On Physics, Physicists and Philosophy

N. Mukunda
Foreword

The Masterclass series of eBooks brings together pedagogical articles on single broad topics taken from Resonance, the Journal of Science Education, that has been published monthly by the Indian Academy of Sciences since January 1996. Primarily directed at students and teachers at the undergraduate level, the journal has brought out a wide spectrum of articles in a range of scientific disciplines. Articles in the journal are written in a style that makes them accessible to readers from diverse backgrounds, and in addition, they provide a useful source of instruction that is not always available in textbooks.

The fifth book in the series, ‘On Physics, Physicists and Philosophy’, is by Prof. N. Mukunda. A very distinguished theoretical physicist, Prof. Mukunda worked at TIFR (Mumbai) for a few years, and then at the Indian Institute of Science (Bengaluru) till 2001. He still remains actively engaged with science and regularly lectures to students around the country. Prof. Mukunda was the founding Chief Editor of Resonance, a journal that he steered with a very deft though unobtrusive hand during the initial years (1996–2000). I had the pleasure of serving on the Editorial Board of Resonance as a young assistant professor in those years, and I learnt a lot just by observing how Prof. Mukunda handled the trickiest of issues with firmness, underlain by an unflappable good-natured aplomb. Prof. Mukunda was also a very regular contributor to Resonance over the years, as the sheer number of articles in this collection testifies. His engagement with pedagogical and educational issues has been widely recognized in India. He was instrumental in establishing and nurturing the science education initiatives and programmes of the three science academies in the country, and also served a long stint as Editor of Publications at the Indian Academy of Sciences, Bengaluru.

Prof. Mukunda is very well known for his research contributions to the areas of classical and quantum mechanics, theoretical optics and mathematical physics. In addition, he has had a deep and abiding interest in philosophical issues in physics and, indeed, in science in general. The book, which will be available in digital format, and will be housed as always on the Academy website, will be valuable to both students and experts as a useful handbook on all kinds of topics in physics, ranging from technical and philosophical issues to sketches of the lives and work of eminent physicists, all rendered in Prof. Mukunda’s inimitable style.

Amitabh Joshi
Editor of Publications
Indian Academy of Sciences
February 2018
About the Author

Professor Mukunda is a well-known theoretical physicist who has made many outstanding contributions to the areas of high energy physics, group theory, classical dynamics, classical and quantum optics and geometric phases in quantum mechanics. He is known for developing and applying group theoretic and phase space based methods to problems in theoretical physics. He worked with Prof E C G Sudarshan for his PhD (1964) at the University of Rochester USA, began his independent scientific career in TIFR Mumbai in 1967, and later (1972) moved to the Center for Theoretical Studies at IISc Bangalore, where he spent almost thirty years. Apart from his passion for research, he has a keen interest in physics teaching and science pedagogy at large. He was the architect of the first set of SERC schools in high-energy physics, and his lectures on group theory in these schools which appeared in the form of a monograph with S Mukhi have become an essential part of the training of anyone aspiring to become a theoretical physicist. He chaired the first Panel on Science Education formed in 1994 by the Indian Academy of Sciences, whose recommendations led to active Academy initiatives in the arena of science education. One among these initiatives is the science journal Resonance, for which Prof Mukunda served as the Founder-Editor. Prof Mukunda also helped shape KVPY in his capacity as its founder member and the first convenor.

Professor Mukunda is known for his unique style in which he practices and teaches physics. The flow of ideas in his lectures unfolds in a beautiful way and leaves a deep impression on his audience. He prepares and delivers his lectures with the same degree of zeal, enthusiasm and care for detail regardless of whether the audience stems from an elite institution or a small undergraduate college in a remote part of the country. As editor of Resonance, he has written several editorials in his characteristic style with insistence of clarity, elegance of expression and readability. Professor Mukunda is a connoisseur of cinema, especially Hollywood films of the black and white era. His demeanour is very simple, his lifestyle is appealingly minimalistic.

The present volume brings together some of the writings of Professor Mukunda which have appeared in Resonance during the last two decades. These articles are wide-ranging in nature and range from book reviews (which are like mini-books in themselves!), reflections on science and philosophy, biosketches of physicists and several articles which bring out the subtle symmetry and invariance aspects of physics. This volume is an invaluable addition to the bookshelf of any serious student of physics not only for physics per se but also as a benchmark for clarity of thought and expression one should strive for.

Arvind
IISER Mohali
Preface

I am sincerely grateful to Professor Ramakrishna Ramaswamy, President of the Indian Academy of Sciences, for thinking of including my articles in Resonance in the Masterclass series of e-books being published by the Academy. Thanks also to Professor Amitabh Joshi, Editor of Publications, and Professor Arvind, for the very kind and generous sentiments expressed in the Foreword and in ‘About the Author’.

My formal association with Resonance as a member of the Editorial Board lasted for about eighteen years. Coincidentally the thirty one articles put together in this book also span eighteen years. The articles have been arranged in five groups in a natural way, based on the nature of content, length etc. Some are relatively long, others like Article-in-a- Box are short biographical pieces. A few book reviews have also been included, hoping that they may be useful for younger readers. It was a pleasure to write each one of them, paying attention not only to content but also to language.

Some of the articles written around 2005, the centenary of Albert Einstein’s ‘annus mirabilis’, were inspired by the need to communicate to Resonance readers the enormous significance of the occasion. They had to be written.

My coauthors – V Balakrishnan, S Chaturvedi, K S Mallesh, V Natarajan, and R Simon – have very graciously agreed to have coauthored articles included in this collection, it is a pleasure to thank them. It was a rewarding experience to work with each one of them.

Let me hope that my readers derive some enjoyment as well as some benefit from this collection.

N Mukunda
Contents

Biographical Pieces

Einstein’s Miraculous Year
   with Vasant Natarajan and V Balakrishnan

Eugene Paul Wigner – A Tribute

The Life and Work of Niels Bohr – A Brief Sketch

Bohr and Dirac

Enrico Fermi – Life, Personality and Accomplishments

Sir William Rowan Hamilton – Life, Achievements, Stature in Physics

General Articles

The Story of the Photon

The Einstein–Podolsky–Rosen Paper – An Important Event in the History of Quantum Mechanics
   with S Chaturvedi and R Simon

The 2005 Nobel Prize in Physics: Optics
   with Vasant Natarajan

Space, Time and Relativity
   with S Chaturvedi and R Simon

Space and Time in Life and Science
   with Vasant Natarajan and V Balakrishnan

Symmetries and Conservation Laws in Classical and Quantum Mechanics: Classical Mechanics
   with K S Mallesh, S Chaturvedi, V Balakrishnan and R Simon
Symmetries and Conservation Laws in Classical and Quantum Mechanics: Quantum Mechanics

with K S Mallesh, S Chaturvedi, V Balakrishnan and R Simon

States of Physical Systems in Classical and Quantum Mechanics

with K S Mallesh, S Chaturvedi and R Simon

Reflections

Science and the Human Condition

Einstein’s Life and Legacy

The Philosophy of the Physical Sciences

Brief Life Sketches

Abdus Salam

About Wolfgang Pauli

Paul Adrien Maurice Dirac – An Appreciation

Lagrange and Classical Mechanics

Max Planck – Founder of Quantum Theory

Israel Moiseevich Gelfand – An Appreciation from a Physicist’s Viewpoint

Richard Phillips Feynman – Physicist and Teacher Extraordinary

Arnold Sommerfeld – Physicist and Teacher Beyond Compare

Book Reviews

“The Story of Spin” – From Spectroscopy to Relativistic Quantum Mechanics

Profile of a Polymath
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erwin Schrödinger, “What is Life? The Physical Aspect of the Living Cell”</td>
<td>271</td>
</tr>
<tr>
<td>‘Mind from Matter? An Essay on Evolutionary Epistemology’</td>
<td>275</td>
</tr>
<tr>
<td>Science and Hypothesis</td>
<td>279</td>
</tr>
<tr>
<td>Books by and about Werner Heisenberg</td>
<td>281</td>
</tr>
</tbody>
</table>
Biographical Pieces
Einstein’s Miraculous Year*

Vasant Natarajan, V Balakrishnan and N Mukunda

With each passing year, the young Albert Einstein’s achievements in physics in the year 1905 seem to be ever more miraculous. We describe why the centenary of this remarkable year is worthy of celebration.

Introduction

The revolution of the earth around the sun has given us a natural unit of time, the year. Since time immemorial, notable events in human affairs have been marked out by the year in which they occurred. Commemorations are customarily held every twenty-five years after the event. Of these, the centenary is very special. If the centenary of an event is celebrated, it signifies two things: on the one hand, a hundred years is a sufficiently long period to claim that the importance of the event has stood the test of time; at the same time, it is a period short enough to be almost within living memory, so that the historical setting of the event can be recalled reliably.

In science, too, there have occurred many notable events and discoveries that justify centenary celebrations. But there are a select few that are more than notable: they are watershed events for the human race itself, in a far more profound sense than mere political events (however tumultuous the latter may appear to be when they occur). They separate distinct eras in humankind’s understanding of the universe in which it lives. The year 1905 was, without question, such a vintage year. The current year, 2005, marking the centenary of that remarkable year, has been declared the International Year of Physics by organizations such as the UN and UNESCO. It is being observed by special programmes, lectures and seminars in a large number of countries, including India.

1905 was Albert Einstein’s *Annus Mirabilis* or ‘Miraculous Year’. Between March and December that year, the 26-year-old Einstein published six seminal papers in the journal *Annalen der Physik* that advanced – indeed, revolutionized – our understanding of the physical universe in major ways in three different directions. In the order in which they appeared, the papers (see Box 1) dealt with (i) the ‘light-quantum’ or the photon concept and an explanation of the photoelectric effect, (ii) the theory and explanation of Brownian motion, and (iii) the Special Theory of Relativity, a radically new view of space and time. Einstein himself regarded the first as truly revolutionary; it was the second major step in the development of quantum theory. In contrast, both Brownian motion and Special Relativity belong to the realm of classical physics. In addition, in 1905, Einstein discovered the equivalence of mass and energy, encapsulated in

---

Vasant Natarajan, V Balakrishnan and N Mukunda

Box 1.
The three seminal papers published by Einstein in his “miraculous year” in Annalen der Physik are, in chronological order:

(i) *On a heuristic point of view concerning the production and transformation of light*, Vol. 17, pp. 132–148. Received March 18, 1905.


The relation $E = mc^2$ appeared for the first time in

(iv) *Does the inertia of a body depend upon its energy content?*, Vol. 18, pp. 639–641. Received September 27, 1905.

perhaps the most famous formula of all: $E = mc^2$. No single year before or since then has seen such a diversity of fundamental discoveries by a single person, with the exception of the period 1665–66 in which Isaac Newton, also in his early twenties, discovered ‘the system of the world’, and much else besides (see Box 2).

The previous decade had seen three major experimental discoveries in physics. X-rays were discovered by Wilhelm Rontgen in 1895, in Germany; Henri Becquerel in France discovered radio-activity in 1896; and in England, J J Thomson identified the electron in 1897. Shortly after that, in 1900, Max Planck had taken the first step toward quantum theory with his Law of Temperature Radiation. It is helpful to remember also that, at that time, there still remained some prominent scientists – the physicist-philosopher Ernst Mach and the chemist Wilhelm Ostwald among them – who doubted the atomic nature of matter. Galaxies other than our own were unknown, and it was thought that the Milky Way comprised the entire universe. Powered flight of a heavier-than-air vehicle had just been demonstrated by the Wright brothers in 1903. Needless to say, most of the technological marvels we take for granted today (jet aircraft, mobile phones, satellite TV, computers) were completely unknown.

To properly understand the significance of Einstein’s three major achievements of 1905, we have to set the stage by going a bit further back in history.

Box 2.
It is difficult, if not impossible, to make a fair comparison of truly outstanding achievements in any field of human endeavour if these are widely separated in time and circumstance. (Is the greatest batsman to date Bradman or Tendulkar?) And yet human interest in records and extrema is insatiable. What would qualify as the most intense and sustained mental effort by a single person leading to the most profound results? Newton, Gauss, Darwin and Einstein, each at the peak of his creative outburst, would surely qualify to be very near, if not at the top, of this exclusive list. Clearly, proper mental preparation was an essential condition – their minds had to be congenial receptacles and fertile ground for the new ideas to germinate and grow. And each of these great figures did indeed “stand on the shoulders of giants” who preceded them, to see further. For, in science, there is no room for any miraculous revelation – or for unquestionable dogma, for that matter.
Einstein’s Miraculous Year

The Genesis of the Photon

In 1859 Gustav Kirchhoff had posed the following problem: to measure experimentally, and explain theoretically, the energy distribution of ‘Temperature Radiation’ over different frequencies of the radiation. If we have radiation in thermal equilibrium with material bodies at a common absolute temperature $T$, how much of its energy density lies in each small range $(\nu, \nu + d\nu)$ of frequencies? In the years that followed, many physicists – Stefan, Boltzmann and Wien, among others – made important contributions toward the solution of the problem. Wien not only proved a theorem showing that the energy density $\rho(\nu, T)$ must have the form

$$\rho(\nu, T) = a\nu^3 f(\nu/T),$$

but also suggested that the function $f$ had an exponential form, so that

$$\rho(\nu, T) = a\nu^3 e^{-b\nu/T} \quad (a, b = \text{constants}).$$

For a while, Planck believed that Wien’s formula was exact, i.e., valid for all frequencies $\nu$, and made several unsuccessful attempts to derive it from first principles. In late 1900, however, he learnt that the formula was in agreement with experimental observations only for large $\nu$, and not for small $\nu$. At low frequencies the experimental data agreed with the Rayleigh–Jeans Law, according to which $\rho(\nu, T) = (8\pi\nu^2/c^3)(k_B T)$, where $k_B$ is Boltzmann’s constant. This was the unique form for $\rho$ predicted on the basis of Maxwell’s equations of electromagnetism combined with classical statistical physics. Within a few hours of learning of this situation, he had found a formula for $\rho$ that interpolated between these two frequency regimes:

$$\rho(\nu, T) = \frac{8\pi\nu^2}{c^3} \left(\frac{\nu}{e^{\nu/hk_B T} - 1}\right) \rightarrow \begin{cases} (8\pi\nu^2/c^3)(k_B T) & \text{for low } \nu, \\ (8\pi\nu^2/c^3)(h\nu e^{-\nu/hk_B T}) & \text{for high } \nu. \end{cases}$$

This formula (now known as Planck’s Law) involved a new constant of nature, $h$, now called Planck’s constant. It fit the data for all frequencies. Over the next couple of months he constructed a mechanism, an argument, that would lead to the formula. To do this he made the assumption that matter had only discrete, i.e., quantized, energy values, so that it could only absorb and emit discrete amounts of radiative energy. However, radiation itself was assumed to obey Maxwell’s equations exactly. Its energy could therefore vary continuously from zero upward.

In 1905 Einstein presented an incisive analysis of Planck’s Law in the Wien or high-frequency limit, which was known to be the non-classical regime – clearly, this was where something new could be learnt about radiation. He considered Wien radiation with energy $E$ at a frequency $\nu$ contained in a volume $V$, and found the result (in modern notation)

"... the probability that at a randomly chosen instant the total radiation energy will be found in the portion $\nu$ of the volume $V$ is $W = (\nu/V)E/h\nu."
From this he drew the profound conclusion
“... monochromatic radiation of low density (within the range of validity of Wien’s radiation formula) behaves thermodynamically as if it consisted of mutually independent energy quanta of magnitude $h\nu$.”

And then he continued,
“... it seems reasonable to investigate whether the laws governing the emission and transformation of light are also constructed as if light consisted of such energy quanta.”

This is how the concept of the photon was born in 1905, though the name itself was coined much later (in 1926) by the chemist G N Lewis. Einstein then applied the idea to three known phenomena. One of these was the photoelectric effect, discovered by Heinrich Hertz in 1887. Hertz found that if two metal surfaces are held at a high potential difference, light from a primary spark on one surface falling on the other surface leads to another spark. In 1899 J J Thomson showed that when light falls on a metal surface, the objects liberated are electrons. In 1902 Philip Lenard discovered that the energy of these electrons was independent of the light intensity, and found qualitative evidence for an increase in this energy with increasing frequency.

There were thus three features of the photoelectric effect that were not consistent with the wave picture of light. First, the energy transferred by the light to the electron does not depend on the light intensity, which is contrary to expectation because the energy of a wave is proportional to its intensity. Second, the frequency of a wave gives the number of disturbances per unit time. One would therefore expect that a light wave with a higher frequency (and the same intensity) would liberate more electrons, but their energy would not increase. This, too, is not what is observed. Finally, experiments showed that incident light of a frequency lower than a threshold frequency (which depended on the metal) did not liberate any electrons, no matter how much the intensity (and hence energy) of the incident radiation was increased. This was puzzling because, even if one assumed that there was a threshold or energy barrier that the electrons had to overcome to be liberated from the metal, one would expect that increasing the intensity of the light would give an adequate impetus to the electrons. Why should the frequency of the light be involved?

Einstein answered all these questions in his 1905 paper. He used the idea of the light quantum to propose the extremely simple equation

$$E = h\nu - P$$

for the kinetic energy $E$ of the photo-electron. $P$ denotes the work function or the energy used up in liberating the electron from the metal surface. This equation immediately explained the apparently strange experimental results, since the energy of each “bundle of radiation” (which produces the photo-electron) is proportional to its frequency. Increasing the light intensity increases the number of radiation quanta, and thus increases the number of photo-electrons, but not the energy of each light quantum.

Today, the equation above is taught in high school, and it seems so obviously correct – in hindsight! At the time Einstein proposed it, however, it was a truly revolutionary idea that
Einstein’s Miraculous Year

required physicists to give up their well-entrenched ideas on the wave nature of light. It is therefore not surprising that considerable opposition to Einstein’s idea persisted for almost two decades after it was first presented. R A Millikan did extensive experiments from 1905 to 1915 and then said,

“I spent ten years of my life testing that 1905 equation of Einstein’s, and, contrary to all my expectations, I was compelled in 1915 to assert its unambiguous verification in spite of its unreasonableness, since it seemed to violate everything we knew about the interference of light.”

In the meantime, Einstein sharpened his concept of the light quantum. In 1909 he analyzed the energy fluctuations for temperature radiation described by the complete Planck Law (not just the Wien limit), and found that it was the sum of two contributions – one corresponding to a pure Wien Law, and the other to a pure classical Rayleigh-Jeans Law. He then described the Wien contribution in these words:

“If it alone were present, it would result in fluctuations (to be expected) if radiation were to consist of independently moving point-like quanta with energy $h\nu$.”

Around this time, Einstein took yet another revolutionary step. He argued that the Planck notion of quantization was not restricted to light waves alone, but could be extended to oscillations of other kinds. He was motivated by the fact that, similar to the breakdown of classical theory in explaining the blackbody spectrum, there was difficulty in explaining the low-temperature behaviour of the specific heat of solids. In 1907, Einstein suggested that one should treat a crystalline solid as a set of harmonic oscillators of a given frequency, and calculate its average (or internal) energy at a temperature $T$ by assuming that these oscillators had only the discrete energies proposed by Planck, i.e., that the energy of an oscillator was related to its frequency by $E = nh\nu$. The title of his paper, ‘The Planck theory of radiation and the theory of specific heat’, says it all. This was the bold first step toward the correct explanation of the specific heat of solids, and the first time that the notion of quantization was applied to oscillations other than light. Today, these quantized lattice oscillations are called ‘phonons’. Although the complete explanation of the specific heat came from Peter Debye a few years later, Einstein was one of the first physicists to accept the idea of quantization as a general principle.

Later, in 1916, Einstein showed that, besides carrying an energy $h\nu$, the light quantum also carries a linear momentum of magnitude $p = h\nu/c$, directed along its direction of propagation. After this he wrote in 1917 to his close friend Michele Besso,

“I do not doubt anymore the reality of radiation quanta, although I still stand quite alone in this conviction.”

This reflected prevailing continued opposition to the idea of light quanta – not only from Millikan, but also – surprisingly enough – from Planck and Bohr. The reason was the strong belief that the phenomena of interference and diffraction of light implied that the classical Maxwell wave theory had to be the correct description of radiation. Quantum effects had
Vasant Natarajan, V Balakrishnan and N Mukunda

to be limited to matter and its interaction with radiation. In their nomination of Einstein for an academic position in Berlin in 1914, Planck, Nernst, Rubens and Warburg went so far as to add,

“That he may sometimes have missed the target in his speculations, as for example, in his theory of light quanta, cannot really be held against him.”

As good an example of “famous last words” as any! Even later, in 1923, Bohr went to the extent of proposing that energy conservation in individual microscopic events be given up, in order to save Maxwell’s classical description of radiation. But this was a possibility that Einstein had already considered – and rejected, as early as in 1910.

The final widespread acceptance of the photon idea came only in 1925, after A H Compton and A W Simon verified the conservation of energy and momentum in the Compton effect, that is, in direct photon-electron collisions.

Brownian Motion

When microscopic, micron-sized particles such as pollen grains are suspended in a liquid, they show erratic and sudden movements as though they were being kicked around in a random fashion. This ‘Brownian motion’ is named after the botanist Robert Brown, who studied it systematically in 1827–28, but the phenomenon was known even earlier. It had been thought by some that these irregular and jerky movements were evidence for ‘vitalism’, a kind of ‘life-force’. But after Brown’s studies it became clear that no ‘vital forces’ were involved. By the 1850’s the motion was believed to be caused either by internal motions in the fluid, or by collisions with fluid molecules from different directions. Einstein was apparently not too familiar with the precise details of earlier experimental work – or rather, he characterized this work as too imprecise to enable unambiguous conclusions to be drawn. This is essentially why the phrase ‘Brownian motion’ does not appear in the title of his first paper on the subject (see Box 3), although in the text of that paper he says,

“It is possible that the motions to be discussed here are identical with so-called Brownian molecular motion ···”

His aim was far more fundamental: to show that, if the predictions of the theory could be experimentally verified, then

“... an exact determination of actual atomic sizes becomes possible.”

Indeed, the determination of atomic sizes and of Avogadro’s number $N_A$ are recurring themes in Einstein’s early work on statistical physics. He returned to the determination of $N_A$ again and again, proposing several independent methods to estimate this fundamental quantity. It is clear that one of his motivations was to establish beyond all doubt the atomic nature of matter.

Einstein’s analysis of Brownian motion was nothing less than ingenious. Using essentially physical arguments, he threaded his way through carefully, avoiding pitfalls arising from what we now know are mathematical subtleties in the behaviour of certain random processes. A year before A A Markov introduced what are now called Markov processes in the theory of
The title of Einstein’s first paper on Brownian motion was *On the motion required by the molecular kinetic theory of heat of particles suspended in fluids at rest*. This paper was received by *Annalen der Physik* just eleven days after Einstein’s doctoral thesis was completed, although the thesis itself was only published in 1906. The thesis contains results quite as fundamental as those Einstein published in his *Annus Mirabilis*. In fact, the marvellous formula relating the diffusion coefficient, Avogadro’s number, viscosity and the temperature appeared there for the first time. His second paper on Brownian motion, in December 1905, gets right to the point, being titled simply *On the theory of Brownian motion*.

Probability, Einstein had essentially recognized that Brownian motion was a special kind of Markov process, called a diffusion process. He correctly identified the distinct time scales in the problem of a micron-sized object being buffeted incessantly and randomly by much smaller molecules, and this helped him write down the equation governing the probability distribution of (any component of) the position of the larger particle, in the form

\[
\frac{\partial p(x,t)}{\partial t} = D \frac{\partial^2 p(x,t)}{\partial x^2}.
\]

This is the famous diffusion equation (also called the heat conduction equation, as the two are mathematically identical equations), \(D\) being the diffusion coefficient. Einstein also wrote down the fundamental Gaussian solution to this equation. If the particle is taken to start from the origin \(x = 0\) at \(t = 0\), this solution is

\[
p(x,t) = \frac{1}{\sqrt{4\pi D t}} e^{-x^2/4Dt}
\]

for any \(t > 0\). Once these results were in place, the crucial characteristic feature of the diffusion process emerged automatically – namely, that the average value of the *square* of the distance travelled in any given direction by a Brownian particle in a time interval \(t\) is proportional to \(t\), rather than \(t^2\):

\[
\langle x^2(t) \rangle = 2Dt,
\]

where \(D\) is the diffusion coefficient.

Einstein’s deep insight lay in the fact that he concentrated on the mean squared displacement, rather than the instantaneous velocity of the particle, as the quantity to be studied and measured. This is also related to the mathematical subtleties referred to earlier (see *Box 4*). He used an “extremely ingenious” argument that combined thermodynamics with dynamics, to relate \(D\) to the temperature \(T\) of the liquid and its viscosity \(\eta\) according to

\[
D = \frac{RT}{6\pi N_A \eta a},
\]

where \(R\) is the gas constant, for the case of spherical particles of radius \(a\). Therefore

\[
\langle x^2(t) \rangle = \frac{RT}{3\pi N_A \eta a} t.
\]
Here are some of the peculiarities of the ‘sample path’ of a particle undergoing Brownian motion in the strict mathematical sense. Its instantaneous velocity turns out to be unbounded. Its trajectory is a continuous, but extremely jagged, curve. It is an example of a random fractal: it is non-differentiable almost everywhere, and is said to be statistically self-similar. That is, its degree of jaggedness remains unchanged under arbitrarily large magnification of any portion of the curve. The curve is space-filling, in the following sense: if the Brownian motion is restricted to an infinite line or an infinite plane, then every point of the line or plane is sure to be visited infinitely often by the particle as \( t \to \infty \). However, the mean time between successive visits is infinite. If the Brownian motion occurs in three-dimensional space, the so-called fractal dimension of its trail is 2, and not 1 as would be expected of an ordinary regular curve.

This makes it possible to determine \( N_A \) by a measurement of the mean square displacement of a Brownian particle over different intervals of time.

The predictions of Einstein’s theory were checked by Jean Perrin and his students in a series of experiments from 1908 to 1914, and they were all confirmed with “an until then unmatched precision”. With this successful explanation of Brownian motion, resistance to the reality of atoms (almost!) ended. Ostwald acknowledged this in 1908, but while Mach also did so initially, he reverted later to his doubtful attitude and remained unconvinced till the end.

The Polish physicist Marian von Smoluchowski and the French physicist Paul Langevin also did pioneering and extremely significant work on the problem of Brownian motion and related matters concerning deep issues such as macroscopic irreversibility, around the same time as Einstein. (See Resonance, Vol. 10, No. 3, pp. 3–5, 2005). Brownian motion has become a paradigm for a kind of random motion with a staggering variety of applications – for instance, in stock market fluctuations, dynamical friction in star clusters in galaxies, and the dynamics of sand-piles, to name just three of these. The ramifications of Brownian motion in unexpected areas of mathematics and physics are equally astounding – the Gaussian solution written down above leads, via the so-called Wiener measure, to the Feynman or path integral formulation of quantum mechanics, and then on to the modern method of quantization in quantum field theory.

The Birth of Special Relativity

Einstein’s work on the light quantum and on Brownian motion were rooted in specific physical phenomena and problems. So was his work on relativity – in particular, Special Relativity: it sprang from the search for a consistent way to describe the electrodynamics of moving charges, which involves the dynamics of both material particles and radiation in interaction with each other. However, once formulated, the principle and postulate of special relativity transcend specific phenomena. They lead directly to deep insights into the nature of space-time itself, and into fundamental issues such as the symmetry, form-invariance and observer-independence of physical laws.

To appreciate Einstein’s achievement in this regard, we have to go somewhat further back in history. Newton’s magnum opus, Philosophiae Naturalis Principia Mathematica (the Principia, as it is generally known), was first published in 1687. In this great book he gave expres-
Einstein’s Miraculous Year

The transition to definite views on the natures of space and time – the pre-existing background or arena in which all natural phenomena occur. Essentially, space and time were regarded as individually absolute and the same for all observers. Of course, inertial observers and their frames of reference played a distinguished role, and in them Newton’s Laws of mechanics and universal gravitation are obeyed.

Almost two hundred years later, in 1865, Maxwell presented his system of equations which unified electricity, magnetism, and optics (the second grand example of unification, the first being Newton’s unification of celestial and terrestrial gravity!) Light was shown to be a propagating electromagnetic wave, with a speed calculable from electric and magnetic measurements. It soon became clear that there was a clash between Newton’s treatment of space and time, and the Maxwell theory. The speed of light in a vacuum (or free space) could be as predicted only in a sub-class of the Newtonian inertial frames, all of which would have to be at rest with respect to each other. In all other inertial frames, this speed would have to be variable, dependent on the motion of the observer. That is, in any other inertial frame (moving at a constant velocity \( \nu \) with respect to the above subclass of Newtonian inertial frames), the Newtonian law of addition of velocities would imply that the measured speed of light would be between \( c - \nu \) and \( c + \nu \), and not just \( c \) as required by Maxwell’s equations. However, all attempts to detect this frame-dependence of the speed of light failed. The most famous experiments were carried out in 1887 by Michelson and Morley, working at the Case School of Applied Science and Western Reserve University in Cleveland, Ohio. These experiments thus showed that Maxwell was correct, not Newton.

Many scientists attempted to reconcile Newtonian mechanics with the Maxwell theory, the most prominent being Lorentz, Fitzgerald and Poincaré. But their efforts were unconvincing, and ultimately unsuccessful. The definitive answer came with Einstein’s work in 1905, where he re-analyzed the nature of space and time. They are not individually absolute and the same for everybody, as Newton had visualized; rather, it is only the combined space-time continuum which is common to all, but each inertial observer divides it up into a space and a time in her own way. The difference can be illustrated in the following manner. Imagine two events which occur at two different spatial locations at two unequal times. Comparing the observations of two different inertial observers of these two events, one finds the following distinction between the old (or Newtonian, non-relativistic) description, and the new (or Einsteinian, relativistic) description:

<table>
<thead>
<tr>
<th>Spatial separation</th>
<th>Time separation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newtonian view</td>
<td>different</td>
</tr>
<tr>
<td>Einsteinian view</td>
<td>different</td>
</tr>
</tbody>
</table>

In essence, the simultaneity of spatially separated events, and the time interval between events, are not absolute concepts. They are both dependent on (the state of motion of) the observer. Thus Einstein’s resolution of the conflict was to modify Newtonian mechanics while retaining
Maxwell’s theory – the former had to fall in line with the latter. Later in 1905 he obtained, as a consequence of the modified mechanics, the famous formula \( E = mc^2 \). Special relativity was thus found via Maxwell’s theory of electromagnetism. But we must also recognise that Einstein was already aware that this classical Maxwell theory itself was in need of modification, as was indicated by the failure of the Rayleigh–Jeans law for temperature radiation, and the evidence for the quantum nature of light.

As we have already mentioned, Special Relativity is really a basic principle applicable to all of physics (except gravitation)! Here are two expressions of this idea:

From a lecture by Einstein in 1911 – “The Principle of Relativity is a principle that narrows the possibilities; it is not a model, just as the Second Law of Thermodynamics is not a model.”

And from a review by V Bargmann – “… every physical theory is supposed to conform to the basic relativistic principles and any concrete problem involves a synthesis of relativity and some specific physical theory.”

Examples of this are the Dirac equation for the electron, the theory of quantum electrodynamics and the subsequent unified electroweak theory, and the currently accepted quantum chromodynamic theory of strong interactions – in fact, the entire standard model of elementary particle physics, which is ultimately all of fundamental physics except for gravitation.

Life after Twenty Six

It is the centenary of these remarkable achievements of Einstein in 1905 that is being celebrated this year throughout the world. Any one of these three pieces of work by a single person would have established that individual’s reputation for life. What is awesome is that Einstein did all three of them (see Box 5). As Abraham Pais says in his definitive biography of Einstein, “No one before or since has widened the horizons of physics in so short a time as Einstein did in 1905.”

Box 5.
The urge to compare being an irrepressible human quality, one is tempted to ask: which of Einstein’s stupendous achievements is his greatest contribution to physics, at least in hindsight? An extremely difficult question, given the awe-inspiring depth of his insight. An excellent case can be made out in favour of his contributions to each one of the major subjects he tackled: statistical physics, quantum physics, relativity and gravitation. Some underlying themes can be distinguished. To list a few of these, he had the most profound insight into the fundamental role of fluctuations, symmetry, invariance, causality, and into the non-locality inherent in quantum mechanics.

In order to give an illustration of the way Einstein thought about physical problems, and the manner in which he combined physical arguments to arrive at far-reaching results, we summarize in Box 6 a specific instance, namely, his work on the stimulated emission of radiation. This led, when the technology became available, straight to the laser.
To round off the picture, let us recount briefly some of the significant later developments in physics in which Einstein played the leading role or to which he contributed in significant measure.

1909: As we mentioned earlier, by using the complete Planck law, Einstein showed that the energy fluctuations of temperature radiation are the sum of two terms – a non-classical particle like Wien term, and a classical wave like Rayleigh–Jeans term. Einstein described their simultaneous presence thus:

“It is my opinion that the next phase in the development of theoretical physics will bring us a theory of light that can be interpreted as a kind of fusion of the wave and the emission theory · · · The wave structure and the quantum structure are not to be considered as mutually incompatible” [emphasis added].

Thus, this was the first clear recognition of wave-particle duality in physics.

1907–1915: During this decade Einstein steadily built up his General Theory of Relativity. In attempting to bring together Newton’s theory of gravitation and Special Relativity, he saw that it was necessary to supersede both of them. Gravity found a new interpretation as curvature of space-time, and geometry became a dynamical entity, a part of physics influenced by, and influencing, the rest of nature. It should be emphasized that, while Special Relativity amounted to a requirement on all of physics except gravitation, General Relativity is the final classical theory of gravitation itself, with rules for determining the effects of gravity on all other interactions. It is, to quote Landau and Lifshitz, “the classical field theory par excellence”.

1916: Planck’s Law appeared in Einstein’s work many times – in 1905, in 1909, then again in 1916 when he gave a startlingly new derivation of it based on Bohr’s idea of discrete stationary states of atoms, and spontaneous and stimulated emission and absorption of radiation by matter. Already in Rutherford’s exponential law for radioactive decay in 1900, the notion of probability had come into physics in an important way, apart from its use in statistical mechanics. Through his work Einstein showed that this mathematical concept played a role at the most fundamental level in the atomic domain. Almost four decades later, the concept of stimulated emission was exploited in the development of the maser and the laser (see Box 6).

1917: This year saw Einstein applying general relativity to the field of cosmology, but it turned out to be somewhat premature, as Hubble’s discovery of other galaxies and the expansion of the universe was still some years away.

1925: Building on the discovery of Bose statistics by Satyendra Nath Bose in 1924, Einstein gave the first theory of the ideal quantum (or Bose) gas, and predicted the phenomenon that has become known as Bose–Einstein condensation. Parallel to the 1909 energy fluctuation formula for radiation, he now obtained a density fluctuation formula for the material quantum gas – it appeared now as the sum of a non-classical wave term and a classical particle term. This meant that matter too had to exhibit wave-particle duality.
Einstein's 1916 paper, titled On the quantum theory of radiation, is a tour de force in physics. Using simple arguments, he was able to predict several new features of the interaction between matter and radiation: the process of stimulated emission; the relation between the coefficients for emission and absorption (the Einstein A and B coefficients, still used in modern terminology); and the discrete momentum $h\nu/c$ carried by each photon.

He starts the paper with the profound statement, “The formal similarity between the chromatic distribution curve for thermal radiation and the Maxwell velocity-distribution law is too striking to have remained hidden for long”. With this motivation, he proceeds to understand the features of matter-radiation interaction from the point of view of thermodynamic equilibrium. The year is 1916. He is therefore quite familiar with his own hypothesis of radiation quanta, having used it to explain the photoelectric effect; he is aware of the Bohr model to explain the discrete nature of atomic spectra; and he is of course a master at using thermodynamic arguments, right from his doctoral thesis work on Brownian motion. But quantum mechanics itself, or the Schrödinger equation, is not yet in place. Still, Einstein is able to predict many new ‘quantum’ features of radiation.

Einstein considers a gas of atoms at a temperature $T$ and assumes that each atom has only two energy levels. He then makes certain hypotheses about the processes of absorption and emission of radiation for transitions between these levels. He then requires that, under thermal equilibrium, the rate of absorption should be balanced by the rate of emission, so that the equilibrium occupancy of the two levels remains unchanged. He shows that this is possible only if one postulates the new process of stimulated emission, in addition to the known process of spontaneous emission. With this process included, he is able to give a simple, new derivation of Planck’s radiation formula, and further show that the frequency of the emitted radiation is related to the difference in the atomic energy levels by the Bohr principle, $\Delta E = h\nu$. Going further, he states that the exchange of momentum between the atoms and radiation (and the consequent change in velocity of the atoms) should not affect the thermal (Maxwell) velocity distribution. He now uses his deep insight into Brownian motion (this time in momentum space) to show that this is possible only if each “radiation bundle” carries a momentum $h\nu/c$ along its direction of propagation.

