Introductory note to 1900

Jos Uffink

Zermelo returned to the foundations of statistical physics in his *Habilitation* lecture for the University of Göttingen. Here, he refers to Boltzmann's assumption of molecular disorder, but without attempting to describe what Boltzmann understood by this notion. Zermelo understood this and similar assumptions as implying that various stages of molecular motion can be seen as probabilistically independent. At least, this understanding seems to be suggested by the concern he expressed, namely, that due to the assumed equations of motion in the theory of gases, or dynamical systems in general, these various stages are *not* mutually independent but functionally related.

Zermelo proposes a definition of a probability measure (say ρ on phase space) by taking the probability of any region or open set A (which Zermelo calls g) to be the Lebesgue measure μ of that set divided by the total measure of the phase space: $\rho(A) = \mu(A)/\mu(\Gamma)$.

Zermelo then presents three theorems. The first is actually just the well-known Liouville theorem, and does not need comment. The other two are more interesting since they deal again with the question whether any entropy-like function could be defined on phase space which might have the property of changing monotonically in time.

Zermelo's second theorem applies to the case when we take A to be invariant under time evolution, i.e.: $\forall t \in \mathbb{R} : T_t A = A$. For example, A might be the energy hypersurface $\{x \in \Gamma \mid H(x) = \text{const.}\}$ (in which case the probability measure becomes the microcanonical one) or an energy hypershell $\{x \in \Gamma \mid a < H(x) < b\}$ etc. The theorem states: For every differentiable function S on phase space, the phase average value of $\frac{dS}{dt}$ over such an invariant region A is zero. In other words:

$$\frac{1}{\mu(\Gamma)} \int_A \frac{dS(x_t)}{dt} d\mu(x) = 0\,. \tag{1}$$

H.-D. Ebbinghaus, A. Kanamori (Eds.), *Ernst Zermelo - Collected Works/ Gesammelte Werke II*, DOI 10.1007/978-3-540-70856-8_6, Schriften der Mathematisch-naturwissenschaftlichen Klasse der Heidelberger Akademie der Wissenschaften 23,

Zermelo's third theorem extends this result to arbitrary open regions A that need not be invariant.

These last two theorems are analogous to the lemma (Theorem 3.2 on p. 206 of this volume) he presented in his *1896a*. Of course, the difference is that in the present manuscript the result is restricted to differentiable functions, but otherwise both theorems provided here imply the same conclusion: for any such function S, there can be no open set A such that for all $x \in A$, $S(T_t x)$ would be monotonically increasing in t. The present results however seem to be stronger than the lemma based on the recurrence theorem, since the mere non-existence of monotonically increasing functions on an open set A does not by itself imply the vanishing of the phase-average of their derivative.

Zermelo refrained from drawing any conclusion that was critical of Boltzmann. He only stated that these results "may serve as a basis for further considerations by showing us which probabilities and mean values we are permitted to generally replace by other, simpler, ones." Of course, it is hard (at least for me) to resist the temptation of thinking that his reference to other but unspecified "simpler" notions of probability refers to Boltzmann's equivocation of phase-average probabilities and time-average probabilities that he questioned in his response to Boltzmann. But he does not go into details.

Zermelo ends by announcing an intention to attempt a new derivation of Maxwell's distribution law "solely on the basis of the definition of probability given here without any further hypothesis." This intention does not seem to have materialized.

All in all, I would say that this paper contains little of crucial importance to the current debate on the foundations of statistical physics. What it shows is that Zermelo was not in the least bit moved by Boltzmann's replies. He however refrained from mounting or reiterating explicit criticisms of Boltzmann. Zermelo may have chosen to avoid explicitly raising controversy in his *Habilitation* lecture.

