7 Black holes, dumb holes, and entropy

William G. Unruh

It has now been 25 years since Hawking (Hawking 1974, 1975, Bardeen, Carter, and Hawking 1973) first surprised the world of physics with his analysis of quantum fields near black holes. Black holes, as their name implies, were believed to be objects into which things could fall, but out of which nothing could come. They were the epitome of black and dark objects. However, Hawking's analysis predicted that black holes should radiate, the radiation should be continuous and thermal, and the temperature should be inversely proportional to the mass of the black hole. Since black holes can also be said to have an energy proportional to their mass, this result led to opening of a whole new field of black hole thermodynamics.

That black holes could behave like thermodynamic objects had been intimated by results over the the previous five years. Christodolou (1970), Hawking and Ellis (1973, especially Lemma 9.2.2), Misner, Thorne, and Wheeler (1973) and Bekenstein (1973, 1974) had shown that there were certain formal similarities between black holes and thermodynamic objects. In particular, if one assumed positive energy for matter (an uncontested assumption), then – as Hawking most clearly showed – the area of a black hole horizon does not decrease. However, this formal similarity with entropy, which also does not decrease for an isolated system, did not seem to have any real relation with thermodynamics. The entropy of a body does not decrease only if the body is isolated, and not in interaction with any other system. In interaction with other systems, it can, and often does decrease. Furthermore, for ordinary systems, entropy and thermodynamics are primarily of interest in situations in which the system is able to traverse a closed cycle. Black holes on the other hand, have a surface area which always increases when in interaction with other systems. They could never engage in a closed thermodynamic cycle.

Hawking's discovery overturned that picture of black holes. The mass, and thus also the surface area (which is a function only of the mass) of a black hole could decrease, by the emission of thermal radiation into the outside world. That thermal

radiation carries away energy, and thus mass, from the black hole. At least formally, the emission of energy was accompanied by the absorption of negative energy by the black hole. This negative energy violated the assumptions of the Christodolou–Hawking theorems, and could therefore result in the shrinkage of the area of the black hole. Suddenly the formal analogy of area to entropy seemed far from simply formal.

This interpretation of the area as an entropy also fits together with the thermal nature of the emission. Since the black hole has a mass, and thus an energy, and since the radiation emitted was thermal with a universal temperature which depended only on the properties of the black hole, and not of the matter fields which were radiated, one could use the standard thermodynamic relations to derive an entropy:

$$T\,dS = dE \tag{7.1}$$

or

$$\frac{dS}{dM} = \frac{1}{T}. \tag{7.2}$$

Since (in the simplest cases) T was only a function of M, one could (up to a constant of integration) obtain an expression for the entropy. Surprisingly (or not) this entropy is exactly proportional to the surface area of the black hole. Working in units where $G = \hbar = c = k = 1$, this relationship turned out to be just

$$S = \frac{1}{4\pi}A, \tag{7.3}$$

where A is the area of the horizon of the black hole.

One thus had the astonishing situation that an entropy (in statistical mechanics, a measure of microscopic plenitude for a given macroscopic state) arising from purely quantum considerations was determined by a purely classical macroscopic non-quantum feature of the black hole, namely its area. The past twenty-five years have seen a persistent attempt to understand this puzzle, and a persistent failure to do so. This chapter will outline some of the directions which the attempt to do so has taken.

This chapter makes no pretense to be an exhaustive catalogue of the many ways in which people have tried to understand black hole thermodynamics. Instead, it treads a path which I have followed in trying to attack the problem. As a personal odyssey, it is thus as idiosyncratic as could be expected. It also will neglect to mention other approaches, not because I think they should be ignored – they certainly should not – but because of space limitations, and because I am often not sufficiently expert to present them here. I thus apologize in advance to both the reader and to my colleagues for my sins of omission.

In trying to understand black hole thermodynamics, there are at least two separate things one must try to understand: what causes the radiation which the black hole gives off and why is that radiation thermal; and how does the entropy fit in with the standard view we have of the statistical origin of entropy for all systems other than black holes. My own efforts have concentrated on the former question, and thus

this chapter will also concentrate on that same question, coming back to the latter problem in the later sections.

One note about units. I will work throughout in a natural system of units called Planck units. In these units, the fundamental constants – the velocity of light, Planck's constant, Newton's gravitational constant, and Boltzmann's constant – all have the value of unity. In this system of units all dimensioned quantities are expressed in terms of fundamental units. Thus distances are expressed in units of 1.6×10^{-33} cm, time in units of approximately 5×10^{-44} s, mass in terms of 2.1×10^{-5} g, and temperature in terms of 1.4×10^{32} K.

7.1 Thermodynamics of black holes

With Hawking's calculation, we were faced with the necessity of taking seriously the idea that black holes were true thermodynamic objects. One of the key uses of thermodynamics, and the reason for its invention in the nineteenth century, was its explanation of the operation of heat engines. In particular, the second law of thermodynamics proved to be very powerful in its predictions concerning what kinds of engines could be built, and what the ultimate efficiency of any heat engine was. The limitations imposed on the operations of engines by the second law were so powerful that they led throughout the past two centuries to persistent attempts to violate the second law, to create heat engines which were far more efficient in their conversion of heat to work than was allowed by the strictures of this law. It has of course been one of the minor embarrassments of physics that no complete fundamental proof of the second law exists. However, thousands of examples of the failure of attempts to violate it have given us great confidence in its validity, as do very strong statistical plausibility arguments.

If black holes are thermodynamic objects, we now have another system which one can introduce into the operation of a heat engine, and we have to ask once again whether or not machines which now make use of black holes in a fundamental way also obey the second law, or whether closed thermodynamic cycles can be found in the operation of heat machines with black holes which violate that second law.

This investigation began with Bekenstein and Geroch[1] well before Hawking's discovery. When Bekenstein suggested that the Christodolou–Hawking results could be interpreted as showing that black holes had an entropy, Geroch suggested an example of a heat engine which could then be used to violate the second law (including black holes). Fill a box full of high-entropy radiation. Slowly lower the box toward the black hole by means of a rope. As the box is lowered, the tension in the rope far away from the hole can be used to perform work. In an ideal case, the energy extracted from the box goes as

$$W = E \cdot \left(1 - \sqrt{1 - \frac{2M}{R}}\right), \tag{7.4}$$

where R is the radius away from the black hole to which the box has been lowered, M is the mass of the black hole, and E is the energy in the box. In principle, it would be possible to lower the box arbitrarily close to the black hole. It would thus be possible

to extract arbitrarily much of the energy of the radiation in the box in the form of useful work by lowering the box very very near to the horizon. (The requirements on the rope that it not break and on the box that it not to crack open are severe, but since this is an 'in-principle' argument, they will be ignored.) Now, once the box is very near the black hole horizon, open it and allow the radiation to fall into the black hole. The entropy of the radiation within the box is thus lost to the outside world. However, the black hole's entropy increases by an amount proportional to the energy which falls into the black hole, i.e. the energy of the radiation which has not been extracted as useful work. Since that remaining energy can be made as small as desired, the increase in entropy of the black hole is as small as desired, and in particular it can be made smaller than the entropy lost down the black hole when the radiation fell in. Thus, the second law could be violated.