Einstein’s prediction of stimulated emission led, almost forty years later, to the development of the maser and the laser. Today lasers are found everywhere: in your computer’s CD-ROM drive, in the grocery-store scanner, in the doctor’s office, in fibre-optic telecommunications, and in research laboratories. The momentum carried by photons demonstrated in this paper leads to radiation pressure, which is important in situations ranging from isotope separation to laser cooling of atoms. And stimulated emission in a more general avatar, called stimulated scattering of bosons, leads to the phenomenon of Bose–Einstein condensation of a gas, as first shown by Einstein in 1925. This new state of matter was experimentally created in the laboratory in 1995.

To gauge the impact of this paper by Einstein, note that no fewer than four Nobel Prizes in Physics have been awarded in recent times for related developments: in 1964 (laser/maser action), in 1981 (laser spectroscopy), in 1997 (laser cooling and trapping), and in 2001 (Bose–Einstein condensation).

1925–1927: This two-year period saw the creation of quantum mechanics by Werner Heisenberg, Erwin Schrödinger and P A M Dirac. It also witnessed the emergence of the so-called orthodox or Copenhagen interpretation with inputs from many, including Born, Bohr, Heisenberg, Jordan and Pauli. Heisenberg’s Uncertainty Principles and Bohr’s Complementarity Principle formed important components of this interpretation. At crucial stages both Heisenberg
Einstein’s Miraculous Year

and Schrödinger drew inspiration from conversations with and remarks by Einstein. However, even though he had done so much to prepare the ground for the advent of quantum mechanics, Einstein never accepted the orthodox interpretation or the claim of the finality of quantum mechanics.

1927–1930: Initially, Einstein tried to show that quantum mechanics was incorrect, by devising subtle experimental arrangements which could circumvent the uncertainty principles. This happened with respect to the position-momentum uncertainty principle during the 1927 Solvay Conference, and the time-energy uncertainty principle at the next Solvay Conference, in 1930. However, on both occasions Bohr was able to counter Einstein’s arguments and prove the consistency of quantum mechanics. Einstein accepted Bohr’s replies, but remained unconvinced of the finality of quantum mechanics.

1935: Einstein then changed his stand, and in a landmark paper with Boris Podolsky and Nathan Rosen he argued that, while quantum mechanics may well be internally consistent, it was incomplete. They proposed retaining what they called locality and realism in any complete physical theory, both of which are violated by standard quantum mechanics. The most important effect of their paper has been to highlight a key feature of quantum mechanics called entanglement. In fact, in an important contribution by Schrödinger within the year, this term was introduced for the first time; and Schrödinger went so far as to say that this was the key feature, not one of the features, of quantum mechanics. In picturesque language the idea can be conveyed thus: in the quantum mechanics of composite systems, the whole can be greater than the sum of the parts, as the latter cannot capture subtle quantum correlations. Over the decades, experiments of increasing sensitivity have ruled in favour of quantum mechanics and against the Einstein point of view. Today quantum entanglement is referred to as a resource or currency for carrying out quantum computation.

To conclude, the importance of Einstein’s work in 1905 for later developments in physics is amply evident. Usually, advances in physics, or indeed in any part of science, take place in a more-or-less steady and cumulative manner. Each step forward is built on a chain of earlier advances, and is rarely an isolated breakthrough. Occasionally, however, there occur major advances, steps into stunningly new ways of thinking (paradigm shifts, in fashionable language), which completely alter the landscape of the subject. This happened with each of Einstein’s achievements in 1905. It happened again with Niels Bohr’s atomic model in 1913, with General Relativity in 1915, and with the advent of quantum mechanics in 1925–27. Cause enough for celebration!

Suggested Reading


In addition, the reader may consult the following articles


Eugene Paul Wigner – A Tribute*

N Mukunda

One of our last surviving links with the period of the creation and development of quantum mechanics was broken with the passing of Eugene Wigner on 1 January 1995 at Princeton in USA. Wigner was remarkably talented and wide-ranging in his interests, and his work touched innumerable aspects of modern physics. In every area that he turned to, he discovered new and profound insights and interesting viewpoints, often understood and carried further by others much later. He was as much at home in fundamental problems of physics and their mathematical analysis as in engineering and technological matters. In this tribute, I shall first describe briefly his life and career, then turn to a sketch of his work, and conclude with an attempt to capture his personality and philosophy of science and life.

A Brief Life Sketch [1]

Eugene Paul (Jeno Pal in Hungarian) Wigner was born on 17 November 1902 in Budapest, Hungary, to Elisabeth Einhorn and Anthony Wigner. He thus belonged to the same generation as Werner Heisenberg, Enrico Fermi and Paul Dirac. Leo Szilard and John von Neumann were Wigner’s classmates at the Lutheran High School in Budapest – ‘at that time, perhaps the best high school of Hungary and probably also one of the best of the world’ [2]. Wigner retained great regard for his mathematics teacher L. Ratz, who also recognized and encouraged von Neumann’s unusual talents.

After a year spent at the Technical Institute in Budapest, in 1921 Wigner joined the Technische Hochschule in Berlin to train as a chemical engineer. He completed his doctorate in 1925 and then worked for a year and a half as a leather chemist. By this time he had become very much a part of the Berlin physics scene; his break came with an appointment as assistant to Richard Becker for 1926–27. This was followed in 1927–28 by a position as assistant to David Hilbert in Göttingen, and then as Privatdozent at Göttingen during 1928–30. At this point he moved to the United States, where he spent the rest of his life.

Wigner’s career in the US began as a lecturer in mathematical physics during 1930 at Princeton University, quickly elevated to a Professorship from 1930 to 1936. The year 1937–38 was spent as a professor at the University of Wisconsin at Madison. Upon return to Princeton, he became Thomas D. Jones Professor of Mathematical Physics in 1938, a position he held until 1971. The academic year 1957–58 was spent at Leiden in the Netherlands.

In 1937 Wigner became a naturalized citizen of the United States. He took his citizenship very seriously, and played a very active role in public affairs and matters of government policy. As his contribution to the war effort, he spent 1942–45 at the Metallurgical Laboratory of the University of Chicago, the last two years as the head of the theory group there. Earlier he had joined Szilard and Fermi in persuading Einstein to write the famous August 1939 letter to President Franklin Roosevelt that led to the setting up of the Manhattan Project. He was present at the University of Chicago’s Stagg Field Squash Courts on 2 December 1942 to witness the world’s first controlled nuclear fission reaction set up under Fermi’s leadership. During 1946–47 he served as Director of what later became the Oak Ridge National Laboratory in Tennessee. In 1952 he was full-time adviser to the Du Pont Company to design the Savannah River heavy-water plutonium production reactors. Soon after, in 1954 he was appointed to the General Advisory Committee of the United States Atomic Energy Commission, and served also on many panels of the Science Advisory Committee to the President of the United States.

Of the many awards that came to Wigner, we must mention the Medal for Merit, the Franklin Medal for 1950, the Enrico Fermi Award of the USAEC for 1958, the Atoms for Peace Award for 1960, the Max Planck Medal for 1961, and the 1963 Nobel Prize in physics (shared with Maria Mayer and Hans Jensen) for his wide range of contributions to quantum mechanics.

Wigner’s first marriage, to Amelia Franck in 1936, was followed by a second one in 1941 to Mary Annette Wheeler, a professor of physics. His sister Margit Balasz nee Wigner married Paul Dirac in 1937. It appears that Dirac was so shy that he once introduced his wife to an old friend as Wigner’s sister. In response (!) Wigner referred to Dirac as ‘my famous brother-in-law’ [3]. There is a charming account by Margit of her first meeting with Dirac in Wigner’s company. At a meeting in Budapest, the von Neumanns had invited Margit to visit and stay with them in Princeton. And then [4]:

‘Eugene insisted, “If you come to Princeton, you must stay with me. What would people say if you did not stay with your brother?” I was not terribly thrilled with the idea. The von Neumanns had a lovely home,..., while my brother liked to appear, and act, like a pauper. We sailed in the fall; Eugene had a two-bedroom apartment, proudly boasting that he furnished it to the cost of under $25. It looked like it ... It was soon after my arrival; we were having lunch at one of these restaurants, when a tall, slender young man entered the dining room, looked at Eugene and greeted him. He looked lost, and sad. I asked who he was, still standing undecided and none too happy looking. I was told, he was an English physicist, whom Eugene knew in Göttingen, where they used to have their meals together. “He does not like to eat alone”. “So why don’t you ask him to join us?” That was how I met Paul Dirac. That was the fall of 1934. The Institute for Advanced Studies had no building of its own as yet. Its members, like Einstein, von Neumann and Dirac as a visiting member, had adjoining rooms in a large university building, called Fine Hall. I remember so well: to the left was Einstein’s room, in the middle Eugene’s and to the right of him, Dirac’s.’
Eugene Paul Wigner – A Tribute

Wigner, Szilard and von Neumann formed the famous Hungarian trio who contributed so decisively to intellectual life in the United States in the 1930s and later. There is a story that during a meeting of scientists connected with the war effort there was so much confusion due to many languages being used that someone got up and exclaimed: ‘Gentlemen, let us use one language we can all understand – Hungarian!’

When Wigner died he left behind his third wife Eileen, a son and two daughters.

Contributions to Science and Engineering

Wigner’s work in physics is characterized by hard mathematical analysis based on simple yet profound physical assumptions. While there is a down-to-earth practical quality to some of his work, in others he dealt with the most fundamental issues with great refinement – he was both an artist and an engineer, and quantum mechanics was his medium. To quote John A. Wheeler [5]: ‘In the work of Eugene Wigner one sees the basic harmony between the conceptual framework of physics and the structure of the mathematics associated with that physics.’ On the other hand, his grasp of technology is best conveyed by this passage from Lawrence Dresner and Alvin M. Weinberg [6]: ‘.... the facility with which he could pass back and forth between engineering and physics – from a discussion of the probable distribution of energy levels in U\(^{235}\) to a critical examination of the blueprints of the concrete foundations for the Hanford reactors, or from a group theoretical argument in transport theory to the design of aluminium fuel elements!’

Wigner’s first important work in physics, completed during his apprenticeship with Becker, was a powerful treatment of quantum many-fermion systems. Around the time of the move to Göttingen, and following a suggestion by von Neumann, he undertook the major task of introducing group theoretical methods into quantum mechanics. By 1928 he had published six landmark papers on the subject; he shares with Hermann Weyl the credit for making this an essential and characteristic component of quantum physics which pervades all its applications. During the 1930s he worked in solid-state physics and at the frontiers of the developing subject of nuclear physics, making a major effort to understand the forces between nucleons, and developing the compound nucleus model to explain resonance phenomena in neutron-induced nuclear reactions. His development later of the R-matrix theory of nuclear reactions was a response to a comment by Fermi that the compound nucleus model needed a firm theoretical foundation. Probably his most remarkable work in mathematical physics – the study of the unitary representations of the inhomogeneous Lorentz group – grew out of a suggestion made to him by Dirac in 1928. This was completed in Madison in 1937, and subsequently became the basic framework for all relativistic quantum theories. He came back to problems of nuclear structure in his supermultiplet theory of 1937, and later in his statistical treatment of nuclear spectroscopy based on random matrices.

In the midst of all this, in the 1940s he worked on the theory of neutron chain reactors and the design of plutonium breeder reactors.
Wigner’s concern with the structure of quantum mechanics has led to a series of incisive insights over many years. In the early 1960s he turned to problems of interpretation and epistemology raised by the standard interpretation of quantum mechanics. At this point it is convenient to present briefly and selectively sketches of Wigner’s work under several broad areas. This is admittedly an inadequate, incomplete and possibly superficial way to survey his work, yet it may succeed in conveying some idea of the range and magnitude of his achievements. Before embarking on this, we may recall the following important books published by Wigner: (1) *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press, New York, 1959; the original German version published by Friedrich Vieweg, Braunschweig, 1931); (2) *Nuclear Structure*, with Leonard Eisenbud (Princeton University Press, 1958); (3) *The Physical Theory of Neutron Chain Reactors*, with Alvin M. Weinberg (University of Chicago Press, 1958); and (4) *Symmetries and Reflections – Scientific Essays* (Indiana University Press, 1967). We may also mention that the October 1962 issue of the *Reviews of Modern Physics*, published on the occasion of his 60th birthday, contains many articles surveying Wigner’s work in several areas.

**Structure and Content of Quantum Mechanics**

Any serious user of quantum mechanics is sure to find herself employing repeatedly, either explicitly or implicitly, one or another of the many basic concepts and methods invented by Wigner. One of the earliest is the concept of parity [7]. In classical physics, space inversion is merely a geometrical operation or transformation, a rule to map each point in space to its image by inversion through a chosen origin. The time is left unaffected. A particle trajectory, for example, would be mapped on to another possible trajectory.

*Classical space inversion:*

\[
P : x \rightarrow -x, \quad t \rightarrow t,
\]

\[
x(t) \rightarrow -x(t).
\]

Wigner showed that in quantum mechanics, parity is more than a transformation, it is a physical observable whose value can be experimentally measured. The possible results of measurement are \(\pm 1\), and the corresponding quantum states are said to possess even or odd parity, respectively.

*Quantum space inversion:*

\[
P\psi(x,t) = \psi(-x,t),
\]

\[
\psi(-x,t) = \pm \psi(x,t) \Rightarrow P = \pm 1,
\]

even/odd states.

It was this role of parity in quantum mechanics that was shown by Wigner to be the explanation for Laporte’s selection rule in atomic spectroscopy [8]: the matrix elements of the electric...
dipole moment operator, and hence the corresponding transitions, vanish unless the two concerned states have opposite parities.

The deep connection between invariance principles and conservation laws, both in classical physics and in the quantum domain with specifically new and subtle features, remained a lifelong concern for Wigner, something he came back to time and again. In the particular case of rotational symmetry, the general programme of incorporating group theoretical methods into quantum mechanics led to Wigner’s impressive body of results concerning angular momentum in quantum mechanics [9]. The detailed representation theory of the rotation group SO(3) and its covering group SU(2), which is basic to quantum mechanics, was developed by him in a form suited to practical application. The angular momentum addition theorem, the concept of tensor operators, the Wigner–Eckart theorem for their matrix elements, explicit expressions for the Clebsch–Gordan coupling coefficients (also called the Wigner 3\text{j} symbols), leading on to the intricate Racah–Wigner calculus for coupling of tensor operators and computing the resulting matrix elements, the generalizations to other symmetry groups – all these oh-so familiar tools of the trade in atomic, nuclear and particle physics originate from his work.

In his book on group theory, Wigner formulated and proved a fundamental theorem concerning the representation of symmetry operations in quantum mechanics [10]. This is a very deep and subtle result, and a brief explanation would not be out of place. The relation between physical states and wave functions (or Hilbert space state vectors) in quantum mechanics is one-to-many. This is because a change in the overall phase of a wave function is physically unobservable:

\[
\text{vectors } \psi, e^{i\alpha}\psi, e^{i\beta}\psi, \ldots \rightarrow \text{same physical state},
\]

\[
\text{Many-to-one}
\]

\[
\text{vectors} \leftrightarrow \text{physical states}.
\]

Therefore, what physical states correspond to in a one-to-one manner are not vectors but rays: a ray is an equivalence class of vectors, two vectors being declared equivalent if they differ only by a phase. The ray to which a vector \( \psi \) belongs can be unambiguously described by the corresponding projection operator or density matrix \( \rho_\psi \):

\[
\text{vector } \psi \rightarrow \text{ray } \rho_\psi = \psi \psi^*,
\]

\[
\text{rays} \leftrightarrow \text{physical states}.
\]

Rays do not form a vector space, so their geometry is somewhat harder to visualize than that of vectors \( \psi \). Wigner’s theorem then shows that any mapping \( \mathcal{T} \) of rays (i.e., physical states) onto themselves preserving quantum-mechanical probabilities – and any symmetry operation must be of such a nature – can be ‘lifted’ to either a linear unitary or an antilinear unitary (antiunitary) transformation \( T \) on vectors.
Unitary-Antiunitary Theorem:

Symmetry operation \( \mathcal{T} \),

\[
\rho_\psi = \psi \psi^\dagger \rightarrow \mathcal{T}(\rho_\psi) = \rho_\psi' = \psi' \psi'^\dagger
\]

\[
\rho_\varphi = \varphi \varphi^\dagger \rightarrow \mathcal{T}(\rho_\varphi) = \rho_\varphi' = \varphi' \varphi'^\dagger
\]

\( \psi', \varphi', \ldots \) determined up to phases,

\(|(\varphi', \psi')| = |(\psi, \varphi)| \Rightarrow \)

either

\( \psi' = T \psi, \varphi' = T \varphi, \ldots, \)

\( T \) linear unitary,

\( (\varphi', \psi') = (\psi, \varphi) \rightarrow \) unitary alternative;

or

\( \psi' = T \psi, \varphi' = T \varphi, \ldots, \)

\( T \) antilinear unitary,

\( (\varphi', \psi') = (\psi, \varphi)^* \)

\( = (\psi, \varphi) \rightarrow \) antiunitary alternative.

(Here the inner product of the Hilbert space vectors \( \varphi, \psi \) is denoted by \( (\varphi, \psi) \)). This remarkable theorem has been extended and proved under different conditions by others over the decades.

Most symmetries in quantum mechanics turn out to be of the unitary type; time reversal is one example where the antiunitary alternative is realized. The analysis of this transformation in quantum mechanics was given by Wigner [11] in 1932. In Schrödinger’s quantum mechanics, time reversal acts on wave functions thus:

\[
T \psi(x, t) = \psi(x, -t)^*.
\]

Unlike parity, however, this operation does not have the status of a physical observable in quantum mechanics, and its eigenvalues are not invariantly defined and are not experimentally measurable.

Continuing with the theme of symmetry in quantum mechanics, Wigner and von Neumann proved a very interesting result in 1929, which is of great importance especially in molecular physics [12]: if the electronic states in a molecule are classified according to their symmetry, i.e., the representation of the full group of symmetry of the relevant molecule, and if we have two distinct eigenvalues and eigenstates sharing the same symmetry (two electron terms), these eigenvalues will not cross (become accidentally equal) as one varies the internuclear distances in the molecule. On the other hand, electron terms of distinct symmetry can cross. This is a general theorem of quantum mechanics, applicable to a generic hamiltonian possessing some symmetries and dependent on a continuous parameter: as the parameter is varied, distinct eigenvalues ‘of the same symmetry’ will not accidentally cross but will repel each other.
Many years later, Wick, Wightman and Wigner [13] brought to light another aspect of symmetry in quantum mechanics, namely the existence of superselection rules. This amounts to a restriction on the applicability of the superposition principle in quantum mechanics. In general, the Hilbert space of states of a quantum system breaks up into sectors, and the formation of complex linear combinations to produce new states from old is limited to one sector at a time, not cutting across sectors. This is the reason why the phase of a spinor field – a field with half-odd integer spin – is nonobservable. So, for instance, a nontrivial linear combination of states with integer and half-odd integer angular momenta cannot be prepared. As another example, one finds that linear superpositions of states of distinct electric charge are unphysical. It is suspected that these results had long been known to Wigner, and he was persuaded by his coauthors to join them and say so in print.

In the preface to his book on group theory, Wigner relates a conversation with von Laue on the use of group theory as the natural tool with which to tackle problems in quantum mechanics [14]. He says: ‘I like to recall his question as to which results ... I considered most important. My answer was that the explanation of Laporte’s rule (the concept of parity) and the quantum theory of the vector addition model appeared to me most significant. Since that time, I have come to agree with his answer that the recognition that almost all rules of spectroscopy follow from the symmetry of the problem is the most remarkable result’.

The exponential decay law for unstable states has been well known since the days of Rutherford’s experiments on radioactivity. The first properly quantum-mechanical discussion and derivation of this law is due to Weisskopf and Wigner [15]. They were able to provide the basic theory for the natural linewidths and lifetimes of atomic states decaying via transitions to other states with emission of radiation. Their use of second-order perturbation theory along with judicious and delicate assumptions also disclosed that the exponential decay law is only an approximate, not an exact, consequence of quantum mechanics; so departures from it for both very short and very long times are to be expected.

The linear superposition principle of quantum mechanics, already referred to above, finds its most natural expression at the level of state vectors in Hilbert space. The ray space or density matrix description of physical states, which is closer to a classical description, obscures this principle somewhat – it is present but not manifest. In 1932, while studying thermodynamic equilibrium in quantum mechanics, Wigner introduced another description of states for quantum systems possessing classical canonical analogues [16]. Thus, each quantum state is describable by a certain real distribution or function on the classical phase space. In one dimension with classical phase space variables $x$ and $p$, the construction is as follows:

$$
\psi(x) \rightarrow W(x, p) = \frac{1}{h} \int_{-\infty}^{\infty} dx' \psi \left( x - \frac{1}{2} x' \right) \psi \left( x + \frac{1}{2} x' \right)^* \exp( i x' p / \hbar ).
$$

These distributions – named after Wigner – are at the level of density matrices, not state vectors. They are suggestively like classical probability distributions on phase space, such as one uses in classical statistical mechanics. However, since in general $W(x, p)$ can become negative for some arguments, we do not have a classical statistical picture with well-defined probabilities.
This is as it should be, since quantum features must be preserved. This description of states in quantum mechanics turned out to be the counterpart, or companion, to a rule or convention given by Weyl for associating a quantum-mechanical operator with every classical dynamical variable; and these ideas were further extended, particularly by Moyal [17].

Wigner contributed a great deal to the formal description of scattering and reaction processes in quantum mechanics, especially in the context of nuclear physics. One of his results concerns the physical meaning of phase shifts. In general, scattering cross-sections are determined by the squared magnitudes of S-matrix elements, and in these the phases get washed out. On the other hand, the spatiotemporal development of a scattering process described within the limits set by quantum mechanics, involves these phases. The beautiful connection found by Wigner is the expression for time delay caused by interaction and its relation to the energy dependence of the scattering phase shift [18]:

$$\Delta T(E) = 2 \frac{d}{dE} \delta(E).$$

Here $\delta(E)$ is the phase shift at energy $E$; thus, an attractive (repulsive) interaction leads to $\delta(E)$ increasing (decreasing) with energy, hence to a delay (advance) in the appearance of the final-state products of a collision after undergoing interaction.

We conclude this account with a couple of ‘curios’. Classically, one expects the possible states of a system of interacting particles – especially, a two-body system – to separate into two types: unbounded or scattering states, having positive energy, and bound states, having negative energy. In quantum mechanics we expect the energy eigenvalues to behave analogously: a continuum of unbound, positive-energy scattering states sitting on top of a set of discrete negative-energy, bound states. Only the latter have normalizable wave functions. In a remarkable paper in 1929, Wigner and von Neumann produced an example of a two-body potential which possesses a bound state embedded in the continuum [19]! This is an unexpected and essentially quantum-mechanical result. The potential is ‘artificial’ in that it has to be carefully engineered to produce the desired result, and the state involved is unstable even under small perturbations.

The passage from classical to quantum mechanics results, at the level of dynamical variables, in the loss of *commutativity* in multiplication. Thus, for two physical quantities represented by operators $A$ and $B$, in general $AB \neq BA$. However, this departure from the classical is limited in the sense that *associativity* is maintained: for three (or more) quantities multiplied in a given sequence the product is unambiguous: $(AB)C = A(BC) = ABC$. One can ask how quantum mechanics might be modified if one takes the nonclassical path one step further and, along with commutativity, one gives up associativity as well. This was examined by Jordan, von Neumann and Wigner [20] in 1934. It did not, however, lead to any alternatives with sufficiently interesting and flexible properties to give a further extension of quantum mechanics.

Going over this rich list of contributions, one is tempted to say that Wigner took his revenge for not having been involved in the discovery of quantum mechanics, and compensated for it accordingly!
Nuclear Forces, Structure and Reactions

Following the discovery of the neutron by Chadwick in 1932, there was a great deal of work exploring the nature of the strong nuclear forces between neutrons and protons. It was realized that these would be strikingly different from the familiar Coulomb forces between protons, of very short range, and with complicated distance dependences. Further dependences on spin and space exchange were also anticipated. Wigner was one of the earliest contributors to this field, and his name is associated with one of the four basic types of terms in the potential energy expression [21]:

\[
\text{potential energy between proton and neutron} =
\begin{align*}
\text{purely distance-dependent Wigner term} + \\
\text{spin exchange Bartlett term} + \\
\text{space exchange Majorana term} + \\
\text{spin and space exchange Heisenberg term.}
\end{align*}
\]

Thus, the Wigner force is the simplest of all; the others either distinguish between singlet and triplet spin states, or between even and odd orbital angular momenta, or both. Such phenomenological potentials are useful in analysing low-energy nuclear bound states, scattering processes, etc.

The low-energy (in the keV to few MeV range) scattering cross-sections of neutrons off various nuclei were experimentally studied by Fermi and his collaborators, and many other groups, around 1936. They found striking resonance structures in these cross-sections, with sharp maxima separated by very small values in between. Soon after, a theoretical explanation was offered independently by Niels Bohr on the one hand, and by Gregory Breit and Wigner [22] on the other. This is the so-called compound nucleus model. It pictures the scattering and reaction processes as taking place in two steps. At first the incoming low-energy projectile (which could be some light nucleus rather than a neutron) and the target combine to produce a compound nucleus in one of several possible metastable states. In this process the projectile energy is quickly shared with all the nucleons in the compound structure, and then the mode of formation of this structure is ‘forgotten’. In the second step, the decay of the compound nuclear state into various energetically allowed channels is governed by probability laws. It is the probability of occurrence of the first step that shows an extremely sensitive energy dependence and gives rise to the observed resonances. In their work Breit and Wigner derived the famous bell-shaped single-level resonance formula known after their names:

\[
\text{Probability of formation of compound nucleus:}
\frac{\alpha \Gamma_{\lambda}}{\left(E - E_{\lambda}\right)^2 + \frac{1}{4} \Gamma_{\lambda}^2},
\]

\(E = \text{total initial energy,}
\)

\(E_{\lambda}, \Gamma_{\lambda} = \text{average energy, width, of compound nuclear state } \lambda.\)
The partial cross-sections for subsequent decays into each of the several available final channels retain this characteristic energy dependence.

Sometime after this, around 1944, Fermi remarked to Wigner (as was mentioned earlier) that a good theoretical basis for the compound nucleus model was lacking. Thereupon Wigner set about formulating one. This was the starting point of the $R$-matrix theory of nuclear reactions, developed by him largely in collaboration with Eisenbud [23]. The basic idea is to separate the total multidimensional configuration space of all the nucleons in the compound nucleus (i.e., the projectile nucleons plus the target nucleons) into two parts: an interior region where they are all within the range of nuclear forces acting between every pair, and an exterior region where this is not so. In the latter region, one then defines or picks out essentially nonoverlapping subregions, one for each possible (two-body) final channel into which the compound nucleus can decay. Instead of posing a multichannel hamiltonian eigenfunction and eigenvalue problem, a series of matching conditions connecting the interior and exterior channel wave functions and their radial derivatives, across the borders between the interior and each exterior region, are set up. The $R$-matrix elements are quantities that enter these relations, they are a multichannel generalization of the logarithmic derivative of a wave function in a one-dimensional radial problem. The parameters entering the $R$-matrix are the energy values and the various partial decay widths of all possible compound nucleus levels. Thus, the $R$-matrix became simultaneously a convenient method for parametrization of scattering and reaction amplitudes using phenomenologically accessible compound nuclear state energies and widths, and with further developments a way to embody general physical principles, such as unitarity and causality, governing reaction processes. Inter alia this led to a multi-level generalization of the Breit–Wigner resonance expression given above, and to a criterion for the validity of the single-level formula.

Returning to the problem of nuclear forces and structure, in 1937 Wigner came up with the SU(4) supermultiplet theory to systematize the low-lying energy levels of light nuclei [24]. The idea was that the interactions among protons and neutrons, regarded as nucleons possessing the isospin degree of freedom introduced by Heisenberg [25] as early as 1932, might to a good approximation be both spin- and isospin-independent. More generally, it might be invariant under all four-dimensional unitary transformations mixing up the four independent spin-isospin states of a nucleon. (This assumption actually leads to specific spin and isospin dependences in the interaction.) It would then be possible to arrange the energy levels of ‘neighbouring’ nuclei with a common mass number into various unitary irreducible representations (UIRs) of SU(4), consider systematically the breaking of this symmetry, etc. Each UIR of SU(4) is made up of several spin-isospin multiplets in a definite way. While the idea was physically well motivated as a useful first approximation, it was pursued only to a limited extent. However, many years later, in 1964, Wigner’s theory provided the inspiration for a similar SU(6) invariant theory of baryons and mesons in the framework of the quark model [26].

At the other end of the scale from low-lying well-separated energy levels of light nuclei, we have the relatively highly excited and closely spaced levels of heavy nuclei with many degrees of freedom. Here Wigner proposed a completely different physical approach, one which
Eugene Paul Wigner – A Tribute

has stimulated work by many others and led to connections with several other problems [27]. The physical ideas may be motivated as follows. As the excitation energy (of a complicated nucleus) increases, one expects the energy levels to get closer and closer, and one also loses hope of being able to obtain them individually from a first-principles Hamiltonian. Instead, what would be more accessible and physically interesting are various statistical properties of the levels; the probability distributions for successive levels to occur at various energies, for the spacing between neighbouring levels to have different values, and so on. To obtain these statistical features, and at the same time to reflect the fact that one is dealing with a very complex system with many degrees of freedom, Wigner proposed that the basic Hamiltonian (after truncation to a large but finite dimension) be itself regarded as a random matrix, belonging to an ensemble with specified properties. Once one specifies the nature of this ensemble, regarded as a primary input, the statistical properties of the eigenvalues of the Hamiltonian, the spacing distribution, etc., can all be derived, in principle, as secondary consequences. It turns out that in using this approach one must deal with one ‘simple sequence’ of nuclear levels at a time; this is a set of levels possessing the same exactly conserved quantum numbers – ‘belonging to the same symmetry’ – such as the total angular momentum and parity. Properties of different simple sequences are independent. Thus, Wigner’s hypothesis was that the local statistical behaviour of the levels in a simple sequence is given by the properties of the eigenvalue spectrum of a random matrix drawn from a suitable ensemble. The type of ensemble to be used depends on the integer or half-odd integer nature of the total angular momentum, behaviour under time reversal, and presence or absence of rotational symmetry. Later work has shown that there are three natural types of ensembles, in correspondence with the three great families of classical compact simple Lie groups: the Gaussian real orthogonal, the Gaussian complex unitary, and the Gaussian symplectic ensembles. These ensembles consist respectively of real symmetric, complex hermitian and real quaternionic matrices (of suitable dimensions, even in the last case). The probability distribution defining the ensemble is invariant under a real orthogonal, complex unitary or unitary symplectic group of transformations applied to its elements; moreover, the matrix elements of the Hamiltonian are assumed to be independent random variables. It is the combination of these two properties that makes these ensembles Gaussian.

A great deal of sophisticated mathematical analysis has gone into these objects, and this activity continues [28]. One very interesting feature that was recognized very early was that the spacing distribution vanishes as a power of the spacing as the spacing tends to zero. The rate of this vanishing, the power involved, is characteristic for each of the three families of ensembles. The physical meaning of this result – borne out by experiments and reminding us of the no-crossing theorem of Wigner and von Neumann for electron terms of the same symmetry – is that within a simple sequence neighbouring levels do not like to come very close to one another. Had we imagined that the energy levels themselves were independently statistically distributed, there would have been no cause for such level repulsion. This only emphasizes Wigner’s idea that the properties of the ensemble of Hamiltonians must be chosen first, and other properties then obtained as consequences.
Quantum Field Theory, Relativistic Classical and Quantum Mechanics

The rules for canonical quantization – creating a quantum theory from a classical one – were originally invented in the context of nonrelativistic particle quantum mechanics. The first successful application of these rules to a classical field theory came with Dirac’s quantization of the electromagnetic field. This led to his classic 1927 paper in which he treated the processes of emission and absorption of radiation by matter, using quantum principles and the photon concept [29]. The quantized field led to a synthesis of complementary classical particle and field languages, and could describe states with variable numbers of identical particles. The canonical quantization method led to commutation relations of the form

\[ a_r a_s^\dagger - a_s a_r = \delta_{rs}, \]
\[ a_r a_s - a_s a_r = a_r^\dagger a_s^\dagger - a_s^\dagger a_r^\dagger = 0. \]

Here \( a_r \) (\( a_r^\dagger \)) are the annihilation (creation) operators for photons in various states indexed by \( r \). These states are an independent, orthogonal and complete set of modes of the electromagnetic field. The operators \( a_r, a_r^\dagger \) are quantum analogues of the classical complex coefficients in an expansion of the classical field in these modes. In this case the appearance of commutation relations led naturally to Bose–Einstein statistics for photons. Very soon after Dirac’s paper, Jordan and Wigner showed that to describe fermions (such as electrons) obeying Pauli’s exclusion principle and Fermi–Dirac statistics, the particle annihilation and creation operators must obey anticommutation relations [30]:

\[ a_r^\dagger a_s + a_s^\dagger a_r = \delta_{rs}, \]
\[ a_r a_s + a_s a_r = a_r^\dagger a_s^\dagger + a_s^\dagger a_r^\dagger = 0. \]

For a finite number of modes, they proved that up to equivalence there is only one irreducible representation of these relations, and it is finite-dimensional. This uniqueness is similar to a corresponding result in the case of commutation relations. The major difference is that from a mathematical point of view systems of operators obeying the anticommutation relations are quite ‘harmless’, while in the case of commutation relations they are unbounded and the space is infinite dimensional – even for a finite number of modes. Of course, in the Jordan–Wigner case there is no sensible classical limit.

It is interesting to note that Dirac’s initial reaction to this work of Jordan and Wigner was decidedly negative [31]. Wigner later attributed this to Dirac’s being very committed to the Hamiltonian point of view in dynamics – ‘a captive of the Hamiltonian formalism’. However, it became clear very soon that this was the correct way to set up quantum field theory for fermions, and it became part of the foundations of the subject.

The first attempts at uniting quantum mechanics and special relativity were due to Klein and Gordon. This resulted in the wave equation named after them, but it faced problems of interpretation at the one-particle level. The next, spectacularly successful, attempt was Dirac’s
work in 1928 that led to his wave equation for the electron and its series of amazing consequences [32]. Probably soon after, in 1928 itself, Dirac suggested to Wigner a comprehensive study of all possible unitary irreducible representations of the inhomogeneous Lorentz group (IHLG), i.e., of the homogeneous Lorentz group (HLG) supplemented by space–time translations. By about 1932, Majorana had constructed many of these UIRs, and later these were simplified by Dirac and Proca [33]. The solution of this problem posed by Dirac to Wigner became a herculean effort, being completed only in 1937. The result was an all-time classic paper in mathematical physics [34]. In it, Wigner acknowledges the help and guidance he received not only from Dirac but also on mathematical aspects from von Neumann. At some stage Dirac advised Wigner to be careful, and the latter replied [35]: ‘You point out that care is needed in the analysis of the representations of the Lorentz group; I promise you that I will be careful’.

Wigner’s paper contains a detailed analysis of the structure of the HLG and the IHLG, and of general unitary representations (URs) of the IHLG in the context of quantum mechanics; it then turns to a study of the UIRs. The result was that these could be classified into four broad types, depending upon the nature of the possible values of energy–momentum \( p^\mu \) occurring within the UIR, and the allowed ‘states of polarization’ for each energy–momentum. The helicity \( \lambda \) is defined as the component of angular momentum in the direction of momentum. For each kind of \( p^\mu \) (provided it is not identically vanishing) the allowed values of \( \lambda \) are determined by some UIR of a corresponding subgroup of the HLG, the so-called ‘little group’ for that \( p^\mu \); it consists of all elements of the HLG which leave \( p^\mu \) invariant. The pattern of UIRs of the IHLG is displayed in Table 1** (here space inversion or parity has been included in the HLG, except that for neutrinos this operation is undefined) [36].

While many of these UIRs were known earlier to Majorana and Dirac, the so-called infinite-spin or continuous-spin representations in cases (b) and (c) were genuinely new. In his work, Wigner did not carry the investigation of these, or of case (d), to completion. He mentioned their existence, and only remarked: ‘... the last case may be the most interesting from the mathematical point of view. I hope to return to it in another paper. I did not succeed so far in giving a complete discussion of the 3rd class.’ Wigner’s ‘last case’ and ‘3rd class’ correspond respectively to (c) and (d) in our table. We also see that not every mathematically acceptable UIR of the IHLG is acceptable on physical grounds.

