

The New Statistical Mechanics

By

KARL K. DARROW

THIS is the second article upon statistical mechanics which I have published this year in this Journal. The first, which appeared in the January (1943) issue, was devoted to the oldest form of the theory, which is variously known as the old, the classical, or the Boltzmann statistics. The word "statistics," I repeat from the former article, is a synonym for statistical mechanics, objectionable but (because of the length of the alternative) hardly to be avoided. The "new statistics," frequently divided into "the Bose-Einstein statistics" and "the Fermi-Dirac statistics," emerged in the middle twenties and ever since it has been gradually pushing its ancestor aside. In this article I propose to expound the new statistics, laying especial emphasis on the theory of monatomic gases, to which the former article was strictly limited.

A definition of statistical mechanics may well be asked for at this point, especially since in the former article I failed to give one. Like many other things either subtle or familiar, statistical mechanics cannot fully be defined till it is fully understood, by which time a definition may seem nugatory. As an attempt at an advance definition, I suggest that *statistical mechanics is the theory which, starting from the assumption that matter (and, in due course, radiation) is an assemblage of particles, undertakes to explain (1) entropy, (2) temperature, (3) specific heats, and (4) the distribution-in-energy of the particles in thermal equilibrium.* The critical reader may justly say that these are four aspects of a single problem, but I think it well to separate them notwithstanding. The word "particle" often has to be construed as standing for an elaborate structure, but in dealing with monatomic gases (and with radiation) we may let it stand for a point endowed with energy and momentum.

How does the classical statistics succeed in handling these four problems? To take them in reverse order: it does very well with the fourth, for material gases (but not for radiation). It does very well with the third, for monatomic gases (but not for polyatomic gases nor for radiation). It produces an adequate theory of temperature for monatomic gases, identifying the temperature with the mean kinetic energy of the atoms multiplied by a certain factor. It has a very strange adventure with entropy, producing a theory which in part is remarkably successful and in part is disconcertingly

fallacious, and has to be altered in awkward and dubious ways to be made completely successful.

To consolidate these statements and introduce the new theory, I review the Boltzmann statistics.

The N atoms of a gas in a container are represented by N numbered balls, identical in every way except the numbering. In the earlier article a game was proposed in which a collection of numbered baskets was provided for these balls, and the balls were tossed into the baskets in a predetermined way: N_1 of them into the first of the baskets, N_2 into the second, and so on until each of the baskets contained its preassigned number of balls with N_M in the M th or last basket. The set of numbers $N_1, N_2, N_3, \dots, N_M$ was called a "distribution," and the question was asked: in how many different ways can this distribution be realized? It demands a previous answer to another question: how can a given distribution be realized in more than one way? It is the numbering of the balls which makes this possible. If for instance we exchange two balls in different baskets, the distribution is not changed, and yet there is a difference between the second situation and the first, for an inventory of all the balls in all the baskets shows that in those two particular baskets the assortment has been changed. We thus have realized the same distribution in two different ways. If we had exchanged two balls in one basket, this would not have been regarded as a change; we should still be realizing the same distribution in the same way, in the sense of the Boltzmann statistics. It was shown in the earlier article that the number W of ways of realizing a distribution—or in more technical language, the number of complexions in the distribution—is given by the formula:

$$W = N! / \prod N_j! \quad (1)$$

I have said that in the Boltzmann statistics, the balls stand for the atoms of a gas. For what then shall the baskets stand? The baskets stand for compartments in space; but "space" may have several different meanings.

Giving "space" the ordinary meaning: imagine the gas contained in a box, and the box divided mentally (not physically!) into M compartments of equal volume. I called these by the name of "cells" in the previous article, but now, for a reason which will shortly appear, I rebaptize them "regions." These are the baskets. W has its smallest value, which is unity, when all of the atoms are in the same region. It has its greatest value, which is $N! / ([N/M]!)^M$, when in each of the regions there is the same number N/M of atoms. But this corresponds, as nearly as the picture is able to correspond, to the uniform spreading throughout the box which by vast experience we recognize as the natural permanent state of the gas "in equilibrium." The uniform distribution is outstanding because it has the

greatest value of W . If now we baptize W with the name of "probability," we may then say that the state which in Nature is the prevailing one is in the Boltzmann statistics the most probable one. In the theory it is described by the equation:

$$N_j = N/M \quad (2)$$

Now think of "momentum-space" in which there is a dot for every atom, and the Cartesian coordinates of the dot are the momentum-components p_x, p_y, p_z of the atom. The coordinates of the dot determine the energy E of the atom, by virtue (for material particles) of the relation:

$$E = (1/2m)(p_x^2 + p_y^2 + p_z^2) \quad (3)$$

This is a fact of the first importance, as will shortly appear. Let us divide the momentum-space into regions of equal volume. Each of the regions will correspond to a small range of energy-values, as a sample of which we may take any particular one among them. Therefore when we distribute the dots—or let me say simply, the atoms—in any manner among them, we have perforce a certain value of the total energy U of the gas, which we may consider as preassigned if we so wish. Now we are to compare this distribution only with such others as show the same value of U . Among these there is one which is outstanding because it has the greatest value of W . This was shown in the earlier article to be the canonical or Maxwell-Boltzmann distribution, described by the formula:

$$N_j = NA \exp(-BE_j) \quad (4)$$

in which N_j stands for the number of atoms in the region numbered j ; E_j for the value of E appropriate to that region, *i.e.* obtained by substituting into (3) the coordinates of some point in that cell; A and B for constants, whereof A depends on B while B depends upon U/N the average energy of the atoms of the gas. This distribution also is attested by experiment as being truly that of a gas in its normal natural abiding state of equilibrium.

Now I mention the concept of a six-dimensional space which comprehends both the ordinary space and the momentum-space, and is divided into six-dimensional regions of equal volume. By this device one is able to speak of (2) and (4) as two aspects of a single distribution in the " μ -space." This is the distribution outstanding among those with which it may legitimately be compared by reason of having the greatest W -value. It is the most probable distribution, in the sense given in the Boltzmann statistics to the word "probable."

This is the first triumph of the Boltzmann statistics, attained by numbering the atoms. Its other triumphs, and its ultimate confusion, come when

it copes with the task of interpreting entropy. But if I continue longer this review of the conclusions of the former article, the reader's tolerance may be exhausted. Let us make haste to find out how the newer statistics sets forth to find out the most probable distribution.

One of the most appealing features of the new statistics is, that it does not impose on atoms of a single kind that peculiar distinction which I described above as "numbering" them. We therefore now remove the numbers from the atoms, restoring thus to atoms of a single kind—it might for example be helium—that quality of absolute indistinguishability which the classical statistics took away from them in order to achieve its aims.

Having de-numbered the atoms, we start anew to play that very game with numbered balls and numbered baskets which we played in the classical statistics with such remarkable but incomplete success. But now that the atoms are de-numbered, they can no longer be the balls (nor, for that matter, the baskets). Something drastically new must now be done, and is.

In the new statistics, the balls stand for the compartments and the baskets for the populations.

I must define the word "population." It means the number of atoms in a compartment, or as I will say from this point onward, in a "cell." The balls which are tossed into the basket numbered 0 stand for the cells containing no atoms; the balls which go into the basket numbered 1 correspond to the cells containing one atom apiece, and so forth indefinitely. C_i shall be the symbol for the number of balls in the i th basket, which is to say, the number of cells containing i atoms apiece. C shall stand for the total number of cells.

Let the cells at first be compartments of equal volume in the ordinary space, obtained by dividing up (mentally) the box containing the gas. For the number of complexions or inventories corresponding to a given distribution, defined by given values of the quantities C_i , we have as before:

$$W = C!/\Pi C_i! \quad (5)$$

and taking the logarithm:

$$\ln W = C \ln C - \Sigma C_i \ln C_i \quad (6)$$

In using this expression I have again, as often in the previous article, assumed the validity of what I there called "the super-Stirling approximation"; but notice that this no longer means that I assume each of the cells to enclose an enormous number of atoms—it means instead that there is an enormous number of cells having each particular population.¹

¹ Clearly this cannot be so for all populations no matter how great! This is a difficulty which also pops up in the old statistics, though there it is not met until the ordinary space is replaced by the momentum-space.

We seek a set of values C_i such that when it is realized, the quantity W shall have a value stationary with respect to all variations δC_i conforming to *two* conditions: first, that the number of cells shall remain the same, which is to say, $\Sigma \delta C_i$ shall vanish; and second, that the number of atoms shall remain the same, which is to say, $\Sigma i \delta C_i$ shall vanish.

