Classical Black Holes Are Hot*

Erik Curiel[†]

August 16, 2014

Contents

1 Introduction **2**

2 The Laws of Black Hole Mechanics and the Laws of Thermodynamics **5**

3 The Standard Argument Does Not Work **7**

4 Temperature and Entropy in Classical Thermodynamics **9**

5 Taking Black Holes Seriously as Thermodynamical Objects **12**
- 5.1 Irreducible Mass, Free Energy and "Heat" of Black Holes 12
- 5.2 Carnot-Geroch Cycles for Schwarzschild Black Holes 14
- 5.3 The Clausius and Kelvin Postulates for Black Holes 20

6 Problems, Possible Resolutions, Possible Insights, and Questions **21**

References **23**

*I thank The Young Guns of the Spacetime Church of the Angle Brackets for helpful discussion on an earlier draft of this paper, and Ted Jacobson for insightful and critical discussion of my proposed Carnot-Geroch cycle for classical black holes. I also thank audiences at the physics and philosophy departments of the University of Western Ontario, at the 2011 European Philosophy of Science Association, at the Munich Center for Mathematical Philosophy, at the NYU/Columbia/Rutgers Philosophy of Physics Group, at the 2014 British Society for Philosophy of Science, and most especially at the Southern California Group on the Philosophy of Physics, for invaluable feedback on earlier versions of this work. Finally, I thank in particular Harvey Brown, Craig Callender, Robert Geroch, John Manchak, Wayne Myrvold, Chris Smeenk, and Jim Weatherall for many fruitful and enjoyable discussions of all the issues I treat here.

[†]**Author's address**: Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität, Ludwigstraße 31, 80539 München, Germany; **email**: `erik@strangebeautiful.com`

ABSTRACT

In the early 1970s it is was realized that there is a striking formal analogy between the Laws of black-hole mechanics and the Laws of classical thermodynamics. Before the discovery of Hawking radiation, however, it was generally thought that the analogy was only formal, and did not reflect a deep connection between gravitational and thermodynamical phenomena. It is still commonly held that the surface gravity of a stationary black hole can be construed as a true physical temperature and its area as a true entropy only when quantum effects are taken into account; in the context of classical general relativity alone, one cannot cogently construe them so. Does the use of quantum field theory in curved spacetime offer the only hope for taking the analogy seriously? I think the answer is 'no'. To attempt to justify that answer, I shall begin by arguing that the standard argument to the contrary is not physically well founded, and in any event begs the question. Looking at the various ways that the ideas of "temperature" and "entropy" enter classical thermodynamics then will suggest arguments that, I claim, show the analogy between classical black-hole mechanics and classical thermodynamics should be taken more seriously, without the need to rely on or invoke quantum mechanics. In particular, I construct an analogue of a Carnot cycle in which a black hole "couples" with an ordinary thermodynamical system in such a way that its surface gravity plays the role of temperature and its area that of entropy. Thus, the connection between classical general relativity and classical thermodynamics on their own is already deep and physically significant, independent of quantum mechanics.

1 Introduction

I aim in this paper to clarify the status of the analogy between black-hole mechanics restricted to general relativity on the one hand (*i.e.*, with no input from quantum field theory on curved spacetime or from any other type of semi-classical calculation) and classical thermodynamics on the other ("classical" in the sense that no quantum and no statistical considerations come into play). Based on the striking formal similarities of the respective mathematical formulæ of the Zeroth, First, Second and Third Laws of classical thermodynamics and of the mechanics of black holes in stationary, axisymmetric, asymptotically flat spacetimes, the best particular analogies seem to be: (1) that between the surface gravity of a black hole as measured on its event horizon and the temperature of a classical system; and (2) that between surface area of the horizon and entropy.[1] When it is also noted that black holes, like ordinary thermodynamical systems, are characterized by a small number of gross parameters independent of any details about underlying microstructure, and that each version of the First Law states a conservation principle for essentially the same quantity as the other, *viz.*, mass-energy, it becomes tempting to surmise that some deep or fundamental

[1] Both the surface gravity and the surface area in question are defined with respect to the orbits of the Killing fields in virtue of which the spacetime is qualified as 'stationary' and 'axisymmetric'. See Wald (1984, ch. 12) for details.

connection between black holes and thermodynamics is being uncovered. But is it of real physical significance in some sense?

The conventional answer to this question is ‘no’. Because classical black holes are perfect absorbers, they would seem to have a temperature of absolute zero, even when they have non-zero surface gravity. It is only with the introduction of quantum considerations, the standard account runs, in particular the derivation of Hawking radiation, that one finds grounds for taking the analogy seriously. And yet the startling and suggestive fact remains that one can derive laws for black holes formally identical to those of classical thermodynamical systems from the fundamental principles of general relativity itself with no aid from quantum field theory in curved spacetime. Does the use of quantum field theory in curved spacetime offer the only hope for taking the analogy seriously? I think the answer is ‘no’. To attempt to justify that answer, I shall begin by arguing that the standard argument to the contrary is not physically well founded, and in any event begs the question. Looking at the various ways that the ideas of “temperature” and “entropy” enter classical thermodynamics then will suggest arguments that show the analogy between classical black-hole mechanics and classical thermodynamics should be taken seriously indeed, without the need to rely on or invoke quantum mechanics. If this is correct, then there may already be a deep connection between general relativity and classical thermodynamics on their own, independent of quantum mechanics.

My arguments in this paper, however, are not only negative. I do think that the connection between gravitational and thermodynamical phenomena intimated by the formal equivalence of their respective Laws is of real physical significance. My strongest argument in favor of this claim is the construction of the analogue of a Carnot cycle with the heat sink provided by a stationary black hole. In the process, the black hole’s surface gravity and area play, respectively, the physical roles of temperature and entropy of an ordinary heat sink in an ordinary Carnot cycle. There also follows from the construction the existence of a universal constant with the physical dimensions needed to give surface gravity the physical dimension of temperature and area the physical dimension of entropy. If surface gravity and entropy couple to ordinary thermodynamical systems in the same way as temperature and entropy, respectively, do, then there can be no grounds for denying that they physically *are* a real temperature and entropy. To put it more provocatively, if my claim is correct, then gravity on its own, independent of its relation to the other three known fundamental forces so successfully treated by quantum field theory, already is a fundamentally thermodynamical phenomenon.[2] I want to stress, nonetheless, that I do not consider quantum effects to be irrelevant when considering possible relations between gravitational physics and thermodynamics. I want only to argue for the idea that the analogy between the laws of classical thermodynamics and those of black hole mechanics in classical general relativity is robust and deep in its own right.

I should perhaps say, by way of background, that I am curious about this question in the first place in part because of my curiousity about the larger question of the relation between thermodynamical characteristics of a physical system and the possibility of always being able to or indeed always being required to find an underlying statistical interpretation of those thermodynamical characteristics.

[2] If one could show that the sorts of arguments I give here could be translated into the framework of Newtonian gravitational theory, that would provide even stronger support for this last claim.

That the laws of black hole mechanics follow from the fundamental theory itself (in this case, general relativity), and are not as with classical thermodynamics an independent adjunct connected to the underlying fundamental (Newtonian) theory through the use of statistical devices, could suggest that thermodynamics is itself more of the nature of a fundamental theory than has been thought since the advent of statistical mechanics—or at least that thermodynamical characteristics and quantities of physical systems may be fundamental to them in some way analogous to that of other fundamental characteristics and dynamical quantities, such as the possession of a stress-energy tensor, for example, and its satisfaction of some form of covariant conservation principle. Contrarily, these sorts of results may also perhaps lend support to the idea that general relativity is an effective field theory, and the Einstein field equation only an equation of state, *à la* Jacobson (1995), and perhaps Bredberg, Keeler, Lysov, and Strominger (2011) and Lysov and Strominger (2011). If that is true, then the entire program of "quantizing gravity" may be misguided from the start. Yet another possibility, contrary to that just mentioned, is that one may take my arguments as showing that the signature of quantum gravity, in particular the traces of whatever statistical quantities it may give us for making traditional sense of the thermodynamical phenomena I discuss here, show up already in purely classical, non-statistical theory.[3] Finally, and I think most importantly, my arguments lend *prima facie* support to projects (especially in cosmology) that want to attribute entropy generically to "gravitational degrees of freedom", as in the work of Clifton, Ellis, and Tavakol (2013), and as required by Penrose's Conformal Curvature Hypothesis (Penrose 1979).[4]

I do not intend to investigate these larger issues here, however. I intend to investigate only the status of the analogy between the laws of classical thermodynamics on the one hand and those of black hole mechanics in classical general relativity on the other. I mention these larger issues only to give some of my motivation for this work, and to place it in the context of important work being done in many branches of theoretical physics today.

