The principle of identicality and the foundations of quantum theory. II. The role of identicality in the formation of quantum theory

Peter D. Pešić
St. John's College, Santa Fe, New Mexico 87501

(Received 4 February 1988; revised 11 March 1991; accepted for publication 3 April 1991)

The place of the concept of indistinguishability in the formation of quantum theory is examined, including particularly Planck's earliest formulations of the quantum hypothesis, as well as the use of this concept in the developments surrounding the new quantum theory.

I. INTRODUCTION

We define the principle of identicality as the hypothesis that each fundamental class of physical states is composed of completely indistinguishable entities which can be characterized by certain exactly equal observable parameters (such as mass or charge). In these two papers on the development and significance of this principle, the first paper (I) has considered the development of the concept of "entirely similar particles" in the course of Gibbs' consideration of the entropy of interdiffusing gases.[1] This paper addresses the significance of this concept in the formation of quantum theory. Planck's early work contains an important reconsideration of Gibbs' work and involves an implicit assumption of the identicality of energy states in a cavity, both in the sense of the physical equality and the absolute indistinguishability of those states. By 1924–1927 there were many examples of the invocation of identicality as an fundamental premise both before and after the formulation of the new quantum mechanics. In retrospect, considerations of identicality seem to have played a central role in the development of quantum theory.

II. THE ROLE OF IDENTICALITY IN PLANCK'S WORK

Looking to the writings of Planck, it is important that considerations of entropy were crucial to the novel way he was led to his *ansatz* for the blackbody energy spectrum.[2] Thermodynamics early became Planck's field of specialization; he admired its absolute and universal character for he "had always regarded the search for the absolute as the loftiest goal of all scientific activity."[3] In the preface to the first edition (1897) of his *Treatise on Thermodynamics*, he emphasizes that his approach differs from earlier writers in that "it does not advance the mechanical theory of heat, but, keeping aloof from definite assumptions as to its nature, starts direct from a few very general empirical facts, mainly the two fundamental principles of Thermodynamics."[4] His preface concludes with the hope that, through this thermodynamic approach, a new and "uniform theory of nature, on a mechanical basis or otherwise" might be attained. Planck's work on blackbody radiation moved to realize just this hope, even to its delicate premonition that the new mechanics might not be so "mechanical" as the old. It is also worth noting that this devotion to thermodynamics along with an aloofness to particular mechanical assumptions stood him in good stead, for it freed him from the paralyzing effect of what Lord Kelvin called (in 1900) the "Maxwell–Boltzmann doctrine of [equi]partition of energy" which Kelvin himself considered a "cloud" on the "beauty and clearness of the dynamical theory, which asserts heat and light to be modes of motion."[5]

It is with special interest, then, that one turns to his discussion of the so-called Gibbs paradox (which he, like Gibbs, never refers to by either of those terms). After remarking, as Gibbs had, that "the increase of the entropy does not depend on whether the gases are chemically alike or not," he notes that "by making the two gases the same, there is evidently no increase of the entropy, since no change of state ensues." Then follows a remarkable conclusion:

> It follows that the chemical difference of two gases, or, in general, of two substances, cannot be represented by a continuous variable; but that here we can speak only of a discontinuous relation, either of equality or inequality.[6]

So Planck concludes that each chemical species must be characterized by discrete quantities in order that it have a distinct identity as such. This may be one of the first expressions of a fundamental discontinuity indicated by considerations of entropy. He adds that such discontinuity is characteristic of *chemical* changes, whereas physical properties "may always be regarded as continuous." Planck allowed this remark to remain in the book through its 1926 English translation. This may reflect his ambivalence to the discontinuities which he had postulated.[7] In contrast, Gibbs held that the hypothesis of "entirely similar particles" is one that cuts to the foundation of atomic theory.[8]

These considerations and Planck's later introduction of a novel counting of the modes of black body radiation turn on the issues of (1) equality and (2) indistinguishability, which together comprise what we will call *identicality*. It should be noted that Boltzmann's view contained (1) but not (2); his particles are perfectly equal in their innate properties but still distinguishable in the sense of each having a distinct trajectory. In the above quotation, Planck only adds to this the observation that such perfect equality leads to a fundamental discreteness in the physical parameters of the particles, in the spirit of Gibbs' argument. But the core of the notion of identicality lies in (2) which leads to the necessity of a "sum over histories"; it is when he uses such indistinguishability that Planck really reaches the realm of the quantum.

