

## II Groups and Statistical Mechanics

At about the beginning of the present century, two scientists, one in the United States and one in France, were working along lines which would have seemed to each of them entirely unrelated, if either had had the remotest idea of the existence of the other. In New Haven, Willard Gibbs was developing his new point of view in statistical mechanics. In Paris, Henri Lebesgue was rivalling the fame of his master Emile Borel by the discovery of a revised and more powerful theory of integration for use in the study of trigonometric series. The two discoverers were alike in this, that each was a man of the study rather than of the laboratory, but from this point on, their whole attitudes to science were diametrically opposite.

Gibbs, mathematician though he was, always regarded mathematics as ancillary to physics. Lebesgue was an analyst of the purest type, an able exponent of the extremely exacting modern standards of mathematical rigor, and a writer whose works, as far as I know, do not contain one single example of a problem or a method originating directly from physics. Nevertheless, the work of these two men forms a single whole in which the questions asked by Gibbs find their answers, not in his own work but in the work of Lebesgue.

The key idea of Gibbs is this: in Newton's dynamics, in its original form, we are concerned with an individual system, with given initial velocities and momenta, undergoing changes according to a certain system of forces under the Newtonian laws which link force and acceleration. In the vast majority of practical cases, however, we are far from knowing all the initial velocities and momenta. If we assume a certain initial distribution of the incompletely known positions and momenta of the system, this will determine in a completely Newtonian way the distribution of the momenta and positions for any future time. It will then be possible to make statements about these distributions, and some of these will have the character of assertions that the future system will have certain characteristics with probability one, or certain other characteristics with probability zero.

Probabilities one and zero are notions which include complete certainty and complete impossibility but include much more as well. If I shoot at a target with a bullet of the dimensions of a point, the chance that I hit any specific point on the target will generally be zero, although it is not impossible that I hit it; and indeed, in each specific case I must actually hit some specific point, which is an event of probability zero. Thus an event of probability one, that of my hitting *some* point, may be made up of an assemblage of instances of probability zero.

Nevertheless, one of the processes which is used in the technique of the Gibbsian statistical mechanics, although it is used implicitly, and Gibbs is nowhere clearly aware of it, is the resolution of a complex contingency into an infinite sequence of more special contingencies—a first, a second, a third, and so on—each of which has a known probability; and the expression of the probability of this larger contingency as the sum of the probabilities of the more special contingencies, which form an infinite

sequence. Thus we *cannot* sum probabilities in all conceivable cases, to get a probability of the total event—for the sum of any number of zeros is zero—while we *can* sum them if there is a first, a second, a third member, and so on, forming a sequence of contingencies in which every term has a definite position given by a positive integer.

The distinction between these two cases involves rather subtle considerations concerning the nature of sets of instances, and Gibbs, although a very powerful mathematician, was never a very subtle one. Is it possible for a class to be infinite and yet essentially different in multiplicity from another infinite class, such as that of the positive integers? This problem was solved toward the end of the last century by Georg Cantor, and the answer is “Yes.” If we consider all the distinct decimal fractions, terminating or non-terminating, lying between 0 and 1, it is known that they cannot be arranged in 1, 2, 3 order—although, strangely enough, all the *terminating* decimal fractions can be so arranged. Thus the distinction demanded by the Gibbs statistical mechanics is not on the face of it an impossible one. The service of Lebesgue to the Gibbs theory is to show that the implicit requirements of statistical mechanics concerning contingencies of probability zero and the addition of the probabilities of contingencies can actually be met, and that the Gibbsian theory does not involve contradictions.

Lebesgue’s work, however, was not directly based on the needs of statistical mechanics but on what looks like a very different theory, the theory of trigonometric series. This goes back to the eighteenth-century physics of waves and vibrations, and to the then moot question of the generality of the sets of motions of a linear system which can be synthesized out of the simple vibrations of the system—out of those vibrations, in other words, for

which the passing of time simply multiplies the deviations of the system from equilibrium by a quantity, positive or negative, dependent on the time alone and not on position. Thus a single function is expressed as the sum of a series. In these series, the coefficients are expressed as averages of the product of the function to be represented, multiplied by a given weighting function. The whole theory depends on the properties of the average of a series, in terms of the average of an individual term. Notice that the average of a quantity which is 1 over an interval from 0 to  $A$ , and 0 from  $A$  to 1, is  $A$ , and may be regarded as the probability that the random point should lie in the interval from 0 to  $A$  if it is known to lie between 0 and 1. In other words, the theory needed for the average of a series is very close to the theory needed for an adequate discussion of probabilities compounded from an infinite sequence of cases. This is the reason why Lebesgue, in solving his own problem, had also solved that of Gibbs.