There are other motivations behind this project as well. Although philosophers of physics have recently begun to work on issues arising from proposals for theories of quantum gravity, some of which take as their starting points the seemingly thermodynamical character of gravitational phenomena as exemplified by the laws of black hole mechanics, almost no philosophical work has been done investigating the nature of this seemingly thermodynamical character as revealed by the structures of general relativity and of quantum field theory formulated on curved, relativistic spacetimes. Because general relativity and quantum field theory are well entrenched, clearly and rigorously articulated physical theories, I believe it behooves philosophers to study it, if not before, at least in conjunction with work done on quantum gravity.

[3] I thank Fay Dowker for elucidating this possibility in a very helpful way in conversation.

[4] I thank Harvey Brown for his relentless, bearish, and, most of all, extraordinarily helpful asking of me, "So what?"

2 The Laws of Black Hole Mechanics and the Laws of Thermodynamics

Within the context of general relativity, one can derive laws describing the behavior of black holes in stationary, asymptotically flat spacetimes bearing a remarkable resemblance to the classical laws of equilibrium thermodynamics. I restrict attention to the asymptotically flat case, because those are the simplest natural analogue of an isolated system for black holes in general relativity.[5] I restrict attention to stationary black holes because those are the simplest natural analogue of an equilibrated system for black holes in general relativity.

Now, for the laws themselves:[6]

Zeroth Law

> [**Thermodynamics**] The temperature T is constant throughout a body in thermal equilibrium.[7]
>
> [**Black Holes**] The surface gravity κ is constant over the event horizon of a stationary black hole.

First Law

> [**Thermodynamics**]
> $$\mathrm{d}E = T\mathrm{d}S + p\mathrm{d}V + \Omega\mathrm{d}J$$
> where E is the total energy of the system, T the temperature, S the entropy, p the pressure, V the volume, Ω the rotational velocity and J the angular momentum.[8]
>
> [**Black Holes**]
> $$\delta M = \frac{1}{8\pi}\kappa\delta A + \Omega_{\text{BH}}\delta J_{\text{BH}}$$

[5] The generalization of the idea of a black hole and of the Four Laws to the non-asymptotically flat case by Hayward (1994), by the use of so-called dynamical trapping horizons, is of great interest, but to treat them would take us beyond the scope of this paper. Also, I will not discuss the so-called Minus-First Law of Brown and Uffink (2001); much work has been done to prove, or at least argue for, its correlate in black-hole mechanics (though not referred to as such in that literature), that perturbed black holes tend to settle down to equilibrium, and, in particular, that the sorts of perturbations I consider here do not destroy the event horizon. There are now strong plausibility arguments in favor it (Hollands and Wald 2012), but its status in black-hole mechanics is still, to my mind, very much up for grabs, though, as a betting man, my money is on there being arguments for it at least as strong as for the Third Law (which, perhaps, is not to say very much).

[6] For proofs of the laws, see Wald (1984, ch. 12), Israel (1986), Wald (1994), and Wald and Gao (2001).

[7] This is not the standard formulation of the thermodynamical Zeroth Law, which is "If two systems are in thermal equilibrium with a third, then each is in thermal equilibrium with the other." Because the formulation I use and the standard formulation are essentially equivalent when the Kelvin Postulate (4.2 below) is assumed, however, and the appropriate translation of the Kelvin Postulate is a theorem in classical black-hole mechanics (5.3.2 below), this is not a problem for my arguments.

[8] Strictly speaking, this is not the First Law, but rather the Gibbs Relation, which is equivalent to the First Law for thermodynamical systems in equilibrium. Since all my arguments involve only systems in equilibrium, and, as is standard in thermodynamical arguments, systems that deviate from equilibrium only by quasi-stationary effects (for which the Gibbs relation still holds), this is not a problem.

where M is the total black hole mass, A the surface area of its horizon, Ω_{BH} the "rotational velocity" of its horizon,[9] J_{BH} its total angular momentum, and 'δ' denotes the result of a first-order, linear perturbation of the spacetime.[10]

Second Law

[**Thermodynamics**] $\delta S \geq 0$ in any process.

[**Black Holes**] $\delta A \geq 0$ in any process.

Third Law

[**Thermodynamics**] $T = 0$ is not achievable by any process.[11]

[**Black Holes**] $\kappa = 0$ is not achievable by any process.

The most striking architectonic similarity between the characterization of ordinary thermodynamical systems (in equilibrium) by the laws of thermodynamics and the characterization of black holes (in equilibrium, *i.e.*, stationary) is that in each case the behavior of the system, irrespective of any idiosyncracies in the system's constitution or dynamical history, is entirely captured by the values of a small number of physical quantities, 6 for ordinary thermodynamical systems, 4 for black holes: in the former case, they are temperature, entropy, pressure, volume, angular velocity and angular momentum;[12] in the latter, they are surface gravity, area, angular velocity and angular momentum. The Zeroth and Third Laws suggest that we take the surface gravity of a black hole as the analogue of temperature. The Second Laws suggest that we take area as the analogue of entropy. This is consistent with the First Law, if we treat $\frac{1}{8\pi}\kappa\delta A$ as the Gibbsian "heat" term for a system in thermal equilibrium. Indeed, if we do so then the analogy for the First Law becomes exact: relativistically, energy just is mass, so the lefthand side terms of the First Law for ordinary systems and for black holes are not just analogous, they are physically identical; likewise, $\Omega_{\text{BH}}\delta J_{\text{BH}}$ as a work term in the law for black holes is physically identical to the corresponding term in the law for ordinary systems.

[9]See Wald (1984, ch. 12, §3, pp. 319–320).

[10]For an exact definition and thorough discussion of the perturbations used, see Wald and Gao (2001). There is an oddity about this formulation of the law, however, that I have not seen addressed in the literature but is surely worth puzzling over. While the δ acting on M is the same as that acting on J_{BH}, it is not the same as that acting on A. The δ acting on M and J_{BH} represents a perturbation of a quantity taken asymptotically at spatial infinity; the other represents perturbations taken "at the event horizon". I know of no other physically significant equation where different differential operators act on different mathematical spaces in such a way that, as in this case, there's no natural mapping between them. What's going on here?

[11]I actually think this is a defective statement of the Third Law of thermodynamics. (See, *e.g.*, Schrödinger 1960, Aizenman and Lieb 1981 and Wald 1997 for a discussion of some of its problems.) Schrödinger (1960) provides a far more satisfactory statement of the Third Law, which I think carries over well into black-hole thermodynamics. I do not have room to go into the matter here, though.

[12]Of course, the First Law guarantees that not all these quantities will be independent, and, if one is considering a particular species of thermodynamical system, then one may have available an equation of state that will further reduce the number of independent quantities, but all that is beside the point for my purposes.

Now the force of the question motivating this paper should be clear: the mathematical analogy is perfect, and there are already some indications that the analogy may reach down to the level of physics, not just mathematics. But how far should we take the analogy? What can it mean to take seriously the idea that the surface gravity of a black hole is a physical temperature, and its area a physical entropy?

3 The Standard Argument Does Not Work

There are well-known difficulties with taking the surface gravity of a classical black hole to represent a physical temperature. One important method for defining the thermodynamic temperature of an object derives from the theory of thermal radiation from black bodies. If a normal black body immersed in a bath of thermal radiation settles down to thermal equilibrium, it will itself emit thermal radiation with a power spectrum characteristic of its equilibrium temperature as measured using a gas thermometer. This power spectrum can then be used to define a temperature scale. It is this definition of thermodynamic temperature that is almost always (at times implicitly) invoked when the claim is made that if one considers classical general relativity alone then black holes, being perfect absorbers and perfect non-emitters, have an effective temperature of absolute zero.[13]

To try to be a little more precise, I will offer a reconstruction of the standard argument. It is not given in exactly this form by anyone else in the literature, but I think it captures both the spirit and the letter of the orthodox view. Put a Kerr black hole in a box with perfectly reflective sides, which are far from the event horizon (in the sense that they are many times farther away from the event horizon "in natural spacelike directions" than its own "natural" diameter). Pervade the box with thermal radiation. According to classical general relativity, the black hole will absorb all incident thermal radiation, and emit none, until eventually all thermal radiation in the box (outside the event horizon) has vanished, so the black hole must have a temperature of absolute zero. Thus, the surface gravity κ, which is never zero for a non-extremal Kerr black hole, cannot represent a physical temperature of the black hole in classical general relativity. Conventional wisdom holds, as a result, that if the formal similarities mentioned above were all there were to the matter then they would most likely represent a merely accidental resemblance or perhaps would indicate at best a superficial relationship between thermodynamics and black holes, but in any event would not represent the laws of classical thermodynamics as extended into the realm of black holes.[14]

In 1974, using semi-classical approximation techniques Hawking discovered that stationary, axisymmetric black holes appear to radiate as though they were perfect black-body emitters in thermal equilibrium with temperature $\frac{\hbar}{2\pi}\kappa$, when quantum particle-creation effects near the black hole horizon are taken into account (Hawking 1974; Hawking 1975). It is this result that is generally taken to justify the view that the resemblances between the laws of black hole mechanics and the laws of classical thermodynamics point to a fundamental and deep connection among general relativ-

[13] See for example the remarks in Bardeen, Carter, and Hawking (1973), Carter (1973) and Wald (1999).