Über die Anwendung der Wahrscheinlichkeitsrechnung auf dynamische Systeme*

1900

Auf verschiedenen Gebieten der mathematischen Physik, insbesondere in der kinetischen Gastheorie, entsteht das Problem, die Bewegung solcher mechanischen Systeme zu untersuchen, welche eine zwar endliche, aber sehr grosse Anzahl von Freiheitsgraden besitzen und dementsprechend einer komplicierten, turbulenten Bewegung unterworfen sind. In einem solchen Systeme gemäss den bekannten Principien der Mechanik die Bewegung eines jeden Punktes vollständig zu beschreiben, wäre zwar wünschenswert als eine durchaus sichere Grundlage für alle weiteren Betrachtungen, ist aber fast immer unausführbar und für den vorliegenden Zweck meist auch unnötig, weil sich diese individuelle Bewegung der Beobachtung ebenso entzieht wie der Rechnung und nur der allgemeine, durchschnittliche Grundcharakter des Vorganges eine anschauliche physikalische Bedeutung besitzt. Man wird sich also auf die Betrachtung gewisser Durchschnittsgrössen beschränken, welche physikalisch messbaren Zuständen entsprechen, wie z. B. die mittlere lebendige Kraft in der Gastheorie der Temperatur; man wird aber auch diese Durchschnittswerte nicht genau als Funktionen der Zeit bestimmen können, weil uns der wahre, die ganze Bewegung bestimmende Anfangszustand des mechanischen Systemes, gegeben durch die sämtlichen Koordinaten und Geschwindigkeiten, nicht vollständig bekannt ist, sondern eben nur die anfänglichen Durchschnittswerte, welche, streng genommen, auch zur Bestimmung des sichtbaren Vorganges nicht ausreichen. Vielmehr wird man sich auch hier mit Annäherungen und Wahrscheinlichkeiten begnügen müssen, und als praktisch gewiss wird man schon solche Veränderungen betrachten dürfen, deren Wahrscheinlichkeit bei gewissen, durch die Natur des Systemes nahe gelegten Grenzübergängen, in
318 der Gastheorie z. B. bei unbegrenzt wachsender Anzahl der Moleküle, nach | 1 konvergiert. Wie soll man aber diese Wahrscheinlichkeiten, diese Mittelwerte definieren, wie soll man mit ihnen operieren, um wenigstens im Sinne der Wahrscheinlichkeitsrechnung zu mathematisch zuverlässigen Ergebnissen zu gelangen? Hier hat man sich in verschiedener Weise zu helfen gesucht, man hat (wie z. B. *Helmholtz*) die turbulente Bewegung als eine cyklische aufgefasst, man hat das wirkliche kinetische Potential durch vereinfachte Ausdrücke ersetzt (wie *J. J. Thomson*), die nur von den physikalischen Durchschnittswerten abhängen, oder man ist auch von der Annahme, dass die Bewegung der Moleküle eine „ungeordnete" sei, ausgegangen (wie *Boltzmann*). Aber alle solche Auskunftsmittel, so plausibel sie auch scheinen, so förderlich sie sich auch für den gerade vorliegenden Zweck erweisen mögen, scheinen mir

* Göttinger Habilitationsvorlesung, gehalten am 4. März 1899.

On the application of the calculus of probabilities to dynamical systems*

1900

In different areas of mathematical physics, and in particular in the kinetic theory of gases, we face the problem of investigating the motions of mechanical systems that possess a finite, albeit very great, number of degrees of freedom, and that, accordingly, are subject to complicated, turbulent motions. While it would certainly be desirable to describe completely the motion of each point in such a system according to the well-known principles of mechanics in order to place all further considerations on an absolutely firm footing, it is almost always infeasible to do so and, for the present purpose, also mostly unnecessary since this individual motion defies both observation and computation and only the general, average basic character of the process has a concrete physical meaning. For this reason, one will limit oneself to the consideration of certain average magnitudes that correspond to physically measurable states such as the mean living force in the gas theory of temperature; but it will not be possible to determine precisely these average values as functions of time either since the true initial state of the mechanical system that determines the entire motion and that is given by all of the coordinates and velocities is not completely known to us, but only the initial average values, that, strictly speaking, are not sufficient either to determine the visible process. Rather, one will have to make do with approximations and probabilities also in this case, and one may already consider such changes as practically certain whose probability converges to 1 for certain transitions to limits suggested by the nature of the system, such as for an indefinitely increasing number of molecules in the theory of gases. But how is one to define these probabilities, these mean values? How is one to operate with them in order to obtain results that are mathematically reliable, at least by the standards of the calculus of probabilities? Several approaches have been taken to answer these questions. *Helmholtz*, e.g., conceived of the turbulent motion as a cyclic motion, and *J.J. Thomson* replaced the real kinetic potential by simplified expressions dependent solely on the physical average values, while *Boltzmann*, e.g., also proceeded from the assumption that the motion of the molecules is an “unordered” motion. But it seems to me that such devices, as plausible as they may appear and as advantageous as they may turn out to be for the particular purpose at hand, are

* Habilitation lecture at Göttingen, delivered on March 4, 1899.