Relativistic quantum systems described by any UIR of the IHLG are called ‘elementary systems’. Truly elementary particles, able to exist in isolation, are described using them. Examples are photons, neutrinos, electrons and muons. The phrase ‘elementary systems’ conveys the meaning that all their properties are revealed by studying their behaviour under all elements of the IHLG – there is no internal structure involved. In the above listing, only cases (a) and (b) for finite helicity are realized in nature.

The UIRs of case (d) are actually UIRs of the HLG \( SO (3,1) \) (or of the closely related group \( SL (2, C) \)). It remained for Harish-Chandra and for Gel’fand and Naimark to determine

** At the time this article was written it was generally believed, as indicated in Table 1, that neutrinos are massless. Over the past decade this situation has changed, some of them definitely have (very small) nonzero masses.
<table>
<thead>
<tr>
<th>Nature of $\rho'$</th>
<th>Little group within HLG (SL(2, C))</th>
<th>Number of polarization states, spectrum of $\lambda$</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Time-like</td>
<td>SO(3)(SU(2))</td>
<td>$2s + 1$ for $s = 0, 1/2, 1 \ldots$</td>
<td>Massive particles with zero or finite Spin, $s = 0$ for $\pi$ meson $s = 1/2$ for electron</td>
</tr>
<tr>
<td>(positive or negative)</td>
<td></td>
<td>$\lambda = s, s - 1, \ldots, -s$</td>
<td></td>
</tr>
<tr>
<td>(b) Light-like</td>
<td>E(2), two-dimensional</td>
<td>One: $\lambda = 0$</td>
<td>No known particles</td>
</tr>
<tr>
<td>(Positive or negative)</td>
<td></td>
<td>Two: $\lambda = \pm s, s = 1/2, 1 \ldots$</td>
<td>$s = 1$ for photons $s = 1/2, \lambda = -1/2$ for neutrinos</td>
</tr>
<tr>
<td>Euclidean group</td>
<td></td>
<td>Infinite: $\lambda = 0, \pm 1, \pm 2, \ldots$ or $\lambda = \pm 1/2, \pm 3/2, \ldots$</td>
<td>No known particles</td>
</tr>
<tr>
<td>(c) Space-like</td>
<td>SO(2, 1)(SL(2, R))</td>
<td>One: $\lambda = 0$</td>
<td>Imaginary mass, unphysical</td>
</tr>
<tr>
<td>(d) Vanishing</td>
<td>HLG(SL(2, C))</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In his contribution to RMP, Dirac made the following comments on Wigner’s work [39]. ‘The problem of working out all unitary representations of the IHLG has been dealt with by Wigner, taking the mathematical point of view that two representations are equivalent if they are connected by a unitary transformation. He decomposes the representations into their irreducible constituents and finds that the irreducible constituents provide theories of elementary particles with various spins. This work does not lead to any interaction between particles. To bring in interaction, one must depart from the point of view of looking at two representations as equivalent if they are connected by a unitary transformation, a point of view which involves looking upon all unitary transformations as trivial. To a physicist, some unitary transformations are trivial, whereas others (for example, the $S$ matrix) are far from trivial, so a physicist cannot look upon two representations connected by a unitary transformation as necessarily equivalent.’ The point is that for any really interesting relativistic quantum system, such as a relativistic quantum field theory, it is not only important to know which UIRs of the IHLG are present, but also how they are put together. However it must be pointed out that as early as 1949 Wigner himself had drawn attention to this situation [40]: ‘The elementary systems correspond mathematically to irreducible representations of the Lorentz group and as such can
be enumerated... However, in the description by irreducible states, the form of almost all physically important operators remains unknown and, in fact, depends on the system, the types of interactions, etc. This leads to a rather strange dilemma: in the customary description the form of the physically important operators is known but the time dependence of the states is unpredictable or difficult to calculate. In the description just mentioned, the situation is opposite: the time dependence of the states follows from the invariance properties, but the form of the physically important operators is hard to establish.’

Wigner returned on many occasions to a description of the results of his classic work. He also constructed with Bargmann a unified set of wave equations whose solutions would lead to UIRs of types (a) and (b) in our table [41]. His work with Newton on the problem of position is particularly interesting, so I describe it in a little detail [42].

The starting point of nonrelativistic particle quantum mechanics is the set of positions and momenta as primary dynamical variables, out of which all other variables are built up. (Later work has shown that these positions and momenta can be derived as secondary objects starting from suitable quantum-mechanical representations of the Galilei group). Now, from Wigner’s point of view in the relativistic context, the primary things are the UIRs of the IHLG. After having set them up, one must examine within which UIRs one can construct position operators with physically desirable properties. Such an analysis was first undertaken by Newton and Wigner. They were able to show that in every finite mass and finite spin UIR (case (a)) a unique set of position operators possessing several physically reasonable properties does indeed exist. However, contrary to naive expectation, they do not form the space components of a relativistic four-vector. This illustrates the fact that in quantum theory the unitary transformation law is more basic than the geometric one or manifest covariance. In the massless case with finite nonzero helicity even this much cannot be done. Thus, neither photons nor neutrinos can be localized in space.

In other related work we mention the study by Inonu and Wigner of the process of ‘group contraction’ by which the IHLG goes over in the nonrelativistic limit to the Galilei group [43]; Salecker and Wigner’s analysis of deep conceptual problems in bringing together quantum mechanics and general relativity, caused by quantum limitations on position measurements [44]; and van Dam and Wigner’s construction of classical relativistic direct-interaction theories resting upon integro-differential equations for particle trajectories [45]. One of Wigner’s conclusions was that while special relativity and quantum mechanics could at least conceptually be combined, with general relativity and quantum mechanics there was no common ground at all.

**Interpretation of Quantum Mechanics**

In the early 1960s Wigner turned to a serious examination of the problems of interpretation of quantum mechanics, and a clear expression of the orthodox position which essentially coincided with his own [46]. As evidence for the latter, here is his own statement: ‘The orthodox view is very specific in its epistemological implications ...A large group of physicists finds it
difficult to accept these conclusions and, even though this does not apply to the present writer, he admits that the far-reaching nature of the epistemological conclusions makes one uneasy.’ He also often said that he was adding hardly anything new to London and Bauer’s classic 1939 exposition [47]. He accepted the treatment of measurement theory that had been articulated by his friend von Neumann [48] as early as 1932, and wanted to restate it for a new generation and extract its ultimate consequences for epistemology.

Wigner emphasized that the state vector of a quantum system changes in two mutually exclusive ways – continuous, deterministic Schrödinger evolution when not subject to observation, and discontinuous, probabilistic, collapse when measurements are made. He went to much length to show that the linear Schrödinger equation – even including the apparatus and the system’s coupling to it – can never produce the macroscopically desired collapse phenomenon, and stressed repeatedly that pure states and mixtures have very different physical properties. He also presented a pragmatic answer to the question ‘What is the state vector?’. It was that it codifies in a compact way all past information about a system, on the basis of which we can state the probabilistic connections that quantum mechanics gives among a series of measurements carried out subsequently and sequentially in time; all the consequences of quantum mechanics are just such statements. So, as the orthodox view claims, ‘the laws of quantum mechanics can be expressed only in terms of probability connections’, and cannot be formulated in terms of objective reality.

Pursuing this analysis further, Wigner came to the conclusion that human consciousness is an essential external ingredient needed to make complete sense of quantum mechanics. The collapse of the state vector occurs when and only when an observation is registered in some individual consciousness: ‘It is at this point that the consciousness enters the theory unavoidably and unalterably. If one speaks in terms of the wave function, its changes are coupled with the entering of impressions into our consciousness’. And again: ‘... it was not possible to formulate the laws of quantum mechanics in a fully consistent way without reference to the consciousness’. In support of this declaration, Wigner appeals to Heisenberg and says: ‘W. Heisenberg expressed this most poignantly (Daedalus, 1958, 87, 99): “The laws of nature which we formulate mathematically in quantum theory deal no longer with the particles themselves but with our knowledge of the elementary particles ... The conception of objective reality... evaporated into the... mathematics that represents no longer the behaviour of elementary particles but rather our knowledge of this behaviour”’. As one can imagine, this line of thinking led Wigner inexorably to a kind of solipsism, and to the delineation of two kinds of reality – the content of one’s own consciousness, the only absolutely real, and everything else external to oneself, including every other person’s consciousness. To support the former he turned to Schrödinger; ‘... the most eloquent statement of the prime nature of the consciousness with which this writer is familiar and which is of recent date is on page 2 of Schrödinger’s Mind and Matter: “Would it (the world) otherwise (without consciousness) have remained a play before empty benches, not existing for anybody, thus quite properly not existing?”’ But there was a sign of asymmetry – the only absolutely real, one’s own consciousness, does depend on food, air and water for its own survival and functioning,
as we are painfully aware; so he made a case for devising experiments which might show up the effects of consciousness on matter. In talking of the first kind of reality, Wigner also realized and stated its obvious limitations – its awakening with birth and infant growth, its extinction at death. So he argued for a deep study of the former phase, to understand the nature of consciousness.

Wigner felt that the development of quantum mechanics had widened the outlook of most physicists, and also in a sense made them inward-looking: ‘Until not many years ago, the “existence” of a mind or soul would have been passionately denied by most physical scientists ... Even today, there are adherents to this view though fewer among the physicists than – ironically enough – among biochemists’. He also saw that quantum mechanics reinforces the circumstance that any observation and interpretation of measurement rests on previously constructed and understood theory. Thus, we are linked in a chain to the very beginnings of our acquisition of knowledge of our surroundings and its regularities – indeed to phylogenesis and ontogenesis.

Today it may seem that these conclusions of Wigner were premature. Certainly, efforts are aplenty to find more ‘acceptable’ interpretations of quantum mechanics, without appeal to ourselves as essential prerequisites. Was Wigner then ‘a victim of his generation’? Should we smile at these conclusions which he found inescapable? Or was he only being ruthlessly honest and expressing clearly what others hesitated to put into words?

**Solid-State Physics, Reactor Theory and Technology**

I will touch upon these areas only briefly. Wigner’s interest in problems of solid-state physics and materials science stemmed from a very early date. There must have been links to his original training as a chemical engineer; later on his detailed knowledge of properties of materials played a key role in his work on reactors. Among his gifted students in solid-state science in the 1930s we may mention John Bardeen, Gregory Wannier and Frederick Seitz. It was Wigner who suggested to Wannier [49] ‘that there ought to be a way to reconcile the local and the band concept for electrons, and that such a reconciliation would probably be useful in understanding the spectra of insulators’. Wigner also worked on radiation damage in solids – the detailed microscopic picture of lattice defects occurring when materials are irradiated with neutrons, the resulting changes in heat and electrical conductivity and ductility, and also the ways in which the material seems to recover from the damage as time goes on [50].

Wigner was the source of much of the theory and the major technological developments connected with nuclear reactors. His contributions began in 1940. As briefly mentioned earlier, he was a leader at the University of Chicago Metallurgical Laboratory during 1943–45. He contributed to the development of research reactors, power reactors and plutonium production reactors. On the theoretical front he made major contributions to the spectrum of the Boltzmann equation, neutron thermalization, thermal utilization and resonance absorption. All very practical contributions ‘which one would hardly, *a priori*, have associated with the same man who introduced group theory into quantum mechanics’ [51].
Wigner was a gifted and articulate expositor of science and its principles to general audiences. However, he frequently indulged in a kind of mock humility — as his Princeton colleagues explained his language [52], ‘A piece of work is “amusing” if it is correct and beautiful; it is “interesting” if it is wrong and messy.’ And in describing the epistemology of quantum mechanics to an audience of non-physicists, he said of himself the writer [53]: ‘He realizes the profundity of his ignorance of the thinking of some of the greatest philosophers and is under no illusion that the views to be presented will be very novel. His hope is that they will appear sensible.’ He could convey sharp ideas pithily: ‘Someone once said that philosophy is the misuse of a terminology which was invented just for this purpose’.

These apart, his grasp of and concern for the grand principles of science were very deep. The role of invariance principles and their associated conservation laws captivated him — he dwelt upon them at length on many occasions [54], and said: ‘A large part of my scientific work has been devoted to the study of symmetry principles in physics....’ He titled his Nobel lecture ‘Events, laws of nature, and invariance principles’. He often described as a miracle the fact that human understanding could uncover laws of nature, and separate them from the accidents of initial conditions. The laws provide structure and coherence to events, and, in turn, the symmetry principles provide these qualities to laws; thus one has the ascending progression: events to laws to symmetry principles.

Turning to the role of mathematics in natural science, he expressed wonder at the way in which mathematical concepts and connections show up in unexpected ways and places, and also at the fact that tentative theories turn out upon further development to be far more accurate than could reasonably have been expected at the outset. This led him to conclude that, since we do not quite know why we succeed so well so often, we must be cautious and not immediately regard a successful explanation as the truth!

Pondering on the likely future of science, Wigner wondered whether it might not wind down under its own weight, and lose its attractiveness to the young. The increasing extent of science makes it go beyond the reach of any one individual. But the response to this cannot just be an increase in team efforts, because this can never capture true creative thinking in the individual subconscious. There is a need here to find deeper ways of sharing information and insight, of harmonizing the collective conscious with the subconscious in each individual.

Continuing on the theme of the growth of science and the emergence of large collaborative efforts, he argued for protecting the individual and giving value and esteem to little science: ‘One does not have the satisfaction which creative work, as we know it today, provides, if one’s activities are too closely directed by others’. About the emergence of deep insights, ‘It is hard to imagine how they can be developed other than in comparative solitude’. And as for the pleasures of pursuing science: ‘It has been said that the only occupations which bring true joy and satisfaction are those of poets, artists, and scientists, and, of these, the scientists are apparently the happiest.’
Eugene Paul Wigner – A Tribute

Through the description of his work I have tried to convey the fact that Wigner acknowledged very graciously his debt to some of his most gifted contemporaries. He was also generous in his assessment of them. Of von Neumann he wrote: ‘… whenever I talked with the sharpest intellect whom I have known – with von Neumann – I always had the impression that only he was fully awake, that I was halfway in a dream.’ And about Richard Feynman: ‘He is a second Dirac, only this time more human’.

Two persons that Wigner had been very close to – Enrico Fermi and von Neumann – both died in their fifties. Wigner described and contrasted their attitudes to the inevitable. With Fermi, ‘On a heroic scale was his acceptance of death …He was so completely composed that it appeared superhuman’. But with von Neumann it was very different: ‘It was heartbreaking to watch the frustration of his mind, when all hope was gone, in its struggle with the fate which appeared to him unavoidable but unacceptable’. These experiences must have affected Wigner deeply; at a convocation address to an audience of young students soon after, he said: ‘Our culture is committing a sin by covering our eyes against the realization that none of us will be here always’. And to a general audience some time later: ‘The recognition that physical objects and spiritual values have a very similar kind of reality has contributed in some measure to my mental peace’. These various expressions seem related.

Wigner was a physicist who achieved rare range and depth in his life and work. He was a product of the old world who flowered during the golden age of theoretical physics, and carried the fragrance of his subject to the new world. The time seems past when such another can appear.

Suggested Reading

[1] In the following, frequent references will be made to the following three sources: (a) E P Wigner, Group Theory and its Application to the Quantum Mechanics of Atomic Spectra, Academic Press, New York, 1959; (b) Reviews of Modern Physics, Vol. 34, No. 4, 1962 (c) E P Wigner, Symmetries and Reflections – Scientific Essays, Indiana University Press, 1967. For brevity we shall refer to these as GT, RMP and SR, respectively.


N Mukunda


[26] See, for instance, the reprint group 3 in F J Dyson, ref. 24.


[35] Quoted in ref. 13.
For quantum–mechanical purposes, one has to deal with the group SL(2,C), the universal covering group of the HLG, and its subgroups.


P A M Dirac, in RMP, p. 592.


G Wannier, in RMP, p. 645.

SR, p. 126.

Dresner, Lawrence and A M Weinberg, in RMP, p. 747.

V Bargmann *et al.*, in RMP, p. 588.

SR, p. 186.

In addition to the articles included in SR, see also R M F Houtappel, H van Dam, and E P Wigner, The conceptual basis and use of the geometric invariance principles, *Rev. Mod. Phys.*, Vol. 37, p. 595, 1965.
The Life and Work of Niels Bohr – A Brief Sketch*

N Mukunda

The life and work of Niels Bohr are briefly reviewed covering: his family life, background, and education; the importance of his stay in England after PhD, and the crucial contact with Rutherford; the period of the Old Quantum Theory initiated by his work on atomic structure; and his role as mentor to the next generation of physicists through his Institute in Copenhagen. The contacts with Einstein, their debates on quantum mechanics, their deep mutual respect, and their personalities, are described.

The Atomic Hypothesis

The idea of atoms as the ultimate indivisible building blocks of matter is about 2500 years old. In India it was suggested by Kanada (6th century BC), and in Greece by Leucippus (∼450 BC) and his student Democritus (460 BC–370 BC). Over the centuries there have been both supporters and opponents of this idea. After the advent of modern science in the 17th century, the atomic hypothesis was revived in quantitative terms by John Dalton around 1803. However the actual experimental discovery of atomic structure and the foundations of its theoretical understanding resulted from the work of Ernest Rutherford and Niels Bohr about a century ago. Together they truly ushered in the atomic and nuclear age. Bohr along with Albert Einstein, by general consent, were the two greatest physicists of the 20th century.

The Early Years

Niels Henrik David Bohr was born to affluent parents on 7 October 1885 in Copenhagen in the small country of Denmark. Other notable physicists of his generation were Max Born, Hermann Weyl and Erwin Schrödinger. Niels’ sister Jenny was two years older; while his brother Harald, to whom he was extremely close, was two years younger. Their father Christian Bohr was a professor of physiology at the university, and an important cultural figure in Copenhagen. In the father’s eyes, Harald was ‘brilliant’ but Niels was ‘special’! Family life was close, highly cultured and intellectual, with Niels and Harald being exposed to scientific, literary and philosophical discussions from a very young age. This may well be contrasted with the cases of Einstein and Paul Dirac, for example.

Niels and Harald attended the Gammelholm School in Copenhagen, then went on to the university. From the beginning it seemed understood that they would both pursue academic careers in physics and mathematics respectively. Niels completed undergraduate and master’s

degrees in 1907 and 1909, and the PhD in 1911 for a thesis based on a profound study of the Lorentz–Drude electron theory of metals. During his years at the university he was influenced by the philosophical ideas of Benedict Spinoza and Denmark’s own Søren Kierkegaard. He was also concerned from an early age with the difficulties of communication, problems of words and language, and ambiguities in understanding. This led to his being extremely careful in writing and in the choice of words, all through his life.

**Trip to England**

At this point, Niels (hereafter Bohr) was very keen to work with J J Thomson – the discoverer of the electron – at the Cavendish Laboratory in Cambridge in England, and to describe his own work to the great man. Bohr reached Cambridge in the autumn of 1911 and met Thomson soon after. However things did not go well, partly because Bohr began by pointing out mistakes in Thomson’s papers, and partly because Thomson took an inordinate amount of time to read Bohr’s paper. Earlier that year, in May 1911, Rutherford at Manchester had announced the discovery of his model of the atom, with a massive but tiny nucleus at the centre and electrons a great distance away, like a miniature solar system. By June 1911 the famous Rutherford scattering formula was also in hand. At the end of 1911 Rutherford visited Cambridge, and Bohr was deeply impressed by his work as well as his personality. (A wonderful account is given by R F Peierls, ‘Rutherford and Bohr’, *Resonance*, Vol. 15, No. 5, 2010). Bohr decided to go to Manchester to work with Rutherford. The immediate and deep friendship and mutual respect between the two were to flower greatly in the coming years. On the physics front, Bohr saw very soon that the Rutherford atom model would lead to the explanation of the properties of chemical elements, the structure of the Periodic Table, with the behaviour of electrons determining all of chemistry. On the other hand, he also saw that alpha and beta rays, two forms of radioactivity, originated from the nucleus.

Bohr also understood that there was a need to create a new mechanics for the atom, different from the classical Newtonian one. He remembered how Planck in 1900 and Einstein in 1905 had shown the failure of classical ideas in understanding the properties of radiation. And his own supreme achievement became the initiation of a new mechanics for the atom in which Planck’s quantum of action plays a key role. In his own words:

“In the spring of 1912 I became convinced that the electronic constitution of the Rutherford atom was governed throughout by the quantum of action.”

As Paul Dirac would say many years later, the steps taken by Bohr were the most difficult ones in the passage from classical to quantum physics.

In due course Bohr’s ideas also led to an understanding of the laws of spectroscopy.

In April 1912 Bohr moved to Manchester. Over the next four months, up to July 1912, he worked incredibly hard on his ideas sketched above, all based on the Rutherford atom model, more precisely on developing his own ‘quantum version’ of it. Before leaving Manchester the draft of a first paper was done and shown to Rutherford.
Return to Denmark

Returning to Copenhagen, Bohr and Margrethe Norlund were married on 1 August 1912, the beginning of a 50-year relationship of deep love, harmony and mutual support. They were to have six sons, the eldest of whom, Christian born in 1916, was to die in a tragic boating accident in 1934. Another died in childhood.

In the autumn of 1912, Bohr was appointed assistant professor at the University in Copenhagen. In 1916, after a special Chair was created, he became professor of theoretical physics. Several times in this period Rutherford tried to lure him to Manchester on a more permanent basis, but Bohr only stayed there for a limited period during 1914–1916. Work on his atom model continued, and in 1913 the trilogy of papers: ‘On the Constitution of Atoms and Molecules : Parts I, II, III’ appeared in the Philosophical Magazine in July, September and November. When Einstein came to know about Bohr’s results he remarked:

“Then it is one of the great discoveries”

“Then the frequency of the light does not depend at all on the frequency of the electron... And this is an enormous achievement. The theory of Bohr must be then right.”

The first time Bohr met Einstein (as well as Planck) was in April 1920, when he was invited to Berlin. After this meeting here is what they wrote to one another:

Einstein to Bohr: “Not often in life has a person, by his mere presence, given me such joy as you did... I have the pleasure of seeing your youthful face before me, smiling and explaining. I have learned much from you, especially also about your attitude regarding scientific matters.”

Bohr in reply said: “To me it was one of the greatest experiences ever to meet you and talk with you and I cannot express how grateful I am for all the friendliness with which you met me on my visit to Berlin.”

Period of Old Quantum Theory – Nobel Prize – Later Work

The Old Quantum Theory spanned the quarter century 1900–1924, and clearly the second half was dominated by the impact of Bohr’s work of 1912–1913. In this phase there were major contributions by Arnold Sommerfeld, Paul Ehrenfest, and Einstein. This entire effort paved the way for the later development of quantum mechanics proper.

An important event in 1921 was the inauguration of the Institute of Theoretical Physics in Copenhagen in the month of March with Bohr as Director. Bohr had worked tirelessly towards its creation from 1917 onwards. Through the 1920’s and 1930’s it became a meeting place for talented physicists from many parts of the world, to visit and work for varying periods. In fact, during the period of the Old Quantum Theory and the subsequent discovery and consolidation of the new quantum mechanics, the three centres in Munich, Göttingen and Copenhagen formed an important triangle – Munich emphasizing the physical aspects, Göttingen more mathematically oriented, and Copenhagen the philosophical and interpretational problems. In 1965, the Institute was renamed the Niels Bohr Institute.
The Nobel Prizes in Physics for 1921 and 1922 were announced jointly on November 10, 1922, and awarded to Einstein and Bohr respectively. Bohr immediately wrote to Einstein:

“...To me it was the greatest honour and joy ... that I should be considered at the awarding of the prizes at the same time as you. ... I have felt it as the greatest good fortune that ... the fundamental contribution that you have made ... should be recognized, also quite publicly, before I was considered for such an honour...”.

To which Einstein replied two months later: “Your cordial letter reached me shortly before my departure from Japan. I can say without exaggeration that it pleased me as much as the Nobel Prize. I find especially charming your fear that you might have got the prize before me – that is truly ‘bohrisch’...”

After the advent of quantum mechanics in its two forms – matrix mechanics in 1925 and wave mechanics in 1926 – Bohr’s role changed to that of a guide and elder statesman. The most rapid advances came from the younger group of Heisenberg, Dirac, Pauli, Fermi (along with crucial contributions by the older Schrödinger and Born); and Bohr concentrated on the problems of interpretation of the new mathematical formalisms. His own later significant contributions to physics include the enunciation in 1927 of his Complementarity Principle; the analysis with Rosenfeld in 1933 of the inner consistency of the mathematical formalism of quantum electrodynamics; the theory of the compound nucleus in 1936; and the study with Wheeler in 1939 of the mechanism of nuclear fission.

Debates with Einstein

Bohr’s later interactions and debates with Einstein form an important part of the history of 20th century physics. At the 1927 Solvay Conference, Einstein brought up ‘thought experiments’ to try and show that Heisenberg’s position-momentum uncertainty principle could be violated. At the 1930 Solvay Conference he similarly attacked the time-energy uncertainty principle. On both occasions Bohr was able to answer him and to show the inviolability of these uncertainty principles and the inner consistency of quantum mechanics. In 1935 Einstein (with Podolsky and Rosen) tried to argue that quantum mechanics, while correct, was incomplete. Once again Bohr was able to counter him. (This particular episode is the starting point of a serious study of the phenomenon of entanglement in quantum mechanics, so important today). Later Bohr wrote an extended article in the 1949 volume ‘Albert Einstein – Philosopher Scientist’ edited by Paul A. Schilpp, reviewing his years of discussions and debates with Einstein, which is also a classic of the physics literature.

During the second World War, as Denmark was occupied by Nazi Germany, Bohr had to be “smuggled out” to the USA under a false name – Nicholas Baker! In 1941 Heisenberg, who had been so close to him earlier, visited Bohr in Copenhagen, but this encounter must have been painful to both. Michael Frayn’s play ‘Copenhagen’ tries to reconstruct what might have happened. His attempt in 1944 to talk with Winston Churchill was a fiasco, due to problems of communication.
Some Quotes from Bohr

As mentioned earlier, Bohr was always conscious and concerned about problems of language and communication. His talks and articles are quite dense and difficult to read, prompting Einstein to say: “He utters his opinions like one perpetually groping and never like one who believes he is in possession of definite truth”. Nevertheless some of his most famous pieces should be mentioned, to tempt the brave reader: (i) The lecture at Como in September 1927 announcing his Complementarity Principle; (ii) the 1932 lecture in Copenhagen to an audience of light therapists on ‘Light and Life’; and (iii) ‘Light and Life revisited’ on 21 June 1962 at the University of Cologne.

There are many statements of Bohr which are at the same time both profound and often subtly humorous. Here are a few:

- Strictly speaking, the conscious analysis of any concept stands in a relation of exclusion to its immediate application.
- Every sentence I say must be understood not as an affirmation, but as a question.
- It is difficult to predict, especially about the future.
- Tomorrow is going to be wonderful, because tonight I do not understand anything.
- You are not thinking, you are just being logical.
- If quantum mechanics has not profoundly shocked you, you haven’t understood it yet.
- There are some things so serious you have to laugh at them.
- The opposite of a fact is falsehood, but the opposite of one profound truth may very well be another profound truth.

On 18 November 1962, a little while after lunch, Bohr passed away.

Comparing Bohr and Einstein

In this brief sketch of his life and work, Einstein has played a very prominent role. It is appropriate then to conclude with some sentences from Abraham Pais, who had known them both very well, taken from an address to the Royal Danish Academy of Sciences and Letters in December 1989 titled ‘Reflections on Bohr and Einstein’:

“Both would speak with intense enthusiasm and optimism about work they were engaged in, and had enormous powers of concentration... In younger years, Einstein’s spectrum of scientific activities was broader than Bohr’s ... Both men were indefatigable workers, driving themselves on occasion to states of exhaustion which would lead to illness, more serious in the case of Einstein.”
...Their prime concern was always with what they did not understand, rather than with past achievements. ... From the point of view of science, Bohr was more spectator than actor in his later years. ... Both Bohr and Einstein were a-religious. ... Einstein had a lifelong interest in philosophy.... As for Bohr, less well read in philosophy than Einstein, philosophizing was part of his nature from boyhood on... . How to avoid ambiguity – that was the problem that worried Bohr. ... They differed in their views regarding the interpretation of quantum mechanics. They argued frequently about it, particularly over the concept of Complementarity. ... By and large, however, similarity outweighed disparity. Both had a deep need for simplicity, in thought and in behaviour. ... They took science very seriously, but to them it was ultimately a game. The greatest similarity, though, was that Einstein and Bohr were both scientists without whom the birth of that uniquely twentieth-century mode of thought, quantum physics, is unthinkable.”
We present an account of the work of Niels Bohr and Paul Dirac, their interactions and personalities.

1. Introduction

In this essay I would like to convey to my readers something about the personalities and work of Niels Bohr and Paul Dirac, juxtaposed against one another. Let me hope that the portraits I will paint of these two great figures from the world of physics will be faithful to the originals. The year 1985 was celebrated as the centenary of Bohr’s birth, while Dirac passed away in October of the previous year. There was a gap of almost a generation between them. Let us also recall that Einstein’s life spanned the period 1879 to 1955; so he was just six years older than Bohr.

For Bohr and Dirac, the most important work of their lives was bound up with the strange story of the quantum—the struggle to adapt and alter the fabric of classical physics to accommodate Planck’s quantum of action. This called for an overhauling of all three components of the classical scheme—matter, motion and radiation. Naturally Bohr appeared on the scene at an earlier phase of the struggle than did Dirac, and several others were also involved, but here our focus will be on these two.

2. Planck’s Interpolation

Some of you may remember that Planck made his momentous discovery sometime in the evening of Sunday, October 7, 1900 (incidentally, Bohr’s fifteenth birthday). The experimental physicist Heinrich Rubens and his wife had visited the Plancks for tea that afternoon. Rubens told Planck of his and Kurlbaum’s measurements of the black-body radiation spectrum in the far infrared limit, where he had found definite deviations from the Wien radiation law. This law was a theoretical one which had been proposed in 1896 by Wien, and which Planck had believed to be exactly valid. Soon after the Rubens left, Planck set to work to find an interpolation between Wien’s Law, known to be valid at high frequencies, and the low frequency measurements just reported to him by Rubens, which incidentally agreed with the theoretical results of Rayleigh and Jeans. It was thus that Planck arrived at his celebrated radiation law.

It is somewhat staggering to realise that quantum theory was born or discovered in this way in the space of a few hours!

The quantum of action was thus first discovered via the thermodynamic properties of light, and in the succeeding years the first insights into its significance came largely through statistical arguments as well as the wave–particle duality of light. In all of this, of course, Einstein played a leading role. However, the connection of Planck’s discovery to the structure of matter, its stability and its mechanics had to wait for Bohr’s magic touch in the years 1912–13.

3. Bohr at Cambridge and Manchester

During his doctoral work on the electron theory of metals, completed in 1911, Bohr had realised very clearly that there was a need for a radical departure from the laws of classical electrodynamics in the atomic domain. It was extremely fortunate for him that in April 1912 he went to work briefly with Rutherford at Manchester, after a disappointing stint with J J Thomson at Cambridge. At Manchester he came to know of Rutherford’s model of the atom in which the positively charged core of the atom, the nucleus, containing practically all the mass, occupied a negligible volume at the centre of the atom. This was in contrast to Thomson’s model, in which the positive charge was spread out uniformly over a finite volume of atomic dimensions. Many problems and possibilities immediately became clear to Bohr. On the one hand, in order to produce in this model a length scale of the order of the atomic size, and also to ensure stability of the electron orbits, it was essential to bring in Planck’s constant. On the other hand, it now appeared that all the chemical properties of an element should depend only on one datum, namely the number of peripheral electrons, i.e., the atomic number rather than the mass number. In fact, Bohr saw that while chemistry was determined by the outermost electrons of the atom, all radioactive processes like α and β emission originated from the nucleus, deep inside the atom. It appears that at this stage Bohr took Rutherford’s model more seriously than Rutherford himself did.

Turning to the structure of the atom, Bohr assumed that the electrons moved in concentric circular rings around the nucleus. Classical electrodynamics could never explain the stability of such an arrangement; but Bohr had already anticipated the need for a fundamental departure from classical ideas in this realm. He was familiar with Planck’s method of quantizing the motion and the energy of simple harmonic motion, and he now adapted it to the motion of an electron in the Coulomb field of the nucleus. As much by inspiration as by deduction he was able to arrive at the right order of magnitude for atomic sizes, and at the expression $E_n = -A/n^2$ for the allowed energies of an electron bound in an atom. Here the integer $n$ takes values 1, 2, 3, ... For all this of course, Planck’s constant was essential, but at that time the exact form of the quantum condition was beyond him.

At this stage another important event occurred – he was called upon to investigate the passage of α-particles through matter and to analyze the processes by which they ionized the atoms of matter, losing energy and slowing down as they did so. This was a matter of practical importance in Rutherford’s laboratory. The fact that he could give a satisfactory classical account of
Bohr and Dirac

this process, whereas classical ideas failed completely within the atom, led him to the following truth: however deep the break with classical ideas might be, the new theory would have to agree with the old one in the limit of low frequencies or large quantum numbers. This was the origin of the famous Correspondence Principle, which played such a major part in subsequent developments.

4. Connection to Rydberg’s Formula

At this point in his thinking, Bohr had dealt only with the structure and stability of the atom, and had not yet connected up with atomic spectroscopy or radiation phenomena. He returned from Manchester to Copenhagen in July 1912, married Margrethe Norlund in August 1912, and set about writing up the ideas conceived in Manchester. It was only in early 1913 that his mind suddenly turned to problems of atomic radiation. Atomic spectroscopy was a well-developed field with a lot of data on the characteristic spectral lines and frequencies associated with various elements. There also existed several empirical formulae, giving simple expressions for many series of spectral lines. H M Hansen, a colleague of Bohr’s at the University of Copenhagen, asked him in early 1913 if he knew of Rydberg’s formula which expresses every frequency as the difference of two terms, and which for hydrogen took the simple form

\[
\nu_{mn} = R \left( \frac{1}{n^2} - \frac{1}{m^2} \right)
\]

where both \( n \) and \( m \) were integers. Bohr had not known this even though it had been around since 1890, and Rydberg worked at the nearby University of Lund in southern Sweden. So this query and information from Hansen came as a complete surprise to Bohr. But at the same time he saw that it gave the missing clue to the problem of quantization in the atom. He compared his own formula \( E_n = -A/n^2 \) for quantized electron energies in an atom with individual terms in Rydberg’s expression and immediately realized that each spectral line corresponded to a transition of an electron from one allowed state to another, accompanied by the emission of a quantum of radiation. In the Planck–Einstein spirit, it was Bohr who first saw the Rydberg law as an expression of the conservation of energy,

\[
h\nu_{mn} = E_m - E_n, E_m = -hR/m^2.
\]

By demanding agreement with classical theory for large \( n \), Bohr was able to completely pin down the quantization condition as well as to calculate the value of Rydberg’s constant. The break with classical physics came with the fact that none of the spectral frequencies \( \nu_{mn} \) coincided with any of the classical orbital frequencies, but such a break was essential to explain the stability of the atom, as anticipated by Bohr. In fact, he said that Rydberg’s formula gave him such a transparent clue that he immediately saw the quantum picture of the emission of radiation. He was sure he was on the right track inspite of the total breakdown of classical physics; at the same time the Correspondence Principle was obeyed.
In 1913 he published his three famous papers on the constitution of atoms and molecules, where he stated his two fundamental postulates: (1) the electron could only be in one of a special set of stationary states which had to be chosen out of all possible classical motions by imposing quantum conditions; (2) the transition of the electron from one such state to another is a non-classical and non-visualizable process, during which a single quantum of radiation is emitted or absorbed according to the Rydberg–Bohr frequency condition.

5. Reactions to Bohr’s Work, Later Developments

Many predictions of Bohr’s theory were checked in Rutherford’s laboratory, but the English physicists, in particular Fowler and Jeans, were skeptical and accepted his ideas only reluctantly. It seems that in Göttingen there was a sense of scandal and bewilderment. But both Einstein and Sommerfeld saw immediately the significance of Bohr’s ideas.

I have devoted a considerable amount of space to recounting this early phase of Bohr’s work, because it was the foundation of all else that followed. Indeed, though the quantum of action was discovered in the properties of radiation, the route to the new quantum mechanics was via the mechanics of the atom. And the application of Planck’s ideas to the dynamics of matter, which Dirac was to later describe as the most difficult first step, was taken by Bohr.