[14] The remarks of Wald (1984, p. 337), for example, are exemplary in this regard.

ity, quantum field theory and thermodynamics, and in particular that κ *does* in fact represent the physical temperature of a black hole, and therefore A its entropy.[15]

I have two problems with this orthodoxy. First, I find the physical content of the standard argument not to stand up to scrutiny. While it is true that the Kerr black hole in the box, according to classical general relativity, will emit no blackbody radiation while it absorbs any incident on it, that is not the end of the story. Classical general relativity does tell us that the Kerr black hole will emit some radiation, *viz.*, *gravitational* radiation, while it is perturbed by the infalling thermal radiation, and that gravitational radiation will in fact couple with the thermal radiation still outside the black hole. If we are trying to figure out whether purely gravitational objects, such as black holes, have thermodynamical properties, we should surely allow for the possibility that gravitational radiation, or, indeed, the exchange of "gravitational energy" in any form, may count as a medium for thermodynamical coupling.[16] Indeed, just as electromagnetic radiation turned out to be a medium capable of supporting a physically significant coupling of electromagnetic systems with classical thermodynamical systems, it seems *prima facie* plausible that gravitational radiation may play the same role for gravitational systems. Just as "heat" for an electromagnetic system may be measured by electromagnetic radiation, at least when transfer processes are at issue, so it may be that "heat" for a gravitational system may be measured by gravitational radiation, or any form of exchange of gravitational energy, again at least when transfer processes are at issue. Electromagnetic energy is just not the relevant quantity to track when analyzing the thermodynamic character of purely gravitational systems.

Second, I do not think this definition of temperature is the appropriate one to use in the context of a purely classical description of black holes, for the electromagnetically radiative thermal equilibrium of systems immersed in a thermal bath is essentially a *quantum* and *statistical* phenomenon, by which I mean one that can be correctly modeled only by using the hypothesis that thermal energy is exchanged in discrete quanta and then computed correctly only with the use of statistical methods. To use that characterization of temperature to argue that we must use quantum mechanics in order to take surface gravity seriously as a physical temperature, therefore, is to beg the question. If my qualm is well founded, it follows that the standard argument does not bear on the strength of the analogy as indicating a real physical connection between classical general relativity and thermodynamics. After all, if one is trying to determine the status of the analogy between *classical* gravitational theory and *classical* thermodynamics independently of any quantum considerations, then the most appropriate characterizations of temperature to use are those grounded strictly in classical thermodynamics itself.

[15] See again, for example, the remarks of Wald (1984, p. 337). Indeed, some of the most important researchers in the field make even stronger claims. Unruh and Wald (1982, p. 944), for example, claim that "the existence of acceleration radiation [outside the event horizon, a fundamentally quantum phenomenon,] is vital for the self-consistency of black-hole thermodynamics."

[16] I use scare-quotes for 'gravitational energy' because that is an infamously vexed notion in classical general relativity, with no cogent way known to localize it, and indeed strong reasons to think there can be no localization of it in general. (See, *e.g.*, Curiel 2013.) I will discuss this issue, and the potential problems it may raise for my arguments, in §6.

There is yet another *prima facie* problem, however, with trying to interpret surface gravity as a true temperature and area as a true entropy, which my arguments so far do not address: neither has the proper physical dimension. In geometrized units, the physical dimension of temperature is mass (energy), and entropy is a pure scalar. The physical dimension of surface gravity, however, is mass^{-1}, and that of area mass^2. There are no purely classical universal constants, moreover, available to fix the dimensions by multiplication or division.[17] The only available universal constant to do the job seems to be $\hbar$, which has the dimension mass^2.[18] I cannot address this problem at this stage of my arguments. Remarkably, however, it will turn out as a natural sequela to my construction of the appropriate analogue of a Carnot cycle for black holes, in §5.2, that the existence of a universal constant in the classical regime with the proper dimension is guaranteed.

4 Temperature and Entropy in Classical Thermodynamics

I think there are grounds for taking the analogy very seriously even when one restricts oneself to the classical theories, without input from or reliance on quantum theories. To make the case more poignant, imagine that we are physicists who know only classical general relativity and classical thermodynamics, but have no knowledge of quantum theory. How could we determine whether or not to take black holes as thermodynamical objects in a substantive, physical sense, given that we know the deep formal analogy between the two sets of laws? In such a case, we ought to look to the way that temperature and entropy are introduced in classical thermodynamics and the various physical roles they play there. If the surface gravity and area of black holes can be introduced in the analogous ways and play the analogous physical roles, I contend that the global analogy is already on strong ground. In other words, the surface gravity and area must play the same role in the new theory *vis-à-vis* other theoretical quantities as temperature and entropy do in the original theory *vis-à-vis* the analogous theoretical quantities there. If, moreover, it can be shown that surface gravity couples to ordinary classical thermodynamical systems in the same formal way as ordinary temperature does, then there are no grounds for denying that it is a true physical temperature.[19] And if area for black holes is related to surface gravity and to the proper analogue of heat in the same way as entropy is to ordinary temperature and heat, and if it is required for formulating an appropriately generalized Second Law, then there are no grounds for denying that it is a true physical entropy. Indeed, it was exactly on grounds such as these that physicists in the 19th century concluded that the power spectrum of blackbody radiation itself encoded a *physical* temperature and entropy, not merely that there was an analogy between thermodynamics and the theory of blackbody radiation. Planck (1926) himself had doubts about the thermodynamical character of blackbody radiation until

[17] All the classical universal constants, such as the speed of light and Newton's gravitational constant, are dimensionless. This is actually a puzzling state of affairs, that surely deserves investigation.

[18] I am grateful to Ted Jacobson and Carlo Rovelli for pushing me on the issue of the physical dimensions of the quantities, and on the seeming need to introduce $\hbar$ to make things work out properly.

[19] Since entropy directly mediates no coupling between thermodynamical systems, the same argument is not available for it. This is one of the properties of entropy that makes it a truly puzzling physical quantity: there is no such thing, not even in principle, as an entropometer.

he had satisfied himself on these points.

There are three fundamental, related ways that temperature is introduced in classical thermodynamics, which themselves ground the various physical roles temperature can play in the theory (how it serves as the mediator of particular forms of coupling between different types of physical system, *e.g.*). The first derives from perhaps the most basic of the thermodynamic characteristics of temperature and is perhaps most definitive of the cluster of ideas surrounding the concepts of "temperature" and "heat": it is that when two bodies are brought into contact, heat will spontaneously flow from the one of higher temperature to the one of lower temperature.[20] The second arises from the fact that increase in temperature is positively correlated with increases in the capacity of a system to do work. This fact allows one to define an empirical scale of temperature through, *e.g.*, the use of a gas thermometer: the temperature reading of the thermometer is made directly proportional to the volume of the thermometric gas used, which is itself directly proportional to the work the gas does on its surrounding container as it expands or contracts in response to its coupling with the temperature of the system being measured. The utility of such a scale is underwritten by the empirical verification that such empirical scales defined using a multitude of different gases under a multitude of different conditions are consistent among one another.[21] The third arises from an investigation of the efficiency of reversible, cyclic engines, *viz.*, Carnot engines, which yields a definition of the so-called absolute temperature scale associated with the name of Kelvin.[22] It is the possibility of physically identifying the formally derived absolute scale with the empirically derived scale based on capacity to do work (increase in volumes, *e.g.*) that warrants the assertion that they both measure the same physical quantity.[23]

Likewise, there are (at least) three ways that entropy enters classical thermodynamics. The first historically, and perhaps the most physically basic and intuitive, is as a measure of how much energy it takes to transform the heat of a thermal system into work: generally speaking, the free energy of a thermodynamical system is inversely proportional to its entropy.[24] The second is as that perfect differential $\mathrm{d}S$ into which temperature, as integrating factor, transforms exchanges of heat $\mathrm{d}Q$ over the course of quasi-stationary processes (Fermi 1956, ch. IV): the integral of $\mathrm{d}Q$ along a quasi-stationary path between two equilibrium states in the space of states of a thermodynamical system is not independent of the path chosen, whereas the integral of $\frac{\mathrm{d}Q}{T}$ is. (Indeed, Sommerfeld 1964 uses this fact to conclude that entropy is a true physical property of a thermodynamical system,

[20]It is important for some of my later arguments to note that this characterization of comparative temperature does not preclude processes in which heat at the same time flows from the colder body to the hotter. It says only that it is always the case that heat flows from hotter to colder, irrespective of what may or may not happen in the reverse direction.