doch einem prinzipiellen Bedenken zu unterliegen: es werden hier Hypothesen zu Grunde gelegt, die sich auf den ganzen Verlauf der Bewegung beziehen sollen, während doch die verschiedenen Phasen derselben nicht von einander unabhängig sind, sondern durch die Gesetze der Mechanik (in uns freilich noch unbekannter Weise) notwendig zusammenhängen, sodass die Frage berechtigt scheint, ob denn eine Annahme, die für einen gegebenen Zeitpunkt im Sinne der Wahrscheinlichkeitsrechnung zulässig sein mag, es auch für spätere Zeiten bleiben werde oder ob nicht vielmehr das System vermöge seiner eigenen Konstitution mit Notwendigkeit anderen Zuständen zustreben werde, in denen diese Voraussetzung nicht mehr erfüllt ist? In der That scheinen mir auch gewisse Widersprüche, die sich bei manchen gastheoretischen Betrachtungen ergeben, der Nichtbeachtung solcher Bedenken zuzuschreiben zu sein.

Um also eine zuverlässigere Grundlage für die Behandlung solcher Probleme zu gewinnen, scheint mir die Forderung unabweisbar, eine bestimmte Definition der Wahrscheinlichkeit an die Spitze zu stellen, die, wenn auch in gewissem Grade willkürlich, doch im Verlaufe der Untersuchung nicht mehr geändert oder durch neue Annahmen ergänzt werden darf, und man wird ferner unbedingt festhalten müssen an dem *Laplace*schen Wahrscheinlichkeitssatze, nach welchem zwei notwendig wie Ursache und Wirkung mit einander verbundene Ereignisse a und b, sodass das Eintreten des einen von ihnen das des anderen bedingt, auch immer gleich wahrscheinlich sein müssen. Diesen Anforderungen werden wir genügen, wenn wir die Wahrscheinlichkeit irgend eines dynamischen Zustandes, welcher bei der Bewegung unseres Systemes zu irgend einer Zeit t eintreten soll, definieren durch die Wahrscheinlichkeit desjenigen Anfangszustandes zu einer Zeit $t = 0$, aus dem der betrachtete selbst hervorgegangen sein muss.

Nehmen wir an, dass auf unser System ausschliesslich Potentialkräfte wirken, welche allein von der augenblicklichen Konfiguration, d. h. von der *Lage* der Punkte, abhängen sollen, so ist der Bewegungszustand zu einer gegebenen Zeit t und damit auch die ganze Bewegung vollständig bestimmt durch das System aller Koordinaten $q_1, q_2, \dots q_n$ und der zugehörigen Bewegungsmomente (oder Impulskoordinaten) $p_1, p_2, \dots p_n$, die wir alle zusammen abgekürzt durch (q, p) bezeichnen wollen. Ebenso ist der Anfangszustand zur Zeit $t = 0$ bestimmt durch das System der entsprechenden Grössen $\overline{q}_\lambda$ und $\overline{p}_\lambda$ oder abgekürzt durch $(\overline{q}, \overline{p})$, während die Bewegungsgleichungen sich unter der gemachten Voraussetzung in der *Hamilton*'schen Form schreiben lassen:

$$\frac{dq_\lambda}{dt} = \frac{\partial H}{\partial p_\lambda}\,, \quad \frac{dp_\lambda}{dt} = -\frac{\partial H}{\partial q_\lambda} \qquad (\lambda = 1, 2, \dots n)\,. \tag{1}$$

Wird nun nach der „Wahrscheinlichkeit" eines Anfangszustandes $(\overline{q}, \overline{p})$ gefragt, so kann dies wie immer bei „geometrischen Wahrscheinlichkeiten" nur in folgendem Sinne verstanden werden. Wir suchen die Wahrscheinlichkeit dafür, dass unser Anfangszustand in gegebenen Grenzen liegt, d. h. einem vorgeschriebenen „Gebiete" g_0 von „möglichen Anfangszuständen" angehört, still vulnerable to the one basic concern: the hypotheses forming the basis here are supposed to refer to the entire course of motion, while the various stages of the motion are not mutually independent but necessarily connected with one another by virtue of the laws of mechanics (in a way that is, however, not yet known to us), so that it seems legitimate to ask whether an assumption that may be permissible by the standards of the calculus of probabilities for a given instant in time is also permissible for later instants, or whether the system will rather, by necessity and due to its own constitution, tend towards other states in which this presupposition no longer holds. In fact, it seems to me that certain contradictions arising in many a gas-theoretic consideration must also be attributed to a failure to heed such concerns.

