Models of Black Hole Evaporation
– or –
What to Do When You Can't Solve Equations

Erik Curiel

Munich Center For Mathematical Philosophy
Ludwig-Maximilians-Universität

Black Hole Initiative
Harvard University

Smithsonian Astrophysical Observatory
Radio and Geoastronomy Division

erik@strangebeautiful.com

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 312032894.
(Funded by the German Research Foundation (DFG) – project number 312032894.)

Principle Models

Black Hole Evaporation

Epistemic Value, Understanding and Realism

L. Patton, "Fishbones, Wheels, Eyes, and Butterflies:..." (p. 4):

> *The structuralist, structural realist, and methodological realist accounts all rely on one thing: a history of successful scientific theories in which not only posits about entities, but also formal elements like variables and relations, are indispensable instruments.*
>
> *I want to begin by raising the question: what happens, then, if precisely the problem that faces scientists is the inability to solve the equations, the inadequacy of our current theory to provide a model of a given domain in which the results we can prove in that domain can be represented as consequences of a deductive axiomatic system?*
>
> *What happens if a scientific theory is successful, but not in the sense that the equations of the theory allow it to be represented [explicitly] as a deductively complete and categorical system?*

(ask me about '[explicitly]' during discussion, if you're curious) the human condition (or, life in the big theoretical city):

for every theory we have, we know almost nothing about the nature of "generic solutions" to the equations of motion or field equations—individual models that we have no idea how to construct in any way graspable by the human mind, or indeed even how to identify if one bit us on the ass

(if one holds that a theory is characterized in a strong sense by its family of "models", then the overwhelming majority of the content of a theory is something that we do not know, are in fact nowhere near knowing, and have good reason to think we will never know in anything resembling thoroughness and detail—epistemologically speaking, to postulate the very idea of a theory as a well defined structure in this sense is an act of faith) what to do, then, when you can't solve equations, and:

1. either other mathematical difficulties prohibit approximative methods and numerical simulation (as in classical general relativity and semi-classical gravity);
2. or one does not want detailed information about individual models or individual possible solutions (because, *e.g.*, the only ones we can derive are trivial in one way or another—as in GR and SCG),
3. and narrow classes of solutions (depending on exact symmetries, *etc.*) just won't do the job?

especially when it is difficult to experimentally access regimes in which one expects the behavior of interest to manifest itself—as in GR and SCG

Then one can try to prove general results; but what exactly is it that one is then doing?

theorems

(1) either assume something of a suitably generic character (*e.g.*, an energy condition), and try to do derive or argue for something of a suitably generic character

1. non-existence claims (non-singular cosmological models)
2. rigidity/stability claims (topology—past-singular closed cosmological models satisfying the strong energy condition)
3. scarcity/genericity claims (measure—cosmological models with a Killing field)
4. non-constructive existence claims (classic singularity theorems)
5. non-constructive behavioral claims (formation of closed trapped surfaces under gravitational collapse)

principle models

(2) or construct models based on general principles—*principle models*:

1. fix general principles and generic conditions (the framework of SCG)
2. characterize general structures that embody the principles and conform to the conditions (a quantum field satisfying the semi-classical Einstein field equation on an asymptotically flat black hole spacetime)
3. derive a statement about generic features of the character and behavior of the general structures (Hawking effect: black holes evaporate)

what exactly is it that one is doing with a principle model?

