

## Memorial to the Classical Statistics

By KARL K. DARROW

ONE of the most elusive and perplexing, hazy and confusing of the parts of theoretical physics is that which bears the name of "statistical mechanics".\* On the principle that a tree is to be judged by its fruits, this must be ranked as high as the tree which bore the golden apples of the Hesperides; for among its fruits are the Maxwell-Boltzmann distribution-law, the black-body radiation law, the value of the chemical constant, the Fermi distribution-law for the electrons in metals, the alternating intensities in band-spectra—and indeed the tree might lay a valid claim to the whole of quantum-theory. The singular thing is that such wonderful fruits should have grown from, or should have been grafted upon, so badly-rooted a tree. To change the metaphor, one frequently feels that the superstructure is sustaining the foundations, and the premises are flowing from the consequences, rather than the other way about. Perhaps anyone who feels this way should be disqualified from writing about the subject; but on the present occasion, the attempt is going to be made.

Statistical mechanics—hereinafter to be called "S.M." at times for short—did not of course arise from any desire to solve the problems suggested above, which came late. It seems to have sprung from attempts to answer older questions, of which the following may serve as an example. Consider a gas in a box, with an electric fan or something of the sort fitted inside to stir it up. The gas having been stirred up, the fan is stopped, leaving it in a state of surging and whirling about within the confines of the box. Very shortly, however, the surging and the whirling cease, the gas having passed of itself into a state of tranquillity and uniformity—uniform density, uniform pressure, uniform temperature. From this state it never departs, unless stirred up afresh. There is a tendency of the gas to go of itself from the state of surging into the state of uniformity, and no tendency at all for it to go from the state of uniformity back into the state of surging. This is very unlike the behavior of a pendulum, which having fallen from one end of the arc of its sweep to the middle thereof, moves on to the opposite end, retraces its path and returns to its first situation. Why should the gas behave that other way?

\* I acknowledge with gratitude the incentive given me by Smith College to explore this subject, by offering me the opportunity of giving a course on statistical and chemical physics in the spring semester of 1942.

For an answer to this question and others of the kind, S.M. offers the following statement:

BASIC THEOREM OF STATISTICAL MECHANICS

*A system is more likely to be found in a state of greater probability than in a state of lesser probability*

It may be that no reader of these lines has ever seen the basic theorem of S.M. set forth with such merciless candor, though in many a sober treatise there is an elaborate statement which when analyzed turns out to be just this and nothing more. Of course it is a tautological statement, and has no value except insofar as it may help to drive some contradictory notion out of the student's mind or to prepare that mind for some meaning or other which is not yet in the statement but may be added to it later. Actually it can serve both these offices.

To be expelled from the mind of the student is first of all the idea that S.M. is going to give him a description of the way in which the gas proceeds from the surging state to the uniform. From an astronomer he may learn the orbit of the moon from apogee to perigee. From an authority on ballistics he may find the trajectory of the bullet from the muzzle of the gun to the bull's eye on the target. From a railroad office he may get a timetable showing the passage of the train from mile to mile over the rails from Boston to Chicago. All this sort of thing is out of the range of statistical mechanics! If a railroad acted like a surging gas and its timetable were devised in the spirit of S.M., one would go to the office and be told that the trains were enormously more likely to be in Boston than in Chicago or anywhere in between. From this one would be expected to infer that at any moment chosen at random the chance of finding a train anywhere along the line except in Boston would be practically nil—unless indeed one got a train and put it on the rails at Springfield, and even this would be of little use for getting to Chicago, since at every subsequent instant the train would almost certainly be in Boston. Not a very useful timetable, and not a very useful railroad!

S.M. thus starts off with a renunciation. It renounces the prospect of telling just how the gas proceeds from the surging state to the uniform state. To that smooth unbroken sequence of times and places whereby the moon finds its way through the heavens and the bullet through the air and the train along the rails, there is no counterpart presented.

This of course is a serious matter, for the smooth unbroken sequence is inseparably linked—or almost inseparably linked—with the notion of cause-and-effect, the notion of natural law, the notion of man as a being who can foretell the future. Mechanics harmonizes with these notions; for mech-

anics is the science which professes that, given the positions and momenta and the forces in a system of particles at 10 A.M. sharp, it can predict the positions and momenta of all the particles at 11 A.M. and every instant in between and all through the endless future. Statistical mechanics, for all the implications of its name, is nowhere nearly so audacious.

Suppose the electric fan, or whatever stirring-gadget was employed, was stopped an instant before 10 A.M. sharp. S.M. limits itself to affirming that at 11 A.M. the most probable state for the gas in the container—the immensely most probable state, the almost-certain state—is the uniform state. It also says the same thing exactly for 10:15 A.M., and for 10:01 A.M., and even for 10 A.M. sharp. If at 10 A.M. sharp the gas is in a state of wild and furious surging, S.M. does not deny the fact, but sees no reason for revising its own affirmation. If at 10:01 A.M. the gas is settling down but has not yet quite reached the uniform state, that again does not deter S.M. from standing by its assertion. Whenever a freak of chance or act of man may have produced one of the states which it calls improbable, S.M. just says “wait, and you shall have the state which *I* am going to talk about.” To further questions it can only say “I know my limits”—and that is what its basic theorem says for it.

If now the negative aspect of the basic theorem is sufficiently clear, we may address ourselves to the task of giving the theorem a positive meaning. For this there is but one way: the word “probability” must be replaced with some word or phrase or mathematical expression which does have a meaning. After this is done we can of course restore the word “probability” as an equivalent for that other word or expression. The basic theorem will then be tautological upon the surface only, for actually it will have the meaning conferred upon it by the definition of its key-word.

Various meanings have been offered for the key-word, by various people who have been successful in getting useful results out of statistical mechanics. Until 1924 the dominant meaning was that imposed by Gibbs and Boltzmann. From this meaning arises the form of S.M. which is called “the classical statistics”. (The word “statistics”, by the way, is a bad but common abbreviation for “statistical mechanics”.) This is the topic of the present article. In 1924 there was proposed a novel meaning for the key-word, which led to results sometimes agreeing with, sometimes differing from, those attained by the classical statistics. Where the results of the two agreed, they agreed with experiment also; where the results of the two disagreed, experiment sustained the new one. This event has left the classical statistics in a strange situation, in which one cannot exclude the possibility that all of its remarkable achievement is due to a happy but deceptive chance. The classical statistics may indeed be only a past episode in the history of scientific thought, and it is for this reason that I have given to the article the

strange and sombre title "Memorial to the Classical Statistics". Yet even as a past episode, it is worthy of remembrance; its didactic value may yet be great; and perhaps the human mind may some day stretch its powers to the point of conceiving the classical and the new statistics as aspects of a single whole, as it has lately stretched itself to the extent of uniting the wave-picture and the corpuscular picture of matter and of light.

#### THE MAXWELL STATISTICS

Since the main concern of S.M. is with the "most probable state", one sees that its principal content must be made up of assertions about that most probable state. Maxwell made such an assertion. He wrote down a formula for the distribution-in-velocity of the molecules of a gas. It is the formula now called "the Maxwell-Boltzmann distribution-law", which is so well known to the readers of this journal that I will not bother to write it down until there is actual need for having it on the page. Maxwell might have said bluntly: "This is the distribution which I will assume for the most probable state"; and having said so, left it at that. He did not leave it at that, and presumably he would have been dissatisfied so to leave it, as most of us would be. Instead, he postulated a pair of attributes for the most probable state, and showed that if these are the attributes, then the distribution is according to that formula.

The attributes which Maxwell postulated are "isotropy" and "independence".

The former is easy enough. One assumes that in the most probable state, the distribution of velocities of the molecules is isotropic. Nothing can usefully be added to this simple statement.