Therefore, in order to place the treatment of such problems on a firmer footing, it seems to me indispensable to first establish a particular definition of probability that, while to a certain extent being arbitrary, must neither be altered over the course of the investigation nor amended by new assumptions. Furthermore, it will be necessary to always hold on to *Laplace*'s probability theorem, according to which two events a and b that are necessarily connected with one another so that the occurrence of one of them brings about the occurrence of the other like cause and effect are also always equally probable. We will meet these requirements by defining the probability of some dynamic state that is supposed to occur when the system moves at some time t in terms of the probability of that initial state at time $t = 0$ from which the state under consideration itself must have arisen.

Let us assume that only those potential forces are acting on our system that are supposed to depend solely on the momentary configuration, i. e., on the *position* of the points. Then the state of motion at a given time t, and hence also the entire motion, is completely determined by the system of all coordinates $q_1, q_2, \dots q_n$ and the associated momenta of motion (or impulse coordinates) $p_1, p_2, \dots p_n$, summarily referred to here as (q, p) for short. In the same way, the initial state at time $t = 0$ is determined by the system of the corresponding magnitudes $\overline{q}_\lambda$ and $\overline{p}_\lambda$, or $(\overline{q}, \overline{p})$ for short, while, given the assumptions made, we can write the equations of motion in *Hamilton*ian form:

$$\frac{dq_\lambda}{dt} = \frac{\partial H}{\partial p_\lambda}\,, \quad \frac{dp_\lambda}{dt} = -\frac{\partial H}{\partial q_\lambda} \qquad (\lambda = 1, 2, \dots n)\,. \tag{1}$$

As is invariably the case with "geometric probabilities", any question about the "probability" of an initial state $(\overline{q}, \overline{p})$ can only be understood in the following way. We are trying to find the probability that our initial state lies within specified limits, i. e., belongs to a prescribed "region" g_0 of "possible welche einem Systeme von Ungleichungen genügen

$$\overline{g}_\mu\,(\overline{q}_1, \overline{q}_2, \ldots \overline{q}_n; \overline{p}_1, \overline{p}_2, \ldots \overline{p}_n) \equiv \overline{g}_\mu\,(\overline{q}, \overline{p}) < 0\,. \tag{2}$$

Diese „Wahrscheinlichkeit für das Gebiet g_0“ werden wir nun, da alle Anfangszustände als von einander unabhängig anzusehen sind, zweckmässigerweise, wenn auch nicht ganz ohne Willkür, proportional setzen dem über den Bereich g_0 erstreckten $2n$-fachen Integrale

$$\gamma_0 = \int^{(g_0)} d\overline{q}_1 \ldots d\overline{q}_n d\overline{p}_1 \ldots d\overline{p}_n \equiv \int^{(g_0)} dq\,dp\,,$$

das wir als „die Ausdehnung des Gebietes g_0“ bezeichnen wollen, und den absoluten Wert der Wahrscheinlichkeit erhalten wir dann durch Division in die Ausdehnung $\varGamma_0$ des Bereiches G_0 aller überhaupt in Betracht kommenden Anfangszustände, indem wir es als gewiss ansehen, dass unser Zustand $(\overline{q}, \overline{p})$ dem Gebiete G_0 angehöre. Nun entspricht aber jedem Anfangszustande $(\overline{q}, \overline{p})$ zur Zeit $t = 0$ eine ganze Bewegung, also auch ein ganz bestimmter Zustand $(q, p)_t$ zu einer beliebig vorgeschriebenen Zeit t und somit der Gesamtheit aller Anfangszustände von g_0 ebenfalls eine Gesamtheit von Zuständen $(q, p)_t$,
319 welche wieder ein kontinuierliches Gebiet g_t erfüllen werden, von dem | wir sagen wollen, dass es „in der Zeit t aus dem Gebiete g_0 hervorgehe“, und auch die „Ausdehnung“ γ_t dieses Gebietes g_t wird sich als $2n$-faches Integral bestimmen lassen. In dem hier betrachteten Falle nun, wo die Bewegungsgleichungen sich in der Form (1) schreiben lassen, gilt der Satz von *Liouville*, dass die Ausdehnung des Gebietes g_t der Ausdehnung des entsprechenden Gebietes g_0 gleich, mithin von der Zeit t unabhängig ist:

$$\gamma_t = \int^{(g_t)} dq\,dp = \int^{(g_0)} dq\,dp = \gamma_0\,. \tag{3}$$

Ebenso entspricht auch dem Bereiche G_0 zur Zeit t wieder ein Bereich G_t von der gleichen Ausdehnung $\varGamma_t = \varGamma_0$. Durch den Bruch

$$w_g = \frac{\gamma_0}{\varGamma_0} = \frac{\gamma_t}{\varGamma_t}$$

soll aber die Wahrscheinlichkeit gemessen werden, dass unser Anfangszustand $(\overline{q}, \overline{p})$ dem Gebiete g_0 angehöre, und nach unserem oben aufgestellten Reduktionsprinzipe zugleich auch die Wahrscheinlichkeit dafür, dass der spätere Zustand $(q, p)_t$ dem Gebiete g_t angehöre, denn beide Ereignisse sind notwendig mit einander verbunden. Diese Wahrscheinlichkeit ist also durch die Gebiete g_t und G_t allein bestimmt, gerade als ob wir es nur mit einem Anfangszustande zu thun hätten, und von der Zeit selbst gänzlich unabhängig, sodass wir den Satz haben:

initial states" satisfying a system of inequalities

$$\overline{g}_\mu\left(\overline{q}_1, \overline{q}_2, \ldots \overline{q}_n; \overline{p}_1, \overline{p}_2, \ldots \overline{p}_n\right) \equiv \overline{g}_\mu\left(\overline{q}, \overline{p}\right) < 0 \,. \tag{2}$$

As all initial states are mutually independent, it will be advantageous, albeit somewhat arbitrary, to now set the "probability for the region g_0" proportional to the $2n$-fold integral extended over the domain g_0

$$\gamma_0 = \int^{(g_0)} d\overline{q}_1 \ldots d\overline{q}_n d\overline{p}_1 \ldots d\overline{p}_n \equiv \int^{(g_0)} dq\, dp \,,$$

which we will call "the extension of the region g_0". We then obtain the absolute value of the probability by division by the extension $\mathit{\Gamma}_0$ of the domain G_0 of all initial states that are actually possible by taking it for granted that our state $(\overline{q}, \overline{p})$ belongs to the region G_0. But now to every initial state $(\overline{q}, \overline{p})$ at time $t = 0$ there corresponds an entire motion, and hence also a specific state $(q, p)_t$ at an arbitrarily prescribed time t. Consequently, there corresponds to the totality of all initial states of g_0 also a totality of states $(q, p)_t$ which, in turn, will fill a continuous region g_t, which, as we shall say, "arises in time t from the region g_0", and we will also be able to determine the "extension" γ_t of this region g_t as a $2n$-fold integral. Now, in the case considered here, where it is possible to write the equations of motion in form (1), *Liouville*'s theorem holds according to which the extension of the region g_t is equal to the extension of the corresponding region g_0, and hence independent of time t:

$$\gamma_t = \int^{(g_t)} dq\, dp = \int^{(g_0)} dq\, dp = \gamma_0 \,. \tag{3}$$

Likewise, to the domain G_0 at time t there also corresponds a domain G_t of identical extension $\mathit{\Gamma}_t = \mathit{\Gamma}_0$. But by means of the fraction

$$w_g = \frac{\gamma_0}{\mathit{\Gamma}_0} = \frac{\gamma_t}{\mathit{\Gamma}_t}$$

we want to measure the probability that our initial state $(\overline{q}, \overline{p})$ belongs to the region g_0, and, by the reduction principle presented above, also the probability that the later state $(q, p)_t$ belongs to the region g_t, for both events are necessarily connected with one another. Hence, this probability is solely determined by the regions g_t and G_t, just as if we were only dealing with an initial state, and is itself entirely independent of time, so that we have the theorem:

Satz I. Die Wahrscheinlichkeit für eine gegebene Begrenzung eines Bewegungszustandes ist von der Zeit unabhängig, d. h. stets dieselbe, ob der betrachtete Zustand im Anfange oder in irgend einer anderen Phase der Bewegung eintreten soll.

