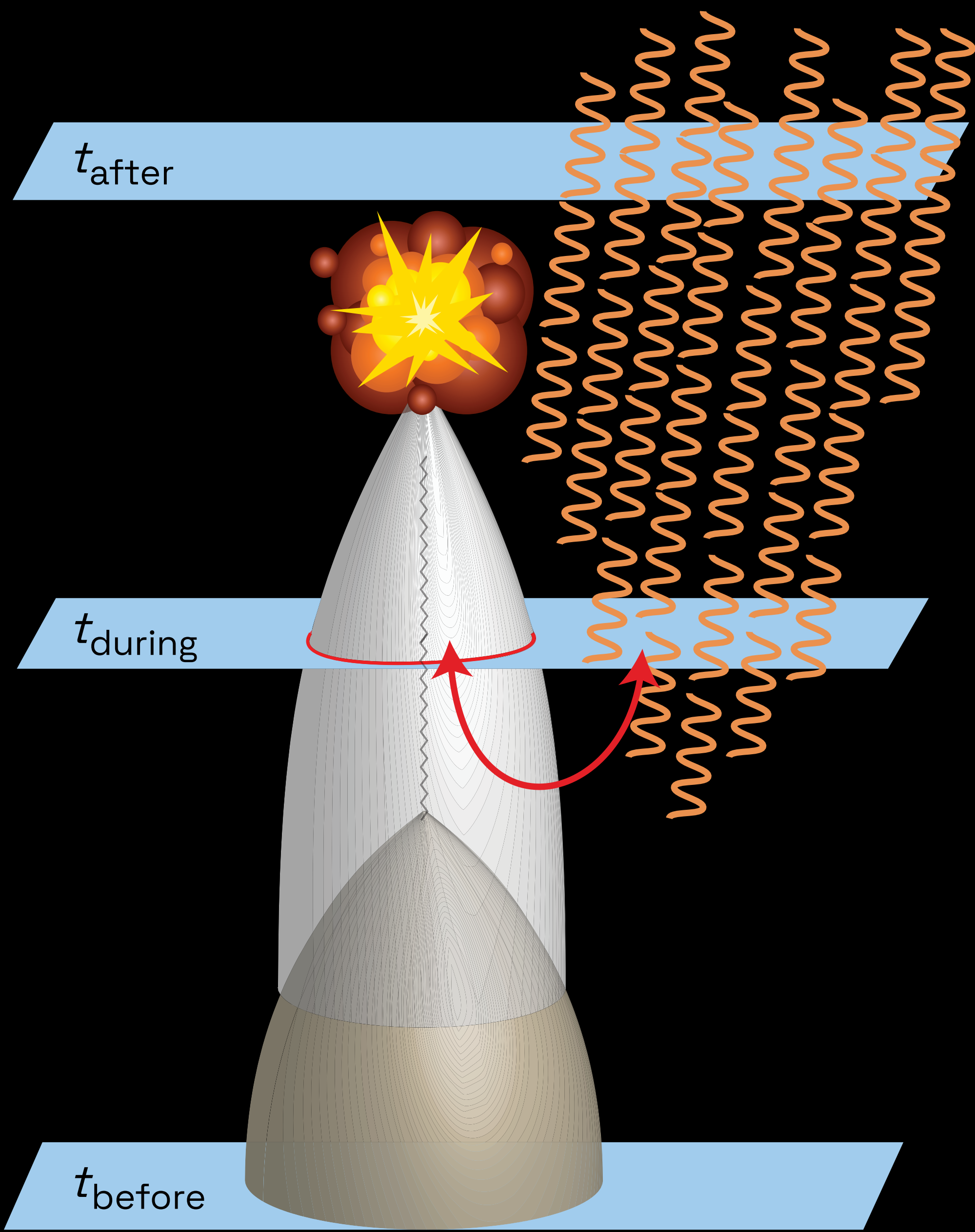


FIGURE 1:
The black hole formation and evaporation process, with time running vertically upwards. At early times, t_{before}, we have a star (brown section) shrinking in radius with time. Eventually an event horizon (gray section) forms. The black hole radiates, shown in the orange waves, during the evaporation process (as illustrated at the instant in time t_{during}). The radiation is entangled with the black hole interior (heuristically indicated by the red arrow). Eventually the black hole evaporates completely, and we are only left with the radiation at t_{after}. Credit: Netta Engelhardt