There may be no other theoretical issue in which the *self-consistency* of classical mechanics is so directly at issue. For instance, the question of the stability of atoms against energy loss through radiation requires the empirical knowledge that atoms are, indeed, stable. Often descriptions of quantum theory rely on experimental discoveries (such as the photoelectric effect, discrete atomic spectra, specific heats, etc.) in a way which implies that there was not *internal* difficulty with classical theory.

The exception would seem to be the "ultraviolet catastrophe" of the classical Rayleigh–Jeans blackbody spectrum, for there an unphysical divergence at high frequency emerges from classical assumptions even before one embarks on comparison with experimental results. However, Planck's derivation of the blackbody spectrum rests on the question of the correct counting of the modes of vibration. M. J. Klein, T. S. Kuhn, and O. Darrigol have discussed in detail the way in which Planck at first rejected and then began to use Boltzmann's statistical approach.[9] In his earlier work, Planck had "regarded the principle of the increase of entropy as no less immutably valid than the principle of the conservation of energy itself, whereas Boltzmann treated the former merely as a law of probabilities—in other words, as a principle that could admit of exceptions."[10] But through Boltzmann's criticisms in 1897 Planck was forced to abandon his original plan to derive the irreversible approach to equilibrium in the cavity from reversible electrodynamics. It was at that point that he adopted Boltzmann's statistical approach; this statistical character marked his original formulation of the quantum hypothesis and every succeeding development of it, as Darrigol has noted.[11]

Boltzmann had in 1872 already introduced discrete "energy elements" in order to give a convenient solution to the problem of the entropy of a gas in terms of the probability of individual energy states of the molecules, or "complexions," as he called them.[12] He refers to this procedure as a "fiction" and takes the limit to continuous energies at the end of his calculation. Planck followed this path to his quantum hypothesis, only he did not regard it as a "fiction" and did not allow $h \to 0$ at the end. He also followed Boltzmann's basic approach of counting the complexions that make up the state of the N resonators that he hypothesizes to probe the state of the radiation field, including the assumption that any particular complexion is as likely as any other. But here Planck diverges from Boltzmann, for in his counting Boltzmann distinguishes the individual atoms whereas Planck's counting does not distinguish the resonators. At this point Planck uses for the first time the *indistinguishability*, which is the core of the principle of identicality and leads to the combinatorial factors essential to the final form of the blackbody distribution. Indeed, Planck's resonators are imaginary entities, as A. Needell and Kuhn have pointed out, and so their indistinguishability really is part of their definition.[13] Nevertheless, even though identicality enters here through the imaginary resonators, the new counting appropriate to them remained a part of the theory even when it turned to consider real atoms. In that sense the resonators formed a bridge from mathematical energy elements to fully identical atoms and light quanta and brought to light the novel counting appropriate for such identical beings. A. Pais has pointed out that even in 1901 Planck realized that this novel counting was a hypothesis which requires turning to experience for proof; Darrigol has emphasized that the combinatorics "were not a provisional artifice to be replaced later by a more detailed specification of the dynamics."[14] This is borne out by Planck's own statement that "since elementary disorder and the lack of any detailed control belongs to the essence of entropy, only combinatorial or probabilistic considerations can provide the necessary basis for its calculation."[15] That is, Planck realized quite early that the counting itself and its attendant probabilities was at the heart of the quantum hypothesis; in so doing he grasped the principle of identicality as it was made manifest in blackbody radiation.