Den Bereich G_0 aller „möglichen" Anfangszustände, auf den wir unsere Betrachtung beschränken, dessen Wahrscheinlichkeit wir $= 1$ annehmen wollen, wählen wir zweckmässig so, dass er bei der Bewegung unseres Systemes stets *in sich selbst übergeht*, d. h. mit allen seinen späteren Phasen G_t identisch ist. Ein solches Gebiet $G_0 = G_t$, das wir als ein „invariantes Gebiet" bezeichnen wollen, wird immer begrenzt durch Integrale der Bewegungsgleichungen (1), d. h. durch Gleichungen $G(q,p) = \text{const.}$, welche der Bedingung genügen:

$$\frac{dG}{dt} = \sum_{\lambda=1}^{n} \Big(\frac{\partial G}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial G}{\partial p_\lambda} \frac{\partial H}{\partial q} \Big) \equiv (G, H) = 0 \ . \tag{4}$$

So wird z. B. durch das „Integral der lebendigen Kraft" $H = \text{const.}$ eine Reihe von „invarianten Gebieten" $H < c$ oder $a < H < b$ definiert, welche die *Energie* des Systemes nach oben oder unten begrenzen und gelegentlich als Gebiete „aller möglichen" Zustände interpretiert werden können.

Von der Definition der „Wahrscheinlichkeit" gehen wir jetzt über zu der des „mittleren" oder „wahrscheinlichen Wertes". Es sei nämlich $S \equiv S(q,p)$ eine eindeutige und stetige Funktion des „Zustandes", d. h. der $2n$ Variablen q_λ, p_λ, und g ein beliebiges Gebiet von der endlichen Ausdehnung γ, so bezeichnen wir als „den mittleren Wert von S im Gebiete g" den Ausdruck:

$$\overline{S}_g = \frac{1}{\gamma} \int^{(g)} S(q,p) dq\, dq \quad \left(\gamma = \int^{(g)} dq\, dp \right) . \tag{5}$$

Dieser Mittelwert ist eindeutig gegeben durch das vorgeschriebene Gebiet g und wird sich im allgemeinen mit der Zeit stetig ändern, wenn g im Verlaufe der Bewegung vom g_0 in g_t übergeht, und es ist dabei immer

$$\frac{d\overline{S}_{g_t}}{dt} = \frac{1}{\gamma} \int^{(g_t)} \frac{dS}{dt} dq\, dp = \frac{1}{\gamma} \int^{(g_t)} (S, H) dq\, dp \ , \tag{6}$$

weil nach (3) γ ebenso wie das $2n$-fache Differential $d\gamma = dq\, dp$ von t unabhängig ist. Ist aber $g = g_t$ ein „invariantes" Gebiet, z. B. das Gebiet G „aller möglichen" Zustände, so ist auch $\overline{S}_g$ von der Zeit unabhängig, also

$$\gamma \frac{d\overline{S}_g}{dt} = \int^{(G)} \frac{dS}{dt} dq\, dp = \int^{(G)} (S, H) dq\, dp = 0 \ , \tag{6a}$$

und wir haben den Satz:

Theorem I. The probability for a given boundary of a state of motion is independent of time, i. e., always the same, whether the considered state is supposed to occur at the beginning or at some other phase of the motion.

It is useful to choose the domain G_0 of all "possible" initial states to which we restrict our consideration and whose probability we shall take to be $= 1$ so that it *is* always *transformed into itself* when our system moves, i. e., so that it is identical with all its later phases G_t. Such a region $G_0 = G_t$, which we shall call an "invariant region", is always bounded by integrals of the equations of motion (1), i. e., by equations $G(q, p) =$ const. satisfying the condition:[1]

$$\frac{dG}{dt} = \sum_{\lambda=1}^{n} \Big(\frac{\partial G}{\partial q_\lambda} \frac{\partial H}{\partial p_\lambda} - \frac{\partial G}{\partial p_\lambda} \frac{\partial H}{\partial q_\lambda} \Big) \equiv (G, H) = 0 \,. \tag{4}$$

Thus, e. g., the "integral of the living force" $H =$ const. defines a series of "invariant regions" $H < c$ or $a < H < b$ by which the *energy* of the system is bounded upwards or downwards and which can, on occasion, be interpreted as regions of "all possible" states.