The particular distributions discussed by Gibbs have themselves a dynamical interpretation. If we consider a certain very general sort of conservative dynamical system, with  $N$  degrees of freedom, we find that its position and velocity coordinates may be reduced to a special set of  $2N$  coordinates,  $N$  of which are called the generalized position coordinates and  $N$  the generalized momenta. These determine a  $2N$ -dimensional space defining a  $2N$ -dimensional volume; and if we take any region of this space and let the points flow with the course of time, which changes every set of  $2N$  coordinates into a new set depending on the elapsed time, the continual change of the boundary of the region does not change its  $2N$ -dimensional volume. In general, for sets not so simply defined as these regions, the notion of volume generates a system of measure of the type of Lebesgue.

In this system of measure, and in the conservative dynamical systems which are transformed in such a way as to keep this measure constant, there is one other numerically valued entity which also remains constant: the energy. If all the bodies in the system act only on one another and there are no forces attached to fixed positions and fixed orientations in space, there are two other expressions which also remain constant. Both of these are vectors: the momentum, and the moment of momentum of the system as a whole. They are not difficult to eliminate, so that the system is replaced by a system with fewer degrees of freedom.

In highly specialized systems, there may be other quantities not determined by the energy, the momentum, and the moment of momentum, which are unchanged as the system develops. However, it is known that systems in which another invariant quantity exists, dependent on the initial coordinates and momenta of a dynamical system, and regular enough to be subject to the system of integration based on Lebesgue measure, are very rare indeed in a quite precise sense.<sup>1</sup> In systems without other invariant quantities, we can fix the coordinates corresponding to energy, momentum, and total moment of momentum, and in the space of the remaining coordinates, the measure determined by the position and momentum coordinates will itself determine a sort of sub-measure, just as measure in space will determine area on a two-dimensional surface out of a family of two-dimensional surfaces. For example, if our family is that of concentric spheres, then the volume between two concentric spheres close together, when normalized by taking as one the total volume of the region between the two spheres, will give in the limit a measure of area on the surface of a sphere.

Let us then take this new measure on a region in phase space for which energy, total momentum, and total moment of momentum are determined, and let us suppose that there are no other measurable invariant quantities in the system. Let the total measure of this restricted region be constant, or as we can make it by a change in scale, 1. As our measure has been obtained from a measure invariant in time, in a way invariant in time, it will itself be invariant. We shall call this measure *phase measure*, and averages taken with respect to it *phase averages*.

However, any quantity varying in time may also have a *time average*. If, for example,  $f(t)$  depends on  $t$ , its time average for the past will be

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 f(t) dt \quad (2.01)$$

and its time average for the future

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(t) dt \quad (2.02)$$

In Gibbs' statistical mechanics, both time averages and space averages occur. It was a brilliant idea of Gibbs to try to show that these two types of average were, in some sense, the same. In the notion that these two types of average were related, Gibbs was perfectly right; and in the method by which he tried to show this relation, he was utterly and hopelessly wrong. For this he was scarcely to blame. Even at the time of his death, the fame of the Lebesgue integral had just begun to penetrate to America. For another fifteen years, it was a museum curiosity, only useful to show to young mathematicians the needs and possibilities of rigor. A mathematician as distinguished as W. F. Osgood<sup>2</sup> would have nothing to do with it till his dying day. It was not until about 1930 that a group of mathematicians—Koopman, von

Neumann, Birkhoff<sup>3</sup>—finally established the proper foundations of the Gibbs statistical mechanics. Later, in the study of ergodic theory, we shall see what these foundations were.

Gibbs himself thought that in a system from which all the invariants had been removed as extra coordinates almost all paths of points in phase space passed through all coordinates in such a space. This hypothesis he called the *ergodic hypothesis*, from the Greek words *ἔργον*, “work,” and *ὁδός*, “path.” Now, in the first place, as Plancherel and others have shown, there is no significant case where that hypothesis is true. No differentiable path can cover an area in the plane, even if it is of infinite length. The followers of Gibbs, including at the end perhaps Gibbs himself, saw this in a vague way, and replaced this hypothesis by the *quasi-ergodic* hypothesis, which merely asserts that in the course of time a system generally passes indefinitely near to every point in the region of phase space determined by the known invariants. There is no logical difficulty as to the truth of this: it is merely quite inadequate for the conclusions which Gibbs bases on it. It says nothing about the relative time which the system spends in the neighborhood of each point.