Thus the "ultraviolet catastrophe" rests on the issue of identicality. Klein has also argued that Planck attached no special significance to Rayleigh's brief paper of 1900 which showed the classical distribution law for the blackbody, and which included Rayleigh's speculation concerning the failure of equipartition.[16] In explanation, Klein adduces Planck's lack of sympathy for "the Boltzmann–Maxwell doctrine" of which Rayleigh speaks. All this speaks against the account sometimes found in textbooks which describes Planck's hypothesis as a response to the "ultraviolet catastrophe." As Klein has noted, it was Ehrenfest who much later coined this term (1911) and spoke of its implications.[17]

By 1911, Planck found further confirmation that the hypothesis of quanta is deeply connected with the fundamental nature of the entropy. The introduction of h gives a clear and unambiguous elementary volume of phase space without which statistical mechanics has an unavoidable arbitrariness. In Nernst's theorem Planck found a definitive indication that only with the quantum hypothesis does the entropy of a system have such a definite, finite value. "For the present," he wrote, "I would consider this proposition as the very quintessence of the hypothesis of quanta."[18] Gibbs had noted that the adoption of generic phase is only necessitated when the number of particles can vary; otherwise the difference between the specific and generic phase would be absorbed into "the arbitrary constant of integration which is added to entropy." Though Gibbs was not disturbed by the arbitrariness he considered inherent in entropy, for Planck it was another case in which "the search for the absolute" guided him toward the quantum.[19]

A helpful overview of Planck's understanding of indistinguishability is found in lectures which he gave at Columbia University in 1909. In surveying theoretical physics his concern is with "the elimination of the individuality of the particular physicist and therefore with the production of a common system of physics for all physicists."[20] In accord with this he unfolds the development of thermodynamics emphasizing the way in which Boltzmann eliminated all "anthropomorphic elements" (such as notions of "usable work") in the statistical definition of entropy. He then argues that this concept of entropy as probability necessitates counting "a finite number of a priori equally likely configurations (complexions) through each of which the state considered may be realized." These must be

> numerous discrete homogeneous elements—for in perfectly continuous systems there exist no reckonable elements—and hereby the atomistic view is made a fundamental requirement. We have, therefore, to regard all bodies in nature, in so far as they possess an entropy, as constituted of atoms ...[21]

Then the application of entropy to the case of radiation implies an "atomic conception" in which "certain energy elements play an essential role." These are the quanta. Further, these energy elements

> must actually be of the same kind, or they must at least form a number of groups of like kind, e.g., constitute a mixture in which each kind of element occurs in large numbers. For only through the similarity of the elements does it come about that order and law can result in the larger from the smaller. If the molecules of a gas be all different from one another, the properties of a gas can

never show so simple a law-abiding behavior as that which is indicated by thermodynamics.[22]

The last sentence shows that Planck's sense of the "sameness" of the energy elements in blackbody radiation carries over into his sense of the "sameness" of molecules. Here he stands on the verge of the quantum statistics later elaborated by Bose and Einstein. Pais has noted that L. Natanson (1911) "was the first to state that distinguishability has to be abandoned in order to arrive at Planck's law" and Klein has discussed Ehrenfest's seminal work dating from that year on the lack of independence of particles.[23] However, Planck's arguments indicate that he himself may have been aware of this essential point much earlier, since they underlie the novel way in which he had counted the complexions of blackbody radiation in 1900.

III. IDENTICALITY AND THE DEVELOPMENT OF THE QUANTUM THEORY

Klein has pointed out that most of Gibbs' contemporaries, specifically including Ehrenfest, did not appreciate the significance of the indistinguishability that Gibbs introduced. Nevertheless, Ehrenfest found himself confronting just this matter already in 1914 in his rederivation of the Planck radiation formula. As Klein points out, Ehrenfest realized that "if one is to describe radiation by a particle representation, then the "particles" one uses must have properties fundamentally different from those of any particles previously used in physical theory ... [they] must also have lost that most basic of properties, their individuality ... " This, in turn, Klein connects to the "fusion" of wave and particle properties which Einstein suggested in his 1909 studies of the fluctuations in black-body radiation.[24] Klein has also discussed how the work of Ehrenfest and Trkal (1920) on the Gibbs paradox was an important step towards Ehrenfest's full realization that the identicality of particles does not only consist in their complete equivalence (which was already present in classical theory) but even more in their lack of *independence*.[25]

In de Broglie's celebrated thesis (1924) one also notices this theme of indistinguishability and the attendant lack of independence in the new context of the phase wave he identifies with the atoms. Indeed, it becomes more explicable how this can result through what de Broglie himself calls "a new hypothesis":

> If two or several atoms have perfectly identical phase waves, so that one can say that they are carried by the same wave, their movement can no longer be considered entirely independent and these atoms can no longer be treated as distinct unities in probability calculations. ... their identity of structure dispenses us from taking their individuality into account.[26]

Darrigol has discussed how de Broglie's theory helped show that indistinguishable particles could not have a track or a history by connecting those particles with a phase wave; it was exactly the tracklessness of the waves that clarified this quality.[27] One might say that the wave–particle duality emerged here as a response to the need to connect together the different facets of identicality that had emerged.