[21]Planck (1926, §1, p. 1) remarks that quantitative exactness is introduced into thermodynamics through this observation, for changes of volume admit of exact measurements, whereas sensations of heat and cold do not, nor even comparative judgments of hotter and cooler on their own.

[22]See, *e.g.*, Fermi (1956, §§8–10).

[23]Maxwell (1888, chs. VIII, XIII) gives a wonderfully illuminating discussion of the physical basis of the equivalence of the absolute temperature scale with the one based on gas thermometry.

[24]Again, the discussion of Maxwell (1888, ch. XII) about this idea is a masterpiece of physical clarification and insight.

whereas heat content is not.) The third also arises from the analysis of the efficiency of Carnot cycles (Fermi 1956, ch. IV).

Now, the following fundamental theorem of classical thermodynamics provides the basis both for the definition of the absolute temperature scale and for the introduction of entropy as the perfect differential derived from exchanges of heat when that temperature is used as an integrating factor.

Theorem 4.1 *Any two reversible, cyclic engines operating between temperatures T_2 and T_1 (as measured using gas thermometry) have the same efficiency. The efficiency of any non-reversible engine operating between T_2 and T_1 is always less than this.*

This theorem is a direct consequence of either of two classical postulates, which can be argued on physical grounds both to be equivalent to each other and to directly imply the Second Law of thermodynamics (for the proofs of which statements see, *e.g.*, Fermi 1956):[25]

Postulate 4.2 (Lord Kelvin) *A transformation whose only final result is to transform into work heat extracted from a source that is at the same temperature throughout is impossible.*

Postulate 4.3 (Clausius) *A transformation whose only final result is to transfer heat from a body at a given temperature to a body of a higher temperature is impossible.*

I claim that these last two postulates, and the fact that they provide grounds for proof of the efficiency theorem and the Second Law, encode essentially all that is of physical significance in the ways I sketched that both temperature and entropy enter into classical thermodynamics.

The Clausius Postulate captures the idea that when two bodies are brought into thermal contact, heat flows from the body of higher temperature to the other. The Kelvin Postulate captures the idea that the capacity of a body to do work on its environment tends to increase as its temperature increases. If one could show that appropriately formulated analogues to these two propositions about classical black holes hold in general relativity, with surface gravity playing the role of temperature and area that of entropy, one would have gone a long way towards showing that surface gravity *is* a true thermodynamical temperature and area a true entropy. If one could moreover show that those analogues imply the properly formulated analogue of the efficiency theorem, and so use them to define an absolute temperature scale for black holes that was essentially equivalent to surface gravity, and then showed that surface gravity so characterized played the role of integrating factor for the gravitational analogue of heat, turning it into the perfect differential of area, the result would be even more secure. Finally, if one could show that ordinary thermodynamical systems equilibrate with black holes in a way properly mediated by their ordinary temperature and by the black hole's surface gravity, the analogy would have been shown to be far more than analogy: it would be physical equivalence in the strongest possible sense. I prove all these propositions in §5 below.

[25] Historically, these two propositions were themselves taken to be (equivalent statements of) the Second Law, and the principle of entropy increase was not accorded the fundamental status it is today. See, *e.g.*, Maxwell (1888, ch. VIII).

5 Taking Black Holes Seriously as Thermodynamical Objects

What is needed, first, is a way to characterize "thermal coupling" between black holes and ordinary thermodynamical systems: granted that "heat" in the gravitational context is gravitational energy of a particular form, such as that carried in the form of gravitational radiation or that responsible for red-shift effects in monopole solutions, then one also realizes that black holes *are not perfect absorbers*. When there is an ambient electromagnetic field, the black hole will radiate gravitationally as it absorbs energy and grows from the infalling electromagnetic radiation. So to conclude that surface gravity is a physical temperature, one need show only that the gravitational energy exchanged between a black hole and other thermodynamical systems in transfer processes depends in the appropriate way on the surface gravity of the event horizon.[26] This approach has *prima facie* physical plausibility: to take the energy in gravitational radiation, *e.g.*, to be the gravitational equivalent of heat is the same as to take the energy in electromagnetic radiation to be the electromagnetic equivalent of heat—it is what couples in the appropriate way to the average kinetic energy of molecules in ordinary thermodynamical systems, *viz.*, what makes it increase and decrease, and that with respect to which equilibrium is defined.

Just as the concept of "thermal coupling" had to be emended in the extension of classical thermodynamics to include phenomena associated with radiating black bodies, so we should expect it to be in this case. In classical thermodynamics before the inclusion of black-body phenomena, thermal coupling meant immediate spatial contiguity: heat was known to flow among solids, liquids and gases only when they had surfaces touching each other.[27] In order to extend classical thermodynamics to include black-body phenomena, the idea of thermal coupling had to be extended as well: two black bodies thermally couple when and only when the ambient electromagnetic field each is immersed in includes direct contributions from the electromagnetic radiation emitted by the other. They do not need to have surfaces touching each other.

In order to characterize the correct notion of thermal coupling among systems including black holes (or more generalized purely gravitational systems, such as cosmological horizons), we first need to characterize an appropriate notion of "heat" for black holes, and the concomitant notion of free energy. That will put us in a position to formulate the appropriate generalizations of the Clausius and Kelvin Postulates for such systems, and to construct the appropriate generalization of Carnot cycles for them.

5.1 Irreducible Mass, Free Energy and "Heat" of Black Holes

In analyzing the ideas of reversibility and irreversibility for processes involving black holes, Christodoulou (1970) introduced the *irreducible mass* M_{irr} of a black hole of mass M and angular momentum

[26] I will discuss in §6 below the fact that there is no well defined notion of localized gravitational energy in general relativity, and how that may bear on my arguments.

[27] This fact, perhaps, contributed to the historical idea that heat was a fluxional, perhaps even fluid, substance, such as phlogiston or caloric.

J:[28]

$$M_{\mathrm{irr}}^2 := \frac{1}{2}[M^2 + (M^4 - J^2)^{\frac{1}{2}}]$$

(From hereon, I shall drop the subscripted 'BH' on terms denoting quantities associated with black holes, except in cases where ambiguity may arise.) Inverting the definition yields

$$M^2 = M_{\mathrm{irr}}^2 + \frac{1}{4}\frac{J^2}{M_{\mathrm{irr}}^2}$$

and so, for a Kerr black hole,

$$M > M_{\mathrm{irr}}$$

(Clearly, $M_{\mathrm{irr}} = M$ for a Schwarzschild black hole.) Thus, the initial total mass of a black hole cannot be reduced below the initial value of M_{irr} by any physical process. A simple calculation for a Kerr black hole, moreover, shows that,

$$A = 16\pi M_{\mathrm{irr}}^2 \tag{5.1.1}$$

Thus, it follows from the Second Law that M_{irr} itself cannot be reduced by any physical process, and so any process in which the irreducible mass increases is a physically irreversible process. In principle, therefore, the free energy of a black hole is just $M - M_{\mathrm{irr}}$, in so far as its total mass M represents the sum total of all forms of its energies, and M_{irr} represents the minimum total energy the black hole can be reduced to.[29]

In classical thermodynamics, it makes no sense to inquire after the absolute value of the quantity of heat a given system possesses. In general, that is not a well defined property accruing to a system. One rather can ask only about the amount of heat tranferred between bodies during a given process.[30] Consider, then, a classical thermodynamical system with total energy E and free energy E_{f}. $E - E_{\mathrm{f}}$ is the amount of energy unavailable for extraction, what Kelvin called its dissipated energy, E_{d}. Say that through some quasi-stationary process, we know now what, both E and E_{d} change so that it now has less free energy than it did before; therefore, the entropy of the system must have increased, which can happen only when it absorbs heat, which will in general be the difference between the total change in energy and the change in free energy. If they both change so that it has more free energy, the same reasoning applies, and it must have given up a quantity of heat equal to that difference.

These remarks suggest defining the "quantity of heat transferred" to or from a black hole during any quasi-stationary thermodynamical process to be the change in its free energy, which is to say

[28] I will discuss only Kerr black holes, not Kerr-Newman black holes that also have electric charge, as the ensuing technical complications would not be compensated by any gain in physical comprehension.

[29] Some—*e.g.*, Wald (1984, ch. 12, §4)—interpret $M - M_{\mathrm{irr}}$ as the rotational energy of a Kerr black hole, in so far as extracting that much energy from a black hole would necessarily reduce its angular momentum to zero. Based on the arguments I will give in this section, I prefer to think of it as a thermodynamical free energy, which cannot necessarily be decomposed in a canonical way into different "forms", *e.g.*, that much heat and that much rotational energy, *etc.*

[30] See Maxwell (1888, chs. I, III, IV, VIII, XII).

the change in total black hole mass minus the change in its irreducible mass, $\Delta M - \Delta M_{\text{irr}}$.[31] If, for instance, the irreducible mass of a black hole does not change, while the total mass decreases, then it would have given up a quantity of heat. As a consistency check, it is easy to see that, according to this definition, when an ordinary thermodynamical system in equilibrium is dumped into a Kerr black hole, the black hole absorbs the quantity of heat the ordinary matter contained as characterized by the Gibbs relation, *viz.*, its temperature times its entropy, as only that energy contributes to its total mass without directly changing its angular momentum. Based on this characterization of "quantity of heat transferred", I claim that the appropriate notion of thermal coupling for systems involving black holes is any interaction where there is a change in the black hole's free energy. For purely gravitational interactions, this includes emission and absorption of that part of the energy of gravitational radiation not due to angular momentum, energy exchange due to simple monopole- or multipole-moment couplings in the near-stationary case, and so on.

