Of Particular Significance
Conversations About Science with Theoretical Physicist Matt Strassler

https://profmattstrassler.com/articles-and-posts/relativity-space-astronomy-and-cosmology/black-holes/black-hole-information-paradox-an-introduction/

Black Hole Information Paradox: An Introduction

This article represents a lightning introduction to the black hole information paradox. Many details are omitted for brevity; longer articles will (eventually) explain them. Also, caution! the current understanding of the problem is so confused that the very last portion of this article should not be considered reliable or stable — it is likely to change in future.

[I thank Professor Joe Polchinski for consultations on the physics and for checking my illustrations for errors.]

The Two Conflicting Theories:

Quantum Theory (sometimes called "Quantum Mechanics") is the mathematics that is currently believed to underlie all physical processes in nature. It can't be used to predict precisely what will happen, but only the probability for any particular thing to happen. But probabilities only make sense if, when you add up all the probabilities for ***all*** of the different things that can possibly happen, you find the sum is equal to **one**. A quantum theory where this isn't true makes no sense. One consequence of this is that in a quantum theory, *information is never truly lost, nor is it truly copied*; at least in principle, you can always determine how a system started (its "initial state") from complete information about how it ends (its "final state"). See Figure 1, which shows two particles colliding, and several particles exiting from the collision, carrying off, in scrambled form, the information about the nature and properties of the two initial particles.

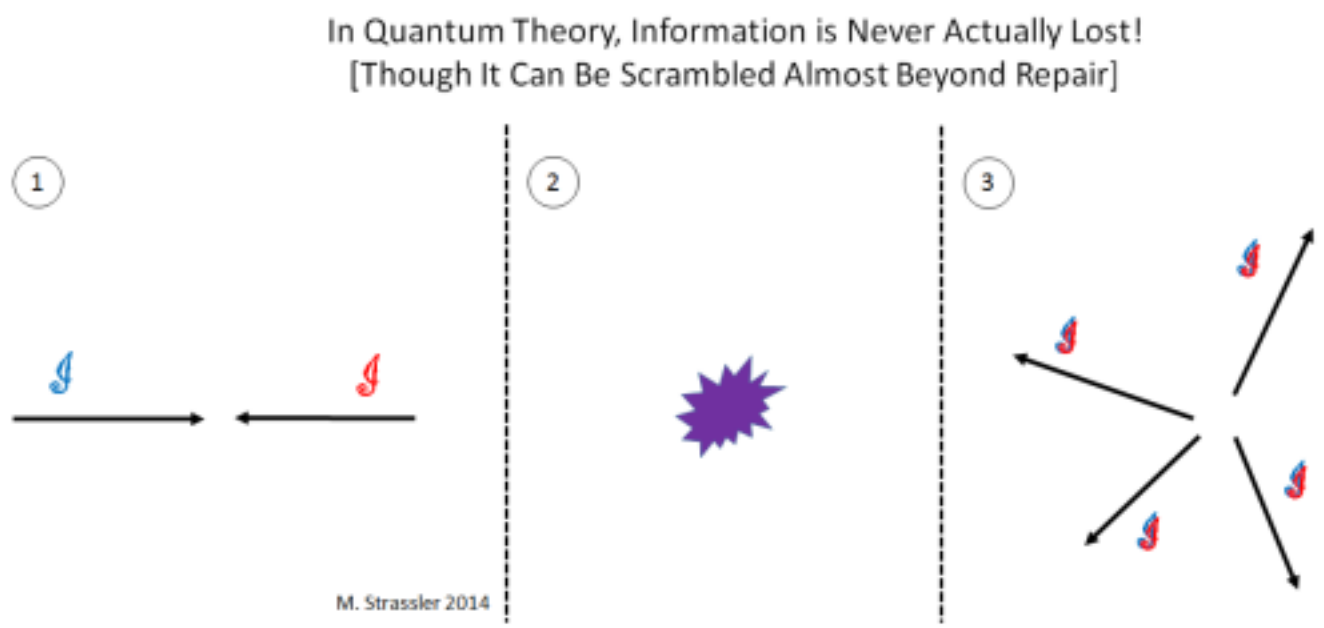

Fig. 1: In any type of quantum theory, information that goes in must come back out, scrambled but complete.

General Relativity is the name for Einstein's theory of gravity, in which gravity can be thought of as an effect of the warping of space and time. General relativity is not a quantum theory; it predicts exactly what happens, not probabilities for various things to happen.

It was gradually understood, over the years from 1915 to 1958, that extremely compact and massive objects form black holes. Gravity becomes immensely strong in their vicinity… strong enough to warp space-time dramatically, with the effect that any object that gets too close, and crosses the black hole's horizon — a surface of no-return — can never escape. See Figure 2, which shows a black hole horizon forming when two shells of matter become sufficiently compact. The information about these two shells goes inside the horizon and never can come out… in general relativity.