Bose's rederivation of Planck's results emphasized the central significance of the correct counting of identical states. Although Bose himself said he had not known in 1924 that his new statistics involved "something which was really different from what Boltzmann would have done," he yet felt that it was "evident." Darrigol has suggested that Bose was following the counting that Planck had applied to distribute energy elements over his imaginary resonators but now Bose applied to distribute light-quanta over cells; if so, this would show the further way in which Planck's resonators were a bridge to complete identicality that we suggested above.[28] By 1925 Einstein described the premises underlying what he called the "simplicity" of the strange ways of counting used by Planck and by Bose. Einstein notes that *particles* are no longer statistically independent; Bose's *ansatz* requires only statistical independence of the *cells* in phase space. In Einstein's words, this expresses "a certain hypothesis on a mutual influence of the molecules which for the time being is of a quite mysterious nature." Here Einstein realizes the interdependence that results from complete identicality and which entails such phenomena as the Bose–Einstein condensation. Einstein goes on to use the partitioning of indistinguishable particles in establishing his theory of the quantum gas. It is important to Einstein that his quantum gas (unlike a Boltzmann gas) does satisfy Nernst's theorem, just as it had been for Planck in the case of the "radiation gas."[29] P. A. Hanle and also Darrigol have discussed Einstein's use in 1925 of a paradox very similar to that of Gibbs in which the entropy of mixing different gases is reconciled with the entropy of a single gas by means of de Broglie's waves.[30]

Indeed, that version of the paradox had already been considered in 1921 by Schrödinger who entered into a controversy which followed on the work of Ehrenfest and Trkal mentioned above. Planck had rejoined with a defense of an absolute definition of entropy (rather than the reliance only on entropy differences used by Ehrenfest and Trkal). As Hanle has discussed, Schrödinger was seeking to defend and deepen the results of Ehrenfest and Trkal.[31] In 1925 he did so by using the classical counting of Maxwell and Boltzmann to enumerate the modes of de Broglie waves and thereby to arrive at the statistics of Bose and Einstein without using their novel counting procedures. In this result can be seen the way in which consideration of the de Broglie waves could be seen as the missing ingredient that connects the classical counting of Maxwell and Boltzmann to the new statistics of Bose and Einstein. This general equivalence was also independently shown by Jordan, as Darrigol has pointed out.[32]

Thus important use was made of the principle of identicality well before the full elaboration of matrix and wave mechanics. Pais has noted that the new counting based on indistinguishability leads to quantum statistics without any reference to symmetries of wave functions or even to Planck's constant. Such an account, suggests Pais, "does facilitate the understanding of the founders' reasonings ... ."[33] In November 1925 Bohr, Heisenberg, and Jordan noted that the newly-born quantum mechanics implemented Bose–Einstein statistics by eliminating any possibility of distinguishing photons. In the spring of 1926 Heisenberg wrote that "it is a characteristic trait of atomic systems that their constituent parts, the electrons, be equal and subject to equal forces." This he formulates in the symmetry of the Hamiltonian as a function of the constituent coordinates and parameters. The same year marks the publications of Dirac and Fermi on the statistics that bear their names.

By 1925 Pauli had advanced his exclusion principle in its final form, and by 1927 Heitler and London had elaborated their theory of the exchange forces responsible for homopolar valence based on the identicality of the constituents.

Thus the notion of identical particles yielded rich fruit even before the new quantum theory was fully formulated. But the fact that all these developments came so closely together leaves the fundamental question: What is the relation between identicality and the rest of the quantum theory?