The latter is a little harder to grasp. Perhaps it can best be exhibited by describing a couple of imagined cases for which it would not be valid. Suppose for instance that all of the molecules have the same speed—the same magnitude of velocity, though their velocity-vectors be pointed in all directions. Let this common value of speed be denoted by  $V$ , and let any direction chosen at random be made the axis of  $x$  in an ordinary coordinate-frame. If a molecule happened to be travelling with such a velocity that the component thereof along the  $x$ -axis,  $v_x$  let us call it, was just equal to  $V$ , then it would be a certainty that  $v_y$  and  $v_z$ , the  $y$  and  $z$  components of the velocity, were both of them zero. If a molecule happened to be travelling in such a way that  $v_x$  was zero, then either  $v_y$  or  $v_z$  or both of them would have to be different from zero, and the square root of the sum of the squares of  $v_y$  and  $v_z$  would have to be equal to  $V$ . There would consequently be a *correlation* between the values of the three components, and the probable—may even the possible—values of any one of them would be affected by those of the other two. If the molecules had a uniform distribution of speeds up

to a maximum value of  $V$ , there would still be a correlation of a similar sort, though not so marked a one: the higher the velocity-component in the  $x$ -direction, the lower would the  $y$  and the  $z$  components be likely to be. One could imagine distributions for which the higher the velocity-component in the  $x$ -direction, the higher would the  $y$  and the  $z$  components be likely to be.

The "assumption of independence" is, that in the most probable state there is no correlation at all. Whether the  $x$ -component of the velocity of a molecule is high or low is a detail which has no influence whatever on the possible or the probable values of the  $y$  and the  $z$  components. Low values of  $v_y$  go just as well and just as abundantly with low values of  $v_x$  as with high, and reversely.

The Maxwell-Boltzmann law, as I said, is the distribution-law which conforms to both the assumption of isotropy and the assumption of independence. So the question arises: do those two assumptions have the quality of plausibility and of convincingness, which make the average person say "Surely these must be the attributes of the most probable state of a gas!" I do not know what result a referendum on this question would give, but it is my guess that most physicists would feel more satisfied with these than they would with the Maxwell-Boltzmann distribution-law if it were tossed out to them with the bare affirmation "This is assumed to be the attribute of the most probable state". Clearly this is how Maxwell felt, and there is no better guide than the intuition of a Maxwell.

The foregoing question is something else than the question whether the assumptions, and the Maxwell-Boltzmann distribution-law which follows from them, are truly the attributes of the most probable state. It is a strange historical fact that not for many years after the promulgation of this famous law, and not till after both of its sponsors were dead, was there any proper test of it. The derivations of the law were exercises in abstract and unenumerated thought. Nevertheless experiment—applied to thermionic electrons, to molecules of ordinary gases, to thermal neutrons—came at long last to justify Maxwell. To any who may feel that the assumption of independence is in itself too reasonable to require any proof, I disclose that in other forms of statistics this assumption is declared to be false, except as an approximation.

The "Maxwell statistics" therefore consists in the main of the statement:

*The most probable state of a gas is that in which isotropy and independence prevail among the velocity-vectors of the molecules.*

We now require some terminology and some notation.

I take for granted an understanding of the terms "velocity-vector" and "distribution-in-velocity", these being learned by physicists out of kinetic theory if not out of S.M. A velocity-vector may be replaced by a point

which serves just as well for all of its purposes and even better for some. Let the velocity-components  $v_x$ ,  $v_y$ , and  $v_z$  be laid out along the axes of a Cartesian coordinate-frame, and the vector for any molecule be drawn from the origin: the point at its tip is the point in question. Point and coordinate-frame are said to be "in velocity-space". Statistical mechanics prefers as a rule to deal with the momenta of the molecules rather than their velocities. This is for valid and powerful reasons, one of which is that the transition to the case of photons becomes much easier.<sup>1</sup> In the case of material gases it makes no practical difference, since the momentum of a molecule is its velocity-vector multiplied by the mass of the molecule which is practically a constant, and every statement about the distribution-in-velocity can with the utmost ease be translated into a statement about the distribution-in-momentum and *vice versa*. The momentum-vector may be replaced by the point at its tip, having coordinates  $p_x$ ,  $p_y$  and  $p_z$  in a coordinate-frame in "momentum-space". If we consider together with these the three coordinates  $x$ ,  $y$ ,  $z$  of the molecule in ordinary space, we may say that we are locating the molecule in six-dimensional space. I have yet to meet someone who claims that he can visualize a six-dimensional space, and yet there is no doubt that the phrase fulfills a psychological need and has a practical value. The six-dimensional space of these particular six variables is called "the  $\mu$ -space".

It seems odd to bring in the  $\mu$ -space before considering by itself the three-dimensional "ordinary" or "coordinate-space" in which the gas is located. Is there nothing to be said about the most probable distribution of the molecules in the coordinate-space? Well, "every schoolboy knows" that the state to which a gas tends and in which it remains is a state of uniform density. Maxwell, I think, accepted this as one of the facts behind which one cannot, or does not, go. For a complete statement of the Maxwell statistics I therefore offer the following:

*A gas is very much more likely to be in its "most probable state" than in any other. The most probable state is that in which isotropy and independence prevail among the momentum-vectors, while the distribution in coordinate-space is uniform.*

So in the Maxwell statistics the distribution-in-momentum of the molecules is *derived* from assumptions ostensibly more basic, while the distribution-in-ordinary-space is simply *affirmed*. If a theory could be devised in which both were derived from assumptions apparently more basic, one would be likely to feel that something had been gained. Now this is a characteristic, and one of the principal virtues, of Boltzmann's theory known as the "Boltzmann Statistics" or as the "Classical Statistics".

<sup>1</sup> Another reason has to do with "Liouville's theorem," for which unfortunately I cannot make room without overloading this article.

## THE BOLTZMANN STATISTICS

Boltzmann invented a way of appraising the probability of any imagined state of a gas, which has the following very remarkable features:

(a) It gives so sharp a definition to the key-word "probability", that not only can the state of maximum probability be identified, but the ratio of the probabilities of any two states can be computed.

(b) For the distribution-in-momentum of the molecules in the most probable state, it derives a formula identical with that which springs from the Maxwell statistics. This of course is why the formula is known as the Maxwell-Boltzmann law.

(c) For the distribution-in-space of the molecules in the most probable state, it derives the uniform distribution.

All this does not entail that the Boltzmann statistics is necessarily right. It does, however, lead to consequences, which it is the privilege and the affair of experimental physics to verify or to reject.

I can now write down a phrase into which the Boltzmann statistics, and equally well those which came later, can be fitted:

*The probability of a state is the number of different ways in which the state can be realized.*

This is another of those oracular sayings which acquire a meaning only after some meaning is given to the key-word, which in this case is *ways*. I could now rewrite the basic theorem without the word "probability", and so can the reader; but the only effect would be to transfer the mystery out of the word "probability" and into the word "way". Boltzmann, however, assigned a meaning to the latter word. It is this meaning which we now must strive to realize.

For this purpose I propose a game of which the outfit consists of a sack, an enormous number  $N$  of balls, and a smaller number  $M$  of baskets. The game is played by reaching into the sack, drawing out the balls one after another, and tossing them into the baskets. All of the balls feel precisely alike to the hand, so that there is never the least inclination to put one aside and pick up another as one's hand gropes around in the sack. Nevertheless when one looks at the balls after they have fallen into the baskets, one sees that they are nicely adorned with the integer numbers running from 1 to  $N$ . Incidentally the baskets also are numbered. It is this numbering which gives point to the game.

Someone or other—someone who might be designated as the caller, after the man who calls the figures of a square-dance—has prescribed a sequence of  $M$  numbers  $N_1$  and  $N_2$  and  $N_3$  and so on to  $N_M$ , all of them positive integers and totalling up to  $N$ . A single inning of the game consists in drawing all of the balls out of the sack one after another, and dropping the first  $N_1$  which come out into the basket I, the next  $N_2$  which emerge into the

basket II, and so on until every one of the balls is reposing in one or another basket. Now along comes the umpire with pencil and pen, and he writes down on one sheet of his pad of paper the numbers of all the balls which are in basket I, and on a second sheet the numbers of all which are in II, and so on until he has got an inventory of the contents of all of the baskets. *The inventory does not state the order in which the balls in any basket were dropped into that basket.* That order is blotted out and forgotten. *The inventory states which balls are in which baskets,* and lets it go at that.

This does not seem a very entertaining game, but entertainment is not what it is for. The present question is: how many different inventories can there be, consistent with that sequence of figures  $N_1, N_2, N_3, \dots, N_M$  which the caller prescribed at the start?