Beside the notions of *average* and of *measure*—the average over a universe of a function 1 over a set to be measured and 0 elsewhere—which were most urgently needed to make sense out of Gibbs' theory, in order to appreciate the real significance of ergodic theory we need a more precise analysis of the notion of *invariant*, as well as the notion of *transformation group*. These notions were certainly familiar to Gibbs, as his study of vector analysis shows. Nevertheless, it is possible to maintain that he did not assess them at their full philosophical value. Like his contemporary Heaviside, Gibbs is one of the scientists whose physico-mathematical acumen often outstrips their logic and

who are generally right, while they are often unable to explain why and how they are right.

For the existence of any science, it is necessary that there exist phenomena which do not stand isolated. In a world ruled by a succession of miracles performed by an irrational God subject to sudden whims, we should be forced to await each new catastrophe in a state of perplexed passiveness. We have a picture of such a world in the croquet game in *Alice in Wonderland*; where the mallets are flamingos; the balls, hedgehogs, which quietly unroll and go about their own business; the hoops, playing-card soldiers, likewise subject to locomotor initiative of their own; and the rules are the decrees of the testy, unpredictable Queen of Hearts.

The essence of an effective rule for a game or a useful law of physics is that it be storable in advance, and that it apply to more than one case. Ideally, it should represent a property of the system discussed which remains the same under the flux of particular circumstances. In the simplest case, it is a property which is *invariant* to a set of *transformations* to which the system is subject. We are thus led to the notions of *transformation*, *transformation group*, and *invariant*.

A transformation of a system is some alteration in which each element goes into another. The modification of the solar system which occurs in the transition between time  $t_1$  and time  $t_2$  is a transformation of the sets of coordinates of the planets. The similar change in their coordinate when we move their origin, or subject our geometric axes to a rotation, is a transformation. The change in scale which occurs when we examine a preparation under the magnifying action of a microscope is likewise a transformation.



The result of following a transformation  $A$  by a transformation  $B$  is another transformation, known as the *product* or *resultant*  $BA$ . Note that in general it depends on the order of  $A$  and  $B$ . Thus if  $A$  is the transformation which takes the coordinate  $x$  into the coordinate  $y$ , and  $y$  into  $-x$ , while  $z$  is unchanged; while  $B$  takes  $x$  into  $z$ ,  $z$  into  $-x$ , and  $y$  is unchanged; then  $BA$  will take  $x$  into  $y$ ,  $y$  into  $-z$ , and  $z$  into  $-x$ ; while  $AB$  will take  $x$  into  $z$ ,  $y$  into  $-x$ , and  $z$  into  $-y$ . If  $AB$  and  $BA$  are the same, we shall say that  $A$  and  $B$  are *permutable*.

Sometimes, but not always, the transformation  $A$  will not only carry every element of the system into an element but will have the property that every element is the result of transforming an element. In this case, there is a unique transformation  $A^{-1}$ , such that both  $AA^{-1}$  and  $A^{-1}A$  are that very special transformation which we call  $I$ , the *identity transformation*, which transforms every element into itself. In this case we call  $A^{-1}$  the *inverse* of  $A$ . It is clear that  $A$  is the inverse of  $A^{-1}$ , that  $I$  is its own inverse, and that the inverse of  $AB$  is  $B^{-1}A^{-1}$ .

There exist certain sets of transformations where every transformation belonging to the set has an inverse, likewise belonging to the set; and where the resultant of any two transformations belonging to the set itself belongs to the set. These sets are known as *transformation groups*. The set of all translations along a line, or in a plane, or in a three-dimensional space, is a transformation group; and even more, it is a transformation group of the special sort known as *Abelian*, where any two transformations of the group are permutable. The set of rotations about a point, and of all motions of a rigid body in space, are non-Abelian groups.

Let us suppose that we have some quantity attached to all the elements transformed by a transformation group. If this quantity is unchanged when each element is changed by the same

transformation of the group, whatever that transformation may be, it is called *an invariant of the group*. There are many sorts of such group invariants, of which two are especially important for our purposes.