- these models often involve behavioral claims (in the sense of a general result)
- they are not individual solutions to equations of motion or field equations
- usually, no exact individual solutions are known that represent any thing like such systems, especially not in the generality postulated
- they do not otherwise represent individual systems in any straightforward sense
- sometimes, they are not grounded in exact, rigorous general results
- they are not approximations to or idealizations of solutions
- there are no clearly specifiable families of solutions they correspond to

- they rather represent general features that we expect certain (often loosely characterized, merely postulated) families of solutions to have
- almost always, one can (loosely) characterize such (postulated) families in several different, often mutually contradictory ways
- thus, they have an interpretive looseness and a flexibility of consequence to them not characteristic of exact or approximative solutions

black hole evaporation by the Hawking process in SCG is a beautiful example

there are no exact solutions of the SCEFE known for a spacetime representing an evaporating black hole

and the approximate ones we have are severely limited in their scope and in their potential to yield general insight, by dint of various kinds of almost absurdly idealizing assumptions (*e.g.*, two-dimensional spacetime) deriving the principle model:

1. topological assumptions (domain of outer communication is simply connected)
2. causal assumptions (no CTCs)
3. stability assumptions ("small perturbations do not destroy the event horizon")
4. assumptions about asymptotic behavior and structure (instantaneous symmetries, "almost-conserved" quantities, energy fluxes that don't contribute to curvature)
5. behaviorial assumptions about the coupling of the metric and the matter field ("almost-stationary", adiabatic)
6. arguments *cum* assumptions about the singularity structure of the expectation value of the renormalized stress-energy tensor operator for a given kind of quantum field near the event horizon

$\Longrightarrow$ the principle model:

complex, largely heuristic arguments based on subtle assumptions relating quasi-local intra-spacetime behavior to global, asymptotic behavior

> "local negative energy fluxes associated with the positive thermal Hawking flux decrease the global ADM mass, which in turn decreases the area of the event horizon—radiating black holes evaporate"

Hawking (1975), "Particle Creation by Black Holes":

> *It should be emphasized that these pictures of the mechanism responsible for the thermal emission and area decrease are heuristic only and should not be taken too literally. It should not be thought unreasonable that a black hole, which is an excited state of the gravitational field, should decay quantum mechanically and that, because of quantum fluctuation of the metric, energy should be able to tunnel out of the potential well of a black hole.*

interpretive looseness and flexibility of consequence as a general feature of principle models: not a unique behavior, nor system, nor even kind of system is picked out

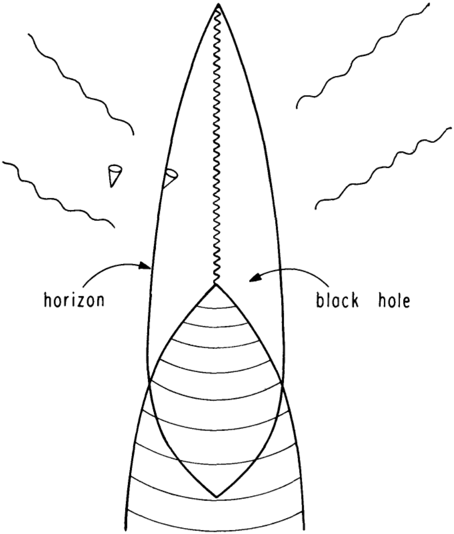


Figure 1 A space–time diagram showing the formation of a black hole by the collapse of matter and its subsequent evaporation by emission of Hawking radiation.

(taken from Wald 1984, "Black Holes, Singularities, and Predictability", in *Quantum Theory of Gravity: Essays in honor of the 60th birthday of Bryce S. DeWitt*, ed. S. Christensen)