Bohr was fully aware of the limitations of his theory. It was necessary to generalize the quantum condition from the circular motions of a single particle to the motions of general mechanical systems; to analyze the relationship between classical and quantum aspects of atomic phenomena; and to explore the many applications of his theory. To do all this, he gradually built up a school around himself in Copenhagen. One of his earliest collaborators was Kramers from Holland, who joined him in 1916. By 1919, he had an Institute of his own. Meanwhile his programme had also been taken up by the groups at Göttingen and Munich, led respectively by Max Born and Sommerfeld. The three centers worked in an atmosphere of friendly cooperation with frequent exchanges of ideas, and sharing of successes, hopes and people. Pauli and Heisenberg, among others, travelled frequently from one of these centers to another. In 1915, Sommerfeld found the general form of the quantum conditions for any so-called multiply-periodic system, and soon Bohr adopted Sommerfeld’s mathematical methods. Instead of a picture of electrons moving in concentric circular orbits in a plane, Bohr could now deal with shells of electron orbits, tackle complex atoms and their spectra, and go on to elucidate the structure of the periodic table. This was of course, a great shot in the arm for chemistry. One must remember that Bohr did all this even before the Pauli exclusion principle and the electron spin had been discovered. In all this work the Correspondence Principle was the constant guide, being used both brilliantly and judiciously. In 1921 the Correspondence Principle was extended to dispersion by Ladenburg, and Kramers followed this up in Copenhagen. In this work he was joined by Heisenberg. (Along the way, Bohr collected the Nobel Prize for 1922.) But not all the data could be satisfactorily explained by the theory. Bohr remained acutely aware how far he was from a logically consistent framework which was able to explain his two postulates and at the same time be in harmony with the Correspondence Principle. In fact, the period 1923–
1925 witnessed a crisis in the old quantum theory. To this period belongs a famous paper of Bohr, Kramers and Slater. In this, Bohr tried to give an overall picture of radiative processes taking place in the atom, and the authors suggested that classical causality had to be replaced by a purely statistical description. This paper had a deep influence on Heisenberg, as it showed even more clearly the inadequacy of the classical picture of atomic processes.

As is well known, the resolution of the crisis came with Heisenberg’s discovery of matrix mechanics in June–July 1925. This was a direct outgrowth of his work with Kramers in Copenhagen on dispersion, and of the influence on him of the Bohr–Kramers–Slater work. But all that is another story.

6. Dirac at Cambridge

Meanwhile, back at the ranch in Cambridge, a young Paul Dirac had joined R H Fowler as a research student in 1923, after getting a degree in electrical engineering. For two years he worked on applying Hamiltonian methods to multiply periodic systems in the framework of the Rutherford–Bohr model, but that did not lead to any significant success. Then in September 1925, his lucky break came when, by a somewhat roundabout route, he learnt of Heisenberg’s discovery of matrix mechanics. This was the spark that ignited him. He soon elaborated, practically in isolation, his own version of quantum mechanics, giving it a particularly abstract and elegant structure. One might remember here that Heisenberg’s achievement had been aided by continuous contact and exchange of ideas with Bohr, Born, Pauli, Kramers and Sommerfeld. In any case, once the key step had been taken by Heisenberg, progress towards the establishment of a mathematically satisfactory quantum mechanics was extremely rapid and was essentially finished by early 1927. Schrödinger’s discovery of wave mechanics had come in early 1926, and its equivalence to Heisenberg’s version soon after. One of Dirac’s key contributions in this phase was the exposure of the link between classical and quantum mechanics. This was the most beautiful expression of the Correspondence Principle and, said Dirac, it had given him the most pleasure of all his discoveries.

7. Interpreting Quantum Mechanics

From 1925 to 1927, the most important advances were being made by Dirac in Cambridge, Heisenberg, Born and Jordan in Göttingen, and Schrödinger in Zurich. During this period, Bohr was in a sense watching from a distance, with a critical but approving attitude. He had inspired and oriented the work of the others; and the new theory had attained the goals he had set himself all along. The departure from classical physics he had sensed and foreseen for so long was now explicitly expressed; relations among physical quantities could no longer be maintained in the classical numerical sense, but only in a more abstract algebraic sense. Every physical attribute of a system could not always be reduced to a number. When the stage was set to find the physical meaning of the mathematical structure, Bohr re-entered the scene. The deeper understanding of the situation needed Bohr and his philosophical bent of mind. Indeed
Heisenberg said of him:

Bohr was primarily a philosopher, not a physicist, but he understood that natural philosophy, in our day and age, carries weight only if its every detail can be subjected to the inexorable test of experiment.

In early 1927, between the two of them, Bohr and Heisenberg developed what we now call the ‘Copenhagen interpretation of quantum mechanics’. In this, they were greatly aided by the transformation theory of quantum mechanics, which had just been developed by Dirac and Jordan. Heisenberg’s contribution was the uncertainty relations. Bohr’s was the complementarity idea. According to the latter, every classical concept retains its usefulness in quantum mechanics, but not necessarily simultaneously. According to Bohr, this was the greatest lesson of quantum mechanics – that the classical concepts, each individually valid, might be mutually exclusive. In later years he would say that physics had by its simplicity shown the way to this profound idea, but that the idea itself was applicable to much more complex situations, such as the relation between physics and life.

Einstein critically attacked the Copenhagen interpretation at the two Solvay Congresses of 1927 and 1930, and it was Bohr who answered him each time and proved the logical consistency of quantum mechanics. Finally Einstein had to concede, saying only that he still felt there was an unreasonableness about it all. Of Bohr himself he said:

His is a first-rate mind, extremely critical and far-seeing, which never loses track of the grand design,
and

He is truly a man of genius, it is fortunate to have someone like that.

8. Dirac’s Later Work

Turning our attention now to Dirac for a while, I have already recounted how he burst on to the scene in late 1925. Thereafter, he kept going like a house on fire, with a steady and staggering profusion of fundamental ideas and discoveries. One of his most important papers, on the quantum theory of the emission and absorption of radiation, was written at Bohr’s Institute in Copenhagen; so he too had been drawn into Bohr’s circle. By applying the principles of quantum mechanics to the electromagnetic field, Dirac brought to a successful conclusion the work begun by Planck in 1900, and also inaugurated quantum field theory. Then there was the discovery of the new statistics named after him and Fermi, the relativistic theory of the electron, the prediction of the positron and the general concept of antimatter, the idea of the magnetic monopole, and many more. In the midst of all this, he wrote the classic book The Principles of Quantum Mechanics, often compared with Newton’s Principia. It would take a great deal of space to do justice to all that Dirac accomplished in this period. Just as Bohr had
Bohr and Dirac

made the preceding era a heroic one, Dirac turned this one into the ‘Golden Age of Theoretical Physics’.

There is a charming anecdote from the Solvay Congress of 1927, which is worth recalling. In the interval between two sessions, Bohr asked Dirac what he was working on, to which Dirac replied that he was looking for a satisfactory relativistic wave equation for the electron, which would combine special relativity and quantum mechanics properly. Bohr then told him that such an equation had already been found by Klein and Gordon, but before Dirac could explain why he was not satisfied with it, the bell rang and they had to go back to the next session. Dirac later said:

... It rather opened my eyes to the fact that so many physicists were quite complacent with a theory which involved a radical departure from some of the basic laws of quantum mechanics, and they did not feel the necessity of keeping to these basic laws in the way that I felt.

9. Dirac’s Style

Dirac’s style is essentially mathematical, and he turned out to be a master craftsman in the art of theoretical physics. He created with ease the mathematical tools that he needed. Bohr on the other hand was somewhat like Faraday. As Heisenberg said,

... his insight into the structure of the theory was not a result of a mathematical analysis of the basic assumptions, but rather of an intense occupation with the actual phenomena, such that it was possible for him to sense the relationships intuitively rather than derive them formally.

For Dirac, considerations of mathematical beauty and symmetry were of the highest importance, and he was matchless in the art of manipulating and working with the abstract. Bohr, on the other hand, was much more concerned with the problems of interpretation and communication, the difficulties and ambiguities inherent in language, and other such philosophical questions.

Dirac’s writings have a characteristic and unmistakeable directness, simplicity and beauty. Bohr, on the other hand, is much harder to read because each long sentence of his contains a great deal of thought in a highly compressed form. He spent a lot of effort in the choice of each important word. Bohr’s style of work was to have a junior collaborator sit at a desk and take down notes while he himself kept pacing up and down the room, forming and changing and reforming his phrases and sentences. Watching him at one such session, Dirac apparently said something to the following effect:

Professor Bohr, when we were young we were taught never to start a sentence until we knew how to finish it.

Bohr’s speech and handwriting were, respectively, inaudible and illegible. On both counts, Dirac was far superior. As Bohr himself said:
Whenever Dirac sends me a manuscript, the writing is so neat and free of corrections that merely looking at it is an aesthetic pleasure. If I suggest even minor changes, Paul becomes unhappy and generally changes nothing at all.

As I recalled earlier, Bohr was very deeply interested in the problems of biology, which he saw as a fertile field of application for his Principle of Complementarity. In fact, for him physics was a far simpler problem. In Dirac’s writings I have been able to find a reference to biology. In his paper of 1931 concerned with the magnetic monopole, he says,

There are at present fundamental problems in theoretical physics awaiting solution, e.g., the relativistic formulation of quantum mechanics and the nature of atomic nuclei (to be followed by more difficult ones such as the problem of life)...

At another time he is supposed to have said that his equation for the electron explained all of chemistry and most of physics. Presumably for him, the problem of life was just one more of the things that theoretical physics would deal with in good time!

Bohr created and inspired an international school of theoretical physics; and his influence upon others was as much by direct contact and involvement in their struggles as through his writings. Dirac, on the other hand, worked largely on his own. He did not create a school of any kind, although his influence on others through his writings and ideas has been enormous.

10. Bohr’s Later Work

In the years following the creation and completion of quantum mechanics, Bohr turned to the problems of nuclear physics while Dirac was more concerned with relativistic quantum field theory and later on with gravitation and cosmology as well. However, there is a classic contribution by Bohr along with Rosenfeld in 1933 to quantum field theory. They analyzed the consistency of applying the principles of quantization to the electromagnetic field – something which Dirac had done in 1927 – and demonstrated the logical necessity of doing this if the quantum mechanics of particles and, in particular, Heisenberg’s uncertainty relations were to be maintained.

11. Two Personalities

As human beings, there is a great deal worthy of admiration in both Bohr and Dirac, and a touching simplicity and sincere modesty in their dealings with others. Dirac was always most ready to acknowledge his debt to others. And in seminars, it seems that Bohr would always preface his questions with the statement that he only wished to better understand the speaker’s point of view. Bohr concerned himself with political matters and spoke a great deal on philosophical issues as well, while Dirac seems to have avoided both these areas. Bohr was quite categorical that quantum mechanics was complete; and the most valuable lesson it had taught us was that of complementarity. He was anxious to extend its application to other fields.
Bohr and Dirac

such as reason and instinct, heredity and environment, physics and biology. His debate with Einstein, begun in the 1927 Solvay Congress, continued for more than two decades, and he adhered to his point of view. In the 70's however, Dirac had this to say,

\[
\ldots \text{the present form of quantum mechanics should not be considered as the final form. It is the best that one can do up till now. But one should not suppose that it will survive indefinitely into the future. And I think that it is quite likely that at some future time we may get an improved quantum mechanics in which there will be a return to determinism and which will, therefore, justify the Einstein point of view.}
\]

One is left speculating on what Dirac actually had in mind.

Physicists are familiar with many lovely sayings and stories about and by Bohr and Dirac. And they are all really a reflection of their greatness as human beings. Bohr was always a synthesizer of conflicting points of view. On one occasion he said,

\[
\text{The opposite of a correct statement is a false statement. But the opposite of a profound truth may well be another profound truth.}
\]

On another occasion he is quoted as saying,

\[
\text{There are things that are so serious that you can only joke about them.}
\]

One of Dirac’s most celebrated statements was about the value of mathematical beauty in physics. He said,

\[
\ldots \text{it is more important to have beauty in one’s equations than to have them fit experiment \ldots It seems that if one is working from the point of view of getting beauty in one’s equations, and if one has really a sound insight, one is on a sure line of progress.}
\]

This reminds us of the poet John Keats saying, “What the imagination seizes as beauty must be truth – whether it existed before or not.”

Bohr paved the way from the world of classical physics to the world of the quantum, guiding everybody through the most difficult period with his unerring instinct and intuition. And when the great victory had been won, it was he who most comprehensively assessed the impact it had for the nature and goals of science. Dirac was one of the chief architects of the victory, and he went on to raise theoretical physics to unparalleled heights of imagination and beauty. As much for their heroic labours as for their great human qualities, Bohr and Dirac will always rank among the greatest scientists of all time.
Enrico Fermi – Life, Personality and Accomplishments

N Mukunda

A brief account of the life of Enrico Fermi, against the background of others of his generation, is given. His role in the revival of Italian Physics, and influence on many others, are described. Fermi’s major contributions to quantum physics, in both theory and experiment, and aspects of his unique personality, are recounted.

Introduction

The opening years of the twentieth century saw the birth of several exceptionally gifted persons, all of whom, a generation later and in their twenties, played stellar roles in the creation, formalization, consolidation and interpretation of quantum mechanics: Wolfgang Pauli in 1900, Werner Heisenberg and Enrico Fermi in 1901, Paul Dirac and Eugene Wigner in 1902, and John von Neumann in 1903. Just a few years later, within the same decade, came John R. Oppenheimer and George Gamow in 1904, Hans Bethe and Ettore Majorana in 1906, Rudolf Peierls in 1907, and Lev Landau and Victor Weisskopf in 1908. One is tempted to ask – why and how did this near simultaneous appearance of such great talent in such profusion come about? After reflection, we are likely to agree with what Heisenberg said in a talk in 1973 celebrating the five hundredth birthday of Copernicus:

“To what extent are we bound by tradition in the selection of our problems? ...Looking back upon history..., we see that we apparently have little freedom in the selection of our problems. We are bound up with the historical process, and our choice seems to be restricted to the decision whether or not we want to participate in a development that takes place in our time, with or without our contribution. ... one may say that a fruitful period is characterized by the fact that the problems are given, that we need not invent them. This seems to be true in science as well as in art.” And, one might add, within science, not in physics alone.

Early Years to PhD at Pisa

Enrico Fermi was born on 29 September 1901 in Rome, Italy, to parents of modest means. His grandfather Stefano came from a farming background, around Piacenza in northern Italy, and later became a county secretary. His father Alberto worked in the Italian State Railway system, and settled in Rome. Enrico’s mother, Ida de Gattis, a school teacher, was fourteen

years younger than Alberto. Enrico was the youngest of three children, with Maria born in 1899 and Giulio in 1900.

Being just a year apart, Giulio and Enrico were extremely close to one another, with shared interests and talents in building machines and gadgets. However, tragedy struck the family when Giulio died in 1915 after a freak incident in a hospital. During his school years, Enrico taught himself physics and mathematics from a book written by a Jesuit priest in 1840. He and Enrico Persico were great friends in this period. He was also guided in his reading by a friend of his father, and was soon recognized to be exceptionally gifted. In November 1918, after school, Enrico gained admission to the prestigious Scuola Normale Superiore in Pisa, an institution set up by Napoleon. The period 1918–1922 spent in Pisa, with his class fellow Franco Rasetti and enlivened by many pranks, were most happy and lively years for Enrico (hereafter Fermi) – the first world war had ended, and Trieste and Trento had been won back from Austria though at great human cost.

Fermi completed his PhD at Pisa in 1922, just around the time the fascist movement was taking control of Italy. He was clearly far ahead of his contemporaries and even his teachers. With support from fellowships, he then spent six months in Gottingen with Max Born, and somewhat later in Leiden with Paul Ehrenfest. However he felt somewhat of an outsider to the group around Born, wondering how he compared with others there and looking for appreciation and encouragement from Born.

The Rome Group – ‘Senator Corbino’s Boys’

Upon return to Italy, Fermi taught mathematics and physics as an Instructor at Rome during 1924, then at Florence till 1926. In the Florence period he completed a piece of work in quantum statistics which immediately established his reputation. At that point came the move to Rome. The initiative was taken by Professor Orso Mario Corbino, head of the physics department at Rome, and member of the Italian Senate. Fermi had met Corbino earlier, after finishing at Pisa. Corbino was from Sicily in southern Italy, and he had a vision and ambition to revive Italian physics which had declined compared to the days of Galileo and Volta. In comparison, Italian mathematics was doing very well, thanks to figures like Gregorio Ricci-Curbastro, Tullio Levi-Civita, Luigi Bianchi, Vito Volterra and others. By October 1926, Fermi was professor at Rome, then Rasetti came from Florence to join the group being built up by Corbino. Soon after, Edoardo Amaldi, Emilio Segrè and Ettore Majorana joined as Fermi’s students, completed their Ph.D’s, and became members of the Department. The years up to the late 1930’s saw a great flowering of Italian physics around Fermi, the members of the group being referred to as ‘Senator Corbino’s boys’ or the ‘via Panisperna boys’. As the acknowledged leader, Fermi was ‘the Pope’.

His finest achievements in the Rome period were his 1933 theory of beta decay; and soon after, the experimental work on slow neutron induced radioactivity.

A charming incident in this period is a conversation between Fermi and Majorana as reported by S Chandrasekhar:
Majorana: There are scientists who ‘happen’ only once in every 500 years, like Archimedes or Newton. And there are scientists who happen only once or twice in a century, like Einstein or Bohr.

Fermi: But where do I come in, Majorana?

Majorana: Be reasonable, Enrico! I am not talking about you or me. I am talking about Einstein and Bohr.

The Rome group lasted till the mid to late 1930s. We will pick up Fermi’s story later. As for the others – Rasetti left Rome for Université Laval in Quebec, Canada, in 1939; then moved to Johns Hopkins University in Baltimore, USA from 1947 to 1967. Emilio Segrè left Rome for Palermo in Italy in 1936, then to the University of Berkeley, USA, in 1938. Edoardo Amaldi left Rome in 1939, served in the Army 1939–1941, then stayed on in Italy and played an important role in the revival of Italian physics after the second world war. Majorana, in some sense a tortured genius, spent some time with Heisenberg in Leipzig in 1933; lived a secluded hermit’s life 1933 to 1937; took up the Chair in Theoretical Physics at Napoli in 1937; and disappeared on a boat trip from Palermo to Napoli in March 1938.

Some other well-known Italian physicists inspired by Fermi were Giulio Racah, Ugo Fano and Gian Carlo Wick.

Family Life – Personality – The Nobel Prize

At this point, let us return to the early 1920’s, and describe important events in Fermi’s family life. His wife Laura (nee Capon, born 1907) met him for the first time in 1924 as part of a hiking group with several other young people. She came from a non-observing Jewish family, her father was a naval officer, and she recalled later that Fermi showed a protective attitude towards her from the very beginning. Their second meeting was in late 1926 after he had joined the Rome physics department. They became engaged in early 1928, and married in July that year. For a while they lived a frugal life, on his monthly salary of US$ 90.00. Later, conditions improved – in 1929 Fermi was named to Mussolini’s new Accademia d’Italia, bringing him a title and a higher salary. They had two children – a daughter Nella born in 1931, and son Giulio in 1936. Here are some descriptions of Fermi’s nature and personality, taken from Laura’s engaging biography of him titled Atoms in the Family: he felt that

‘... a losing cause is worth no effort’;

‘... qualities that were to become assets in his scientific prose – the going straight to the point with no flourishes, the simplicity of style, the avoidance of any word not strictly essential’;

‘He was of simple tastes and, moreover, he thought that complaining was an idle form of expression directed to no purpose’;

‘he would never seek money, never ask or strive for more of it. “Money,” he used to tell me, “has the tendency of coming of its own will to those who don’t look for it. I don’t care for money, but it will come to me”.’
During the years of rising fascism, Fermi often felt it was better to emigrate to the USA; but Laura, feeling more strongly rooted in Rome, was reluctant to do so. Finally, in 1938 the persecution of Jews started, while politically Italy became subservient to Nazi Germany. All this culminated in their decision to leave for good. The timing was unusual. On November 10, 1938, the Nobel Prize in physics for 1938 was announced; it was given ‘To Professor Enrico Fermi of Rome for his identification of new radioactive elements produced by neutron bombardment and his discovery, made in connection with this work, of nuclear reactions effected by slow neutrons.’

That year only two Nobel awards were given – to Pearl Buck for literature, to Fermi for physics. Fermi, Laura and their children went by train from Rome to Stockholm, attended the Nobel ceremonies, then went by ship via England to the USA, reaching New York on January 2, 1939. Fermi had been offered a professorship at Columbia University; he was to work there till April 1942.

Columbia–Chicago–Los Alamos

In late 1938, the chemists Otto Hahn and Fritz Strassmann in Berlin discovered that when bombarded by a neutron, the uranium nucleus split into fragments of substantial sizes – the fission process. The physicists Lise Meitner and Otto Frisch soon confirmed that a great deal of energy was also released. Fermi and his group in Rome had missed all this in their work. He now immediately hypothesized that many neutrons would also be released during fission; if they could be saved and slowed down, they could cause further fissions; and in principle a controlled self-sustaining chain reaction with enormous energy release could be created. He immediately started experiments at Columbia to realize this.

Meanwhile, over July and August 1939, Leo Szilard and Eugene Wigner met Einstein and persuaded him to sign a letter drafted by them addressed to President Franklin Roosevelt, informing him about the possibility of creating new extremely powerful bombs based on the nuclear fission chain reaction. The names of Fermi and Szilard, and their work, are repeatedly mentioned in the letter. It was received by Roosevelt on October 11, 1939, and an ‘Advisory Committee on Uranium’ was soon set up. Two years later, on December 6, 1941, the decision to make an all-out effort in atomic energy research was taken by the US Government, leading to the ‘Manhattan Project’.

As the work at Columbia grew in scale, it was shifted to the University of Chicago in 1942. Fermi moved there in April 1942, his family in June. On December 2, 1942, on the campus of the University of Chicago, the team led by Fermi achieved ‘the first self sustaining chain reaction and thereby initiated the controlled release of nuclear energy’. In 1944 Fermi moved to Los Alamos to work on the atomic bomb project, under the leadership of Oppenheimer. After the success of the project and the end of the Second World War, he returned at the very end of 1945 to Chicago, where he lived and worked for the rest of his life.
Magnitude of Fermi’s Work

Fermi had a very down-to-earth extremely physical approach to problems, and was able to see the essentials of any situation very quickly. The range of his work is amazing encompassing both theory and experiment. While we have referred to some of his important work already, it is worth describing some of them further.

In 1922, while still a student in Pisa, he contributed a beautiful and important idea in the framework of general relativity, which has come to be called ‘Fermi transport’ (see article by Joseph Samuel, ‘Fermi Transport’, Resonance, Vol. 19, No. 1, pp. 62–72, 2014) or ‘Fermi–Walker differentiation’ after an extension made by A G Walker in 1932. This is distinct from covariant differentiation, and is a rule for carrying or evolving vectors and tensors along a given world-line in space time. It makes precise the idea of a ‘non-rotating’ or ‘non-spinning’ coordinate frame, and is an appealing and useful concept.

Fermi’s major contribution to quantum statistics has already been briefly mentioned. The Pauli Exclusion Principle, applied to electrons bound in an atom, had been enunciated in late 1925 (The translation of Fermi’s paper from German appears in the Classics Section, ‘Quantization of an Ideal Monoatomic Gas’, Resonance, Vol. 19, No. 1, pp. 82–96, 2014). In February 1926, while in Florence, Fermi generalized this to the case of an ideal monatomic gas, and obtained a distribution function differing from both the classical Boltzmann distribution and the quantum Bose distribution of 1924. Later, in August 1926, Dirac showed how both Bose and Fermi distributions come from the basic principles of quantum mechanics, from two contrasting symmetry properties of wave functions for indistinguishable particles. It appears that when Fermi told Dirac he had obtained the result earlier, Dirac replied that he had seen Fermi’s paper but then forgotten it! This statistical law is now named after both of them as the Fermi–Dirac statistics.

Here is an interesting rejoinder to this story. The Fermi Golden Rule is a famous formula in quantum mechanics for the transition rate for a system to go from some given initial state to any chosen final state under the action of a perturbation. It was apparently first obtained by Dirac, but has ever since carried Fermi’s name!

A very interesting application of his new statistics to the cloud of electrons bound in an atom followed in 1927. It was an approximate method which took into account the exclusion principle and the Coulomb attraction to the nucleus, to calculate an effective potential in which the electrons move. As it was independently proposed also by L H Thomas, it is called the Thomas–Fermi model for the atom, applicable for heavy atoms.

In 1930 summer Fermi spent two months at the University of Michigan at Ann Arbor giving a series of lectures on the new quantum electrodynamics. This later appeared as ‘Quantum Theory of Radiation’ in Volume 4 of the Reviews of Modern Physics in 1932. It is probably the most beautiful account of this subject at the stage it had reached at that time. This was Fermi’s writing at its best, so beautiful that one cannot resist quoting what many outstanding physicists said about it:
Wigner: ‘His article on the Quantum Theory of Radiation in the Reviews of Modern Physics (1932) is a model of many of his addresses and lectures: nobody not fully familiar with the intricacies of the theory could have written it, nobody could have better avoided those intricacies’.

Bethe: ‘... It is an unsurpassed example of simplicity in a difficult subject. It appeared after a group of extremely complicated papers on the subject, and preceded another group of papers that were equally complicated. Without Fermi’s luminous simplicity I think that many of us would have been unable to explore field theory in depth. I am certainly one of them.’

Weisskopf: ‘Fermi was unique in his way of doing physics. He had a very special way of attacking problems. He always managed to find the simplest and most direct approach, with the minimum of complication and sophistication. In the early 1930’s, when I tried in vain to understand the new quantum electrodynamics I was lucky enough to find in Reviews of Modern Physics’ Fermi’s article called ‘Quantization of radiation in the Coulomb gauge’. I studied it, and from then on I understood field theory. I know I am not the only one who reached this result and who has this opinion’.

Finally we come to Fermi’s monumental 1933 theory of beta decay (see the article by G Rajasekaran, ‘Fermi and the Theory of Weak Interaction’, Resonance, Vol. 19, No. 1, pp. 18–44, January 2014.). After the two classical theories of gravitation and electromagnetism, this is the next fully quantitative and detailed theory of a fundamental interaction in nature, completely in the quantum domain. It is astounding that so soon after quantum field theory in the form of quantum electrodynamics was founded, and the Pauli neutrino hypothesis had been presented as a way to preserve energy and angular momentum conservation in beta decay, Fermi constructed this next quantum field theoretic description of a fundamental interaction in nature. Ideas of particle creation and annihilation and (vector) current current interaction, which would dominate elementary particle physics for decades to come, were all present in Fermi’s theory. Truly the starting point of a great saga in physics of the twentieth century.

Two more gems of his pedagogical skills deserve mention. The first is a book on ‘Thermodynamics’ based on his 1936 Columbia lectures, published in 1937 and available as a Dover paperback. The second is the course in Nuclear Physics given by him at Chicago in 1949, written up by three of those who attended: Jay Orear, A H Rosenfeld and R A Schluter. It is amazing to see the range and depth of the topics covered, both theoretical and experimental, in such an authoritative manner.

Fermi died at the tragically young age of fifty three, on November 29, 1954. In Wigner’s words again:

‘The same simplicity and realism, which was manifest in Fermi’s scientific work, manifested itself also in his human relations. Although he never engaged in subtle analyses of personalities, he knew what he could expect of his friends and colleagues and he seldom went wrong in his estimates. On a heroic scale was his acceptance of death. ... He was so completely composed that it appeared superhuman.’
Of very few can it be said that ‘20th century physics would have been only a shadow of what it became’, in their absence. Enrico Fermi was certainly one of them.

Suggested Reading


Sir William Rowan Hamilton – Life, Achievements, Stature in Physics

N Mukunda

Against the background of the development of physics, and in particular of mechanics, over the centuries since Galileo and Newton, we describe the life and work of William Rowan Hamilton in the 19th century. The depth of his ideas which brought together the understanding of ray optics and classical mechanics, and the remarkable ways in which his work paved the way to the construction of quantum mechanics in the 20th century, are emphasized.

Introduction

Any student of science, indeed any well-educated person today, is aware that the foundations of modern science rest ultimately on the work and ideas of a small number of larger-than-life figures from the 15th century onwards – Nicolaus Copernicus (1473–1543), Johannes Kepler (1571–1630), Galileo Galilei (1564–1642) and Isaac Newton (1642–1727). Of course the scientific revolution was born out of a much larger social phenomenon – the Renaissance – and many more persons such as René Descartes, Christian Huygens... – were involved. But if one is asked to narrow the choice to as few as possible, these four would be the irreducible minimum. Their great books remain everlasting classics of the era of the birth of modern science – Copernicus’ *De revolutionibus orbium coelestium*; Kepler’s *Mysterium Cosmographicum* and *Astronomia nova*; Galilei’s *Il Saggiatore* and *Dialogue Concerning the Two Chief World Systems*; and finally Newton’s *Philosophiae Naturalis Principia Mathematica*, a culmination of this phase of the scientific revolution. Newton’s *Principia* was built upon and synthesized the earlier ideas of Kepler, Galileo and Descartes on the nature of motion. But of his three Laws of Motion, the third on the equality of action and reaction, and the Law of Universal Gravitation, were uniquely his own.

Among the various branches of modern science it is understandable that it was physics, and within physics the discipline of mechanics, that first achieved a secure mathematical foundation permitting a systematic growth and elaboration in the succeeding centuries. Over the course of the 18th century, the greatest contributors to mechanics were Leonhard Euler (1707–1783), Joseph Louis Lagrange (1736–1813), (known in Italy as Giuseppe Luigi Lagrangia, as they too would like to claim him as their own), and Pierre-Simon Laplace (1749–1827). Lagrange’s *Mécanique analytique* and Laplace’s *Mécanique céleste* decorate the 18th century as the *Principia* does the 17th.

Progress in the field of mechanics has been truly impressive over the centuries, with each individual’s contributions influenced by and influencing those of many others. Those who have so far been named above are only the most illustrious ones from that period that always spring to mind. The pattern of classical mechanics has served as a model for the other major areas in physics, such as electromagnetism, thermodynamics, statistical mechanics, and later the twin theories of relativity followed by quantum mechanics. In the growth of mechanics itself the interplay between physical ideas and mathematical formulations, with conceptions in the two reinforcing one another, has been of immense importance. From the physics standpoint, the field of optics has grown from the earliest times side-by-side with mechanics, with much give and take.

During the 19th century the greatest contributions to classical mechanics have been from William Rowan Hamilton (1805–1865), Carl Gustav Jacob Jacobi (1804–1851), and somewhat later Henri Poincaré (1854–1912) particularly in the realm of what is called ‘qualitative dynamics’. A profoundly critical account of mechanics as viewed from century’s end is Ernst Mach’s *The Science of Mechanics* published in 1883, which had a deep influence on Albert Einstein.

Here we celebrate the life and work of Hamilton, in particular the amazing circumstance that so many of his pioneering ideas proved crucial for the creation of quantum mechanics in the 20th century.

**A Brief Life Sketch**

Hamilton was born at midnight of 3–4 August 1805 in Dublin, Ireland, to Archibald Hamilton, a solicitor by profession, and Sarah Hutton Hamilton. (Their families were originally from England and Scotland.) He was the fourth of nine children. There were many distinguished scientists on his mother’s side, suggesting that his scientific genius came from her. While he was still in his teens, Hamilton’s mother and father passed away, in 1817 and 1819 respectively.

Hamilton displayed precocious and amazing gifts very early. He could read English by three; knew Greek, Latin and Hebrew by five; and by twelve the major European languages as well as Persian, Arabic, Sanskrit and Hindustani. His interest in mathematics was sparked off at age fifteen. By seventeen he had read both Newton’s *Principia* and Laplace’s *Mécanique céleste*.

In 1823 he entered Trinity College, Dublin, as an undergraduate, completing his studies there in 1827 and excelling in both science and classics. In 1824 he submitted a paper ‘On Caustics’ to the Royal Irish Academy. After receiving positive suggestions from (of course) a Committee, he revised and enlarged it by 1827, while still in college, to ‘A Theory of Systems of Rays’. This established his reputation. It was here that he presented his concept of characteristic functions in optics. Later there were three supplements, in 1828, 1830 and 1832.

Already in 1827 he was appointed Andrews Professor of Astronomy at Trinity College, and Director of the Dunsink Observatory, five miles away from the centre of Dublin, where he lived for the rest of his life. Designated also as the Astronomer Royal of Ireland, he was spared observational duties, and was left free to concentrate on theoretical work.
Sir William Rowan Hamilton – Life, Achievements, Stature in Physics

Hamilton experienced two early disappointments in romantic relationships – with Catherine Disney in 1824 who married instead a wealthy clergyman fifteen years her senior, but who retained a relationship and correspondence with Hamilton for many years; and with Ellen de Vere around 1830, who however felt she could ‘not live happily anywhere but at Curragh’. Finally he married Helen Maria Bayly in 1833. They had two sons and a daughter, but the marriage was an unhappy one.

The period 1827 to 1834–1835 comprises his ‘sunshine years’, when he did his finest and most creative work in mathematical physics. He also became close to the poets William Wordsworth and Samuel Taylor Coleridge in England. In 1835 he was knighted, and in 1836 became President of the Royal Irish Academy.

The range and depth of Hamilton’s work are staggering. After the initial work on geometrical optics, based on Fermat’s Principle and properties of systems or bundles of rays, he turned to the analogous ideas in dynamics. The highlights here are his version of the Action Principle, and a new and amazingly fruitful form for the equations of motion. The physical content is the same as in Newton’s original equations of motion, or in the Euler–Lagrange version of them, but the mathematical shape given to them proved unbelievably powerful. The optico–mechanical analogy is an expression of the similarities in his treatments of geometrical optics and Newtonian particle dynamics. The theory of the Hamilton–Jacobi equations is an especially elegant part of this work. The discovery of quaternions came in 1843; this was more purely mathematical in character, though Hamilton had hopes it could be applied in physics.

The last phase of his life was unfortunately a disaster. Hamilton became an alcoholic, and his home and lifestyle disintegrated completely. Shortly before his death he learned that he had been elected the first Foreign Associate of the newly established United States National Academy of Science. He died on 2nd September 1865 in Dublin, from a severe attack of gout. He had lived all his life in Dublin.

Next, in briefly describing his major achievements and discoveries, the separation into different areas is only for convenience in presentation. There are deep interconnections between them, as must have existed in his own mind. Again for convenience we will use notations and terminology familiar to students today, so that his ideas can be more readily appreciated, and will not strictly follow the chronological order.

The Action Principle

One of the earliest ‘minimum’ or ‘extremum’ or ‘variational’ principles in the era of modern science is Fermat’s principle of least time in optics, formulated in 1657. As given by Cornelius Lanczos in his 1949 book The Variational Principles of Mechanics,

“The path of a light ray is distinguished by the property that if light travels from one given point M to another given point N, it does so in the smallest possible time.”
Thus, in a transparent medium with variable refractive index \( n(x) \), and denoting \( M, N \) by \( x_1, x_2 \) respectively, the ray follows a path \( x(s) \) determined by

\[
\delta \int_{x_1}^{x_2} n(x(s)) ds = 0, \quad x(s_1) = x_1, \quad x(s_2) = x_2.
\] (1)

We assume an isotropic medium, with refractive index dependent only on position and not on direction.

Here the parameter \( s \) measures path length, \( \left( \frac{dx(s)}{ds} \right)^2 = 1 \), and the variation \( \delta x(s) \) of \( x(s) \) leaves the end points fixed.

Probably the earliest analogous idea in mechanics was expressed in 1747 in a somewhat imprecise form by Maupertuis. From Sommerfeld’s Mechanics,

“Among all possible motions, Nature chooses that which reaches its goal with the minimum expenditure of action.”

Sommerfeld adds:

“This statement of the principle of least action may sound somewhat vague, but is completely in keeping with the form given to it by its discoverer.”

Many others contributed to this stream of thought – d’Alembert with his Principle of Virtual Work, then Euler and Lagrange who succeeded in making Maupertuis’ idea more precise. They identified ‘action’ as the time integral of an expression which in the notation familiar today is \( p_j \frac{dq^j}{dt} \) – the \( q^j \) are coordinates and the \( p_j \) are corresponding momenta. However in their formulation it was understood that both the actual and imagined varied motions are energy conserving: \( \delta E = 0 \). Then in 1834–35, Hamilton gave a new and much more flexible version of the Principle in which the condition \( \delta E = 0 \) was completely avoided. For conservative systems subject only to holonomic constraints and where forces are derivable from a potential function \( V(q) \), Hamilton’s Principle says that the actual motion is such that variations about it obey:

\[
\delta \int_{t_1}^{t_2} L(q, \dot{q}) dt = 0,
\]

\[
L(q, \dot{q}) = T(q, \dot{q}) - V(q),
\]

\[
T(q, \dot{q}) = \text{kinetic energy}.
\] (2)

The dot here denotes the derivative with respect to time.