The first are the so-called *linear invariants*. Let the elements transformed by an Abelian group be the terms which we represent by  $x$ , and let  $f(x)$  be a complex-valued function of these elements, with certain appropriate properties of continuity or integrability. Then if  $Tx$  stands for the element resulting from  $x$  under the transformation  $T$ , and if  $f(x)$  is a function of absolute value 1, such that

$$f(Tx) = \alpha(T)f(x) \tag{2.03}$$

where  $\alpha(T)$  is a number of absolute value 1 depending only on  $T$ , we shall say that  $f(x)$  is a *character* of the group. It is an invariant of the group in a slightly generalized sense. If  $f(x)$  and  $g(x)$  are group characters, clearly  $f(x)g(x)$  is one also, as is  $[f(x)]^{-1}$ . If we can represent any function  $h(x)$  defined over the group as a linear combination of the characters of the group, in some such form as

$$h(x) = \sum A_k f_k(x) \tag{2.04}$$

where  $f_k(x)$  is a character of the group, and  $\alpha_k(T)$  bears the same relation to  $f_k(x)$  that  $\alpha(T)$  does to  $f(x)$  in Eq. 2.03, then

$$h(Tx) = \sum A_k \alpha_k(T) f_k(x) \tag{2.05}$$

Thus if we can develop  $h(x)$  in terms of a set of group characters, we can develop  $h(Tx)$  for all  $T$  in terms of the characters.

We have seen that the characters of a group generate other characters under multiplication and inversion, and it may similarly be seen that the constant 1 is a character. Multiplication by a group character thus generates a transformation group of

the group characters themselves, which is known as the *character group* of the original group.

If the original group is the translation group on the infinite line, so that the operator  $T$  changes  $x$  into  $x + T$ , Eq. 2.03 becomes

$$f(x + T) = \alpha(T)f(x) \tag{2.06}$$

which is satisfied if  $f(x) = e^{i\lambda x}$ ,  $\alpha(T) = e^{i\lambda T}$ . The characters will be the functions  $e^{i\lambda x}$ , and the character group will be the group of translations changing  $\lambda$  into  $\lambda + \tau$ , thus having the same structure as the original group. This will not be the case when the original group consists of the rotations about a circle. In this case, the operator  $T$  changes  $x$  into a number between 0 and  $2\pi$ , differing from  $x + T$  by an integral multiple of  $2\pi$ , and, while Eq. 2.06 will still hold, we have the extra condition that

$$\alpha(T + 2\pi) = \alpha(T) \tag{2.07}$$

If now we put  $f(x) = e^{i\lambda x}$  as before, we shall obtain

$$e^{i2\pi\lambda} = 1 \tag{2.08}$$

which means that  $\lambda$  must be a real integer, positive, negative, or zero. The character group thus corresponds to the translations of the real integers. If, on the other hand, the original group is that of the translations of the integers,  $x$  and  $T$  in Eq. 2.05 are confined to the integer values, and  $e^{i\lambda x}$  involves only the number between 0 and  $2\pi$  which differs from  $\lambda$  by an integral multiple of  $2\pi$ . Thus the character group is essentially the group of rotations about a circle.

In any character group, for a given character  $f$ , the values of  $\alpha(T)$  are distributed in such a way that the distribution is not altered when they are all multiplied by  $\alpha(S)$ , for any elements in the group. That is, if there is any reasonable basis of taking an average of these values which is not affected by the

transformation of the group by the multiplication of each transformation by a fixed one of its transformations, either  $\alpha(T)$  is always 1, or this average is invariant when multiplied by some number not 1, and must be 0. From this it may be concluded that the average of the product of any character by its conjugate (which will also be a character) will have the value 1, and that the average of the product of any character by the conjugate of another character will have the value 0. In other words, if we can express  $h(x)$  as in Eq. 2.04, we shall have

$$A_k = \text{average} [h(x) \overline{f_k(x)}] \quad (2.09)$$

In the case of the group of rotations on a circle, this gives us directly that if

$$f(x) = \sum a_n e^{inx} \quad (2.10)$$

then

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx \quad (2.11)$$

and the result for translations along the infinite line is closely related to the fact that if in an appropriate sense

$$f(x) = \int_{-\infty}^{\infty} a(\lambda) e^{i\lambda x} d\lambda \quad (2.12)$$

then in a certain sense

$$a(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\lambda x} dx \quad (2.13)$$

These results have been stated here very roughly and without a clear statement of their conditions of validity. For more precise statements of the theory, the reader should consult the following reference.<sup>4</sup>

Beside the theory of the linear invariants of a group, there is also the general theory of its metrical invariants. These are the systems of Lebesgue measure which do not undergo any change when the objects transformed by the group are permuted by the operators of the group. In this connection, we should cite the interesting theory of group measure, due to Haar.<sup>5</sup> As we have seen, every group itself is a collection of objects which are permuted by being multiplied by the operations of the group itself. As such, it may have an invariant measure. Haar has proved that a certain rather wide class of groups does possess a uniquely determined invariant measure, definable in terms of the structure of the group itself.