Some care must be taken in applying this definition to Schwarzschild black holes, however. Because $M = M_{\text{irr}}$ for a Schwarzschild black hole, one can never give up heat while remaining Schwarzschildian. Schwarzschild black holes, essentially, have achieved heat death—one cannot extract energy from them without perturbing them in an appropriate way. Similarly, they cannot absorb heat: if it absorbs ordinary heat from a classical thermodynamical system, say, being thrown into it, unless it acquires angular momentum in the process, then after it settles down again to staticity it will once again have its total mass equal to its irreducible mass. In this case, I think it still makes sense to say the black hole has absorbed heat, in so far as, between the time the system is thrown in and the time the black hole equilibrates again, its irreducible mass will not be equal to its total mass. The maximum of this difference, during the equilibration process, will presumably equal the energy of the system the black hole absorbed. There are many challenges one could reasonably pose to the approximations involved in attempting to carry out such a calculation with anything approaching rigor (which I have not done), but they are all the same sort of challenge one could pose to the analogous problem in classical thermodynamics, so there is no problem here peculiar to black-hole thermodynamics.

5.2 Carnot-Geroch Cycles for Schwarzschild Black Holes

As I remarked at the end of §4, the strongest evidence that the formal equivalence of the laws of black holes and those of ordinary thermodynamical systems in fact constitutes a true physical equivalence, and that surface gravity is a physical temperature and area a physical entropy, would consist in a demonstration that black holes thermally couple with ordinary thermodynamical systems in such a way that κ plays the same role in that coupling as ordinary temperature would if the system at issue were coupling with another ordinary thermodynamical system and not with a black hole, and the same for area. My proposed construction of the appropriate analogue for a Carnot cycle

[31] I thank Harvey Brown for drawing to my attention the fact that Carathéodory (1909), in his ground-breaking axiomatization of classical thermodynamics, introduced the notion of heat in a way very similar to this, not as a primitive quantity as is usually done, but as the difference between the internal and the free energies of a system.

including black holes, which I give in this subsection, will kill three birds with one stone: not only will it show that κ can be characterized as the absolute temperature of the black hole using the same arguments as classical thermodynamics uses to introduce the absolute temperature scale; it will do so by showing that in the coupling of black holes with ordinary thermodynamical systems, κ does in fact play the physical role of temperature and area that of entropy; and it will have as a natural corollary the existence of a universal constant that renders the proper physical dimensions to surface gravity as a measure of temperature and area as a measure of entropy.[32]

I call the constructed process a "Carnot-Geroch cycle" both to mark its difference from standard Carnot cycles, and because it relies essentially on the mechanism at the heart of the most infamous example in this entire field of study, Geroch's putative counter-example against Bekenstein's original claim that one should think of the area of a black hole as its physical entropy.[33] I will first sketch the steps of the proposed cycle informally, then work through the calculations.

Reversible Carnot-Geroch Cycle Using a Schwarzschild Black Hole as a Heat Sink

1. start with a small, empty, essentially massless, perfectly insulating box "at infinity", one side of which is the outer face of a piston; in particular, the box is "small" in the sense that it will experience negligible tidal forces as it is lowered toward the black hole; very slowly ("quasi-statically", so that the process is well approximated as an isentropic process) draw the piston back through the inside of the box, so filling the box with fluid from a large heat bath consisting of a large quantity of the fluid at fixed temperature T_0, so the fluid does work against the piston as it moves; when the piston has withdrawn part but not all of the way to the opposite side of the box, quickly seal the box, leaving the space opened by the piston filled with a mass of the fluid M_0 in thermal equilibrium at temperature T_0, and with entropy S_0; assume the entire energy of the box is negligible compared to the mass of the black hole
2. very slowly, lower the box towards the black hole using an essentially massless rope; during this process, an observer inside the box would see nothing relevant change; in particular, as measured by an observer co-moving with the box, the temperature, volume and entropy of the fluid remain constant[34]

[32] I am grateful to Ted Jacobson for bringing to my attention after I wrote this paper the insightful analysis of Sciama (1976), in some ways quite similar to mine. (See Jacobson 2003 for a *prècis* of Sciama's analysis.) Sciama, however uses quantum systems all the way through and assumes that the analogy between black holes and ordinary thermodynamical systems is merely formal without reliance on quantum effects.

[33] According to Israel (1973), Geroch first proposed the example during the question-and-answer period at an informal colloquium at Princeton in 1970 at which Bekenstein was presenting an early version of his work. I cannot resist pointing out that my construction is essentially a jiu jitsu move against Geroch's original intent, turning the force of the example against itself, using Geroch's proposed mechanism to show that area really is an entropy.

[34] The mass-density distribution of the fluid would change, increasing towards the side facing the black hole; this, however, does not affect the analysis, since this is what one expects for a system in thermal equilibrium in a quasi-static "gravitational field". In any event, given our assumption about the size of the box, this effect, even if relevant to the physics of the process, would be negligible.

3. at a predetermined fixed proper radial distance from the black hole, stop lowering the box and hold it stationary
4. very slowly, draw the piston back even further, so lowering the temperature of the fluid to a fixed, pre-determined value T_1 while keeping its entropy the same; the value of the temperature is to be fixed by the requirement that the change in total entropy vanishes during the next step (*i.e.*, entropy of black hole plus entropy of everything outside black hole does not change after the fluid is dumped into the black hole)
5. open the box and eject the fluid out of it by using the piston to push it out, so the fluid falls into the black hole delivering positive mass-energy and positive entropy to it, and the piston returns to its initial state; by the way the temperature of the fluid was fixed in the previous step, this is an isentropic process
6. pull the box back up to infinity (which takes no work, as the box now has zero mass-energy, and so zero weight), so it returns to its initial state

Because the total entropy remains constant during every step in the process, these cycles are reversible in the sense of classical thermodynamics. Because the irreducible mass of the black hole increases, however, it is not an irreversible process in the sense of black-hole mechanics.[35]

Now, let us make the following assumptions: first, that it makes sense to attribute a physical temperature T_{BH} and entropy S_{BH} to a black hole (though we do not yet know what they are); second, that the entropy of ordinary thermodynamical systems and the entropy of the black hole are jointly additive; and third, that the appropriate temperature at which to eject the fluid into the black hole for the entire cycle to be isentropic (T_1 in step 5) is that one would expect for a thermally equilibrated body in thermal contact with another at temperature T_{BH} sitting the given distance away in a nearly-static gravitational field. It will then follow that the physical temperature must be $8\pi\alpha\kappa$ and the physical entropy $\dfrac{A}{\alpha}$, where κ is the black hole's surface gravity, A its area, and α is a universal constant, the analogue of Boltzmann's constant for black holes (to be derived below).

Let the static Killing field in the spacetime be ξ^a (timelike outside the event horizon, null on it). Let $\chi = (\xi^n \xi_n)^{\frac{1}{2}}$, and $a^a = (\xi^n \nabla_n \xi^a)/\chi^2$ be the acceleration of an orbit of ξ^a. Then a standard calculation[36] shows that

$$\kappa = \lim(\chi a)$$

where the limit is taken as one approaches the event horizon in the radial direction, *i.e.*, near the black hole χa is essentially the force that needs to be exerted "at infinity" to hold an object so that it follows an orbit of χ^a, which is to say, to hold it so that it is locally stationary. Thus χ is essentially the "redshift factor" in a Schwarzschild spacetime.

Let the total energy content of the box when it is initially filled at infinity be E_0 (as measured with respect to the static Killing field). In particular, E_0 includes contributions from the rest mass

[35] In Curiel (2014a), I propose another form of Carnot-Geroch cycle for a Kerr black hole, one that exploits its angular momentum in such a way as to make the process both reversible in the sense of classical thermodynamics and physically reversible according to black-hole mechanics.

[36] See, *e.g.*, Wald (1984).

of the fluid M_0, and from its temperature T_0 and entropy S_{b}; let W_0 be the work done by the fluid as it pushes against the piston in filling the box. By the Gibbs relation and by the First Law of thermodynamics, therefore, we can compute the quantity of heat Q_{b} initially in the box:

$$Q_{\mathrm{b}} = T_0 S_{\mathrm{b}} = E_0 + W_0$$

As the box is quasi-statically lowered to a proper distance ℓ from the event horizon, its energy as measured at infinity becomes χE_0, where χ is the value of the redshift factor at ℓ. Thus, the amount of work done at infinity in lowering the box is

$$W_\ell = (1 - \chi) E_0$$

(Recall that we assumed the box to be so small that χ does not differ appreciably from top to bottom.) This is not standard thermodynamical work, as the volume of the fluid, as measured by a co-moving observer, has not changed. It is rather work done by "the gravity of the black hole".