Such a set is the following:

$$C_i = C \alpha e^{-i\beta} \quad (7)$$

α and β standing for constants yet to be determined; for taking the first variation of $\ln W$ from (6), we find:

$$\begin{aligned} \delta(\ln W) &= -\Sigma \delta C_i (1 + \ln C_i) \\ &= -\Sigma \delta C_i (1 + \ln C \alpha) + \beta \Sigma i \delta C_i \end{aligned} \quad (8)$$

and the required condition is fulfilled. Assuming without proof that the *stationary* value of W is also a maximum value, and referring to W as the "probability" of the distribution of cells among populations, we have come to the startling conclusion that the most probable distribution is the one given by (7)!

I call this a startling conclusion, because it contravenes our inbred conviction that the natural distribution of a gas in ordinary space is the uniform distribution. Of course, in the last two sentences I have used the word "distribution" in two senses, and this must be rectified at once. What I have just called "the uniform distribution" is the uniform distribution in the *old* sense—the same number of particles in every cell. In the new sense of the word, this is a distribution in which all of the cells have the same population, and therefore in which one basket contains all of the balls. Definitely, this is not, in the new statistics, the most probable distribution! Indeed it is not even a conceivable distribution, for the number of cells is infinite.

To mitigate this clash of theory with experience we can do nothing else than assume our cells to be so tiny that in any region of the gas large enough to be surveyed by observation, there is a mighty number of the cells. Then at worst we can take it from experience that in the normal natural abiding state of the gas the number of atoms in each region will be the same if all the regions are of equal volume, while within each region we can distribute the atoms among the cells as the new statistics tells us to. However, it may yet be possible to come to this conclusion from the theory. In preparation for the effort, I sketch the procedure for evaluating the constants α and β in the distribution (7).

A similar task was set before us in the earlier article: that of evaluating the constants of the Maxwell-Boltzmann law in terms of the total number and the total energy of the atoms. Here for any region we are to evaluate the constants α and β in terms of the number of cells C and the number of

atoms N . The task is greatly eased by the opportunity of using two well-known formulae:

$$1 + x + x^2 + \dots = (1 - x)^{-1} \quad (9)$$

$$1 + 2x + 3x^2 + \dots = (1 - x)^{-2} \quad (10)$$

The start is made from the two self-evident equations:

$$C = \sum C_i = C \sum \alpha e^{-i\beta} \quad (11)$$

$$N = \sum i C_i = C \sum i \alpha e^{-i\beta} \quad (12)$$

By putting x for $e^{-\beta}$ and using (9) and (10), the student can easily win through to the results,

$$\alpha = (1 - e^{-\beta}), \quad C/N = e^{\beta} - 1 \quad (13)$$

and then to the final form of the distribution-law (7):

$$C_i = C \frac{C}{N + C} \left(\frac{N}{N + C} \right)^i \quad (14)$$

and finally, after consulting (5), to the expression for the number of ways W_{\max} in which this the most probable distribution—"most probable" in the eyes of the new statistics—can be realized. Its logarithm is:

$$\ln W_{\max} = C \ln \frac{N + C}{C} + N \ln \frac{N + C}{N} \quad (15)$$

This is the most important formula of the new statistics, as will presently be clear.

Divide now the space containing the gas into "regions" of equal size, each comprising the same number C of cells, which number shall be great. For the benefit of those to whom the memory of the previous article may still be vivid, I say now that insofar as there is any correspondence of the new to the old statistics, these "regions" correspond to the "cells" of the older theory. This is the reason why, in my recent brief synopsis of the old statistics, I used the word "region" to replace the word "cell" used in the prior article. Let the subscript j be the marker for these regions, so that N_j shall stand for the number of atoms in the j th region. Put N_j for N in (15). Now each member of (15) refers explicitly to the j th region, and on the left I should put $(\ln W_{(\max)j})$, but for two purposes—one of which is brevity, while the other will appear in due time—I put $\ln W_j$ instead:

$$\ln W_j = C \ln \frac{N_j + C}{C} + N_j \ln \frac{N_j + C}{N_j} \quad (16)$$

The quantity W_j is an odd sort of "probability" relating only to the contents of the region j . It is, to repeat, the number of ways in which the

most probable distribution of the C cells among the possible populations can be realized, there being N_j atoms in the region. Nothing of the sort appeared in the old statistics.

Now form the product of all of the quantities W_j . This is a "probability" relating to the entirety of all the regions, therefore to the whole of the gas. It is the total number of ways in which the most probable distribution can simultaneously be achieved within each of the regions. It is taken to be the total number of ways in which the most probable distribution of the gas-as-a-whole can be realized; for in the new statistics we have no other way of defining the most probable distribution of the gas-as-a-whole, than this way of subdividing first into cells and then into regions comprising many cells. The symbol for this product shall be W , for although I have already used that symbol in this article, its former meaning is now taken over by W_j , and it is free again. We have:

$$\ln W = \sum \ln W_j \quad (17)$$

the quantities W_j being still those functions of N_j which were shown in (16).

We seek now a set of values of N_j such that when it is realized, $\ln W$, and therefore W also, shall have a value stationary with respect to all variations δN_j conforming to the sole condition that the total number of atoms shall remain the same, which is to say, $\sum \delta N_j$ shall vanish.

It may be recalled that a similar problem arose in the old statistics. I treat it here in a more general and hardly less simple way, by writing the self-evident equation:

$$\delta \ln W = \sum \frac{d(\ln W_j)}{dN_j} \delta N_j \quad (18)$$

For the fulfilment of our wish it is a sufficient condition that all of the derivatives on the right-hand side should have the same value; since than $\delta \ln W$ will be $\sum \delta N_j$ multiplied by a constant, and when one vanishes so will the other. For this it is in turn a sufficient condition that all of the independent variables N_j should have the same value.

Uniform spreading of the atoms among the regions, with equal numbers in all regions of equal size, is therefore the condition in which $\ln W$ has a "stationary" value, which as always is assumed to be a maximum value. With the new definition of probability, the state of uniform spreading becomes the most probable in the new statistics, as with the old definition of probability it was in the old.

We go into the momentum-space to see whether the Maxwell-Boltzmann law results from the new statistics.

The momentum-space is now to be divided into regions of equal size,

each large enough to comprise a great number C of cells and small enough so that the function E of equation (3) may be deemed sensibly constant throughout it; E_j shall stand for the value of E appropriate to the region. (It is convenient to imagine the regions as layers separated from one another by concentric spheres having the origin for their common centre). For each of the regions W_j , the number of ways in which the most probable distribution of the cells among the possible populations can be realized, is given again by (16); and W , the number of ways in which the most probable distribution can simultaneously be achieved within each of the regions, is given again by (17).

We now seek a set of values of N_j such that when it is realized, $\ln W$ shall have a value stationary with respect to all variations δN_j conforming to two conditions: first, that the total number of atoms shall remain the same, which is to say, $\Sigma \delta N_j$ shall vanish; and second, that the total energy of the gas shall remain the same, which is to say, $\Sigma E_j \delta N_j$ shall vanish.

Referring back to (18), we see that for the fulfilment of our wish the following is a sufficient condition:

$$\frac{d(\ln W_j)}{dN_j} = P + QE_j \quad (19)$$

P and Q standing for constants; for when these substitutions are made into every term of the summation on the right of (18), the expression to which $\delta \ln W$ is there equated may be regrouped into one term proportional to $\Sigma \delta N_j$ and one proportional to $\Sigma E_j \delta N_j$, and vanishes when it ought to vanish.

Gone is the comfortable ease with which we disposed of the corresponding problem in the ordinary space! There we did not even have to know what sort of function W_j is of N_j ; whatever it might be, we were able to conclude that N_j must be the same for every region. Here the outcome must depend upon the functional relation between W_j and N_j . There is, however, no ground for apprehension, for though the function in question looks rather involved in equation (16), its derivative is surprisingly simple, and we come with ease to the condition which we seek:

$$\ln(N_j + C) - \ln N_j = P + QE_j \quad (19)$$

which may be rewritten thus:

$$\frac{C}{N_j} = -1 + e^{P+QE_j} \quad (20)$$

This is *not* the Maxwell-Boltzmann law, but approaches that desired law in what I will call the "limit of extreme rarefaction," where the number of cells in the region exceeds manifold the number of atoms. As C/N_j grows

greater and greater, the first term on the right recedes into relative insignificance; and with an ever-increasing degree of approximation, we have:

$$N_j = NAe^{-BE_j} \quad (21)$$

with NA put for Ce^{-P} and B for Q —which is the Maxwell-Boltzmann law and the law confirmed by experiment.

