Black Holes and Reality

Siddharth Muthukrishnan

February 26, 2024

Black holes are weird. Here are places in the universe where matter has collapsed to such density that a surface forms around it—the event horizon—such that even light cannot escape from within it.

Weirdness is not paradox. The paradoxes arise when we start to think about the quantum physics of black holes. The central paradox here is the black hole information paradox. Quantum physics tells us that black holes emit an incredibly faint, ghostly glow of radiation from just outside its horizon. This is Hawking radiation (Hawking 1975). As a black hole radiates, it slowly evaporates away, becoming smaller and smaller. According to quantum physics, this process must preserve *information*: it must allow for the complete reconstruction of the past from the future. This means the radiation from the black hole can't always be entirely featureless. Because as the black hole becomes smaller and smaller, the radiation becomes the bulk of what might contain information about what formed and fell into the black hole, and so it must at some point contain enough detail to allow for the reconstruction of that information. However, according to Hawking, the radiation from a black hole must always be *thermal*, i.e., always be featureless. This because the radiation emerges from just outside the horizon, and as such there's no way the radiation could have been causally influenced by what fell in, since all that stuff is behind the horizon. Obviously, the radiation from the black hole can't simultaneously be richly detailed (Shakespeare's collected works may have fallen in!) and be entirely featureless.

This is the black hole information paradox (see, e.g., Wallace (2020) for a more precise statement). It reveals a deep tension that arises when one applies quantum physics to black holes. It has irritated and inspired theoretical physicists for half a century. It has been a particularly fruitful direction of investigation because the paradox arises at energy scales we otherwise think we have a good theoretical handle on. And so the information paradox is not immediately just a special case of the well-known difficult problem of giving a complete and consistent theory of quantum gravity—a problem typically thought to arise only at much higher energies.

Our theories of black holes yield paradox; is that their death-knell? Not quite. Maybe not all parts of the theory are relevant to empirical tests and the empirically testable parts don't traffic in paradox. Or maybe a more careful interpretation of the theory can avoid paradox. These are routes taken by black hole complementarity—a collection of ideas intended, in some way or other, to sand-off the bite of the information paradox. In particular, I articulate two broad ways—*operational* and *descriptive*—of understanding black hole complementarity and evaluate their plausibility.

According to *operational* black hole complementarity, the information paradox does not lead to any violations of quantum mechanics that can be detected by an observer.

Really? Can't there be any experiment where the paradox is empirically manifest? Let's try the following. Say, we have two experimenters, Alice and Bob. Alice dives into the black hole; Bob hovers outside the black hole. Alice carries a quantum system with her. If the black hole radiation carries the information about what fell in, then eventually Bob will be able to recover from the radiation the state of the quantum system that Alice carried in with her. Recovered state in hand, Bob dives into the black hole. Now that Bob is inside the black hole along with Alice, Alice can send the original quantum system to Bob. (She couldn't do so earlier since she's was behind the horizon.) Bob would then verify that the state Alice had is the same as the state Bob recovered from the radiation. This would then mean that Bob would have verified a violation of the quantum no-cloning theorem, which says no process can produce a copy of an arbitrary quantum system. For that is what the black hole would have done: copied Alice's state onto the radiation. An observer, Bob, would've seen a violation of quantum mechanics.

But operational complementarity tells us that can't happen, and so the experiment above shouldn't be doable, even in principle. This is the case. Bob just does not have enough time to receive the quantum system from Alice before he crashes into the singularity at the heart of the black hole. This remains the case even when Alice's information exits the black hole as quickly as possible (Hayden and Preskill 2007). That is, the amount of time that Alice has to send the message to make sure Bob gets it before he hits the singularity is just shy of the amount of the time Bob has to recover the information and jump in.

This type of near-miss is frequently seen in the black hole complementarity literature: a promising experiment is proposed for an observer to detect a violation of quantum mechanics near a black hole, but some subtle barrier or other emerges to thwart such detection. A prominent example is the firewall paradox (Almheiri et al. 2013), which suggested a way an observer near a black hole could observe the violation of entanglement monogamy, an important constraint on quantum entanglement. But Harlow and Hayden (2013) showed that the experiment suggested by Almheiri et al. (2013) would require solving a computational problem that would take much longer than the lifetime of the black hole. This is a computational barrier against operationalizing the black hole information paradox.

