VI

DISMANTLING PLANCK'S BLACK-BODY THEORY: EHRENFEST, RAYLEIGH, AND JEANS

If Planck's *Lectures on the Theory of Thermal Radiation* exemplifies classical radiation theory, then it is the culmination of that tradition. In 1905, the year before the book appeared, Jeans had argued and Einstein had remarked in passing that the only radiation law compatible with classical theory was very different from either Planck's or Wien's. That claim could be and was readily set aside, most explicitly by Rayleigh, from whose earlier work Jeans's argument derived. But the status of what has since usually been called the Rayleigh–Jeans law nevertheless began to change during 1906. Ehrenfest then argued that Planck's resonators could not fulfill the primary function for which they had been developed, a point Planck conceded in a strange paragraph added, probably in proof, to the end of the *Lectures*. Properly conducted, Ehrenfest continued, resonator theory should yield the same result as Jeans's theory of the behavior of radiation in an otherwise empty cavity. To achieve Planck's result, conservation of energy and of the number of vibrators or vibration modes would have to be supplemented by some additional constraint, foreign to classical theory. One such constraint, Ehrenfest pointed out in closing, would be to restrict the energy of each vibration mode to integral multiples of the energy element $h\nu$. Einstein, by a very different route, arrived simultaneously at a far stronger and more restrictive proof of the same result. Previously, Einstein declared, he had thought Planck's theory incompatible with the light-particle hypothesis he had himself introduced during the preceding year. Now he realized that, properly understood, Planck's theory required that hypothesis.

Though Einstein and Ehrenfest doubtless helped here and there to prepare the way for a new attitude towards the significance of Planck's work, only Max von Laue (1879–1960) appears to have found their analysis of Planck's theory convincing from the start. But conviction did follow rapidly in the aftermath of Lorentz's presentation of a new proof of Jeans's result in the early spring of 1908. By the end of the following year, Lorentz, Wien, and Planck himself had been persuaded that radiation theory required discontinuity. Arnold Sommerfeld (1868–1951) and Jeans, among others, were at least moving towards that position during 1910, the year in which Lorentz provided a particularly effective and widely read set of arguments for it. By the years 1911 and 1912, with which this volume closes, all or virtually all those physicists who had devoted significant attention to cavity radiation were persuaded that it demanded some Planck-like theory, which would, in turn, require the development of a discontinuous physics. Though no one claimed to know what the shape of the new physics would be, the men concerned all recognized that there could be no turning back.

This chapter traces the way in which Ehrenfest's recourse to the work of Rayleigh and Jeans prepared the way for recognition of the central role of energy discontinuity in Planck's theory. The next considers the breathtaking series of papers that led Einstein to insist in 1906 that discontinuity was the fundamental prerequisite for Planck's success. Chapter VIII then describes both the events that led Lorentz to embrace a discontinuous radiation theory during 1908 and also the impact of his newfound conviction on other physicists, especially Planck and Wien. The final chapter of Part Two then examines the status of the quantum discontinuity during 1911 and 1912, attempting in the process to set black-body theory in the context of the other applications that had meanwhile been proposed for the quantum. It suggests that by 1912 physicists had learned very nearly all they could from the examination of cavity radiation and that the leading edge of quantum research had suddenly shifted to the previously neglected problem of the specific heat of solids. Finally, a brief epilogue on Planck's so-called second theory concludes the volume.

The origin of the Rayleigh-Jeans law, 1900–1905

The claim that black-body radiation should conform to the distribution law that has since been variously attributed to Rayleigh and Jeans was not made until 1905. But the main conceptual foundations for that claim can be found in a two-page note published by Rayleigh in the June 1900 issue of the *Philosophical Magazine.*[1] Commenting on

current attempts to discover "the law of complete radiation," Rayleigh called attention to a fundamentally implausible characteristic of the Wien distribution law, recently rederived by Planck and still generally supported by experiment. The Wien law, he said, makes the density of radiant energy at a particular wavelength proportional to $\lambda^{-5}\,e^{-a/\lambda T}$. If it were correct, then energy would no longer increase with temperature when λT was large compared with the known constant a. Furthermore, Rayleigh pointed out, the cutoff should occur within an accessible experimental range. If the Wien-Planck law held, the infrared intensity at 60 μ recently studied by Rubens should not increase further for temperatures above 1000° absolute. No such cutoff seemed likely, and Rayleigh therefore proposed that the law be modified. At the end of an argument partly theoretical and partly ad hoc, he suggested replacing the factor λ^{-5} in the Wien–Planck law with $\lambda^{-4}T$, yielding a distribution law with the form

$$u = b\lambda^{-4}T\,e^{-a/\lambda T}. \tag{1}$$

That proposal satisfied the displacement law and also permitted intensity to increase with temperature at all wavelengths.

Rayleigh's 1900 note is both cryptic and incomplete. For that reason and to avoid functionless repetition, the first part of the following sketch is based on the fuller argument he developed in 1905.[2] Completeness aside, the latter differs from its predecessor only in specifying the value of the proportionality constant in the distribution law. Though Rayleigh's expression for that constant was smaller by a factor of eight than the one derived below, the error was at once corrected by Jeans and acknowledged by its author.[3]

As an expert on the theory of sound, Rayleigh chose to represent the electromagnetic field in a cavity by the vibrations of an elastic medium, initially a vibrating string. If a string, fixed at both ends, has a length L, it can vibrate only in modes of wavelength $\lambda = 2L/k$, where $k = 1, 2, 3, \ldots$. The same condition applies to the vibration of an elastic fluid, for example air, in a rigid cubical box of side L, except that permissible modes are governed by three integers, k, l, m, with $\lambda = 2L/\sqrt{k^2 + l^2 + m^2}$. If each triple of integers k, l, m represents a point at distance $R\,(= \sqrt{k^2 + l^2 + m^2})$ from the origin, then the number of modes lying in the wavelength range λ to $\lambda + d\lambda$ is given by the number of points lying in the first octant of a spherical shell contained between radii R and $R + \mathrm{d}R$. Since $\lambda = 2L/R$, the number of modes must be

$$\frac{4\pi}{8} R^2 \, \mathrm{d}R = \frac{4\pi L^3}{\lambda^4} \, \mathrm{d}\lambda. \tag{2}$$

As represented in equation (2), Rayleigh's count of modes yields the λ^{-4} in the formula toward which he was proceeding. To discover the dependence of energy density on temperature, he appealed next to a very different set of considerations, which lead to what he called "the Maxwell–Boltzmann doctrine of the partition of energy." Subsequently better known as the equipartition theorem, it specifies that in any mechanical system each degree of freedom will on the average possess the same kinetic energy. In terms of the constants Planck had introduced in January 1901, furthermore, that energy must be just $\frac{1}{2}kT$, i.e., one-third the total translatory energy of a molecule. If the theorem applies to the vibrations of an elastic solid, then each of the distinct modes enumerated by equation (2) corresponds to one of the infinite degrees of freedom of the vibrating medium. It should possess mean kinetic energy $\frac{1}{2}kT$ and mean total energy kT, since the average kinetic and potential energies of a linear vibrator are equal. Finally, for the electromagnetic case involving transverse vibrations, the number of modes indicated by equation (2) must be doubled to permit two independent states of polarization, and energy kT must then be attributed to each of the resulting modes. With the resulting energy for all modes divided by L^3 the energy per unit volume in the range λ to $\lambda + \mathrm{d}\lambda$ is

$$u_\lambda \, \mathrm{d}\lambda = \frac{8\pi kT}{\lambda^4} \, \mathrm{d}\lambda \left(= \frac{8\pi\nu^2}{c^3} kT \, \mathrm{d}\nu \right). \tag{3}$$

That equation, usually in its wavelength form, is what came after 1905 to be known as the Rayleigh-Jeans law, but it is not the law proposed in 1900 by Rayleigh. Long a critic of the equipartition theorem, which he thought valid only under restricted conditions, Rayleigh remarked early in his note that any general treatment of equilibrium energy relations "is hampered by the difficulties which attend the Maxwell–Boltzmann doctrine of the partition of energy." He also suggested, however, that, "although for some reason not yet explained the doctrine fails in general, it seems possible that it may apply to the graver modes." By "graver modes" he meant long wavelength vibrations, to which alone he thought a form like equation

(3) might apply. To obtain an equation that could apply also to the shorter wavelengths at which equipartition broke down, he suggested that the factor $\lambda^4 T$ from equation (3) be multiplied by the exponential term from the Wien distribution law. The result was equation (1), and Rayleigh's note concluded with the hope that the agreement of this quasi-empirical law with observation "may soon receive an answer at the hands of the distinguished experimenters who have been occupied with this subject."

The "distinguished experimenters" learned of Rayleigh's proposal with surprising speed, and some of them responded.[4] Lummer and his colleague E. Jahnke (1863–1921) cited the proposed new law in a paper submitted to the *Annalen der Physik* late in July 1900.[5] Rubens shortly found that for large values of λT the energy density of radiation did increase with temperature as Rayleigh's law required. Though it probably played no significant role in his work, Planck knew of that result before he presented his own newly invented law to the Physical Society in October. But he soon also knew the result of experiments for smaller values of λT. In a paper submitted to the Academy of Sciences late in the month, Rubens and Kurlbaum compared a number of proposed radiation formulas with their data and concluded that Rayleigh's was satisfactory only in the limit where it coincided with Planck's.[6] Since the law, as proposed, was almost totally ad hoc, there was no further reason to take it seriously. Less than six months after it had been suggested, it was set aside. Even the usually scrupulous Planck made no mention of Rayleigh's contribution until 1906, by which time a very different set of concerns had led to the rederivation, not of Rayleigh's law, but of equation (3), originally a step on the way to it.

What led physicists back to Rayleigh's mode count was not the black-body problem at all but anomalies in the specific heat of gases.[7] If equipartition were admitted, then kinetic theory predicted that γ, the ratio of the specific heat at constant pressure to that at constant volume, should be given by $(n + 2)/n$, with n the number of degrees of freedom of a gas molecule. To bring that result into agreement with experiment it was necessary to suppose that molecules were composed of atomic mass-points rigidly connected, and even the agreement achieved in that way was far from satisfactory. In addition, it was difficult to conceive of atoms as dimensionless points, especially because the existence of spectra suggested that they must have a complex internal structure. Since there were many observed spectral lines, each presumably corresponding to a separate degree of freedom, n ought to be very large and γ close to unity, which it was not for any known gas.