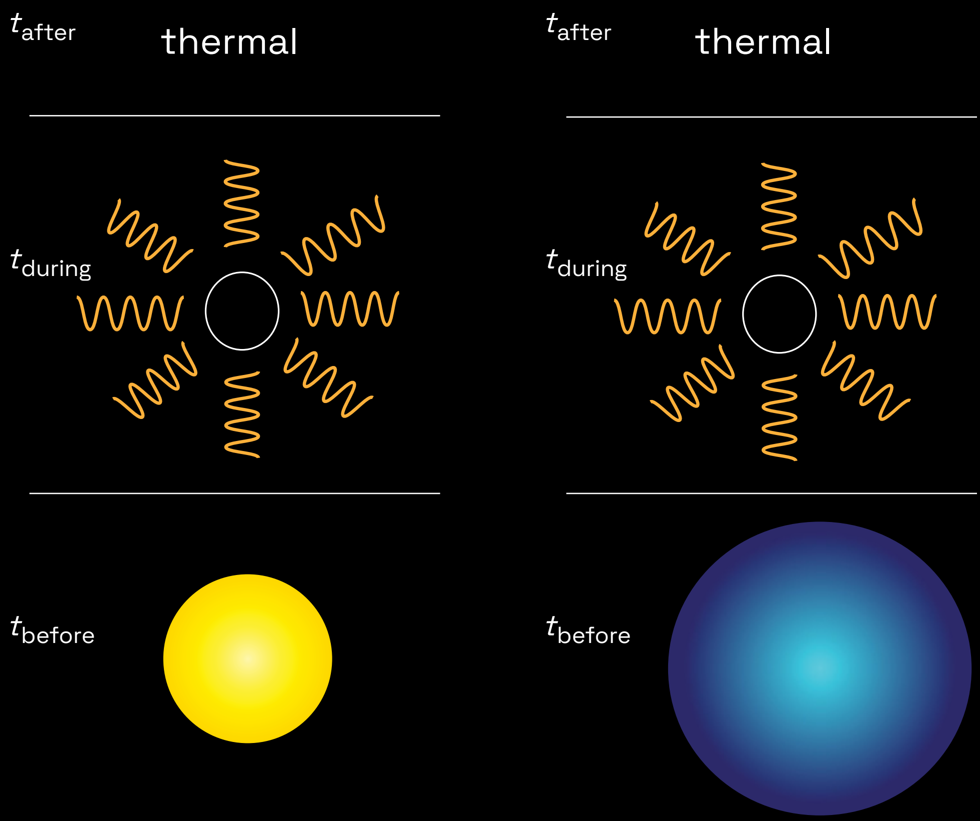


FIGURE 2:
A series of "snapshots" of the black hole evaporation process as analyzed by Hawking. Two very different stars collapse into black holes, which then radiate and evaporate. Once evaporation is complete, the universe is in a thermal state in both cases.
Credit: Netta Engelhardt

an unprecedented problem: it is *in principle* impossible to reverse-engineer the information that went into a black hole that has evaporated. Even if we knew the exact equations of motion of the universe and the exact state of the universe after evaporation, we *still* would not be able to ascertain the information that went into the black hole. The radiation emitted by the black hole as it evaporates must be thermal, and indistinguishable between any two evaporated black holes—even if they were originally formed from two very different stars!

An immediate consequence of this phenomenon is that if we knew the exact and precise state of the entire universe now (down to its fundamental particles), it would be in principle impossible to know what the universe was like a few years ago. Since our business as physicists is to use existing information to predict the evolution of the universe both forwards and backwards in time, this represents a catastrophic and unprecedented loss of determinism in physics. There is no other process which is known to result in *net* information loss.

Thus we appear to require one of two unappealing options: either strong quantum gravity effects are needed to describe the large-scale dynamics of regions of the universe that look just like Mercury and the Sun, or physics is not a deterministic science. This is the black hole information paradox.

This paradox has been a guiding post for progress on quantum gravity since its discovery by Hawking in 1975. Developments in string theory in the 1990s and 2000s provided the first conclusive evidence that information is *not* lost. How information can be conserved, however, remained a mystery. Is semiclassical gravity violated at the event horizon of a black hole? How can this be, given that interactions between quantum effects and gravity must be extremely weak there?

A NEW PERSPECTIVE

In 2019, the tide turned with a set of two simultaneously submitted papers by myself and my collaborators Almheiri, Marolf and Maxfield, and in parallel, Penington. We executed a semiclassical gravity analysis of black hole evaporation that was consistent, by a famous litmus test, with information conservation. This test, known as the Page curve, tracks the behavior of the von Neumann entropy of the radiation. This entropy, which is different from the standard entropy of thermodynamics, measures how “entangled” (or, correlated) a system is with its complement. Given some quantum system, say, n qubits, we can divide it up into two complementary subsystems: R and B. R will stand for the radiation of a black hole and B for the remaining black hole. When R is the trivial empty set, *i.e.*, R contains zero qubits, R is trivially uncorrelated with B: the von Neumann entropy of R vanishes. If we repartition the system so that R has progressively more qubits, we at first expect its von Neumann entropy to increase. Analogously, as the black hole evaporates into radiation, the data in B must end up in R. Eventually, we can repartition the system so that B has zero qubits and R has all of the qubits. That is, the black hole has fully evaporated.