The answer is obtained in what must seem, to anyone meeting for the first time such a question, a strangely devious way.

First we evaluate the whole number of different *orders* in which the balls can be drawn from the sack. This is  $N$ -factorial or  $N!$ ; for the first ball to emerge may be any one of the  $N$ , and the next may then be any one of the  $(N-1)$  remaining, and the next may then be any one of the  $(N-2)$  remaining, and so on to the end.

If each order corresponded to a different inventory,  $N!$  would be our answer. Clearly this is so, if and only if there are as many baskets as balls and one ball in every basket. In all other cases  $N!$  is larger, and often colossally larger, than the number which we seek. It is necessary now to see that this great multitude of  $N!$  different orders falls into groups composed of  $X$  orders apiece, all of those in a single group corresponding to a single inventory—necessary to see this, and to calculate  $X$ ; whereupon we shall find that  $X$ , the “number of orders per inventory”, is the same for all of the inventories—so that the number which we seek is  $N!$  divided by this common value of  $X$ .

It seems to be helpful to think of some one inventory, and of some one order which leads to that inventory. By a certain amount of mental effort, which varies from person to person, it can be seen that this particular order is but one among  $N_1! N_2! N_3! \dots N_M!$  different orders all leading to the very same inventory. For think of the  $N_1$  numbered balls which lie in the first of the baskets: there are  $N_1!$  different orders in which they could have come out of the sack, and every one of these corresponds to the very same inventory. Think next of the  $N_2$  numbered balls which rest in the second basket: they might have come out of the sack in  $N_2!$  different orders, without changing the inventory. Think now of the contents of both of these baskets at once. Each of the  $N_2!$  orders in which the second basketful may come out of the sack may follow on any one of the  $N_1!$  orders in which it is possible for the first basketful to emerge. The product  $N_1! N_2!$  is therefore the total

number of ways in which the first  $(N_1 + N_2)$  of the balls might have come out of the sack without changing the inventory.

The process of proof need not be carried further.  $X$  has been evaluated. It bears no earmark of whatever particular inventory the student may have chosen to adopt at the beginning. It depends only upon the sequence of numbers  $N_1, N_2, \dots, N_M$  fixed by the caller, which sequence I will hereafter term a "distribution".

The number of inventories—or "complexions", to use a commoner word—for the distribution  $N_1, \dots, N_M$  is therefore given by the formula,

$$W = N! / N_1! N_2! \dots N_M! = N! / \prod N_i! \quad (1)$$

The theorem to which we are advancing affirms that this number has its maximum value for the uniform distribution—the distribution in which the caller assigns the same number of balls,  $N/M$ , to each of the baskets.

The usual argument for this statement may be put as follows: Let us assume the uniform distribution, with  $A = (N/M)$  balls in each basket, and compare its value of  $W$  with that of one of the neighboring distributions such as the one in which there are  $(A + 1)$  balls in the first of the baskets,  $(A - 1)$  in the second and  $A$  in each of the rest. It is not even necessary to get out a pencil and paper to see that  $W$  for the latter is less than  $W$  for the former, being in fact just  $A/(A + 1)$  times as great. The same is evidently true for disarrangements of the uniform distribution which involve more than two baskets and more than one ball per basket. The conclusion is clinched by the obvious fact that when all of the balls are in any one basket,  $W$  has its least possible value, *viz.* unity. (To unite this formally with the previous statements, one must follow the mathematicians' practice of using a symbol  $0!$  or "zero-factorial" and giving it the value unity).

We shall have to play this not so very entertaining game on several occasions in S.M., altering the meaning of the balls and the meaning of the baskets from one occasion to the next. The reader has probably guessed that the balls stand for the molecules. The guess is right in the classical statistics, wrong in the newer forms. To get at the meaning of the baskets, suppose the gas contained in a box of volume  $V$ , the interior of which is divided up by impalpable coordinate-planes into compartments or cells all of the same volume  $V_0$ . The baskets stand for the cells.

Now we have the theorem that  $W$  is greatest for the uniform distribution of the balls in the baskets, and the assertion that the most probable state of a gas is the state of uniform density, all ready to be fitted together. The process of fitting-together is of the simplest.  $W$  is christened the "probability" of the state described by the "distribution"  $N_1, N_2, \dots, N_M$ , the quantities  $N_i$  now standing for the numbers of molecules in the various cells. Not only is the state of uniformity the most probable one by this definition,

but so long as the number of molecules  $N$  is many times as great as the number of compartments  $M$ —a condition easy to realize—those distributions which are markedly far from uniform have probabilities which are fantastically smaller than the value of  $W$  for the uniform state.

The Boltzmann statistics manages thus to derive the assertion aforesaid—the assertion that the uniform distribution in ordinary space is of all the most probable—from a principle which (at least in appearance) is more fundamental. It has indeed a couple of bothersome points—more than a couple perhaps, but there are two in particular which the newer statistics will attempt to assuage. One of these is the size to be assigned to the cells  $V_0$ ; but we are borrowing trouble to think too much of that now, since whatever choice be made so long as  $N/M$  be large will not affect the achievement just cited. The other is, that one would much rather think of the molecules of a gas (of a single chemical kind) as being alike absolutely, than as being distinguished one from another by a mysterious something-or-other represented in this theory by numbers painted on balls. In the Boltzmann statistics, however, the numbers must stay on the balls.

We go over into the momentum-space, setting up a coordinate-frame and representing the molecules by dots, the coordinates of which are the momentum-components  $p_x, p_y, p_z$  of the molecules in question. To each position of a dot corresponds an energy-value, equal to  $(1/2m)(p_x^2 + p_y^2 + p_z^2)$ ; we will call it  $E$ .  $E$  vanishes at the origin, and has a constant value over any spherical surface centered at the origin. To any distribution of the dots will correspond a specific value for the total energy of the gas. For this we need a symbol different from  $E$ ; and as we shall have a good deal to do with thermodynamics later on, I choose the thermodynamical symbol  $U$ . The average energy of the molecules of the gas will then be  $U/N$ , to be denoted by  $\bar{U}$ .

The entry of  $E$  and  $U$  into the situation is of the first importance. It is in fact all that will save us from the highly unwanted conclusion that the most probable distribution in the momentum-space is the uniform one, just as it was in the coordinate-space. To see why it makes so great a difference is not altogether easy. I think that the reflections which follow may give an inkling of the reason.

The momentum-space must be taken either as infinite or as finite. If we take it as infinite and demand a distribution of uniform density, then the density goes to zero and at the same time the energies of the molecules go to infinity, producing an impossible situation. Let us then take it as finite, blocking off all of the parts of it which lie beyond a certain sphere centered at the origin. Assume a uniform distribution within the sphere. This will correspond to a certain value of  $U$ . (The student may suppose, if it makes him happier, that the  $U$ -value was preassigned and the radius of the "certain sphere" chosen accordingly.) The  $W$ -value of this distribution will surely

be greater than that for any non-uniform distribution, whether of the same  $U$  or of a different  $U$ , confined within the sphere. However, by blocking off the whole of the momentum-space beyond the sphere, we have barred a whole lot of distributions corresponding to the same  $U$  and having some of their dots beyond the sphere. By no means have we proved that the  $W$ -value for the uniform distribution within the sphere is greater than that for any and all of the barred distributions. Now if we can agree that the blocking-off of part of the momentum-space is a silly thing to do and unacceptable to Nature, the argument for the uniform distribution is spoiled, and we have to look for a new idea.

At this point it seems best to go through the mathematical process for finding the distribution of greatest  $W$  in the coordinate-space and the momentum-space, just as that process is presented in the textbooks.

We return to equation (1) and make it a manageable one by having recourse to that godsend of statistical mechanics, the "Stirling approximation", which may be written thus:

$$\ln N! = N \ln N - N + \ln \sqrt{2\pi N} \quad (2)$$

This is valid only for large values of  $N$ , though writers on S.M. never seem to remember how large the values must be. For still larger values of  $N$  we can drop off the last two terms, arriving at a sort of super-Stirling approximation which however itself is commonly called *the* Stirling approximation:

$$\ln N! = N \ln N \quad (3)$$

Putting (3) into (1), we find:

$$\ln W = N \ln N - \sum N_i \ln N_i \quad (4)$$

Defining some quantities  $w_i$  by the equations  $N_i = N w_i$ , we make this over into:

$$\ln W = - N \sum w_i \ln w_i \quad (5)$$

having availed ourselves of the obvious fact that  $\sum w_i$  is equal to unity.