A helpful and troublesome coincidence between two different quantities

When in the earlier article I used the section-heading repeated just above, it referred to the near-equality between the logarithms of two different numbers: one number being that of all the complexions compatible with the *most probable distribution* (of numbered atoms sprinkled among numbered cells in ordinary space) and the other being that of all the complexions compatible with *all conceivable distributions altogether*. The most probable distribution had so great a share of all conceivable complexions, that no grave error was committed in pretending (so long as we were dealing with $\ln W$) that it actually had them all without exception!

A similar coincidence occurs in the new statistics, and will now be set forth.

Consider the j th region by itself. In (15) I have given the expression for $\ln W_{\max}$, the logarithm of the number of ways in which the *most probable distribution* of cells among populations can be realized. This is now to be compared with $\ln W_{\text{tot}}$, the logarithm of the total number of ways in which *all possible distributions* of cells among populations can be realized. Note that I say "all possible" and not "all conceivable" distributions! The only possible ones are those which are compatible with the fixed number N_j of atoms. This limitation prevents us from proceeding by the easy route of the earlier article. Indeed in order to solve the problem "in how many ways can all possible distributions of cells among populations be realized?" it is necessary, or at any rate customary, to restate it in a very different manner, which is the following:

In how many different ways can N_j un-numbered balls be distributed among C numbered baskets? Two ways are considered as different unless $n_i = n'_i$ for every value of i (n_i and n'_i standing for the populations of the i th basket in the two ways).

Notice that again the balls stand for the atoms and the baskets for compartments in space, as they did in the old statistics! We are playing a new game with the old baskets and the old balls, instead of playing the old game with new balls and new baskets as we have just finished doing. It has to be a new game, for the numbers have been removed from the old balls and the old game is therefore unplayable.

This is, to put it mildly, one of the less perspicuous problems of the

"theory of probability." I cannot do better than repeat, with slight changes in wording, the process of solution given by the Mayers in their book.² "A distribution is characterized by the number of balls in each of the numbered baskets, since the balls are indistinguishable. Consider the arrangement in a line of the symbols $z_1, z_2, \dots, z_C, a_1, a_2, \dots, a_N$, as for instance,

$$z_1 a_2 a_4 z_8 a_8 a_9 z_5 z_4 z_9 a_6 \dots$$

Such an arrangement could be used to define an assignment of N numbered balls, the a 's, to C numbered baskets, the z 's, by adopting the convention that the balls to the right of each numbered z belong to the basket of that number. For instance, the above corresponds to balls 2 and 4 in basket 1, balls 8 and 9 in basket 8, no balls in baskets 5 or 4, and ball 6 in basket 9. One must observe the convention that the row starts with a z , and we shall consider only arrangements of the symbols which start with z_1 . However, $(C - 1)!$ such arrangements of the symbols correspond to one arrangement of numbered balls in the same baskets, since permutations of the $(C - 1)$ groups of each z_i with its following a 's correspond to the same arrangement of numbered balls in the baskets. In addition all $N!$ permutations of the a 's correspond to the same distribution of un-numbered and indistinguishable balls in the baskets. In all, each distribution of the indistinguishable balls among the numbered baskets corresponds to $N!(C - 1)!$ arrangements of the symbols, and the $(N + C - 1)$ symbols (after the first) may be arranged in $(N + C - 1)!$ different ways."

Thus we come to the formula³ for W_{tot} , the total number of ways in which *all possible distributions* of the cells of a region among the populations can be realized; it is,

$$W_{\text{tot}} = \frac{(N_j + C - 1)!}{N_j!(C - 1)!} \quad (22)$$

Dropping the "ones" for the amply sufficient reason that they are insignificant by comparison with N_j and C , and taking the logarithm with use of the super-Stirling approximation, we find:

$$\ln W_{\text{tot}} = (N_j + C) \ln (N_j + C) - N_j \ln N_j - C \ln C \quad (23)$$

which with a little regrouping of terms is found to be the very same expression appearing in (15) for $\ln W_{\text{max}}$.

² J. E. and M. G. Mayer, "Statistical Mechanics" (John Wiley & Sons, 1940); p. 438. Reprinted by permission.

³ For the historian of science it is interesting to note that the formula (22) was used by Planck in his earliest derivation of the black-body radiation law. His un-numbered balls were quanta of energy, his baskets were linear oscillators, and his $k \ln W$ was the entropy of the system of C oscillators sharing N_j quanta among themselves. Cf. *Naturwissenschaften*, April 2, 1943.

I review the situation. We began by dividing space (ordinary space, or momentum-space, or μ -space) into what were called cells in the earlier article and are now designated "regions." We wanted to reach, as most-probable-distribution of the atoms among the regions, the uniform spread in ordinary space and the Maxwell-Boltzmann law in momentum-space. To do this by the method of the new statistics, we divided each region into many "cells." The first stage of the argument then consisted in taking a typical region, and ascertaining the most probable distribution of the cells among the populations. We then evaluated W_{\max} , the number of ways in which this distribution could be realized. Inserting $\ln W_{\max}$ into the argument we continued into the second stage, and attained the wanted result. But now it turns out that in the first stage we might have omitted to ascertain the most probable distribution of the cells among the populations in the typical region. Anybody might win through to the same desirable outcome without even suspecting that there is a most probable distribution of cells among populations. All he needs is to evaluate W_{tot} , the number of ways in which all possible distributions of cells among populations within the region can be realized. He may then replace $\ln W_{\max}$ by $\ln W_{\text{tot}}$, and proceed with the second stage as before. Since the two logarithms are practically equal, the outcome is the same.

There are accordingly two routes to the result, which do not merge until the argument is carried partway to the conclusion. Is one of them right and the other wrong? Or to ask a milder question: is either to be preferred to the other?

So far as I can see, neither can be proved wrong, and the question must be asked in the milder form. For myself I stand by the preference exhibited in this article, for the basic reason that along this route each of the stages of the argument consists in finding a most probable distribution: first for the cells among the populations of each region by itself, and then for the atoms among the regions. By the other route the two stages are differently handled, since in the first stage one considers all the distributions (of cells among populations in each region by itself) and then in the second stage the most probable distribution (of atoms among regions). There is also the minor advantage, that the value of W_{\max} is much easier to derive than the value of W_{tot} , or at least so it seems to me.⁴ However, many physicists of eminence have preferred the second route. Anyone may say of course that the question is foolish, since the number of complexions subsumed under the most probable distribution is so large a fraction of the total number of complexions altogether that no danger arises from confusing them. This is what the equations have been saying, and now I have said it again in words.

⁴ It was the other way about in the somewhat similar case which was treated in the earlier article.

Yet any policy which leaves this basic law unsaid, or even fails to emphasize it, is (I think) a bad one for all but the very few to whom it is already obvious.

Another question: may the old statistics be regarded as the limiting form of the new statistics, in the limiting case of "extreme rarefaction" where in every region the number of atoms is very much smaller than the number of cells?

It may seem that this question has already been answered with a *yes*, in view of the fact that in the limiting case the new statistics gives the same distribution-law—in momentum-space as in ordinary space—as the old one does. Nevertheless the answer is *no*. Mathematically this appears in the following way. In the old statistics the Maxwell-Boltzmann law springs from the denominator $\prod N_j!$ in the right-hand member of equation (1), which turns up as $(-\sum N_j \ln N_j)$ in the expression for $\ln W$. Now we look at the equation (23) and see the term $-N_j \ln N_j$ appearing with several other terms in the right-hand member. In the limit of extreme rarefaction it outweighs all the others and survives by itself. We summed it over all the regions and so arrived again at $(-\sum N_j \ln N_j)$, from which again the Maxwell-Boltzmann law emerged. But in this method of the new statistics each term of the summation comes by itself from the corresponding region, whereas in the method of the old statistics the whole summation arrived upon the scene *en bloc* or all in a single piece. The former method does not pass into the latter method in the limiting case. The conclusions agree in the limit, but the methods do not.

I have mentioned this because not infrequently one finds in print the careless statement that the old statistics is the limiting case of the new statistics, or words to that effect. Actually one can find more potent ways of contradicting that statement, as for example by emphasizing that the old statistics numbers the atoms and the new one leaves them un-numbered, and in no way can the one policy be regarded as a limiting case of the other. More convincing yet would it be to show that the new statistics and the old lead to results which definitely differ even in the limiting case of extreme rarefaction. This is what I next undertake to show as an incident of the explanation of entropy which the new statistics affords.