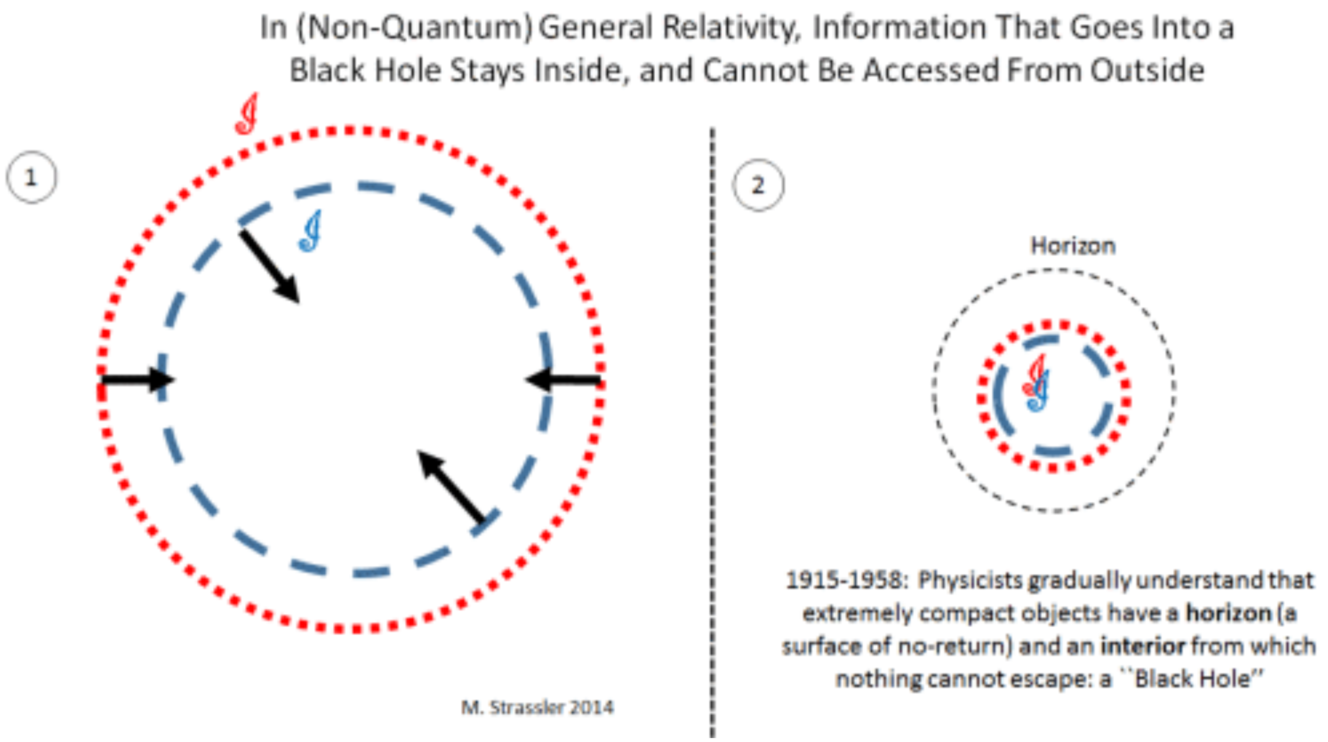

Fig. 2: In general relativity, what goes into a black hole never comes out. (1) Two shells of matter are collapsing under their own weight. (2) A horizon (a point of no-return) forms when they become sufficiently compact. The information about the two still-collapsing shells is locked behind the horizon forever.

Caution! It is impossible to draw black holes, and the information inside them, without being misleading. My illustrations can't illustrate how space and time are warped; for instance, to understand the whole story, you have to account for the fact that clocks inside the black hole run very differently from clocks just outside, which in turn run very differently from clocks far away. So don't take my illustrations, which illustrate conceptual but not technical points, too seriously!

A horizon is not an object, but a place beyond which escape is impossible. A famous analogy is to a boat approaching a waterfall, in an increasingly fast current. Once the boat passes a curve of no-return (see Figure 3), its engine will be unable to fight the current, and it will inevitably go over the waterfall. But the captain of the boat will not notice anything when crossing this curve; it is just an ordinary part of the river, whose importance will only become clear when the captain seeks to escape disaster. Similarly, in general relativity you will notice nothing when crossing the horizon; it's only when you try to escape the black hole that you will discover that — oops! — you went too close.