We might now convert this into an equation for  $W$ , but this would be a waste of time and energy, since whenever  $W$  has a maximum so also will  $\ln W$ . With  $\ln W$ , therefore, we operate from now on. Making small variations in the quantities  $N_i$ , and making therefore small variations—call them  $\delta w_i$ —in the quantities  $w_i$ , we find in first approximation for the ensuing change in  $\ln W$ ,

$$\delta \ln W = - N \sum (1 + \ln w_i) \delta w_i \quad (6)$$

Now we are restricting ourselves to variations in the quantities  $N_i$  which leave unchanged the total number of molecules in the cells, or of balls in the baskets—to variations, therefore, for which

$$\Sigma N_i = N = \text{constant}, \quad \Sigma \delta w_i = 0 \quad (7)$$

*This restriction being introduced into (6),  $\delta \ln W$  proceeds to vanish if and only if  $w_i$  has the same value for all of the cells.* Now, the vanishment of  $\delta \ln W$  is a necessary condition for having a maximum of  $W$  at the situation in question. I do not refer to it as a sufficient condition, because it admits of a minimum or of what is technically known as a “stationary” value of  $W$  in the situation in question. However it has already been shown, without the aid of the Stirling approximation, that the expression to which we are approximating is greater for the uniform distribution than for the neighboring non-uniform ones. It may therefore be accepted that here we have a maximum of  $W$  for the uniform distribution, and have reached the old result in a new way; an achievement nearly useless, were it not a prelude to the performance in momentum-space.

I continue to use the symbols  $W$  and  $N$  and  $N_i$  and  $w_i$ , but now with reference to the distribution of the representative dots in momentum-space. A new symbol,  $E_i$ , shall signify the energy of a molecule in the  $i$ th cell of the momentum-space. We wish at all costs to avoid the conclusion that the stable distribution in the momentum-space is the uniform one. Boltzmann managed to avoid it, and his was the following way:

Let us write, for the number of molecules in the  $i$ th cell, the expression:

$$Nw_i = NA \exp(-BE_i) \quad (8)$$

and insert it into (6). We shall find:

$$\delta \ln W = -N \Sigma (1 + \ln A - BE_i) \delta w_i. \quad (9)$$

Of the three terms on the right, two vanish for all variations in which the total number of molecules remains the same. The third does not—but it *will* vanish for a restricted class of these variations, to wit, those and those only for which the total energy of all the molecules remains the same; for  $N \Sigma w_i E_i$  is precisely that total energy.

Some writers at this point ask the student to imagine a gas in a container being completely cut off from energy-interchange with the container-walls and with the whole of the outside world, and therefore being limited to the particular  $U$ -value with which it started out. Others import the word “temperature” which I am desperately (and vainly) trying to keep out until I am ready to bring it formally into the discourse, and aver that the gas is nearly or quite so limited if the walls of the container have the same temperature as the gas itself. The student may take his choice, but must suppose that

under such conditions Nature rejects that distribution which so to speak is "stable against" every conceivable variation, and elects that peculiar distribution which is stable not against any conceivable variation but only against the possible ones. Perhaps this is because the uniform distribution would entail the consequences mentioned on page 117, or perhaps there is no sense in saying that it is "because" of this or "because" of that. Anyhow, the peculiar distribution is the one which the data sustain.

However I have not really defined the peculiar distribution as yet, having merely thrown the symbols  $A$  and  $B$  into equation (8) as though they stood for completely disposable constants. It can readily be seen that at the most there can be but one disposable constant, for  $A$  and  $B$  interlinked by the obvious equation:

$$\Sigma w_i = A \Sigma \exp(-BE_i) = 1 \quad (10)$$

But even  $B$  is not disposable, if the total energy  $U$  and the average energy per molecule  $\bar{U}$  are preassigned; for there is another obvious equation:

$$\bar{U} = \Sigma E_i w_i = A \Sigma E_i \exp(-BE_i) \quad (11)$$

What with equations (10) and (11), there is no longer anything disposable about the constants  $A$  and  $B$ . The peculiar distribution in the momentum-space is completely defined. It is the Maxwell-Boltzmann distribution-law obtained from the Maxwell statistics, and sometimes known as the "canonical" distribution.

To summarize now the Boltzmann statistics as on page 113 the Maxwell statistics was summarized:

*A gas is more likely to be found in its most probable state than in any other. The probability of a state is found by imagining it as a distribution of numbered molecules among cells, in the coordinate-space and in the momentum-space. That of any distribution is measured by the number of inventories compatible therewith. By this criterion the most probable distribution in coordinate-space is the uniform one, and by this criterion carefully hedged about, the most probable distribution in momentum-space is the Maxwell-Boltzmann or canonical one. It is necessary to liken molecules of a single kind to numbered balls, differing in no way except the numbering.*

This point was reached by statistical mechanics about fifty years ago. Had it not been for Planck's wish and tenacious will to explain the black-body radiation-law, it might have been the stopping-point.

#### A HELPFUL AND TROUBLESOME COINCIDENCE BETWEEN TWO DIFFERENT QUANTITIES

Let us return to the game with the sack, the balls and the basket, played in the manner which led to good results when applied to the molecules in the coordinate-space.

The most probable distribution is the one evoked by the caller, when he calls for an equal number of balls in every basket. If there are  $N$  balls and  $M$  baskets, this means  $N/M$  balls to each basket, and a maximum number of inventories which I will call  $W_{\max}$ . Looking back to equation (1), we see that  $W_{\max}$  is a fraction the numerator of which is  $N$ -factorial, while the denominator is  $(N/M)$ -factorial raised to the power  $M$ . Taking logarithms and using the super-Stirling approximation, we find:

$$\ln W_{\max} = N \ln M \quad (12)$$

The logarithm of the probability of the most probable distribution (of numbered balls in numbered baskets, or molecules in equal cells of coordinate-space) is equal to the logarithm of the number of baskets (cells), multiplied by the number of balls (molecules).

Next suppose the caller, in a fit of uncontrollable zest for the game, calling in succession every one of the conceivable distributions. What is the total number of inventories compatible with all of them together? To sum over every conceivable expression of the type of (1) seems a hopeless assignment, but there is a short-cut to the result.

Fix a particular order for the drawing of the balls from the sacks—it may as well be the very order of their numbering. The first of the balls to be drawn may be tossed into any one of the baskets, giving  $M$  distinct “possibilities”. The second may be tossed into any one of the baskets, the same or another, giving in conjunction with the fate of the first  $M^2$  different possibilities. The third may be tossed—but we leap to the conclusion. There are  $M^N$  possibilities altogether, and *these are the inventories*. Thus the total number of inventories consistent with all of the distributions, which I will call  $W_{\text{tot}}$ , is a number whereof the logarithm is,

$$\ln W_{\text{tot}} = N \ln M \quad (13)$$

But this is the same as the expression for  $\ln W_{\max}$ !

The meaning of this strange coincidence can only be, that when  $N$  and  $N/M$  are both so great that the super-Stirling approximation is a good one, then the *logarithm* of the number of inventories belonging to the most probable distribution is nearly as great as the *logarithm* of the total number of inventories belonging to all of the distributions put together—so nearly as great, that either *logarithm* is a good approximation to the other.

In the foregoing very important paragraph, I have italicized the word “logarithm” because if it were left out the statement would become a false one. The statement is not true if applied to the numbers themselves.  $W_{\text{tot}}$  is manyfold greater than  $W_{\max}$ , and the ratio between the two actually increases with rising  $N$ . So does the difference between  $\ln W_{\text{tot}}$  and  $\ln W_{\max}$  increase with rising  $N$ , but not so fast as either by itself; wherefore the truth

of the statement. The student may convince himself of this by applying the second-degree Stirling approximation (equation 2)<sup>2</sup>.