At this point, R is again uncorrelated with B. We thus expect that the von Neumann entropy of R increases and then decreases as a function of the number of qubits in R.

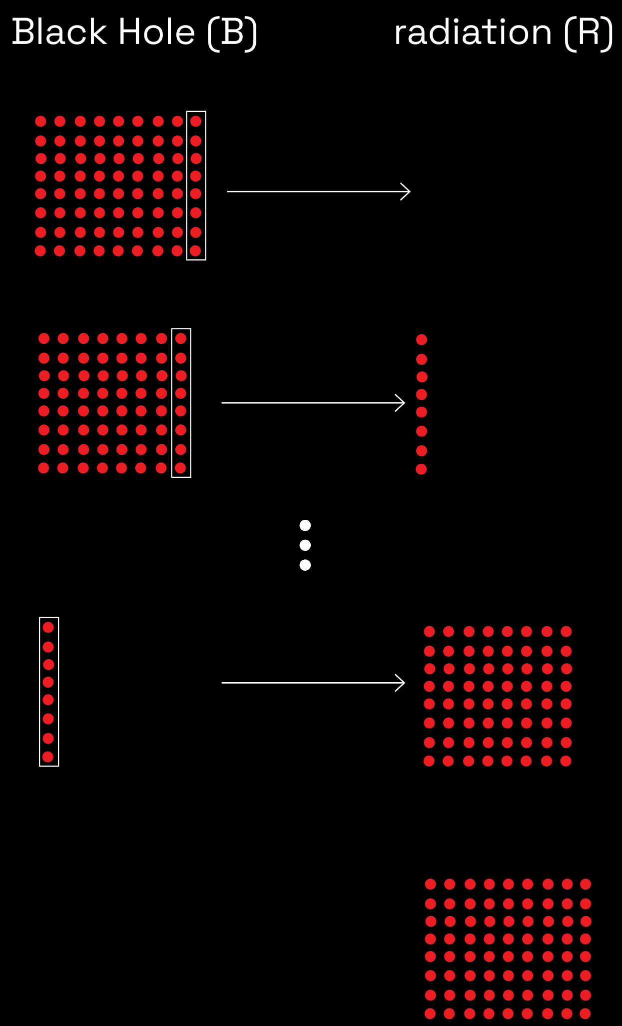


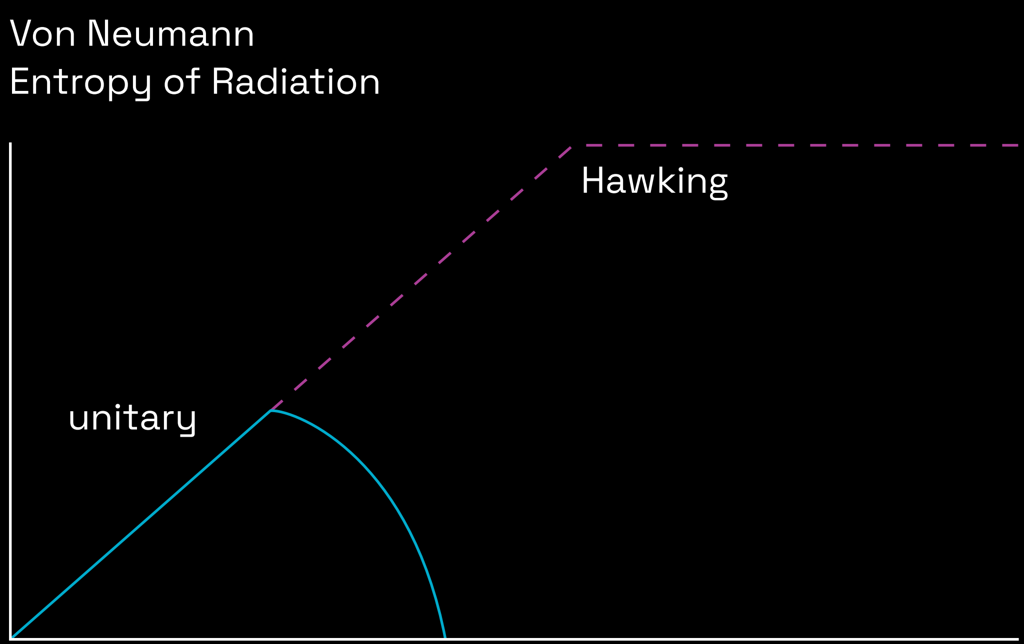
FIGURE 3:
A caricature of how a black hole evaporation process should look when information is conserved. Here the black hole B starts out (at the top) as a quantum system with some number of qubits. More and more of the system is transferred into R until eventually R is the entire system. Credit: Netta Engelhardt