We now move from the definition of the "probability" to that of the "mean" or "probable value". For let $S \equiv S(q, p)$ be a single-valued and continuous function of the "state", i. e., of the $2n$ variables q_λ, p_λ, and let g be any region of finite extension γ. Then by "the mean value of S in the region g" we refer to the expression[2]

$$\overline{S}_g = \frac{1}{\gamma} \int^{(g)} S(q, p) dq\, dp \quad \left(\gamma = \int^{(g)} dq\, dp \right) . \tag{5}$$

This mean value is uniquely given by the prescribed region g and, in general, will change continuously with time, if g is transformed from g_0 into g_t over the course of the motion, and we always have

$$\frac{d\overline{S}_{g_t}}{dt} = \frac{1}{\gamma} \int^{(g_t)} \frac{dS}{dt} dq\, dp = \frac{1}{\gamma} \int^{(g_t)} (S, H) dq\, dp \,, \tag{6}$$

since, by (3), γ as well as the $2n$-fold differential $d\gamma = dq\, dp$ is independent of t. If, however, $g = g_t$ is an "invariant" region, e. g., the region G of "all possible" states, then $\overline{S}_g$, too, is independent of time, and hence

$$\gamma \frac{d\overline{S}_g}{dt} = \int^{(G)} \frac{dS}{dt} dq\, dp = \int^{(G)} (S, H) dq\, dp = 0 \,, \tag{6a}$$

and we have the theorem:

[1] ⟦On the right side of the following equation (4), Zermelo erroneously omits three lower indices λ.⟧

[2] ⟦In the middle of the following formula (5), Zermelo erroneously writes "*dqdq*" instead of "*dqdp*".⟧

Satz II. Ist S eine beliebige eindeutige und differentiierbare Funktion des Bewegungszustandes, so hat der Mittelwert von $\frac{dS}{dt}$ in jedem invarianten Gebiete G den Wert Null.

Es kann also in diesem Gebiete die Grösse S ebenso gut zunehmen wie abnehmen, und es kann keine solche Funktion S existieren, die für alle oder auch nur für die überwiegend meisten Zustände eines invarianten Gebietes vermöge der Bewegungsgleichungen (1) beständig zunehmen oder beständig abnehmen würde.

Ein analoger Satz gilt aber auch für nicht invariante Gebiete, nämlich für solche, welche durch vorgeschriebene numerische Werte von S begrenzt werden, d. h. für Gebiete $a < S < b$. Wir beweisen den Satz zunächst für den einfacheren Fall $S < c$, auf den sich der allgemeine zurückführen lässt. Ist nämlich unser Gebiet $g = g_t$ kein invariantes, so wird es in der Zeit τ in ein anderes Gebiet $g' = g_{t+\tau}$ übergehen. Beide Gebiete haben die gleiche Ausdehnung

320 |
$$\gamma = \int^{(g)} dq\, dp = \int^{(g')} dq\, dp$$

und haben für kleine Werte von τ ein Stück g'' mit einander gemein, während die Ausdehnungen der Restgebiete $g - g''$ und $g' - g''$ mit τ gleichzeitig verschwinden. Daher wird die Differenz der beiden Integrale

$$\begin{aligned}
\gamma\,(S_{g'} - S_g) &= \int^{(g')} S\, dq\, dp - \int^{(g)} S\, dq\, dp \\
&= \int^{(g')} (S-c) dq\, dp - \int^{(g)} (S-c) dq\, dp \\
&= \int^{(g'-g'')} (S-c) dq\, dp - \int^{(g-g'')} (S-c) dq\, dp\,,
\end{aligned}$$

wo c eine beliebige Konstante sein kann. Ist aber das Gebiet g definiert durch die Ungleichung $S < c$, so werden in den beiden schmalen Randgebieten $g' - g''$ und $g - g''$ die Werte von S nur wenig von c verschieden sein und die beiden Integrale über $S - c$ von der Ordnung τ^2 verschwinden, so dass wir haben, wie behauptet:

$$\underset{\tau=0}{Lim}\, \gamma \frac{\left(S_{g'} - \overline{S}_g\right)}{\tau} = \gamma \frac{dS_g}{dt} = \int^{(g)} \frac{dS}{dt} dq\, dp = 0\,. \tag{7}$$

Ferner ist dasselbe Integral $\int \frac{dS}{dt} dq\, dp$, erstreckt über ein Gebiet $a < S < b$ nichts anderes als die Differenz der beiden analogen Integrale über die Gebiete

Theorem II. If S is any single-valued and differentiable function of the state of motion, then the mean value of $\frac{dS}{dt}$ takes the value zero in every invariant region G.

Hence, in this region, the magnitude S may just as well increase as decrease and there can exist no function S that continuously increases or continuously decreases for all or at least the great majority of states of an invariant region by dint of the equations of motion (1).