I have called this both a helpful and a troublesome coincidence. It may be deemed a helpful one, because the expression for the total number of inventories is easier to derive and easier to remember than the expression for the number of inventories belonging to the most probable distribution. If therefore one has good ground for believing (as here is the case) that the logarithms of the two are approximately equal, one may serenely remember and use  $\ln W_{\text{tot}}$  instead of  $\ln W_{\text{max}}$ . The troublesome feature is, that some expositors speak of  $\ln W_{\text{tot}}$  throughout and never allude to  $\ln W_{\text{max}}$ , thus confusing the student to an extent which (if my experience is typical) may well be serious. I shall later dwell on the fact that  $\ln W$  for any distribution is regarded as a measure of the entropy of that distribution, and  $\ln W_{\text{max}}$  therefore as a measure of the entropy of the most probable distribution. Some people imply that  $\ln W_{\text{tot}}$  is the true measure of the entropy of the gas, instead of being an approximation to it. They commit no numerical error in so doing, but they blot out the most remarkable quality of the Boltzmann statistics, to wit, the clear distinction which it makes between the most probable distribution and those of lesser probability. This mistake is more commonly made in treating the newer statistics. Here I am not so sure that it is a mistake, but I think so.

#### MEANINGS OF THE WORD "STATE"

The word "state", which turns up continually in this essay, is one of those words of which a proper definition is hardly less than a full description of the theory which employs it. When the theory changes so also does the meaning of the word. In the welter of statistical theories, the word "state" has several different meanings. In thermodynamics also it has more than one meaning, but one is preeminent.

Thermodynamics usually concerns itself with gases (not to speak of liquids and solids) which are in what I earlier called a "uniform" state: uniform density, uniform pressure, uniform temperature. For a gas of a single kind ("kind" being a word which it is the business of chemistry to define) it is a fact of experience that any two of these three variables suffice to define the third and also all of the other variables which thermodynamics cares about. Of these others there are two in particular which I mention at this point, the

<sup>2</sup> Actually if one goes from the "most probable state"  $N_i = \text{const.} = N/M$  to the "next most probable" in which one ball is taken out of one of the baskets and put into another, the change in  $W$  is in the ratio of  $(N/M)$  to  $(N/M) + 1$ , which is practically no change at all when  $N/M$  is so high as is commonly taken. This shows that the statement could not be true if it were made about the numbers  $W_{\text{max}}$  and  $W_{\text{tot}}$  rather than about the logarithms thereof. It certainly looks as though the statement could not be true even when made of the logarithms, but this is evidently one of the cases where "intuition" is a fallible guide.

energy and the entropy. This makes five altogether, and any two of the five suffice to determine the state,—THE STATE, the uniform state, the only one about which thermodynamics really knows or cares. When asked about what I earlier called “a surging state”, thermodynamics mutters something to the effect that the entropy of such a state is smaller than that of THE STATE, and then puts an end to the conversation by refusing to commit itself further. Thermodynamics takes no cognizance of the molecular structure of matter. A gas might be a continuum, for all that it knows or cares.

Statistical mechanics talks about a mental image of the gas, in the form of a flock of dots in the coordinate-space and another flock of dots in the momentum-space, or one may call them a single flock of dots in the  $\mu$ -space. In Boltzmann statistics, the “state” of this image is what I have been calling the “distribution”. The most probable state of the image—to wit, the one with the greatest number of inventories or complexions—is identified with THE STATE of thermodynamics. All of the rest belong to the category of which thermodynamics would say, that the entropy is smaller than it is for THE STATE. But since according to S.M. they belong to a category for which the probability is smaller than it is for THE STATE, one sees a connection between entropy of the gas and probability of the image beginning to take shape.

Now it is time to make a formal introduction of the concepts of entropy and temperature—the latter word having already sneaked into this article two or three times in spite of all my efforts to keep it out.

#### FORMAL ENTRANCE OF ENTROPY AND TEMPERATURE

For a substance, meaning now a gas, of a single kind, entropy and temperature are defined by the equation,

$$dU = TdS - PdV \quad (14)$$

$P$  stands for pressure,  $V$  for volume, and  $S$  for entropy. For energy I use the symbol  $U$  already employed in that sense—but notice that formerly it stood for the kinetic energy of the molecules! To use the same symbol in both senses implies that the energy of the gas is entirely the kinetic energy (of translatory motion) of the molecules. This identification turns out to be valid for the “monatomic” gases, which are luckily numerous and well-studied. To these we confine ourselves throughout this article.  $T$  stands for the temperature called absolute; this being the only kind of temperature which will ever figure in this article, the adjective henceforth is discarded. Density was the fifth variable in my list given above, but volume is usually preferred to it. To make them equally useful, the quantity of gas must be stated; here it will be taken as one gramme-molecule.

It is evident that the equation is a comparison between two states. Do not go astray by supposing that these are like two of the states which we have been considering, having the same  $U$  and  $V$  and differing in the number of inventories! These on the contrary are two examples of THE STATE—of the thermodynamic state, of the most probable state—of a gas, differing in the values of some at least among the five variables. The quantities  $dU$  and  $dS$  and  $dV$  are the differences between the  $U$ -values and the  $S$ -values and the  $V$ -values of the two states, while  $P$  and  $T$  may be taken as referring to either, the smallness of the difference between the two states—implied by the differential notation—permitting of this.

It is also evident from my wording that the one equation is being used to define the two quantities  $S$  and  $T$ . This is unluckily no verbal slip, nor is it a temporary shortcut to be replaced by a royal road as the argument proceeds. The meanings of entropy and temperature are so coiled up together in thermodynamics, that it is impossible to take them apart un mutilated. One cannot seize either by storm and then invest the other, at least not without the aid of statistical theory: one has to surround them both in a single campaign. As Eddington has vigorously written, this is a common thing in physics. Electric force is defined as that which acts on electric charge, electric charge as that which is acted upon by electric force, and so on. . . . Common as it may be, it is probably nowhere else so harassing as in thermodynamics. There are three ways of intruding upon the vicious circle.

*First*, to apprehend both concepts in a single mental act. This is the counsel of perfection.

*Second*, to use a temporary definition of temperature, with the promise of confirming or correcting it later. The ideal-gas thermometer is the device used for this purpose in thermodynamics. Anyone trained in this way is likely to think for the rest of his life of temperature as the primary concept, entropy as a derived one—as indeed was the case, when thermodynamics started.

*Third*, to produce a theory which makes a pronouncement as to the nature of entropy.

This last is the major office of statistical mechanics. To those who accept it, entropy becomes the primary concept and temperature the derived one, and both are visualized by the aid of the key-word "probability" of the basic theorem, interpreted in some particular way.

#### OLD STATISTICAL THEORY OF ENTROPY

*In the classical statistics, the entropy of a distribution is considered to be the logarithm of the number of inventories or complexions compatible with that dis-*

tribution, multiplied by a constant (always denoted by  $k$ ) which is adjusted to bring about agreement with experiment:

$$\text{entropy } S = k \ln W. \quad (15)$$

To illustrate this doctrine and to evaluate  $k$ , I now take the student back to the coordinate-space, where a box of volume  $V$  populated with  $N$  molecules is divided mentally into  $M$  equal cells of volume  $V_0$ , and the most probable distribution is characterized by the value  $N \ln M$  for the logarithm of the probability. The entropy—or no, not the entire entropy of the gas, but merely what I will call “the contribution of the volume of the entropy” and denote by  $S_c$ —is then supposed to be  $kN \ln M$ , or:

$$S_c = kN \ln V - kN \ln V_0 \quad (16)$$

Reverting to the equation (14) in which the definitions of entropy and temperature were tangled up together, and rearranging it, we get:

$$TdS = dU + PdV \quad (17)$$

Now, an “ideal gas” is defined by two attributes. First, there exists between its pressure and its volume and its temperature the relation  $P = aT/V$ , wherein  $a$  stands for a constant. Second, its energy  $U$  depends upon the temperature only, and not upon any other variable, in particular not upon the volume. Therefore we may write:

$$TdS = C_v dT + (aT/V)dV \quad (18)$$

$C_v$  here standing for something of which we need only know that it is a function of  $T$  alone. Integrating, we find:

$$S = R \ln V + (\text{function of temperature}) + \text{constant} \quad (19)$$

and lo! it is seen that the dependence of entropy on volume is precisely of the sort which the theory is fitted to explain.