Now, when the box is held at the proper distance ℓ from the black hole and the piston slowly pushes or pulls so as to change the temperature of the fluid from T_0 to T_1 (as measured locally), the piston does work (as measured at infinity)

$$W_1 = \chi(E_0 - E_1)$$

where E_1 is the locally measured total energy of the fluid after the fluid's (locally measured) volume has been changed by the piston. When the fluid has reached the desired temperature T_1, the box is opened and the piston pushes the fluid quasi-statically out of the box, so it will fall into the black hole; in the process, the piston does work W_2 (as measured at infinity).[37] Now, by the First Law, the total amount of energy the fluid has as it leaves the box is

$$E_1 - \frac{W_2}{\chi} = T_1 S_{\mathrm{b}} \tag{5.2.2}$$

as measured locally.

In order to compute the total amount of energy and the total amount of heat dumped into the black hole as measured at infinity, we must compute the temperature of the box as measured from there. It is a standard result (Tolman 1934, p. 318) that the condition for a body at locally measured temperature T to be in thermal equilibrium in a strong, nearly static gravitational field is that the temperature measured "at infinity" be χT. Thus the temperature of the box as measured from infinity will be χT_1. It follows from equation (5.2.2), therefore, that the total amount of heat dumped into the black hole is

$$\chi T_1 S_{\mathrm{b}} = \chi E_1 - W_2$$

[37] One may worry that this process cannot be quasi-static, not even in principle, in so far as the phase-space volume available to the fluid as it is expelled from the box and before it is absorbed by the black hole is, in principle, unbounded, *i.e.*, the entropy of the fluid increases by an arbitrary amount. A superficial, but I think still adequate, answer to this problem is that one can arrange a telescopically extending mechanism from the box to the black hole to ensure that the volume available to the fluid never changes. A deeper and I think more satisfying answer is that, when the fluid passes the event horizon, as all of it must do, its available phase-space volume only decreases, and arbitrarily so. I thank Tim Maudlin for pushing me on this point.

But $\chi E_1 = \chi E_0 - W_1$ and $\chi E_0 = E_0 - W_\ell$, so

$$\chi T_1 S_{\mathrm{b}} = E_0 - W_\ell - W_1 - W_2$$

The expression on the righthand side of the last equation, however, is just the total amount of energy in the box as measured at infinity, and so $\chi T_1 S_{\mathrm{b}}$ is the total amount of energy the black hole absorbs, as measured from infinity, which is entirely in the form of heat.

Now, because we have assumed that the entropy for the fluid and for the black hole is additive, the total change in entropy is

$$\Delta S = -S_{\mathrm{b}} + \frac{\chi T_1 S_{\mathrm{b}}}{T_{\mathrm{BH}}}$$

For the process to be isentropic,

$$\Delta S = 0$$

and so

$$\frac{\chi T_1 S_{\mathrm{b}}}{T_{\mathrm{BH}}} = S_{\mathrm{b}} \tag{5.2.3}$$

It follows immediately that $T_1 = \frac{T_{\mathrm{BH}}}{\chi}$, precisely the temperature one would expect for a thermally equilibrated body in thermal contact with another body at temperature T_{BH} a redshift distance χ away. Write Q_{BH} for the amount of heat the black hole absorbs ($= \chi T_1 S_{\mathrm{b}}$), so equation (5.2.3) becomes

$$\frac{Q_{\mathrm{BH}}}{T_{\mathrm{BH}}} = S_{\mathrm{b}}$$

Now, in the limit as the box, and so the heat and entropy it contains, becomes very small (while the temperature remains constant), we may think of this as an equation of differentials,

$$\frac{\mathrm{d}Q_{\mathrm{BH}}}{T_{\mathrm{BH}}} = \mathrm{d}S_{\mathrm{b}} \tag{5.2.4}$$

This expresses the well known fact that temperature plays the role of an integrating factor for heat. Since $\mathrm{d}Q_{\mathrm{BH}}$ is the change in mass of the black hole, $\mathrm{d}M_{\mathrm{BH}}$, due to its being the entirety of the energy absorbed, there follows from the First Law of black-hole mechanics[38]

$$\frac{8\pi \mathrm{d}Q_{\mathrm{BH}}}{\kappa} = \mathrm{d}A \tag{5.2.5}$$

Thus, κ is also an integrating factor for heat. It is a well known theorem that if two quantities are both integrating factors of the same third quantity, the ratio of the two must be a function of the quantity in the total differential, and so in this case

$$\frac{T_{\mathrm{BH}}}{\kappa} = \psi(A) \tag{5.2.6}$$

[38] At least two conceptually distinct formulations of the First Law of black-hole mechanics appear in the literature, what (following Wald 1994, ch. 6, §2) I will call the physical-process version and the equilibrium version. The former fixes the relations among the changes in an initially stationary black hole's mass, surface gravity, area, angular velocity, angular momentum, electric potential and electric charge when the black hole is perturbed by throwing in an "infinitesimally small" bit of matter, after the black hole settles back down to stationarity. The latter considers the relation among all those quantities for two black holes in "infinitesimally close" stationary states, or, more precisely, for two "infinitesimally close" black-hole spacetimes. Clearly, I am relying on the physical-process version, for the most thorough and physically sound discussion and proof of which see Wald and Gao (2001).

for some ψ. (It is also the case that $\frac{T_{\text{BH}}}{\kappa} = \phi(S_{\text{b}})$ for some ϕ, but we will not need to use that.) It follows from equations (5.2.4) and (5.2.5) that

$$\frac{1}{8\pi}\psi(A)\mathrm{d}A = \mathrm{d}S_{\text{b}} \tag{5.2.7}$$

and so integrating this equation yields the change in the black hole's area, ΔA as a function of S_{b}, say $\Delta A = \theta(S_{\text{b}})$. (From hereon, we fix some arbitrary standard value for A, and so drop the 'Δ'.)

In order to complete the argument, and make explicit the relation between A and S_{b}, and at the same time fix the relation between κ and T_{BH}, consider two black holes very far apart, and otherwise isolated, so there is essentially no interaction between them. Perform the Geroch-Carnot cycle on each separately. Let A_1 and A_2 be their respective areas, θ_1 and θ_2 the respective functions for those areas expressed using $S_{\text{b}1}$ and $S_{\text{b}2}$, the respective entropies dumped into the black holes by the cycles, and let $\theta_{12}(S_{\text{b}12})$ be the function for the total area of the black holes considered as a single system, expressed using the total entropy $S_{\text{b}12}$ dumped into the system. Both the total area of the black holes and the total entropy dumped in are additive (since the black holes, and so the elements of the Carnot-Geroch cycles, have negligible interaction), *i.e.*,

$$\theta_1(S_{\text{b}1}) + \theta_2(S_{\text{b}2}) = \theta_{12}(S_{\text{b}12}) = \theta_{12}(S_{\text{b}1} + S_{\text{b}2})$$

Differentiate each side, first with respect to $S_{\text{b}1}$ and then with respect to $S_{\text{b}2}$; because θ_{12} is symmetric in $S_{\text{b}1}$ and $S_{\text{b}2}$,

$$\frac{\mathrm{d}\theta_1}{\mathrm{d}S_{\text{b}1}} = \frac{\mathrm{d}\theta_2}{\mathrm{d}S_{\text{b}2}}$$

Since the parameters of the two black holes and the two cycles are arbitrary, it follows that there is a universal constant α such that

$$\frac{\mathrm{d}\theta}{\mathrm{d}S_b} = \frac{\mathrm{d}A}{\mathrm{d}S_b} = \alpha$$

for all Schwarzschild black holes. It now follows directly from equations (5.2.6) and (5.2.7) that

$$T_{\text{BH}} = 8\pi\alpha\kappa$$

and from equation (5.2.3) that

$$S_{\text{BH}} = \frac{A}{\alpha}$$

up to an additive constant we may as well set equal to zero.[39] α is guaranteed by construction to have the proper dimensions to give T_{BH} the physical dimension of temperature (mass, in geometrized units), and S_{BH} the physical dimension of entropy (dimensionless, in geometrized units).