THEORY OF ENTROPY

For a substance of a single kind in a single phase, the basis of thermodynamics is the single equation,

$$dU = TdS - PdV \quad (31)$$

in which there are five variables: pressure P , volume V , absolute temperature T , energy U and entropy S . Two may be varied independently, and

any two of the five may be taken as these two, the remainder becoming the dependent variables.

From (31) we deduce, to begin with,

$$T = (\partial U/\partial S)_V \quad (32)$$

an equation which shows that if ever someone sets up a theory in which entropy is expressed as a function of energy and *vice versa*, it is *per se* a theory of absolute temperature. This, however, will find its due place later. What is of instant importance is a second deduction,

$$(\partial S/\partial T)_V = T^{-1}(\partial U/\partial T)_V \quad (33)$$

for making use of which we take note of the fact (not explicitly stated till now) that when the volume does not change no mechanical work is done upon or by the substance, and therefore all of the change in energy is that brought about by the inflow or outflow of heat. This fact is expressed in another equation,

$$(\partial U/\partial T)_V = H_v \quad (34)$$

H_v standing for the amount of heat that must be fed into the substance to raise its temperature, at constant volume, by one degree—the “heat-capacity at constant volume,” as some would call it. Combining the two,

$$(\partial S/\partial T)_V = H_v/T \quad (35)$$

Envisage now the entropy S as a function of volume and temperature, and view the equation:

$$dS = (\partial S/\partial V)_T dV + (\partial S/\partial T)_V dT \quad (36)$$

An equivalent for the coefficient of dT has been provided, and now it is needful to find one for the coefficient of dV . To do this we use the function $(U - TS)$, to be denoted by A , which by aid of (31) is seen to have the following differential:

$$dA = -PdV - SdT \quad (37)$$

Out of this one draws the following two deductions,

$$(\partial A/\partial V)_T = -P, \quad (\partial A/\partial T)_V = -S \quad (38)$$

Differentiating both sides of each of these equations, the former with respect to T while holding V constant, the latter with respect to V while holding T constant, one gets two expressions for what is one and the same quantity, to wit, the second derivative $\partial^2 A/\partial T\partial V$. Equating these two expressions, and saying goodbye to A which has fulfilled its purpose, one has,

$$(\partial S/\partial V)_T = (\partial P/\partial T)_V \quad (39)$$

which by the way is one of four equations collectively known as Maxwell's relations—a memorial of Maxwell's creative work in early thermodynamics, as the name of the distribution-in-energy law is of his work in early statistics.

Substituting from (35) and (39) into (36), we find:

$$dS = (\partial P/\partial T)_v dV + (H_v/T)dT \quad (40)$$

a usable and a useful expression for the entropy S —usable, that is to say, to anyone who knows the heat-capacity H_v and the derivative $(\partial P/\partial T)_v$, as functions of volume and temperature, for the substance in question.

Now take it on faith that there exists a gas having the following qualities: *first*, its pressure and volume and absolute temperature are linked together by the equation,

$$PV = LT \quad (41)$$

L being a constant; whence follows,

$$(\partial P/\partial T)_v = L/V \quad (42)$$

second, if the attempt is made to express its energy U as function of T and any one of the remaining variables (to wit, P or V or S), then the latter variable drops right out of the picture, leaving U as a function of T alone. I suppose this seems a needlessly longwinded way of saying that U does not depend on P or V , but it is necessary to provide for the fact that U may be expressed as (say) a function of S and V , whereupon it will be found that neither variable drops out of the picture. This is one of the features that make the science of thermodynamics very like a maze.

third, the heat-capacity H_v is independent of all the variables.

With these stipulations, (36) becomes:

$$dS = (L/V)dV + (H_v/T)dT \quad (43)$$

integrating which, we readily find that for the peculiar kind of gas presented just above as an article of faith, the entropy is given by the formula:

$$S = L \ln V + H_v \ln T + C \quad (44)$$

The symbol C stands for one of the most useless things in the world: an arbitrary additive constant of integration. The only purpose normally served by such a constant is, to prevent people from thinking that the equation is right if the constant is left off. Its presence means that the absolute value of S is undeterminable, is beyond the reach of experiment to determine. Nevertheless this constant is one of the principal themes of statistical theory; and we shall see that in defiance of what I have just said, and no part of which I retract, it *does* make sense to assign a particular value to this constant, and remarkably good sense at that.

I propose to begin very soon on the proof that the new statistics, applied to a flock of atoms which are merely mass-points, gives an excellent description of just such a gas. However, there is a detail, or rather an element of the structure, waiting to be inserted correctly—a trivial one in appearance, but in all of thermodynamics and all of statistics there is nothing further removed from the trivial. It is the dependence of the entropy on the quantity of the substance, the dependence of S upon N the number of atoms. To put a question seemingly so simple that it almost answers itself: given two samples of the same kind of gas under identical conditions, one comprising twice as many atoms as the other, what is the ratio of their entropies?

This is a remarkable question, because it seems so absurdly simple and is actually so very complex.

Before the advent of statistical theory, anyone versed in thermodynamics would probably have answered it by replying either that S is proportional to N , or that the question has no meaning. The first reply is suggested by the consequences of the fact that thermodynamics proposes no way of measuring the entropy of a gas (or other substance), but only ways of measuring the entropy-difference between two states. Let V_1, T_1 and V_2, T_2 stand for the values of V and T in two states of one gas. Equation (43) informs us that the entropy-difference is $L \ln (V_2/V_1) + H_v \ln (T_2/T_1)$. The constant C has vanished; the remaining terms are proportional to N because L and H_v are proportional to N . The entropy-difference is therefore proportional to N . It seems reasonable to conclude that S is proportional to N , but so long as there is no specific assertion about C the conclusion is not binding; and the proper reply is actually, that the question has no meaning.

But the statistical theories do make assertions about C , and the question is on the verge of acquiring a meaning; so it might be a good idea to ask in advance what sort of answer we should like to have. It seems natural to expect S to be proportional to N , so that the "double sample" shall have twice the entropy of the "single sample" under identical conditions. But what are "identical conditions?" Here is the catch. No more than two of the three variables P, V, T can be made the same for both the samples. I suppose that almost anyone would choose T for one of these two, so unpalatable would it seem to expect the double sample to have twice the entropy of the single sample if their temperatures differed. But after this is decided, shall we make V the same for both, and accordingly give doubled pressure to the double sample? or shall we make P the same for both, and accordingly give doubled volume to the double sample?

This is no mere quibble, for the choice will determine the dependence of C on N .

The first alternative requires that C be proportional to N . This is

obvious from inspection of (44) when one remembers that L and H_v are both proportional to N . I rewrite that equation accordingly, and put N_0 for the number of atoms in a gramme-molecule, R and C_v and RK_0 for the values of L and H_v and C appropriate to N_0 atoms:

$$S = (N/N_0)R \ln V + (N/N_0) C_v \ln T + (N/N_0)RK_0 \quad (45)$$

S is doubled if N is doubled while V and T stay the same, which is what was intended.

This equation will not suit the second alternative; for if V is doubled along with N , S will be more than doubled. Over and above the doubling, S acquires an extra term $2(N/N_0) R \ln 2$. Now if the constant C just happened to include a term $-2(N/N_0) R \ln 2$, the extra term would be obliterated, and S would just be doubled if N and V were to be doubled while P and T remained the same. Such a term is provided by replacing (45) with the equation:

$$S = (N/N_0)R \ln V + (N/N_0)C_v \ln T - (N/N_0)R \ln N + (N/N_0)RK_0 \quad (46)$$

where the last two terms on the right are to be regarded as forming the constant C . This then is the dependence of C on N which is demanded by the second alternative.

To guide the choice between the two alternatives there is, so far as I know, but the one argument; it is, however, a powerful one, and seems likely to hold the field unchallenged.

We have been thinking of two samples of identical gas at identical temperature. Think of them now as divided by a removable partition. When the partition is taken away, what happens? If the initial pressures are not the same, there is a swirling and a surging, dying away in time into a state in which the pressure is the same throughout the volume now common to the samples, but is not the same as it was before in either separate gas. This is just the sort of trend of events with which one likes to think that an entropy-change, and indeed an entropy-gain, is linked. Notice also that if the partition is replaced, the state of affairs on either side does not become the same as it was before! But now suppose the initial pressures to be the same. The partition can be removed and replaced without entailing any perceptible change in the gas such as one likes to associate with a change in entropy.