Geroch's argument was basically irrelevant at the time, since black holes could not engage in a cyclic processes; the black hole always grew bigger. Also, since classical black holes had no temperature, they could be regarded as zero temperature heat baths, and it is well known that with a zero temperature heat bath, one can convert all of the heat energy in any system to useful work.

However, with the advent of Hawking's discovery of black hole temperatures, this sanguine dismissal of the problem failed. Black holes could participate in closed cycles within a heat engine. You just needed to wait until the black hole had radiated the tiny bit of energy you had deposited into it. Black holes had a finite temperature, and the laws of thermodynamics dictated a maximum efficiency for the conversion of energy to work. Geroch's argument thus became critical. Could this process be used (in principle) to violate the second law of thermodynamics?

Bekenstein (1981) suggested that the reason that the Geroch argument failed was because the assumption that one could extract an arbitrary amount of work by lowering the box arbitrarily near the black hole was wrong. Boxes have finite sizes. One cannot lower a box closer to the black hole than the vertical dimension of the box. He postulated that there existed another law of nature, namely that the ratio of energy to entropy of the box was limited by the dimension of the box, i.e. that for any system,

$$S/E < \frac{H}{2\pi}, \tag{7.5}$$

where H is the minimum dimension of the box. Since the box cannot be lowered to closer than H from the black hole, we have a limit on the maximum amount of work which can be extracted during the lowering of the box. Since S is also limited by that same H we have a limit on the maximum amount of entropy that can be deposited into the black hole. The factor of 2π is chosen to ensure that this maximum deposited entropy is less than the increase in area due to the residual unextracted energy left in the box. It is astonishing that the mass of the black hole does not enter into this equation. Bekenstein has examined a number of systems, and has found that most (if not all realistic) in fact obey a bound very similar to this one.

However, his argument ignores the fact that the radiation would tend to gather at the bottom of the box in the arbitrarily large gravitational field felt by the box when near the horizon, i.e. the radiation would be nearer the horizon than the

dimension of the box when it was released. Furthermore, I was always disturbed by this line of argument. It implies that one must posit a new law of nature (the entropy to energy bound) in order to rescue the second law of thermodynamics for black holes. Furthermore, it would seem that there are cases in which one could imagine this law to be violated. For example, the entropy of a system with given energy contained within a box, but with N different species of massless particles, is that of a single species times $\ln(N)$. By making the number of species of massless particles sufficiently large, one could thus violate the bound. The fact that the real world does not have this sort of a proliferation of massless particles is surely irrelevant to the validity of the second law. It seems unlikely that black hole thermodynamics could be used to place a bound on the number of species of particles.

The framework for an answer came one day at a Texas Symposium in a lunch lineup, when Kip Thorne and Freeman Dyson posed the Geroch argument to me. After I had told them of my objections to Bekenstein's argument, they asked what a correct answer would be. The glimmering of an answer which came to me then was developed into a complete argument by Bob Wald and myself (Unruh and Wald 1982, 1983) the following summer. At the time of Hawking's work, I had independently discovered (Unruh 1976) a closely related feature of quantum fields in flat spacetime (i.e. in the absence of any gravitational field). I had been worried about how one could define particles in a quantum field theory. Fields are not particles, and the particle aspects of quantum fields arise as secondary features of those fields in their interactions with other matter. Due to the quantization of the energy levels of localized matter, the interaction of that matter with the fields takes place by the absorption and emission of discrete quanta of energy. It is that process which makes the field behave as though its excitations were composed of particles.

The question I posed to myself was how one could interpret the excitations of quantum fields in terms of particles in the presence of strong gravitational fields. The principle I came up with was that particles should be defined to be that which particle detectors detect. Although this definition of particles seems to be a tautology, it focusses attention on a productive aspect of the problem. It is easy to design a system (at least in principle) which will detect particles when the state of the field is such that we usually consider it to have particles. One can now take that same detector and ask how it will respond in novel situations. One can then use that response to define what one means by particles in that novel situation. If it reacts in the same way that it would in detecting particles in well-understood situations, we can say that the field acts as though it were composed of particles in that novel situation.

The novel situation I examined was to accelerate a model of a particle detector in the vacuum state of the field. Now, an unaccelerated detector in that same state will see nothing. The detector will not click: it will respond as though there are no particles present. However, I found that if I examined an accelerated detector with the field left in exactly that same state, it would respond exactly as though that state of the field were populated by an isotropic thermal bath of particles. If the detector is left in contact with that state of the field (and continues in its state of constant acceleration), it will come into thermal equilibrium with a temperature just proportional to the acceleration of the detector. That is, an accelerated object immersed into the vacuum state of the field behaves as though it was in a thermal

bath of particles of that field, with the temperature equal to

$$T = \frac{a}{2\pi}\frac{\hbar}{ck}, \tag{7.6}$$

where I have in this case reintroduced Planck's constant, the velocity of light, and Boltzmann's constant. This temperature is such that it requires an acceleration of 10^{24} cm/s^2 to achieve a temperature of 1 K.

This phenomenon of acceleration radiation can be used to explain why the Geroch argument does not work. As one lowers the box full of radiation toward the black hole, one requires a tension in the rope to keep the box from following a geodesic (straight line) and falling into the black hole: the box must be accelerated to hold it outside the black hole. This accelerated box thus sees the field outside the box as though it is in a thermal state with a temperature proportional to the acceleration. In addition, because it is accelerated, the box also sees itself as being in a gravitational field. Finally, in order that the box hold the radiation inside the box, it must also exclude the radiation outside. Thus, we have exactly the situation analyzed by Archimedes in a preprint over 2000 years ago.[2] A body immersed in a fluid will suffer a buoyant force equal to the weight of the displaced fluid. Thus, the box, immersed in the fluid of that thermal radiation and in the apparent gravitational field due to its own acceleration, will feel a buoyant force proportional to the weight (in that apparent gravitational field) of the displaced thermal fluid. Therefore, the tension in the rope holding the box from falling into the black hole is less than it would be if that thermal fluid were absent. Since the tension in the rope is less, the energy extracted by the work done by that tension far away from the black hole as the box is lowered is also less. In fact, the maximum amount of work extracted occurs not when the box has been lowered to the horizon (where the acceleration is infinite and the resultant buoyant force is thus infinite), but when it has been lowered to the 'floating point' where the weight of the displaced thermal fluid is just equal to the weight of the radiation inside the box.

Our calculation showed that the energy remaining to fall down the black hole (i.e. the total original energy in the box minus the work extracted by lowering the box) was precisely what was needed to increase the mass of the black hole, and thus increase its entropy to a value larger than or equal to the entropy contained in the box. That is, the entropy falling into the black hole in the form of radiation was always less than the entropy increase of the black hole due to its increase in surface area. The second law of thermodynamics was valid, even if one included black holes in the operation of the heat engine. This argument – unlike the Bekenstein argument – is universally valid, no matter how many species of field one imagines, since the buoyant force also increases if the box encloses, and thus excludes, more species of field. Black holes are truly thermodynamic objects: they obey the second law of thermodynamics.