Is operational complementarity satisfactory? In particular, what to make of the fact that it is stated in terms of what an observer will or will not be able to do? An instrumentalist about scientific theories would be satisfied: if scientific theories are simply useful instruments, then it is fine if they make reference to the instruments' users. But most philosophers reject instrumentalism. And for good reason: positivist ideas (of which instrumentalism is a piece) faced trenchant critiques in twentieth century philosophy (see, especially, (Quine 1951)).

Can we retain the power of black hole complementarity but avoid instrumentalism? Doing so would soften the blow of the information paradox while allowing us to take our theories as representing the world. This would be what I call *descriptive* black hole complementarity.

How might one articulate a descriptive complementarity principle? Many introductions to relativity theory talk about what observers see, with thought experiments involving observers, trains, and light pulses (e.g., Mermin (2006)). But relativity isn't an operational theory. Observers are dispensable—they're just placeholders for reference frames. In the context of black hole complementarity, the natural thought, then, is that the observers mentioned in operational complementarity (like Alice and Bob) are placeholders for systems of description. One system of description attaches to hovering observers, like Bob. Another system of description attaches to infalling observers, like Alice. Descriptive complementarity would then be the claim that these two descriptions are descriptively consistent: i.e., they can simultaneously be accurate representations of the world consistent with quantum mechanics. If descriptive complementarity is viable, then we can have the benefits of operational complementarity without the instrumentalism.

However, it isn't viable—at least not in the form above. Let's go back to the thought experiment where Alice fell in with a quantum system. But now go descriptive; discard talk of Alice: the quantum system just falls in. The system's quantum state later emerges in the radiation. Forget about Bob collecting that information. Stay descriptive. Let's just take the physics representationally seriously. If we do that, we see then that the black hole cloned the state that fell in. At one time, there was one quantum state in the exterior; at a future time, the state is in the interior and a copy in the exterior in the radiation. That is a violation of the no-cloning theorem. It may well be that this clone can never be observed, but that's immaterial for descriptivists. (A similar argument applies to the entanglement monogamy case.)

What do we learn from this? The answer will depend on your philosophical predilections. Say you lean realist. Then the success of operational complementarity is to be explained by future successful descriptive theories. The failure of descriptive complementarity shows not so much that getting a descriptive story of quantum black holes will fail but more that a particularly natural way of de-operationalizing black hole complementarity will fail. A more sophisticated descriptive theory might well succeed. Perhaps some of the cutting-edge work using holographic ideas (see, e.g., Almheiri et al. (2021) and references therein) to calculate the entropy of black hole radiation will be such a descriptive theory—but the jury is out.

But say you are fine with instrumentalism. Then, if operational complementarity continues to be successful, you might as well take the information paradox to be solved. For what is a paradox? It is a plausible argument to an absurd conclusion (Quine 1962). For an instrumentalist, such absurdity must be operationally relevant. But the success of operational complementarity suggests that such an absurdity would never be so relevant.

References

Almheiri, Ahmed, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. 2021. "The entropy of Hawking radiation." *Rev. Mod. Phys.* 93 (3): 035002. https://doi.org/10.1103/RevModPhys.93.035002. https://link.aps.org/doi/10.1103/RevModPhys.93.035002.

Almheiri, Ahmed, Donald Marolf, Joseph Polchinski, and James Sully. 2013. "Black holes: complementarity or firewalls?" *Journal of High Energy Physics* 2013 (2): 62.

Harlow, Daniel, and Patrick Hayden. 2013. “Quantum computation vs. firewalls.” *Journal of High Energy Physics* 2013 (6): 85. https://doi.org/10.1007/JHEP06(2013)085.

Hawking, Stephen W. 1975. “Particle creation by black holes.” *Communications in mathematical physics* 43 (3): 199–220.

Hayden, Patrick, and John Preskill. 2007. “Black holes as mirrors: quantum information in random subsystems.” *Journal of High Energy Physics* 2007 (09): 120–120. https://doi.org/10.1088/1126-6708/2007/09/120. https://doi.org/10.1088.

Mermin, N David. 2006. *It’s about time: understanding Einstein’s relativity.* Princeton University Press.

Quine, Willard V. O. 1951. “Two Dogmas of Empiricism.” *Philosophical Review* 60 (1): 20–43. https://doi.org/10.2307/2266637.

———. 1962. “Paradox.” *Scientific American* 206 (4): 84–99.

Wallace, David. 2020. “Why Black Hole Information Loss Is Paradoxical.” In *Beyond Spacetime: The Foundations of Quantum Gravity,* edited by Nick Huggett, Keizo Matsubara, and Christian Wüthrich, 209–236. Cambridge University Press. https://doi.org/10.1017/9781108655705.013.