The second alternative is in harmony with these facts, the first is not. So to the question "is the entropy of a gas of $2N$ atoms double the entropy of a gas of N atoms?" the acceptable answer is: "yes, if the volume of the double gas is twice that of the single gas, their temperatures being the same." Now, this is also the answer given by the new statistics; for as we shall presently see, it leads to a formula like (46). It is not the answer given by the old statistics, which (as I said in the earlier article) leads to a formula

like (44). This is one of the dominant reasons for preferring the new statistics to the old.

Now I proceed to the theory of entropy and temperature derived from the new statistics.

NEW STATISTICAL THEORY OF ENTROPY

Entropy is identified with the quantity $\ln W$, multiplied by a constant k which as yet is disposable:

$$S = k \ln W \quad (47)$$

It is now to be shown that for the picture of a gas which is a flock of mass-points in the "most probable state" as defined by the new statistics, and in the limit of extreme rarefaction, this expression becomes the same as (46), with further consequences of much value.

As in the previous article, I separate the entropy into S_v the "contribution of volume to entropy" which springs from the sprinkling of the mass-points in ordinary space, and S_m the "contribution of temperature to entropy" which springs from the sprinkling of the mass-points in momentum-space. This is an artificial separation and worse than artificial, for it leads to a fault in a detail which is not trivial. Nevertheless I think that for ease of exposition the procedure is justified, and the detail will be made correct at the end of the argument.

We must now take (16) down to the "limit of extreme rarefaction." I repeat this equation:

$$\ln W_j = (N_j + C) \ln (N_j + C) - C \ln C - N_j \ln N_j \quad (16)$$

The journey toward the limit is menaced by some of the oddest pitfalls, and must be travelled with care. I recall that by Taylor's expansion, $\ln (N_j + C)$ is equal in first approximation to $(N_j/C + \ln C)$ when N_j is small by comparison with C . Making this substitution into (16), one finds that the right-hand member consists of six terms. The two largest of these, $C \ln C$ and $-C \ln C$, destroy one another. The smallest, N_j^2/C , is to be neglected (if we couldn't neglect it, the dependence of entropy upon N would be hopelessly misrepresented). All of the remaining three terms must be kept, for even the smallest—which is N_j —will play a perceptible part in the check of theory with experiment. We have:

$$\ln W_j = N_j \ln C - N_j \ln N_j + N_j \quad (48)$$

The quantity $\ln W$ is the summation of $\ln W_j$ over all the regions. Notice that we are interpreting entropy in such a way, that the entropy of the gas in the container is the sum of the entropies of the portions thereof in the individual regions. This is why we are destined to come to a result

in harmony with the "second alternative" aforesaid, wherein the total entropy of two gases of identical P and T is the sum of their separate entropies! It is otherwise in the old statistics, and that was the source of the troubles of that elder theory. But to proceed:

$$\ln W = \sum \ln W_j = N \ln C - \sum N_j \ln N_j + N \quad (49)$$

In the ordinary space and in the most probable state, N_j is the same for all the regions, as we have found already (page 363) and therefore is equal to N divided by V/V_0 , or to NV_0/V ; here V_0 is the volume of the *region* (not the cell!). The next step is to put this into (49), realizing now that each term becomes the same and the whole summation is V/V_0 times the typical term. One is agreeably surprised to find that V_0 tumbles out of the expression: this is a feature of the new statistics—the regions have but an intermediate and an auxiliary quality, the size assigned to them is gone from the final equations. In its place appears the volume of the *cell*, which is V_0/C , and which I denote by q_c . For S_c we have:

$$S_c = Nk \ln V - Nk \ln N - Nk \ln q_c + N \quad (50)$$

Note the last three terms, for future comparison with the two last of (46); but at this moment note especially the first, and compare it with the first of (46). Entire agreement is attained by assigning to k the value,

$$k = R/N_0 \quad (51)$$

as in the old statistics. The "Boltzmann constant" k is the "gas-constant" R divided by the "Avogadro number" N_0 .

Seeking now the "contribution of temperature to entropy," S_m , we turn to the momentum-space. Here the most probable distribution is given by (21), and is to be inserted into (50):

$$\begin{aligned} \ln W &= N \ln C - \sum N_j \ln N_j + N \\ &= N \ln C - N \ln A + N \sum ABE_j e^{-BE_j} + N \end{aligned} \quad (52)$$

It will be recalled from the earlier article, or failing this can easily be seen, that,

$$N \sum A^{-BE_j} = N, \quad N \sum AE_j e^{-BE_j} = U \quad (53)$$

U standing as heretofore for the total energy of the gas. The expression (52) is simplified of aspect, and multiplying it by k , we find for S_m ,

$$S_m = k \ln W = kN \ln C - kN \ln A + kBU + kN \quad (54)$$

Though I have spoken of this as the contribution of temperature to entropy, the temperature is nowhere to be seen! It is waiting on the doorstep; but before allowing it in, I wish to operate on the quantity $\ln A$.

This quantity, by (53), is given thus:

$$-\ln A = \ln \sum e^{-BE_j} \quad (55)$$

The so-called "partition-function," which is the sum appearing on the right, is made of terms contributed one from each region, the term from the j th region being $\exp(-BE_j)$ —herein E_j stands (it is proper to say) for the average value in the j th region of the function,

$$E = (1/2m) (p_x^2 + p_y^2 + p_z^2) \quad (56)$$

If the regions were of unit volume, this summation would be (approximately) equal to the triple *integral* of $\exp(-BE_j)$ over the whole of momentum-space. But the volume of each region is $C_j q_m$, wherein q_m stands for the volume of the *cell*. This signifies that while the integral contains one term for each unit of volume, the summation comprises $1/C_j q_m$ terms for each unit of volume. The summation is accordingly $1/C_j q_m$ times as great as the integral. Denoting the integral by I , we have in place of (55),

$$-\ln A = \ln I - \ln q_m - \ln C_j \quad (57)$$

and in place of (54),

$$S_m = kN \ln I - kN \ln q_m + kBU - kN \ln N + kN \quad (58)$$

We note with satisfaction that the size of the *region* has disappeared, even as it did while we were operating in ordinary space!

The next step is to consult a table of definite integrals for the value of the integral I (or to work it out one's self, if one's memory of the mathematical technique is vivid). The tables give:

$$I = (2\pi m/B)^{3/2} \quad (59)$$

Before returning the table of integrals to the library, the student should also look up the value of the definite integral $\int_0^\infty x^2 e^{-ax} dx$; for with its aid he will be able to find a very simple relation between B and U . I have already said that either determines the other, and now for this special case we shall find the relationship. The procedure consists in going back to the second of equations (53), realizing that

$$\sum E_j \exp(-BE_j) = (1/C_j q_m) \iiint E \exp^{-BE} dp_x dp_y dp_z \quad (60)$$

and performing the triple integration over the whole of momentum-space, a feat which is not so hard as it looks. Multiplication by A , as indicated in (53), removes the factor $(1/C_j q_m)$, and the simple conclusion is,

$$U = 3N/2B \quad (61)$$

a relation valuable in two ways.

In the first place, (61) enables us to eject either B or U from the expression for S_m . It seems more sensible to do away with B , leaving S_m expressed as a function of the energy of the gas; but I will let the reader do that for his own instruction. In view of the peculiar significance of B this is for our purposes the better one to keep.

Now it is high time indeed to show what is that peculiar significance. Differentiating S with respect to U , we find kB for the derivative. Turning back to (32) we are reminded that this derivative is $1/T$ by the definition of the absolute temperature T . Now the temperature has stepped across the threshold, and S_m assumes the form:

$$S_m = (3/2) Nk \ln T - Nk \ln N + Nk \ln [(2\pi mk)^{3/2} e^{5/2}/q_m] \quad (62)$$

Notice that the term $+ kN$ has been absorbed into the final term, so that $e^{3/2}$ has been replaced by $e^{5/2}$ in the argument of the logarithm: this is a usage with which the student must become familiar. (Some writers also incorporate $-Nk \ln N$ into this final term, which thereby acquires a factor N in the denominator of the argument; the term then ceases to be a constant, which is why I do not follow this policy.)

Comparing (62) with (46) we see that S_m embodies correctly the dependence of entropy on temperature, provided that C_v (the specific heat per gramme-molecule) is equal to $(3/2)kN_0$. Since a value for k —to wit, R/N_0 —has already been forced upon us as a necessary and a sufficient condition for making S depend correctly on the volume, this new requirement is that C_v should be equal to $(3/2)R$. Now this is a fact of experience for the gases called monatomic!