The most important application of the theory of the metrical invariants of a group of transformations is to show the justification of that interchangeability of phase averages and time averages which, as we have already seen, Gibbs tried in vain to establish. The basis on which this has been accomplished is known as the ergodic theory. The ordinary ergodic theorems start with an ensemble  $E$ , which we can take to be of measure 1, transformed into itself by a measure-preserving transformation  $T$  or by a group of measure-preserving transformations  $T^\lambda$ , where  $-\infty < \lambda < \infty$  and where

$$T^\lambda \cdot T^\mu = T^{\lambda+\mu} \tag{2.14}$$

Ergodic theory concerns itself with complex-valued functions  $f(x)$  of the elements  $x$  of  $E$ . In all cases,  $f(x)$  is taken to be measurable in  $x$ , and if we are concerned with a continuous group of transformations,  $f(T^\lambda x)$  is taken to be measurable in  $x$  and  $\lambda$  simultaneously.

In the mean ergodic theorem of Koopman and von Neumann,  $f(x)$  is taken to be of class  $L^2$ ; that is,

$$\int_E |f(x)|^2 dx < \infty \tag{2.15}$$

The theorem then asserts that

$$f_N(x) = \frac{1}{N+1} \sum_{n=0}^N f(T^n x) \tag{2.16}$$

or

$$f_A(x) = \frac{1}{A} \int_0^A f(T^\lambda x) d\lambda \tag{2.17}$$

as the case may be, converges in the mean to a limit  $f^*(x)$  as  $N \rightarrow \infty$  or  $A \rightarrow \infty$ , respectively, in the sense that

$$\lim_{N \rightarrow \infty} \int_E |f^*(x) - f_N(x)|^2 dx = 0 \tag{2.18}$$

$$\lim_{A \rightarrow \infty} \int_E |f^*(x) - f_A(x)|^2 dx = 0 \tag{2.19}$$

In the “almost everywhere” ergodic theorem of Birkhoff,  $f(x)$  is taken to be of class  $L$ ; which means that

$$\int_E |f(x)| dx < \infty \tag{2.20}$$

The functions  $f_N(x)$  and  $f_A(x)$  are defined as in Eqs. 2.16 and 2.17. The theorem then states that, except for a set of values of  $x$  of measure 0,

$$f^*(x) = \lim_{N \rightarrow \infty} f_N(x) \tag{2.21}$$

and

$$f^*(x) = \lim_{A \rightarrow \infty} f_A(x) \tag{2.22}$$

exist.

A very interesting case is the so-called *ergodic* or *metrically transitive* one, in which the transformation  $T$  or the set of transformations  $TA$  leaves invariant no set of points  $x$  which has a

measure other than 1 or 0. In such a case, the set of values (for either ergodic theorem) for which  $f^*$  takes on a certain range of value is almost always either 1 or 0. This is impossible unless  $f^*(x)$  is almost always constant. The value which  $f^*(x)$  then assumes almost always is

$$\int_0^1 f(x) dx \tag{2.23}$$

That is, in the Koopman theorem, we have the limit in the mean

$$\text{l.i.m.}_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N f(T^n x) = \int_0^1 f(x) dx \tag{2.24}$$

and in the Birkhoff theorem, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N f(T^n x) = \int_0^1 f(x) dx \tag{2.25}$$

except for a set of values of  $x$  of zero measure or probability 0. Similar results hold in the continuous case. This is an adequate justification for Gibbs' interchange of phase averages and time averages.