It is understood that the variations \( \delta q^j(t) \) vanish at the terminal times \( t_1 \) and \( t_2 \), but are otherwise unrestricted; and of course \( \delta t = 0 \) in between, i.e., ‘time is not varied’. Thus the actual motion of the system in time is directly characterized, and one immediately obtains from (2) the Euler–Lagrange differential equations of motion (EOM) in time:

\[
\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^j} - \frac{\partial L}{\partial q^j} = 0.
\] (3)

So all the EOM are determined by one single function \( L \). It was in fact Hamilton who gave the name ‘Lagrangian’ to the integrand in the definition of action above. In this extremely flexible
Sir William Rowan Hamilton – Life, Achievements, Stature in Physics

and convenient form, Hamilton’s Principle continues to be used in fundamental physics to this day.

Soon after this work of Hamilton, in the years 1834–37 Jacobi refined and completed this formulation of mechanics, and in 1842–43 gave his celebrated ‘Lectures on Dynamics’ in Konigsberg, where Hamilton’s ideas were carried much further. However, Jacobi dwelt more on mechanics than on optics. (These lectures in English translation have recently been published by Hindustan Book Agency, Delhi.) In particular, Jacobi gave yet another formulation of the Principle of Least Action, in which however the dependence of coordinates on time was set aside, and instead, the path of the system in its configuration space was viewed purely geometrically. This version however has not been very fruitful in later times, though in spirit it is close to Fermat’s Principle in optics where again time plays no essential role.

In Feynman’s well-known three – volume Lectures on Physics, Lecture 19 in Volume 2 is a beautiful account of Hamilton’s Principle. As Feynman recalls, he had always been fascinated by it, ever since the time in school when his teacher Mr Bader described it to him one day after school hours.

**The Canonical Equations of Motion**

The Euler–Lagrange EOM (3) can be presented in an interesting form:

\[ p_j = \frac{\partial L(q, \dot{q})}{\partial \dot{q}_j}, \quad \dot{p}_j = \frac{\partial L(q, \dot{q})}{\partial q_j}. \] (4)

We regard the first set of equations as definitions of (canonical) momenta, and the second set as the ‘true’ EOM in the Newtonian sense. From the physical point of view, then, the content of the EOM (4) (or (3)) is the same as of Newton’s EOM; however the former are invariant in form under any change in the choice of the (generalized) coordinates \( q^j \). This is also evident from Hamilton’s Principle (2). Such changes in \( q^j \) are called point transformations, and we have the freedom to choose the coordinates to suit the analysis of a given system.

In 1835, Hamilton went one step further and cast the EOM in a remarkably symmetrical form. Treating the \( q^j \) and the \( p_j \) as independent variables, and defining a function \( H(q, p) \) as the Legendre transform of the Lagrangian,

\[ H(q, p) = p_j \dot{q}^j - L(q, \dot{q}), \] (5)

he obtained the system

\[ \dot{q}^j = \frac{\partial H(q, p)}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial H(q, p)}{\partial q_j}. \] (6)

These are Hamilton’s canonical EOM, they form the basis of his treatment of mechanics. For nonrelativistic systems with kinetic energy quadratic in the velocities, \( H(q, p) \) is the total energy. Equations of this form were apparently used by Lagrange in 1809 in the context of perturbation theory, however it is Hamilton who made them the foundation of dynamics. In his
honour, the function \( H(q, p) \) is called the Hamiltonian of the system. To avoid misunderstanding, we should mention that we implicitly assume that we are able to express the Lagrangian velocities as functions of the momenta defined in equation (4), and the coordinates. This means that the Lagrangian is nonsingular in the Dirac sense; for singular Lagrangians a comprehensive extension of Hamiltonian dynamics has been created by Dirac.

The mathematical space on which Hamilton’s Principle (2) and the Euler–Lagrange EOM (3) are formulated is the configuration space \( Q \) for which \( q^j \) are independent coordinates. For the canonical EOM (6) however, the basic space or ‘carrier space’ is the ‘phase space’ whose dimension is twice that of \( Q \) and for which the \( q^j \) and \( p_j \) together are independent coordinates. It is sometimes said that the construction of the phase space, for a given configuration space \( Q \), is amongst Hamilton’s profoundest discoveries. It turns out that the passage from configuration space \( Q \) to its associated phase space is intrinsic and independent of, indeed prior to, the choice of Lagrangian.

Hamilton’s treatment of the variables \( q \) and \( p \) ‘on the same footing’ leads to a kind of symmetry between them which is profoundly different from the more familiar rotational symmetry among Cartesian spatial coordinates alone coming from Euclidean geometry.

As we said above, the Euler–Lagrange EOM (3) are preserved in form under all point transformations on \( Q \). This then remains true for the canonical EOM (6) as well. However Jacobi then showed that the latter EOM preserve their form – their Hamiltonian form – under an immeasurably larger group of transformations on phase space, the group of all so-called canonical transformations. Point transformations form a (very tiny!) subgroup of the group of all canonical transformations. These transformations were not known to Hamilton, and it is to Jacobi that we owe our understanding of their properties and descriptions.

Going beyond this, in a given system of phase space variables \( q^j, p_j \), the solutions in time of EOM of the Hamiltonian form (6) describe the gradual and continuous unfolding of a family of canonical transformations. Thus, the equations expressing \( q^j(t_2), p_j(t_2) \) at a later time \( t_2 \) in terms of \( q^j(t_1), p_j(t_1) \) at an earlier time \( t_1 \) constitute a canonical transformation dependent on \( t_1, t_2 \) and the system Hamiltonian.

Thus, of the three physically equivalent forms of classical dynamics – the Newtonian, the Euler–Lagrange, and the Hamiltonian forms – it is the last that is the most sophisticated from the structural mathematical point of view. Its profound importance has been further brought out by the circumstance that in both initial forms of quantum mechanics – Heisenberg’s matrix mechanics of 1925 and Schrödinger’s wave mechanics of 1926 – Hamilton’s work is used as the starting point to go from classical to quantum mechanics. Heisenberg’s EOM for quantum mechanics are based on (6), converting them into first order differential equations in time for (non commuting) operators. The Schrödinger wave equation starts from the classical Hamiltonian \( H(q, p) \), constructs a linear operator out of it capable of acting on complex wave functions \( \psi(q) \) by the rule of replacement \( p \rightarrow -i\hbar \frac{\partial}{\partial q} \), and then sets up the time dependent Schrödinger wave equation. Many years after Feynman learned about Hamilton’s Principle of Least Action from Mr Bader, he discovered a third form of quantum mechanics called the Path Integral form. This was based on work by Dirac in 1934 on the role of the Lagrangian in quantum mechanics,
Sir William Rowan Hamilton – Life, Achievements, Stature in Physics

and in it Feynman made essential use of the action as defined by Hamilton in equation (2). It is awe inspiring to realise that Hamilton’s work around 1834–35 played such a crucial role in the birth of quantum mechanics nine decades later! Yet another aspect of this will be described below.

Optico-mechanical analogy and the Hamilton–Jacobi Equations

The presence of analogies between the descriptions of optical and mechanical phenomena was sensed quite early, by John Bernoulli and Maupertuis. As we have seen, Hamilton’s work (R Nityananda, Hamilton’s Optics: The Power of Wavefronts, Resonance, Vol. 21, No. 6, pp. 511–528, 2016.) on geometrical optics preceded his work on mechanics. There were four memoirs on optics, in the years 1827, 1828, 1830 and 1832, in which he developed the idea of characteristic functions as a basis for geometrical optics, based on Fermat’s Principle (1). (The quite remarkable prediction of conical refraction came in the fourth memoir.) Even though it may be a bit demanding, we sketch these ideas here.

Fermat’s Principle (1) leads to a system of second order ordinary differential equations for rays in geometrical optics:

\[
\begin{align*}
\ddot{x}_j &= (\delta_{jk} - \dot{x}_j \dot{x}_k) \partial_k n(x), \\
& \dot{x}_j \dot{x}_j = 1, \\
& \dot{x}_j \dot{x}_j = 0, \\
& j, k = 1, 2, 3. \\
\end{align*}
\]

(7)

Now the dots signify derivatives with respect to distance s along the ray, not with respect to time.

For any choice of ‘initial data’ \(x_j(s_1), \dot{x}_j(s_1)\) we get one definite ray \(x(s)\). Alternatively, we may choose \(x_1 = x(s_1)\) and \(x_2\) independently; then in the generic case we get a ray \(x(s)\) and a value for \(s_2\) such that \(x(s_2) = x_2\). (We cannot choose \(s_2\) independently since \(\dot{x}(s)^2 = 1\) must be obeyed!)

The point characteristic of Hamilton is then defined as

\[
V(x_1, x_2) = \int_{s_1}^{s_2} ds \ n(x(s)), \ x(s_1) = x_1, x(s_2) = x_2.
\]

(8)

This determines the ‘directional’ properties of the ray from \(x_1\) to \(x_2\):

\[
\nabla_{x_2} V(x_1, x_2) = n(x_2) \dot{x}_2, \nabla_{x_1} V(x_1, x_2) = -n(x_1) \dot{x}_1.
\]

(9)

These are the basic equations of Hamiltonian optics, at the level of single rays.

Other kinds of characteristic functions in optics were defined by Hamilton as various Legendre transforms of the point characteristic \(V(x_1, x_2)\).
Next we come to systems or families of rays, families of solutions to (7), built up in a very special manner. While (7) is a system of ordinary differential equations for \( x(s) \), now we have to deal with a partial differential equation for a real function \( S(x) \) on space:

\[
|\nabla S(x)| = n(x). \tag{10}
\]

(Many years later this equation acquired the name ‘eikonal equation’, and \( S(x) \) the ‘eikonal’).

For a given \( S(x) \), the choice of initial data

\[
x(s_1) = x_1, \quad \dot{x}(s_1) = \frac{1}{n(x_1)}(\nabla S(x))_{x=x_1} \tag{11}
\]

leads via (7) to a definite ray \( x(s) \). Analysis shows that, since \( S(x) \) obeys (10), \( x(s) \) is determined by a system of first order ordinary differential equations:

\[
n(x(s))\dot{x}(s) = (\nabla S(x))_{x=x(s)}, \quad x(s_1) = x_1. \tag{12}
\]

Allowing the initial point \( x_1 \) to vary over a suitably chosen two-dimensional ‘transverse region’ in physical space, for example a surface \( S(x) = \) constant, we get a family or bundle of rays determined by \( S(x) \), filling out some three-dimensional region in space.

In this picture, wave fronts are defined as surfaces of constant \( S(x) \), while rays (within the family!) are trajectories orthogonal to the wave fronts. Again within the family corresponding to \( S(x) \), Hamilton’s point characteristic (8) becomes

\[
V(x_1, x_2) = S(x_2) - S(x_1). \tag{13}
\]

Hamilton then transferred these ideas to mechanics. Now time enters as an independent variable, leading to some important changes. The discussion becomes ‘dynamical’, and not simply geometrical as in optics.

The canonical EOM (6) can be solved for given definite data at an initial time \( t_1 \), leading to a definite phase space trajectory:

\[
q^i(t_1), p_j(t_1) \text{ at } t_1 \to q^i(t), p_j(t) \text{ for } t \geq t_1. \tag{14}
\]

Alternatively we may specify a solution by giving some initial and some final data:

\[
t_1, q^i(t_1), t_2, q^i(t_2) \to q^i(t), p_j(t) \text{ for } t_1 \leq t \leq t_2. \tag{15}
\]

It now turns out that the action appearing in Hamilton’s Principle (2), evaluated for the solution (15) of the canonical EOM, is the analogue of the point characteristic \( V(x_1, x_2) \) in (8) in optics:

\[
S(q(t_1), t_1; q(t_2), t_2) = \int_{t_1}^{t_2} dt \, L(q(t), \dot{q}(t)); \tag{16a}
\]

along trajectory
This is called Hamilton's Principal function. It acts in the sense of (16b), as the Generating Function for the canonical transformation connecting \( q_j(t_2), p_j(t_2) \) to \( q_j(t_1), p_j(t_1) \). That every canonical transformation can be described in this way (or a variant thereof) via a Generating Function was shown by Jacobi. We should also mention a beautiful extension of Jacobi's result by Constantin Carathéodory many years later.

This function obeys two partial differential equations with respect to the time variables:

\[
\begin{align*}
\frac{\partial S}{\partial t_2} + H\left(q(t_2), \frac{\partial S}{\partial q(t_2)}\right) &= 0, \\
\frac{\partial S}{\partial t_1} - H\left(q(t_1), \frac{\partial S}{\partial q(t_1)}\right) &= 0,
\end{align*}
\]

These are the famous Hamilton–Jacobi (H–J) partial differential equations of mechanics.

The general H–J problem is the search for a solution \( S(q, t) \) to the partial differential equation

\[
\frac{\partial S(q, t)}{\partial t} + H\left(q, \frac{\partial S(q, t)}{\partial q}\right) = 0, \quad \text{given } S(q, t_1) = S_1(q)
\]

presented as an initial value problem. This is the mechanics analogue to (10) in optics. The solution to (18) is called a Hamilton Principal function. The link to phase space trajectories obeying the canonical EOM (6) is as follows: Since \( S(q, t) \) obeys (18), the canonical EOM simplify to just the set

\[
\frac{dq^i}{dt} = \left(\frac{\partial H(q, p)}{\partial p_j}\right)_{p = \frac{\partial S(q)}{\partial q}}, \quad q = q(t).
\]

Thus \( S(q, t) \) determines a special family of phase space trajectories, corresponding to the selected set of initial conditions \( \left(q^i, p_j = \frac{\partial S(q)}{\partial q^i}\right) \) at time \( t_1 \), with \( q \) varying over configuration space \( Q \).

Let us mention that since we have (implicitly) assumed that \( H(q, p) \) has no explicit time dependence, there is a time independent version of the H–J equation, whose solutions are called Hamilton characteristic functions.

The Hamilton–Jacobi equations appear also in the semi-classical approximation to the Schrödinger wave equation of quantum mechanics.

It is widely felt that the Hamilton–Jacobi equation is the most beautiful form of classical mechanics. (Remember though that it is a single partial differential equation, unlike the Euler–Lagrange or Hamiltonian EOM (3,4,6)). At the start of his discussion of this equation, Lanczos places this quotation from the old testament:
‘Put off thy shoes from off thy feet, for the place whereon thou standest is holy ground’ (Exodus III, 5)

Hamilton’s optico-mechanical analogy was a great source of inspiration to Schrödinger in the creation of wave mechanics. As for the construction of special families of rays in optics or of phase space trajectories in mechanics we must quote an eloquent passage from Paul Dirac. In his paper titled ‘The Hamiltonian form of field dynamics’ (Canadian Journal of Mathematics, Vol. 3, p. 1, 1951) he says:

‘In classical dynamics one has usually supposed that when one has solved the equations of motion one has done everything worth doing. However, with the further insight into general dynamical theory which has been provided by the discovery of quantum mechanics, one is led to believe that this is not the case. It seems that there is some further work to be done, namely to group the solutions into families (each family corresponding to one principal function satisfying the Hamilton–Jacobi equation). The family does not have any importance from the point of view of Newtonian mechanics; but it is a family which corresponds to one state of motion in the quantum mechanics, so presumably the family has some deep significance in nature, not yet properly understood’.

Quaternions

Hamilton’s creation of the algebra of quaternions came at a later stage in his life, after the work in optics and mechanics. Complex numbers were known, and could be represented on a Euclidean plane. Hamilton’s idea was to extend this to higher spatial dimensions, so as to be applicable in physics. After much effort he finally succeeded in 1843, though now it is known that Benjamin Olinde Rodrigues had found essentially the same results in 1840. In place of the single pure imaginary unit \( i \) for complex numbers, quaternions are built using three such units \( i, j, \) and \( k \). Their essential algebraic or composition properties became suddenly clear to him during a walk with his wife along the Royal Canal from Dunsink to a meeting of the Royal Irish Academy on 16th October 1843:

\[
\begin{align*}
  i^2 &= j^2 = k^2 = ijk = -1. 
\end{align*}
\]  

He is said to have immediately carved these formulae with his penknife on the stone of Broome Bridge as he passed it. In a letter to Peter Guthrie Tait many years later, on 15th October 1858, he described what had happened in these words:

‘I then and there felt the galvanic circuit of thought close, and the sparks which fell from it were the fundamental equations between \( i, j, k \); exactly such as I have used them ever since.’

Indeed he went on to claim: ‘I still must assert that this discovery appears to me to be as important for the middle of the nineteenth century as the discovery of fluxions (the calculus) was for the close of the seventeenth’.

In 1958 on behalf of the Royal Irish Academy the Irish mathematician-political leader Eamon de Valera unveiled a plaque under Broome Bridge to commemorate the event.

Hamilton wrote extensively on quaternions, his ‘Lectures on Quaternions’ being published in 1853. They are the first example within mathematics of algebraic quantities obeying a non-commutative but associative rule of multiplication. Recall here that physical quantities are represented in quantum mechanics by generally non commuting operators. One little gem in this book is Hamilton’s construction of a pictorial representation for elements of the group SU(2) and their (non commutative) composition law, using great circle arcs on a sphere $S^2$. These ‘turns’ of Hamilton have been used in physics in recent times.

However, quaternions did not prove as useful for physics as Hamilton had hoped. The methods of vector algebra and vector calculus later pioneered by Oliver Heaviside and Josiah Willard Gibbs turned out more useful and capable of extension to any number of dimensions.

**Concluding Remarks**

This account has hopefully succeeded in conveying to the reader the depth and profound beauty of Hamilton’s contributions to theoretical physics. His exalted status among the greatest physicists of all time rests on his magnificent achievements. There is no better way to conclude than by quoting from two of his countrymen, one belonging to his times and the other from a few generations later:

‘Hamilton was gifted with a rare combination of those qualities which are essential instruments of discovery. He had a fine perception by which the investigator is guided in his passage from the known to the unknown... But he seems, also, to have possessed a higher power of divination – an intuitive perception that new truths lay in a particular direction, and that patient and systematic search, carried on within definite limits, must certainly be rewarded by the discovery of a path leading into regions hitherto unexplored...’

– Memorial address, November 1865, by Charles Graves, President Royal Irish Academy.

‘Hamilton lived in the heroic age of mathematics... It was not Hamilton’s ambition to polish the corners of structures built by other men. Newton had the theory of gravitation and planetary motions, Lagrange had his dynamical equations, Laplace had the theory of potential. What monument would Hamilton create to make his memory imperishable? Hamilton realised that optics and dynamics are essentially a single mathematical subject. He was able to characterise or describe any optical or dynamical system by means of a single characteristic or principal function... Hamilton liked to refer to himself in the words Ptolemy used of Hipparchus: a lover of labour and a lover of truth...’

—John Lighton Synge, 1943.
N Mukunda

Suggested Reading


General Articles
The Story of the Photon*

N Mukunda

An account of the story of the light quantum or photon is given, from its inception in 1905 to its final acceptance in 1924. Necessary background information on radiation theory and historical details are included.

Introduction

The photon, so named by the physical chemist Gilbert Norton Lewis in 1926, is a child of the 20th century. It is the ‘particle of light’ – or ‘light quantum’ – first hypothesized by Albert Einstein in 1905, and then used by him to explain, among other things, the photoelectric effect. The story of the photon is rich in history, development of ideas, experiment and personalities. In this account an attempt will be made to convey something of each of these aspects; the fundamental motivations and currents of ideas will be described as carefully as possible, and only selected derivations will be presented.

During the year 1905, aptly called ‘Einstein’s Miraculous Year’, he submitted five research papers for publication and also completed his Ph.D. thesis. Of the former, three have become all-time classics. In chronological sequence they are: the light quantum paper (March), the paper on the Brownian Motion (May), and the paper establishing the Special Theory of Relativity (June). Einstein himself felt that of these only the first was truly path-breaking, for he wrote in a letter of May 1905 to his friend Conrad Habicht: “I promise you four papers ...... the first of which I could send you soon.... The paper deals with radiation and the energetic properties of light and is very revolutionary, as you will see. ....”.

Radiation Theory from Kirchoff to Planck – a Capsule

The study of (electromagnetic) radiation forms a glorious chapter in the history of physics. The first major step was taken in 1859 by Gustav Kirchoff (the ‘grandfather’ of the quantum theory) when he proved the following result: if radiation and material bodies are in equilibrium at a common (absolute) temperature \( T \), the former being reflected, scattered, absorbed and emitted by the latter, then the energy density of the radiation per unit frequency interval is a universal function of frequency and temperature, independent of the particular material bodies present:

\[
\rho(\nu, T)\Delta \nu = \text{energy of radiation per unit volume in the frequency range } \\
\nu \text{ to } \nu + \Delta \nu, \text{ at temperature } T \\
= (\text{universal function of } \nu \text{ and } T) \times \Delta \nu. \tag{1}
\]

N Mukunda

For the proof, Kirchoff used the Second Law of the then young science of thermodynamics; and he posed the determination and understanding of the function $\rho(\nu, T)$ as a major experimental and theoretical challenge. Such radiation is variously called ‘black-body’ or ‘temperature’ or ‘thermal’ radiation.

Twenty years later, in 1879, the experimentalist Josef Stefan measured the total energy density of thermal radiation by ‘summing’ over all frequencies, and then conjectured that it was proportional to $T^4$:

$$u(T) = \text{total energy density of thermal radiation}$$

$$= \int_0^\infty d\nu \rho(\nu, T) = \frac{\sigma}{c} T^4. \quad (2)$$

Soon after, in 1884, Ludwig Boltzmann was able to give a thermodynamic proof of this result, using Maxwell’s result that the pressure of radiation is one third of its energy density. (See Box 1.) Once again, this was an outstanding and early application of thermodynamics to radiation problems – more were to follow. The constant $\sigma$ in (2) is named jointly after Stefan and Boltzmann.

**Box 1. Thermodynamics and the Stefan – Boltzmann Law**

Consider thermal radiation, at temperature $T$, enclosed in a spatial volume $V$, and treat $T$ and $V$ as independent variables. The total energy $U = V u(T)$ where $u(T)$ is the energy density including all frequencies. The pressure, according to Maxwell, is one third the energy density: $p = \frac{u}{V} = \frac{u(T)}{3}$. (In contrast, for a classical (nonrelativistic) ideal gas of $n$ particles the total energy $U = \frac{1}{2}nkT$ is volume independent; while from the ideal gas law the pressure is two-thirds the energy density, $p = \frac{2}{3} \frac{U}{V}$.) The Second Law of Thermodynamics implies that the expression

$$dS = \frac{1}{T}(dU + pdV)$$

must be a perfect differential. Writing this out as

$$dS = \frac{1}{T} \left( u(T) dV + V \frac{du(T)}{dT}dT + \frac{u(T)}{3} dV \right),$$

this means that

$$\frac{\partial}{\partial T} \left( \frac{4u(T)}{3T} \right) = \frac{\partial}{\partial V} \left( \frac{V}{T} \frac{du(T)}{dT} \right),$$

which simplifies to

$$T \frac{du(T)}{dT} = 4 u(T).$$

The solution is the Stefan–Boltzmann Law:

$$u(T) = \text{Constant} \times T^4.$$
The Story of the Photon

From the 1860’s onwards many guesses were made for the form of the function \( \rho(\nu, T) \). In 1893 Wilhelm Wien constructed a clever thermodynamical argument and proved that the dependences of \( \rho(\nu, T) \) on its two arguments were correlated by a scaling law:

\[
\rho(\nu, T) = \nu^3 f(\nu/T),
\]

so the original Kirchoff problem became that of finding the form of the universal function \( f(\nu/T) \) involving only one argument. He followed this up soon after in 1896 by offering a guess for the form of \( f(\nu/T) \), inspired by the Maxwell velocity distribution in a classical ideal gas: with two constants \( \alpha \) and \( \beta \) he suggested

\[
f(\nu/T) = \alpha e^{-\beta \nu/T}, \\
\rho(\nu, T) = \alpha \nu^3 e^{-\beta \nu/T}.
\]

Early experiments by Friedrich Paschen (reported in January 1897) gave support to the Wien Law (4). They were done in the near infrared part of the spectrum, with wavelengths \( \lambda \) in the range \((1 \text{ to } 8) \times 10^4 \text{ Å}\) and temperatures \( T \) in the range 400 to 1600 K; and showed the validity of the Wien Law in the high frequency limit.

Now we turn to Max Karl Ernst Ludwig Planck, successor to Kirchoff and the ‘father’ of the quantum theory. His major goal was the theoretical determination of Kirchoff’s universal function \( \rho(\nu, T) \). For a while he believed that the Wien Law (4) was correct for all \( \nu \) and was the answer to Kirchoff’s problem; his task was to find a proper theoretical basis for that law. In the 1890’s he carried out many fundamental investigations on the interaction of Maxwell’s electromagnetic waves with matter; he was a master of thermodynamics as well. However during 1900 new experiments showed deviations from the Wien Law (4) in the low frequency limit, and there were new theoretical developments as well. In February 1900 the experiments of Otto Lummer and Ernst Pringsheim in the far infrared region \( \lambda = (1.2 \text{ to } 1.8) \times 10^5 \text{ Å} \) and \( T = 300 \text{ to } 1650 \text{ K} \) showed disagreement with the Wien Law (4). In June 1900 Lord Rayleigh applied the equipartition theorem of classical statistical mechanics to thermal radiation treated as a system on its own and derived the result

\[
f(\nu/T) = c_1 T/\nu, \\
\rho(\nu, T) = c_1 \nu^2 T, \ c_1 \text{ a constant.}
\]

(\text{After further work by Rayleigh in May 1905 calculating } c_1 \text{ and a later correction by James Hopwood Jeans in June-July 1905, this Rayleigh–Jeans Law attained its final exact form}

\[
f(\nu/T) = \frac{8\pi k}{c^3} \cdot \frac{T}{\nu}, \\
\rho(\nu, T) = \frac{8\pi \nu^2}{c^3} \cdot k T,
\]

with \( c \) the vacuum speed of light and \( k \) the Boltzmann constant). Slightly later, by October 1900, Heinrich Rubens and Ferdinand Kurlbaum did experiments in the deep infrared, \( \lambda = \)

79
(3 to 6) × 10^5 Å, \( T = 200 \) to 1500K, and found again deviations from the Wien Law (4) but agreement with the Rayleigh expression (5).

Sunday, October 7, 1900 is the birthdate of the quantum theory. On the afternoon of that day, Rubens visited Planck’s home for tea, and told him of his and Kurlbaum’s latest experimental results. After he left, Planck set to work. He realised that Wien’s Law could not be the final answer to Kirchoff’s problem. While it was obeyed at high enough frequencies, it failed at the low frequency end where the Rayleigh form was valid. What Planck achieved that evening was a mathematical interpolation between these two limiting forms. His strategy seems round-about but was, in retrospect, fortunate. He had in earlier work related the Kirchoff function \( \rho(\nu, T) \) to the average energy \( \bar{E}(\nu, T) \) of a charged material oscillator with natural frequency \( \nu \) and at a temperature \( T \), by balancing the effect on it of incident radiation and its own emission of radiation. This ‘Planck link’ reads

\[
\rho(\nu, T) = \frac{8\pi\nu^2}{c^3} \bar{E}(\nu, T). \tag{7}
\]

Planck translated the limiting forms of \( \rho(\nu, T) \) in the high \( \nu \) (Wien) and low \( \nu \) (Rayleigh) limits into corresponding limiting forms for \( \bar{E}(\nu, T) \); converted this into limiting forms for the entropy \( S(\bar{E}) \) of the material oscillator (written as a function of energy) at high and low \( \bar{E} \), respectively; and then by solving a simple differential equation found a formula interpolating between these limiting expressions. Translating all this back into the original problem his result for Kirchoff’s function \( \rho(\nu, T) \) is the Planck radiation law we all know so well:

\[
\rho(\nu, T) = \frac{8\pi\nu^2}{c^3} \cdot \frac{h\nu}{e^{h\nu/kT} - 1}. \tag{8}
\]

A new fundamental constant of nature with the dimensions of action, Planck’s constant \( h \), entered his result; and there was agreement with experiment at all measured frequencies. On October 19, 1900, Planck announced his formula following a talk given by Kurlbaum. In the high frequency limit we recover the Wien result (4) from (8) with

\[
\alpha = \frac{8\pi h}{c^3}, \quad \beta = h/k. \tag{9}
\]

Comparing (7) and (8) it follows that Planck’s formula implies that the average energy of a material oscillator \( \bar{E}(\nu, T) \) must have a value differing from the result \( kT \) of the equipartition theorem:

\[
\bar{E}(\nu, T) = \frac{h\nu}{e^{h\nu/kT} - 1}. \tag{10}
\]

During the period October to December 1900 Planck tried very hard to find a theoretical basis for this formula. Finally, “...as an act of desperation,... to obtain a positive result, under any circumstances and at whatever cost”, he invented the concept of irreducible packets or quanta of energy for matter, and in mid-December 1900 he presented the following statistical derivation of (10). He imagined a large number, \( N \), of identical (but distinguishable!) material oscillators,
with a total energy $E$ and at a temperature $T$. Assuming that this total energy $E$ was made up of $P$ (indistinguishable!) packets or quanta of energy $\epsilon_0$ each, (so that $E = PE_0$ and the energy of each oscillator is an integer multiple of $\epsilon_0$), he counted the number of ways $W$ (number of micro states or complexions) in which these packets could be distributed over the $N$ oscillators. By a simple combinatorial argument, followed by an application of the Boltzmann entropy relation $S = k \ln W$, he computed the entropy $S/N$ per material oscillator, connected it up to the temperature $T$, and finally arrived at the result (10) he was after, with the identification $\epsilon_0 = h\nu$.∗

Einstein’s State of Preparedness

It is time now to turn to Einstein. Already since 1897 during his student days at the Eidgenossische Technische Hochschule in Zurich he had become familiar with Kirchhoff’s work on thermal radiation. From his teacher Heinrich Friedrich Weber in 1899 he learnt about Wien’s theorem (3) and the resulting Wien Displacement Law. He was also familiar with Planck’s work, and while he had full faith in the experimental validity of the Planck law (8), he was acutely conscious of the absence of a proper theoretical basis for it. (See Box 2 for a brief account of Einstein’s involvement with Planck’s Law). During the period 1902–1904 he rediscovered for himself the foundations and key concepts of statistical physics, obtaining independently many of Josiah Willard Gibbs’ results. He invented on his own the concept of the canonical ensemble, derived the equipartition law for energy, found ways to use the ‘Boltzmann Principle’ $S = k \ln W$, and found the formula for energy fluctuations for a mechanical system at a given temperature. (See later.)

The empirical validity of the Planck Law (8) and the realisation that it could not be derived from the classical Maxwell theory of electromagnetic radiation convinced him that the picture

---

**Box 2: Einstein and the Planck Law**

Here is a chronological list of the many occasions and ways in which Einstein ‘played’ with the Planck radiation law and ‘teased out’ its deep consequences:

**1905:** Examines the volume dependence of entropy of radiation in the Wien limit, abstracts the light quantum idea, applies it inter alia to the photoelectric effect.

**1909:** Calculates energy fluctuations for thermal radiation using the complete Planck Law; arrives at the earliest ever statement of wave-particle duality in nature; considers also momentum fluctuations of a mirror placed in thermal radiation, due to fluctuations in radiation pressure.

**1916:** Derives the Planck Law based on Bohr’s theory of stationary states and transitions, and processes of absorption, induced and spontaneous emission of radiation by matter. Extends the 1905 analysis to show that individual light quanta are directed in space and carry momentum.

**1924–25:** Extends Bose’s derivation of the Planck Law to matter, finds particle-wave duality for matter, predicts Bose–Einstein condensation.

---

∗The symbol $k$ for Boltzmann’s constant first appeared in the Planck Law (8) in 1900. The formula $S = k \ln W$ was given the name ‘Boltzmann’s Principle’ by Einstein.
of radiation given by the latter had to be modified by incorporating quantum features in some way. As he was to say much later: “Already soon after 1900, i.e., shortly after Planck’s trailblazing work, it became clear to me that neither mechanics nor thermodynamics could (except in limiting cases) claim exact validity”.

Einstein independently derived, in his March 1905 paper, the Rayleigh–Jeans Law (6): he started from the ‘Planck link’ (7) between radiation and matter, used the equipartition law to substitute \( kT \) for the average energy \( \overline{E}(\nu, T) \) of the material oscillator, and directly obtained (6)! Thus there were two theoretically well-founded, but experimentally invalid, routes to the Rayleigh–Jeans result: one applying equipartition directly to radiation; and another using the ‘Planck link’ and then applying equipartition to the material oscillator.

Added to all this, it should be mentioned that in the course of some work on the molecular theory of gases done in 1904, Einstein had realised the importance of the volume dependence of thermodynamic quantities, in particular of the entropy. The relevance of this will become clear presently.

The ‘Light Quantum’ Paper of 1905

Einstein’s views, circa 1905, on the radiation problem may be summarised as follows: the Planck Law is experimentally accurate but has no proper theoretical basis; the Rayleigh–Jeans limit has a proper classical theoretical foundation but is experimentally unacceptable; the Wien limit is a guess, with no derivation from first principles or classical basis, and is experimentally valid only at high frequencies. He also declared right away that, in spite of the success of Maxwell’s wave theory in explaining typical optical phenomena, he believed it was necessary to replace it by a different picture in which radiant energy is made up of discontinuous spatially localized quanta of finite energy, which could be absorbed and emitted only as complete units.

Einstein then took a ‘phenomenological’ attitude to the radiation problem: since Wien’s Law (4) is experimentally valid in a definite domain and has no classical underpinnings, an examination of this domain from the thermodynamical point of view – involving radiation on its own and not using the ‘Planck link’ at all – should reveal key nonclassical features of radiation.

Apart from the independent derivation mentioned above of the Rayleigh–Jeans Law, in his paper Einstein recalls some results of Wien on the entropy of radiation. He then uses this to calculate the volume dependence of the entropy of thermal radiation in the Wien limit; gives the corresponding calculation for a classical ideal gas; compares the two results; and then draws his epoch-making conclusions about the existence and nature of radiation quanta. The Wien limit calculation given by Einstein is essentially equivalent to the following.

Consider thermal radiation at temperature \( T \) and between frequencies \( \nu \) and \( \nu + \Delta \nu \), contained in a spatial volume \( V \). The total energy, \( E \) say, of this radiation is given, when the Wien limit is applicable, by

\[ E = \frac{8 \pi k^4 T^4}{3 h^3} V \]
The Story of the Photon

\[ E = V \alpha v^3 e^{-\beta v/T} \cdot \Delta v = N \frac{V}{e^{-\beta v/T}}, \]
\[ N = \alpha v^3 \Delta v. \quad (11) \]

Treating \( E \) and \( V \) as the independent thermodynamic variables, the inverse temperature is
\[ \frac{1}{T} = \frac{1}{\beta v} (\ln N + \ln V - \ln E). \quad (12) \]

The entropy \( S(E, V) \) of this portion of Wien radiation is obtained by integrating the basic thermodynamic relation
\[ \frac{\partial S(E, V)}{\partial E} = \frac{1}{T} = \frac{1}{\beta v} (\ln N + \ln V - \ln E), \quad (13) \]
the dependences of \( S(E, V) \) on \( \nu \) and \( \Delta \nu \) being left implicit.
This leads to
\[ S(E, V) = \frac{E}{\beta v} (\ln N + \ln V + 1 - \ln E), \quad (14) \]
(apart from a function of \( V \) alone which must vanish since \( S(E, V) \rightarrow 0 \) as \( E \rightarrow 0 \)). If we now compare the values of the entropy for two different volumes \( V_1 \) and \( V_2 \), keeping \( E \) (and of course \( \nu \) and \( \Delta \nu \)) fixed, we find:
\[ S(E, V_1) - S(E, V_2) = \frac{E}{\beta v} \ln \left( \frac{V_1}{V_2} \right) = k \ln \left( \frac{V_1}{V_2} \right)^{E/\hbar \nu}, \quad (15) \]
where the value of the Wien constant \( \beta \) was taken from (9).

Einstein then follows up the derivation of the result (15) by a detailed calculation of a similar entropy difference for a classical ideal gas of \( n \) molecules. For this he exploits the ‘Boltzmann Principle’ \( S = k \ln W \) relating entropy to statistical probability; omitting the details of his argument, he arrives at the result **
\[ S(E, V_1) - S(E, V_2) = k \ln \left( \frac{V_1}{V_2} \right)^n \quad (16) \]

Comparison of the two results (15) and (16) leads to his profound conclusion:

**The entropy of a classical ideal gas of \( n \) particles has the form \( S(E, V) = nk(\ln V + 3/2\ln(2E/3nk)) \). While the volume dependence is similar to that in (14), the energy dependence is quite different.**
“...We (further) conclude that monochromatic radiation of low density (within the range of validity of Wien’s radiation formula) behaves thermodynamically as if it consisted of mutually independent energy quanta of magnitude $h\nu$.

