How Problematic is Black Hole Information?

Pierce Comerford

January 5, 2024

Word count: 2093

The question whether a black hole can be considered a thermodynamic object—in particular, whether it can be said to have a meaningful temperature and entropy—has provoked controversy since at least the time of Wheeler [1,2], and each new contribution to the puzzle only appears to beget a new question of an equally controversial nature [1,3]. Hawking, initially a critic of Bekenstein's realist interpretation of his formulation of black hole entropy [1,4,5], later proved (to his own annoyance) [1] that black holes radiate with a temperature proportional to their surface gravity [6].

It is debated whether Hawking's proof settles the above question [2,7]. His proof, and later similar proofs [8], follow a standard setup of quantum field theory in curved spacetime, and are equivalent to showing that the Bugolyubov coefficients between vacuum states $|0_a\rangle$ in some coordinates (x_a^i) and $|0_b\rangle$ in coordinates (x_b^i) are nonzero – thus, since one coordinate system in general relativity has no special physical relevance over another, a vacuum in infalling coordinates at the black hole is a thermal bath of particles to observers at infinity. This raises all of the same ontological questions for a quantum gravity theory as does the Unruh effect [9,10], and comes with the same set of pitfalls for interpretation of the result. Quite apart from asking whether the emitted particles are 'real particles' and the extent to which they constitute 'radiation', there is also the question whether we can assign to the black hole the status of an 'emitter'. The temperature of the thermal bath is proportional to the surface gravity of the black hole and thus identical to Bekenstein's result [4], but this seems to be miraculous: there is no obvious way to connect this entropy to the one described by Bekenstein, who actually interprets his quantity S as referring to the "equivalence class of all black holes which have the same mass, charge, and angular momentum, *not to one particular black hole*"[1]. Still, Hawking's result *does* appear to quell Wheeler's original worry that black holes violated the second law of thermodynamics [1], since the radiation should allow the black hole to thermalize with its surroundings, at least in the semiclassical regime used in [6]. But it almost immediately raises a new problem in that formation and subsequent total evaporation of a black hole represents a pure-to-mixed state transformation [3,11,12], violating the unitary principle of quantum mechanics. Hawking had such reverence for the implications of this problem that [11] was original published as 'The Breakdown of Physics in Gravitational Collapse', that is, that the unitarity violation prevents us from properly time-evolving a state even in principle, which he believed constituted an undermining of the predictive power of a physical theory [3].

It is worth comparing this to other theoretical results which make similar claims. For instance, in a recent paper [13], Purcell *et. al.* construct a system

[1] My emphasis.

where the location of a phase transition corresponds to an incomputable (Chaitin) number. If the system is allowed to evolve, the phase transition may be found and the location measured, but it is in principle impossible to (accurately) predict the value of the variable ahead of time. This is similar to earlier results on the undecidability of the spectral gap problem in one and two dimensions [14,15], and both cases deal with ideal systems with no connection to real physical systems – in the first case, the lattice is infinite, and the second two are of course of lower dimension.

But less us suppose they may be generalised to apply to more realistic cases. How different are they from everyday cases in which exact computation is in principle possible, but not achievable in practice – the motion of projectiles in the Earth's atmosphere, for example [16] – or for which there is no nonanalytic solution – the solution to Schrödinger's equation for transhydrogenic atoms, or most cases of many-body mechanics? There, at least, it is well-accepted that the solutions to these problems can be computed to arbitrary accuracy using some decision procedure. While a Chaitin construction Ω is incomputable, it can be bounded from below to arbitrary precision, although it is impossible to compute whether a given step Ω_s in some (computable) sequence $\lim_{s\to\infty} \Omega_s = \Omega$ is a good approximation [13]. Practically this is dissimilar in some ways to a non-analytic but computable solution, but arguably not in any way that is important: QED, the "jewel of physics" [17] gives a prediction for the anomalous magnetic dipole moment with a precision of about one part in a billion [18]; Calude *et. al.* compute 64 bits of a Chaitin construction in [19], which is one part in $2^{40} \approx 10^{19}$— clearly far better than the celebrated result from QFT. Practically, then, these undecidable results should not alarm us much more than nonanalytical ones.

Is the nonunitarity violation in Hawking's paper disastrous enough to call it a "breakdown of physics", or is it a similarly benign result? The pure-to-mixed state evolution derived in [11] breaks unitarity: given some state $\psi(t_0)$ of the system there exists t for which we cannot find $\psi(t_0 \pm t)$, *i.e.* we cannot pre- or retrodict the system on either side of black hole evaporation. The moniker of an *information* paradox suggests, informally, that what we know about the system is lost at some point in the formation-to-evaporation process. But whereas, in the above cases, we have a well-defined rule for time-evolving the system, we are now missing the map between states at different times altogether.