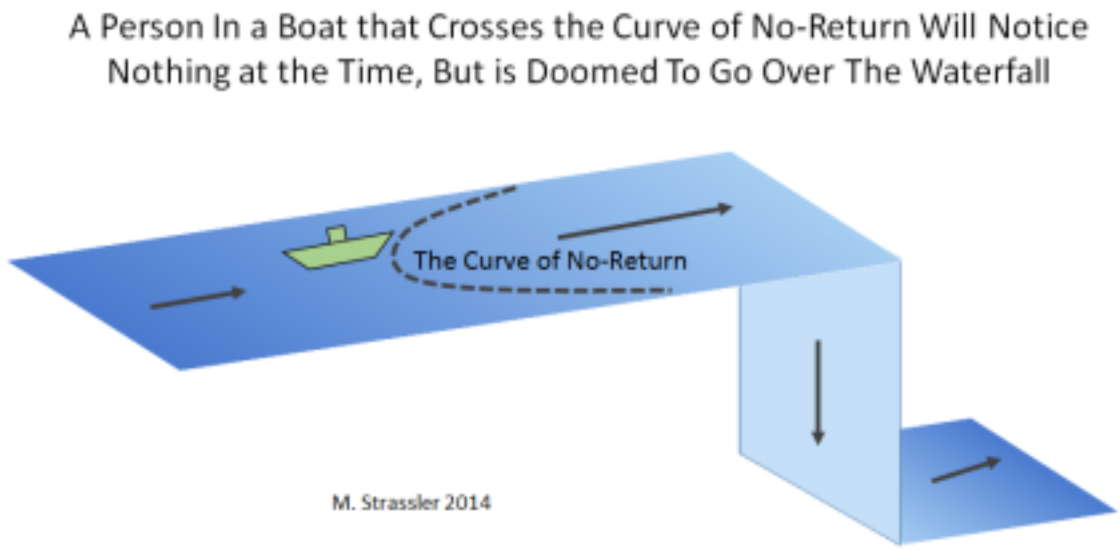

Fig. 3: You won't notice anything when you cross the point of no-return! By the time you sense danger it will be too late! It is the same with a black hole horizon; it's not a *thing*, it's a *place*.

The "Black Hole Information Paradox"

The paradox arose after Hawking showed, in 1974-1975, that black holes surrounded by quantum fields actually will radiate particles ("Hawking radiation") and shrink in size (Figure 4), eventually evaporating completely. Compare with Figure 2, where the information about the two shells gets stuck inside the black hole. In Figure 4, the black hole is gone. Where did the information go? If it disappeared along with the black hole, that violates quantum theory.

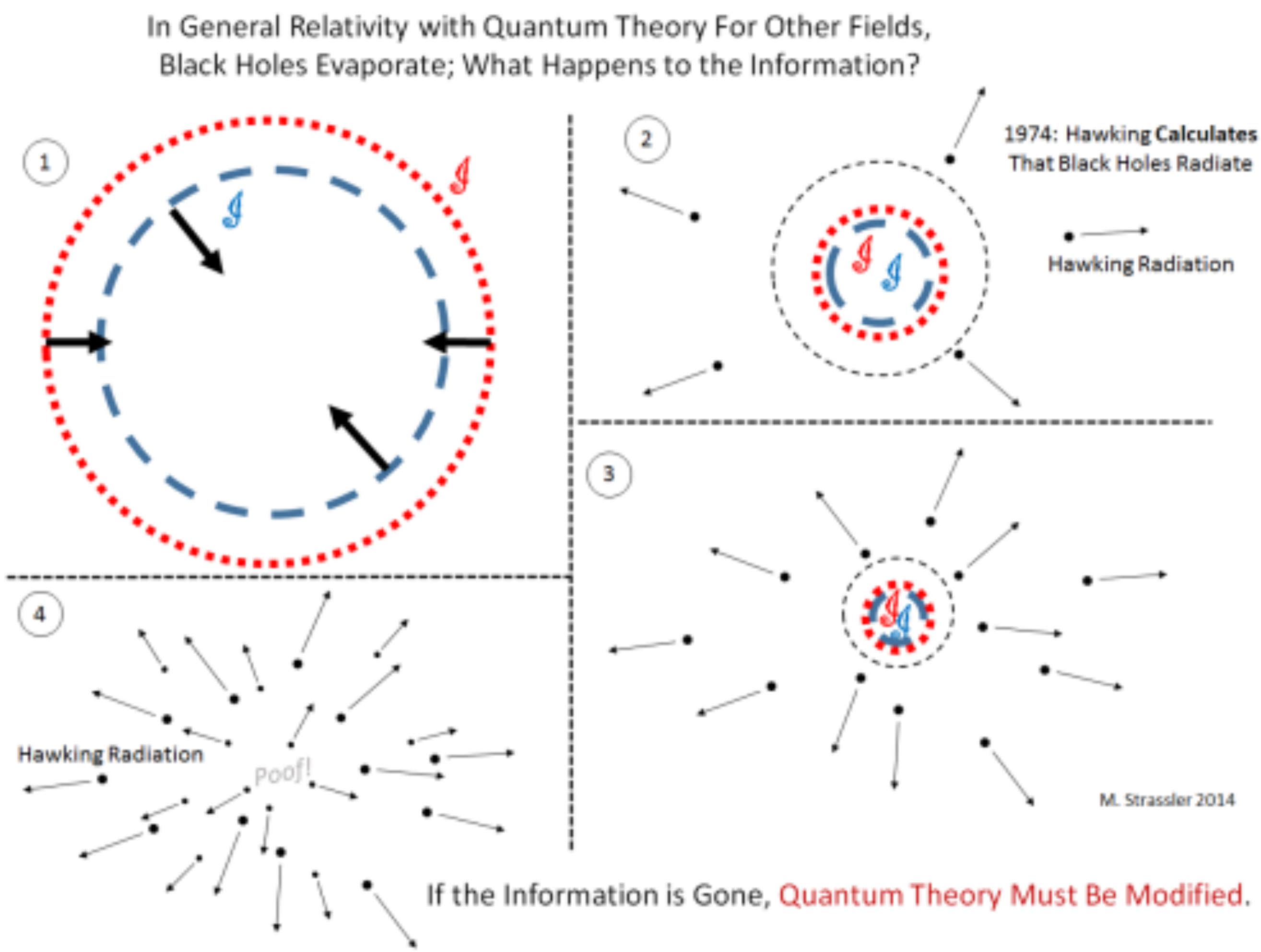

Fig. 4: (1) Shells of matter collapse; (2) A horizon forms, and Hawking radiation (in the form of particles of zero or low mass, such as photons or neutrinos or gravitons) emerges from the horizon; (3) the Hawking radiation carries off energy, causing the black hole's size and mass to shrink; (4) eventually the black hole evaporates completely, leaving only the Hawking radiation behind. Naively at least, it seems that the information about what went into the black hole is gone, violating the principles of quantum theory. Must quantum theory be changed?