I said that the relation of U and B expressed in (61) is valuable in two ways. The second is only the first seen from a different viewpoint, for which I rewrite (61) in the form:

$$U/N = U = (3/2)kT \quad (63)$$

For the flock of mass-points distributed in momentum-space in the manner indicated as the most probable by the new statistics (as, for that matter, by the old) the average energy is $(3/2)k$ times the absolute temperature. This is the very result obtained from simple kinetic theory for the ideal-gas scale of temperature. The statistical theory therefore identifies the absolute scale of temperature with the ideal-gas scale, which is as it should be. It is therefore an adequate theory of temperature and (as we lately saw) of the specific heat of monatomic gases.

Now I have given an expression for S_v , the "contribution of volume to entropy," which is (50); and an expression for S_m , the "contribution of temperature to entropy," which is (62); and it seems natural to proceed by

adding the two and identifying their sum with the entropy of the gas. But each of the summands contains the so-ardently-wanted term $-Nk \ln N$, and therefore the sum must contain a term $-2Nk \ln N$, which is $-Nk \ln N^2$. This term is not at all of the wanted form, and its mere presence in $(S_m + S_e)$ spoils the chance of identifying that sum with entropy. We have in fact come to a result in contradiction with equation (46) and with the assumption on which that equation was founded, viz. that if two samples of a gas are at the same pressure and temperature their entropies are in proportion to their volumes. The reasoning has not been suited to its aim.

The origin of this final misadventure lies in the circumstance that in the hope of making easier the exposition, I made what now has proved to be an undue separation between the two "contributions" to the entropy. The gas was mentally divided into groups of atoms, each occupying a certain region of limited size and distribution among the cells of that region according to the law of the new statistics. In computing S_e I defined the j th region as a small piece of ordinary space, and then counted all the atoms in that region regardless of the fact that they have very diversified momenta. In computing S_m I defined the j th region as a small piece of momentum-space, and then counted all the atoms in that region regardless of the fact that they are sprinkled all through the total volume of the container. I may properly say that I used a six-dimensional region throughout, but in the first stage it was a region limited in ordinary space and comprising the whole infinity of momentum-space, while in the second stage it was a region limited in momentum-space and comprising the whole volume of the box in ordinary space. I should instead have carried through the operation in a single stage, using a six-dimensional region limited in both ordinary space and momentum-space. It may seem that this procedure must either lead to the same result as the other, or must be much more difficult, or both. Neither is the case.

Instead of writing down a number of new equations which would look precisely like the old ones, I invite the student to go back to page 368 and recommence the argument at the words "We go into the momentum-space. . . ." If he will replace "momentum-space" by " μ -space," he need make no other change as far along as equation (21); the argument is just the same. Now let him turn ahead to page 379, and equation (52): this is valid for the μ -space as it was for the momentum-space, and so are equations (53). The novelty, however, is latent in the first of equations (53), which reappears as (55), and which I now rewrite for one more time:

$$-\ln A = \ln \sum e^{-BR_i} \quad (64)$$

On page 380, the summation was shown to be equal to $(1/C_j q_m)$ times a certain integral denoted by I ; the integral was over the three dimensions of momentum-space; q_m was the size of the elementary cell in momentum-

space. In μ -space, however, the corresponding integral is over the six dimensions, and may be written thus:

$$\iiint dx dy dz \iiint e^{-\beta E} dp_x dp_y dp_z$$

This sixfold integral is nothing but the product of V the volume of the container (resulting from the first three integrations) by the integral heretofore denoted as I (resulting from the last three integrations). It is to be multiplied by $(1/C_j h^3)$, C_j now standing for the number of cells in the six-dimensional region and h^3 for the volume of the six-dimensional cell (I explain the curious symbol later). The product is the reciprocal of A , and therefore:

$$k \ln W = kN \ln V + \ln I + kB U - kN \ln N + kN - kN \ln h^3 \quad (65)$$

The term $(-kN \ln N)$ appears just once, and not twice as it did in the sum S_e and S_m : all is well in this regard. The presentation of I , of B and of U as functions of T follows just the same lines as above.

(Notice, for future reference, that we should have attained to the same result had we ignored the ordinary space, operated in the momentum-space exclusively, and assigned the value h^3/V to the volume of the elementary cell in momentum-space.)

So, identifying $k \ln W$ with S , we come to the consummation of the new statistical theory of entropy, the equation:

$$S = kN \ln V + (3/2)kN \ln T - kN \ln N + kN \ln [(2 \pi m k)^{3/2} e^{5/2} / h^3] \quad (66)$$

The dependence on volume is right; it was qualitatively so to start with, was made exactly so by choice of the value of k as R/N_0 . The dependence on temperature is exactly right, since it is a fact of experience that for monatomic ideal gases the specific heat at constant volume is $(3/2)R$ per gramme-molecule. The dependence on number of atoms is exactly right, that is to say, it makes S proportional to N for given P and T . The additive constant is fixed in value absolutely, or will be when we assign a numerical value to h^3 ; for k is a universal constant, e the base of natural logarithms, and m the mass of an atom of the gas.

For the benefit of such as may still be interested in comparing the old statistics with the new, I recall that the old statistics in its theory of entropy furnished the first and the second terms of (66), and apparently furnished also the fourth term though with $e^{3/2}$ in place of $e^{5/2}$. The third term it omitted, thereby lending itself to the untenable doctrine that entropy should be proportional to N for given V and T (and not for given P and T). Since there was no term $kN \ln N$, I committed no error when in the previous article I deduced S_e and S_m separately and then added them together to get S .

This procedure is right in the old statistics, becomes wrong in the new. I suppose that this is what some expositors mean when they say that in the new statistics there is a correlation between positions and momenta, or words to that effect. I say that the old statistics *apparently* furnished a term equal to the fourth of (66) except for the power to which e is raised. Actually the old statistics gives an additive term $Nk \ln [(2\pi mke)^{3/2}/Q]$ and the new statistics gives an additive term $Nk \ln [(2\pi mk)^{3/2} e^{5/2}/Q]$, but Q in the former case is the volume of the *region* and in the latter case is the volume of the *cell*. Giving the same value h^3 to Q in the two cases is positively *not* doing the same thing. However by doing this notwithstanding, and by "tampering" with the old statistics in a certain way which I described at the end of the previous article, it is possible to produce an expression exactly like (66).

SIZE OF THE ELEMENTARY CELL

We have reached the final step, which consists in assigning a value to the size of the elementary cell in μ -space. For this I have used the symbol h^3 , implying (as everyone has guessed already) that it is taken to be the cube of Planck's constant h so promiscuously found in Nature. What arguments can be advanced to justify this choice?

It may be remarked very simply, that since the volume of the elementary cell has the "dimensions" of the cube of the product of length by momentum, and since these are also the dimensions of h^3 , and since both h^3 and that volume are very fundamental things, what could be more natural than to identify them the one with the other? This was the argument used when formula (66) was first derived from the old statistics with the aid of judicious tampering.

An argument more precise of aspect may be adduced from wave-mechanics. Imagine the box containing the gas to be a cube, its edges—these being of length L so that $L^3 = V$ —being along the coordinate-axes x, y, z . The doctrine of wave-mechanics avers that the momentum-components p_x, p_y, p_z of any atom are perforce integer multiples of $h/2L$; for this is the condition that the waves which are associated with the atom shall form a stationary wave-pattern with nodes at the walls of the cube, and upon this condition wave-mechanics is insistent.⁵ Now let us reenter the momentum-space, and place a dot at every point for which p_x, p_y and p_z are integer multiples of $h/2L$. The dots form a cubic lattice, and it would seem very

⁵ The wave-length of the waves associated with a particle moving parallel to the x -axis is h/p_x , and there must be an integer number of half-wave-lengths between the walls of the cube which are perpendicular to the axis of x and face one another at a distance L . The same may be said, *mutatis mutandis*, of a particle moving parallel to the axis of y or z , with momentum p_y or p_z ; while if an atom is moving obliquely so that two or all three of its momentum-components differ from zero, each of these components is to be treated as if it alone existed.

vice if I could say at once that the elementary cube of this lattice has the volume h^3/L^3 which is h^3/V . However, this cannot be said, for there is an obstinate factor which makes the elementary cube have the volume $h^3/8L^3$. People get around this by remarking that since an atom reverberating to and fro between the walls of the cube changes the sign of one of its momentum-components whenever it strikes against one of the walls, therefore every dot is one of a group of eight dots all of which correspond to the same motion of the atom, and all eight should be counted as though they were one.⁶ Therefore in the region of momentum-space enclosed between any two spheres centred at the origin (such as we used in determining the distribution-in-momentum) we are to count one-eighth of the dots. The number so obtained is the same as the number of cells of volume h^3/V contained in the region. Thus it comes to the same thing whether one says that the atoms are distributed among one-eighth of the dots or among cells of volume h^3/V . Now I recall my remark (page 383) that equation (66) down to the last detail can be derived by playing the game of balls and baskets by the rules of the new statistics *in the momentum-space alone*, provided that to the elementary cell in this space we assign the volume h^3/V . For doing this last, wave-mechanics has now offered a kind of retroactive basis. There seem to be flaws in the basis, but they are of a kind which cannot be mended (if at all) without a thorough study of a very hard subject, to wit, the art of interpreting wave-mechanics in the ordinary language of space and time.⁷ I think it will be better to proceed at once to the test by experiment.