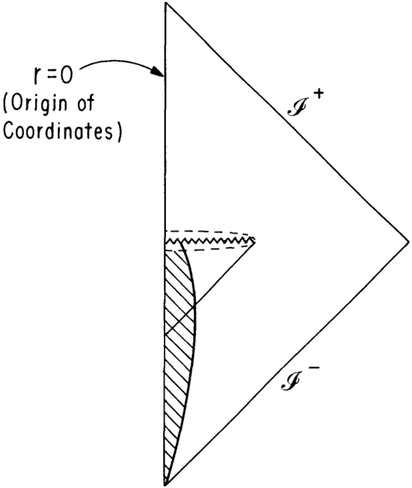
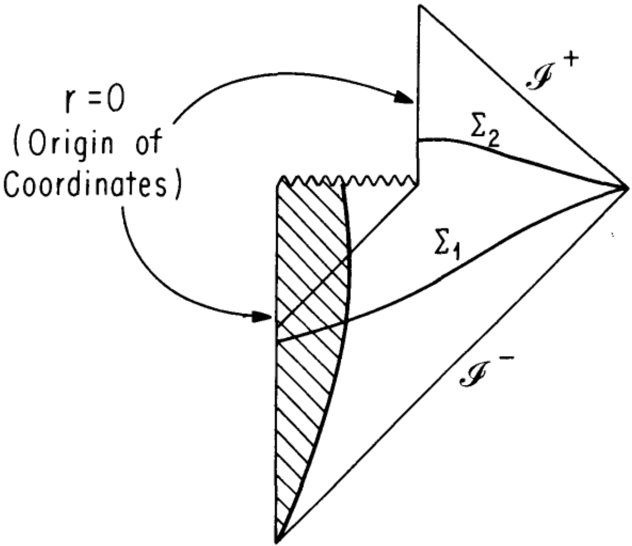


Figure 2 A conformal diagram showing the causal structure of a space–time where a black hole forms and evaporates in the manner of possibility (2) discussed in the text. Information cannot propagate outward through the black hole horizon without violating causality. The dotted lines near the classical singularity indicate a region where the classical description of space–time structure breaks down, thereby perhaps avoiding the presence of a singularity and allowing the information contained in the black hole to escape. However, this would require an enormous release of information with total energy only of order of the Planck energy. $\mathscr{I}^-$ and $\mathscr{I}^+$ are past null infinity and future null infinity.

(*ibid*.)



(*ibid*.)

Principle Models

Black Hole Evaporation

Epistemic Value, Understanding and Realism epistemic value:

1. what can we learn?
2. what can we do with them?

depends of course on the character and the epistemic warrant of the general principles or conditions one has assumed in the construction of the model

1. general behavioral claims that cannot easily be captured by theorems (Cosmic Censorship: "the maximal globally hyperbolic future developments of generic, physically reasonable initial data are future inextendible to regular Lorentzian manifolds")
2. or that are such that the theorems that do imply or characterize them are technically abstruse and have no perspicuous physical interpretation, and so the general behavioral claims ("radiating black holes evaporate in a weakly predictable way, but not retrodictably") are epistemically more valuable for the purposes of further exploratory investigation than the technically secure results ("Let $(\mathcal{M}, g_{ab})$ be a time orientable spacetime and let Σ_1 and Σ_2 in it be closed achronal sets such that Σ_1 is connected and Σ_2 is edgeless. Suppose there exists a point $p \in D^+(\Sigma_1)$ such that $p \notin J^-(\Sigma_2) \cup J^+(\Sigma_1)$. Suppose further that the set K, defined by $K := \Sigma_1 - D^-(\Sigma_2) \cap \Sigma_1$, has the property that $J^+(K) \cap \Sigma_2$ has compact closure. Then Σ_2 cannot be contained in $D^+(\Sigma_1)$.")
3. especially when the general behavioral claims go beyond "strict interpretations" of the technical results, (*e.g.*, one has to make further assumptions, not rigorously formalizable in our current state of knowledge, about the character of the propagation of quantum fields in a curved spacetime)

- principle models thus seem to give us *understanding* in some sense—but what sense is that?
- what can we *not* learn?