The next step is to adjust the value of the constant  $k$ . The constant  $a$  aforesaid is proportional to the amount of gas in the box, proportional therefore to  $N$ : it is the constant ratio of  $a$  to  $N$  to which  $k$  must be equated. For the amount of gas let us choose one gramme-molecule. Then  $a$  assumes the value always symbolized by  $R$  and called the “gas-constant”, and  $N$  assumes the value usually symbolized by  $N_0$  and called the “Avogadro number”. Both of these are known from experiment, and  $k$  is fixed by the equation

$$k = R/N_0 \quad (20)$$

The constant  $k$  is named in Boltzmann's honor, though in his time its value was not known because the value of  $N_0$  was only vaguely apprehended.

Now we have settled what I called “the contribution of volume to en-

ropy". It remains to interpret the rest of the right-hand member of (19), which I will call "the contribution of temperature to entropy". To do this we must re-enter the momentum-space.

From (15) and (5) and (8) we get, for the entropy  $S_m$  of the flock of dots in the momentum-space:

$$\begin{aligned} S_m &= -k \ln W = -kN \sum w_i \ln w_i \\ &= -kNA \sum (\ln A - BE_i) e^{-BE_i} \end{aligned} \quad (21)$$

Refreshing our memory from (10), we see that the first term of this expression reduces to  $-kN \ln A$ . Refreshing our memory from (11), we see that the second term reduces to  $+kNB\bar{U}$  or  $kBU$ . Referring now to one gramme-molecule of gas, I put  $R$  for  $Nk$ , and find:

$$S_m = -R \ln A + kBU \quad (22)$$

$S_m$  is hereby given as a function of  $U$ , but a more complicated function than appears on the surface, since  $A$  depends upon  $B$  (equation 10) and  $B$  upon  $V$  (equation 11). Yet when we differentiate  $S_m$  with respect to  $U$ , and in so doing take account of these complications, it turns out that we might as well have been oblivious of them! for the result is the same as though  $A$  and  $B$  were constants:

$$dS_m/dU = kB \quad (23)$$

Now the temperature, which has so often slipped into this argument in ways more or less surreptitious, is about to make its formal and ceremonious entry into the statistical picture. We turn back to equation (17), and deduce:

$$dS/dU = 1/T \quad (24)$$

The derivative here standing on the left is the derivative of entropy with respect to energy under the condition of constant volume: a thermodynamicist would write it  $(\partial S/\partial U)_v$ . It is therefore properly to be identified with the derivative in (23), and we make the two identical by putting:

$$B = 1/kT \quad (25)$$

Now taking the entropy  $S$  to be the sum of  $S_c$  and  $S_m$ , we find:

$$S = S_c + S_m = -R \ln A + U/T + R \ln V - R \ln V_0 \quad (26)$$

and this is to be compared with (19), the thermodynamic expression for entropy, which I repeat to make the comparison easier:

$$S = \int (C_v/T) dT + R \ln V + \text{constant} \quad (27)$$

Comparing these, we see first of all that  $R \ln V$  appears in both, as was already stated. It also seems at first glance that  $(-R \ln A + U/T)$  is to be identified with the integral in (27), and that  $-R \ln V_0$  is to be identified with the constant in (27). This however is not necessarily the case, for  $(-R \ln A + U/T)$  may prove to include constant terms. Indeed they do; and we must proceed to evaluate both  $A$  and  $U$  in terms of  $T$  in order to round off the task.

I recall equation (10) and write it thus:

$$1/A = \Sigma \exp(-E_i/kT) \quad (28)$$

This is a summation, to which each cell contributes one term having the value of  $E$  appropriate to that cell— $E_i$  for the  $i$ th cell. Of the volumes of these cells I have thus far said nothing, except that all are equal. I continue to say nothing further, but I give to their common volume the symbol  $H_0$ . Let us now form the integral:

$$\iiint \exp(-E/kT) dp_x dp_y dp_z, \quad E = (1/2m)(p_x^2 + p_y^2 + p_z^2) \quad (29)$$

the range of integration extending over the whole of momentum-space. This integral may be described as follows. Let the momentum-space be divided into cells of unit volume. Each of these cells of unit volume makes a contribution

$$\exp(-\bar{E}/kT)$$

to the integral,  $\bar{E}$  standing now for the average value of  $E$  in the cell in question. The integral is the sum of all of these contributions. Now let us inquire how much of a contribution is made by this same cell of unit volume to the summation (28). This second contribution is made up of  $1/H_0$  terms, one for each of the cells of volume  $H_0$  which occupy the cell of unit volume. The values  $E_i$  corresponding to these cells will not be exactly equal to the value  $\bar{E}$  corresponding to the entire cell of unit volume; but to the degree of approximation which is now being used, the difference may be neglected. The summation (28) is then equal to  $1/H_0$  times the integral (29). Now the value of the integral (29) is given in all tables of definite integrals, and in terms of our symbols it amounts to

$$(2\pi mkT)^{3/2}$$

so we come to the conclusion:

$$\begin{aligned} \ln A &= -\ln (2\pi mkT)^{3/2} + \ln H_0 \\ &= -\frac{3}{2} \ln T - \ln (2\pi mk)^{3/2} + \ln H_0 \end{aligned} \quad (30)$$

Now we have attended to every term in (26) except the term  $U/T$ . Nearly every reader will remember that the average kinetic energy of an atom of a monatomic gas at temperature  $T$  is  $\frac{3}{2}kT$ . I therefore leave out the derivation of this result, except for showing the student how to begin on it: the first step is to go back to equation (11) where an expression was given for  $\bar{U}$ , and in that expression to replace the summation  $\sum E_i \exp(-BE_i)$  by  $(1/H_0)$  times the integral  $\iiint E \exp(-BE) dp_x dp_y dp_z$ . It follows that  $U/T$  is  $(3/2)Nk$ , which for one gramme-molecule of gas is  $(3/2)R$ , which I write as  $R \ln e^{3/2}$ .

The picture of entropy for a monatomic gas limned by the Boltzmann statistics, is now completed. Entropy is the function which follows:

$$S = \frac{3}{2} R \ln T + R \ln V + R \ln \frac{(2\pi m k e)^{3/2}}{V_0 H_0} \quad (31)$$

The dependence on volume is correct, *i.e.*, just the same as in the thermodynamic formula. The dependence on temperature is correct, for  $(3/2)R$  is the value of the specific heat at constant volume per gramme-molecule of a gas, the quantity  $C_v$  of equation (18). The additive constant, as to the value of which thermodynamics says nothing, is fixed when the volumes  $V_0$  and  $H_0$  of the elementary cells in the ordinary space and the momentum-space are fixed.

#### MIXTURES OF GASES

Now we will go through the mental operation which is called "considering a mixture" of two different monatomic gases,  $N'$  atoms of the one and  $N''$  atoms of the other, in the same box and (necessarily) in the same momentum-space. Let me denote by  $U'$  and  $U''$ , respectively, the energies of these two gases; and by  $N'_i = N'w'_i$  and  $N''_i = N''w''_i$ , respectively, the numbers of atoms of the two kinds in the  $i$ th cell of momentum-space.

If we seek the most probable distribution of the first gas in the momentum-space, making the stipulation that we will admit only such variations of the quantities  $w'_i$  as leave  $N'$  and  $U'$  unchanged—well, of course, we get the same result as before, the distribution (8), with  $N'$  in place of  $N$  and (let me say)  $A'$  in place of  $A$  and  $B'$  in place of  $B$ .  $A'$  will depend upon  $B'$  and  $B'$  will depend upon  $U'/N'$ . If we do the like with the second gas, we get anew to the distribution (8) with  $N''$ ,  $A''$  and  $B''$  in place of  $N'$ ,  $A'$  and  $B'$ .  $A''$  will not be the same as  $A'$  nor will  $B''$  be the same as  $B'$ , unless it happens that  $U''/N''$  is equal to  $U'/N'$ . There is no cause for surprise in this. In acting this way we are only treating each gas by itself, and have as yet done nothing which can be regarded as "considering a mixture".