However, an analogous theorem also holds for regions that are not invariant, namely for those bounded by prescribed numerical values of S, i. e., for regions $a < S < b$. We first prove the theorem for the simpler case $S < c$, to which the general case can be reduced. For if our region $g = g_t$ is not invariant, then it will be transformed into a different region $g' = g_{t+\tau}$ during time τ. Both regions have the same extension

$$\gamma = \int^{(g)} dq\,dp = \int^{(g')} dq\,dp$$

and share a segment g'' for small values of τ, while the extensions of the remainder regions $g - g''$ and $g' - g''$ vanish along with τ. Hence, the difference of the two integrals becomes[3]

$$\begin{aligned}\gamma\left(\overline{S}_{g'} - \overline{S}_g\right) &= \int^{(g')} S\,dq\,dp - \int^{(g)} S\,dq\,dp \\ &= \int^{(g')} (S-c)dq\,dp - \int^{(g)} (S-c)dq\,dp \\ &= \int^{(g'-g'')} (S-c)dq\,dp - \int^{(g-g'')} (S-c)dq\,dp\,,\end{aligned}$$

where c may be any constant. But if the region g is defined by the inequality $S < c$, then, in the two narrow boundary regions $g' - g''$ and $g - g''$, the values of S will differ only slightly from c and the two integrals over $S - c$ will vanish with order τ^2 so that we have, as stated,

$$\underset{\tau=0}{Lim}\,\gamma\frac{\left(\overline{S}_{g'} - \overline{S}_g\right)}{\tau} = \gamma\frac{d\overline{S}_g}{dt} = \int^{(g)} \frac{dS}{dt}dq\,dp = 0\,. \tag{7}$$

Furthermore, the same integral $\int \frac{dS}{dt}dq\,dp$, extended over a region $a < S < b$, is but the difference of the two analogous integrals over the regions $S < b$

[3] ⟦In the following formula and in formula (7) below, Zermelo erroneously writes "$S_{g'}$" and "S_g" for "$\overline{S}_{g'}$" and "$\overline{S}_g$", respectively.⟧

$S < b$ und $S < a$ und muss daher gleichfalls verschwinden, auch wenn a und b sich beliebig wenig unterscheiden. So haben wir:

Satz III. Ist uns der Wert S_0 einer eindeutigen Funktion S des Bewegungszustandes mit beliebiger Annäherung σ vorgeschrieben, so hat in dem dadurch definierten Gebiete $S_0 - \sigma < S < S_0 + \sigma$ die mittlere Zunahme $\frac{dS}{dt}$ derselben Funktion den Wert Null.

Die hier entwickelten Sätze könnten als Grundlage für weitere Betrachtungen dienen, indem sie uns lehren, welche Wahrscheinlichkeiten und Mittelwerte wir ganz allgemein durch andere, einfachere ersetzen dürfen. Um nun aber weitere Gleichungen oder Differentialgleichungen zu finden, die zu einer wirklichen Bestimmung der Mittelwerte als Funktionen der Zeit und zu einer Beschreibung des physikalischen Vorganges führen können, wird man diese allgemeinen Betrachtungen verlassen und zu specielleren Annahmen übergehen müssen. Als ein Beispiel hierfür beabsichtige ich, demnächst für das sog. *Maxwell*'sche Gesetz von der Geschwindigkeitsverteilung der Gasmoleküle eine neue Ableitung zu versuchen, die ohne jede weitere Hypothese allein auf der hier gegebenen Definition der Wahrscheinlichkeit beruhen soll.

(Eingegangen 24. März 1900.)

and $S < a$, and hence must also vanish, however small the difference between a and b may be. We thus have:

Theorem III. If the value S_0 of a single-valued function S of the state of motion is prescribed with arbitrary approximation σ, then, in the region $S_0 - \sigma < S < S_0 + \sigma$ thereby defined, the mean increase $\frac{dS}{dt}$ of the same function has value zero.

The theorems developed here may serve as a basis for further considerations by showing us which probabilities and mean values we are permitted to generally replace by other, simpler ones. However, in order to find further equations or differential equations that can lead to an actual determination of the mean values as functions of time and to a description of the physical process we will have to move from these general considerations to more particular assumptions. To illustrate this by an example, I intend to attempt a new derivation of the so-called *Maxwell* law of the velocity distribution of gas molecules soon which is supposed to be based solely on the definition of probability given here without any further hypothesis.

(Received on March 24, 1900.)