What is different about this result is that it seems to present a strong barrier for a consistent notion of causation — and so do many attempts at a resolution. To illustrate this point we may look at a pedagogically deterministic system. Consider a ten-pin bowling setup determined entirely by Newtonian mechanics, *à la* the billiards setup in [20]. Here, the fate of each pin depends not only on the trajectory of the bowling ball, but also on the trajectory of each other pin. Suppose however that this is a magic bowling ball, and that the first pin it impacts disappears without a trace, such that examining the system after a round we could only conclude that we had engaged in a game of nine-pin bowling. Like the pure-to-mixed state operator in [11], this is not an invertible transformation. We know how to describe the magic bowling ball, but playing the footage of our strike in reverse shows us a completely different series of events than would be retrodicted by examining the positions and momenta of the nine pins we end up with: we can easily describe nine pins moving back into position, but there is no reason to suspect that a tenth pin would appear at the last moment to knock away the bowling ball. We may say that we have lost information by losing a pin, but we also may put it as there being an asymmetry: bowling is causal in one direction, but not another.

Introduction of magic bowling balls certainly seems to lead to a “breakdown of physics”, much moreso than incomputable or nonanalytic time-evolutions: they assert that something effectively happens *for no reason*. This might be a joyful notion for Epicureans, who need a “random swerve” as the source of free will [21,22], but it is a disaster for physics. The sudden appearance of a tenth pin is unlike the sudden appearance of a virtual particle due to vacuum fluctuations in quantum field theory (or any other ‘random’ process in quantum mechanics), since that process is still fully determined by the equations of motion of the system. Most physicists, then, seek to find a solution to the paradox that allows unitary transformations [12] by somehow allowing that all of the information about the black hole is encoded in the outgoing Hawking radiation, that is, the nine pins carry some ‘memory’ of the tenth one.

There are, of course, other motivations for why the evolution must be unitary, one of the strongest being evidence from AdS/CFT correspondence [23]. Page [24,25] argues that a unitary evolution of a black hole implies that the von-Neumann entropy of the system must increase for a time t_{Page} and then decreases sharply to zero. A similar phenomenology can also be reached by allowing small amounts of nonunitarity, however, by considering the black hole as a system of quantum circuits [26], and moreover these present a difficulty in terms of testing, since the Hawking temperature of nearby black holes is much lower than that of the CMB [7]. And either way this clashes with Hawking's result [11]. Much work on the paradox is on trying to develop a new internal black hole mechanics which corrects the thermal spectrum slightly.

More especially, unitarity-preserving models of interior black hole mechanics that amend the pitfalls of Hawking's semiclassical have the issue that an observer falling into the event horizon should thermalise in a non-local way, violating the speed of information transfer. This has its own set of complications for causation, since it can imply, among other things, retrocausality. The proposed resolution by Susskind *et. al.* is the notion of black hole complimentarity: to infalling observers, information is absorbed, never to be revealed again to the outside world. To the outside world, the information is reflected in a way that does not violate information transfer speed. [27].

This fix has its own set of physical problems (see, for instance, the firewall paradox [28]), but how good is it as a philosophical fix? What is the difference between absolute abolition of superluminal information and merely making it inaccessible? To some extent the difference could be considered minor: the inaccessibility could be seen as having the same strength (or weakness) as the mutual inaccessibility of events over a spacelike interval – it is an extra stipulation equal in rank. But complimentarity presents a more sinister threat in that it does not provide a 'splitting' mechanism for the information, and instead posits that the same information can have two destinies at once (even if this is not detectable), which at least superficially violates Leibniz' Law [29]. Unlike the many-worlds approach, this does not present a new notion of particle history or future [30]; it doesn't separate worlds, it simply *ad-hoc* adds to physics the possibility that information may be separated *from itself* by a spacelike interval.

It should also imply that, after an observer falls into a black hole, she can write a report on all of the information collected, noting that all of the information must never have been reflected. Eventually the information must be radiated to outside observers, or else the resolution does not work, and such an observer will possess a report which blankly contradicts its own existence. The observer is left with irrefutable evidence of two mutually contradictory facts, limiting his ability to deliberate [20] and especially placing limits on what he reasonably expects a physical theory to do. We would suppose that if incomputable, undecidable or nonanalytic predictions are ok but acausal ones are not, then one for which possible physically reasonable rational explanations present a manifest contradiction should also be disallowed. Amending our theories with a complimentarity principle, even if mathematically self-consistent, threatens the ability of our theories to allow for conclusions that make sense to a reasonable observer, in just the same way that a magic bowling ball (the unresolved paradox simpliciter) would.