The problem was an old one, and Boltzmann himself had commented on it in the well-known letter he had sent to *Nature* in 1895:

> But how can the molecules of a gas behave as rigid bodies? Are they not composed of smaller atoms [the vibrations of which cause radiation through the ether]? Probably they are; but the *vis viva* of their internal vibration is transformed into progressive and rotatory motion so slowly, that when a gas is brought to a lower temperature, the molecules may retain for days, or even for years, the higher *vis viva* of their internal vibrations corresponding to the original temperature. This transference of energy, in fact, takes place so slowly that it cannot be perceived amid the fluctuations of temperature of the surrounding bodies. The possibility of the transference of energy being so gradual cannot be denied, if we also attribute to the ether so little friction that the earth is not sensibly retarded by moving through it for many hundreds of years.[8]

Despite its plausibility, no one attempted to explore Boltzmann's idea until 1901. Then it was taken up or redeveloped by James Jeans, who thereafter worked on the subject for a number of years. In his first paper Jeans referred to the possibility of interaction between molecules and the ether, and he then continued: "That such an interaction must exist is shown by the fact that a gas is capable of radiating energy." On that assumption he undertook to demonstrate that

> a slight deviation from perfect conservation of energy [among the molecules of a gas] may result in a complete redistribution of the total [molecular] energy, and it will appear that this new distribution of energy will lead to values for the ratios of the two specific heats which are not open to the objections mentioned above.[9]

Jeans's technical argument, which has no relevance here, was expanded and refined at considerable length in Chapters 8 through 10 of the first edition of his well-known *Dynamical Theory of Gases*, published in 1904. Those chapters constituted the most original, if also the least enduring portion of his book, and Jeans continued to develop the theory they presented after the volume appeared, preparing two major papers on the interaction between matter and ether for publication in 1905.[10] In the first of these he redeveloped Rayleigh's method of mode counting in order to find the number of ether vibrations with periods comparable to or less than the time occupied by a collision between molecules. Only those modes would, he thought, take up energy rapidly; higher modes would respond extremely slowly; and the total energy in the ether could be shown to be very small.

Even before those papers appeared, Rayleigh, in a letter to *Nature*, opened a continuing discussion with Jeans by questioning the methods the latter had used in the *Dynamical Theory of Gases* to reconcile equipartition with observed specific heats.[11] A slow radiation of molecular energy into the ether might, he remarked, have the result Jeans desired if the molecules radiated into empty space. But if ether and molecules were confined together in a perfectly reflecting enclosure, there would be no net dissipation; all the infinitely many modes of ether vibration would receive their share of energy, so that the total energy would be dissipated over the infinity of higher modes. Since black-body measurements showed that nothing of the sort occurred, equipartition must break down for short wavelengths. "A full comprehension here would," Rayleigh concluded, "probably carry with it a resolution of the specific heat difficulty." Equipartition, not interaction with the ether, was, in his view, the source of the difficulty examined by Jeans.

The discussion that followed is neither uninteresting nor entirely one-sided, but only a few of its details are relevant here. To show that equipartition must break down, Rayleigh rederived his previous formula in a second letter, now with the proportionality constant and without the exponential term.[12] (This is the point at which he produced a constant eight times too large, and it is the resulting discrepancy between his formula and the long wavelength limit of Planck's that led him to wonder "how another process also based on Boltzmann's ideas, can lead to a different result."[13]) If the formula he had derived were accurate, Rayleigh pointed out, then a continuous ether, characterized by an infinite number of vibration modes, would absorb infinite energy. An atomic ether, he conceded, might reach equilibrium with finite total energy, but it would nevertheless absorb essentially all the initial translational energy of the molecules of the gas. Jeans responded at once, refusing to acknowledge difficulties with equipartition and insisting that the only possible equilibrium distribution for the energy of the ether was the one he had introduced, equation (3) above.[14] Though he agreed that such a distribution could never be physically realized, Jeans maintained that no paradox need result. Millions of years might, he thought, be required to transmit energy from the lower to the higher modes of ether vibration; a genuine equilibrium might never be achieved. For him, the physical situations studied in black-body experiments were therefore not cases of equilibrium at all. That is the position he redeveloped systematically in a paper read to the Royal Society on 16 November 1905, the first of his technical works in which he transferred the skills developed through research on specific heats to the problem of radiation alone.[15] His concerns and those that motivated Rayleigh's 1900 note had at last begun to intersect.

This debate, conducted in the columns of *Nature* during 1905, is the source of the Rayleigh–Jeans law together with the claim that no other equilibrium distribution of radiant energy can be compatible with classical theory. But Jeans was the only physicist except Einstein to make or accept such a claim at the time, and the strength of his convictions should probably be understood in terms of the subject's relation to his own earlier work on the specific heat of gases. His contributions to the latter subject had been since 1901 a primary source of his still developing professional identity, and the results he had presented during 1904 in his *Dynamical Theory of Gases* were impressive and in most respects plausible. His radiation theory, on the other hand, though apparently required to preserve his work on gases, had a very different status when evaluated separately. Its theoretical basis was in several respects doubtful, and much experimental evidence spoke against it as well.

By 1905 Planck's law was known to be in excellent quantitative agreement with experiment, and nothing in Jeans's argument suggested how that could possibly be the case. Indeed, when Lorentz briefly defended Jeans's position in 1908, experimentalists insisted at once that long-available experimental data permitted its rejection out-of-hand, an episode to be considered in Chapter VIII. Nor was it only experiments on the distribution of radiant energy that became, in principle, inexplicable from Jeans's viewpoint. By denying that these or related experiments dealt with equilibrium situations, Jeans also denied the relevance of thermodynamic arguments to them. New derivations would have to be found for Kirchhoff's law, the Stefan–Boltzmann law, and the Wien displacement law. Like the Rayleigh–Jeans law itself, the old derivations were valid only under unrealizable conditions; they could not apply to any of the situations actually encountered in laboratory research. With respect to the Stefan–Boltzmann law, Jeans acknowledged the difficulty explicitly, and he quickly began to seek a new, non-equilibrium derivation of the displacement law, from which the Stefan–Boltzmann law followed.[16] Though his arguments were of interest and developed with considerable skill, their acceptance would have exacted a high price for very little positive achievement.

Given such grounds for skepticism, Jeans's arguments for his radiation law could readily be set aside. His derivation presupposed the equipartition theorem, and that theorem was, as Rayleigh's position indicates, a source of much question and debate.[17] The standard routes to equipartition depended upon the explicit consideration of the process of collision between complex molecules. The resulting theorem was therefore applicable only to gases, and it, in any case, depended on special hypotheses and approximations which were themselves often doubted. At the beginning of the century those doubts were of special urgency because, as previously noted, both spectra and specific heats appeared to provide direct evidence against the applicability of equipartition even to gases. To consider still more general mechanical systems, including the elastic ether postulated by Rayleigh and Jeans, required some form of ergodic hypothesis: the phase point of the system in question must in time densely fill all the space compatible with the system's constraints. That hypothesis had generally been considered, even by its authors, to be extremely dubious.[18]

Besides, even if equipartition could be shown to apply to general mechanical systems, one might legitimately doubt its applicability to the electromagnetic displacements postulated by Maxwell's theory. No attempt to provide a mechanical model of the ether had succeeded, and physicists had increasingly stopped trying to provide one. The elastic fluid considered by Rayleigh and Jeans was not, in any case, a suitable model for electromagnetic phenomena. Indeed, their model did not even provide a coupling mechanism for the transmission of energy from one mode to the next. It could therefore represent radiation only in a perfectly conducting enclosure, one which was known to preserve the frequency distribution of the radiation initially introduced. How could it provide any basis at all for a theory of black-body radiation?

Planck's model, in contrast, seemed free of such difficulties, at least until the middle of 1906. It deployed Maxwell's equations rather than an elastic fluid ether; it had no apparent recourse to any doctrine like equipartition;[19] and it provided, or seemed to, the coupling mechanism absent in the theory developed by Rayleigh and Jeans. There were important questions about the way Planck's law was to be derived from his model, but they did not at all detract from the adequacy of the model itself. It was apparently founded on secure physical principles, as the Rayleigh–Jeans model clearly was not. Not surprisingly, therefore, when Planck briefly described the Rayleigh–Jeans law in his *Lectures* he treated it as simply "another interesting corroboration of the black-body radiation law for *long wavelengths* as well as of the relation between the radiation constant k and the absolute mass of ponderable molecules."[20] Planck mentioned Jeans's theses that the law represented the only possible equilibrium distribution for all frequencies and that experimental black-body radiation could therefore not be in an equilibrium state, but he quickly dismissed them for many of the reasons outlined above. When that eminently reasonable evaluation was prepared, no one but Jeans himself, and probably also Einstein, would have been inclined to quarrel with it. In any case, the experimentalists who, in 1900, had responded at once to Rayleigh's ad hoc proposal remained for some time entirely silent about Jeans's. They could have refuted it at once, but apparently did not think it worth the effort. The Rayleigh–Jeans law and what came to be called the "ultraviolet catastrophe"[21] did not yet pose problems for more than two or three physicists.

Ehrenfest's theory of quasi-entropies

That the theses first enunciated by Jeans finally did become central to physics is due to their repeated rederivation by a variety of different techniques, often applied to more realistic radiation models, including Planck's. Among such rederivations, the earliest were supplied by Einstein and Ehrenfest. Here we consider Ehrenfest's first, since it resulted from a critical study of Planck's radiation theory, and it made use as well of the earlier work of Rayleigh and Jeans. Einstein's more incisive results were, by contrast, the end product of his independent development, from 1902, of a generalized statistical thermodynamics. Their consideration will occupy the whole of the next chapter.