Maybe the information came back out with the Hawking radiation? The problem is that the information in the black hole can't get out. So the only way it can be in the Hawking radiation (naively) is if what is inside is copied. Having two copies of the information, one inside, one outside, also violates quantum theory.

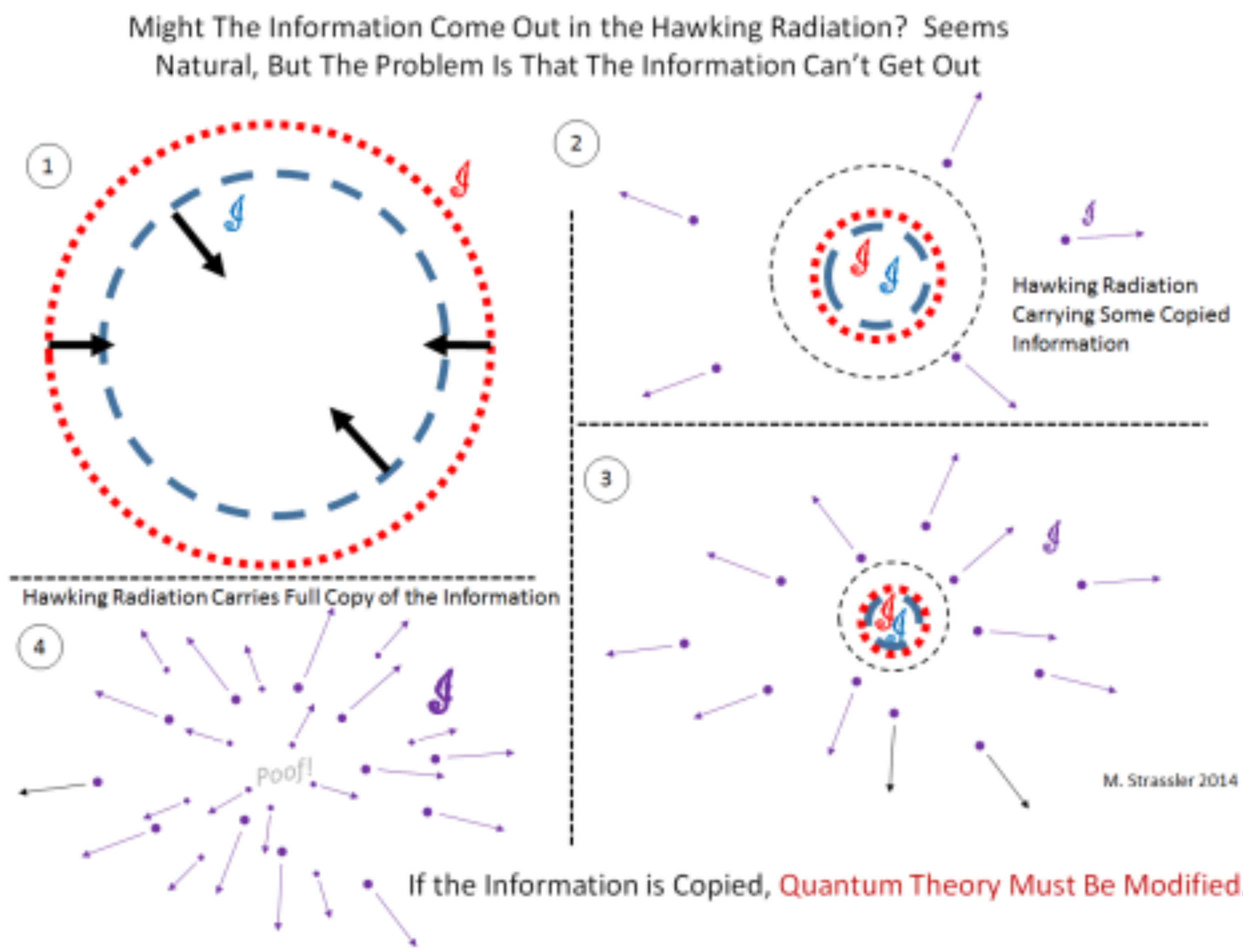

Fig. 5: Maybe the information is stored in the Hawking radiation? But since the information can't get out of the black hole, it must somehow be copied if it is to be put in the Hawking radiation. And that violates quantum theory also!

Of course, it may simply be that quantum theory is incomplete, and that the physics of black holes forces us to extend that theory, much as Einstein extended Newton's laws of motion in his theory of relativity. And this is what Hawking believed for three decades.