We conclude, then, by suggesting that, while unitary resolutions to the paradox should be focused on for their apparently better faithfulness to certain important scientific principles than nonunitary alternatives, the speculative nature of the mathematical and physical work currently being done is fraught with difficulties of a similarly unfortunate nature. A further essay may suggest some better philosophical frameworks for looking at the complimentarity principle in a noncontradictory way, but we posit that the shift in framework would have radical consequences for how we look at identity, rational deliberation, continuity, and causation.

Bibliography

[1] C. R. Almeida, The thermodynamics of black holes: from Penrose process to Hawking radiation, Eur. Phys. J (2021).

[2] E. Curiel, The many definitions of a black hole, Nature Astronomy **3**, 27 (2019).

[3] G. Belot, J. Earman, and L. Ruetsche, The Hawking information loss paradox: The anatomy of a controversy, Brit. J. Phil. Sci **50**, 189 (1999).

[4] J. D. Bekenstein, Black holes and entropy, Phys. Rev. **7**, (1973).

[5] J. Bardeen, B. Carter, and S. Hawking, The four laws of black hole mechanics, Commun. Math. Phys **31**, 161 (1973).

[6] S. W. Hawking, Particle creation by black holes, Commun. Math. Phys **43**, 199 (1975).

[7] E. Curiel, *Singularities and Black Holes*, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman, Summer 2023 (https://plato.stanford.edu/archives/sum2023/entries/spacetime-singularities/; Metaphysics Research Lab, Stanford University, 2023).

[8] D. Deeg, Quantum Aspects of Black Holes, Ludwig-Maximilians-Universität München, 2006.

[9] M. Kuhlmann, *Quantum Field Theory*, in *The Stanford Encyclopedia of*

Philosophy, edited by E. N. Zalta and U. Nodelman, Summer 2023 (https://plato.stanford.edu/archives/s
field-theory/; Metaphysics Research Lab, Stanford University, 2023).

[10] S. Carroll, *Spacetime and Geometry: An Introduction to General Relativity* (2004).

[11] S. W. Hawking, Breakdown of predictability in gravitational collapse, Phys. Rev. D. **14**, (1976).

[12] D. Wallace, Why black hole information loss is paradoxical, (2018).

[13] J. Purcell, Z. Li, and T. Cubitt, *Chaitin Phase Transition*.

[14] T. Cubitt, D. Perez-Garcia, and M. M. Wolf, Undecidability of the spectral gap, Forum of Mathematics, Pi **10**, (2022).

[15] J. Bausch, T. S. Cubitt, A. Lucia, and D. Perez-Garcia, Undecidability of the spectral gap in one dimension, Phys. Rev. X **10**, 031038 (2020).

[16] S. Weinberg, *Dreams of a Final Theory: The Search for the Fundamental Laws of Nature* (Random House, 2010).

[17] R. Feynman and A. Zee, *QED: The Strange Theory of Light and Matter* (Princeton University Press, 2014).

[18] G. Gabrielse, D. Hanneke, T. Kinoshita, M. Nio, and B. Odom, New determination of the fine structure constant from the electron g value and qed, Phys. Rev. Lett. **97**, 030802 (2006).

[19] C. S. Calude, M. J. Dinneen, and C.-K. Shu, *CDMTCS Research Report Series Computing a Glimpse of Randomness Computing a Glimpse of Randomness*, in (n.d.).

[20] A. Fernandes, A deliberative account of causation, Philosophy and Phenomenological Research **95**, 686 (2017).

[21] D. C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting*

(MIT Press, 1984).

[22] T. L. Carus, *On the Nature of Things* (Hackett Publishing Company, Indianapolis, 2001).

[23] D. Harlow, Jerusalem lectures on black holes and quantum information, Reviews of Modern Physics **88**, (2016).

[24] D. N. Page, Information in black hole radiation, Physical Review Letters **71**, 3743 (1993).

[25] D. Page, Hawking radiation and black hole thermodynamics, (2004).

[26] P. Hayden and J. Preskill, Black holes as mirrors: quantum information in random subsystems, (2007).

[27] L. Susskind, L. Thorlacius, and J. Uglum, The stretched horizon and black hole complementarity, Physical Review D **48**, 3743 (1993).

[28] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, An apologia for firewalls, Journal of High Energy Physics **2013**, (2013).

[29] P. Forrest, *The Identity of Indiscernibles*, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta and U. Nodelman, Winter 2024 (https://plato.stanford.edu/archives/win2024/entries/identity-indiscernible/; Metaphysics Research Lab, Stanford University, 2024).

[30] L. Vaidman, *Many-Worlds Interpretation of Quantum Mechanics*, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Fall 2021 (https://plato.stanford.edu/archives/fall2021/entries/qm-manyworlds/; Metaphysics Research Lab, Stanford University, 2021).