Paul Ehrenfest first learned of the black-body problem and of Planck's theory from Lorentz during a student visit to Leyden in the spring of 1903.[22] As a student of Boltzmann's he found the subject attractive, a fact attested by a series of notebooks entries made in late June after his return to Vienna.[23] But his notebooks also suggest that more urgent concerns—he completed his dissertation and married during 1904—postponed its persistent pursuit until the spring of 1905. Scattered references to the subject appear from early April of that year and are virtually continuous from mid-June. By the latter date, furthermore, Ehrenfest's attention had focused upon an especially puzzling feature of Planck's work, the one he would discuss in a paper presented to the Vienna Academy in early November. In addition, he had established most of the main elements of the approach to cavity

radiation, that would lead him to the more consequential conclusions he submitted to the *Physikalische Zeitschrift* late in June 1906.[24]

Ehrenfest's principal approach to the study of radiation theory, and especially of Planck's puzzling conclusions, is disguised in his papers but emerges explicitly in his notebooks around the beginning of May 1905. Though he occasionally considered Maxwell's equations or electron theory,[25] he preferred to investigate kinetic or mechanical analogues of the radiation problem, evaluating their adequacy by their success in producing forms equivalent to Kirchhoff's law, the Stefan–Boltzmann law, and the Wien displacement law.[26] One early model, considered in mid-May, involved a weightless leaf spring, loaded with a mass M at one end and forced to vibrate when that mass was struck by molecules of mass m.[27] Other models of the same general type recur until the end of the year, sometimes involving non-uniform or composite elastic strings and sometimes a network of particles connected by coil springs.[28] Though Ehrenfest regularly introduced some mechanism, usually gas molecules, to redistribute energy over the various vibration modes of his model, his problem otherwise closely resembled the mechanical one considered by Rayleigh and Jeans. Apparently, furthermore, he was aware of the resemblance. A first cryptic reference to Jeans appears in his notebooks in mid-July though it was not for another eight months that Jeans's theory became central to his thought.[29]

By mid-June, also, Ehrenfest had discovered an intriguing puzzle in Planck's theory. The earliest hint of its recognition is the phrase, "Noteworthy that [there is] only *one* maximum," which constitutes the initial entry under a heading, "About black radiation."[30] What Ehrenfest had in mind is partially indicated by a series of entries that quickly follow, the first of them immediately after an item dated 18 June: "Planck found many functions that always increase—and accordingly [produce] different stationary states—Clarify!! . . . Boltzmann's H corresponds to entropy in those cases where the latter is defined—but what is the case with Planck? // Planck's stationariness-formula is sufficient to make $dS/dt = 0$ but [is] not necessary!"[31] With only one brief item intervening, Ehrenfest then turned for the first time to the role of "complexion theory" in Planck's work, noting that, even in the absence of any possible recourse to Liouville's theorem, Boltzmann's combinatorial definition permits the introduction of "Abstract Entropy Theory," a phrase he placed in quotation marks and underlined.[32] At about the time he wrote it, he must also have entered into the correspondence with Planck of which the only surviving portion is the letter of 6 July quoted in Chapter V. (Reasons to suppose the exchange was extended will emerge below.) A notebook entry just before 22 July records Ehrenfest's reaction to what Planck there said: "Can one seriously believe that the magnitude of the electronic charge can by itself provide that a pre-established quantum [of] total energy will be pulled about [*zerzaust*] in a certain manner[?]—Try to prove the contrary."[33]

Ehrenfest's notebook entries suggest that, when the last remark was written, his attention had temporarily shifted to other research topics. Taking up the black-body problem again in late October, he would rapidly proceed towards a deeper resolution of the puzzles he had posed between mid-June and mid-July. But that earlier month of concentrated work had already given him a command of both the problem and the concepts basic to the paper he presented to the Vienna Academy in early November.[34] There, under the title "On the Physical Presuppositions of Planck's Theory of Irreversible Radiation Processes," he argued that Planck's proof of the electromagnetic H-theorem lacked an essential step present in Boltzmann's equivalent theorem for gases. Planck has produced, Ehrenfest said, an entropy function Σ that can only increase or remain constant for all possible "natural" radiation states. But he has failed to prove that, "for given total energy, Σ will remain constant only if the stationary state reached by the radiation conforms to a distribution uniquely determined by the total energy." In fact, Ehrenfest continued, no such proof can be given in the absence of additional assumptions. Under the conditions required by Planck's theory, "it is possible to specify infinitely many radiation fields which: 1. all possess the same total energy; 2. [are] 'natural'; 3. [are] stationary, so that, in particular, the value of Σ does not increase when they are present; but which nevertheless, 4. correspond to infinitely many different distribution functions."[35] At the very end of his paper Ehrenfest specified, though without discussion, the additional conditions that appeared to ensure that Planck's theory would yield a unique distribution. They were, of course, complexion theory and the relationship $\varepsilon = h\nu$.[36] Ehrenfest stated he would devote a later paper to them.

In the event, Ehrenfest's argument proved more important than his conclusion, and it took, in outline, the following form. During 1900 Planck had made use of two different expressions for entropy, $\Sigma_1(U)$ and $\Sigma_2(U)$, equations (III-20) and (IV-15). Each specifies entropy as a function of the energy (and wavelength) alone. But the two are

different: maximizing the first yields the Wien distribution, $\phi_1(\lambda)$; maximizing the second leads to Planck's distribution, $\phi_2(\lambda)$. Now imagine two different states, Z_1 and Z_2, of the radiation and resonators, each specified by a particular set of electromagnetic field functions, $\mathbf{E}(x, y, z, t)$ and $\mathbf{H}(x, y, z, t)$, together with the set of functions $f_i(t)$ for the various resonator moments. Both sets of functions are assumed to satisfy the conditions of natural radiation as well as Maxwell's field equations and those governing Planck's resonators. What differentiates the two is that the first, specifying Z_1, corresponds to the Wien distribution $\phi_1(\lambda)$, a total density of radiant energy $\rho_1(= \int \phi_1 \, d\lambda)$, and a total energy E_1; whereas the second corresponds to the Planck distribution $\phi_2(\lambda)$, a total density of radiant energy ρ_2, and a total energy E_2. Using the function $\Sigma_1(U)$, Planck has shown that, if the state Z_1 ever occurs, then Σ_1, and consequently U, must thereafter remain constant. But if U is constant, then Σ_2 must be constant as well, for both Planck's entropy functions increase with energy. Precisely the same argument applies to the state Z_2. If it ever occurs, then not only Σ_2 but Σ_1 must thereafter be stationary. It follows that, if radiation in an initial state Z evolves to Z_1 in accordance with the equations governing the field and resonators, then, while Σ_1 rises to an absolute maximum, Σ_2 will rise to a stationary value, and conversely for an initial state that evolves to Z_2. Since Planck's criterion for an entropy function has been simply that it increase steadily to a stationary state, he has no basis on which to choose between them.[37†]

What made this situation especially serious, in Ehrenfest's view, is that the two states Z_1 and Z_2 may readily be adjusted so that they correspond to the same density of radiant energy and thus, by the Stefan–Boltzmann law, to the same temperature. The relevant arguments are the dimensional ones traceable to Lorentz and mentioned by Planck in his July letter. Multiplying all the field and resonator functions specifying Z_1 by m^2 yields a new set of functions specifying a new state Z_1'. It satisfies all the preceding conditions on Z_1 except that the total energy, density of radiant energy, and distribution function are given by $E_1' = m^4 E_1$, $\rho_1' = m^4 \rho_1$, and $\phi_1'(\lambda) = m^4 \phi_1(\lambda)$. If m^2 is now chosen so that $\rho_1' = \rho_2$, two simultaneously stationary radiation states, Z_1' and Z_2, have been determined, which correspond to the same temperature but to different distributions of radiant energy in the field. Unlike Boltzmann's arguments, Ehrenfest concludes, Planck's cannot determine a unique distribution function without recourse to some special assumptions.

A preliminary version of this argument and of the conclusions drawn from it are presumably what Ehrenfest communicated to Planck at the end of June. Planck's answer suggests that he was unimpressed, and one can easily see why. Excepting their generality, which the article does not altogether justify,[38†] Ehrenfest's conclusions add nothing to what Planck had known in 1900. By March of that year he had persuaded himself that any function $S(U)$ would have the characteristic properties of entropy if it satisfied the equation, $\partial^2 S/\partial U^2 \propto -f(U)$, with f any positive function that approaches zero as U goes to infinity. To find a unique distribution law would therefore require some additional argument to determine f. Planck's use of combinatorials was, in fact, his second such additional argument within the year. But Planck had very likely failed to recognize that the radiation state that maximized one entropy function was also a stationary state for all the others. That circumstance posed a further puzzle which, between late October 1905 and mid-February 1906, led Ehrenfest to a set of conclusions far more important than those described in his Academy paper. What aspects of the physical situation could account for the existence and strange interrelationships of the various functions that could serve as entropy? By the time his first paper was presented, Ehrenfest, in his notebooks, had begun to call for a "general theory of [such] quasi-entropies," one that would explain both their "always increasing" and the "uniqueness of the final state" to which each led. That theory was also to relate somehow to "complexion theory."[39]

Roughly two weeks before he first used the term "quasi-entropies," Ehrenfest recorded in his notebooks an important clue to the puzzling behavior of the functions for which he coined the name, and a similar clue appeared as an aside in his November paper. In the latter he wrote that, when several possible entropy functions exist with maxima corresponding to different radiation states, then the state actually reached "is not determined [simply] by the total energy but depends in some scarcely specifiable manner on the other initial conditions of the motion."[40] The equivalent notebook entry is quickly followed by a sketch of a model that would, by stages, lead him to understand how initial conditions might function. In that model, resonators are represented by numerous small masses, and these are interconnected by a network of springs that represent the ether. In addition Ehrenfest introduced a multitude of gas molecules able to excite the resonators by collision and thus to transfer energy between the various vibration modes of the ether-resonator net.[41]

What processes within the model, Ehrenfest quickly asked, correspond to increases of entropy? The first part of his notebook entry is a list of processes that, like diffusion, produce no such increases. Then he continued:

> H increases only if the mol[ecules in the phase-space cell] do $d\omega$ experience different events—
> a) Collisions with each other
> b) " with a very rough wall
> ??!! c) Enorm[ously] long time—[42]

In that alphabetized list, the first two items (the third will be considered in the next section) are independent mechanisms that can produce entropy increase. Either can cause H to change in the absence of the other. The effect of each can therefore be examined separately, a fact Ehrenfest immediately exploited by considering molecules that do *not* ordinarily collide and, in the process, significantly changing his model. Five items after the one just quoted, Ehrenfest wrote:

Heuristic for a Thermod[ynamic]-Kinetic Theory of Cavity Radiation:

Partial oscillations [i.e., ether vibrations at individual frequencies] pass each other without [interaction]	Molecules of different sort travel without any collisions
– o –	– o –
Resonators	Catalytic substances[43]

Individual *non-colliding* molecules here take on a new role as analogues to individual modes of free ether vibration. The function of transmitting energy from mode to mode is transferred to resonators, on the one hand, and to a catalytic mechanism that promotes interaction between otherwise mutually transparent molecules, on the other.