7.2 A problem in the derivation of black hole evaporation

Despite this beautiful constancy of the black hole thermodynamics, it was clear from the beginning that the derivation that Hawking gave for the thermal emission from

black holes was flawed at a very fundamental level. The derivation took the background gravitational field of a black hole as a given. On this background spacetime, one looked at the evolution of quantized fields. The first calculations were for scalar fields, and that is what I will use in this chapter, but the calculations for other fields quickly followed and gave exactly the same results.

Let us look at how Hawking actually carried out his calculation. While this may seem to be somewhat technical, it is important to understand why I made the statement above that his derivation was clearly wrong (even though, as I will argue, the results are almost certainly correct). A field (a concept first introduced, with much controversy, into physics by Faraday) is something which exists at all places in space at all times. It is characterized, not like particles by positions, but rather by the whole set of values that the field takes at each point in space and at each instant in time. Those values at different places and times are related to each other via the equations of motion that the field obeys. The field thus does not move or travel, but one can talk about excitations – places where the field has non-zero values, or certain patterns in the values of that the field takes at different points – travelling. The field, whether classical or quantum, is a deterministic system, in which the field values now are determined by the values at an earlier time. Since the disturbances in the field travel, the state of those disturbances is determined by the state of the disturbances at an earlier time in different places. We are interested in understanding where the radiation given off by the black hole came from. To do so, we can propagate those disturbances back in time to see what aspect of the early state of the field must have caused the disturbance we are interested in. In particular, disturbances which are now, at late times, coming out of the black hole must have been caused by disturbances which came in toward the black hole at a much earlier time.

Although the above is a classical description, in which the magnitude of the fields at each point has some definite value, the same is true of a quantum field where the amplitude of the field is a quantum operator with the set of all such operators being in a certain quantum state. The state of the quantum field can be divided into in-going and out-going parts, the state of the out-going parts having been determined by the state of the in-going components far in the past. The gravitational field for a black hole is such that the spacetime becomes flat far from the black hole, and one can thus use the usual interpretation of the behaviour of quantized fields in a flat spacetime, as long as one concentrated on the behaviour of the fields far from the black hole. The in-going part of the fields were assumed to be in the vacuum state, the state of lowest energy, or the state in which, under the conventional picture, there exist no particles corresponding to that field. One then uses the field equations to propagate those fields into the region of the black hole and then back out to become out-going components of the field in a region far from the black hole again. Far from the black hole those out-going fields could again be analysed in the conventional way and the particle content of the resultant out-going state could be determined.

The result was that if one began with the in-going components of the field in its vacuum state and with some (non-quantized) matter collapsing together to form a black hole, then very quickly (on a time scale of the light travel time across a distance corresponding to the circumference of the black hole) the out-going portions of the quantum field would settle down to a state such that the out-going field looked

like a thermal state with a temperature of $1/8\pi M$ (for a spherically symmetric black hole).

However, one could, for any particle emitted out of the region of the black hole (or for any mode of the field), trace back, by use of the equations of motion of the field, the disturbances which corresponded to that particle to the in-going aspects of the field which must have created that particle. It was in this determination that one obtains nonsensical answers. Consider a particle which was emitted from the black hole at a time $t - t_0$ after the formation of the black hole at time t_0. That particle would have a typical energy of order of $1/M$ and a wavelength of order M, the radius of the black hole. However, the initial fluctuations in the field which must have caused this particle would have had an energy of order $(1/M)e^{(t-t_0)/M}$ and a wavelength of order $Me^{-(t-t_0)/M}$. Let us consider a solar mass black hole, for which the typical wavelength of the emitted particles would be a few kilometres, and typical frequency of the order of kilohertz, and consider a particle emitted one second after the black hole had formed. This emitted excitation of the field must have had its origin in the incoming fluctuations in the field with a frequency of e^{10^4} and wavelength e^{-10^4} (units are irrelevant since any known system of units would make only a negligible change in that exponent of 10^4). It is clear that at these scales the very assumptions which went into the derivation are nonsense. At these frequencies, a single quantum of the field has an energy not only larger than that the energy of the black hole, but also inconceivably larger than the energy of the whole universe. At these scales, the assumption that the quantum field is simply a small perturbation, which does not affect the gravitational field of the black hole, is clearly wrong: Hawking's calculation contains its own destruction. If it is correct, all of the assumptions which were necessary to make his derivation are clearly wrong. But it is equally clear that his results are far too appealing to be wrong. But how can the derivation be nonsense, and the results still be correct? This is a puzzle which has bothered me ever since Hawking's original paper. The problem of course is that this derivation (or essentially equivalent ones, like certain analyticity arguments) were the only evidence for the thermal nature of black holes. Black holes have never been seen to evaporate, nor are we likely to get any such experimental evidence for the validity of this result.

Thus, the picture of the origin of black hole radiation painted by the standard derivation is that we have a mode of the field with such an absurdly high energy. This mode is in its lowest energy state, the vacuum state and is travelling toward the region in which the black hole will form. It enters the collapsing matter just before it forms a black hole. It crosses through the matter, and emerges from the matter just (exponentially 'just' – e^{-10^4} s or less) before the horizon forms. It is trapped against the horizon of the future black hole, exponentially red-shifting, and being torn apart by the horizon, until finally its frequency and wavelength are of the characteristic scale of the black hole. At this point part of this mode escapes from the black hole, leaving part behind to fall into the black hole. The part that escapes is no longer in its ground, or lowest energy state. Instead it is excited and behaves just like an excitation which we would associate with a particle coming out of the black hole. Because of the correlations that existed between the various parts of this mode in the initial state,

there are correlations between the part of the mode which escapes from the black hole, and the part left behind inside the black hole. These correlations ensure that the excitation of the field coming out of the black hole is precisely thermal in nature.

The natural question to ask is whether the predication of thermal radiation really does depend on this exponential red-shifting, or whether this apparent dependence is only a result of the approximations made in the derivation. In particular, could an alteration in the theory at high frequencies and energies destroy the prediction of thermal evaporation? That is, is the prediction of thermal radiation from black holes robust against changes to the theory at those high energies which we know must occur? There seemed to be nothing in that original derivation which would indicate that the prediction was robust, except for the beauty of the result. (Of course physics is not supposed to use the appeal of a result as an argument for the truth of that result.)

7.3 **Dumb holes**

Is there some way of testing the idea that the truth of the thermal emission from black holes is independent of the obvious absurdity of assuming that the quantum field theory was valid at arbitrarily high frequencies and short distances? Although any ultimate answer must depend on our being able to solve for the emission of radiation in some complete quantum theory of gravity, one can get clues from some analogous system. One such analogue is what I have called dumb holes, the sonic equivalent to black holes (Unruh 1980).