If black hole evaporation is to be unitary, then the von Neumann entropy of the radiation should start out at zero, increase for a while, then—once the black hole has fully evaporated—return to zero. The resulting curve is known as the Page curve. However, Hawking's calculation shows that the von Neumann entropy of the radiation increases monotonically until the black hole has finished evaporating! The radiation, according to a semiclassical gravity treatment of the horizon, is now correlated with something that does not exist in the universe.

In 2019 we found that there exists a different semiclassical analysis from Hawking's that yields the Page curve. There was, however, a catch: while our calculation was within the regime of semiclassical gravity, and assumed that the standard picture of semiclassical gravity is an accurate description of the physics, the rules for how to compute certain quantities were vastly different from the standard rules of semiclassical gravity. By analogy, suppose you are asked to compute the pressure of an ideal gas in a cylinder. You may be tempted to compute the average velocity or momentum of the molecules of the gas and then use that to deduce the pressure. However, since the average velocity is zero, you would be led astray! Instead, we know that in the limit where thermodynamics is emergent from statistical mechanics, we must use $PV = nRT$, which is valid thermodynamically, but inherited from statistical mechanics. In complete analogy with the ideal gas law, we used the "quantum extremal surface formula," proposed by myself and A. Wall in 2014, rather than the Hawking formula (analogous to the erroneous zero average velocity calculation). The logic is identical: in both cases, you use an alternative formula which follows from the underlying microscopics of the statistical mechanics of your system.

This unusual approach gave us precisely the loophole we needed: the basic constructs of semiclassical gravity—space and time and its curvatures—can be consistent with information conservation, but only if we use the correct equations inherited from quantum gravity.

FIGURE 4:
The two curves corresponding to the unitary calculation of the von Neumann entropy of the radiation of an evaporating black hole (blue)—the Page curve—and the Hawking calculation of the von Neumann entropy of an evaporating black hole (purple). Credit: Netta Engelhardt



This insight resulted in an explosion of progress across the field of black hole information: finally, there might be a way of having our cake and eating it too! We can have standard spacetime and geometry at the event horizon of a black hole without paying the price of determinism of physics.

TOWARDS A RESOLUTION

A significant question remained, however: why are the equations for various quantities modified by quantum gravity when a black hole is involved, but not modified for the Sun or Mercury? Last summer, my collaborators at MIT (Chris Akers, Daniel Harlow and Shreya Vardhan) and I, together with Penington, proposed a resolution for the distinguishing feature between black holes and other objects. Our resolution was predicated on an older insight by Daniel Harlow and Patrick Hayden that even though the information about the black hole interior must escape in its radiation, actually processing the radiation to distill information about the black hole is incredibly complex. To be precise, this "decoding" process of the black hole radiation would require a quantum computer to implement a circuit whose size is exponential in the size of the black hole. For a black hole with the mass of the Sun, this would be exponential in 10^{77}! Black holes in general are extremely complex objects, which sets them apart from other astrophysical phenomena with similar curvature scales as those at the horizon of an astrophysical black hole. We proposed that semiclassical gravity is valid at low curvatures *and* low complexity; in our quantitative models, we saw that the modifications to the calculations required by the 2019 calculation of the Page curve can be attributed exactly to complexity in toy models of black holes.

We will likely be exploring the consequences of these developments on quantum gravity for years to come. Just as the black hole information problem has served as a point of inspiration for a vast landscape of developments in quantum gravity, I predict — with confidence since the fundamental theory of our universe is, in fact, predictive! — that its resolution will do the same.

NETTA ENGELHARDT is the Biedenharn Career Development Associate Professor of Physics in the Center for Theoretical Physics, and works on quantum gravity, primarily within the framework of the AdS/CFT correspondence. Her research focuses on understanding the dynamics of black holes in quantum gravity, leveraging insights from the interplay between gravity and quantum information via holography. Her current primary interests revolve around the black hole information paradox, the thermodynamic behavior of black holes, and the cosmic censorship hypothesis (which conjectures that singularities are always hidden behind event horizons).

Netta Engelhardt grew up in Jerusalem, Israel, and Boston, MA. She received her BSc in physics and mathematics from Brandeis University and her PhD in physics from the University of California, Santa Barbara. She was a postdoctoral fellow at Princeton University and a member of the Princeton Gravity Initiative prior to joining the physics faculty at MIT in July 2019.