First recorded in early November, the representation of radiation in a reflecting enclosure by the molecules of a "collision-free" gas played an important role in Ehrenfest's thought from the end of the following month. Then, in two notes written just before 1 January 1906, he explicitly employed the model to explore the behavior of entropy when a radiation-filled cavity expands without doing work.[44] The same model reappeared in February, at which time the functions previously attributed to rough walls were performed by spheres fixed in the cavity. Together, these spheres and the collision-free gas constituted, for Ehrenfest, a "Simplified Model for [a] Quasi H-Theorem," and that is the role in which the model was introduced in the paper Ehrenfest submitted to the *Physikalische Zeitschrift* at the end of June.[45]

What he then pointed out he must have recognized at least five, and very possibly eight, months before.[46†]

If the molecules of a collision-free gas interact only with rough walls or fixed elastic spheres, but not with each other, then only their directions of motion, not their individual speeds, will change over time. In the absence of a mechanism for redistributing energy among them, whatever velocity distribution they possess initially will be preserved. Boltzmann's H and other suitably selected functions may nevertheless increase with time to a stationary value as molecules are redistributed in position and in the directions of their motions. Such functions may, that is, behave as quasi-entropies. But only for some specially chosen initial velocity distribution—itself varying with the choice of entropy function—will a stationary value thus reached be an absolute maximum. Boltzmann's H, in particular, will reach such a maximum only if the molecules conform to the Maxwell–Boltzmann distribution from the start.

That was precisely the behavior that had so puzzled Ehrenfest when he discovered it in Planck's model of radiation interacting with resonators. Placing it in the more familiar gas-theoretic context provided clues to its source. A collision-free gas could proceed to any of an infinite number of different final states (and thus possess no unique equivalent for Boltzmann's H) because no mechanism was available to change the value of one of the parameters characterizing the state of a molecule, in this case its energy or speed. At least for a collision-free gas, the existence of numerous quasi-entropy functions results from the incompleteness of the available mechanisms of equilibration. If only Planck's model were similarly incomplete, its strange behavior could be at once explained. By the time he submitted his June paper, Ehrenfest had identified the required incompleteness. Once it was recognized, he went on to claim, Planck's model could be seen to be equivalent to the one employed by Rayleigh and Jeans. Properly conducted, even recourse to complexion theory should lead to the Rayleigh–Jeans law.

The impotence of resonators

Burbury's scrupulous 1902 analysis of Planck's pre-1900 radiation theory, last mentioned in Chapter V, may well have helped Ehrenfest to identify the element that that theory lacked. In an addendum attached to his Vienna Academy paper in proof, Ehrenfest called special attention to Burbury's three-year-old article, which he had not previously

seen. Close to its end Burbury noted in passing what he took to be an inadequately developed aspect of Planck's proof:

> Planck has given no account of interchanges of energy between systems of different vibration periods. His method is in fact based on the assumption or proof (art. 6) that waves of different period from that of a resonator pass the resonator unaffected, so that no interchange of energy takes place. This, however, is not quite rigorous. If the difference of periods, though not zero, be very small, some very small interchange of energy between the wave and the resonator will consistently with the equations of page 433 take place. . . .
>
> Now Planck does not investigate the law of these slow interchanges. He assumes that an entropy function exists for them, and that it is precisely the same function (but with variable ν) which has been defined above for systems having the same period. That may be true, but it cannot, I think, be accepted as an axiom. It seems to me that this branch of the subject requires further elucidation.[47]

Though the second sentence of the passage is not quite fair to Planck, Burbury's central point is sound. Planck had recognized that a damped resonator responds to incident frequencies near but not coincident with its resonant frequency, and he had implicitly relied on such interactions to redistribute energy among the various vibration modes of the field, thus opening a path to equilibrium. But Planck had not investigated the field-mediated interaction between resonators at slightly different frequencies, and his use of an analyzing resonator tuned precisely to the frequency of the field resonator under investigation precluded his doing so. Though resonators at nearby frequencies might, in fact, exchange energy, Planck's theory did not take such exchanges into account. Under those circumstances, if Planck had been able to exhibit a function that could only increase to a stationary value with time, the increase must be due to field alterations taking place at a single frequency, for example to changes in phase, direction, or polarization of individual partial vibrations.

With or without Burbury's intervention, Ehrenfest's June paper opened with a related but significantly stronger point:

> 1. The frequency distribution of the radiation introduced into the model [described by Planck] will not be influenced by the presence of arbitrarily many Planck resonators, but will be permanently preserved.
>
> 2. A stationary radiation state will [nevertheless] result from emission and absorption by the oscillators in that the intensity and polarization of all rays of each color will be simultaneously equilibrated in magnitude and direction.
>
> In short: radiation enclosed in Planck's model may in the course of time become arbitrarily disordered, but it certainly does not become

blacker.—For the discussion to come the following formulation is especially suitable: Resonators within the reflecting cavity produce the same effect as an empty reflecting cavity with a single diffusely reflecting spot on its wall.[48]

Ehrenfest here announces the incompleteness of the equilibrating mechanism that permits Planck's radiation model to support numerous quasi-entropies. Fixed linear resonators cannot alter the frequency distribution of the energy in the radiation field.

Ehrenfest's notebooks are frustratingly elusive with respect to the date at which he recognized this surprising and initially counter-intuitive theorem. Its first explicit enunciation, to be considered near the end of this section, does not appear until late May 1906, but it is then attached to a particular model of a problematic process with which Ehrenfest had been concerned for at least the preceding six months. He had surely been aware of potential difficulties about frequency redistribution by November 1905, at the latest, and he may have recognized the general theorem at that time. Reasons for wishing that his recognition of the impotence of resonators could be dated more precisely will shortly appear.

As early as June 1905, Ehrenfest had recorded the general solution of the differential equation for a damped resonator of frequency ν_0 driven by a sinusoidal field of frequency ν.[49] The resulting resonator moment consisted, he noted, of two superimposed vibrations, one of fixed amplitude at the exciting frequency ν, the other a damped oscillation at a field-independent frequency almost coincident with ν_0. Presumably, at this point, he took the latter to represent the new frequency to which interaction with the resonator had shifted some field energy previously at the frequency ν.[50†] Doubts about that manner of understanding energy redistribution however, began to appear almost at once.

One likely source of these doubts, to which Ehrenfest alluded in his paper of June 1906, derived from two of the models for resonators he had mentioned in his notebooks twelve months before. Two nearby entries point out that entities defined simply by Planck's resonator equation can be physically represented either by tiny tuned perfect conductors or by suitably chosen bits of dielectric.[51] In print, that remark was accompanied by highly general theoretical reasons why no change in frequency distribution could result from the introduction of resonators of this sort. But even in the absence of such reasons, existing research on dispersion and reflection would have suggested that

neither way of modeling resonators could produce significant alterations of the distribution of energy with frequency except conceivably over very long periods of time. At best the effect would be vastly more gradual than the rapid changes of phase, direction, and polarization due to interaction between a resonator and the incident field. Possibly Ehrenfest had such slow changes in mind when he closed the list of items that could cause changes in H with "???!! c) Enorm[ously] long time."[52]

Ehrenfest could have seen these difficulties as early as the summer of 1905, and by November he had apparently recognized others. Almost immediately after his first list of entropy-increasing factors, he wrote in his notebook: "[How] to specify resonators which transform each incident wave into a small spectrum (dep[endent] on temp–) (Diff[erential] equ[ation] non-linear)," and four items later, "Black body: changes any given energy quant[um] into a precisely det[erminate] spectrum (dependent on T)."[53] These entries read like proposals for the introduction of resonators with properties, like non-linearity, not possessed by Planck's. Very possibly, when he wrote them, Ehrenfest had realized that the apparently "new" frequency near ν_0, introduced when a linear resonator is first struck by an incoming "monochromatic" wave of frequency ν, is actually present through all past time in the Fourier representation of that wave and is thus not really a new frequency at all. Late in the following month, December, his notebook entries display his awareness of another problem of only slightly less force. The energy spectrum of a given field is not obtained simply by squaring the amplitudes of the field's individual Fourier components. One must instead, as Planck had shown in 1898, average the square of the new field over a time that is long compared with the period of its significant components. The corresponding measure of field intensity demands analyzing resonators with bandwidths too large to discriminate between a resonator's natural frequency ν_0 and those field frequencies ν sufficiently near ν_0 to cause significant resonator excitation.[54] In his published paper Ehrenfest would emphasize that the plausibility of Planck's frequency-changing mechanism was due, in part, to confusing the properties of the distribution of field strength, with those of the distribution of energy.[55]

Early in 1906 Ehrenfest had thus developed a far deeper understanding of the "Physical Presuppositions of Planck's Theory of Irreversible Radiation Processes" than was presented in the paper with that title he had read to the Vienna Academy at the beginning of the preceding November. By February, at the latest, he knew that the existence of quasi-entropy functions could be due to an incomplete equilibrating mechanism, and he had earlier had reason at least to suspect the nature of the particular incompleteness that characterized Planck's model. Those discoveries need not, for him, have suggested that no classical model of black-body radiation could succeed. A footnote in his published paper points out that molecular collisions could redistribute energy between fixed resonators and that a non-linear resonator equation would have the same effect.[56] But the discoveries did raise fundamental questions about the adequacy of Planck's theory, and Ehrenfest is likely to have responded to them by renewing the correspondence he had initiated with its author during the preceding June or July. That he, in fact, did so is strongly suggested by a surprising "Conclusion" added, presumably in proof, to the first edition of Planck's *Lectures*. Planck specially directed his readers' attention to it at the end of his brief preface, dated "Easter 1906." The first numbered paragraph of Ehrenfest's June article refers to it as well.[57†]