IV. IDENTICALITY AND THE FOUNDATIONS OF QUANTUM THEORY

The view of this question which has prevailed until now seems to be that indistinguishability is a consequence of other assumptions which themselves constitute the fundamental postulates of quantum theory. For example, Landau and Lifshitz argue that "by virtue of the uncertainty principle, the concept of the path of an electron ceases to have any meaning" and therefore the impossibility of distinguishing an electron follows.[34] Though his discussion of the amplitude turns on the question of the distinguishability of final and intermediate states, Feynman considers identicality a "beautiful consequence" of quantum theory.[35] Weisskopf has notably emphasized identicality as a crucial result of quantum theory, but seems also to think it a consequence rather than a fundamental principle.[36] However, Yukawa has written that the "invisible mold" from which emerge identical electrons "is a manifestation, in its most fundamental form of the rule that prevails in the natural world."[37] Thus the place of the principle of identicality in the foundations of quantum theory remains a fundamental question for further consideration.

ACKNOWLEDGMENTS

This paper was stimulated by discussions in the study group on quantum theory at St. John's College in Santa Fe. I thank my colleagues P. Le Cuyer, J. Steadman, R. Swentzell, and H. von Briesen for helpful discussions. I also thank E. M. Purcell for his encouraging suggestions as well as M. J. Klein and P. Le Cuyer for their comments on the manuscript. I wish also to thank St. John's College and the Alfred P. Sloan Foundation for their support.

[1] P. D. Pešić. "The principle of identicality and the foundations of quantum theory I. The Gibbs paradox," Am. J. Phys. **59**, 971–974 (1991).

[2] See the invaluble treatments of Planck's work by M. J. Klein in "Max Planck and the beginnings of the quantum theory," Arch. Hist. Exact Sci. **1**, 459–479 (1960); "Planck, entropy, and quanta, 1901–1906," Natural Philos. **1**, 81–108 (1963), and "Thermodynamics and quanta in Planck's work," Phys. Today **19**(11), 23–32 (1966); a brief general account of the history is given by M. Jammer, *The Conceptual Development of Quantum Mechanics* (McGraw-Hall, New York, 1966), pp. 10–28.

[3] M. Planck, *Scientific Autobiography* (Philosophical Library, New York, 1949), pp. 34–35.

[4] M. Planck, *Treatise on Thermodynamics* (Dover, New York), p. viii.

[5] Lord Kelvin, "Nineteenth century clouds over the dynamical theory of heat and light," Philos. Mag. **2**, 1–40 (1901); see also M. J. Klein, *Paul Ehrenfest* (North-Holland, Amsterdam, 1972), Vol. 1, p. 110.

[6] M. Planck, Ref. 4. p. 222.

[7] E. Schrödinger, *What is Life?* (Doubleday, New York, 1958), pp. 132–160.

[8] See A. Haas and P. S. Epstein, in A. Haas, *A Commentary on the Scientific Works of J. Willard Gibbs* (Yale U.P., New Haven, 1936), Vol. 2, pp. 35–38, 172–178, 287–296.

[9] M. J. Klein, Ref. 2; T. S. Kuhn, *Black-Body Theory and the Quantum Discontinuity, 1894-1912* (University of Chicago, Chicago, 1987), pp. 57–60, 98, 101, 105; O. Darrigol, "Statistics and combinatorics in early quantum theory," Hist. Stud. Phys. Sci. **19**, 17–80 (1988).

[10] M. Planck, Ref. 3. p. 32.

[11] See O. Darrigol, Ref. 9, p. 53.

[12] L. Boltzmann, "Über die Beziehung zwischen den zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht," Wien. Ber. **76**, 373–435 (1877) [reprinted in L. Boltzmann, *Wissenschaftliche Abhandlungen* (Chelsea, New York, 1968), Vol. 2, pp. 164–223] and L. Boltzmann, "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen," S.-B. Akad. Wiss. Wien PT II **66**, 275–370 (1872) (*Wissenschaftlichen Abhandlungen*, Vol. 1, pp. 316–402), translated in S. G. Brush, *Kinetic Theory* (Pergamon, Oxford, 1966), Vol. 2, pp. 88–175. See also M. J. Klein's commentary in Ref. 2 and A. Kastler's valuble historical survey, "On the historical development of the indistinguishability concept for microparticles," in *Old and New Questions in Physics, Cosmology, and Theoretical Biology*, edited by A. van der Meerwe (Plenum, New York, 1983), pp. 607–623 as well as A. Pais, *'Subtle is the Lord ...'* (Oxford U.P., Oxford, 1982), p. 371.