(Einstein actually wrote $R\beta\nu/N$ for this last expression, which is just $h\nu$). Note carefully the explicit mention that this refers to radiation in the Wien limit; indeed the use of the complete Planck Law does not lead to such a result! Note also the conclusion that the energy quanta are mutually independent, reflecting the comparison being made to the classical ideal gas.

Thus was the concept of ‘light quanta’ first arrived at, with its stated limitations. Nevertheless, right away Einstein abstracts the key idea and boldly extrapolates it beyond these limitations to formulate his ‘heuristic principle’:

“If monochromatic radiation (of sufficiently low density) behaves, as concerns the dependence of its entropy on volume, as though the radiation were a discontinuous medium consisting of energy quanta of magnitude $h\nu$, then it seems reasonable to investigate whether the laws governing the emission and transformation of light are also constructed as if light consisted of such energy quanta”. Thus he proposes that in the processes of emission and absorption and interaction of light with matter, the same particulate nature should be seen!

Einstein concluded his paper by applying his ‘heuristic principle’ to three experimental observations: the Stokes rule in photo luminiscence, the photoelectric effect, and lastly the ionization of gases by ultraviolet light. We look next briefly at some highlights of the second of these applications.

The Photo-Electric Effect

This effect was discovered accidentally by Heinrich Hertz in 1887 while studying sparks generated by potential differences between metal surfaces. (Remember at that time the electron was not yet known!). After Joseph John Thomson discovered the electron in 1897, he turned to the photo electric effect and in 1899 could state that it was the electron that was ejected when ultraviolet light shone on a metal surface. In experiments around 1902 Philip Lenard studied the dependence of the ejected electron’s energy on the intensity and frequency of the incident radiation – independent of the former, increasing with the latter.

In his 1905 paper Einstein proposed the following ‘simplest conception’ for what happens: a light quantum transfers all its energy to a single electron, independent of other quanta present and disappearing in the process; the electron emerges from the metal surface carrying with it the photon’s energy except for what it has to ‘pay’ to leave the metal. He then proposed the following famous and simple equation (in modern notation) for the maximum energy of the emitted electron:

$$E_{\text{max}} = h\nu - P,$$

where $\nu$ is the frequency of incident radiation and $P$– the work function characteristic of the metal – the energy lost by the electron in the release process.
The Story of the Photon

The most extensive series of experiments to test (17) were carried out by Robert Andrews Millikan in the decade up to 1915, even though he was extremely skeptical about the light quantum hypothesis itself. In his 1915 paper he said: “Einstein’s photoelectric equation.... appears in every case to predict exactly the observed results.... Yet the semicorpuscular theory by which Einstein arrived at his equation seems at present wholly untenable”. Many years later, in 1949, he reminisced in these words: “I spent ten years of my life testing that 1905 equation of Einstein’s and contrary to all my expectations, I was compelled in 1915 to assert its unambiguous verification in spite of its unreasonableness, since it seemed to violate everything we knew about the interference of light”.

We discuss reasons for the widespread opposition to the photon idea later; let us conclude this section by quoting from the 1921 Physics Nobel Award citation to Einstein: “... for his services to theoretical physics and in particular for his discovery of the law of the photoelectric effect”.

Wave-Particle Duality, Photon Momentum

We saw that in 1905 Einstein worked only with the Wien limit of the Planck Law, not the latter in its entirety. In 1909 he went back to the Planck Law itself. As was mentioned earlier, in 1904 he had derived on his own the energy fluctuation formula on the basis of the canonical ensemble construction:

\[
(\Delta E)^2 \equiv \langle E^2 \rangle - \langle E \rangle^2 = k T^2 \frac{\partial}{\partial T} \langle E \rangle. \tag{18}
\]

(We take temperature \( T \) and volume \( V \) as the independent variables, and leave implicit the dependences of the average energy \( \langle E \rangle \) on these). Considering thermal radiation contained in the frequency range \( \nu \) to \( \nu + \Delta \nu \) and in a unit spatial volume, at temperature \( T \), the Planck Law gives:

\[
\langle E \rangle = \frac{8 \pi \nu^2}{c^3} \cdot \frac{\Delta \nu}{e^{\hbar \nu/kT} - 1},
\]

\[
(\Delta E)^2 = kT^2 \cdot \frac{8 \pi \hbar \nu^3 \Delta \nu}{c^3} \cdot \frac{e^{\hbar \nu/kT}}{(e^{\hbar \nu/kT} - 1)^2} \cdot \frac{\hbar \nu}{kT^2}
\]

\[
= \frac{8 \pi \hbar^2 \nu^4 \Delta \nu}{c^3} \left( \frac{1}{(e^{\hbar \nu/kT} - 1)^2} + \frac{1}{(e^{\hbar \nu/kT} - 1)} \right)
\]

\[
= \frac{c^3}{8 \pi \nu^2 \Delta \nu} \langle E \rangle^2 + \hbar \nu \langle E \rangle. \tag{19}
\]

At this point the reader is encouraged to check that if \( \langle E \rangle \) had been given purely by the Rayleigh–Jeans expression (6), only the first term on the right would have been obtained; while if \( \langle E \rangle \) was given solely by the Wien expression (4) only the second term on the right
would have appeared. Recalling that the Rayleigh–Jeans Law is the unambiguous result of classical Maxwell wave theory and the equipartition theorem, while the Wien Law led to the light quantum hypothesis, we see in the energy fluctuation formula (19) a synthesis or duality of wave and particle aspects of radiation. In Einstein's words: "...It is my opinion that the next phase in the development of theoretical physics will bring us a theory of light that can be interpreted as a kind of fusion of the wave and the emission theories... (The) wave structure and (the) quantum structure... are not to be considered as mutually incompatible...."

Fourteen years later, in 1923, Prince Louis Victor de Broglie would suggest a similar particle-wave duality for the electron.

The next time Einstein turned to the Planck Law was in 1916 when he gave a new derivation of it based on Bohr's 1913 theory of stationary states of atoms (and molecules) and transitions between them accompanied by emission or absorption of radiation. In his work, Einstein introduced the famous $A$ and $B$ coefficients characterising the interaction between matter and radiation, and corresponding to the three distinct processes of absorption, induced emission and spontaneous emission of radiation by matter. Planck's radiation law was shown to be the result of equilibrium among these processes, given Bohr's postulates and the Boltzmann distribution for the numbers of molecules in the various energy or stationary states. While we will not reproduce this beautiful work here, let us mention that at the same time Einstein completed his physical picture of the light quantum – not only was it a localized parcel of energy $h\nu$, it was directed and carried a momentum $\frac{h\nu}{c}$ in its direction of motion as well. (Initial steps in this direction had earlier been taken by Einstein in 1909, by considering the momentum fluctuations of a mirror immersed in thermal radiation, as a result of fluctuations in the radiation pressure.) This result was derived by carefully analysing both energy and momentum balances when a molecule makes a transition from one energy level to another via emission or absorption of radiation, and demanding stability of the Planck distribution for radiation on the one hand, and of the Boltzmann distribution for molecules on the other.

It is interesting to realise that it took the discoverer of special relativity from 1905 to 1916 to complete the picture of light quanta. Remember though that the creation of the General Theory of Relativity had kept him busy upto November 1915.

In any case, with this additional insight into the kinematical properties of the light quantum Einstein was fully convinced of its reality. In 1917 he wrote to Besso: "With that, (the existence of) light quanta is practically certain". And two years later: "I do not doubt any more the reality of radiation quanta, although I still stand quite alone in this conviction".

**Opposition to the Light Quantum – the Compton Effect**

Why was there such prolonged and widespread reluctance to accept the idea of light quanta? In the cases of the electron, proton and neutron, all of which were experimental discoveries, the concerned particles were quickly accepted into the body of physics. But it was indeed very different with the photon.
The Story of the Photon

One reason may have been Einstein’s own sense of caution which he expressed in 1911 in this way: “I insist on the provisional character of this concept (light quanta) which does not seem reconcilable with the experimentally verified consequences of the wave theory”. On several occasions people like Max von Laue, Arnold Sommerfeld and Millikan misinterpreted Einstein’s statements to mean that he had gone back on his ideas! Apart from that the main reason seems to have been a near universal feeling that Maxwell’s description of radiation should be retained as far as free radiation was concerned, and the quantum features should be looked for only in the interaction between matter and radiation. Indeed Planck said in 1907: “I am not seeking the meaning of the quantum of action (light-quantum) in the vacuum but rather in places where absorption and emission occur, and (I) assume that what happens in the vacuum is rigorously described by Maxwell’s equations”. And again in 1909: “I believe one should first try to move the whole difficulty of the quantum theory to the domain of the interaction between matter and radiation”. It is also amusing to see what Planck and others said in 1913 while proposing Einstein for election to the Prussian Academy of Sciences: “In sum, one can say that there is hardly one among the great problems in which modern physics is so rich to which Einstein has not made a remarkable contribution. That he may sometimes have missed the target in his speculations, as, for example, in his hypothesis of light quanta, cannot really be held too much against him, for it is not possible to introduce really new ideas even in the most exact sciences without sometimes taking a risk”.

The situation changed decisively only after the discovery of the Compton effect by Arthur Holly Compton in 1923. This is the scattering of a photon by a (nearly) free electron; the validity of the energy and momentum conservation laws convinced most skeptics of the reality of light quanta. The relation between the change in frequency of the photon and the scattering angle is very simply calculable in the photon picture, and agrees perfectly with experiment; classical explanations do not work. (Today in the language of quantum field theory we say the incident photon is annihilated and the final photon with different frequency and momentum gets created, while the electron continues to exist throughout). In a popular article in 1924 Einstein remarked: “The positive result of the Compton experiment proves that radiation behaves as if it consisted of discrete energy projectiles, not only in regard to energy transfer but also in regard to Stosswirkung (momentum transfer)”.

Except for one lone but important dissenter – Niels Henrik David Bohr. He continued to doubt the reality of light quanta, wanted to retain the Maxwellian picture of radiation, and to relegate quantum features exclusively to matter and not to radiation. As part of this line of thinking, in an important paper in 1924, Bohr and his coauthors Hendrik Anton Kramers and John Clarke Slater proposed giving up both causality and energy – momentum conservation in individual elementary processes, but retaining them only statistically. Fortunately these two ideas were experimentally tested right away – by Walther Bothe and Hans Geiger and by Compton and A W Simon respectively – and in both respects Bohr’s proposals failed.

The light quantum idea was here to stay.
Bose Statistics – the Photon Spin

It was emphasized earlier that from the very beginning Einstein was conscious of the fact that there was no theoretically well founded derivation of the Planck Law (8). Even his own derivation of 1916 relied on the Bohr theory for matter and interaction processes between matter and radiation. In June 1924 Satyendra Nath Bose working at Dacca University (now Dhaka in Bangladesh) sent Einstein a four page paper containing a novel logically self-contained derivation of the Planck Law, treating thermal radiation as a statistical mechanical system on its own and taking the photon picture to its logical conclusion. Einstein immediately recognised the depth of Bose’s ideas; helped in publishing his paper after translating it into German; and then followed it up with a paper of his own applying Bose’s method to the ideal material quantum gas. The key point in Bose’s method was a new way of counting complexions or microstates for an assembly of photons, in the process giving new meaning to the concept of identity of indistinguishable particles in the quantum world. In contrast to Einstein’s conclusion drawn from the Wien Law that light quanta have a certain mutual independence, Bose statistics shows that photons – because of their identity in the quantum sense – have a tendency to clump or stick together. And basically this difference accounts exactly for the Planck Law and its difference from the Wien limit.

In his paper sent to Einsein, Bose apparently made another radical suggestion – that each photon has an intrinsic angular momentum or helicity of exactly one (quantum) unit, which could be either parallel or antiparallel to its momentum direction. But – revolutionary as he was – Einstein found this suggestion too revolutionary and removed it in the published version of Bose’s paper!

Conclusion

Soon after the above events, modern quantum mechanics was discovered during 1925–26; and in 1927 Paul Adrien Maurice Dirac completed the task of quantising the classical Maxwell field, something which Einstein had foreseen as early as in 1917. And with that the photon was here to stay. What better way to end this account than to turn to Einstein himself in his old age:

“All the fifty years of conscious brooding have brought me no closer to the answer to the question, ‘What are light quanta?’ Of course today every rascal thinks he knows the answer, but he is deluding himself”.

Suggested Reading


The Story of the Photon


The Einstein–Podolsky–Rosen Paper – An Important Event in the History of Quantum Mechanics

S Chaturvedi, N Mukunda and R Simon

1. Introduction

In 1935, Albert Einstein, Boris Podolsky and Nathan Rosen published a paper titled ‘Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?’ in the Physical Review, Volume 47, pages 777–780. Generally known as the EPR paper, it has had an enormous influence, despite its short length, on the interpretation and further analysis of quantum mechanics. It immediately led to a reply from Niels Bohr in a paper with exactly the same title (except that ‘Be’ was replaced by ‘be’) in the same year, on pages 696–702 of the next volume of the same journal.

The aim of this article is to describe the background to and content of the EPR paper, viewed critically and keeping in mind a reader who begins her exploration of this subject with a study of this famous paper. We cover briefly: the discovery of quantum mechanics and the development of its traditional interpretation; Einstein’s unhappiness with this interpretation; Bohr’s Complementarity Principle expressed in a symbolic manner; a resumé of Einstein’s general viewpoints on the notions of separability and objective reality; the structure and main arguments of the EPR paper; Bohr’s point of view and rejoinder; Bell’s analysis of locality and realism; and an instructive example due to Hardy bringing out sharply the difficulties with realism in quantum theory. We conclude with a brief summary and the lessons to be learnt from this famous episode in physics.

2. Discovery of Quantum Mechanics and Development of its Interpretation

Quantum mechanics was discovered in three major steps over the space of less than a year during 1925–26: Heisenberg’s matrix mechanics in the summer of 1925, Dirac’s more symbolic quantum mechanics in autumn-winter 1925, and Schrödinger’s wave mechanics in winter-spring 1926. In the work of Heisenberg and of Dirac, the emphasis was on the description of observables or dynamical variables for quantum systems, especially the noncommutative nature of their multiplication. In Schrödinger’s work the concept of a general state and its description by means of a wave function was emphasized, of course along with an equation of motion determining its evolution in time. This latter was the quantum mechanical replacement

S Chaturvedi, N Mukunda and R Simon

for Newton’s equation of motion in classical mechanics. In Dirac’s hands, the wavefunction concept led to the formulation of the fundamental Superposition Principle of quantum mechanics. In June 1926, Max Born proposed the interpretation of the wave function in terms of probabilities, something at variance with Schrödinger’s own initial expectations.

The traditional interpretation of quantum mechanics emerged during 1926 and early 1927, involving intensive discussions among Bohr, Heisenberg and Pauli. Both Heisenberg’s Uncertainty Principle and Bohr’s Complementarity Principle resulted from these discussions. In September 1927, at a conference in Como to observe the centenary of the death of Alessandro Volta, Bohr made the first public presentation of his Complementarity Principle, but failed to communicate his ideas effectively. Einstein was not present at this conference.

Einstein was unhappy with the traditional interpretation of quantum mechanics, and till the end of his life he was unwilling to accept it. His initial attitude was that quantum mechanics was incorrect and internally inconsistent. He believed he could create experimental situations where the Uncertainty Principles would be violated. He made two important attempts to do so. The first was at the 5th Solvay Congress in October 1927, where he proposed thought experiments to show how the position-momentum uncertainty principle

$$\Delta q \Delta p \geq \frac{\hbar}{2}$$

(1)
could be ‘beaten’. The second was three years later, at the 6th Solvay Congress in October 1930; this time his proposed thought experiment was to disprove the energy-time uncertainty relation

$$\Delta t \Delta E \gtrsim \hbar$$

(2)
known as the Bohr Uncertainty Principle. On both occasions, Bohr was able to pinpoint the errors in the argument and thus rescue the interpretation and consistency of quantum mechanics.

After these episodes Einstein altered his stand: he conceded that quantum mechanics was internally consistent but claimed that it was incomplete. He believed that there exist situations in Nature which could not be described in the framework of quantum mechanics. It was this train of thought that ultimately led to the 1935 EPR paper.

3. The Standard Interpretation

At this point it is useful to recapitulate very briefly the standard interpretation of quantum mechanics. Let $S$ denote some physical system. Its quantum mechanical description involves two sets of mathematical quantities, with associated physical meanings:

(a) Physical quantities or observables or dynamical variables:
   represented (generally) by noncommuting hermitian operators $\hat{A}, \hat{B}, \ldots$;

(b) (pure) States – describable by vectors $|\psi\rangle, |\varphi\rangle, \ldots$ in a Hilbert space $\mathcal{H}$, subject to the Superposition Principle.
The operators $\hat{A}, \hat{B}, \ldots$ act on the Hilbert space $\mathcal{H}$ characteristic of the system. Given that $S$ is in the state $|\psi\rangle$, suppose an experiment is set up to measure $\hat{A}$. We can ask: In principle, what can the results of the measurement be? Knowing that the state is $|\psi\rangle$, what will the results be, and what is the probability for each possible result?

To answer these questions, we need to study the eigenvalues and eigenvectors of $\hat{A}$. Assume for simplicity that the eigenvalues are discrete and nondegenerate. Denote them and the corresponding eigenvectors by $a_j$ and $|a_j\rangle$ respectively. Then we have

$$\hat{A}|a_j\rangle = a_j|a_j\rangle, \quad a_j \text{ real},$$

$$\langle a_j|a_k\rangle = \delta_{jk}.$$  

(4)

and $\{|a_j\rangle\}$ forms an orthonormal basis for $\mathcal{H}$ (reality of the eigenvalues and orthogonality and completeness of the eigenvectors are assured by the hermiticity of $\hat{A}$). Then expand $|\psi\rangle$ in this basis:

$$|\psi\rangle = \sum_j \psi_j|a_j\rangle, \quad \psi_j = \langle a_j|\psi\rangle,$$

$$\langle \psi|\psi\rangle = \sum_j |\psi_j|^2 = 1.$$  

(5)

We assumed here that $|\psi\rangle$ is normalised to unit length. The answers to our questions are: the possible results of measurement of $\hat{A}$ are the eigenvalues $a_j$. In the (pure) state $|\psi\rangle$, the probability of obtaining the result $a_j$ is $|\psi_j|^2$. To this is added the collapse postulate: If the result $a_j$ is obtained, then $|\psi\rangle$ collapses to (is to be replaced by) $|a_j\rangle$, which is to be used for discussing further observations. Clearly, further repeated measurements of $\hat{A}$, on the collapsed state $|a_j\rangle$, will result in the same eigenvalue $a_j$. The experimental verification of the predicted probabilities $|\psi_j|^2$ therefore requires making measurements on a large collection of identically prepared copies of the state $|\psi\rangle$. For the time dependent state vector $|\psi(t)\rangle$, the equation of motion is the time-dependent Schrödinger equation. This is to be used only in between measurements.

For a composite system $S = A + B$ made up of subsystems $A$ and $B$, the state space is the tensor product of the individual spaces:

$$\mathcal{H}_{A+B} = \mathcal{H}_A \otimes \mathcal{H}_B.$$  

(6)

We come back to this later.

4. The Complementarity Principle

Bohr’s Complementarity Principle cannot be easily formulated with the same precision with which Heisenberg’s Uncertainty Principle is formulated. We give here a schematic account of Bohr’s Complementarity Principle. Every announcement of an experimental result ‘$R$’ must be accompanied by a statement of the experimental set up ‘$E$’ that led to it, so we must always
speak of and deal with pairs \((E, R)\). In turn, \(E\) is a combination of the system \(S\) (in some state \(|\psi\rangle\)), and apparatus \(A\) or \(B\) or ... constructed or designed to measure an observable \(\hat{A}\) or \(\hat{B}\) or .... \(A\) itself symbolically represents the apparatus and the actual carrying out of the experiment. So we ultimately deal with triples \((S, A, R)\), where \(A\) specifies \(\hat{A}\), and the state \(|\psi\rangle\) is left implicit.

In the classical case we say: we can delete \(A\) from the triple \((S, A, R)\), and we can claim and imagine that the system \(S\) possessed the value \(R\) for the physical quantity \(\hat{A}\) at the time of the measurement. We say that this is so, independent of the apparatus \(A\) and the act of measurement. So we can simply deal with pairs \((S, R)\), the relevant values of time being left implicit.

In quantum mechanics, on the other hand, we cannot do so, we cannot remove \(A\) from the description. In the language of Bohr the entire triple \((S, A, R)\) is a ‘phenomenon’ not reducible to anything more elementary. So we have the situation:

\[
(S, A, R) \begin{cases} 
\text{Classical: delete } A, \text{ keep } (S, R), \text{ say } S \\
\text{Quantum: cannot delete } A, \text{ retain triplet as a whole.}
\end{cases}
\]  

(7)

Thus: if the experimental arrangements \(A_1, A_2\) to measure \(\hat{A}_1, \hat{A}_2\) are mutually exclusive (for example, Stern-Gerlach apparatuses to measure the spin components \(S_x, S_y\) of a spin \(\frac{1}{2}\) object in two different directions), it means that \(\hat{A}_1\) and \(\hat{A}_2\) do not commute and cannot be measured simultaneously. We can have either \((S, A_1, R_1)\) or \((S, A_2, R_2)\) at a given time, one phenomenon or the other. We cannot think of \(S\) possessing values \(R_1\) and \(R_2\) for \(\hat{A}_1\) and \(\hat{A}_2\) simultaneously.

For Bohr, \(S\) was quantum and \(A\) classical, and he said the whole \(S+A\) was ‘unanalysable’. In von Neumann’s treatment, \(S, A\) and \(S+A\) are all quantum systems, so \(S+A\) is subject to a Schrödinger equation. But then the collapse rule cannot be derived from the Schrodinger equation, at least not in any simple manner.

5. Einstein’s General View Points

It was mentioned earlier that Einstein never accepted the standard interpretation of quantum mechanics. He also did not agree with Bohr’s Complementarity Principle. About this he said:

“Of the ‘orthodox’ quantum theoreticians whose position I know, Niels Bohr’s seems to me to come nearest to doing justice to the (EPR) problem .... Bohr’s principle of complementarity, the sharp formulation of which I have been unable to achieve despite much effort I have expended on it”.

In contrast to the Complementarity Principle, he believed in ‘objective reality’: physical systems possess numerical values for their properties independent of our observations of them. He also insisted upon some other important ideas, even if they were not always precisely expressed. Here are two of them:
Separability: This was a necessary ingredient of any theory in physics. Two systems $S_1$ and $S_2$ which are spatially far away from each other must be ‘independent’, they cannot ‘influence’ one another. There cannot be any action at a distance.

This requirement is meaningful even nonrelativistically, special relativistic locality is more precise and refined.

Real state of a system: this is indicated only qualitatively: it is not something ‘...immediately accessible to experience, and its application is always hypothetical (comparable to the notion of force in classical mechanics, if one doesn’t fix a priori the law of motion)’.

It is demanded that the real state of a system $S_2$ be independent of a spatially separated system $S_1$. A well-known and often quoted statement of Einstein, amounting to what is termed ‘Einstein locality’, is:

“But on one supposition we should, in my opinion, absolutely hold fast: the real factual situation of the system $S_2$ is independent of what is done with the system $S_1$ which is spatially separated from the former”.

It is time now to see what EPR attempted in their historic paper.

6. The EPR Paper

The aim was to show that quantum mechanics was incomplete. Since we wish to present their work in a critical manner (to hopefully orient and assist any reader of their paper), various key words will be underlined for emphasis. To achieve their objective, they introduced three notions or concepts, followed by two statements about them. The concepts are

- (a) a complete theory,
- (b) elements of physical reality (or epr),
- (c) counterpart of epr in physical theory.

None of these concepts is defined comprehensively with full meaning, because according to EPR, that much is not needed. Next come the two statements involving these concepts:

(1) A necessary condition for a theory to be complete is that every epr must have a counterpart in the theory.

Thus, as stated above, no sufficient conditions for completeness are given or attempted. What about epr’s which play a role in statement (1)? Only a sufficient condition which can help us recognise some of them is given:

“The epr’s cannot be determined by a priori philosophical considerations, but must be found by an appeal to results of experiments and measurements”.

The sufficient condition constitutes statement (2):

(2) Sufficient condition to identify an epr: “If, without in any way disturbing a system, we can predict with certainty (...) the value of a physical quantity, then there exists an epr corresponding to this physical quantity”.

95
S Chaturvedi, N Mukunda and R Simon

So the scheme of ideas can be depicted thus:

(a) Complete theory
No full definition, only a necessary Condition (1) involving all epr’s
(b) epr’s
No general definition, only a sufficient Condition (2) to identify some of them

(c) Counterpart of epr in theory
No precise meaning, simply left to be inferred

It seems reasonable to infer that ‘epr’ and ‘counterpart’ together mean the assignment of a definite numerical value to the concerned physical quantity, under specified circumstances. The strategy now is the following: Come up with a situation in quantum mechanics, some system $S$ in some state $|\psi\rangle$, such that using the sufficient condition (2) some epr’s can be found; then show that according to quantum mechanics they cannot all have counterparts in the quantum description, i.e., condition (1) is not satisfied; thus conclude that quantum mechanics is incomplete.

The EPR paper dealt with position and momentum eigenstates for two particles moving in one dimension. Since these are not normalisable states, in an equally good version Bohm considered a total system $S = A + B$, where $A$ and $B$ are two spin $\frac{1}{2}$ particles moving in physical space. They are supposed to be initially close together, then move far apart in such a way that the total spin wave function is unchanged. The spin wave function $|\psi\rangle$ is chosen to be the singlet state which is invariant under all spatial rotations, and can therefore be written in (infinitely) many equivalent ways. We choose the following two expressions:

\[
|\psi\rangle = \frac{1}{\sqrt{2}} \left( |z, +\rangle|z, -\rangle - |z, -\rangle|z, +\rangle \right)
= \frac{1}{\sqrt{2}} \left( |x, +\rangle|x, -\rangle - |x, -\rangle|x, +\rangle \right). \tag{9}
\]

Here, in each term on the right hand side, the first factor is the state of $A$ and the second factor that of $B$; for either particle, $|z, \pm\rangle$ are eigenstates of $S_\parallel$ with eigenvalues $\pm 1/2, |x, \pm\rangle$ are eigenstates of $S_\perp$ with eigenvalues $\pm 1/2$. The situation can be depicted as follows:

\[
\begin{array}{ccc}
\times & \leftarrow & \bigcirc \\
A & |\psi\rangle \text{ formed here} & B
\end{array} \tag{10}
\]
Now, in effect, EPR argue as follows: If we wish, we can measure $S_z^{(A)}$; the possible results are $\pm 1/2$; after the measurement the wave function collapses to $|z, \pm\rangle|z, \mp\rangle$; thus ‘without disturbing $B$’ we can infer that $S_z^{(B)}$ has value $\mp 1/2$; therefore $S_z^{(B)} = \mp 1/2$ is an epr. On the other hand, if we wish, we can start by measuring $S_x^{(A)}$, getting the result $\pm 1/2$; and then following a parallel line of reasoning, we conclude that $S_x^{(B)} = \mp 1/2$ is an epr. By locality, the ‘real state’ of $B$ should be unaffected by what is measured at $A$. At $A$ we can choose what to measure, either $S_z^{(A)}$ or $S_x^{(A)}$ but not both. Since we have shown the existence of two epr’s for $B$, both $S_z^{(B)}$ and $S_x^{(B)}$ should have had definite values already. As they do not commute, quantum mechanics cannot account for this. Therefore quantum mechanics is incomplete.

One sees from the EPR argument that for ‘an epr to have a counterpart in the theory’ means ‘for the concerned physical quantity to have a definite numerical value’, i.e., for the state to be the relevant eigenstate of the operator.

The expression for $|\psi\rangle$ in equation (9) cannot be written in the product form $|\varphi\rangle_A |\chi\rangle_B$, for any choices of $|\varphi\rangle$ and $|\chi\rangle$. Such states are said to be entangled.

Later accounts say that the EPR paper was drafted by Podolsky and not seen by Einstein in its final form. (For example, as a rule Einstein always spoke of the ‘psi-function’, never of the wavefunction!) He felt his views were not well presented. Here is his version of the incompleteness argument. As we saw earlier, one of his basic requirements was that the ‘real state’ of $B$ should be independent of the spatially separated $A$. Then: quantum mechanics would be complete if and only if there is a one-to-one correspondence between real states of $B$ and wavefunctions $\psi_B$ for $B$ (upto overall phases, and limiting ourselves to pure states). For the composite system $S = A + B$ with spatially separated parts $A$ and $B$, suppose the overall wavefunction $\psi_{AB}$ is not a product but is entangled. Then the wavefunction we ascribe to $B$ after a measurement on $A$ followed by collapse depends on what is measured on $A$. Schematically:

$$\psi_{AB} \text{ for } S = A + B$$

\[
\begin{aligned}
\text{measure something on } A, \\
\text{use collapse, get } \psi_B \text{ for } B \\
\text{assume incompatible} \\
\text{measure something else on } A, \\
\text{use collapse, get } \psi'_B \text{ for } B
\end{aligned}
\]

\begin{equation}
B \text{ is not disturbed}
\end{equation}

The ‘real state’ of $B$ is thus not represented by a unique wavefunction, independent of operations on $A$. The wavefunction for $B$, found via collapse, depends on what is measured on $A$ which is far away. This nonuniqueness of the wavefunction assigned to $B$ shows that quantum mechanics is incomplete. In effect, Einstein believed that, being spatially separated, $A$ and $B$ have their respective individual ‘real states’, which together determine the state of $S = A + B$. 

For clarity let us repeat the two forms of the incompleteness argument:

**EPR paper:** we show the existence of two epr’s for B but no $\psi_B$ in quantum mechanics can accommodate both of them

**Einstein version:** B has one ‘real state’, but quantum mechanics does not give us a unique $\psi_B$ to represent it. \(^{(12)}\)

Clearly, both forms involve making two mutually incompatible measurements on A. On this aspect, EPR admit:

"... one would not arrive at our conclusion if one insisted that two or more physical quantities can be regarded as simultaneous elements of reality only when they can be simultaneously measured or predicted”.

Clearly they do not require this stronger condition while identifying two epr’s for B – they imagine two mutually incompatible experiments being carried out on A, only one of which can actually be carried out at a given time. They however say: if one adopted the above stronger condition, only one epr can be identified for B, but it depends on what is measured at A which is far away, and this is unreasonable: “No reasonable definition of reality could be expected to permit this”.

So the two options, both of which they find unacceptable, are these:

**Weaker:** Consider two mutually exclusive measurements on A; imagining either one or the other being carried out at some time, claim to have found two epr’s for B; both should have counterparts in a complete theory; quantum mechanics has no such simultaneous eigenstates, so it is incomplete.

**Stronger:** Since the two measurements contemplated on A are incompatible, we can carry out only one or the other, then infer the existence of a corresponding epr for B. But B is far away from A, so an epr for it cannot depend on what is measured at A. It is unreasonable to permit such dependence.

### 7. Bohr’s Response

As we mentioned in Section 1, Bohr’s response was contained in a paper in the very next volume of the Physical Review in the same year, 1935, as EPR. According to Pais,

“(Bohr) did not believe that the EPR paper called for any change in the interpretation of quantum mechanics”.

Bohr’s paper is not easy reading; here is an attempt to convey what he probably had in mind. Consistent with Pais’ statement quoted above, we say: If $S = A + B$ is in an entangled (i.e. nonproduct) state, there are no mutually independent (pure) states for A and B even if they are spatially far apart; separability of state with respect to subsystems does not hold in quantum mechanics. Large spatial separation of A and B may imply absence of physical interaction between them, but it does not imply existence of independent pure states for A and B. When $\psi_{AB}$ is entangled, the pure state obtained for B via collapse after a measurement on A does
depend on what is measured on $A$, and is to be used for making predictions about any (later) measurements made on $B$.

Thus quantum correlations between $A$ and $B$ in an entangled state $\psi_{AB}$ are distance independent. This means that quantum mechanics is nonlocal at the wave function level. As parts of a total $S = A + B$, $A$ and $B$ do not always ‘possess’ independent individual wavefunctions. Nevertheless these correlations cannot be used to send messages. $A$ knows the resulting state of $B$ at the end of his measurement, but not $B$. In other words, even though the result of a measurement at $A$ plus wavefunction collapse allows us to predict a certain measurement result at $B$ with certainty, the former is uncertain and governed by probabilities; that uncertainty then remains for any prediction concerning $B$.

Statistically speaking, quantum mechanics does obey locality, in the following precise sense. For any observable $\hat{B}$ of $B$, whether we make some measurement on $A$ (and retain all results) or not, the expectation value is the same, namely:

$$\langle \hat{B} \rangle_{\psi_{AB}} = \text{Tr}_B(\hat{B} \hat{\rho}_B),$$
$$\hat{\rho}_B = \text{Tr}_A |\psi_{AB}\rangle\langle \psi_{AB}|.$$  \hspace{1cm} (13)

Here $\text{Tr}_A$ is to be read as partial trace with respect to the subsystem $A$, which amounts to ‘ignoring’ $A$ or treating all states of $A$ on equal footing.

One is tempted to ask the question: where did EPR ‘go wrong’? Why does quantum mechanics not abide by their ‘innocent looking’ conditions? The answer is that those conditions are not really so innocent! Bohm suggests that in addition to their two statements (1), (2) in Section 6, they made two more implicit assumptions:

(3) The world can be correctly analysed in terms of distinct and separately existing epr’s;
(4) Every one of these epr’s must be a counterpart of a precisely defined mathematical quantity appearing in a complete theory.

Here we see the word ‘counterpart’ again! Presumably the meaning of (4) is that if we have an epr, then some variable in the theory must have a corresponding numerical value. Of course, as Bohm says, quantum mechanics does not abide by (3) and (4); it does not allow us to work wholly with a set of dynamical variables always possessing definite numerical values.

One can even take the following attitude: given the singlet state $|\psi\rangle$ of eq. (9) for the pair of spin $1/2$ particles $A + B$, no single unambiguous epr has been shown to exist for $B$. From this point of view, Einstein’s view described in Section 6 seems preferable as a criticism of quantum mechanics.

One is finally left with the feeling that EPR’s criticism reduces to their not liking quantum mechanics, as it does not agree with their prejudices about any theory.

8. Bell’s Analysis

We saw in Section 5 that Einstein insisted upon both locality and objective reality as general requirements for any physical theory. Their combination is called ‘local realism’, and it was
analysed in precise fashion by John Bell. He found an inequality which any local realist theory should obey, but which quantum mechanics does not. The inescapable conclusion is that quantum mechanics violates local realism.

Here is Bell’s argument, referring again to the two spin $1/2$ system $A + B$, assumed to be in the singlet state $|\psi\rangle$ of equation (9). Quantum mechanics does not allow us to imagine that any spin component of $A$ (or of $B$) has a definite value, in the absence of some measurement. Imagine now that quantum mechanics can be extended to, or embedded within, some more encompassing theory involving some hidden variables $\lambda$. If $\lambda$ were known, we suppose that we could then say: for any three-dimensional unit vectors $a$ and $b$:

\[
a \cdot \sigma^{(A)} \text{ has the numerical value } A(a, \lambda) = \pm 1,
\]

\[
b \cdot \sigma^{(B)} \text{ has the numerical value } B(b, \lambda) = \pm 1.
\]

(14)

Realism is expressed by the possibility of assigning definite numerical values $A(a, \lambda)$, $B(b, \lambda)$ to $a \cdot \sigma^{(A)}$, $b \cdot \sigma^{(B)}$ (for all choices of $a$, $b$) if $\lambda$ were known. Locality is expressed by the $b$-independence of $A(a, \lambda)$ and the $a$-independence of $B(b, \lambda)$.

Let $\rho(\lambda)$ be the probability distribution of $\lambda$, possibly dependent on $|\psi\rangle$.