> § 190. **Conclusion:** The theory of irreversible radiation processes developed here explains why, in an irradiated cavity filled with oscillators of all possible frequencies, the radiation, regardless of its initial conditions, reaches a stationary state: the intensities and polarizations of all its components are simultaneously equilibrated in magnitude and direction. But the theory still is characterized by an essential gap. It treats only the interaction between radiation and oscillator vibrations at the same frequency. At a given frequency the continuous increase of entropy to a maximum value, required by the second law of thermodynamics, is therefore proven on purely electro-dynamic grounds. But for all frequencies taken together the maximum reached in this way is not the absolute maximum of the system's entropy, and the corresponding state of the radiation is not in general the [state of] absolutely stable equilibrium (cf., paragraph 27). The theory does not at all elucidate the manner in which the radiation intensities corresponding to different frequencies are simultaneously equilibrated, that is, the way in which the initial arbitrary distribution settles in time to the normal distribution characteristic of black radiation. The oscillators which provide the basis for the present treatment influence only the intensities of the radiation corresponding to their own natural frequencies. They are not able, however, to change its frequency if their effects are restricted to the emission and absorption of radiant energy.[58]

When Planck wrote that passage, the article in which Ehrenfest announced the identical discovery had not yet been composed, and Planck may therefore conceivably have been led to it independently by his work on dispersion theory. But, if so, the discovery must have come very late, after Planck's research had shifted from dispersion

theory to relativity and after his manuscript of the lectures delivered in the Winter Semester, 1905–1906, was largely completed. Had Planck made the discovery earlier, he would surely have presented it within the body of his book, the structure of which would have required significant revision. In its published form, the *Lectures* opens with two chapters on the behavior of radiation in a cavity that contains no resonators. Paragraph 27, to which Planck calls attention in the passage just quoted, emphasizes that, though entropy will increase in such a cavity, it will not reach an absolute maximum "since the given total energy can be arbitrarily distributed over the various colors of radiation."[59] Resonators are then introduced in the third chapter to provide a mechanism for redistributing energy over frequency, and they retain that role until the last two paragraphs of the book.

Those paragraphs, which constitute Planck's new "Conclusion," must have been devastating to write, for they invalidate, not the details of his presentation, most of which remain worthwhile, but the overall structure of the argument in which those details had been embedded through the preceding two hundred and twenty pages. Much of the awkwardness that inevitably results could have been avoided if Planck had recognized the impotence of resonators in time. Under these circumstances, it seems probable that Ehrenfest's intervention was required to bring Planck to his "Conclusion," and that likelihood is reinforced by a footnote attached to the close of the preceding quotation. It cites Ehrenfest's 1905 paper to the Vienna Academy, an article in which, as previously emphasized, nothing whatsoever is said about the mechanisms that lead to entropy change. Except as an inadvertently disguised acknowledgment of Ehrenfest's role, that citation is extraordinarily difficult to understand.[60†]

If, however, Planck accepted Ehrenfest's still unpublished theorem, he said nothing about how it was to be proven or why his initial intuition of the effect of damped resonators was wrong. Ehrenfest's remarks on that subject in his June paper are too cryptic to be helpful, a prelude to the quite different subject to which we shall shortly find his article primarily devoted. Soon after it appeared, with the increasing recognition of discontinuity and the consequent elimination of the damping term from the resonator equation, the problem vanished from physics, only to reappear in a quite different context, scattering theory. As a result, the claim that damped resonators can alter the frequency distribution of radiant energy is still widely believed, even among physicists. A brief discussion of the considerations that invalidate it may therefore eliminate confusion and facilitate understanding of the concepts that connect the concerns of Ehrenfest's notebooks with his brief published remarks.

A radiation-damped resonator will oscillate naturally at its resonant frequency ν_0, but it can also absorb energy from a wave at a nearby frequency ν. That such a resonator will alter the frequency distribution of radiant energy is plausible, for, if it absorbs energy from an incident wave of frequency ν, it will, one supposes, reradiate that energy at its natural frequency ν_0 after the initial wave has passed. Two sets of considerations, both adumbrated in Ehrenfest's notebooks, are relevant to showing that nothing of quite that sort occurs. A damped resonator will reradiate only at frequencies already contained in the exciting wave. At each of the frequencies to which it does respond, furthermore, it will reradiate only as much energy as it absorbed at that frequency.

Planck's equation for the moment f of a damped resonator, equation (I-8b), may for convenience be rewritten in the form,

$$\ddot{f} + 2\alpha\dot{f} + \omega_0^2 f = \beta Z(t), \tag{4}$$

with $Z(t)$ the electric field parallel to the resonator axis, ω_0 the resonator's natural angular velocity, α its decay rate (equal to $\sigma\omega_0/2\pi$), and β a constant equal to $3c^3\sigma/2\pi\omega_0$. The right-hand side of that equation, $\beta Z(t)$, may generally be represented by a Fourier integral, assumed convergent,

$$\beta Z(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} E(\omega)\,\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega, \tag{5}$$

with $E(\omega) = E^*(-\omega)$ to ensure that $Z(t)$ is real. The general solution of equation (4) is then given by

$$f(t) = A\,\mathrm{e}^{-\alpha t}\cos(\omega_0 t + \theta) + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{E(\omega)\,\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega}{\omega_0^2 - \omega^2 + 2\mathrm{i}\alpha\omega}, \tag{6}$$

where A and θ are arbitrary constants.

In equation (5), the function $E(\omega)$ may be chosen so that there is no net field $Z(t)$ during all time prior to some selected instant t_0. By complex integration clockwise about a contour consisting of the real axis and an infinite semicircle in the lower half plane, it is then easily shown that, for $\alpha > 0$, the integral on the right of equation (6) must vanish for all $t \le t_0$. Therefore, unless the resonator has been excited by

some source other than the field, equation (6) provides a physically permissible solution of the resonator equation only if the arbitrary constant A is set equal to zero. The resonator's natural angular velocity ω_0 does not then appear in the spectrum of $f(t)$ unless it is already represented in the spectrum of the driving field. Among those frequencies that are present in the incoming wave, the resonator does, of course, respond most strongly to the ones closest to ω_0. But that is only to say that it *both* absorbs *and* reradiates more strongly near its resonant frequency than it does elsewhere. Since the techniques required to show the conservation of the energy of wave plus resonator can be applied separately to each frequency, no net redistribution of energy can occur.

Some cryptic entries in Ehrenfest's notebooks suggest that he reached this conclusion by a somewhat different route, one less likely to be familiar to most readers. Characteristically, he again called on a mechanical model the motion of which was to be represented in what are usually called normal coordinates.[61] Such coordinates are always available for motions consisting exclusively of displacements from equilibrium provided that the restoring forces are linear functions of the displacements. For present purposes, the special feature of normal coordinates is that the variation of each of them occurs at a single frequency and can be represented by a form like $C_i\sin(2\pi\nu_i t + \theta_i)$. Here, ν_i is some linear combination of the original displacement frequencies, and both C_i and θ_i are constants determined by the masses of the system and their equilibrium positions. The total energy is then given, in normal coordinates, by one-half the sum of the squares of the individual coordinates and of their first derivatives, so that a certain unchanging energy, $\frac{1}{2}C_i^2$, is attributed to each of the coordinates over all time. In the original displacement coordinates, on the other hand, the energy attributed to each coordinate oscillates slowly at a frequency given by some linear combination of the various displacement frequencies.

Reduced to standing wave solutions, the Rayleigh–Jeans problem of radiation in an empty reflecting cavity is clearly equivalent to a mechanical vibration problem treated in normal coordinates. Ehrenfest appears to have recognized their relation by late March 1906 when he wrote in his notebooks, "Energy distribution over normal vibrations.... An H-theorem on this basis."[62] If a linear resonator weakly coupled to the normal modes of the empty cavity is introduced, the new system thus formed can again be reduced to normal coordinates, and the problem treated as before. In either case, the energy in each mode remains constant; no energy is transferred from one vibration frequency to another. Some such theorem is what Ehrenfest must have anticipated when he wrote the following entry in his notebook just before 30 May: "By returning to the normal vibrations of the system [consisting of] ether plus resonators, show that no change in the 'color distribution' can ever be brought about in this way."[63] That is his first fully explicit expression of his conviction about the impotence of resonators, though it presumably does not represent the basis for his earlier doubts on the subject of energy redistribution.

It is against this background that Ehrenfest's very brief published remarks about the problem of redistributing energy must be read. Early in his June paper he refers readers to the closing paragraphs of Planck's *Lectures* (which can scarcely have appeared) and proceeds to his own, previously quoted, statement about the equivalence of Planck's model to a perfectly conducting cavity containing a diffusely reflecting spot. Next, he mentions the treatment of the Rayleigh–Jeans cavity in the *Lectures*, and he quotes from it the statement that in such an empty cavity "there can be no talk…of a tendency towards the equilibration of the energy allotted to individual partial vibrations."[64] Then he continues:

> This conclusion applies immediately to the Planck model. So long as the oscillators are defined only by the linear homogeneous differential equation,* which Mr. Planck establishes for them, they are essentially identical with small specks of complete conductors or suitable dielectrics. In that case every state of motion of the Planck model is again [like the empty cavity] a superposition of the normal vibrations of this more complicated system. There can, therefore, in this case also, be no talk of a tendency towards the equilibration of the energy allotted to individual partial vibrations.[65]

The asterisk in that passage leads to a footnote that admits the possibility of energy redistribution in the presence of moving molecules or with a non-linear oscillator equation; Planck noted an equivalent escape route in the last paragraph of his "Conclusion." To Ehrenfest, however, these possibilities were not of current interest; his immediate object was to analyze Planck's theory rather than produce one of his own. Having pointed out, at the start, that Planck's success could not be due to the use of resonators, he therefore devoted the body of his paper to exploring the potential of another apparently special aspect of Planck's approach. It was the use of combinatorials or what Ehrenfest had previously labeled "Abstract Entropy Theory."