Complementarity: Saving Quantum Theory

However, others felt that it was general relativity, not quantum theory, that would need to be changed. And a proposal was made in 1992, called "complementarity", that suggested that the information was, in a sense, both inside and outside but without violating quantum theory. (This proposal was developed by Susskind and his younger co-workers.) Specifically, observers who remain outside the black hole see the information accumulate at the horizon, and then come flying outward in the Hawking radiation. Observers who fall into the black hole see the information located inside. (See Figure 6.) Since the two classes of observers cannot communicate, there is no paradox.

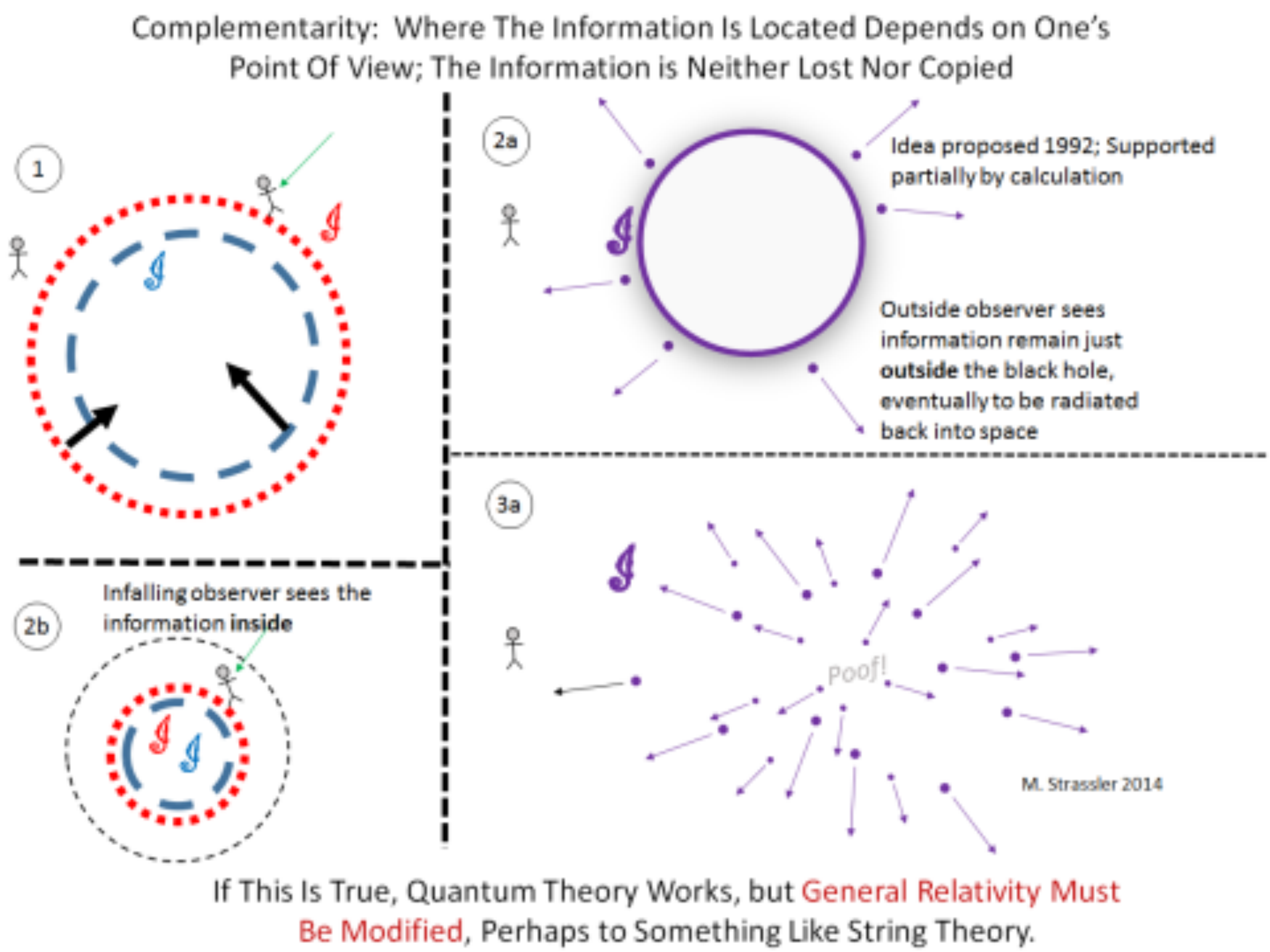

Fig. 6: Complementarity suggests it depends on one's perspective. Observers outside see the information (2a) being stored just outside, where it can later be (3a) transferred to the Hawking radiation. Only observers falling in (2b) see the information is being inside.

Still, this suggestion is potentially self-contradictory, and requires a number of strange things be true. Among them is something called "holography", an idea developed by 't Hooft and further by Susskind. The idea is that the physics of the three-dimensional interior of the black hole, where gravity obviously plays a role, can instead be viewed, via a rather mysterious transformation, as physics just above the two-dimensional horizon, where it is described by two-dimensional equations that do not include gravity at all! See Figure 7.