[13] See T. S. Kuhn, Ref. 9, p. 117; A. A. Needell in M. Planck, *The Theory of Heat Radiation* (Tomash-AIP, New York, 1988), pp. xv–xvi.

[14] See Ref. 11.

[15] See M. Planck, Ref. 13, p. 382 (first edition).

[16] Lord Rayleigh, *Scientific Papers* (Cambridge U.P., Cambridge, 1903). Vol. 4, pp. 483–485.

[17] M. J. Klein, Ref. 5, pp. 249–250.

[18] M. Planck, Ref. 13, p. 6 (second edition): see also Klein's commentary in Ref. 2.

[19] A. Kastler, Ref. 12, points out that Boltzmann particles also obey Nernst's theorem, contrary to the calculations of both Planck and Einstein. This seems to be a case in which equality of particles is required but not true indistinguishability.

[20] M. Planck, *Eight Lectures on Theoretical Physics*, translated by A. P. Wills (Columbia U.P., New York, 1915), p. 7.

[21] M. Planck, Ref. 20, p. 45.

[22] M. Planck, Ref. 20, p. 52.

[23] A. Pais, *Inward Bound* (Oxford U.P., Oxford, 1986), p. 283; M. J. Klein, "Ehrenfest's contributions to the development of quantum statistics," Proc. Acad. Amsterdam B **62**, 41–62 (1959); A. Kastler, Ref. 12. See also J. Mehra and H. Rechenberg, *The Historical Development of Quantum Theory* (Springer-Verlag, New York, 1982), Vol. 1, Pt 2, pp. 613–619.

[24] M. J. Klein, Ref. 5, pp. 136–137, 256–257. Further historical background is recounted by M. Jammer, Ref. 2, pp. 338–345.

[25] See M. J. Klein, Ref. 23, and the brief discussion in paper (I).

[26] L. de Broglie, *Recherches sur la théorie des quanta* (Masson, Paris, 1963), pp. 103–104, 98–99; see O. Darrigol, "The origin of quantized matter waves," Hist. Stud. Phys. Sci. **16**, 197–253 (1986).

[27] O. Darrigol, Ref. 26, pp. 205–209; P. A. Hanle, "The Schrödinger–Einstein correspondence and the sources of wave mechanics," Am. J. Phys. **47**, 644–648 (1979).

[28] O. Darrigol, Ref. 26, pp. 209–213, as well as his forthcoming paper on the early history of indistinguishability in Hist. Stud. Phys. Sci.

[29] A. Pais, Ref. 12, pp. 423–434.

[30] P. A. Hanle, Ref. 27; O. Darrigol, Ref. 27.

[31] P. A. Hanle, "The coming of age of Erwin Schrödinger: His quantum statistics of ideal gases," Arch. Hist. Exact Sci. **17**, 165–191 (1977). See also J. Mehra and H. Rechenberg, Ref. 23, Vol. 5, pp. 361–366 and W. T. Scott, *Erwin Schrödinger* (University of Massachusetts Press, Amherst, MA, 1967), pp. 22–26.

[32] O. Darrigol, Ref. 26, pp. 217–229.

[33] A. Pais, Ref. 20, pp. 280–285.

[34] L. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Addison-Wesley, Reading, MA, 1958), p. 204.

[35] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures in Physics* (Addison-Wesley, Reading, MA, 1965), Vol. 3, pp. 1–10.

[36] V. F. Weisskopf, *Physics in the Twentieth Century* (MIT, Cambridge, MA, 1972), pp. 24–51, 295–297.

[37] H. Yukawa, *Creativity and Intuition* (Kodansha, Tokyo, 1973), pp. 72–74. See also H. Margenau, Philos. Sci. **11**, 187 (1944); H. Margenau, *The Nature of Physical Reality* (McGraw-Hill, New York, 1950), pp. 427–447; E. Cassirer, *Determinism and Indeterminism in Modern Physics* (Yale U.P., New Haven, 1956), p. 184 ff; N. R. Hanson, *Patterns of Discovery* (Cambridge U.P., Cambridge, 1965), pp. 128–135; A. S. Eddington, *The Philosophy of Physical Science* (University of Michigan, Ann Arbor, MI, 1958), pp. 128–129.