Where the transformation  $T$  or the transformation group  $TA$  is not ergodic, von Neumann has shown under very general conditions that they can be reduced to ergodic components. That is, except for a set of values of  $x$  of zero measure,  $E$  can be separated into a finite or denumerable set of classes  $E_n$  and a continuum of classes  $E(y)$ , such that a measure is established on each  $E_n$  and  $E(y)$ , which is invariant under  $T$  or  $T^t$ . These transformations are all ergodic; and if  $S(y)$  is the intersection of  $S$  with  $E(y)$  and  $S_n$  with  $E_n$ , then

$$\text{measure}_E(S) = \int \text{measure}_{E(y)}[S(y)] dy + \sum_{E_n} \text{measure}_{E_n}(S_n) \tag{2.26}$$

In other words, the whole theory of measure-preserving transformations can be reduced to the theory of ergodic transformations.

The whole of ergodic theory, let us remark in passing, may be applied to groups of transformations more general than those isomorphic with the translation group on the line. In particular, it may be applied to the translation group in  $n$  dimensions. The case of three dimensions is physically important. The spatial analogue of temporal equilibrium is spatial homogeneity, and such theories as that of the homogeneous gas, liquid, or solid depend on the application of three-dimensional ergodic theory. Incidentally, a non-ergodic group of translation transformations in three dimensions appears as the set of translations of a mixture of distinct states, such that one or another exists at a given time, not a mixture of both.

One of the cardinal notions of statistical mechanics, which also receives an application in the classical thermodynamics, is that of *entropy*. It is primarily a property of regions in phase space and expresses the logarithm of their probability measure. For example, let us consider the dynamics of  $n$  particles in a bottle, divided into two parts,  $A$  and  $B$ . If  $m$  particles are in  $A$ , and  $n - m$  in  $B$ , we have characterized a region in phase space, and it will have a certain probability measure. The logarithm is the entropy of the distribution:  $m$  particles in  $A$ ,  $n - m$  in  $B$ . The system will spend most of its time in a state near that of greatest entropy, in the sense that for most of the time, nearly  $m_1$  particles will be in  $A$ , nearly  $n - m_1$  in  $B$ , where the probability of the combination  $m_1$  in  $A$ ,  $n - m_1$  in  $B$  is a maximum. For systems with a large number of particles and states within the limits of practical discrimination, this means that if we take a state of other than maximum entropy and observe what happens to it, the entropy almost always increases.



In the ordinary thermodynamic problems of the heat engine, we are dealing with conditions in which we have a rough thermal equilibrium in large regions like an engine cylinder. The states for which we study the entropy are states involving maximum entropy for a given temperature and volume, for a small number of regions of the given volumes and at the given temperature assumed. Even the more refined discussions of thermal engines, particularly of thermal engines like the turbine, in which a gas is expanding in a more complicated manner than in a cylinder, do not change these conditions too radically. We may still talk of local temperatures, with a very fair approximation, even though no temperature is precisely determined except in a state of equilibrium and by methods involving this equilibrium. However, in living matter, we lose much of even this rough homogeneity. The structure of a protein tissue as shown by the electron microscope has an enormous definiteness and fineness of texture, and its physiology is certainly of a corresponding fineness of texture. This fineness is far greater than that of the space-and-time scale of an ordinary thermometer, and so the temperatures read by ordinary thermometers in living tissues are gross averages and not the true temperatures of thermodynamics. Gibbsian statistical mechanics may well be a fairly adequate model of what happens in the body; the picture suggested by the ordinary heat engine certainly is not. The thermal efficiency of muscle action means next to nothing, and certainly does not mean what it appears to mean.

A very important idea in statistical mechanics is that of the Maxwell demon. Let us suppose a gas in which the particles are moving around with the distribution of velocities in statistical equilibrium for a given temperature. For a perfect gas, this is the Maxwell distribution. Let this gas be contained in a rigid

container with a wall across it, containing an opening spanned by a small gate, operated by a gatekeeper, either an anthropomorphic demon or a minute mechanism. When a particle of more than average velocity approaches the gate from compartment *A* or a particle of less than average velocity approaches the gate from compartment *B*, the gatekeeper opens the gate, and the particle passes through; but when a particle of less than average velocity approaches from compartment *A* or a particle of greater than average velocity approaches from compartment *B*, the gate is closed. In this way, the concentration of particles of high velocity is increased in compartment *B* and is decreased in compartment *A*. This produces an apparent decrease in entropy; so that if the two compartments are now connected by a heat engine, we seem to obtain a perpetual-motion machine of the second kind.

It is simpler to repel the question posed by the Maxwell demon than to answer it. Nothing is easier than to deny the possibility of such beings or structures. We shall actually find that Maxwell demons in the strictest sense cannot exist in a system in equilibrium, but if we accept this from the beginning, and so not try to demonstrate it, we shall miss an admirable opportunity to learn something about entropy and about possible physical, chemical, and biological systems.