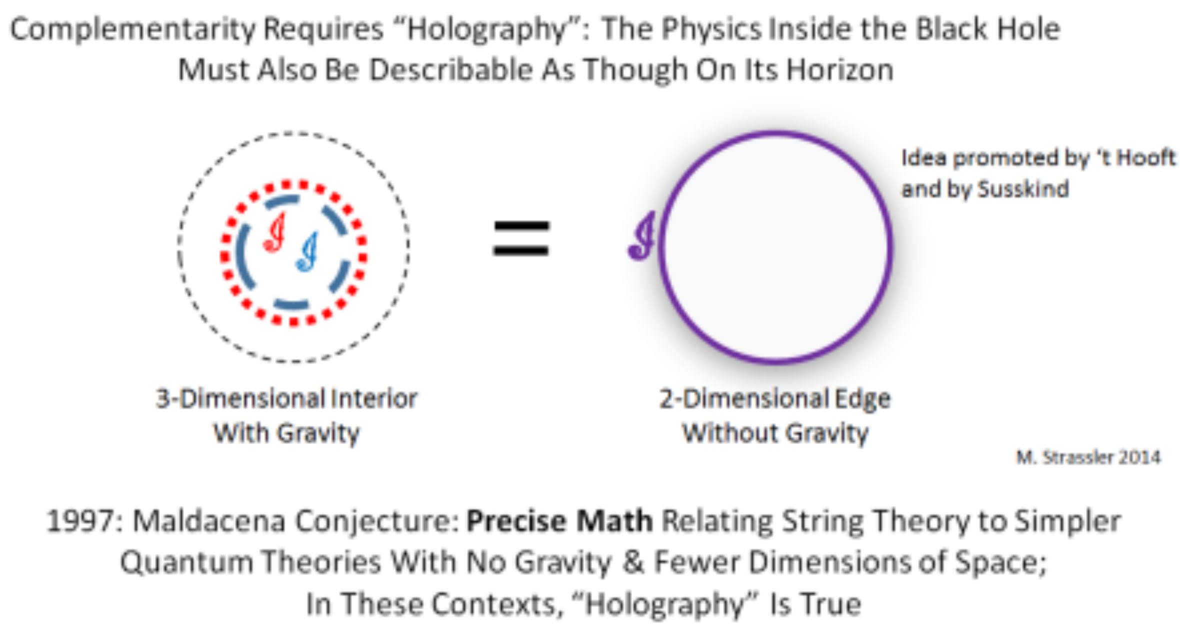

Fig. 7: Complementarity requires that everything happening in the interior of a black hole can be equally-well described as though it were all just outside the black hole. Remarkably this is not impossible, as was demonstrated in the late 1990s and early 2000s; string theory, which contains a quantum version of general relativity, exhibits this property in at least some circumstances.

Crazy at it sounds, considerable evidence arose in the late 1990s that it is true, at least in some situations! In 1997, Maldacena conjectured (and hundreds of people checked, in various ways) that under the right circumstances, string theory (a theory that is a quantum generalization of general relativity, and is a candidate for a theory of the laws of nature in our universe) is actually equivalent to a quantum theory (specifically, a "quantum field theory") **without** gravity and with fewer dimensions. This relationship, known variously as "AdS/CFT" or the "field/string" correspondence, deserves an article all its own (stay tuned for that.)

The success of holography gave additional credence to the complementarity idea. Furthermore, the field/string correspondence allowed for a very strong argument (perhaps a proof?) that small black holes can form and evaporate in the string theory via a process that *can be described by the corresponding quantum field theory* (though not explicitly) — and which therefore, as in all processes in any quantum theory, ***does preserve information***! By 2005, even Hawking had come around to this point of view — that in fact, as the complementarity proposal had suggested, black holes do not cause information to be lost, and that general relativity, but not quantum theory, must be modified.

The Firewall and the Current Turmoil

Still, there were loose ends in the complementarity proposal. Black hole evaporation is so subtle that there were still no quantum theory equations for complementarity that could describe the evaporation process. While trying to find such equations, Almheiri, Marolf, Polchinksi and Sully discovered that in fact (at least under reasonable assumptions) complementarity contains a self-contradiction, which shows up when a black hole has evaporated about halfway. The argument is extremely subtle, involving the kind of "quantum entanglement" that Einstein called "spooky" and that gets used in quantum computers. But crudely speaking, by the halfway point, so much information has departed the black hole in the Hawking radiation that there's not enough left at the horizon for holography to represent the black hole's interior. Consequently, instead of an in-falling observer smoothly entering the black hole through the harmless horizon, as in Figure 6, the observer finds there's no interior at all, and does so the hard way, by being fried to a crisp by a so-called "firewall" that hovers just outside the horizon (Figure 8).