Let us however seek the most probable distribution of the two gases, mak-

ing the stipulation that we will admit only such variations of the quantities  $w_i'$  and  $w_i''$  as leave  $N'$  and  $N''$  and the sum of the energies  $U'$  and  $U''$ —not however the individual energies  $U'$  and  $U''$ —unchanged. In acting this way we are doing something which may be regarded as “considering a mixture”, since we are allowing for the possibility that energy may pass from the one gas to the other and the other to the one. Equally well are we considering the case of two gases separated by a partition through which energy may pass, but not the atoms. Since in such a case we really ought to take into account the atoms and the energy of the partition also, we must appease the critics by providing that the partition shall be very thin.

Choose any set of values of the quantities  $N_i'$ , which is to say, any particular distribution of the first gas; and choose any set of values of the quantities  $N_i''$ , which is to say, any particular distribution of the second gas. Go back to equation (1) and put primes on all the symbols  $N$ ,  $N_1$ ,  $N_2$ ,  $\dots$  on the right-hand side of that equation. The resulting expression gives the total number of inventories or complexions of the first gas. Take off the primes and affix double primes to each of these symbols. The resulting expression gives the total number of inventories or complexions of the second gas. Every complexion of either may coexist with any complexion of the other. Therefore the total number of complexions of the pair of gases is the product of the two expressions. It is this product which is  $W$  for the pair of gases, be they mixed or side-by-side.

With use of the Stirling approximation, the logarithm of  $W$  for the pair is the sum of two such expressions as we have seen in (5):

$$\ln W = -N'\Sigma w_i' \ln w_i' - N''\Sigma w_i'' \ln w_i'' \quad (32)$$

and its variation is:

$$\delta \ln W = -N'\Sigma(1 + \ln w_i')\delta w_i' - N''\Sigma(1 + \ln w_i'')\delta w_i'' \quad (33)$$

Let us now give a trial to the tentative distribution,

$$w_i' = A' \exp(-B'E_i), \quad w_i'' = A'' \exp(-B''E_i) \quad (34)$$

On substituting this into (33) we find that if  $B'$  is unequal to  $B''$ , the distribution has a stationary value of  $W$  with respect only to such variations as leave the energies of the two gases separately unchanged—the result which we had before. If however  $B'$  and  $B''$  are the same, then  $W$  is stationary with respect to variations which leave the sum of the energies unchanged, either being allowed to gain or lose so long as the other loses or gains by an equal amount. Since each  $B$  is controlled by the corresponding  $U/N$ , the distribution (33) has a stationary value of  $W$  for variations of the type in question if and only if the average energy of the atoms of each gas is the same. Since each  $B$  controls the corresponding  $A$ , this condition of

equal average energy makes the distributions of the two gases just the same.

We have already seen that  $kB$  is the reciprocal of the temperature: for it is the reciprocal of  $(\partial U/\partial S)_v$  in our statistical picture, and the definition of absolute temperature  $T$  is precisely that  $T$  is this derivative. The statement to which we have come is, that *the most probable state of the mixture is the one in which  $T$  is the same for both components*. It is often expressed in this way: classical statistics shows that for two (or more) gases in equilibrium with each other, the temperature must be the same. It is indeed a fact of experience, and a most important one, that when two systems (be they gases or be they not) are in thermal equilibrium, their temperatures *are* the same. This has not hitherto been mentioned, and yet we seem to have derived it. Quite a rabbit for the magician of the classical statistics to have pulled out of the hat!

However, skeptical people who see a rabbit pulled out of a hat are inclined suspect that either the rabbit was in the hat beforehand, or else there is no rabbit. Let us inquire into the contents of the hat and see whether we can find the rabbit there.

The first (and the last) question to be asked is: what is the difference between "different" kinds of gas in the statistical picture?

To the physicist or the chemist, different kinds of gas will be (for example) mercury and helium. These differ in their spectra, boiling-points, chemical properties, and quantities of other features. None of these features however appears in the theory, and therefore none of them can contribute to the result. The atoms also differ in mass, and for a moment this seems to be a difference of which the statistical picture takes account, since the letter  $m$  appears in some of our equations. However, it appears only in the ultimate equations, those such as (29) in which the distribution-in-momentum is expressed. It does not appear in the original form of the Maxwell-Boltzmann distribution-in-energy, the form shown in equation (8). It appears in particular in the last term of equation (31), but not elsewhere. Apart from this it may be said that *in the classical statistics, all gases are the same gas*.

This is a paradox, but only one of two. The other paradox is, that *in the classical statistics two parts of the same gas are different gases*. This second paradox arises from the numbering of the molecules, which is an essential feature of the classical statistics.

Therefore in the statistical picture a mixture of  $N'$  atoms of mercury and  $N''$  atoms of helium is distinguished by the fact that the mercury atoms bear one set of integer numbers (say those from 1 to  $N'$ ) and the helium atoms another set (say those from  $N' + 1$  to  $N' + N''$ ). But if the atoms were all helium atoms or all mercury atoms, they would also be divisible in many

different ways into a set of  $N'$  atoms bearing one set of numbers and a set of  $N''$  atoms bearing another set of numbers. Each set would obviously have to have the same distribution, with the same  $A$  and the same  $B$ , as any other set or as the totality of all the atoms. This conclusion, which is self-evident in the case in which all the atoms are called "mercury", remains true when some of the atoms are called "mercury" and others are called "helium". We have done nothing but change the names of some of the atoms; we have not imported into our theory anything which differentiates one kind of atom from another kind. No wonder we have arrived at the conclusion that all kinds have the same distribution-in-energy, the same  $A$ , the same  $B$  and the same temperature! The rabbit was indeed in the hat, but it does not look like so much of a rabbit.

The classical statistics therefore doesn't recognize any of the real differences between atoms of different kinds, except for alterations in the last term of (31); but it does make an artificial difference which creates the astonishing result, that any two samples of the same gas are different gases! At this point we may begin to wonder whether this peculiarity, which has led to so apparently brilliant a result in respect of the equality of temperatures in thermal equilibrium, might elsewhere lead us astray. It does; and here appears the rift in the lute of classical statistics.

#### THE RIFT IN THE LUTE

Let us imagine two boxes of equal size separated by a common partition, each containing a gas consisting of  $N$  atoms, both gases at the same temperature. We will baptize one gas "mercury" and the other gas "helium". Let an opening be made through the partition. It is known that in such a situation in Nature, the two gases diffuse into one another, the final and permanent condition being that in which the mercury and the helium are equally distributed between the two boxes. The process of diffusion is an example of what in thermodynamics is called an "irreversible" process. The state of uniform mixing ought to correspond to the most probable state in the statistical picture. But what does the statistical theory say?

The statistical theory says nothing about diffusion and nothing about mixing. The statistical theory takes account of nothing but the facts that the mercury had at its disposal the volume  $V$  before and the volume  $2V$  after the breaking of the partition, and ditto for the helium. The value  $V$  contains  $M$  cells ( $M = V/V_0$ ) and the volume  $2V$  contains  $2M$  cells. The (approximate) probabilities of the uniform distribution are  $M^N$  before and  $(2M)^N$  after. The latter is greater than the former; the entropy goes up by  $Nk \ln 2$  for each gas, by  $2 Nk \ln 2$  for the two of them, when the private preserve of each is thrown open to the other. This gain is what is called the

"entropy of mixing" though as we have seen it is really the "entropy of expansion". It is the alteration in the second term of the righthand member of (31).

But now suppose both of the boxes hold helium. One may indeed continue to suppose that when the partition is opened each one of the two samples of helium undergoes an expansion, doubling its volume. The entropy would then go up by  $2Nk \ln 2$ . However this looks so silly a thing to say that no one, I feel almost secure in affirming, has ever said it. The natural thing to say is, that the  $2N$  atoms of helium distributed through the two boxes at uniform temperature and uniform pressure have just the same entropy-value whether or not the partition is broken.