As a consistency check, it is easy to compute that the total work performed in the process,

$$W_T = W_0 + W_\ell + W_1 + W_2$$

[39] In contradistinction to classical thermodynamical systems, geometrized units for the entropy of black holes can be naturally constructed: let a natural unit for mass be, say, that of a proton; then one unit of entropy is that of a Schwarzschild black hole of unit mass. Why does classical black-hole thermodynamics allow for the construction of a natural unit for entropy when purely classical, non-gravitational thermodynamics does not?

equals the total change in heat of the box during the process, $Q_\mathrm{b} - \chi T_1 S_\mathrm{b}$, exactly as one should expect for a Carnot cycle. One can use the total work, then, to define the efficiency of the process in the standard way,

$$\eta := \frac{W_T}{Q_\mathrm{b}} = 1 - \frac{\chi T_1 S_\mathrm{b}}{Q_\mathrm{b}}$$

from which it follows that

$$\eta = 1 - \frac{8\pi\alpha\kappa}{T_0}$$

Thus, one can use the standard procedure for defining an absolute temperature scale based on the efficiency of Carnot cycles, and one concludes that the absolute temperature of the black hole is indeed $8\pi\alpha\kappa$.

Unfortunately, one cannot use similar arguments as in the classical case to prove the analogue of theorem 4.1, as the Carnot-Geroch Cycle for Schwarzschild black holes is not reversible in the physical sense. Under restricted conditions, however, the Carnot-Geroch cycle for Kerr black holes *is* physically reversible, and so in that case one can use the classical arguments to prove the analogue of theorem 4.1, as I plan to discuss in future work (Curiel 2014a).

5.3 The Clausius and Kelvin Postulates for Black Holes

Although I consider the construction of the Carnot-Geroch Cycle and the arguments based on it to be the most decisive in favor of conceiving of classical black holes as truly thermodynamical objects, I think it is still worthwhile to show that the appropriately translated analogues of the Clausius and Kelvin Postulates hold for black holes as well. Because those Postulates provide the ground for all ways of introducing temperature and entropy in classical thermodynamics, to show that they hold of black holes as well will show that the physical behavior of black holes conforms as closely as possible to that of classical thermodynamical in all fundamental respects.

The standard arguments in favor of the Clausius and Kelvin postulates (as given, *e.g.*, in Fermi 1956, ch. 3), which rely on the impossibility of constructing a *perpetuum mobile* of the second kind, do not translate straightforwardly into the context of general relativity, where there is no general principle of the conservation of energy. Remarkably enough, however, as with the Second Law, both Postulates follow as theorems of differential geometry.

Theorem 5.3.1 (Clausius Postulate for Black Holes) *A transformation whose only final result is that a "quantity of heat" (as defined in §5.1) is transferred from a black hole at a given surface gravity to a system at a higher temperature is impossible.*

Assume that such a transformation as described in the antecedent of the theorem were possible. Then the change in irreducible mass of the black hole would have to be strictly greater than the change in its total mass during the interaction, with no other change in the spacetime than that another system absorbed heat. In particular, its irreducible mass must increase. However, it follows from equation (5.1.1) and the Second Law that an increase in irreducible mass must yield an increase in the black hole's area, and so its entropy, violating the assumption that nothing else thermodynamically relevant in the spacetime changed.

Theorem 5.3.2 (Kelvin Postulate for Black Holes) *A transformation whose only final result is that a "quantity of heat" (as defined in §5.1) is extracted from a stationary black hole and transformed entirely into work is impossible.*

The argument is essentially the same as for the Clausius Postulate for black hole. Again, for such a process to occur, the irreducible mass of the black hole would have to increase, but that would necessitate a change in the area of the black hole, violating the conditions of the theorem.

6 Problems, Possible Resolutions, Possible Insights, and Questions

I conclude the paper with a brief discussion of some *prima facie* problems with my arguments, suggestions for their resolutions, an examination of what insights my conclusions, if correct, may offer, and some general questions that I think need to addressed, possibly with the help of my arguments and conclusions.

As is well known, the surface gravity κ is well defined only for stationary black holes; does this mean that my analysis cannot apply to non-stationary black holes? Yes, it is the case that my analysis cannot apply to non-stationary black holes, but that is no problem. Non-stationary black holes are ones out of equilibrium, and so this presents the same situation as obtains in classical equilibrium thermodynamics. I think we often forget that, strictly speaking, temperature in ordinary thermodynamics is well defined only for bodies in (or quite close to) thermal equilibrium. One way to see this is to note that, for systems far from equilibrium, different kinds of thermometric device will return very different readings, as fine details of their different couplings to the system which are negligible for equilibrium become non-trivial, in particular phenomena manifesting themselves at temporal and spatial scales below the hydrodynamic scale.[40]

Another problem is that it seems as though we can attribute heat to a Schwarzschild black hole only when it is being perturbed. Again, the situation is in fact much the same as in classical thermodynamics, wherein it never makes sense to attribute a definite quantity of heat to an isolated system in equilibrium. The only definite claims we can make, as Maxwell himself so insightfully and eloquently discussed (footnote 30), are about the quantification of heat *transfer*. In any event, one *can* extract both "heat" and work from a Schwarzschild black hole by perturbing it; indeed, this is in excellent analogy with ordinary thermodynamical systems that have reached heat death, from which heat and work can be extracted only if one perturbs them properly. In fact, the analogy is even better than that brief remark suggests: stationary classical black holes do not "radiate heat", but neither do ordinary classical thermodynamical systems in equilibrium; classical systems exchange heat only when they are in direct contact (contiguous) with another system at a different temperature, but the same holds for stationary classical black holes, in so far as their immediately

[40] See, *e.g.*, Benedict (1969, §§4.1–4.4, pp. 24–9). This reference is not the most up-to-date with regard to the international agreement on defining the standard, practical methods for the determination of temperature, but I have found no better reference for the nuts and bolts of thermometry. See Curiel (2005) for a discussion of the details.

contiguous environment is "at the same temperature", *viz.*, has essentially the same effective surface gravity as measured at infinity, as the black hole does. Still, one may protest, in the construction of the Carnot-Geroch Cycle, I ignored perturbations to the black hole from the lowering of the box, so how can one say, given my definitions and arguments, that energy was extracted from it? Given the assumption that the total energy of the box is negligible compared to the mass of the black hole, I claim it is a good approximation to ignore any perturbations to the black hole while still accounting for the (relatively negligible) amount of energy the box gains by being lowered through the black hole's "gravitational field".

Another potential problem: it is clear that black holes have, by the standard definition, negative specific heat, since their surface gravity decreases as their mass-energy increases. Standard arguments, however, conclude that two bodies with negative specific heat cannot thermally equilibrate. There is, though, a hidden assumption in the standard arguments, to wit, "conservation of heat"—it is always assumed, that is to say, that for two bodies in thermal contact one can gain heat only if the other loses it, and that in the same amount. Heat, however, is not a substance, as everyone from Maxwell (1888) to Planck (1926) to Sommerfeld (1964) is at pains to emphasize, and so obeys no conservation law. There is no reason why two bodies with different temperatures in thermal contact cannot both "gain or lose heat from or to each other" at the same time. When two black holes in quasi-stationary orbit[41] about each other equilibrate, the temperatures of both bodies simultaneously decrease as they both gain heat from the other, the one of higher temperature decreasing more quickly than the other, so they will eventually reach the same temperature.

A potentially more serious problem with my analysis is that it is difficult to see what sense can be made of "exchange" between a global energetic quantity (in the case of stationary, asymptotically flat black holes, ADM mass) on the one hand, and localized stress-energy of ordinary systems on the other. A more poignant way of posing the problem is to note that gravitational energy is strictly non-local in the precise sense that there is no such thing as a gravitational stress-energy tensor (Curiel 2013), and so it satisfies no general conservation law. How, then, can one talk about exchange for such a *recherché* quantity?[42] There are, I think, two responses to this problem, one stronger than the other. The first, weaker, response is that one always has in place a quasi-local notion of mass-energy in stationary and axisymmetric spacetimes, which suffices for the purposes of my arguments, just as it does in Newtonian gravitational theory (*à la* the "Poynting integral" of Bondi 1961). The stronger response, which is more to the point, is that heat is not a localized form of energy in classical thermodynamics either—it is not a perfect differential (as the discussion of Sommerfeld 1964 makes particularly clear), and so it also has no corresponding conservation law—just like gravitational energy—and yet we feel no inconsistency in talking there about exchange of energy for a quantity that can be represented only as a total magnitude, with no corresponding localized density. Sauce for the goose is surely sauce for the gander.

My arguments, I think, have not only residual problems; they also open the possibility for

[41] There are no solutions to the Einstein field equation representing two Kerr black holes in stable orbit about each other (Manko and Ruiz 2001).