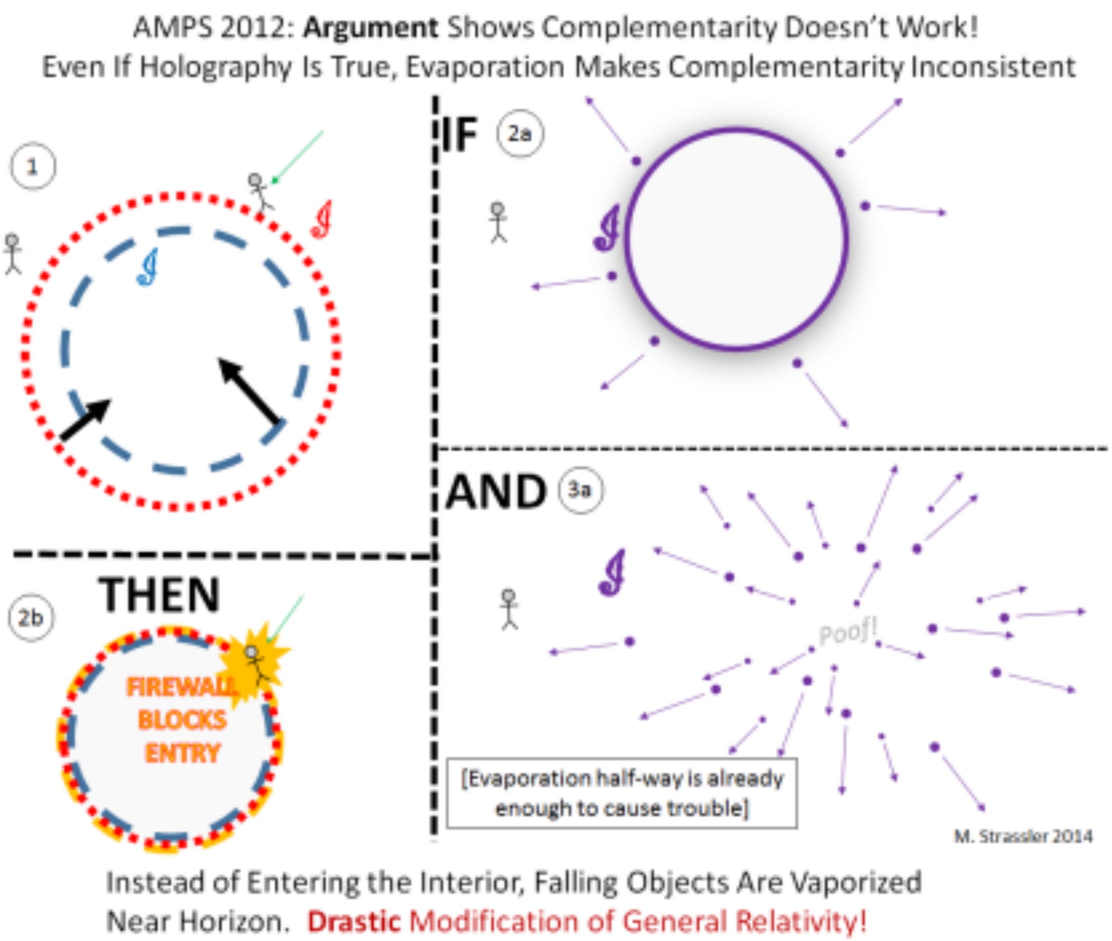

Fig. 8: But as a black hole evaporates, complementarity itself runs into a serious paradox. If the outside observer sees evaporation occur without loss (or copying) of information, than an in-falling observer crashes into something at the horizon — whimsically called a "firewall" — where there wasn't supposed to **be** anything! This violates expectations from general relativity in a very big way!

The possibility of a firewall would involve a very drastic modification of general relativity. It it were right, it would say that the picture provided of black holes by general relativity, that a black hole has a substantial interior, and that a horizon is nothing but a point of no-return (as in Figure 3) and not a place where special things happen as you pass it, is 100% wrong, once a black hole has done some significant amount of evaporating.

So now **the paradox is baaaaack**! And worse than ever. It seems that if quantum theory and complementarity are right, general relativity isn't just requiring some small modification — it requires major surgery! And there's no sign of such surgery in string theory, which provided the example of holography. But the field/string correspondence suggests quantum theory can describe black hole formation and evaporation, so information isn't lost. So can complementarity be replaced with something else? Or is one of the arguments that creates this paradox actually wrong?

Everyone's confused. There are lots and lots of proposals as to how to get out of this conundrum. You're not hearing about most of them. The media told you about Hawking's because he's famous, but he's really just one of many, many voices tossing ideas around. All of these ideas suffer from the same thing: not enough equations to provide evidence and details of how they're supposed to work. And since not having enough equations is what led to the firewall paradox, we can hardly try to get out of this situation by relying on yet another argument that lacks equations for its details!

But even though Hawking is just one person making a proposal, and even though his proposal lacks equations and is likely to be, at best (in my view), incomplete, and more likely just wrong, you probably want to know what he suggested. It's hard to figure that out without equations, but here's my best effort (see Figure 9). Hawking points out that although exteriors of black holes quickly become simple, the interiors can become very complex. Complex systems, like weather, can exhibit chaos, which can make them unpredictable even before you think about quantum theory. He seems to suggest that the complexity itself destabilizes the horizon and allows the information, having been scrambled inside the black hole, to leak back out. Since this would violate Hawking's own theorems about general relativity, I assume this means that general relativity must be modified. And since his argument rests on AdS/CFT (i.e. field/string correspondence), I assume he believes that this must occur in string theory. Since what went into a black hole does eventually come out, these holes are not really black after all — so call them "grey holes" or "metastable gravitational bound states" or "apparently-black holes", but "black" is perhaps not really the right term.