Let me introduce the analogy by means of a story. Consider a world in which fish swimming in an ocean have become physicists. These fish experience their world, not through sight (they are blind) but through their ears: through sound alone. In this world there exists a particularly virulent waterfall, in which the water, as it runs over the falls, achieves a velocity greater than the velocity of sound. This waterfall will act as an attractor to fish and to sound. Sound waves travelling by this waterfall, will be pulled toward the waterfall, just as light passing by a star is pulled toward the star and bent. Fish swimming near this waterfall will be attracted to it. However, as long as a fish swims outside the boundary surface at which the velocity of water equals the velocity of sound, its voice and the shouts it makes will travel out to any other fish also swimming outside that surface. However, the closer the fish comes to the surface, the longer it will take the sound to get out, since its velocity out is partially cancelled by the flow of the water over the waterfall; the sound waves are swept over the falls along with anything else. Now consider a fish which falls over those falls and calls for help as it does so. The closer to the surface where the velocity of the water equals the velocity of sound, the slower the effective speed of sound getting out is, and the longer the sound takes to get out. A sound wave emitted just at the brink of that surface will take a very long time to escape. Because the sound takes longer and longer to get out, the frequency of the sound emitted by that fish will be bass-shifted to lower and lower frequencies. Any sound emitted after the fish passed through the surface will never get out, and will be swept onto the rocks below along with the fish itself (for a picture and a description of this process, see Susskind 1997).

This description is very similar to that of light around a black hole. The surface where the water speed exceeds the velocity of sound is an analogue of the horizon

of a black hole. The bass-shifting of the sound emitted by a fish falling through that surface is analogous to the red-shifting of the light emitted by an object falling through the horizon of a black hole. These dumb holes (holes unable to speak or to emit sound) are thus an analogue to black holes and, since the physics of fluid flow is believed to be well understood, an analogue which one could hope to understand in a way that black holes are not yet understood.

These physicist fish will develop a theory of their surroundings, and in particular will develop a theory of sound waves very similar to our theory of light within General Relativity, Einstein's theory of gravity. This theory will furthermore have in it structures, these waterfalls, which share many of their features with black holes in our world.

It is important to point out that there are features of these dumb holes which are not analogous. If a fish stays stationary outside the waterfall, the frequencies of its sound are not bass-shifted, while the light from such a stationary source is red-shifted. The flow of time is not altered by the presence of the waterfall as is the flow of time in the gravitational field of the black hole. The size of the waterfall, of the sonic horizon, is not a function of how much matter or energy has fallen over the falls, while the black hole size is a direct measure of the amount of energy which has fallen into the black hole. Despite these differences however, the dumb holes still turn out to be ideal subjects for understanding the quantum behaviour of black holes.

Let us consider sound waves in slightly more detail. They are small perturbations in the density and velocity of the background flow of the fluid. As is well known, if one examines only the lowest-order perturbations around a background fluid flow, perturbations which represent sound waves in that fluid, then those perturbations obey a second-order linear partial differential equation. If we furthermore assume inviscid flow, so that the perturbations are conservative (do not lose energy due to dissipation) then the equations are homogeneous second-order equations in space and time derivatives. This can be cast into the form

$$\partial_\mu a^{\mu\nu}\, \partial_\nu \Phi = 0, \tag{7.7}$$

where the tensor $a^{\mu\nu}$ depends on the background fluid flow, and Φ is some scalar such as the density perturbation or the velocity potential. But by defining

$$g = \det(a^{\mu\nu}), \tag{7.8}$$

$$g^{\mu\nu} = \frac{1}{\sqrt{g}} a^{\mu\nu}, \tag{7.9}$$

this is cast into the form of the perturbation of a scalar field Φ on a background metric $g_{\mu\nu}$: the equations of motion for the sound waves on a background fluid flow obey exactly the same equations as a massless field on a background spacetime metric. Furthermore, if the fluid flow has a surface on which it becomes supersonic, the metric associated with that fluid flow is the metric of a black hole with that supersonic surface as the horizon of the black hole.

Those small perturbations of sound waves in the fluid flow can be quantized just as can the scalar field on a background metric (quantized sound waves are often

called phonons). Furthermore, one can use Hawking's argument to conclude that such a dumb hole should emit radiation in the form of phonons in exactly the same way as a black hole will emit thermal radiation. The fish physicists would come up with a theory that their dumb, sonic, holes would emit thermal radiation, just as black holes do.

Of course in the case of dumb holes, there is no analogue of the mass or energy of the dumb hole. The structure of the sonic horizon is just a consequence of the peculiarities of the flow set-up, and not of the amount of energy that has gone down the dumb hole. There is no analogy to the energy of a black hole, and the relationship between that energy and the temperature that occurs with a black hole does not occur for a dumb hole. Dumb holes would not be thermodynamic objects as black holes are. They are however hot objects. The temperature of dumb holes is an exact analogue of the temperature of black holes.

Because of the pivotal role that the temperature plays in black hole thermodynamics, the existence of a dumb hole temperature makes them a sufficient analogue of black holes to, one hopes, give us clues as to the origin of the radiation from black holes. In particular, in the case of black holes we have no idea as to the form that a correct theory at high energies would have, or at least no idea as to how we could calculate the effects that the alterations of any such theory at high frequencies would have on the thermal emission process. However, for dumb holes we have a good idea as to the correct theory.

Real fluids are made of atoms. We know that once the wavelengths of the sound waves are of order the interatomic spacing then the fluid picture becomes inapplicable. Of course, calculating the behaviour of 10^{23} atoms in a fluid is as far out of our reach as is calculating the effects of any putative theory of quantum gravity on the emission process around black holes. This fact kept me from being able to use dumb holes to understand black hole thermal emission for over ten years. However, Jacobson (1991) realized that there are approximations that can be made. One of the key effects of the atomic nature of matter on the behaviour of sound waves is to alter the dispersion relation (the relation between wavelength and frequency) of the sound waves. While sound waves at lower frequencies have a direct proportionality between wavelength and the inverse frequency, just given by a fixed velocity of sound, at higher frequencies this relation can become much more complex. The effective velocity of sound at high frequencies can increase or decrease from its value at low frequencies. Thus, the location of the sonic horizon, which is determined by where the background flow equals the velocity of sound, can be different at different frequencies of the sound waves.

This allows one to ask, and answer, the more limited question: 'If we alter the dispersion relation of the sound waves at high frequencies, does this alter the prediction of thermal radiation at low frequencies?' The calculation proceeds in precisely the same way as it does for a black hole. Consider a wave packet travelling away from the dumb hole in the future, and at a time when it is far from the dumb hole. In order to represent a real particle, we choose that wave packet so that it is made purely of waves with positive frequencies (i.e. their Fourier transform in time at any position is such that only positive and no negative frequencies non-zero). We now propagate that wave backward in time, to see what configuration of the field

would have created that wave packet in the future. Going backward in time, that wave packet approaches closer and closer to the (sonic) horizon of the dumb hole. In order to have come from infinity in the past, and if the velocity of sound does not change with frequency, that wave must remain outside the horizon (nothing from inside can escape). But in this propagation backward in time, the wave is squeezed up closer and closer to the horizon, in a process which grows exponentially with time. It is only when we get to the time at which the sonic hole first formed that those waves can escape and propagate (backwards in time) back out to infinity. By that time that exponential squeezing against the horizon has produced a wave with with an incredibly short wavelength, and thus incredibly high frequency. One can now analyse that resultant wave packet to see what its frequency components are. In general they are very high. However, now one can find that that packet which began with purely positive-frequency components, also has negative-frequency components. Those negative-frequency components are a direct measure of the number of particles which were produced in that mode by the process of passing by the dumb hole. In particular, Hawking showed for black holes (and the same follows for dumb holes) that the ratio of amplitudes for those negative- and positive-frequency components is given by $e^{\omega/T}$, where T is the Hawking Temperature of the black hole, proportional to the inverse mass of the black hole, or, in our case, is the temperature of the dumb hole, proportional to the rate of change of velocity of the fluid as it passes through the dumb hole horizon.