TEST BY EXPERIMENT OF THE NEW STATISTICAL FORMULA FOR ENTROPY

Enough has been said already to cover the first three terms of the formula (66), which correctly give the dependence of entropy S upon volume V , temperature T , and number of atoms N . The present question is: what does experiment say of the fourth term, the additive constant which involves the mass m of the atom and the universal constants k and h ?

Having treated this question at length in the June 1942 issue of this Journal, I will here give only the barest outline. For this purpose I rewrite (66), by the aid of the equation of state of the perfect gas,

$$PV = NkT \quad (67)$$

⁶ If (a, b, c) are the coordinates of one dot, those of the other seven of its group are: $(a, -b, c)$; $(a, b, -c)$; $(a, -b, -c)$; $(-a, b, c)$; $(-a, -b, c)$; $(-a, b, -c)$; $(-a, -b, -c)$.

⁷ In previous pages I said that the proper way of playing the game of balls and baskets is to play it in the six-dimensional space, with N_i representing a definite number of atoms located in a six-dimensional region which is composed of a narrowly-limited region in ordinary space and another narrowly-limited region in momentum-space. Wave-mechanics, in the current interpretation, will not allow this; it claims that, if the N_i atoms are located in a limited region of momentum-space, they are spread all over the box containing the gas.

so as to give entropy S as function of *pressure* and temperature:

$$S = -kN \ln P + (5/2)kN \ln T + kN \ln \left[\frac{(2\pi m)^{3/2} (ke)^{5/2}}{h^3} \right] \quad (68)$$

Notice that here every term is strictly proportional to N , in accordance with the "second alternative" of page 377.

Let P and T be so chosen that the gas is in equilibrium with its solid crystalline phase. To keep this choice in mind, I will replace T by T_s , signifying "temperature of sublimation" at pressure P . Let the N atoms of gas now be cooled to the absolute zero. First they will condense, still at temperature T_s , into the crystalline solid. In so doing they will disgorge the "heat of sublimation," L per gramme-molecule, amounting to NL/N_0 ; and their entropy will decline by $NL/N_0 T_s$, since the process is reversible. Let the cooling continue. As the crystal declines in temperature from any T down to $(T - dT)$, it disgorges heat in the amount of $(N/N_0)C_p dT$ and entropy in the amount of $(N/N_0)(C_p/T)dT$; here C_p stands for the specific heat (per gramme-molecule) of the crystal. The pressure is supposed to remain the same throughout the entire process. When the crystal arrives at absolute zero, its entropy has the value:

$$S_0 = -kN \ln P + (5/2)kN \ln T_s + kN \ln \left[\frac{(2\pi m)^{3/2} (ke)^{5/2}}{h^3} \right] - (N/N_0)(L/T_s) - (N/N_0) \int_0^{T_s} (C_p/T) dT \quad (69)$$

The right-hand member of this equation embodies the new statistical theory of entropy. If on the left I put the value *zero* for S_0 , I express what is known as "Nernst's Heat Theorem" or the "Third Law of Thermodynamics." If experiments say that the right-hand member of (69) is equal to zero, they ratify not indeed the statistical theory by itself or the Third Law by itself, but the assumption that both are true. Now, this is what the experiments *do* say. Better to describe the situation, they say that the first three terms on the right of (69) are equal to the last two terms with sign reversed. All of the noble gases have been tested with suitable accuracy, and eight or nine of the metals with accuracy not so high, yet better than "order-of-magnitude accuracy." For further details I must refer to my article already cited.⁸

⁸ I cannot refrain from mentioning a detail of the statistical theories, which is amusing if one sees it at once and confusing if one sees it belatedly (mine was the latter experience). It pertains to the power to which e is raised in the third term on the right in (69). If in the new statistical theory we leave out the term N in (49), thus stopping with a first approximation instead of going on to the second, we arrive ultimately at $e^{3/2}$ instead of $e^{5/2}$. If in the old statistical theory as modified by Tetrode we use the first-order Stirling approximation instead of the second-order one for $N!$, we arrive ultimately at $e^{3/2}$ instead

THEORY OF RADIATION

Black or total radiation, which is the electromagnetic radiation within a cavity enclosed by walls at a uniform temperature, may be regarded as a monatomic gas of which the atoms are called "photons." It has two peculiarities. First, the relation between energy and momentum is not the same for a photon as for a material atom. If by p I represent the magnitude $\sqrt{p_x^2 + p_y^2 + p_z^2}$ of the momentum, then the energy E is given no longer by the familiar equation (3), but rather by this one:

$$E = pc \quad (70)$$

c standing, of course, for the speed of light. This is no insignificant change, but recedes into secondary importance when compared with the other contrast. Not only the distribution-law for the photons, but the actual total number of photons itself, is fixed by Nature when the temperature of the walls of the cavity is fixed by the observer. To the quantity called N , the number of atoms in a container of volume V , no specific value has ever yet been assigned in these pages; for with a material gas it may be raised or lowered at will, by pumping gas into or out of the box. In this section, however, it will have to have a value, for Nature has given it one.

Can the theory achieve what Nature demands of it? It can, and this is the way.

The momentum-space is divided as heretofore into regions of equal volume, each containing C cells of volume q_m . A distribution is described by giving the number of photons in each region, N_j standing for the number in the j th region. The probability W of a distribution is given as always by the formula (16) and this is it:

$$\ln W = \Sigma \ln W_j = \Sigma[(N_j + C) \ln (N_j + C) - C \ln C - N_j \ln N_j] \quad (71)$$

We are not now proceeding to the limit of extreme rarefaction! Radiation presents itself to us under conditions remote from this limit, and must be treated without recourse to the approximation hitherto used in these pages.

When the quantities N_j are altered by the small amounts or "variations" δN_j , W undergoes the slight alteration or variation given thus to first approximation:

$$\delta W = \Sigma \left(\frac{\partial W}{\partial N_j} \right) \delta N_j = \Sigma [\ln (N_j + C) - \ln N_j] \delta N_j \quad (72)$$

of $e^{5/2}$ (see the text preceding equation (35) of the prior article on page 134 of the January issue of this Journal). Thus in both cases we arrive at $e^{3/2}$ or $e^{5/2}$, according as we pause at a first approximation or go on to a second; *but* I discern no mathematical or physical similarity whatever in the two situations in which these approximations are made.

In the quest for the "most probable distribution" this quantity is required to vanish for variations which are controlled by a certain condition.

On a previous page (368) where we were dealing with *material* atoms in *ordinary* space, the sole condition was that the total number of atoms should remain the same (and equal to N). This led to the uniform distribution, $N_j = \alpha$; here α stands for a constant, which turns out to be the product of N by the ratio of the volume V_0 of the region to the volume V of the box.

On another previous page (369), where we were dealing with *material* atoms in *momentum* space, the condition imposed was twofold: that the number of atoms N and the total energy U of the atoms should remain the same. This led to the distribution (20), which in the limit of extreme rarefaction became the Maxwell-Boltzmann or canonical distribution $N_j = NA \exp(-BE_j)$; here E_j stands for the energy-value appropriate to the j th region, and A and B for two constants which were shown to be determined by N and U .

In this case where we are dealing with *photons* in *momentum* space, the condition which leads to the right result is simple but surprising. We must admit only such variations as leave the total energy constant, but we must *not* require that the total number of photons should likewise remain the same. Applying this strange condition, we find it taking the form,

$$\ln(N_j + C) - \ln N_j = BE_j \quad (73)$$

with only one constant, which is going to be controlled by the total energy U . Rewriting this:

$$\frac{N_j}{C} = \frac{1}{e^{BE_j} - 1} \quad (74)$$

One sees immediately that N , which is the sum of all the quantities N_j , is no longer at liberty to take whatever value the experimenter pleases!