Then the correlation between components of $A$-spin and $B$-spin in general directions is

\[
P(a, b) = \int d\lambda \rho(\lambda) A(a, \lambda) B(b, \lambda).
\]

(15)

Take four directions $a, a', b, b'$ to get two preliminary inequalities:

\[
|P(a, b) - P(a', b')| =
\]

\[
|\int d\lambda \rho(\lambda)(A(a, \lambda)(B(b, \lambda) - B(b', \lambda))|
\]

\[
\leq \int d\lambda \rho(\lambda)|B(b, \lambda) - B(b', \lambda)|;
\]

\[
|P(a', b) + P(a', b')| =
\]

\[
|\int d\lambda \rho(\lambda) A(a', \lambda)(B(b, \lambda) + B(b', \lambda))|
\]

\[
\leq \int d\lambda \rho(\lambda) |B(b, \lambda) + B(b', \lambda)|.
\]

(16)

Adding these gives the Bell inequality:

\[
|P(a, b) - P(a', b')| + |P(a', b) + P(a', b')| \leq \int d\lambda \rho(\lambda) \times
\]

\[
\{ |B(b, \lambda) - B(b', \lambda)| + |B(b, \lambda) + B(b', \lambda)| \} = 2,
\]

(17)

the final result following from the fact that if one term within the curly brackets is 0 the other is 2 and conversely. Thus local realism implemented via hidden variables entails

\[
|P(a, b) - P(a', b')| + |P(a', b) + P(a', b')| \leq 2.
\]

(18)
Now in the singlet state (9) quantum mechanics gives the value

\[ P_{QM}(a, b) = -a \cdot b \]  

(19)

We choose \( a, b, a', b' \) to be coplanar with \( b \) making an angle \( \pi/4 \) with \( a \), and \( b' \) another \( \pi/4 \) with \( a' \) (and so \( 3\pi/4, \pi/2 \) with \( a, b \) respectively). Then the left hand side of (18) is \( 2 \sqrt{2} \), in violation of that inequality.

Many experiments over the years, with increasing precision, have shown violation of such local realist inequalities, and in agreement with quantum mechanics. Nature does not respect local realism!

9. The Hardy state

Even realism is untenable in quantum mechanics! We know this already from the discussion of Section 7, but here is another striking illustration, due to Hardy. We consider again a pair \( A + B \) of spin 1/2 particles. For each, we contemplate measurements of one of two non-commuting variables, \( \sigma_x \) and \( \sigma_z \). We search for a pure state \( |\psi\rangle \) obeying three conditions:

(i) If measurement of \( \sigma_z^{(A)} \) yields \(+1\),
    it must then yield \( \sigma_z^{(B)} = +1 \);

(ii) If measurement of \( \sigma_z^{(B)} \) yields \(+1\),
    it must then yield \( \sigma_x^{(A)} = +1 \);

(iii) If measurement of \( \sigma_x^{(A)} \) yields \(+1\),
    it must then yield \( \sigma_x^{(B)} = +1 \); 

(20)

In an obvious notation, these three conditions lead to the following three expressions for \( |\psi\rangle \):

\[ |\psi\rangle = N_1(|z+\rangle|z+\rangle + |z-\rangle|\alpha\rangle) \]
\[ = N_2(|x+\rangle|z+\rangle + |\beta\rangle|z-\rangle) \]
\[ = N_3(|x+\rangle|x+\rangle + |x-\rangle|y\rangle). \]

(21)

In each term, the first/second factor is an \( A \) state/\( B \) state; \( |z\pm\rangle \) are \( \sigma_z \) eigenstates; \( |x\pm\rangle \) are \( \sigma_x \) eigenstates,

\[ |x\pm\rangle = \frac{1}{\sqrt{2}}(|z+\rangle \pm |z-\rangle); \]

(22)
$S$ Chaturvedi, $N$ Mukunda and $R$ Simon

$N$’s are normalization factors; $|\alpha\rangle$, $|\gamma\rangle$ are some $B$-states, and $|\beta\rangle$ is some $A$-state. Simple algebra then gives:

$$|\alpha\rangle = |z+\rangle + 2|z-\rangle,$$
$$|\beta\rangle = \sqrt{2}|z-\rangle,$$
$$|\gamma\rangle = -\frac{1}{\sqrt{2}}|z-\rangle,$$

$$N_1 = \frac{1}{\sqrt{6}}, N_2 = \frac{1}{\sqrt{3}}, N_3 = \frac{\sqrt{2}}{3}. \quad (23)$$

We have as a consequence:

$$\langle z + |\langle x - ||\psi\rangle = \frac{1}{2\sqrt{3}}. \quad (24)$$

Classical realism would mean, on the basis of (20): if $\sigma^A_z$ definitely has the value +1, then $\sigma^B_z$ definitely has value +1, then $\sigma^A_x$ has definitely the value +1, then $\sigma^B_x$ also definitely has value +1. But from quantum mechanics we see that equation (24) says that there is an 8.33% probability that $\sigma^A_z = +1$ and $\sigma^B_x = -1$. This is an illustration of the loss or lack of realism in quantum mechanics.

An even more striking illustration not involving any probabilities at all is due to Greenberger, Horne and Zeilinger, and involves a set of three spin half particles. But let us leave that to the curious reader!

10. Concluding Comments

One should remember that the EPR paper was written seventy years ago, when quantum mechanics was barely a decade old. Since that time, by and large physicists have grown accustomed to the counter-intuitive features of quantum phenomena. Even though our account of EPR has been intentionally critical, it must be admitted that it has highlighted specific features of quantum mechanics – the issues of realism, locality and entanglement. It immediately inspired Schrödinger’s ideas on entanglement, which is a reflection of the enormous richness contained in the tensor product rule $H_{A+B} = H_A \otimes H_B$ for composite system state spaces. Today entanglement is the key resource for quantum computation, and for other quantum information processes. The ideas of realism and locality which Einstein would never give up led to Bohm’s efforts to find an almost classical interpretation of quantum mechanics, and then to Bell’s incisive analysis of the full implications of local realism. Now we know that quantum correlations are very subtle, and go beyond classical limits even without involving large spatial separations. Thus quantum correlations at the wave function level are distance-independent. Thus all this has resulted from the brief EPR paper.

We hope that young readers of our account will feel confident in looking at the literature in this area of physics with a good sense of direction to guide them.
Acknowledgements: We thank Vidyanand Nanjundiah for a critical reading of the manuscript and many useful suggestions.

Suggested Reading


The 2005 Nobel Prize in Physics: Optics*

Vasant Natarajan and N Mukunda

The 2005 Nobel Prize in Physics has been awarded in the area of optics, or more specifically in laser physics. One half of the prize (theory part) has been given to Roy Glauber of Harvard University “for his contribution to the quantum theory of optical coherence,” which became important soon after the invention of the laser. The second half of the prize (experimental part) is jointly awarded to two physicists, John Hall of the National Institute of Standards and Technology (NIST) in Boulder, USA, and Theodor Hänsch of the Max-Planck Institute for Quantum Optics in Garching, Germany. They have been cited “for their contributions to the development of laser-based precision spectroscopy, including the optical frequency comb technique.”

India has a rich tradition of research in optics dating back to the pioneering work of C V Raman (Nobel Prize for the Raman Effect in 1930). In the 1950’s there appeared S Pancharatnam’s fundamental studies on polarization optics in the course of which he discovered the geometric phase in its earliest form. Then in 1961 came the crystal optics work of G N Ramachandran and S Ramaseshan. Towards the end of the theory part of this article we will describe briefly the remarkable 1963 discovery of the Diagonal Coherent State Representation and the Optical Equivalence Theorem, central to quantum optics, by E C G Sudarshan working in the USA.

1. Quantum Theory of Optical Coherence

The understanding of the nature and properties of light has fascinated humankind for a very long time; its progress is an important part of the history of physics. It may be useful to very briefly remind the reader of some of the more ‘recent’ events in this history, starting with the work of Maxwell in mid 19th century. With such a background, one can understand better the work for which the ‘theory part’ of the 2005 Physics Nobel has been given.

Maxwell 1 succeeded in unifying the laws of electricity and magnetism into a single theory, and then went on to show that light was an electromagnetic wave. Thus as a result of his work three previously separate fields of physics became one. Around the same time, the field of statistical mechanics, as the foundation for thermodynamics, was also being developed. Around 1900, however, it became clear that the combination of statistical ideas and the classical


1The year 2006 marks the 175th birth anniversary of Maxwell, and is being celebrated as Maxwell Year in Scotland. Resonance featured Maxwell in the May 2003 issue.
Maxwell description of electromagnetic radiation led to an impasse: it could not explain the experimental results concerning black body or thermal radiation, i.e. radiation in equilibrium with material bodies at a common temperature.

It was Planck’s solution of this problem that led to the birth of quantum theory in late 1900, the dawn of the 20th century (Nobel 1918). Planck’s work involved two steps: first, a mathematical interpolation amounting to inspired guess work that led to his famous radiation formula which fitted experiment beautifully; second, a derivation of this formula based on the hypothesis that (electrically charged) material oscillators could emit and absorb radiation energy only in discrete amounts or quanta. This was a revolutionary idea.

Each of the later advances in the understanding of light has been equally stunning. In 1905 Einstein was able to argue from the non-classical limit of Planck’s formula that radiation in its own nature has a lumpy or particle-like aspect, in contrast to the classical continuous Maxwell picture. He then presented an explanation of the photoelectric effect as one piece of evidence in support of his conclusions (Nobel 1921). A few years later in 1909 he studied the energy fluctuations of Planck radiation and deduced that radiation simultaneously possesses the seemingly contradictory, or dual, particle and wave properties. Then in 1916 he presented a startlingly new derivation of Planck’s law based on the processes of emission and absorption of radiation by matter, and also showed that light quanta – photons – carry momentum in addition to energy. In 1924, S N Bose gave yet another derivation of Planck’s law based on a deep understanding of the identity of light quanta; the work was immediately appreciated and taken further by Einstein. This series of events came to a triumphant conclusion with Dirac in 1927 showing how to apply the principles of the just discovered quantum mechanics to the classical Maxwell theory.

As this implies, the Dirac theory of the quantised electromagnetic field came after a satisfactory quantum mechanics for matter had been developed. The first steps here were (apart from Planck in 1900) again taken by Einstein in 1907, in his theory of specific heats; and then by Niels Bohr in 1913 with the theory of stationary electronic states in the hydrogen atom (Nobel 1922). (This Bohr theory was a vital component of Einstein’s 1916 work on radiation). There followed what was later called the period of the Old Quantum Theory when Bohr’s initial ideas were tried to be extended to more complex material systems. By about 1923 this effort ran into severe problems, and the situation was resolved only with the discovery of quantum mechanics by Heisenberg, Dirac and Schrödinger independently during 1925–26 (Nobels 1932, 1933).

Returning to radiation, after Dirac the theory of quantum electrodynamics – QED – was further developed by many leading physicists of that time including Heisenberg, Pauli, Peierls and Landau. However it was now found that when one went beyond the lowest level of approximation it was plagued by severe mathematical inconsistencies – the so-called problem of divergences. Calculations gave meaningless infinite answers for quantities which should have been finite. This was the situation through most of the 1930’s and early 1940’s, until the discovery of the method of renormalization independently by Tomonaga, Schwinger and Feynman.
The 2005 Nobel Prize in Physics: Optics

(Nobel 1965), completed by around 1947. The impetus given to this effort by the experimental measurement of the Lamb shift (Nobel 1955) is emphasized in the second Section of this article.

With the arrival of the renormalization procedure resulting in a finite QED, it became clear that our understanding of the fundamental nature of light and its interaction with matter had reached a level of completion. All later work including what will be described below is within that framework.

Meanwhile within the arena of classical optics many new developments had been taking place. They could be regarded as a completion of the earlier elementary treatments of diffraction and interference of classical wave amplitudes. It was realised that essentially all earlier classical optical effects could be described in terms of the two-point amplitude correlation function; and via this object the concepts of partial coherence and its propagation were brought into the field. (Analogous developments with regard to polarization of light had also taken place.) In this way the role of statistical methods in optics came to be much better appreciated. Some of the early names are those of Fritz Zernike (Nobel 1953), van Cittert, Blanc-Lapierre and Dumontet. From about the mid-1950’s the whole subject was developed in a systematic manner largely by Emil Wolf. After the invention of intensity interferometry by Hanbury-Brown and Twiss in 1956, it became clear that it was necessary to go beyond the two-point amplitude correlation function (adequate to describe Young-type interference phenomena) to higher order correlation functions. Thus intensity correlations involve correlations among amplitudes at four space-time points, or a four-point function. Correlation functions of all higher orders came into play in the treatment by Mandel of the semi-classical photoelectron counting distribution formula. Here one has a fluctuating classical light beam falling on a photodetector, and one wishes to find the probabilities for various numbers of electrons to be emitted over a given time period. Then, from the experimentally measured statistical properties of the photoelectrons emitted, one obtains information on the statistical properties of the incident light beam.

Note the contrast to the original 1905 Einstein explanation of the photoelectric effect. In that treatment it was light which was regarded as possessing quantum features, and a quantum description of matter was still many years away. After the arrival of quantum mechanics for matter it became possible to account for the photoelectric effect in an alternative semiclassical manner – light can be treated as a statistical fluctuating classical quantity, while the electron is quantum mechanical. The key feature is that quantum ideas are needed at least for one of the two players in the process, light or electrons (ultimately of course for both in a completely satisfactory treatment). In any case, in Mandel’s work the second of the above two viewpoints was adopted.

To give the reader some idea of the kinds of expressions and concepts involved in this development, we present in Box 1 the definitions and interpretations of correlation functions in classical statistical optics. For simplicity we ignore the vector nature of the electric field and treat it as though it were a scalar. (We also omit reference to the magnetic field). The arguments \(x, y, \ldots\) are combined spatial and time coordinates; and classical statistical averages are indicated by angular brackets. Note the separation of the real total electric field into two
mutually conjugate parts, and the use of these parts in defining correlation functions and coherence. Again for simplicity only correlation functions with equal numbers of $E^{(+)}$'s and $E^{(-)}$'s are considered.

These two streams of work – the completion of QED and the growth of classical statistical optics – merged in the early 1960’s and led to the quantum theory of optical coherence, more generally quantum optics, to which many basic contributions were made by R J Glauber. The invention of the laser by that time had made it clear that there was a need to describe, within the overall framework of QED, states of electromagnetic radiation associated with arbitrary, in particular non-thermal, light beams. (The traditional uses of QED, in the realm of elementary particle physics, had only dealt with processes involving small numbers of photons – absorption and emission of single photons, scattering of a photon on an electron, and the like.) The physical principle underlying Glauber’s work, as foreshadowed in the Mandel treatment of photo electron counting, is that all conventional methods of light detection involve absorption of photons from the field being observed. (This is true even in the human and animal visual systems.) Building on this, Glauber was able to arrive at the most useful measure of (partial) coherence of the quantised electromagnetic field at the two-point level, and then to generalize it to correlation functions of all higher orders. This was a specific way to pass from the complete classical hierarchy of correlation functions of various orders – Box 1 – to their quantum counterparts. In then defining and analysing the concepts of partial and of complete coherence, to some finite order or to all orders, he demonstrated the great usefulness of a special set of quantum states called coherent states. These states can be defined both for material oscillators

<table>
<thead>
<tr>
<th>Box 1. Classical Correlation Functions for Fluctuating Electric Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Real classical electric field</strong> $E(x) = E^{(+)}(x) + E^{(-)}(x)$,</td>
</tr>
<tr>
<td>$E^{(+)}(x) =$ complex positive frequency part,</td>
</tr>
<tr>
<td>$E^{(-)}(x) = E^{(+)}(x)^*$ = complex negative frequency part.</td>
</tr>
<tr>
<td>Classical two-point correlation function = statistical average of product $E^{(-)}(y)E^{(+)}(x)$ of two complex field amplitudes</td>
</tr>
<tr>
<td>$= \langle E^{(-)}(y)E^{(+)}(x) \rangle$ : (1b)</td>
</tr>
<tr>
<td>adequate to discuss intensity measurements ($y = x$), Young type interference phenomena.</td>
</tr>
<tr>
<td>Classical four-point correlation function</td>
</tr>
<tr>
<td>$= \langle E^{(-)}(y_1)E^{(+)}(y_2)E^{(+)}(x_1)E^{(+)}(x_2) \rangle$ : (1c)</td>
</tr>
<tr>
<td>needed to discuss Hanbury-Brown-Twiss intensity correlations ($y_1 = x_1, y_2 = x_2$). Mandel's semiclassical photo-electron counting distribution formula involves</td>
</tr>
<tr>
<td>$\langle E^{(-)}(y_1) \cdots E^{(-)}(y_n)E^{(+)}(x_1) \cdots E^{(+)}(x_n) \rangle$ (1d)</td>
</tr>
<tr>
<td>for all $n$, with $y_1 = x_1, \ldots, y_n = x_n$.</td>
</tr>
<tr>
<td>Coherence of order $2n$ holds if the expression (1d) factorises completely as $V(y_1) \cdots V(y_n)V(x_1) \cdots V(x_n)$ for some field amplitude $V$.</td>
</tr>
</tbody>
</table>
Box 2. Correlation Functions for Quantised Electric Field

\[ E(x) = E^+(x) + E^-(x), \text{ all field operators;} \]

\[ E^+ \text{ annihilates one photon, } E^- \text{ creates one photon.} \quad (2a) \]

Two-point correlation function adequate to describe intensity measurements by photon absorption, Young type interference:

\[ G^{(1,1)}(x; y) = \text{Tr}[\rho E^-(y)E^+(x)], \]

\[ \rho = \text{density operator of quantum state.} \quad (2b) \]

Four-point correlation function needed to describe Hanbury–Brown–Twiss intensity correlations:

\[ G^{(2,2)}(x_1, x_2; y_1, y_2) = \text{Tr}[\rho E^-(y_1)E^-(y_2)E^+(x_1)E^+(x_2)]. \quad (2c) \]

Higher order correlation functions:

\[ G^{(n,n)}(x_1, \ldots, x_n; y_1, \ldots, y_n) = \text{Tr}[\rho E^-(y_1)\ldots E^-(y_n)E^+(x_1)\ldots E^+(x_n)]. \quad (2d) \]

Complete coherence \( \equiv \) for all \( n, G^{(n,n)}(x_1, \ldots, y_1, \ldots, x_n, y_n) = V(y_1)^* \ldots V(x_1)^* \ldots \) for some \( V(x) \Rightarrow \text{essentially, } \rho \text{ is a coherent state.} \)

and for the free radiation field. They had been discovered by Schrödinger in 1927, studied by von Neumann in 1930, and used in a specific context within QED by Bloch and Nordsieck in 1937. Glauber’s work amounted to a rediscovery of their enormous usefulness in describing states of radiation in the complete quantum optics context.

Let us first describe briefly the quantum counterparts of the contents of Box 1, assembled in Box 2, and then turn to coherent states. It is of course out of place to attempt to give here anything like a complete resumé of the basic structures of quantum mechanics, much less of QED. We can do no better than make suggestive statements, and try to get across some basic ideas. For simplicity we use the same symbols \( E^{(\pm)}(x) \) in quantum theory as classically.

In quantum theory, however, these are not complex valued numbers any more, but operators which act on quantum state vectors. \( E^{(+)}(x) \) is an operator which acting on a state annihilates or subtracts one photon; \( E^{(-)}(x) \) is the hermitian conjugate (replacement for the classical complex conjugate) of \( E^{(+)}(x) \), and acting on a state it creates or adds one photon. \( E^{(+)} \) and \( E^{(-)} \) do not commute. In the vacuum state there are no photons at all, so \( E^{(+)} \) applied to that state gives zero. States in quantum mechanics may be pure, describable by a single state vector or wave function \( \psi \); or mixed, namely an ensemble of several pure states \( \psi_1, \psi_2, \ldots \), each present with a corresponding probability \( p_1, p_2, \ldots \). In the latter case, the entire ensemble can be represented by what is called a density operator or density matrix \( \rho \), this is the most general quantum state.

The entries in Box 2 can now be hopefully understood.

The symbol ‘Tr’ stands for ‘Trace’ and (along with the presence of \( \rho \)) is the quantum counterpart of classical statistical averaging which was denoted in Box 1 by angular brackets.
One point to note with care is that in the definitions of $G^{(1,1)}$, $G^{(2,2)}$, ..., in Box 2, the $E^{(−)}$ factors (creation operators) always stand to the left of the $E^{(+)}$ factors (annihilation operators). This is the key feature of the Glauber definition – detection by absorption of photons – and we are not free to interchange the sequence of $E^{(−)}$’s and $E^{(+)}$’s since they do not commute. The last sentence in Box 2 brings in the coherent states, so we describe them briefly at this point, aided by Box 3.

### Box 3. Coherent States of Single Mode Radiation Field

States with definite number of photons: $|n\rangle$, $n = 0, 1, 2, \ldots$.  
For any complex number $z$:

Coherent state $|z\rangle$ = superposition of states with definite photon numbers

$$|z\rangle = e^{-|z|^2/2} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}} |n\rangle.$$  

Some important properties of coherent state $|z\rangle$:

- Probability of finding $n$ photons = $e^{-|z|^2} \frac{|z|^{2n}}{n!}$: Poisson distribution.  
- Mean number of photons = average of $n = |z|^2$.  
- Fluctuation in number of photons = average of $n^2$ – (average of $n$)$^2$
  
  $$= |z|^2$$
  
  $$= \text{mean} : \text{characteristic of Poisson distribution.}$$

We limit ourselves to a single mode of the quantum radiation field, so all the photons have the same spatio-temporal characteristics. (The generalisations to several modes or to the entire field are straightforward). For example we may fix the frequency, propagation direction and polarization state for all of them, so only the photon number can vary. In quantum mechanical notation the state with exactly $n$ photons is written as $|n\rangle$: for $n = 0$ we have the vacuum or no photon state $|0\rangle$, for $n = 1$ the one-photon state $|1\rangle$, and so on. Each of these is a pure state and they are mutually exclusive or orthogonal: if we know that there are exactly $n$ photons present, we certainly know the total photon number is not $n'$ for any $n' \neq n$. Given any set of pure states, we can multiply each one by some complex number and add them all up to get a new pure state. This is the fundamental Superposition Principle of quantum mechanics which has no classical analogue. For each complex number $z$, we can produce exactly one pure state using the expression given in (3b) of Box 3. This is the coherent state $|z\rangle$ of the concerned mode. Thus a coherent state has a variable number of photons present, with probabilities given by a Poisson distribution, (3c) of Box 3. These states turn out to be quantum states as close as possible to classical field states in the sense that the unavoidable or inescapable uncertainty principle of quantum mechanics is barely obeyed. They also turn out to be as close to having a definite ‘phase’ – in contrast to a definite photon number – as is possible in the quantum framework.
We will conclude this part of our article by describing two crucial properties of coherent states, justifying their importance for quantum optics. Each coherent state $|z\rangle$ is a pure state. Consider now a mixed state $\rho$ in which all these $|z\rangle$ are present with various probabilities, described by a ‘classical’ probability distribution $\phi(z)$ over the complex plane. It then turns out that the particular definitions of the quantum optical correlation functions given in (2b), (2c), (2d) of Box 2 combine with the very special properties of coherent states to lead to a remarkable result: each quantum correlation function has the same form and the same value as the corresponding classical correlation function calculated for a suitably defined classical statistical state. The key to this lies in two facts: the $E^{(-)}$ factors always stand to the left of the $E^{(+)}$ factors in the definitions of quantum correlation functions; and the $E^{(+)}$ factors act very simply on coherent states $|z\rangle$. This brings out graphically the extreme appropriateness of coherent states in these problems.

Now we come to our final point. A general quantum state $\rho$ can certainly not be reconstructed from the coherent states $\{|z\rangle\}$ via a classical probability distribution $\phi(z)$, as was assumed in the previous paragraph. But in a remarkable and truly fundamental result it was shown in 1963 by E C G Sudarshan that every quantum state $\rho$ can be formally regarded as a ‘generalized ensemble’ over the coherent states $|z\rangle$, except that $\phi(z)$ may not be interpretable as a ‘classical’ probability distribution! This is known technically as the Diagonal Coherent State Representation and the Optical Equivalence Theorem. Referring to the highlighted phrase in the previous paragraph we can say that in case $\phi(z)$ is not a true probability distribution, we have equivalence of forms but not of values for the two families of correlation functions, quantum and classical. For the most general quantum state $\rho$, $\phi(z)$ is not a function in any ordinary mathematical sense, but a singular quantity, a so-called distribution of a particular class that can be precisely characterized. This result is truly basic to the theory of quantum optics, as it is the only way in which we can exhibit the clear distinction between classical and quantum natures of optical fields. States displaying sub-Poissonian photon statistics or antibunching, so-called squeezing and Hanbury-Brown–Twiss anticorrelations are all truly quantum in nature, and correspond to singular, or at least non-positive definite, $\phi(z)$ in the Sudarshan classification. One can say that the need to allow $\phi(z)$ to go beyond the collection of probability distributions in considering all quantum states shows why quantum and classical theories are radically different, the former overstepping the confines of the latter. In fact this is a recurring feature of attempts to express quantum mechanics in the language of classical physics – the range of quantum mechanical possibilities always overflows classical boundaries. It is extremely unfortunate that this result of Sudarshan has not received the credit and recognition that is its due. The interested reader may refer to the article, ‘On Sudarshan’s Diagonal Coherent State Representation’ by C L Mehta [2].

That apart, the reader would have appreciated all the developments that form the backdrop to the ‘theory part’ of the physics 2005 Nobel.
2. Optical Frequency Comb Technique

Lasers have impacted our lives in a countless number of ways. Today they are found everywhere, in computer hard disk drives, CD players, grocery store scanners, and in the surgeon’s kit. In research laboratories, almost everyone uses lasers for one reason or another. However, arguably the greatest impact of lasers in physics has been in high-resolution spectroscopy of atoms and molecules. To see this, consider how spectroscopy was done before the advent of lasers. You would use a high-energy light source to excite all the transitions in the system, and then study the resulting emission “spectrum” as the atoms relaxed back to their ground states. This is like studying the modes of vibration of a box by hitting it with a sledgehammer and then separating the resulting sound into its different frequency components. A more gentle way of doing this would be to try and excite the system with a tuning fork of a given frequency. Then by changing the frequency of the tuning fork, one could build up the spectrum of the system. This is how you do laser spectroscopy with a tunable laser; you study the absorption of light by the atoms as you tune the laser frequency. When you come close to an atomic resonance, you build up a typical absorption curve with a characteristic width called the natural width.

In order to be able to do such high-resolution laser spectroscopy, two things have to be satisfied. First, the atomic resonance should not be artificially broadened. This can happen, for example, due to the Doppler effect in hot vapour, where the thermal velocity causes a frequency shift and broadens the line. Even with atoms at room temperature, the Doppler width can be 100 times the natural width, and can prevent closely-spaced levels from being resolved. The second requirement for high-resolution spectroscopy is that the tunable laser should have a narrow “linewidth”. The linewidth of the laser, or its frequency uncertainty, is like the width of the pen used to draw a curve on a sheet of paper. Obviously, you cannot draw a very fine curve if you have a broad pen.

It is in the above context that the Nobel citation mentions the work of the two laureates in laser-based precision spectroscopy. Their names are quite well known to anyone working in laser spectroscopy. In the early 1970s, Hänsch, then working at Stanford University with Arthur Schawlow (Nobel Prize for laser spectroscopy, 1981), pioneered the use of Doppler-free techniques such as saturation spectroscopy, particularly for spectroscopy in hydrogen. Around the same time, Hall developed many techniques to stabilize the frequency of lasers and reduce their linewidth. Today, two of the most popular techniques for laser stabilization are called the Hansch-Couillaud technique and the Pound-Drever-Hall technique, in honour of these scientists.

In 1976, Hall and coworkers used high-resolution laser spectroscopy in methane to observe for the first time the recoil-induced splitting of a line. In other words, when the molecule absorbs a photon of wavelength \( \lambda \), the photon momentum \( h/\lambda \) imparts a recoil to the molecule. This recoil velocity results in a frequency shift due to the Doppler effect. But this is a small effect, about 2 kHz in a frequency of \( 10^{14} \) Hz, and requires an extremely high resolving power. In the same year, Hänsch and Schawlow independently proposed that the momentum of laser photons could be used to cool atoms to very low temperatures, a technique that is now called...
The field of laser cooling has grown explosively in the last two decades, and two Nobel Prizes have been awarded; the first in 1997 for techniques of laser cooling (Chu, Cohen-Tannoudji, and Phillips), and the second in 2001 for using laser-cooled atoms to achieve Bose–Einstein condensation (Cornell, Ketterle, and Wieman).

But back to spectroscopy. Many advances in physics have been brought about by high-resolution spectroscopy of atoms. Indeed, one might argue that the most obvious manifestation of quantisation (or discreteness) at the atomic scale is the fact that atomic spectra show sharp spectral lines. The well-known Fraunhofer lines were first observed in the solar spectrum as dark lines using a spectrometer that was “high-resolution” for its time. In the early part of the twentieth century, Niels Bohr (Nobel Prize 1922) was able to explain such discrete lines by postulating that an electron in an atom was allowed only certain quantised values of angular momentum and energy. This led to the development of quantum mechanics as a theory in the atomic domain. Further measurements of atomic spectra at higher resolution revealed that many lines were actually doublets. A common example is the yellow light emitted by the ubiquitous sodium vapour lamp; it actually consists of two lines, called $D_1$ and $D_2$, which can be resolved and measured in a high school laboratory today. The origin of this splitting is the interaction between two types of electronic angular momentum – orbital and spin. In 1928, Dirac (Nobel Prize 1933) wrote down his famous equation to describe the electron, which incorporated its spin angular momentum in a natural way. However, even the very successful Dirac theory predicted that the $2S$ and $2P$ states of hydrogen have the same energy. A precise measurement of these levels by Lamb (Nobel Prize 1955) showed that their energies are slightly different, which is now called the Lamb shift. The discovery of the Lamb shift led to the birth of quantum electrodynamics (QED), for which the Nobel Prize was awarded to Feynman, Schwinger, and Tomonaga in 1965.

We thus see that improvement in precision almost always leads to new discoveries in physics. In recent times, one atomic transition that has inspired many advances in high-resolution spectroscopy and optical frequency measurements is the $1S - 2S$ resonance in hydrogen, with a natural width of only 1 Hz. Measurement of the frequency of this transition is important as a test of QED and for the measurement of fundamental constants. However, the wavelength of this transition is 121 nm, corresponding to a frequency of $2.5 \times 10^{15}$ Hz. Since the SI unit of time is defined in terms of the cesium radio-frequency transition at $9.2 \times 10^9$ Hz, measuring the optical frequency with reference to the atomic clock requires spanning 6 orders of magnitude! You can think of this as having two shafts whose rotation speeds differ by a factor of 1 million, and you need to measure the ratio of their speeds accurately. If we use a belt arrangement to couple the two shafts, then there is a possibility of errors in the ratio measurement due to belt slip. Instead, one would like to couple them through a gearbox mechanism with the correct teeth ratio so that there is no possibility of slip (see Figure 1). This is precisely what is achieved by the frequency comb.
Figure 1. Radio frequency to optical frequency link using a frequency comb. © The Nobel Foundation, 2005.

The basic idea of the comb technique is that periodicity in time implies periodicity in frequency. Thus, if you take a pulsed laser that produces a series of optical pulses at a fixed repetition rate, then the frequency spectrum of the laser will consist of a set of uniformly spaced peaks on either side of a central peak. The central peak is at the optical frequency within each laser pulse (carrier frequency), and the peaks on either side are spaced by the inverse of the repetition period (called sidebands). You can produce such a spectrum by putting the laser through a nonlinear medium such as a nonlinear fiber. The larger the nonlinearity, the more the number of sidebands. Around 1999, there was a major development in making nonlinear fibers; fibers with honeycomb microstructure were developed which had such extreme nonlinearity that the sidebands spanned almost an octave\(^2\). If you sent a pulsed laser (operating near 800 nm) through such a fiber, you would get a near continuum of sidebands spanning the entire visible spectrum. The series of uniformly spaced peaks stretching out over a large frequency range looks like the teeth of a comb, hence the name optical frequency comb. The beautiful part of the technique is that the comb spacing is determined solely by the repetition rate, thus by referencing the repetition rate to a cesium atomic clock, the comb spacing can be determined as precisely as possible. In 1999, Hänsch and coworkers showed that the comb spacing was uniform to 3 parts in \(10^{17}\), even far out into the wings.

\(^2\)Hall calls this development the dawn of a new epoch.
Thus the procedure to produce a frequency comb is now quite straightforward. One starts with a mode-locked, pulsed Ti:sapphire laser and sends its output through 20–30 cm of nonlinear fiber. The pulse repetition rate is referenced to an atomic clock, and determines the comb spacing. The carrier frequency is controlled independently, and determines the comb position.

But how does one measure an optical frequency using this comb? This can be done in two ways. One way is to use a reference transition whose frequency $f_0$ is previously known. We now adjust the comb spacing $\Delta$ so that the reference frequency $f_0$ lies on one peak, and the unknown frequency $f$ lies on another peak that is $n$ comb lines away, i.e. $f = f_0 + n\Delta$ (see Figure 2)$^3$. Thus by measuring $n$, the number of comb lines in between, and using our knowledge of $f_0$ and $\Delta$, we can determine $f$. This was the method used by Hänsch in 1999 to determine the frequency of the $D1$ line in cesium (at 895 nm). The measurement of this frequency can be related to the fine-structure constant $\alpha$, which is one of the most important constants in physics because it sets the scale for electromagnetic interactions and is a fundamental parameter in QED calculations.

However, the above method requires that we already know some optical frequency $f_0$. If we want to determine the absolute value of $f$ solely in terms of the atomic clock, the scheme is slightly more complicated. In effect, we take two multiples (or harmonics) of the laser frequency, and use the uniform comb lines as a precise ruler to span this frequency difference. Let us say we align one peak to $3.5f$, and another peak that is $n$ comb lines away to $4f$, then we have determined

$$4f - 3.5f = n\Delta,$$

which yields

$$f = 2n\Delta,$$

---

$^3$It is not necessary that the comb peak aligns perfectly with the laser frequency. A small difference between the two can be measured easily since the beat signal will be at a sufficiently low frequency.
so that we have $f$ in terms of the comb spacing. In 2000, Hänsch and coworkers used this method to determine the frequency of the hydrogen $1S - 2S$ resonance with an unprecedented accuracy of 13 digits. This was the first time that a frequency comb was used to link a radio frequency to an optical frequency.

Currently, one of the most important questions in physics is whether fundamental constants of nature are really constant, or are changing with time. For example, is the fine-structure constant $\alpha$ constant throughout the life of the universe or is it different in different epochs? Now, if you want to measure a very small rate of change $\dot{\alpha} = dc/\alpha dt$, then you can do it in two ways. You can take a large $dt$ so that the integrated change in $\alpha$ is very large. This is what is done in astronomy, where looking at the light from a distant star is like looking back millions of years in time. You can then compare atomic spectra from distant stars to spectra taken in the laboratory today. Alternately, if you want to do a laboratory experiment to determine $\dot{\alpha}$, then you have no choice but to use a small $dt$. Therefore, you have to improve the accuracy of measuring $\alpha$ so that even small changes become measurable. This is what has been done by Hänsch and his group. By measuring the hydrogen $1S - 2S$ resonance over a few years, they have been able to put a limit on the variation of $\alpha$. Similar limits have been put by other groups using frequency-comb measurements of other optical transitions. The current limit on $\dot{\alpha}/\alpha$ from both astronomy and atomic physics measurements is about $10^{-15}$ per year.

In the last few years, several optical transitions have been measured using frequency combs. The primary motivation is to find a suitable candidate for an optical clock to replace the microwave transition used in the current definition. An optical clock will “tick” a million times faster, and will be inherently more accurate. However, since the cesium atomic clock has an accuracy of $10^{-15}$, one has to measure the candidate optical transition to this accuracy to make sure it is consistent with the current definition. The race is on to find the best candidate among several alternatives such as laser-cooled single ions in a trap, ultracold neutral atoms in an optical lattice, or molecules. As shown in Figure 3, the accuracy of clocks has increased by several orders of magnitude in recent times. The applications for more precise clocks of the future range from telecommunications and satellite navigation to fundamental physics issues such as measurement of pulsar periods, tests of general relativity, and variation of physical constants.

In concluding this article, one of the authors (VN) would like to switch to the first person singular and make some comments on the motivations that underlie work in experimental physics. I recently attended a small reception in honour of John Hall after he won the Nobel Prize. In his speech, he mentioned that the thing he enjoyed most about being at NIST was that the management allowed him complete freedom to play with the latest “toys and gadgets”, pleasures that he has carried from his childhood. I remember that, as a child, I too was fascinated by mechanical and electrical gadgets, and the precision with which they were engineered. I think many of us take to experimental research precisely for this reason, that it allows us to take our childhood pleasures of playing with toys into adulthood, and even make a living out of this enjoyment! I cannot think of a greater advertisement for the young readers of this journal to take up a fulfilling career in research.
Figure 3. Improvement in the accuracy of clocks over the last millennium. © The Nobel Foundation, 2005.

Suggested Reading


1. Introduction

Special Relativity is now a hundred years old, and General Relativity is just ten years younger. Even the general literate public probably knows that these two theories of physics – STR and GTR – profoundly altered previous conceptions and understanding of space and time in physics. We will try to describe these changes beginning with earlier ideas which had served physics for almost 300 years, and seeing how they had to be modified in the light of accumulating experience.