Linearization of the simple pendulum

L. H. Cadwell
Department of Engineering-Physics Systems, Providence College, Providence, Rhode Island 02918
E. R. Boyko[a)]
Department of Chemistry, Providence College, Providence, Rhode Island 02918

(Received 25 September 1989; accepted for publication 8 April 1991)

A new method is described to obtain the leading correction to the simple harmonic approximation of a simple pendulum for large-angle motion using an analytical and geometric approach.

I. INTRODUCTION

The simple pendulum provides an excellent opportunity for students to use approximation concepts to study nonlinear motion associated with large-angle deflections. Many introductory texts often ignore the nonlinear motion, while others refer to the variation in the period for arbitrary amplitudes[1,2] with no explanation. Advanced texts[3–5] and recent studies[6–8] present solutions for the period using elliptic integrals and perturbation theory or use a small computer for theoretical analysis. We present a different approach: a search for a geometric method to obtain the leading correction to the simple harmonic approximation of a simple pendulum.

One way to treat the simple pendulum problem is to linearize the equation of motion

$$\frac{d^2\Theta}{dt^2} = -\left(\frac{g}{L}\right)\sin\Theta,$$

where Θ is the angular displacement, L is the length of the pendulum, and g is the acceleration due to gravity. Using the Maclaurin series expansion of $\sin\Theta$ and keeping the first two terms gives

$$\frac{d^2\Theta}{dt^2} = -\left(\frac{g}{L}\right)\left(\Theta - \frac{\Theta^3}{6} + \cdots\right). \qquad (1)$$

Simply neglecting the Θ^3 term would clearly produce linearization, but this would be a poor procedure except for small angles since Θ and Θ^3 are highly correlated. Introducing the variable x,

$$x = \Theta/\Theta_0,$$

where Θ_0 is the amplitude, Eq. (1) can be rewritten as

$$\frac{d^2x}{dt^2} = -\left(\frac{g}{L}\right)\left[x - \left(\frac{\Theta_0^2}{6}\right)x^3 + \cdots\right].$$

A linear differential equation can be obtained by replacing the x^3 term with the approximation

$$x^3 \cong kx, \qquad (2)$$

where k is a constant between 0 and 1. The period of the pendulum, T, can then be written as

$$T = T_0(1 - k\Theta_0^2/6)^{-1/2}, \qquad (3)$$

where $T_0 = 2\pi(L/g)^{1/2}$. Upon expanding Eq. (3) we have

$$T = T_0(1 + k\Theta_0^2/12 + \cdots).$$

The formal expansion of the exact solution for the period is

$$T = T_0(1 + \Theta_0^2/16 + \cdots),$$

which shows that the correct value for k in Eq. (2) is 3/4.

We will consider alternative approaches to evaluate k both graphically and analytically using a variety of approximation criteria. The graphical approach should allow an extension of the type of simple approximation described by Ganley.[9]

II. DETERMINATION OF k

Figure 1 shows a plot of x^3 vs x. Dropping the Θ^3 term in Eq. (1) is equivalent to replacing x^3 by the line $y = 0$. A better line is found by adjusting the slope to any positive value, k. One approach that a first-year physics student might consider who has not had experience with the theory of approximations is to adjust the slope so that the four shaded areas between the straight line and the cubic curve shown in Fig. 1 are equal. This "equal-areas criterion" simply requires that

$$\int_0^1 (kx - x^3)dx = 0. \qquad (4)$$

Solving Eq. (4) gives $k = 1/2$. Comparing the slope with the theoretical value $k = 3/4$, this criterion might be called a fair, but not good, approximation.

A second approach which might be suggested by an advanced student who has had experience fitting data might suggest the least-squares criterion. This method of approximation determines the best (least-squares) line by requiring that the sum of the squares of the discrepancies between the line and the experimental data be a minimum. Here, k is chosen so that the integral

$$\int_{-1}^{1} (kx - x^3)^2\, dx \qquad (5)$$