What happens if we change the dispersion relation at high frequencies for the fluid? In this case, squeezing against the horizon as one propagates the wave packet back in time no longer occurs to the same extent. As the packet is squeezed, its frequency goes up and the velocity of sound for that packet changes. What was the horizon – the place where the fluid flow equals the velocity of sound – is no longer the horizon for these high-frequency waves. They no longer get squeezed against the horizon, but rather can once again travel freely and leave the vicinity of that low-frequency horizon long 'before' (remember we are actually going backward in time in this description so 'before' means after in the conventional sense) we get back to the time at which the dumb hole formed. Again, we can wait until this wave propagates back to a regime in which the velocity of the water is constant, and we can take the Fourier transform of these waves. Again, we can measure the ratio of the positive and negative frequencies in the incoming wave, and determine the particle creation rate. The astonishing answer I found (Unruh 1995) (and which has been amply confirmed by for example Jacobson and Corley (1996) in a wide variety of situations) is that this ratio is again, to a very good approximation, a thermal factor, with precisely the same temperature as one obtained in the naive case. The history of that packet is entirely different (for example, it never came near the time of formation of the dumb hole, as it did in the naive analysis), and yet the number of particles produced by the interaction with the dumb hole is the same: i.e. such dumb holes produce thermal radiation despite the drastic alteration of the behaviour of the fields at high frequencies. This has been calculated for a wide variety of situations, both where the high-frequency velocity of sound decreases over its low-frequency value, and where it increases sufficiently that there is no longer any horizon at high frequencies. The prediction of thermal radiation from a dumb hole

appears to be remarkably robust and independent of the high-frequency nature of the theory.

How can this be understood? The calculation, done for example by Hawking, used exponential red-shifting of the radiation in an essential way. The ultra high-frequency aspects of the theory appeared to be crucial to the argument. Yet we find here that it is not crucial at all. All that appears to be crucial is the the low-frequency behaviour of the theory and the nature of the horizon at low frequencies.

Although the best way of understanding the particle production process by a black hole or a dumb hole has not yet been found, there are at least hints of a possible answer. If we examine the history of such a wave packet from the past to the future, we find that the packet, while near the horizon, suffers an exponential red-shifting. However, the time scale of the red-shifting is of the order of the inverse temperature associated with the hole, $\hbar/kT$. When the packet has a very high frequency, that time scale is very long compared with the inverse frequency of the packet. The packet sees the effects of this red-shifting as occurring very slowly – adiabatically. Now, we know that a quantum system undergoing adiabatic change remains in the state it started in. If the packet started in its ground state, it remains in its ground state during this adiabatic phase. If the packet began in the vacuum, it remains in the vacuum. Thus, during most of the red-shifting that occurs near the horizon, or during the approach of the packet to the horizon, the packet does not notice the fact that something is happening around it. It gets stretched (red-shifted), but so slowly that it remains in its vacuum state (zero particle state). However, eventually it is stretched sufficiently that its frequency is now low enough that the time scale of red-shifting is of the same order as its inverse frequency. The surroundings now change rapidly over the timescale of oscillation of the packet, and it begins to change its state. It is excited from its ground state to a many particle state. Furthermore, because of the correlations across the horizon at these low frequencies, the particles which are emitted out toward infinity are in an incoherent, thermal, state.

Appealing as this scenario is, it still needs to be fleshed out via compelling calculations. However, even at this point it strongly supports one conclusion, namely that the thermal emission process of a black or dumb hole is a low-frequency, low-energy, long-wavelength phenomenon. Despite the apparent importance of high frequencies in the naive calculation, it is only aspects of the theory on scales of the same order as those set by the temperature of the hole (e.g. the inverse mass for black holes) which are important to the emission of thermal radiation. The thermal emission is a low-energy, low-frequency, long-wavelength process, and is insensitive to any short-scale, high-energy or high-frequency physics. This conclusion will play an important role in the following.

7.4 **Entropy and the 'information paradox'**

Having discussed the temperature of black holes, and come to the conclusion that the thermal radiation is created at low frequencies and long wavelengths outside the horizon, let us now return to the issue of the entropy of the black hole. All of our arguments about black holes in the above have been thermodynamic arguments. Black holes have certain macroscopic attributes of mass (energy) and temperature.

This is not the place to enter into a detailed discussion of what entropy really is, but it is important to state the usual relationship of entropy to the internal states of the system under consideration. For all material objects, the work of Boltzmann, Maxwell, Gibbs, etc., showed that this thermodynamic quantity of a system could be related to a counting of the number of microscopic states of a system under the macroscopic constraints imposed by the mode of operation of the heat engine which wished to use the system as part of its 'working fluid'.

This classical analysis of the nature of entropy is closely coupled to the notion of the deterministic evolution of the system. In the evolution of a system, unique initial states evolve to unique final states for the 'universe' as a whole. In quantum physics, this is the unitary evolution of the state of the system: orthogonal initial states must evolve to orthogonal final states. The number of distinct initial states must be the same as the number of distinct intermediate states, and must equal the number of final states. These states may (and will in general) differ in the extent to which a heat engine can differentiate between them. But on a fundamental level the extent to which each state is different from each other state is conserved throughout the evolutions. The law of the increase of entropy then arises when one considers a highly differentiated initial state (as far as the heat engine is concerned). In general this will evolve to a much less differentiated final state. But in principle, if the heat engine were sufficiently accurate in its ability to differentiate between the various states of the system, all systems would evolve with a constant, zero entropy.

However, black holes, as classical objects, have a small number of parameters describing the exterior spacetime (the spacetime in which a heat engine can operate), namely the mass, angular momentum, and the variety of long-range charges which a black hole can support. All aspects of the matter that falls into the black hole fall into the singularity at its centre, taking the complexity of the structure and state of that matter with them. The horizon presents a one-way membrane, impervious to any prying hands, no matter how sophisticated. Arbitrarily complicated initial states all evolve, if the matter falls into the black hole, into simple final states of a black hole.

Does the situation change in a quantum theory of black holes? One difficulty is that no such theory exists, making the question difficult to answer definitively, but one can at least try to understand why the question has proven as difficult to answer as it has.

Ever since black hole thermodynamics was discovered in the 1970s, attempts have been made to try to understand black hole entropy in a statistical mechanical sense. At the semiclassical level, black holes evaporate giving off maximal entropy thermal radiation, no matter how the black hole was formed. They can however be formed in a wide variety of ways. Is the emission from black holes truly thermal, and independent of the formation of the black hole, or are there subtleties in the outgoing radiation which carry off the information as to how the black hole was formed? If one forms a black hole in two distinct (orthogonal) ways, are the final states, after the black hole in each case has evaporated, also distinct? The fact that the radiation in the intermediate times looks thermal is no barrier to this being true. A hot lump of coal looks superficially as though it is emitting incoherent thermal radiation. However, we strongly believe that the final states of the radiation field would be distinct if the initial states which heated up the lump of coal had been distinct.