Complexion theory and the Rayleigh–Jeans law

In his papers on radiation and again in his *Lectures*, Planck had evaluated entropy by applying complexion theory to resonators. But complexion theory, thus applied, was simply probability theory, and its utility should therefore be independent, Ehrenfest pointed out, of recourse to resonators. Developing a representation "that corresponds more nearly to the methods of Rayleigh and Jeans" than to those used by Planck,[66] Ehrenfest therefore proceeded in the third part of his paper to apply Boltzmann's probabilistic definition of entropy directly to the field.

For radiation in any empty cavity, each of the independent modes of oscillation may be conceived as an undamped oscillation with energy

$$\varepsilon_\nu = \tfrac{1}{2}(\alpha_\nu f^2 + \beta_\nu \dot{f}^2) = \tfrac{1}{2}\left(\alpha_\nu f^2 + \frac{1}{\beta_\nu} g^2\right). \tag{7}$$

In these equations, f is the moment of a field oscillator, g its conjugate momentum $(= \partial\varepsilon_\nu/\partial\dot{f})$, and ν its oscillation frequency $(= \sqrt{\alpha_\nu/4\pi^2\beta_\nu})$. Now let $F(\nu, f, g)$ be the still unknown distribution function that specifies the fraction of field oscillators with frequency ν and coordinates in the range f to $f + \mathrm{d}f$ and g to $g + \mathrm{d}g$. In addition let $N(\nu)\,\mathrm{d}\nu$ be the number of field oscillators or of vibration modes in the range ν to $\nu + \mathrm{d}\nu$. The standard generalization to phase space of Boltzmann's probabilistic definition of entropy[67] then yields for the entropy of the field,

$$S = -k \int_0^\infty N(\nu)\,\mathrm{d}\nu \iint_{-\infty}^{\infty} F(\nu, f, g) \log F(\nu, f, g)\,\mathrm{d}f\,\mathrm{d}g. \tag{8}$$

In that expression $N(\nu)\,\mathrm{d}\nu$ is simply the Rayleigh–Jeans mode count, equation (3), rewritten in terms of frequency and multiplied by two for application to perpendicularly polarized transverse waves. Continuing to follow Boltzmann's technique, Ehrenfest finds the equilibrium state by maximizing equation (8) subject to the constraint on total energy E_t and an appropriate normalization condition on F:

$$\begin{gathered} \iint_{-\infty}^{\infty} F(\nu, f, g)\,\mathrm{d}f\,\mathrm{d}g = 1 \\ E_t = \int_0^\infty N(\nu)\,\mathrm{d}\nu \iint_{-\infty}^{\infty} \varepsilon_\nu F(\nu, f, g)\,\mathrm{d}f\,\mathrm{d}g. \end{gathered} \tag{9}$$

Finally, with F thus determined, the total energy of the radiation with frequency between ν and $\nu + \mathrm{d}\nu$ can be written

$$E(\nu)\,\mathrm{d}\nu = N(\nu)\,\mathrm{d}\nu \iint_{-\infty}^{\infty} \varepsilon_\nu F(\nu, f, g)\,\mathrm{d}f\,\mathrm{d}g.$$

Given the examples provided by both Boltzmann and Planck, the required manipulations are straightforward. Their result, however, is a demonstration that F can be maximized only by attributing equal mean energies to each of the vibration modes. If, furthermore, those modes interact with the molecules of a gas, that constant mean energy is just kT. Complexion theory applied directly to the field thus yields the same impossible radiation law that Jeans had derived directly from the equipartition theorem in 1905.[68]

If, however, Ehrenfest's adaptation of complexion theory to the field led only to a known result, his reformulation suggested how other results might be achieved. Planck's introduction of resonators had only confused the fundamental physical issue. Equations (7) and (8) together with the associated maximization techniques are, Ehrenfest emphasized, common to Planck and Boltzmann. If they nevertheless can lead to different distribution functions, he continued, the source of the difference must lie in the choice of constraints, and indeed choices different from equations (9) may be justified.

> Suppose [for example] that radiation in nature is produced only through the intervention of electrons and that these electrons everywhere possess the same definite structure. This structure—perhaps merely the eternally fixed electron cross-section—may in principle be capable of preventing the excitation by natural means of some of the conceivable normal vibration modes of our cavity.[69]

During the year before that passage was written, Ehrenfest's view of Planck's hopes for electron theory had clearly changed.

Ehrenfest ends his paper with a brief general discussion of constraints, showing that they are not ordinarily uniquely determined by the distribution function that their application yields. Other constraints besides Planck's might thus have led to his distribution law. But one can at least demonstrate, Ehrenfest concludes, that Planck's law can follow from complexion theory applied to the field alone. One suitable device is an additional constraint that restricts the value of ε_ν to integral multiples of the energy quantum $h\nu$, "just as if, for each frequency, the vibration energy consisted of 'energy atoms' of numeri-

cal magnitude, $\varepsilon_\nu^0 = 6.548 \cdot 10^{-27} \cdot \nu$ erg."[70] That condition, he adds, can be restated in a form more nearly standard in statistical mechanics. In the two-dimensional phase space f, g,

> the phase point of a proper vibration of frequency ν cannot occupy any position on the surface: instead it may lie only on a family of curves, namely the family of ellipses
>
> $$\tfrac{1}{2}\left(\alpha_\nu f^2 + \frac{1}{\beta_\nu} g^2\right) = mh\nu,$$
>
> where m runs through the series of integers to a value such that $mh\nu$ would exceed the previously specified total energy if m increased further.[71]

Though Ehrenfest notes in passing that this formulation is not quite the same as the one provided in Planck's *Lectures*, nothing in his paper suggests that he thought that more than a slip of Planck's pen might be involved. He was still reading Planck through the spectacles provided by Lorentz and had not, in any case, had the book long enough to assimilate Planck's developed approach.

Ehrenfest's paper includes no proof that his quantization of field oscillators results in Planck's law, but one can easily be supplied. In eq. (7), ε_ν must be equated to $nh\nu$; in the equations that follow, F becomes a function of n rather than g, and the corresponding integrals over g become sums over n. Then $F(\nu, f, n)$ becomes the fraction of those modes with frequency ν that possess energy $nh\nu$ and lie between f and $f + \mathrm{d}f$, and $N(\nu)$, the number of modes at any fixed frequency, is everywhere equal to two. With these changes, Planck's law results straightforwardly from the manipulations Ehrenfest has sketched. The elaborate relationships between damped resonators and field—the core of Planck's approach from late 1894 through 1906—is thus shown to be irrelevant to their author's principal achievement. His law can be derived without recourse to resonators.

VII

A NEW ROUTE TO BLACK-BODY THEORY: EINSTEIN, 1902–1909

Though the implications of the paper Ehrenfest had submitted in late June 1906 were startling, they were not by that date still altogether new. Very differently expressed, they had also been educed in an article "On the Theory of the Emission and Absorption of Light" received by the editor of the *Annalen der Physik* more than three months before. Its author was Albert Einstein, another little-known young physicist who, unable to obtain an academic position, wrote from the Swiss patent office in Berne. Analyzed in classical terms, Einstein said, Planck's black-body model could lead only to the Rayleigh–Jeans law. Planck's radiation law could be derived instead, but only by decisively altering the concepts its author had employed for that purpose. Midway through his paper Einstein wrote:

> We must, therefore, recognize the following position as fundamental to the Planck theory of radiation: The energy of an elementary resonator can take only values which are integral multiples of $(R/N)\beta\nu$ [where R is the gas constant, N Avogadro's number, and β a constant]. During absorption and emission the energy of a resonator changes discontinuously by an integral multiple of $(R/N)\beta\nu$.[1]

That passage is the first public statement that Planck's derivation demands a restriction on the classical continuum of resonator states. In a sense, it announces the birth of the quantum theory.

Though their conclusions largely overlapped, Einstein's paper was in several respects quite different from the one Ehrenfest was to submit a few months later. Its argument was both more general and more compelling. Unlike Ehrenfest, furthermore, Einstein did not suppose that he was simply restating Planck's own premise, and the structure of his argument highlighted the impossible difficulties to be encountered in developing the proposal that fixed size h be attributed to phase-space cells. Still more important, whereas Ehrenfest's paper had been a study of Planck, Einstein's was primarily a study of nature. What

brought Einstein to the black-body problem in 1904 and to Planck in 1906 was the coherent development of a research program begun in 1902, a program so nearly independent of Planck's that it would almost certainly have led to the black-body law even if Planck had never lived. To understand Einstein's involvement with black-body theory one must begin by briefly retracing his steps.

Einstein on statistical thermodynamics, 1902–1903

As Klein has emphasized, Einstein was deeply impressed from the start of his career by the simplicity and scope of classical thermodynamics. In that respect he was like Planck; but for Einstein thermodynamics included the statistical approach he had learned from Boltzmann's *Gas Theory*.[2] His first two papers, published in 1901 and 1902, were attempts to investigate intermolecular forces by applying phenomenological thermodynamics to such phenomena as capillarity and the potential difference between metals and solutions of their salts.[3] Finding his results inconclusive, Einstein quickly abandoned that approach and began instead to develop a statistical thermodynamics applicable not only to gases, the main concern of earlier workers, but to other states of aggregation as well. Presumably, he felt that the statistical approach provided a firmer basis than the phenomenological for conclusions at the molecular level. Whatever his motive, however, the result of Einstein's effort was a series of three brilliant papers, published successively during each of the years 1902, 1903, and 1904. These papers are now little remembered because their principal results were simultaneously established by Gibbs in his *Statistical Mechanics* of 1902.[4] Nevertheless, they provided the starting point for much of Einstein's future work, especially for that on Brownian motion and on the quantum, both traceable from 1905.