For a Maxwell demon to act, it must receive information from approaching particles concerning their velocity and point of impact on the wall. Whether these impulses involve a transfer of energy or not, they must involve a coupling of the demon and the gas. Now, the law of the increase of entropy applies to a completely isolated system but does not apply to a non-isolated part of such a system. Accordingly, the only entropy which concerns us is that of the system gas-demon, and not that of the gas

alone. The gas entropy is merely one term in the total entropy of the larger system. Can we find terms involving the demon as well which contribute to this total entropy?

Most certainly we can. The demon can only act on information received, and this information, as we shall see in the next chapter, represents a negative entropy. The information must be carried by some physical process, say some form of radiation. It may very well be that this information is carried at a very low energy level, and that the transfer of energy between particle and demon is for a considerable time far less significant than the transfer of information. However, under the quantum mechanics, it is impossible to obtain any information giving the position or the momentum of a particle, much less the two together, without a positive effect on the energy of the particle examined, exceeding a minimum dependent on the frequency of the light used for examination. Thus all coupling is strictly a coupling involving energy, and a system in statistical equilibrium is in equilibrium both in matters concerning entropy and those concerning energy. In the long run, the Maxwell demon is itself subject to a random motion corresponding to the temperature of its environment, and, as Leibniz says of some of his monads, it receives a large number of small impressions, until it falls into "a certain vertigo" and is incapable of clear perceptions. In fact, it ceases to act as a Maxwell demon.

Nevertheless, there may be a quite appreciable interval of time before the demon is deconditioned, and this time may be so prolonged that we may speak of the active phase of the demon as metastable. There is no reason to suppose that metastable demons do not in fact exist; indeed, it may well be that enzymes are metastable Maxwell demons, decreasing entropy, perhaps not by the separation between fast and slow particles

but by some other equivalent process. We may well regard living organisms, such as Man himself, in this light. Certainly the enzyme and the living organism are alike metastable: the stable state of an enzyme is to be deconditioned, and the stable state of a living organism is to be dead. All catalysts are ultimately poisoned: they change rates of reaction but not true equilibrium. Nevertheless, catalysts and Man alike have sufficiently definite states of metastability to deserve the recognition of these states as relatively permanent conditions.

I do not wish to close this chapter without indicating that ergodic theory is a considerably wider subject than we have indicated above. There are certain modern developments of ergodic theory in which the measure to be kept invariant under a set of transformations is defined directly by the set itself rather than assumed in advance. I refer especially to the work of Kryloff and Bogoliouboff, and to some of the work of Hurewicz and the Japanese school.

The next chapter is devoted to the statistical mechanics of time series. This is another field in which the conditions are very remote from those of the statistical mechanics of heat engines and which is thus very well suited to serve as a model of what happens in the living organism.

This is a section of [doi:10.7551/mitpress/11810.001.0001](https://doi.org/10.7551/mitpress/11810.001.0001)

# Cybernetics or Control and Communication in the Animal and the Machine

By: Norbert Wiener

## Citation:

*Cybernetics or Control and Communication in the Animal and the Machine*

By: Norbert Wiener

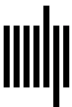
DOI: [10.7551/mitpress/11810.001.0001](https://doi.org/10.7551/mitpress/11810.001.0001)

ISBN (electronic): 9780262355902

Publisher: The MIT Press

Published: 2019

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

© 2019, 1961, 1948 Massachusetts Institute of Technology

First MIT Press paperback edition, February 1965

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by Toppan Best-set Premedia Limited. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Wiener, Norbert, 1894-1964, author.

Title: Cybernetics ; or, Control and communication in the animal and the machine / Norbert Wiener ; forewords by Doug Hill and Sanjoy Mitter.

Other titles: Control and communication in the animal and the machine

Description: [Second edition, 2019 reissue]. | Cambridge, MA : The MIT Press, [2019] | "Reissue of the 1961 second edition." | Includes bibliographical references and index.

Identifiers: LCCN 2019005612 | ISBN 9780262537841 (pbk. : alk. paper)

Subjects: LCSH: Cybernetics. | Control theory. | System theory.

Classification: LCC Q310 .W47 2019 | DDC 003/.5--dc23 LC record available at <https://lccn.loc.gov/2019005612>

10 9 8 7 6 5 4 3 2 1