How does the situation proceed for that lump of coal? If the coal were heated by a pure, zero entropy, state initially, how does it come about that finally the radiation field again has zero (fine-grained) entropy despite the fact that at intermediate times the coal was emitting apparently thermal incoherent radiation? The answer is that the coal has an internal memory, in the various states which the constituents of the coal can assume. When the coal is originally heated by the incoming radiation, the precise state in which those internal degrees of freedom of the coal are left is completely and uniquely determined by the state of the heating radiation. The coal will then emit radiation, in an incoherent fashion: i.e. the radiation emitted will be probabilistically scattered over a huge variety of ways in which it could be emitted. However, each way of emitting the radiation will leave the internal state of the coal in a different, completely correlated state. The coal remembers both how it was heated, and exactly how it has cooled since that time. Thus as the coal cools, its state will help determine how the radiation at late times is given off, and that radiation will then be correlated with what was given off at early times. The state of the emitted radiation, after the coal has cooled, will look – if looked at in a limited region or a limited time – as though it were completely incoherent and thermal. However, there will exist subtle correlations between the radiation in the various regions of space and at various times. It is those correlations which will make the final state a unique function of the initial state of the heating radiation. And the possibility of that final state being uniquely determined is predicated on that coal's internal state being uniquely determined by the initial state and the state of the radiation emitted up to the point of time in question. It is entirely predicated on the coal's having a memory (those possible internal states) and on it remembering how it was formed and how it radiated.

If black holes also behave this way then they also must have a memory, and retain a memory of how they were formed. Furthermore, that memory must affect the nature and state of the radiation which is subsequently given off by the black hole. The radiation given off by the black hole at any time must depend on both the state of the matter which originally formed the black hole, and the state of the radiation which has already been given off by the black hole.

Of course the alternative is simply to accept that black holes are different from other bodies, in that the out-going radiation is unaffected either by how the hole was formed or by how it has decayed since that time. Black holes would then differ fundamentally from all other forms of matter. The final incoherent state of the radiation emitted by a black hole would not be determined by the initial state of the matter that formed the black hole. All initial states would produce the same final, incoherent, mixed state. Of course this is of no practical importance, since those subtle correlations in the emitted radiation responsible for the maintenance of unitarity from even a small lump of coal are entirely unmeasurable.

In the 1980s Banks, Susskind, and Peskin (1984) presented a *reductio ad absurdum* argument that black holes must have a memory, and must preserve unitary evolution. Their argument was that if the black hole really was a memoryless system, then the out-going radiation would be what is technically known as Markovian. They then showed that such a Markovian evaporation process would lead to energy non-conservation, and furthermore that the energy non-conservation would be extreme

for small black holes. For black holes of the order of the Plank mass (10^{-5} g) one would expect radiation to be emitted in quanta whose energy was also of the order of that same Plank mass. Furthermore, one might expect virtual quantum processes to create such Plank mass black holes which would then evaporate creating a severe energy non-conservation in which one might expect energies of that order to be liberated. Since that is of the order of a ton of TNT, one could argue that such non-conservation has been experimentally ruled out, as physicists all might have noticed the presence of an extra ton of TNT in energy liberated in their labs (or indeed even non-physicists might have noticed this). Therefore, they concluded that black holes could not be memoryless: that they must remember how they were formed. However this argument assumed that black holes remembered nothing, not even how much energy went into their formation and evaporation. But black holes clearly have a locus for the memory of the energy, angular momentum, and charge of their formation, namely in the gravitational and other long-range fields which can surround a black hole (Unruh and Wald 1995). There is thus no reason in principle or in experiment why black holes could not be objects very different from other matter, lacking all memory of their formation (apart from those few bits of memory stored outside the horizon).

However, many people have found this conclusion too radical to countenance. While I do not share their horror of black holes being afflicted by this ultimate in Alzheimer's disease, it is certainly important to examine the alternatives.[3]

Various suggestions have been made through the years for a statistical mechanical origin for the black hole entropy. One suggestion has been that the entropy of a black hole is equal to the logarithm of the number of distinct ways that the black hole could have been formed (Thorne, Zurek, and Price 1986). However, this clearly does not identify any locus of memory for the black hole. It is, in fact, simply a statement that the black hole obeys the second law of thermodynamics. If the black hole had many more ways of being formed than the exponential of its entropy, then one could prepare a state which was an incoherent sum of all of the possible ways of forming the black hole. The entropy of this state would just be the logarithm of that number of states. Once the black hole had formed and evaporated, then the entropy of the resulting matter radiation would just be the entropy of the black hole. Thus the entropy of matter, under this process, would have decreased, leading to a violation of the second law of thermodynamics. Similarly, if the entropy of the black hole were larger than the logarithm of the number of all the possible states which went into its formation, the entropy of the world would increase under the evaporation process. However, one of the ways of increasing the size of a black hole would simply be to reverse the radiation which was emitted back to the black hole: the time reverse of the emitted radiation is one of the ways of creating the black hole. Thus the entropy of the emitted radiation also forms a lower bound on the logarithm of the number of ways of creating the black hole. Thus the statement that the number of ways of forming a black hole gives a measure of its entropy is simply a thermodynamic consistency condition, and gives no clues as to the form of the memory (internal states) of a black hole.

Within the past four years, however, string theorists (originally Strominger and Vaffa 1996) have suggested that they may have found a statistical origin for the

black hole entropy, a description of those internal states which would form the black hole's memory. These suggestions are as yet only that – suggestions – but they carry a powerful conviction and hope that they may offer a way of describing black hole entropy. This hope has been adequately broadcast, so what I wish to do in the remainder of this chapter is to raise some potential difficulties which may beset this explanation.

Let me give a cartoon review of the string theory argument. At long distances, and low energy, one can regard string theory as leading to an effective field theory, with the various excitations of the strings corresponding to various types of fields. That low-energy field theory includes gravity, and has, in certain cases, black holes as solutions. In the case where one demands that these solutions to the low-energy field theory retain some of the super-symmetries of the string theory, the fields must be chosen carefully so as to obtain a black hole solution, rather than some singular solution of the field equations. These super-symmetric black holes have, in addition to their gravitational field, a combination of other massless fields. These other fields, like the electromagnetic field, imply that the black hole must also carry certain charges. The mass of these black holes is furthermore a given function of those charges. It is not an independent parameter, as it would be for a generic black hole.

If one wishes to create a model for these black holes within string theory, one requires that the model have sources for those fields, that the configuration of the strings carries those charges. For some of the fields, such sources were no problem. They arise in the perturbative form of string theory. However for other fields, the so-called 'RR fields', there was nothing within standard perturbative theory which carried these RR charges. Polchinski introduced additional structures, called 'D-Branes', into string theory precisely to carry those RR charges: to be sources for the associated fields. D-Branes are higher dimensional surfaces on which open strings could have their endpoints.[4] It is assumed that they will arise in string theory from non-perturbative effects. However, if super-symmetric black holes were to have a string theory analogue, that analogue would have to contain not only the standard strings, but also D-Branes to act as the origin for the charges carried by the black hole.