The first of Einstein's statistical papers developed a theory of statistical thermodynamics for mechanical systems governed by Lagrangian equations of motion with an explicit potential function. Those mechanical equations were required, however, only to justify recourse to Liouville's theorem and to energy conservation, a fact that suggested, Einstein said, that his theory might be redeveloped for systems of a far more general sort.[5] That generalization was supplied in the second paper of Einstein's series. Since in other respects the two papers are closely parallel, attention is here restricted to the second.[6]

Einstein opens that paper by directing attention to a system of which the state can be specified by n independent variables p_i. (For the mechanical systems he had previously treated, these variables were the generalized coordinates and velocities of a Lagrangian description.) If such a system is isolated, its state at one instant must determine the values of the state variables at the next. The behavior of the system over time, that is, must be governed by a set of n equations,

$$\frac{dp_i}{dt} = \phi_i(p_1, p_2, \ldots, p_n), \tag{1}$$

which play the role taken by Lagrange's equations in the mechanical case. Einstein supposes, in addition, that the system of equations (1) possesses one *and only one* independent integral,

$$E(p_1, \ldots, p_n) = \text{Constant}, \tag{2}$$

which is a powerful condition equivalent to an ergodic hypothesis.

For systems of this very general sort, Einstein next develops a series of theorems for which gas theory had had no special need but which are required for the transition to a more widely applicable statistical mechanics or statistical thermodynamics. In particular, Einstein produces, in terms of the functions ϕ_i and E, expressions for those quantities that must correspond to the temperature, entropy, and probability of a state. (In gas theory, temperature could be defined by recourse to the ideal gas thermometer, and an expression for entropy then followed.) These are, of course, precisely the conceptual elements most obviously missing from Planck's discussion of the black-body problem. He had been forced to "define" the probability of a state, and he had noted that his definition, though plausible, could be further justified only by experiment. That uncertainty about probability was transmitted to entropy, the logarithm of probability, and thence to temperature, the derivative of energy with respect to entropy. Einstein's paper, written before he showed any signs of concern with black-body theory, bridged these gaps.

Einstein begins by asking what conditions the trajectories specified by equations (1) and (2) must satisfy in order that the corresponding systems be physical, i.e., possess observable properties. "Experience teaches us," he says, "that an isolated physical system settles after a time into a state such that none of its observable magnitudes changes further; we call such a state stationary."[7] Observables, he goes on to say, are represented by time averages of functions of the microscopic coordinates p_i (consider the pressure of a gas), and his discussion implies

that the time interval over which the average is taken must be long enough so that the coordinates take on all possible combinations of values. Two measurements made at different times will then yield the same value if the trajectories specified by equations (1) return to their starting points with some constant frequency, for the system will then, during each cycle, spend the same proportion of time in the vicinity of any given point specified by the p_i's. Indeed, a slightly weaker condition will produce the same result. Imagine a region Γ in the n-dimensional space of the p_i's; observe the system over some time interval T; and determine the portion τ of that interval during which the system lies in Γ. Then, if for each selected Γ the fraction τ/T approaches a limit with increasing T, the system will possess fixed, observable properties. It will be a physical system.

Though Einstein's argument, to this point, is in several respects defective, it is readily salvaged if observables are taken to be averages not over time but over the members of a suitably arranged set or ensemble of identical systems.[8†] That is the step to which Einstein at once proceeds, and the properties of the set had probably guided his thoughts as he sought the conditions that make an individual system physical. The collection of systems he envisaged is very close to what, following Gibbs, has since been called a microcanonical ensemble.[9] It consists, that is, of a large number N of identical systems, all governed by the equations of motion (1), all independent, and all with the energy constant of equation (2) in the narrow range E^* to $E^* + \delta E^*$. If the systems of this ensemble are distributed through p_i-space in such a way that the number of systems m in a region Γ remains constant during their motion, the ensemble is said to be stationary, and it then possesses the two following properties. The value of m/N is at all times the same as the previously specified limit of τ/T for individual members; all members of the stationary ensemble are therefore physical. In addition, time averages of functions of the p_i's over long intervals T may be replaced by averages over the N members of the ensemble at any instant. Though the statement of these properties is obscure in Einstein's paper, he was clearly aware of both, and he used them.

Examining the properties of his ensemble, Einstein first points out that, if g is an infinitesimal volume in p_i-space, then the number of systems in g at a given instant is

$$dN = \varepsilon(p_1, \ldots, p_n) \int_g dp_1\, dp_2 \cdots dp_n,$$

where $\varepsilon(p_1, \ldots, p_n)$ is the density of system-points in the space of the p_i's. If the ensemble is to be stationary, then this density must obey standard hydrodynamic continuity conditions, and Einstein introduces them to show that ε can depend on the p_i's only through the energy, which he had previously restricted to an infinitesimal range. This part of his analysis he can therefore complete by rewriting the preceding equation in a form fundamental to all that is to come:

$$\mathrm{d}N = \text{Constant} \int_g \mathrm{d}p_1\,\mathrm{d}p_2 \cdots \mathrm{d}p_n. \tag{3}$$

Though it occurs near the beginning of his paper,[10] Einstein's equation (3) displays the aspect of his thought which made it impossible for him to accept Planck's version of black-body theory. For him, as for Planck, the state of a system is specified in terms of the small cell g in which the system's coordinates lie. But, through the physicality condition, Einstein's concept of state carries with it a concept of probability which differentiates it from Planck's. The probability W_g of finding a given system in the state g must be the fraction (τ/T) of time which the system spends in g or, equivalently, the fraction $(\mathrm{d}N/N)$ of the members of the ensemble that are to be found in g at a given time. Thus, W_g is simply the right-hand side of equation (3), appropriately normalized, and it is necessarily proportional to the volume of g. Planck's hope that some physical mechanism yet to be discovered would account for the need to keep cell size constant is therefore unrealizable in principle. To fix the size of cells while retaining continuous trajectories does violence to the concept of probability.

Even before explicitly applying the notion of an ensemble to considerations of probability, Einstein had employed it to define temperature and entropy.[11] Each of the N systems of the ensemble may be conceptually subdivided into interacting sub-systems, a large one Σ and a small one σ, the latter shortly to be called a "thermometer." The large system is specified by the subset $\Pi_1, \ldots, \Pi_\lambda$ of the original n variables, and it has energy H; the small one is specified by the remaining variables $\pi_1, \ldots, \pi_l$ and has energy η. Since the two sub-systems interact, only $E(=H+\eta)$ is a constant, but Einstein assumes that their construction requires $H \gg \eta$, so that H is very nearly constant.

Next Einstein asks how many of the N systems of the ensemble will have the thermometer variables π_i in a specified region, the remaining

variables being unspecified. The answer, he shows, is given by the equation

$$\mathrm{d}N_2 = \text{Constant}\; \mathrm{e}^{-2h\eta}\,\mathrm{d}\pi_1 \cdots \mathrm{d}\pi_l, \tag{4}$$

where h is a parameter that depends only on the total energy E^* and on the structure of the large system Σ. With that structure specified, a new function $\omega(E^*)$ may be defined by

$$\omega(E^*) = \int_{E^*}^{E^*+\delta E^*} \mathrm{d}\Pi_1 \cdots \mathrm{d}\Pi_\lambda. \tag{5}$$

When that function is known, the parameter h required by equation (4) is given by

$$h(E^*) = \frac{1}{2}\frac{\omega'(E^*)}{\omega(E^*)}. \tag{6}$$

The quantity $h(E^*)$ thus defined has, Einstein next shows, the following properties: With the energy and structure of the large system known, h fully specifies, through equation (4), all effects of Σ on the observable properties of the smaller system σ, the thermometer. If two large systems Σ_1 and Σ_2 have the same effect on a thermometer σ, they will also have the same effect on any other thermometer σ'. Finally, two large systems Σ_1 and Σ_2 can have the same effect on a thermometer σ only if that effect is identical as well with the effect on σ of the composite system $\Sigma_1 + \Sigma_2$. These properties are precisely those of the observed quantity temperature, which must therefore be some function of the quantity h. Einstein suggests that an appropriate definition of T (now a symbol for temperature rather than for a time interval) is given by $T = 1/(4h\chi)$ with χ "a universal constant." Shortly afterwards he shows that a molecule of a perfect gas must, according to his theory, have mean energy $3/4h$ so that χ must be equal to $R/2N$ (or one-half of Planck's k).[12]

With temperature thus defined, Einstein proceeds finally to a representation of entropy. In order that work may be done on the previously isolated system Σ, he supposes that the function E of equation (2) depends not only on the p_i's but also on a set of slowly varying parameters λ_i. The change in E corresponding to any small change in the state of the system is then given by

$$\mathrm{d}E = \sum \frac{\partial E}{\partial \lambda_i}\,\mathrm{d}\lambda_i + \sum \frac{\partial E}{\partial p_i}\,\mathrm{d}p_i,$$

and Einstein readily shows that the first summation is the work done on the system during the variation of the λ_i's, the second the added heat. With entropy S defined as $\int dQ/T$, further manipulation yields the formula

$$S = \frac{E^*}{T} + 2\chi \log \int e^{-2hE(p_1,\ldots,p_n)}\, dp_1 \cdots dp_n. \tag{7}$$

In deriving that formula, Einstein supposes that, both before and after variation of the λ_i's, the system under examination interacts with a system far larger than itself, at the same temperature, thus making it susceptible to the treatment he had previously provided for thermometers. As a result, the p_i's in the integral of equation (7) range over the entire p_i-space, and h ceases to be implicitly dependent on δE^*, as it had been before. Finally, Einstein concludes by proving what for present purposes may be taken for granted: S must increase as a system moves from less to more probable states. That is his version of the second law of thermodynamics.

For physical systems of an extraordinarily general sort, Einstein had thus, by the summer of 1903, produced both a generalized measure of the probability of states and also corresponding measures for temperature and entropy. In an eighteen-page article he had shown how to transform a field that had seldom transcended gas theory into a fully general statistical thermodynamics. Only Gibbs's book, published the year before, offers a significant precedent, and the systems considered by Gibbs are less general than Einstein's. That Einstein nevertheless felt the need to go farther is an example of his extraordinary ability to discover and explore problematic interrelationships between what others took to be merely factual generalizations about natural phenomena.