Hitherto I have assumed that all the regions are of equal volume, but I can free myself from this assumption by pointing out that N_j/C is the average number of photons per cell in the portion of momentum-space where E has the value E_j . Now let us carve up the momentum-space into regions separated by spherical shells all centred at the origin. The region extending from the sphere of radius p to the sphere of radius $p + dp$ will be of volume $4\pi p^2 dp$, and will accordingly contain $4\pi p^2 dp/q_m$ cells, if by q_m I denote the volume of a cell. The appropriate value of E will be pc . The number of photons in the region will accordingly be given thus:

$$dN = \frac{4\pi p^2}{q_m} \frac{1}{e^{Bpc} - 1} dp \quad (75)$$

Unrecognizable as it may seem, this is actually a statement about the spectrum of black radiation! This is because a photon of momentum p and energy pc is associated with light-waves of wave-length given by the "Rule of Correlation":

$$\lambda = \frac{h}{p} \quad (76)$$

If I therefore multiply both members of (75) by pc , I have an expression for the amount of energy associated with the waves ranging in wave-length from h/p to $h/(p + dp)$.

There are instruments able to sort out the waves of different wave-lengths with their associated photons; they are called spectroscopes. There are instruments able to indicate the total energy borne by the photons thus sorted out; they are called by such names as bolometer and thermopile. There are people able to use these instruments; and so (75) can be tested. It is customary to rewrite (75) so that either wave-length or frequency becomes the independent variable, in place of p ; but nothing would be gained for the purpose of this article by doing so. The fact of experience is, that (75) is a correct description of black radiation provided that three modifications be made:

a) For q_m we are to write h^3/V , presuming that this comes to the same as though we had operated in six-dimensional space and put h^3 as the volume of the elementary cell therein (page 383);

b) For B we are to put $1/kT$;

c) We must double the right-hand member of (75), the factor 2 being ascribed to the fact that light is polarizable.

Making these modifications, and putting $V = 1$ so that the forthcoming equation shall refer to the radiant energy contained in *unit* volume, we have

$$dN = \frac{8\pi p^2}{h^3} \frac{1}{e^{pc/kT} - 1} dp \quad (77)$$

for the number of photons in unit volume endowed with momenta between p and $p + dp$, energies between cp and $c(p + dp)$. This is the distribution-formula for black radiation of temperature T , commonly known as "Planck's law."

To have derived this law is the first, the great and the historic achievement of the new statistics. Other ways have indeed been found for deriving it, beginning with Planck's own; but the way of the new statistics is smoothest and quickest. Quite different is this story from that of the theory of material gases! There, the distribution law was correctly given by the old statistics long before it was tested. Here, the distribution-law was found

by experiment years before it was explained, and a great puzzle it was. There, the old statistics and the new (in the limit of extreme rarefaction) led to the same result. Here the old statistics was impotent, and the new had to be invented.

Reverting to the identification of B with $1/kT$: this may be proved in the following way⁹. Refer back to equations (74) and (71), and for ease of operation write x_j for $(e^{B E_j} - 1)^{-1}$. We then have:

$$N_j = C x_j \quad (78)$$

$$S = k \ln W = k \sum \{C(1 + x_j) \ln [C(1 + x_j)] - C x_j \ln [C x_j] - C \ln C\} \quad (79)$$

$$U = \sum N_j E_j = C \sum E_j x_j \quad (80)$$

Differentiate S with respect to B and do the like with U , and divide the former derivative by the latter, so as to get the derivative dS/dU . It will be found that this is equal to kB ; and since by definition of absolute temperature it is also equal to T^{-1} , the identification is made.

THE BOSE-EINSTEIN AND THE FERMI-DIRAC STATISTICS

Hitherto in this article, except for one protective allusion, I have spoken as if the new statistics were one and indivisible. There are, however, two branches of it, known respectively as the Bose-Einstein statistics and the Fermi-Dirac statistics. It is the former of which I have treated throughout this essay. The point at which the latter branches off is to be found on page 365, where I introduced the game of balls and baskets, the balls standing for cells and the baskets for populations. On reaching this point the game is to be played with the supplemental assumption that there are only two baskets, those numbered 0 and 1. That is to say: a cell may either be empty or may contain a single atom, but never more than one.

I leave to the student the task of revising equations (7) to (16) accordingly, but I take it upon myself to point out how easily the problem can be solved by the second method—that of pages 370–71, the method involving the counting of *all* the different ways in which un-numbered atoms can be distributed among numbered cells. In the Bose-Einstein case the fundamental formula is (22), which is not very easy to derive. In the Fermi-Dirac case we proceed by playing anew the game of balls and baskets. There are but the two baskets, one being set out to receive the balls corresponding to the empty cells and the other for the cells containing one atom each—the “filled cells,” we may call them. There being in the j th region N_j atoms and C cells all together, the first basket is destined to contain $(C - N_j)$ balls and the second to contain N_j . The question is then: in how many

⁹I am indebted for this proof, as well as for much other assistance in the preparation of the article, to Dr. L. A. MacColl.

ways can C numbered balls be distributed among two baskets, these to contain $(C - N_j)$ and N_j of the balls respectively, two ways being considered as different unless the inventory of each basket is just the same for both ways? But this is the problem set up and solved in the earlier article, though there the balls stood for atoms and the baskets for regions. The answer is:

$$W_{\text{tot}} = C! / (C - N_j)! N_j! \quad (85)$$

instead of equation (22). Using the first or second order Stirling approximation—it doesn't matter which—one comes to the analogue of (23), which is:

$$\ln W_{\text{tot}} = C \ln C - (C - N_j) \ln (C - N_j) - N_j \ln N_j \quad (86)$$

Different as this looks from (23), the two become alike in the limit of extreme rarefaction, and in this limit equation (48) expresses the result both of the Bose-Einstein and of the Fermi-Dirac statistics. Since equation (48) is the parent of the Maxwell-Boltzmann distribution-law and of the expression (66) for the entropy of a monatomic gas, both of these flow from either type of statistics, and experiment does not decide for either over the other.

When we avoid the limit of extreme rarefaction, the two forms of statistics *do* depart from one another. If photons obeyed the Fermi-Dirac statistics, the distribution-law for black radiation would not be (75). We should be obliged, in the denominator on the right-hand side of that equation, to replace the negative sign of the second term by the positive sign. In so doing we should contradict the data of experiment in an unmistakable way; and for photons accordingly, the Fermi-Dirac statistics is to be rejected.

This form of the new statistics being no better than the other for material gases, and definitely wrong for radiation, where is it to be preferred and why?

To answer the first question, I point to the "electron-gas" which pervades the metals and is accountable for their quality of being excellent conductors. Experiment (as I recounted in these pages fourteen years ago¹⁰) confirms that these intra-metallic electrons form a gas which obeys the Fermi-Dirac statistics. It is not, however, the limit of extreme rarefaction which here we meet but the opposite one, the limit of extreme condensation. These electrons are as densely concentrated as the atoms of the solid itself, a degree of condensation never even approached by any ordinary gases. In this limit the distribution-law attains a form entirely different from both the Maxwell-Boltzmann law and the black-radiation law, and very remarkable. I dare not, however, expose this article to the risk of a doubling in length, which a treatment of this topic would probably entail; and I can avoid it with a fairly clear conscience, for the experimental evidence that electrons

¹⁰ This Journal, 8, 672 (1929); also *Physical Review Supplement* (Reviews of Modern Physics) 1, 90 (1929).

obey the Fermi-Dirac statistics has been enlarged but little since my article of 1929.

As for the second question, I can give only the shadow of an answer. The reason for adopting sometimes the Bose-Einstein and sometimes the Fermi-Dirac statistics springs from wave-mechanics, and that requires an article of its own. I can say, without proof, that the choice depends upon the number of elementary particles in the atom. The gas is supposed to conform to the Bose-Einstein or the Fermi-Dirac statistics, according as that number is even or odd. An electron is an elementary particle all by itself, wherefore the preceding paragraph.

For material gases, the crucial number is obtained by adding up the numbers of the protons and the neutrons in the nucleus, and the number of orbital electrons which surround the nucleus and complete the atom. In nature the atoms for which the crucial number is even vastly outnumber those for which it is odd, and the Bose-Einstein statistics is therefore the prevalent one. The principal isotope of nitrogen and the second isotope of hydrogen do indeed belong to the rarer category, but in the gaseous state their atoms always pair themselves into diatomic molecules, a circumstance which restores these gases to the realm of Bose and Einstein. A detail in the band spectrum of a diatomic molecule is available for telling which form of statistics the individual atom would obey if free; it confirms what I have just been saying—but this is an intricate story.