- SCG is a fascinating case: general results and principle models seem to be the only way we have to proceed, given the manifest technical difficulty of, and deep conceptual and physical problems involved in, trying to combine GR and QFT, especially with regard to the divergences and other pathologies accruing to solutions of the SCEFE
- we cannot learn from the models alone which, if any, of the compatible families of "more exact models" are the "right ones"
- in this case, in the absence of experimental probing, what we cannot learn is whether there is any unimpeachable veracity to the results and models—whether there are any systems that behave in any way consistent with the model
- but we can still have some (weak) confidence that we are on the right track, given our epistemic attachment to the general principles (local validity of QFT in curved spacetimes) and to the reasonability of the generic conditions (no causal pathology) used in constructing the principle models, and our epistemic faith in the plausibility of how we combine them (the SCEFE)
- also can gain warrant from forms of consilience and robustness analysis: derivations of same model from different assumptions and from "perturbations" of same assumptions

"Understanding" in this sense, then, is something like "the capacity to operate successfully in the scientific enterprise, in all its aspects and parts", to use our representations and instruments as the basis for the fruitful continuation of the enterprise:

- as part of evidential warrant in testing,
- as basis for predictions and characterizations,
- as inspiration for potentially fruitful new investigations,
- as the grounds for conceptual clarification and innovation in foundational work;
- and perhaps most of all to grasp how our representations and our instruments relate to, inform, and substantively contribute to the constitution of each other,
- and to grasp that in such a way that it founds the work succesfully continuing the enterprise.

insight into epistemology and general structure of scientific knowledge (in physics at least)? of its semantics?

1. the nature of the "representation" of physical systems by principle models is not straightforward
    - they don't represent individual systems, well defined classes of systems, . . .
    - so what do they represent? and how?
2. difficult to make sense of what these can teach us using, *e.g.*, something like the semantic view:
    - the whole point is we don't have access to the models—so, *inter alia*, we cannot characterize their empirical content or judge their empirical adequacy using any of the tools its provides
    - rigidity/stability claims and their ilk, in particular, are relations *among* models, which the semantic view is ill-equipped to handle
3. the ineliminable role of pragmatic factors: often, if not always, the general state of the entire epistemic community bears on whether one is justified in accepting a principle model and using it as the ground for further argument, explanation or prediction in a way that one has grounds for believing has at least *prima facie* epistemic warrant—what widely accepted standards of heuristic argument with regard to a particular theory, for example, constrains what can legitimately be thought of as a "prediction" of the theory? (*n.b.*: these can and do change over time)

1. whatever understanding it gives us, it is not understanding that can be used as evidential warrant for the theory in any straightforward, traditional sense, for we cannot test the truth of the general propositions, only corroborate them in special cases, and often we cannot even do that (the singularity theorems of GR)—not the problem of induction, but rather that we don't know with any exactitude what "the general case" is
2. there is no sense in which it is fruitful to conceive of these as "fictions", whether felicitous or not
3. but neither is there any clear sense in which they may be (or not) "approximately true"
    - what is the quantifiable measure of approximation for general behavioral claims?
    - finding one or a dozen or a million systems that, in some sense, behave approximately like that schematized by the principle model does not show that the principle model captures something "approximately true" about the world in any *representational* sense, only about the application of the general principles and generic conditions to those kinds of systems

principle models go far beyond the “tools used for prediction” of the shallow instrumentalist

but they also fall far short of the representational needs of the naive realist

they rather give us understanding characterized by a fruitfulness of conception—a *pragmatic* virtue valediction in the spirit of Kierkegaard and Stein, following Patton's pregnant remarks:

Realism is not a proposition whose truth-value can be affirmed. It is a way of thinking, an attitude, a style of moving forward in one's cognitive work, shaping and informing but not contributing to the content of one's thought. It is a leap of the absurd into faith, but at the same time a sober reflection that is only one step in the ongoing dialectical dance of one's cognitive capacities in their attempt to explore the world, itself to be given over to instrumentalism—also a leap of the absurd—as the times demand. And that will give over in its turn back to realism, in the Bacchanalic comedy of thought, which, from time to time, settles down into the stately progression of the tragedy of understanding, and back again and again—eventuating in the pragmatic attitude. And so: one is a realist, and one regrets it; one is an instrumentalist, and one regrets it; one is a pragmatist and learns to take joy in passing from one regret to the other.