[42] I thank Jim Weatherall for pushing me on this point.

real insight into existing questions about black-hole mechanics and thermodynamics. Although the following is not a problem peculiar to my analysis, it is a general one in the field I believe my analysis can give some insight into. Black holes have enormous entropy, far more than any reasonably conceivable material system that could form them on collapse (Penrose 1979). There must, therefore, be a correspondingly enormous and discontinuous jump in entropy when a collapsing body passes the point at which an event horizon forms. How can one explain that? It is here that I believe my old-fashioned approach to entropy bears some of its sweetest fruit. More modern characterizations of entropy, whether of a Boltzmannian, Gibbsian, von-Neumann-like, or Shannon-like form, have no explanation for this jump. If, however, one conceives of entropy as a measure of how much work it takes to extract energy from a system, how much free energy a system has, what forms its internal energy (as opposed to free energy) are in, then black holes have enormous energy, only a very small amount of which is extractible, and there is a clear physical discontinuity in extractability of energy when an event horizon forms.

I leave the reader with some questions concerning this entire field that, though not peculiar to my arguments here, I feel strongly need to be investigated further by both philosophers and physicists. The Laws of thermodynamics are empirical generalizations, indeed, the paradigm of such. I know of no other propositions in physics whose support comes *entirely* from experimental evidence, with not even the suggestion of the possibility of a formal derivation from "deeper" physical principles. Also, I know of no other propositions, with the possible exception of the Newtonian inverse-squared distance dependence of gravitational attraction between two bits of matter, that are more deeply entrenched than the Laws of thermodynamics. But, entirely to the contrary, and with the exception only of the Third Law (which is also the most weakly supported by experimental evidence in classical thermodynamics), all the Laws of black-hole mechanics are theorems of differential geometry. They require no input from physical theory at all. One will sometimes see the claim that one or the other of the Laws requires the assumption of the Einstein field equation, but this is not true: all the Laws are independent of the Einstein field equation in the strong sense that one can assume its negation and still derive the Laws; the Einstein field equation enters only when one wants to give a physical interpretation of the quantities involved by way of its asserted relation between the Ricci tensor and the stress-energy tensor of matter.[43] And yet, against what most philosophers would naively expect (and philosophers, as a group, are nothing if not naive when it comes to what counts as real substantiation and confirmation of physical propositions), the Laws of thermodynamics are profoundly more entrenched than those of black-hole mechanics. So what the hell is going on here?

References

Aizenman, M. and E. Lieb (1981). The Third Law of thermodynamics and the degeneracy of the ground state for lattice systems. *Journal of Statistical Physics 24*(1), 279–297. doi:10.1007/BF01007649.

[43] See Curiel (2014b) for a thorough discussion.

Bardeen, J., B. Carter, and S. Hawking (1973). The four laws of black hole mechanics. *Communications in Mathematical Physics 31*(2), 161–170. doi:10.1007/BF01645742.

Benedict, R. (1969). *Fundamentals of Temperature, Pressure and Flow Measurements*. New York: John Wiley & Sons, Inc.

Bondi, H. (1961). [*** something on em energy transfer ***]. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 261*, 1.

Bredberg, I., C. Keeler, V. Lysov, and A. Strominger (2011). From Navier-Stokes to Einstein. arXiv:hep-th/1101.2451v2.

Brown, H. and J. Uffink (2001, December). The origins of time-asymmetry in thermodynamics: The minus first law. *Studies in History and Philosophy of Modern Physics 32*(4), 525–538. doi:10.1016/S1355-2198(01)00021-1.

Carathéodory, C. (1909). Untersuchungen über die Grundlagen der Thermodynamik. *Mathematische Annalen 67*(3), 355–386. doi:10.1007/BF01450409.

Carter, B. (1973). Black hole equilibrium states. In B. DeWitt and C. DeWitt (Eds.), *Black Holes*, pp. 56–214. New York: Gordon and Breach.

Christodoulou, D. (1970). Reversible and irreversible transformations in general relativity. *Physical Review Letters 25*(22), 1596–1597. doi:10.1103/PhysRevLett.25.1596.

Clifton, T., G. Ellis, and R. Tavakol (2013). A gravitational entropy proposal. *Classical and Quantum Gravity 30*(12), 125009. doi:10.1088/0264-9381/30/12/125009. Preprint available at arXiv:1303.5612 [gr-qc].

Curiel, E. (2005). *On the Formal Consistency of Theory and Experiment in Physics*, Volume 2, Chapter 3. Chicago: University of Chicago. Doctoral Dissertation: *Three Papers on How Physics Bears on Philosophy, and How Philosophy Bears on Physics*.

Curiel, E. (2013). On tensorial concomitants and the non-existence of a gravitational stress-energy tensor. This paper has been submitted for review at *Journal of Mathematical Physics*.

Curiel, E. (2014a). Carnot-Geroch cycles for Kerr black holes. Unpublished manuscript.

Curiel, E. (2014b). A primer on energy conditions. Preparing to submit to *British Journal of Philosophy of Science*.

Fermi, E. (1937[1956]). *Thermodynamics*. Dover Publications, Inc. The Dover 1956 edition is an unabridged, unaltered republication of the 1937 Prentice-Hall edition.

Hawking, S. (1974, 01 March). Black hole explosions? *Nature 248*, 30–31. doi:10.1038/248030a0.

Hawking, S. (1975). Particle creation by black holes. *Communications in Mathematical Physics 43*, 199–220. doi:10.1007/BF02345020.

Hayward, S. (1994). General laws of black hole dynamics. *Physical Review D 49*(12), 6467–6474. doi:10.1103/PhysRevD.49.6467. Preprint available at arXiv:gr-qc/9303006v3.

Hollands, S. and R. Wald (2012). Stability of black holes and black branes. arXiv:1201.0463v4 [gr-qc].

Israel, W. (1973). Entropy and black hole dynamics. *Lettere al Nuovo Cimento 2*, 267–269.

Israel, W. (1986). Third Law of black hole mechanics: A formulation of a proof. *Physical Review Letters 57*(4), 397–399. doi:10.1103/PhysRevLett.57.397.

Jacobson, T. (1995, 14 August). Thermodynamics of spacetime: The Einstein equation of state. *Physical Review Letters 75*(7), 1260–1263. doi:10.1103/PhysRevLett.75.1260. Preprint available at arXiv:gr-qc/9504004v2.

Jacobson, T. (2003). Introduction to quantum fields in curved spacetime and the Hawking effect. arXiv:gr-qc/0308048v3.

Lysov, V. and A. Strominger (2011). From Petrov-Einstein to Navier-Stokes. arXiv:hep-th/1104.5502v2.

Manko, V. and E. Ruiz (2001). Exact solution of the double-Kerr equilibrium problem. *Classical and Quantum Gravity 18*, L11–L15. doi:10.1088/0264-9381/18/2/102.

Maxwell, J. C. (1888). *Theory of Heat*. Mineola, NY: Dover Publications, Inc. The Dover edition of 2001 republishes in unabridged form the ninth edition of 1888 published by Longmans, Green and Co., London, and also includes the corrections and notes of Lord Rayleigh incorporated into the edition of 1891.

Penrose, R. (1979). Singularities and time-asymmetry. In S. Hawking and W. Israel (Eds.), *General Relativity: An Einstein Centenary Survey*, pp. 581–638. Cambridge University Press.

Planck, M. (1926). *Thermodynamics*. Dover Publications, Inc. The Dover reprint of the third English edition of 1926, translated by A. Ogg from the 7th German edition of 1922.

Schrödinger, E. (1960). *Statistical Thermodynamics* (Second ed.). Cambridge: Cambridge University Press. A course of seminar lectures, delivered in January–March 1944, at the School of Theoretical Physics, Dublin Institute for Advanced Studies.

Sciama, D. (1976). Black holes and their thermodynamics. *Vistas in Astronomy 19, Part 4*, 385–41. doi:10.1016/0083-6656(76)90052-0.

Sommerfeld, A. (1964). *Thermodynamics and Statistical Mechanics*, Volume V of *Lectures on Theoretical Physics*. New York: Academic Press. Trans. J. Kestin. Edited and posthumously completed by F. Bopp and J. Meixner.

Tolman, R. (1934). *Relativity, Thermodynamics and Cosmology*. New York City: Dover Publications, Inc. A 1987 facsimile of the edition published by the Oxford University Press, at Oxford, 1934, as part of the International Series of Monographs on Physics.

Unruh, W. and R. Wald (1982). Acceleration radiation and the Generalized Second Law of thermodynamics. *Physical Review D 25*(4), 942–958. doi:10.1103/PhysRevD.25.942.

Wald, R. (1984). *General Relativity*. Chicago: University of Chicago Press.

Wald, R. (1994). *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago: University of Chicago Press.

Wald, R. (1997). 'Nernst theorem' and black hole thermodynamics. *Physical Review D 56*(10), 6467–6474. doi:10.1103/PhysRevD.56.6467. Preprint available at arXiv:gr-qc/9704008.

Wald, R. (1999). Gravitation, thermodynamics and quantum theory. arXiv:gr-qc/9901033.

Wald, R. and S. Gao (2001). "Physical process version" of the First Law and the Generalized Second Law for charged and rotating black holes. *Physical Review D 64*(8), 084020. doi:10.1103/PhysRevD.64.084020.