D-Branes are extended surfaces without edges. In order that the black hole be a localized object, it is assumed that our ordinary four dimensions (three space and one time) are all orthogonal to these D-Brane surfaces. In order that the extra dimensions demanded of string theory not embarrass us by not having been observed, these extra dimensions (in which the D-Branes stretch) are assumed to be 'curled up' so that travelling a (very small) distance along one of these extra dimensions always brings one back to the origin. Thus to us, these D-Branes would look as though they were located at a point (or at least a very small region) of our observable three dimensions of space.

In a certain limit, one could carry out a calculation of how such an analogue structure would behave. If one adjusted one of the parameters in the string theory – the so-called 'string coupling constant' – so that the couplings between strings were turned off, then those D-Branes and their attached strings would constitute a free gas of particles. The D-Branes would be held together by a coherent collection of strings which connected them to each other, preserving the localized nature of this black hole analogue.

Now one could ask in how many ways the D-Branes be could distributed in order to produce the same macroscopic total charges. And in how many ways could the open strings connect the various D-Branes together and to each other? In the weak coupling limit, since the D-Branes and their attached open strings have a perturbative description, the number of states of the system could be calculated by solving the low-energy super-gravity conformal field theory to count the degeneracy of these states, and thus their entropy. The answer was that the entropy, as a function of the charges, was precisely the same as the entropy one would have for a super-symmetric, maximal black hole with those same charges.

Now the string theory analogue is *not* a black hole. The only way in which the calculations can be carried out is by assuming that the gravitational constant (proportional to the string coupling constant) is essentially zero. In that case the string analogue would be spread out over a region much larger than the radius of the horizon of the black hole which those charges would form. The string theory calculation is a calculation in which everything is embedded in a flat spacetime.

As a thermodynamic object, a super-symmetric black hole is anomalous. Although it can be said to have both entropy and energy, its temperature is identically zero. It does not evaporate. However, one (Maldecena and Strominger 1997) can also carry out calculations for the D-Brane model if one goes very slightly away from the exactly super-symmetric model. In this case both the black hole and the D-Brane models radiate. The spectrum of the radiation, assuming that the D-Brane is in thermal equilibrium, is again identical in both cases, and in neither case is it thermal. In the case of the black hole, this non-thermality is caused by the non-zero albedo of the black hole. If the wavelength of the radiation is much longer than the diameter of the black hole, most of the radiation scatters from the curvature around the black hole, and is not absorbed. By detailed balance, the emitted radiation is suppressed for exactly those components which would have been scattered if they were incoming radiation. In the D-Brane calculation, the emitted radiation is caused by the collision of open strings of differing temperatures on the D-Branes. The astonishing result is that both calculations give the same, very non-trivial result.

Both of these results, the entropy and the low-temperature emission, make one feel that surely string theory is giving us an insight into the behaviour of black holes. In the case of D-Branes, the entropy has its origin in the standard way, as a measure of the number of ways that the internal configuration (fluctuations of the open strings and configurations of the D-Branes) of the system can vary, constrained by the macroscopic parameters (the charges). The entropy of the D-Branes is an entropy of the same type as the entropy of a lump of coal. The D-Brane analogue has a locus for its memory, that locus being the various configurations of D-Branes and strings which make up the object.

However, it is again important to emphasize that these D-Brane bound states are just analogues to black holes, they are not themselves black holes. As one increases the gravitational constant to a value commensurate with the structure forming a black hole, perturbative calculations in string theory become impossible.[5] Thus the crucial question becomes 'does this D-Brane analogue give us an insight into the entropy of black holes?' Or is the equality of the entropy calculations – one a thermodynamic one from the temperature of the radiation given off by the black hole, and the other

a statistical one from the degeneracy of D-Brane bound states and strings – either an accident or an indication that in any theory the black hole entropy is the maximum entropy that any system could have? One could then argue that since the D-Branes also form such a maximal entropic system, they must have the same entropy as does the black hole.

The key problem is that if string theory is going to give a statistical origin for the entropy of a black hole, it must identify the locus of the memory of the black hole. How can the black hole remember how it was formed and what the state of the radiation was which was previously emitted by the black hole during its evaporation? If one believes that some sort of D-Brane structures are responsible for this memory, where in the black hole are they located? Let us examine some of the possibilities.

(i) *Inside the black hole*: The interior of the black hole, as a classical object, contains a singularity. While the existence of singularities in string theory is unknown, the suspicion is that the theory will not be singular. However, in the field theory description of a charged black hole, it is the singularity which carries the charges associated with the black hole. It is at the singularity that one would therefore expect to find the D-Branes, the carriers of the RR charges, in the string theory formulation. While the presence of the D-Branes might well smooth out the singularity, or make irrelevant the notion of a spacetime in which the singularity lives, one would expect them to live in the region of the spacetime where the curvatures and field strengths approach the string scale. In regions where the curvatures and strengths are much less, the field theory approximation to string theory should be good, and it contains no sources for the charges of the fields making up the super-symmetric black hole.

Could D-Branes living at or near the singularity be the location for the black hole memory (Horowitz and Marolf 1997)? The singularity certainly forms a place which one would expect to be affected by any matter falling into the black hole or by the radiation emitted by the black hole. (That emitted radiation, in the standard calculations, is correlated with radiation which is created inside the horizon and falls into the singularity.) Thus, the formation and evaporation of the black hole could certainly affect the memory of the black hole if this is where it was located. However, to be effective as a source of entropy the memory cannot simply remember, like the unread books of an academic historian, but must also affect the future behaviour of the radiation emitted by the black hole. The radiation emitted late in the life of the black hole must be exquisitely and exactly correlated with the radiation emitted early in its life, in order that the state of the total radiation field be precisely determined by the state of the matter which formed the black hole. The memory must affect, in detail, the radiation which the black hole emits.

But here the low-frequency nature of the emitted radiation gets in the way. As I argued from the behaviour of the dumb hole, the radiation created by black holes is insensitive to the high-frequency behaviour and interactions of the radiation fields. It is a product of the behaviour of the spacetime at the low frequencies and long-distance scales typical of the black hole (milliseconds and kilometres for a solar mass black hole). At these scales the classical spacetime picture of the black hole is surely an accurate one, and at these scales the interior of the black hole is causally entirely separate from the exterior. Anything happening inside the black hole, at its centre,

can have no effect on the detailed nature of the radiation emitted. The D-Brane memory, if it exists at the centre of a black hole, must be sterile.

(ii) *Outside the horizon*: I have argued above that black holes are not completely forgetful. They do have a memory for energy, charges, and angular momentum. Since this memory is encoded in the fields outside the black hole, it would not be expected to be sterile, but can (and we believe does) have a determining effect on the emitted radiation. Could not the rest of the memory be encoded in the same way? Could not the memory of a black hole be in the curvature of the gravitational field or the structures of the other fields outside the black hole? As stated it would be hard to use D-Branes, as a black hole carries no RR charges outside the horizon. However, other string-like structures perhaps could have a wide variety of states for any macroscopic distribution of the gravitational field.