Fluctuation phenomena and black-body theory, 1904-1905

By the time he finished his 1903 paper, Einstein had recognized that his "universal constant" χ could be evaluated in terms of the values of the gas constant and of Avogadro's number. The theory that had led him to the constant was, however, applicable to systems far more general than gases, and it should therefore have a correspondingly general physical basis. That basis, Einstein is likely to have thought, should reflect the statistical nature of the approach that had led him to the constant, thus explaining not only its role as scale-factor for thermodynamic temperature but also its position as a

multiplier in the probabilistic definition of entropy.[13†] Establishing the physical significance of χ was, in any case, the central problem attacked in Einstein's third statistical paper, submitted to the *Annalen* in the spring of 1904. Its solution lay in the phenomenon of energy fluctuation, which had usually been either ignored or dismissed as unobservable in the earlier literature.[14†] Once again Gibbs provides the only precedent for Einstein's treatment, but the precedent is in this case partial. Einstein makes fluctuations physically central and at once suggests their quantitative application. That last step is what brought him, for the first time in print, to the black-body problem.

Einstein's 1904 paper opens with a minor but consequential redefinition of the function $\omega(E^*)$. For its original definition, equation (5), he substitutes

$$\omega(E^*)\, \delta E^* = \int_{E^*}^{E^*+\delta E^*} dp_1 \cdots dp_n.$$

That substitution eliminates the implicit dependence of $\omega(E^*)$ on the size of the interval δE^* and thus on the particular structure of the imagined ensemble. It is therefore available not only for ensembles but also for individual systems.[15†] Einstein at once applies it in a new definition of a system's entropy and then, in the course of a new derivation of the second law, to the behavior of a system in interaction with a large heat bath at temperature T. The bath, he points out, determines only the average, not the instantaneous energy of the system. The latter will fluctuate about the average in a manner governed by the thermometer equation (4). With the aid of the newly defined ω, that equation may be rewritten to specify the probability dW that the system will at a given instant have energy between E^* and $E^* + dE^*$:

$$dW = C\, e^{-E^*/2\chi T} \omega(E^*)\, dE^*,$$

where the normalizing constant is readily determined by equating the integral of dW over all possible energies to unity.

From the last equation it follows that the average energy of the system at temperature T is just

$$\bar{E} = \int_0^\infty CE\, e^{-E/2\chi T} \omega(E)\, dE,$$

so that

$$\overline{(\bar{E} - E)} = 0 = C \int_0^\infty (\bar{E} - E)\, e^{-E/2\chi T}\, dE.$$

Differentiating the last equation by T and equating to zero the average value of the multiplier of the exponential term within the resulting integral yields

$$\overline{\varepsilon^2} \equiv \overline{(\bar{E} - E)^2} = \bar{E}^2 - \overline{E^2} = 2\chi T^2 \frac{d\bar{E}}{dT}. \tag{8}$$

That result is especially remarkable, Einstein points out, because it no longer contains "any magnitudes reminiscent of the hypotheses basic to the theory." The mean square fluctuation of the energy of any system in contact with an infinite heat bath has been expressed in terms of the measurable quantities T and $d\bar{E}/dT$ together with the absolute constant χ. The physical significance of that constant is therefore at last determined. In Einstein's words: "The magnitude $\overline{\varepsilon^2}$ is a measure of the thermal stability of the system; the larger $\overline{\varepsilon^2}$, the smaller this stability. The absolute constant χ thus determines the stability of the system."[16]

Einstein's recognition of the physical role of the constant χ is what directed his attention to the black-body problem. The words with which he introduces the transition are worth examining, for they appear to imply more than they actually say. Immediately after having related χ to thermal stability, Einstein continues:

> The equation just found would permit an exact determination of the universal constant χ if it were possible to determine the energy fluctuation of a system. In the present state of our knowledge, however, that is not the case. Indeed, for only one sort of physical system can we presume from experience that an energy fluctuation occurs. That system is empty space filled with thermal radiation.[17]

What does Einstein have in mind when he states that "for only one sort of physical system can we presume from experience that an energy fluctuation occurs"? Part of the answer he supplies at once by relating χ quantitatively to the displacement-law constant $\lambda_m T$, where λ_m is the wavelength of the radiation of maximum intensity at temperature T. It seems likely, however, that Einstein had in mind a point of greater generality. The equations governing a phenomenon significantly

affected by fluctuations should contain an additional constant not derivable from the macroscopic laws applicable to that phenomenon, e.g., the laws of mechanics or the electromagnetic field. The need for two natural constants in black-body radiation laws had, however, been a recognized puzzle since at least 1900, one to which Lorentz in particular had repeatedly referred.[18] Einstein may well be hinting in the passage just quoted that the entry of a second constant is due, not to the universal electronic charge, but rather to the existence of fluctuations. It is, in any case, by introducing fluctuations in the analysis of a previously unexplained black-body regularity that Einstein illustrates his point.

In a black-body cavity, Einstein points out, the fluctuation of total field energy will be very small if the cavity's dimensions are large compared with the dominant wavelength. (Total energy is proportional to cavity volume; fluctuations, due to the interference of different partial waves at some point, are independent of cavity size and may be in opposite senses at different locations in a large cavity.) But if the dimensions of the cavity equal the wavelength corresponding to maximum intensity, then the mean total energy should be of the same order of magnitude as the mean fluctuation, $\bar{E}^2 \doteq \overline{\varepsilon^2}$. The total energy in such a cavity is given by the Stefan–Boltzmann law as $E = aVT^4$, where V is the cavity volume ($= \lambda_m{}^3$) and a is an experimentally determined constant. Applying equation (8) at once yields

$$\lambda_m = \frac{2}{T}\sqrt[3]{\frac{\chi}{a}} = \frac{0.42}{T}, \tag{9}$$

where the numerical constant on the right is computed from existing measurements of a and R together with established estimates of N. That expression Einstein compares with experimental results that have shown the wavelength of maximum intensity of black-body radiation to be $\lambda_m = 0.293/T$. In view of the order-of-magnitude methods employed to obtain equation (9), the agreement is extraordinary.

By the time he achieved that result Einstein was reading Planck, for he mentions the latter's definition of entropy in the introduction to his paper. About Planck's radiation law, however, he still says nothing at all. As previously noted, Einstein had reasons of his own to doubt that law's derivation, but he was as yet unable to replace it or to understand why it should yield so successful a result. Two more steps in the development of his own research program were needed before he would reach that understanding, and Einstein took the first of them in a famous paper published in the next year, 1905.[19] Both its structure and its content strongly suggest that Einstein had, following his discovery of equation (9), begun to seek a black-body law of his own, that he had quickly encountered paradox, and that he had then dropped the search for a law in favor of an exploration of the paradox itself.

Einstein's new black-body paper was submitted for publication in March 1905, a month before the beginning of the correspondence in *Nature*, through which Rayleigh and Jeans produced the law since known by their names. That law might, therefore, with equal justice be attributed to Einstein, since a few sentences at the start of his paper anticipate their result. Resonators fixed in a black-body cavity that also contains gas molecules should, Einstein points out, acquire mean energy $U = (R/N)T$ when repeatedly struck by those molecules. Planck has shown from electromagnetic theory, Einstein continues, that the equilibrium density u_ν of field energy is related to the resonator energy U_ν by the proportionality factor $8\pi\nu^2/c^3$. The Rayleigh–Jeans law for u_ν results, and Einstein pauses over it only long enough to note its impossible consequence: infinite energy in the radiation field.

Having reached this point, Einstein temporarily abandons the search for a black-body theory. Instead, he introduces Planck's law as the one that "satisfies all experiments to date,"[20] and he then proceeds to explore its high and low frequency limits. For low frequencies, where classical theory and experiment agree, he derives the relationship between black-body and atomic constants, thus rendering Planck's most striking result "to a certain extent independent of his theory of 'black radiation.'"[21] For high frequencies, where experiment and theory diverge to infinity, he develops an argument designed to give physical structure to paradox. The Wien law, he suggests, though clearly not exact, has been well confirmed for large values of ν/T. In the region where it applies, he goes on to show, the entropy of radiation behaves, not like that of waves, but like that of particles. No basis for such behavior can be found in Maxwell's equations. Presumably it is their failure at high frequencies that accounts for the impossible difficulties of the Rayleigh–Jeans law.

To develop these points Einstein first supposes that radiation in a cavity of volume V and temperature T obeys the Wien distribution law, $u_\nu = \alpha\nu^3 \exp(-\beta\nu/T)$. By conventional arguments, familiar from Chapters III and IV,[22] he then shows that, if E is the total radiant

energy at frequency ν, its entropy can be determined from the distribution law and must be given by

$$S = -\frac{E}{\beta\nu}\left(\log\frac{E}{\alpha\nu^3 V} - 1\right), \tag{10}$$

a form equivalent to the one Planck had introduced by definition in 1899. Examining that formula, Einstein directs attention to the manner in which entropy varies for fixed energy and changing volume. If S_0 is the entropy corresponding to volume V_0, then equation (10) may be rewritten

$$S - S_0 = \frac{E}{\beta\nu}\log\left(\frac{V}{V_0}\right). \tag{11}$$

That relationship, Einstein at once points out, is precisely the one that governs "the variation with volume of the entropy of an ideal gas or of a dilute solution."[23]

To show what he has in mind, Einstein next introduces Boltzmann's probabilistic definition of entropy in the form

$$S - S_0 = \frac{R}{N}\log W, \tag{12}$$

where W is the *relative* probability of the state with entropy S compared to that of the state with entropy S_0. If a single molecule is known to be contained in a volume V_0 ($W_0 = 1$), then the probability that it is, in fact, located in a smaller volume V of the same container is just V/V_0. Correspondingly, if a gas has n molecules somewhere in V_0, then the probability that they are all in the smaller volume V is given by $(V/V_0)^n$, so that eq. (12) becomes

$$S - S_0 = n\left(\frac{R}{N}\right)\log\left(\frac{V}{V_0}\right).$$

That equation is identical in form with equation (11). High frequency radiation with energy E therefore behaves like a collection of n particles each with energy $\beta\nu R/N$. Furthermore, though Einstein does not say so for another year, the constant $\beta R/N$ has the same value as Planck's h, so that the energy of Einstein's light-particles is identical with the size of Planck's energy elements. A link between Planck's and Einstein's formulations of statistical radiation theory has at last appeared.