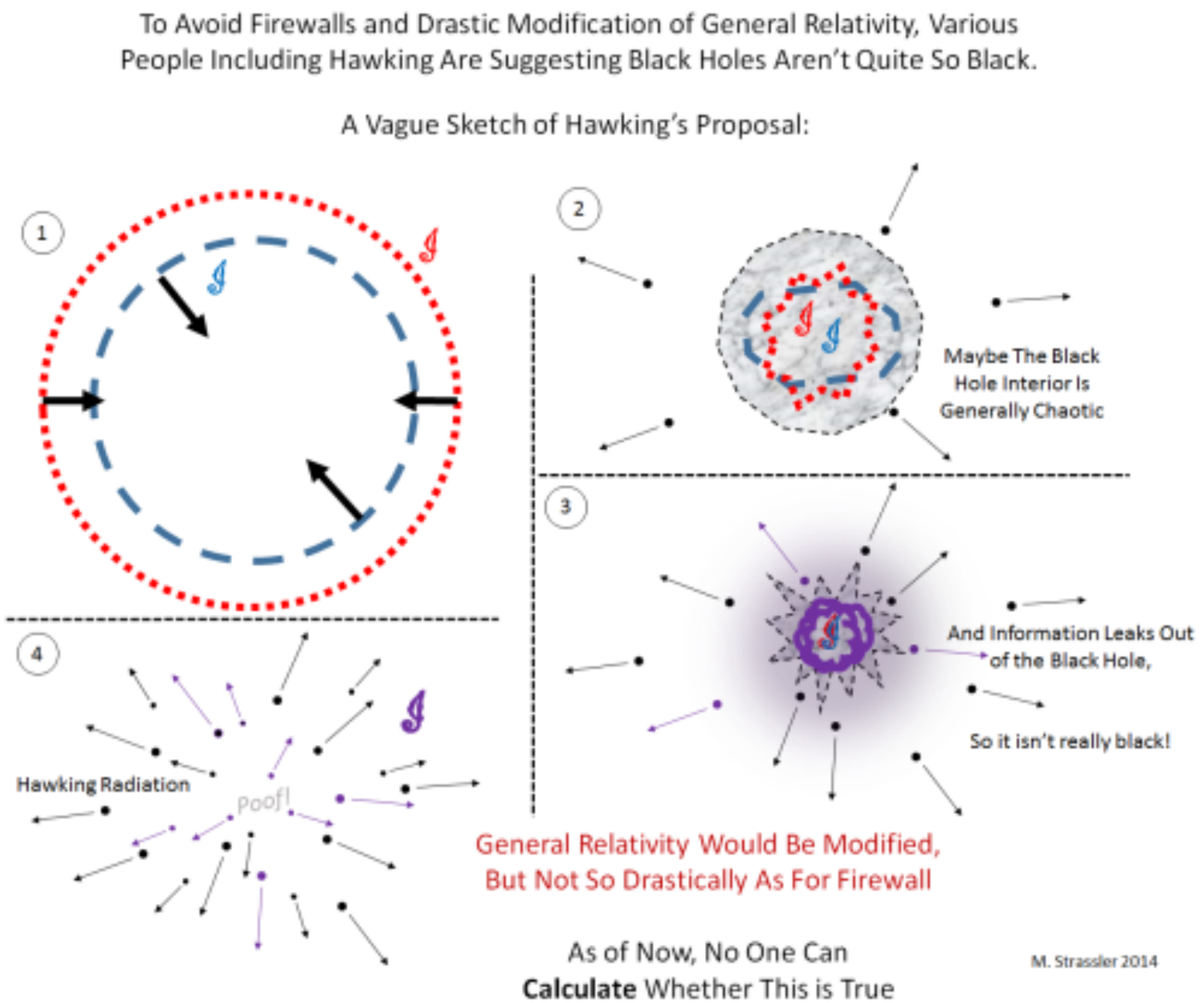

Fig. 9: With apologies to Hawking, since neither I, nor anyone I talk to, has a clear idea of what he has in mind, here's a rough sketch of what I think he's trying to suggest. Keep in mind that there are lots of other proposals out there, and this one is not necessarily the best one so far.

But there are many obvious problems with this proposal — not the least of which is that the firewall puzzle shows up already after the halfway point of black hole evaporation, not just at the end of the evaporation. And thus the black hole is still very large when the information has to be leaking out — which would seem very difficult to reconcile with a proposal like Hawking's. So don't expect a consensus to grow around Hawking's suggestion, especially not without some concrete equations to evaluate.

In any case, everything you learned about black holes is still basically true. Astrophysicists need not be concerned that there will be changes to what they think they know about stellar or galactic "black" holes. At least for a large and not-too-old "black" hole, Hawking's proposal wouldn't actually lead to any changes that you could measure. And if you fell in, you still couldn't get out, nor could you realistically send a message to anyone outside. So even if it turns out that "there are no black holes" in the legalistic sense, there is still a *spectacularly-dark* hole to be found at the center of nearly every galaxy in the universe.

Don't expect this 40-year-old puzzle to be resolved soon. And the resolution will probably come from a young physicist you've never heard of, or from a person not yet born.