Let me first give some completely disparate evidence in favour of this possibility, before expressing my reservations. In the late 1970s, Paul Davies, Steve Fulling, and I (Davies, Fulling, and Unruh 1976) were looking at the stress–energy tensor of a massless scalar field in a two-dimensional model spacetime. Now, as a massless conformal field, the stress–energy tensor obeys the relation that its trace is zero:

$$g_{\mu\nu} T^{\mu\nu} = 0. \tag{7.10}$$

However, when we tried to calculate the expectation value of the tensor it was, as expected, infinite. By regularizing the value (in that case via 'point splitting' – taking the fields which form $T^{\mu\nu}$ at slightly different locations and looking at the behaviour of $T^{\mu\nu}$ as the points were brought together) we found that we were faced with a choice. We could either take the trace to be zero, and then either lose the conservation equation

$$T^{\mu\nu}{}_{;\nu} = 0, \tag{7.11}$$

or demand that the conservation equation be satisfied and lose the trace-free nature of $T^{\mu\nu}$. The most natural choice which follows from the regularization is to keep the trace-free nature and get

$$T^{\mu\nu}{}_{;\nu} = R_{,\mu}. \tag{7.12}$$

That is, the stress–energy tensor is not conserved. However the standard attitude to this problem is to maintain conservation and allow the trace to be non-zero. This is the so-called 'conformal' or 'trace anomaly' (rather than the 'divergence anomaly').

However, if we take the other route, namely abandoning conservation, then we can interpret this equation as saying that in the quantum regime, stress–energy is created by the curvature of spacetime. This interpretation is especially interesting because this conservation equation is really all we need in two dimensions to derive the Hawking radiation. This interpretation would say that the radiation in black hole evaporation is caused by the curvature of spacetime.

The chief argument against this view is that the expectation of the stress–energy tensor is supposed to form the right side of Einstein's equations, and the left side, $G^{\mu\nu}$ satisfies the conservation law as an identity. However, if the metric $g^{\mu\nu}$ is taken as an

operator, then that identity depends on the commutation of various derivatives of the metric and the metric itself: $G^{\mu\nu}$ itself does not necessarily obey the conservation law (the Bianchi identity).

This is not the place to follow this line of reasoning any further (especially since I am not sure that I have anything else to say). The suggestive point is that this interpretation would indicate that the source of the black hole radiation is curvature outside the black hole. Short-scale fluctuations in the curvature, due to the strings and perhaps D-Branes, could then affect the radiation given off by the black hole.

The problem with this locus for the memory is that it is very hard to see how this memory could be affected by the matter which formed the black hole, or which is emitted by the black hole. In particular, the matter which falls into the black hole would have to leave behind all of the quantum correlations and information which distinguishes one state from the other. In the case of the long-range fields (coupled to the mass, charges, and angular momentum), the natural equations of motion of the fields can retain this information in the external fields. However, in the case of all of the other aspects of the fields falling in, there is no known mechanism to strip each of the physical systems of all of their information before they fall into the hole. (The suggestion that, perhaps, because someone looking at matter falling into the black hole never sees it ever crossing the horizon, the information never enters the black hole either, I find very difficult to understand. It would be similar to claiming that because the fish outside the waterfall never hear the fish actually cross the sonic horizon of a dumb hole, the fish and all of the information which constitute the fish, never cross the sonic horizon either.)

We are thus left with the uncomfortable situation that if the memory is located where it can be affected, it cannot affect the out-going radiation, while if it is located where it can affect, it is not affected. In either case it is not efficacious as memory for allowing the black hole entropy to be explainable as having a statistical origin.

(iii) *Inside* and *Outside*: There is of course one other possibility, namely that the memory is located at both the centre and outside at the same time, i.e. that the memory is non-local, in two causally separated (at least at low frequencies and wavelengths) places at once (Susskind and Uglum 1994). For a solar mass black hole this would require a non-locality over distances of kilometres. The structure of the stringy world, acting on macroscopic low-energy phenomena, would have to be such that a single item could exist in two places, separated by a kilometre, and have macroscopic effects (although very subtle and long time scale) over those distances. This is very hard to accept, and would require a radical alteration of our views of how the world operates.

7.5 Conclusion

Black hole evaporation presents us still, twenty-five years after its discovery, with some fascinating and frustrating problems. What causes the radiation? Why do black holes behave like thermodynamic objects? What is the link, if any, between the thermodynamics of black holes and the statistical origin of all other thermodynamic systems?

Dumb holes, the sonic analogue of black holes, suggest that any ultimate understanding of the thermal radiation must concentrate on its creation as a low-energy, large length–scale, phenomenon. The radiation is created on scales of the order of the size of the black hole, and at time scales of the order of the light crossing time for the black hole, by the non-adiabatic nature of the black hole spacetime seen by the fields at such scales.

On the other hand, the only candidate we have for a statistical origin for the black hole thermodynamics, D-Branes and strings, are phenomena at the string scale (the Planck scale). If they are, as would naturally be expected, located inside the horizon, there seems no way that they could affect the low-energy radiation given off outside the horizon. If they correspond to structures lying outside the black hole, it is difficult to see how they could strip off all of the information from matter falling into the black hole.

As with all situations in which alternative explanations for a phenomenon lead to contradictory conclusions and where we do not really understand the physics, this paradox holds within itself the chance that it will uncover fundamental alterations in our view of physics. Black holes are a far from dead subject, and their understanding will lead to some crucial changes in our understanding of nature and our philosophical stance toward what a physical explanation can look like. Truly, black holes form one of the key frontiers where philosophy and physics meet at the Planck scale.

Notes

I would like to thank a number of people with whom I have discussed these issues. In addition to my collaborators (I would especially single out Bob Wald) through the years, I would thank Gary Horowitz for explaining to me much of what little I know about string theory, both in discussions and in his papers. I also thank Lenny Susskind for being a goad. We often disagree, but I at least have learned a lot in trying to understand why I disagree. Finally, I would like to thank N. Hambli for reading earlier versions and preventing me from making at least some mistakes in string theory I would otherwise have done. (All remaining mistakes are my fault, not his.) I would thank the Canadian Institute for Advanced Research for their salary support throughout the past 15 years, and the Natural Science and Engineering Research Council for their support of the costs of my research.

1. Geroch presented a model for extracting all of the energy of a system by lowering it into a black hole in a colloquium at Princeton University in 1972.
2. '*Οχουμένωνά*'. For an English translation of a Latin translation (by William of Moerbeck in 1269) see 'On Floating Bodies' in Heath (1897).
3. For an extensive analysis of the possibilities for the 'loss of information' by black holes, see the review paper by Page (1994). He makes many of the points I do (though pre D-Branes) and gives a calculation of how the correlations in the out-going radiation could contain details of the initial state, despite the apparent thermal and random character of emission process.
4. For a pedagogical (though technical) discussion of D-Branes, see Polchinski (1998).
5. Horowitz and Polchinski (1997) argue that one would expect the entropy of an excited string to (of order of magnitude) smoothly join to the entropy of a black hole as one increased the string coupling, and thus the gravity, from zero: even in the strongly coupled regime, the string entropy and the black hole entropy should be comparable.