What does the classical statistics say about this situation? Its answer can be foretold. Since the two samples of helium are different by virtue of the different numberings of the two sets of atoms, the classical statistics insists that the entropy increase by  $2Nk \ln 2$  when the partition is broken, even though the gases are the same. This is indeed, if I may pervert the poem, "the little rift within the lute, which makes the classical statistics mute." The achievement of predicting the uniform distribution in ordinary space, the achievement of predicting the Maxwell-Boltzmann distribution-law in momentum-space, the achievement of providing the proper relation between temperature and mean kinetic energy—all of these are unsettled by this calamity.

Were I writing a strictly logical article I should quit at this point. Nothing further can apparently be done, except to tamper with the classical statistics in an effort to remove the unwanted result which has sprung forth to plague us. To violate the logic of the classical statistics in order to banish the undesired while keeping the desired results is a very questionable act. In theoretical physics, it is not admissible that the end justifies any and all means. Nevertheless so successful a feat of tampering has been done, that I cannot refrain from mentioning it as I close.

Let me first express in a slightly different way the nature of the "rift". Compare two samples of the same gas at the same temperature, one consisting of  $N$  atoms in a volume  $V$ , the other consisting of  $xN$  atoms in a volume  $xV$ . That which is called entropy in thermodynamics—and therefore that which is entropy, since it is the privilege of thermodynamics to give the definition of entropy—is  $x$  times as great for the latter as for the former. But that which the classical statistics calls entropy—or, as we must admit, miscalls entropy—is not  $x$  times as great for the latter as for the former. It would be, if there were  $x$  times as many atoms but just the same number of cells. However, there are  $x$  times as many atoms but also  $x$  times as many cells into which to put them. The number of complexes is approximately  $M^N$  in the former case and  $(xM)^{xN}$  in the latter,  $M$  standing

for the number of cells in the former box. The thing miscalled entropy is  $kN \ln M$  in the first case and  $(kxN \ln M + kxN \ln x)$  in the second case. It is the term  $kxN \ln x$  which is the rift.

Clearly we could abolish this term by allowing the volume of the cells to swell in the ratio  $x:1$  when going from the former case to the latter. This is the same as making  $H_0$  proportional to the number of atoms in the sample of gas which happens to be under study. Since in equation (31) the volumes  $V_0$  and  $H_0$  (of the elementary cells in ordinary space and in momentum-space) are indissolubly bound together in the product  $V_0 H_0$ , this is the same as making  $V_0 H_0$  equal to some constant multiplied by the number of atoms under study.

Such, if I interpret correctly, was the idea proposed by Sackur in 1912. While it does the task required, it is an "ad hoc" assumption of the most barefaced character. If the gas under study is at first divided into two parts by a partition and the partition is then abolished, the cells must be supposed to swell up at the moment when the partition vanishes.

We can also abolish the fatal term by going back to equation (1) for the number of complexions, and removing the factor  $N!$  in the numerator and replacing it by unity. We then have unity divided by the original denominator, which in the (most probable) case of the uniform distribution is  $(N/M)!$  raised to the power  $M$ , as I remarked on page 121. Using the super-Stirling approximation, we find that the logarithm of one fraction is  $(N \ln M - N \ln N)$ . The factor  $N!$  which we formerly had in the numerator killed off the term  $(-N \ln N)$ , but now that we have taken it out, this term survives. If now we say that  $k$  times the logarithm of  $W/N!$  shall be the picture of entropy in the classical statistics, then the term  $(-kN \ln N)$  comes over into the right-hand member of (31). It may be amalgamated with the last term already standing there; and when this is done, we find  $V_0 H_0$  multiplied by  $N$  exactly as Sackur put it there, and with the same wished-for result.

This, if I interpret correctly, is the idea proposed in 1913 by Tetrode. It does the task required of it, but its drawback is that the removal of the factor  $N!$  from the right-hand member of (1), a drastic piece of surgery as it were, violates the system of the classical statistics.<sup>3</sup>

I was not, however, thinking merely of this achievement when on Page 132 I spoke of "a remarkably successful feat of tampering." To show the

<sup>3</sup> This may seem too strong a statement. We are, after all, only asked to accept  $k \ln (W/N!)$  as our picture of entropy, instead of  $k \ln W$ ; why be reluctant? But in effect, as I see it, we are asked first to accept  $k \ln W f$  as our picture of entropy,  $f$  being an arbitrary function of  $N$ ; and then we are asked so to choose  $f$ , that the dependence of  $k \ln W f$  on  $N$  shall conform to the actual behavior of entropy. This is different from and much less impressive than our original procedure, which consisted in first realizing that  $W$  is the number of complexions, and then discovering that  $k \ln W$  depends on volume and on temperature in just the right ways for entropy.

magnitude of the achievement, I will rewrite equation (31) with two alterations. The first consists in replacing  $R$  with  $Nk$ , so that the expression shall refer not to a gramme-molecule of gas but to any number  $N$  of atoms. The second consists in following Tetrode by affirming that the entropy is not  $k \ln W$ , but  $k$  times the logarithm of  $W/N!$  I follow him still further by using, not the super-Stirling approximation in which  $N \ln N$  is written for  $\ln N!$ , but the better approximation in which  $(N \ln N - N)$  or  $(N \ln N - N \ln e)$  is written for  $\ln N!$  The result is:

$$S = (3/2)Nk \ln T + Nk \ln V - Nk \ln N \\ + Nk \ln [(2\pi mk^{3/2} e^{5/2}/V_0 H_0)] \quad (35)$$

This quantity newly chosen as the picture of "entropy" depends on volume and on temperature in the right way, as did the other. The dependence on  $N$  the number of atoms is now correct, and no wonder, for the new quantity was chosen with that purpose. There is a fourth term in the right-hand member which is proportional to  $N$ , and its value is completely determined if the value of  $V_0 H_0$  is fixed. The value which it takes when  $N$  is made equal to  $N_0$  may be called "the chemical constant"; but this name has been spoiled through being used with several different meanings, and should probably be abandoned.

When to  $V_0 H_0$ , the volume of the elementary cell in six-dimensional space, there is given the value  $h^3$ —the cube of Planck's constant—the resulting value of the fourth term is excellently confirmed by experiments on all of the noble gases, and (with less precision) by experiments on many of the monatomic vapors of metallic elements. This is the achievement known as "the verification of the Sackur-Tetrode formula" and it is indeed a grand one.

Anyone versed in thermodynamics will probably regard this not as a grand result, but as an incomprehensible one! Are we not taught in thermodynamics that nothing is ever measured about entropy except the *differences* between its values under different conditions, so an additive constant like the one in question must drop out of every verifiable equation, and its value can never be found? How then can it make sense to speak of confirming the value of the fourth term on the right-hand side of (35)?

Well, actually it *is* a difference which is measured: the difference between the entropy of the gas at any convenient temperature and volume and the entropy of its solidified crystalline form at the absolute zero. This difference is found to be such, that if for the entropy of the gas one puts the value (35) with  $h^3$  substituted for  $V_0 H_0$ , then for the entropy of the crystalline solid at the absolute zero one finds the value: *zero*. This result—this conclusion that the entropy of a crystal is zero at the absolute zero—is in itself so desirable and welcome that it is taken as the confirmation of the

Sackur-Tetrode formula. By "desirable and welcome" I mean that it is harmonious with the idea that *entropy is a measure of disorder*, an idea plausible in itself and fruitful in its applications. A chemical element perfectly crystallized at the absolute zero is supposed to be the exemplar of supreme order, and therefore its entropy ought to be nil. But this is an enormous subject requiring at least one other article, and I am glad that my attempt at writing such an article stands already in print in the June (1942) issue of this Journal.

Here then is the astonishing history of the Classical Statistics. By a strangely artificial device, the numbering of atoms deemed identical, it arrived at the proper distributions—that is, the distributions ratified by experiment—in ordinary space and in momentum-space. It then proposed a picture of entropy partially right, yet wrong in its dependence on the number of atoms, and therefore fatally wrong. With another artificial and dubious device, it corrected itself by adopting a new picture of entropy, this time depending in the right way upon the number of atoms. With a third artificial device (the introduction of Planck's constant in a peculiar way) it completed the formula for entropy in a manner leading to the consequence that the entropies of solidified crystallized elements are zero at absolute zero. All of these feats and more were subsequently achieved by the New Statistics, in a manner which I hope to explore on a later occasion.