2. Newtonian Space and Time, Inertial Frames

In his great work titled the ‘Principia’ published in 1685, Newton began by expressing in clear terms the natures of space and time as he understood them. The importance of his enunciation of their properties cannot be overestimated, because they provided something tangible which could be used as a basis for further work, and equally importantly which could be examined and criticised in a fruitful manner. Quoting from Einstein:

“... what we have gained up till now would have been impossible without Newton’s clear system”.

Newton was carrying forward what Galileo had begun. Let us start by reminding you of the implicit and explicit assumptions underlying Galilean–Newtonian physics, allowing for some overlaps in the interests of clarity.

1) In principle, there are infinitely many observers, in various states of relative motion, each with a global space-time reference frame. Each of them can assign space time coordinates to every event.

2) For all of them, space is the same: absolute, infinite and Euclidean; time is the same: absolute (and independent of space), linear, flowing uniformly.

In these statements, as Einstein says, ‘absolute’ means the following:

“... physically real ... independent in its physical properties, having a physical effect, but not itself influenced by physical conditions”.

3) Simultaneity of spatially separated events is absolute, the same for all observers.

4) Distance – spatial separation – between simultaneous events is absolute, the same for all observers.

5) Time lapse or interval between any two events is absolute, the same for all observers.

6) There exist special observers – called inertial observers with inertial reference frames – with respect to whom (sufficiently well) isolated bodies are unaccelerated, so Newton’s 1st Law holds. In these special frames, there are no inertial forces (see later), and the laws of mechanics take their simplest form.

7) Inertial frames can in principle be identified by (6): isolated bodies move uniformly in them.

   We should emphasize the importance of inertial frames. In a short essay on Kepler, Einstein says that while working on the laws of planetary motion in the Copernican framework, Kepler faced the problem of choosing a suitable reference frame in astronomical space with respect to which planetary orbits could be described. The sun could be chosen as the origin, but what about the directions of the axes? Einstein then says:

   “This was Kepler’s answer: The apparent motions of the planet Mars are known with great accuracy, including the time of its revolution about the sun (the “Martian year”). It is probable that at the end of each Martian year Mars is at the same spot in (planetary) space. If we limit ourselves for the time being to these points in time, then the planet Mars represents for them a fixed point in planetary space, a point that may be used in triangulation .... This is how Kepler gained the basis for formulating the three fundamental laws with which his name will remain associated for all time to come”.

   Even in the usual classroom derivation of Kepler’s laws starting with Newton’s Law of Motion – the 2nd Law – and universal gravitation, we have to use some coordinate system which we believe is inertial, and this is always taken to be a frame at rest with respect to the ‘fixed stars’. Einstein put it this way:

   “We may look upon the principle of inertia as established, to a high degree of approximation, for the space of our planetary system, provided that we neglect the perturbations due to the sun and planets”.

   From points 6) and 7) above we thus have (we hope!) in principle a method to identify inertial frames. In Galilean–Newtonian physics the way in which bodies acted upon one another was by action at a distance. In fact the only known fundamental force was gravitation described by the instantaneous inverse square law. So on a sufficiently isolated body, we may assume no true forces act, and in any inertial frame it moves uniformly in a straight line.

   So let us accept that a frame with the sun as origin and axes which are stationary with respect to the fixed stars is indeed inertial. If $K$ is an inertial frame, and a frame $K'$ moves uniformly in a straight line with respect to $K$, then $K'$ is also inertial. Essentially this was Galileo’s discovery. The relationships between space and time variables in two inertial frames...
Space, Time and Relativity

$K, K'$ are given by the equations of Galilean relativity. In the simplest case, for relative velocity $v$ in the $x$ direction, they are:

$$K \rightarrow K' : x' = x - vt, \quad y' = y, \quad z' = z, \quad t' = t.$$  \hspace{1cm} (1)

These are relatively simple mathematical equations. In the general case they are best written in three-dimensional vector notation:

$$K \rightarrow K' : \mathbf{x}' = R \mathbf{x} - v t + \mathbf{a}, \quad t' = t + b.$$  \hspace{1cm} (2)

Here $R$ is a (proper) three-dimensional rotation, $v$ the relative velocity, $\mathbf{a}$ a shift in the spatial origin, and $b$ a shift in the zero of time. Equations (2) describe a general element of the Galilei group $\mathcal{G}$, and involve ten independent real parameters. So the set of all inertial frames in Galilean–Newtonian physics is an infinite ten-parameter family, connected pairwise by transformations of the form (2). The associated relativity principle says that all inertial frames are equivalent for the description of mechanical phenomena, gravity included; Newton’s three Laws of Motion hold in all of them. One consequence of the Galilean transformation equations (1),(2) is the Newtonian law for addition of velocities.

3. Noninertial Frames, Inertial Forces

If now we imagine a noninertial frame, i.e., one which is accelerated with respect to (any) inertial frame, Newton’s 1st Law will not hold in it. Even isolated bodies will move with acceleration, and not uniformly in a straight line. We say that there are ‘inertial forces’ acting even on isolated bodies, causing them to accelerate. On a general body there are both ‘true forces’ due to other bodies acting at a distance, and inertial forces due to the frame being noninertial. The latter – since they are purely kinematic in origin – cause mass independent accelerations. The most familiar examples, due to uniform rotation with respect to an inertial frame, are centrifugal and Coriolis forces and accelerations.

Both because of its rotation and its revolution around the sun, a frame rigidly attached to the earth is noninertial. However in comparison, say, to the acceleration due to gravity on the earth’s surface, the noninertiality is quite small. The corresponding centrifugal accelerations are:

- due to earth’s rotation about its own axis:
  \[ v^2/R \sim 3.4 \times 10^{-3}g \text{ at equator}; \]

- due to earth’s orbital revolution:
  \[ v^2/R \sim 6 \times 10^{-4}g, \]
  \[ g \sim 981 \text{ cm/sec}^2. \]

So for many practical purposes – short distance projectile motion, chemistry, atomic physics, ...

... – a laboratory on the earth is a good approximation to an inertial frame.
Should one similarly ‘subtract away’ the sun’s rotation about the galactic centre to get an ‘even better’ inertial frame? Here the centrifugal acceleration is:

- Sun’s motion about galactic centre:
  \[ \frac{\nu^2}{R} \sim 2.4 \times 10^{-11} \text{g}, \]

but by this stage of precision it is time to move on to new concepts and revise our ideas.

We said that since inertial forces are kinematic in origin, they cause mass independent accelerations. From Galileo’s experiment in Pisa, and later Newton’s experiments with pendula, it was known that gravity is similar: it too causes mass independent accelerations. Thus by suitable choice of units, for all bodies the inertial and the gravitational masses are the same:

\[ m_{\text{inertial}} \equiv m_{\text{in}} = m_{\text{gravitational}} \equiv m_{\text{gr}}. \tag{3} \]

In Galilean–Newtonian mechanics this was accepted as a fact, so gravity was curiously similar to inertial forces in causing the same accelerations in all bodies. But again in the Galilean–Newtonian framework we have (we hope!) in principle a way to tell them apart – in a noninertial frame, a (sufficiently well) isolated body experiences (by definition) only inertial forces, not gravity. So by transforming to an inertial frame these inertial forces could be reduced to zero; but gravity, if initially present, would remain nonzero.

### 4. Electricity, Magnetism, STR

Now let us look briefly at the later historical development, over about two centuries, in the form of a time capsule:

<table>
<thead>
<tr>
<th>G-N mechanics</th>
<th>electricity, magnetism</th>
<th>Maxwell’s Equations</th>
<th>STR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gravity 1685</td>
<td>action at a distance (\sim 1800)</td>
<td>field theory (\sim 1865)</td>
<td>1905</td>
</tr>
</tbody>
</table>

Now to some comments. Evidently Maxwell accepted the Newtonian concepts of space and time, which implied Newton’s law for addition of velocities. But his electromagnetic field equations predicted a definite speed, \(c = 3 \times 10^{10}\text{cm/sec}\), for waves which he identified with light. Therefore his equations, unlike mechanics, could not be invariant under Galilean transformations; they could not be valid in all the inertial frames of mechanics. They could only hold in a subset of inertial frames, at rest relative to one another, and which could then define absolute rest. In any inertial frame outside this set, the speed of light would differ from \(c\) and even be direction dependent. Pictorially we have the situation as shown in Figure 1.
Space, Time and Relativity

Figure 1. This figure is not ‘drawn to scale’: The outer oval encloses a tenfold infinity of inertial frames. The inner circle encloses a sevenfold infinity of inertial frames at rest with respect to one another.

Maxwell thus thought that electromagnetic experiments could disclose departures from absolute rest defined by the aether. Thus basically, mechanics would obey the Galilean principle of relativity while electromagnetism would not, leading to a clash between them.

As is well known, the Michelson–Morley experiment gave a null result, not the one predicted by theory. All efforts to measure the speed of the earth with respect to the aether failed. This impasse was resolved by Einstein’s STR of 1905, leading to a new view of space and time, a new understanding of the harmony between mechanics and electromagnetism. The Maxwell equations were invariant under the Lorentz transformations, not the earlier Galilean transformations, so Einstein went on to modify mechanics to make it also Lorentz invariant. The two branches of physics were then ruled by a common new principle of relativity, leading to a deep unification of ideas.

In place of the elementary Galilean equations (1) we now have the Lorentz transformation equations.

\[ K \rightarrow K' : \quad x' = \frac{(x - vt)}{\sqrt{1 - v^2/c^2}}, \quad t' = \frac{(t - vx/c^2)}{\sqrt{1 - v^2/c^2}}, \]

\[ y' = y, \quad z' = z. \]  

(5)

This can be extended to a general ten parameter inhomogeneous Lorentz transformation or Poincaré transformation which replaces equation (2). The important consequences, the physical features of STR, are:

(1) The idea of globally defined inertial reference frames is retained, but these frames are related by Poincaré transformations, not by the elements of \( G \). We have a new ten-parameter relativity group, the Poincaré group \( P \).
In each inertial frame, space is three-dimensional, infinite and Euclidean, while time flows uniformly.

Under a nontrivial change of inertial frame corresponding to Lorentz transformations such as in equation (5), the separation of space-time into space and time gets altered; simultaneity, lengths and time intervals are also altered. But the concepts of absolute past and future are preserved.

The equations of mechanics have to be modified to be invariant under Poincaré transformations, to be in harmony with Maxwell’s field equations of electromagnetism.

5. From STR to GTR – Principle of Equivalence

We can say that the story so far was the first stage in the process of revising the Galilean–Newtonian ideas of space and time. Now let us look at the second stage, the road from STR to GTR, of course going only a short distance along this road! It is clear that STR is the correct language to handle electromagnetism, and all of mechanics except Newtonian gravity. In terms of the entries in equation (4), STR could handle all that stands to its left except gravity.

In 1907 while Einstein was preparing a review article on STR, two beautiful ideas occurred to him:

(a) there seems to be no compelling reason why the physical equivalence of reference frames should be limited to those having only uniform relative motions. Indeed in his review he asks:

“Is it conceivable that the principle of relativity also applies to systems that are accelerated relative to each other?”

(b) Galileo’s discovery of the equality of inertial and gravitational masses, equation (3), was not a mere curiosity but had deep physical significance – gravitational accelerations and inertial accelerations (felt in any noninertial frame), both being independent of mass, were of the same nature. Thus while the latter could be made to vanish by going to any inertial frame, the former (in a limited space-time region of uniform field) could also be made to vanish by going to a suitable uniformly accelerated noninertial frame. He called this ‘the happiest thought of my life’: that a person falling freely in the earth’s gravitational field would feel no weight at all! And then he expanded it to cover all physical phenomena in his celebrated Principle of Equivalence:

“. we shall therefore assume the complete physical equivalence of a gravitational field and a corresponding acceleration of the reference system. This assumption extends the principle of relativity to the uniformly accelerated translational motion of the reference system”.
Let us explore this idea a little bit, first against the background of Newtonian space time, and then see what further changes are brought about by STR.

The truth of the statement that in free fall we are weightless can be seen in the following way. When we stand on the earth’s surface we do feel our weight, the force exerted on our bodies by the earth’s gravitational field. But we certainly do not feel the gravitational pull of the sun, about $6 \times 10^{-4} \text{g}$ as we saw earlier, because (along with the earth) we are in ‘free fall’ as far as the sun is concerned. Here ‘free fall’ includes not just hurtling vertically downward but also motion in any other orbit in a given gravitational field, as happens with astronauts in a spaceship orbiting the earth. In Einstein’s own words:

“For experiments upon the earth tell us nothing of the fact that we are moving about the sun with a velocity of approximately 30 km a second.”

This insight into the meaning of equation (3) can thus be expressed as a mechanical Principle of Equivalence:

Mechanical phenomena in an inertial frame $K$ in which there is no gravity, when viewed from a uniformly accelerated (noninertial) frame $K'$, would appear the same as if $K'$ were inertial but a uniform gravitational field was present.

Here is yet another expression of the same idea:

If in an inertial frame $K$ there is a uniform gravitational field, then by going over to a noninertial frame $K'$ which has a compensating uniform acceleration with respect to $K$, we can reduce gravity in $K'$ to zero.

We should appreciate that in these statements we are transcending – going beyond – Galilean relativity since noninertial frames are involved. Under Galilean transformations a nonzero gravitational field remains nonzero and cannot be eliminated. We should next appreciate that in his Principle of Equivalence Einstein went beyond the above limited statements and replaced ‘mechanical phenomena’ by ‘all physical phenomena’. At that time this essentially meant including electromagnetic phenomena. Let us stress again that it is the nature of gravity that forces us to go beyond Galilean relativity.

The content of Einstein’s Principle of Equivalence can be conveyed in a simplified yet effective manner as follows. We consider two options – I : light is unaffected by gravity; II : light is affected by gravity. Then we consider the behaviours of matter and of light in four different physical situations and reference frames: inertial frame without gravity; uniformly accelerated frame without gravity; inertial frame with uniform gravity; and freely falling frame in uniform gravitational field. We can then see table 1 emerging:

In each box, [a] to [h], we have given first the nature of motion of matter, then the properties of light rays. The second row arises from the first by purely kinematic reasoning as there is no gravity. That is why [a] = [e] and [b] = [f]. When we bring in gravity, since light behaves differently in Option I and in Option II, we see why [c] $\neq$ [g] and [d] $\neq$[h]. The fourth row arises from the third by ‘cancellation of gravity by suitable acceleration’. Now we can see that in Option I, boxes [a], [b], [c], [d] are all distinct, so the behaviours of matter plus light can distinguish among the frames 1, 2, 3 and 4. But with Option II, which is Einstein’s Principle of Equivalence, acceleration can mimic and even cancel gravity! So [f] = [g]; and [e] = [h]
Table 1.

<table>
<thead>
<tr>
<th></th>
<th>I: Gravity affects matter not light</th>
<th>II: Gravity affects matter and light</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Inertial frame, no gravity</td>
<td>Unaccelerated</td>
<td>Unaccelerated</td>
</tr>
<tr>
<td></td>
<td>Straight [a]</td>
<td>Straight [e]</td>
</tr>
<tr>
<td>2 Uniformly accelerated frame, no gravity</td>
<td>Accelerated</td>
<td>Accelerated</td>
</tr>
<tr>
<td></td>
<td>Bent [b]</td>
<td>Bent [f]</td>
</tr>
<tr>
<td>3 Inertial frame, uniform gravity</td>
<td>Accelerated</td>
<td>Accelerated</td>
</tr>
<tr>
<td></td>
<td>Straight [c]</td>
<td>Bent [g]</td>
</tr>
<tr>
<td>4 Freely falling frame, uniform gravity</td>
<td>Unaccelerated</td>
<td>Unaccelerated</td>
</tr>
<tr>
<td></td>
<td>Bent [d]</td>
<td>Straight [h]</td>
</tr>
</tbody>
</table>

because of the cancellation in the frame 4. Matter and light both experience gravity, only two combinations of behaviours are realized; we cannot distinguish frame 1 from 4, or 2 from 3!

6. Gravity and STR

After having thus glimpsed Einstein’s reinterpretation of gravitation in the Newtonian space-time framework – giving it an almost kinematical meaning – let us see what happens when STR and its lessons on the natures of space and time are brought in. In STR, simultaneity of events and instantaneous spatial separations of bodies are both nonabsolute notions, they both change from frame to frame. So also the isolation of bodies and action at a distance lose meaning. The spirit of STR is that all interactions must be local and be transmitted with finite speed. So the Newtonian way of trying to distinguish between gravitational and inertial accelerations is lost. In fact Einstein criticized the idea of inertial frames already in Newton’s scheme in this way:

“The weakness of the principle of inertia lies in this, that it involves an argument in a circle: a mass moves without acceleration if it is sufficiently far from other bodies; we know that it is sufficiently far from other bodies only by the fact that it moves without acceleration.”

So this leaves us no choice but to accept the following: when we go from Newtonian mechanics to STR, and then bring in gravity, there is no way to separate the total mass independent accelerations of bodies into unique gravitational and inertial parts, to disentangle them. So we cannot reach or identify global inertial frames. With inertial frames lost, it becomes meaningless to say that a particular frame is noninertial! We cannot say whether mass independent accelerations are due to ‘real’ gravity or to inertial forces or how much of each, we cannot draw a distinction between them. So, we abolish the distinction, and give up the idea of global inertial frames. In greater detail we can say the following.
Space, Time and Relativity

In Newtonian mechanics, inertial frames were globally well defined, and were definitely useful. So in any generalization, we expect that inertial frames should find a place with possible modifications or limits of range of validity. When we try to bring STR and gravitation together, the spirit of the former tells us: Reference frames are only locally definable. Leaving aside all other forces, as well as interactions between test bodies, all bodies experience the same accelerations at the same space-time locations. Travelling alongside any such ‘freely falling’ body, we can create a local reference frame in the sense of STR in some neighbourhood, defined upto a Lorentz transformation. In such a frame, all other nearby bodies are also unaccelerated, so it is locally an inertial frame. In a limited space-time region viewed from such a frame – there is no gravity, it has been ‘reduced to zero’. More precisely, locally over a region of uniform field:

- Nongravitational phenomena in freely falling frame ≡ phenomena in inertial frame in absence of gravity.

In a general local but not freely falling frame, all bodies experience common accelerations. We can call it gravitational or inertial, we agree as a convention to call it gravity.

You can sense that in trying to reconcile STR with Newtonian gravity, Einstein found that he had to transcend both of them! Inertial frames remained only as a local concept; in general, different local inertial frames do not all combine or mesh together to form a single globally well-defined inertial frame.

7. Non-Euclidean Geometry

We have come quite far from the initial Newtonian ideas of absolute space, absolute time and inertial frames. But there was still a long way to go before Einstein could complete his GTR. Already in 1907 and 1911, arguing on the basis of his Principle of Equivalence of uniform gravity and uniform acceleration, he had shown that light bends in a gravitational field; its frequency is red shifted in passing from regions of lower to regions of higher gravitational potential; clocks run slow and even light slows down in regions of lower gravitational potential. However the calculation of the deflection of light was later found to be incorrect by a factor of 2.

Then in 1912 he switched from uniform acceleration of the reference frame to uniform rotation, and found (using STR) that in such a frame the laws of geometry are not Euclidean! This was something new, and by a leap of the imagination he was led to this idea – the description of the geometry of space-time is a description of gravitation. To pursue this idea and to create a new comprehensive theory of gravitation generalizing Newton’s theory involved a heroic struggle which ended only in November 1915. Along the way he learnt – with much struggle – that in the process of reconciling STR and gravitation, space time coordinates lose their immediate physical meanings in terms of measurements with standard rods and clocks. They become no more than markers like telephone numbers, only helping us distinguish distinct events from one another. Just as there are in general no globally defined privileged inertial

127
reference frames, so also in general there are no privileged coordinate systems either. This meant that the new theory of gravitation should treat all possible choices of coordinate systems on an equal footing, favouring none. By another leap of the imagination, this invariance of the laws of gravitation under arbitrary coordinate changes became for him an expression of the Principle of Equivalence.

The action of gravity on light and matter was in effect conceived of as a two-step process. First, all sources of energy and momentum – both matter and light – act as sources of the gravitational field and influence the geometry of space time. Second, matter and light move and propagate in space time following its geometrical rules! To realize all this Einstein had to master and extend the existing mathematics of Riemannian geometry. At the end, space time was not as in Newton’s view a passive unchangeable stage on which other physical phenomena were played out, it was an active participant in the proceedings. In Einstein’s words:

“It is contrary to the mode of thinking in science to conceive of a thing .... which acts itself, but which cannot be acted upon.”

There can be no unmoved mover!

At the end of the story, as we saw with the inertial frame concept, STR also reappears as a locally valid concept in GTR. Its familiar descriptions of space and time and the action of Lorentz transformations are all valid locally in freely falling frames in limited space time regions or patches, but they cannot be all put together seamlessly to give a global system. From a physical point of view, all nongravitational phenomena must obey the rules of STR, so this theory acts as a restrictive principle. Once a theory passes the STR test, it can more or less automatically be woven into the fabric of GTR.

8. Concluding Comments

As we mentioned at the beginning, STR and GTR are close to a century old. Many major developments have occurred since then, both on the theoretical side and in experiment. Inspired by the beauty of GTR, physicists like Hermann Weyl and Theodore Kaluza and Oskar Klein suggested generalizations of great mathematical beauty. Though their specific proposals turned out to be untenable, their ideas survived and have inspired later work. Einstein himself and Schrödinger too tried to extend GTR in various ways, though without significant success.

We hope the account given here, though incomplete, will inspire our readers and create the confidence to pursue the study of GTR and of the various later developments up to recent times.

Suggested Reading


Space and Time in Life and Science*

Vasant Natarajan, V Balakrishnan and N Mukunda

Space and time are concepts that seem to be embedded in our very consciousness. As we grow up, our ‘intuitive understanding’ of these concepts seems to grow as well. And yet the fact is that our understanding of space-time in the deepest scientific sense is far from complete, although we have covered a considerable distance along the route. There may still be many surprises awaiting us on the road ahead.

1. It Began with Geometry

All of us have some intuitive ideas about the natures of space and time in which we are embedded. Space appears to be the stage on which all events, experiences and phenomena take place, while time is like a background against which this happens. All objects, including ourselves, exist in space and change with time. In this article, we shall describe in simple and qualitative terms how our understanding of space and time has developed over the centuries, and what we have learnt about their properties. We will see that many strands come together in this story – not only from mathematics and physics, but also, in some important respects, from biology.

In one of his essays, Schrödinger\(^1\) says this about the eminent philosopher Immanuel Kant (1724–1804): “Kant ... termed space and time, as he knew them, the forms of our mental intuition – space being the form of external, time that of internal, intuition.” We begin with this quotation both because it is so profound and because some of Kant’s other ideas will appear later on. All of us have at least a preliminary or common-sense understanding of the nature of space through our senses, mainly sight and touch.\(^2\) We see objects arranged in various ways relative to one another; we touch them if they are nearby; we develop a feeling for shape, distance, perspective; and so on. Thus, through direct sensory experiences, space seems to have a definite character, even though the idea of space by itself is quite abstract. Time, on the other hand, is more subtle. We cannot reach it through our senses directly; we can experience it and be aware of its passage only internally through the mind in a non-sensory way. Because we cannot see or touch time, extension in space seems a little easier to grasp than duration in time. Thus, space has to be abstracted from external experience, time from internal experience. Over and above this subtlety, space and time are really not simple concepts at all. An obvious


\(^1\)The great physicist Erwin Schrödinger (1887–1961; Nobel Prize in Physics, 1933), who discovered wave mechanics and was one of the founders of quantum mechanics, was also a wonderfully lucid writer on many profound subjects. His famous books include *Science and Humanism*, *Expanding Universes*, *Statistical Thermodynamics*, *What is Life?*, *Mind and Matter*, *My View of the World*, *Space-Time Structure*, and *Nature and the Greeks*.

\(^2\)It is an interesting fact that about 80% of our sensory input comes from our eyes, and about 40% of our brain capacity is devoted to processing visual information; we are very visual creatures.
Ancient Babylon and Egypt are well known as the civilizations in which the subject of geometry originated. However, there were other nodal centres as well. You are undoubtedly familiar with some of the great and unique achievements of ancient Indian mathematics, such as the invention of the decimal place-value system of numbers, negative numbers, and zero. It is important to note that these remarkable mathematical advances (which took place from about 500 AD onwards) were preceded by equally significant achievements in the Indian sub-continent in even earlier times. In recent years, much new light has been shed on the scope and advanced nature of truly ancient Indian mathematics. It is being gradually recognized that ancient India (especially the Indus Valley Civilization) was in many respects (including mathematics) fully the peer of the ancient Egyptian and Babylonian civilizations. The historians of science J J O’Connor and E F Robertson (see http://www-history.mcs.st-andrews.ac.uk/HistTopics/Indian_mathematics.html) write,

“... the study of mathematical astronomy in India goes back to at least the third millennium BC and mathematics and geometry must have existed to support this study in ancient times.” They go on to quote V G Childe in New Light on the Most Ancient East (Routledge and Kegan Paul Ltd., London, 1952): “India confronts Egypt and Babylonia by the 3rd millennium with a thoroughly individual and independent civilization of her own, technically the peer of the rest.” What is clear is that astronomy was a prime motivating factor behind the development of geometry, trigonometry and related topics in ancient India.

fundamental difference between the two is that space appears to be ‘controllable’ – in the sense that we can move from one point to another, with seemingly arbitrary freedom. On the other hand, time seems to be something over which we have no control – it just marches inexorably on. We are caught up in this march with no apparent choice. The old adage – ‘time and tide wait for no man’ – captures this notion perfectly. This also means that we have memories of what has happened up till now, but no idea of what the future holds (contrary to what soothsayers and horoscope-writers might tell you!)

Let us begin with space. The word geometry, as you know, means ‘measuring the earth (or land)’. The origins of geometry go back several millennia to the great Egyptian civilization (see also Box 1). Geometry arose in ancient Egypt from the repeated need to survey and re-establish boundaries of land holdings after the annual floods of the river Nile. Here is a passage from the Greek traveller and historian Herodotus (482?–434? BC):

“The King moreover (so they said) divided the country among all the Egyptians by giving each an equal square parcel of land, and made this his source of revenue, appointing the payment of a yearly tax. And any man who was robbed by the river of a part of his land would come to Sesostris and declare what had befallen him; then the King would send men to look into it and measure the space by which the land was diminished, so that thereafter he should pay the appointed tax in proportion to the loss. From this, to my thinking, the Greeks learned the art of measuring land.”

So geometry was the product of practical human needs to measure space, much as trade and commerce led to arithmetic. Similarly, the practical need to measure time (for example, to keep track of daily sunrises or the annual floods) gave rise to clocks and calendars. Human scientific and technological progress has required ever-more accurate clocks. A brief history of timekeeping over the centuries is presented in Box 2.
Space and Time in Life and Science

Box 2. A Brief History of Timekeeping

Timekeeping on earth by nature is as old as the planet itself, since all it requires is a periodic process. The periodic rotation and revolution of the earth give rise to daily and yearly cycles, which have been used by nature as a clock much before human beings evolved. Many living organisms including humans show daily (or diurnal) rhythms regulated by the sun, and seasonal patterns that repeat every year. It is therefore natural that the earliest man-made clocks also relied on the sun. One example was the sundial, which consisted of a pointer and a calibrated plate on which the pointer cast a moving shadow. Of course this worked only when the sun was shining. The need to tell the time even when the sun was not out, such as on an overcast day or at night, made it necessary to invent other kinds of clocks. The earliest all-weather clocks were water clocks, which were stone vessels with sloping sides that allowed water to drip at a constant rate from a small hole near the bottom. Markings on the inside surface indicated the passage of time as the water level reached them. Other clocks were used for measuring small intervals of time. Examples included candles marked in increments, oil lamps with marked reservoirs, hourglasses filled with sand, and small stone or metal mazes filled with incense that would burn at a constant rate.

Time measurements became significantly more accurate with the advent of the pendulum clock in the 17th century. Galileo had studied the motion of the pendulum as early as 1582, but the first pendulum clock was built by Christiaan Huygens (1629–1695) only in 1656. As we know from high-school physics, the time period of a pendulum executing small-amplitude oscillations depends only on its length and the acceleration due to gravity. Huygens’ clock had an unprecedentedly small error of less than 1 minute per day. Later refinements allowed him to reduce it to less than 10 seconds a day. While very accurate compared to previous clocks, pendulum clocks still showed significant variations, because a change of just a few degrees in the ambient temperature could change the length of the pendulum (owing to thermal expansion). Many clever schemes were therefore devised in the 18th and 19th centuries to compensate for such changes in length.

In the history of clockmakers, the name of the great horologist John Harrison (1693–1776) stands out. He constructed many ‘marine chronometers’—highly accurate clocks that were used on ships to tell the time from the start of the voyage. A comparison of local noon (that is, the time at which the sun was at its highest point) with the time on the clock (which would give the time of the noon at the starting point) could be used to determine with precision the longitude of the ship’s current position. The British government had instituted the Longitude Prize so that ships could navigate on transatlantic voyages without getting lost. Harrison designed and built several prize-winning clocks based on the oscillations of a balance wheel. These maritime clocks had to maintain their accuracy over lengthy and rough sea voyages with widely-varying conditions of temperature, pressure and humidity. The amazing features of Harrison’s chronometer are exemplified by the fact that, on a voyage from London to Jamaica, the clock only lost 5 seconds, corresponding to an error in distance of 1 mile! Harrison’s heroic story is brilliantly narrated by Dava Sobel in her book, Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time (Penguin Reprint Edition, 1996).

However, all such mechanical clocks suffered from unpredictable changes in timekeeping accuracy owing to wear and tear of the moving parts. Clock accuracies improved dramatically in the first half of the 20th century with the development of the quartz crystal oscillator. These clocks rely on a property of quartz called piezoelectricity, wherein a mechanical deformation produces an electrical signal (voltage), and vice versa. A vibrating crystal therefore produces a periodic electrical signal which acts as a precise clock. (For practical reasons, the frequency of the crystal or ‘resonator’ is usually adjusted to be as close to 32768 or $2^{15}$ Hz as possible. This is then stepped down by successive digital dividers to 1 Hz.) Today, we have the ubiquitous and inexpensive quartz wristwatch that is accurate to within a few seconds over a month, and one does not have to worry about winding the clock every day or replacing the battery more than once a year or so.

Continued...
Life in our technology-driven society needs ever more precise timekeeping. Computers, manufacturing plants, electric power grids, satellite communication—all these depend on ultra-precise timing. The Global Positioning System (GPS), which determines the position of a receiver by triangulating with respect to the three nearest satellites, requires timing accuracy of 1 part in $10^{12}$. Modern atomic clocks, based on the oscillation frequency within an atom, have this kind of accuracy. But scientific and technological needs keep pushing the requirement ever higher. Today’s best atomic clocks have an accuracy of 1 part in $10^{14}$, or an error of 1 second in 3 million years. We have indeed come a long way from the pendulum clock of just a few centuries ago.

The science of geometry must have developed gradually. From the Egyptians this knowledge of length, direction, area and shape passed into the hands of the Greeks, who seem to have had a special gift for this subject and for abstract thinking in general. Around 300 BC, this mathematical knowledge was codified and presented in a beautifully organized form by Euclid of Alexandria (circa 325–265 BC). His *Elements* became the supreme pattern—or paradigm—for mathematical precision and rigour, indeed for all intellectual work, for many centuries. Starting from a few definitions and axioms (or postulates), which seemed self-evident and consistent, and proceeding by a series of logical steps, many theorems were derived as consequences. Thus, results which were implicit in the starting axioms were made explicit, or brought out into the open, by logical arguments. From the economy of the axioms one could pass to the wealth of derived consequences. Achieving this style of thinking was a great intellectual step forward.

### 2. The Newtonian Worldview

We now pass over several centuries and come to the birth of modern science. Galileo Galilei (1564–1642) is generally regarded as the first modern scientist. In his writings he says clearly that mathematics is the language of Nature: “*It is written in mathematical language.*” Soon after, in the latter part of the 17th century, the first clear statements about space and time were given by Isaac Newton (1643–1727) in his monumental three-volume work on the mathematical principles of natural philosophy, usually referred to as the *Principia* for short. The style of this work is also ‘Euclidean’—a small number of definitions and laws are given at the beginning, and many theorems are then proved as consequences. In fact, Newton went so far as to present all his derivations using geometrical arguments (“*con more geometrico*”).

About space and time, he says that although they are commonly thought of in relation to material objects placed in them, he would define them on their own. Here are his famous statements from the opening pages of the *Principia*:

“Absolute, true, and mathematical time, of itself, and from its own nature, flows equally without relation to anything external, and by another name is called duration.”

“Absolute space, in its own nature, without relation to anything external, remains always similar and immovable.”

Newton assumed, of course, that space obeyed the laws of Euclidean geometry.
It must be emphasized that, for the further development of science, it was very important that Newton explained so clearly his views about space and time. They could be examined in a careful way, and made the basis for further progress. Newton’s views were already criticized in his own lifetime by the philosopher George (Bishop) Berkeley (1685–1753) and the great mathematician, physicist and philosopher Gottfried Wilhelm von Leibniz (1646–1716). The latter felt that there was no need to think of space apart from the relation of material objects to one another and declared: “I hold space to be something purely relative as time is.” Much later, in the late 1800’s, the Austrian physicist-philosopher Ernst Mach (1838–1916) also critically examined Newton’s views, not only about space and time, but about mechanics as well. All this only emphasizes the crucial role and importance of Newton’s unambiguous statements. Albert Einstein (1879–1955; Nobel Prize, 1921) himself acknowledged this fact when he stated in one of his essays, “... what we have gained up till now would have been impossible without Newton’s clear system.”

This Newtonian picture of space and time served physics extremely well for over two centuries. The word ‘absolute’ in his statements is important: it strongly suggests that space and time themselves are unaffected by all the physical processes that occur in them. Once again, Einstein conveys the essence of this idea very well: he says ‘absolute’ means “... physically real ... independent in its physical properties, having a physical effect, but not itself influenced by physical conditions.” Following his statements on the absolute natures of space and time, Newton gives his three laws of motion, which we learn at school. Much later, towards the end of the Principia, he introduces his Law of Universal Gravitation. We do not go into Newtonian mechanics or the inverse square law of gravitation, as our focus is on space and time. But we mention here that general relativity (Einstein’s theory of gravitation, unveiled by him in 1915–16) changed precisely the Newtonian assumption of the absoluteness of space and time. We will return to this point later on.

3. Space, Time and Human Intuition

Building on the foundations provided by Galileo and Newton, both mechanics and astronomy made remarkable advances during the 18th century. Towards the end, the laws of electricity and magnetism also seemed to follow the general Galilean–Newtonian pattern. Kant was so impressed by these successes that he tried to present a philosophical justification for them. His main idea was that of the synthetic a priori principles. These are non-trivial statements about the properties of nature which are necessarily true or binding, but they are regarded as not derived from experience. That is the meaning of the phrase “a priori”, namely, something known even before experience. As far as we are concerned here, Kant regarded the concepts of

---

3Thus, Newton treated space and time as quantities that appeared in his laws of motion as independent variables, while other physical quantities were functions of these. We have already noted the controllability of space, in contrast to the ‘inevitability’ of time. These ideas are embodied in Newton’s conception by supposing that objects exist in space, but evolve in time. The evolution in time was determined by his theory of fluxions, which we study today as differential calculus.
If a single loud message comes across from the developments in the physical sciences and the life sciences during the last one hundred years or so, it is surely this: *Do not try to second-guess nature!* As Sherlock Holmes put it so pithily, in *A Study in Scarlet*, "It is a capital mistake to theorize before you have all the evidence. It biases the judgment." This dictum can hardly be over-emphasized when it comes to unlocking the secrets of nature. Time and again, nature has corroborated the great geneticist and biologist J B S Haldane (1892–1964) who said, “Now my own suspicion is that the Universe is not only queerer than we suppose, but queerer than we can suppose”. Both relativity and quantum mechanics, the two cornerstones of our current understanding of the physical universe, show us in dramatic fashion how misleading our ‘intuition’ can be, with regard to the most fundamental aspects of nature. We know now that (i) space is *not* Euclidean, (ii) the Newtonian assumptions of space and time as separate absolutes are not strictly correct, and (iii) space-time is quite complex even at the level up to which we understand it currently (and is likely to turn out to be even more so at the fundamental level). Now, philosophy may lead to some insights into, and systematization of, scientific knowledge. But it is not, in *itself*, a reliable way to discover scientific facts. This is why it is not appropriate to begin scientific enquiry with a pre-conceived set of notions that are postulated to be ‘absolute truths’ simply because they appear to be self-evident or based on ‘common sense’. As Einstein pointed out, “Common sense [in this context] is the collection of prejudices acquired by age eighteen”.

As we have explained subsequently in the main text, what is meant by all this is the following: the ‘intuition’ we acquire with regard to the natural world around us is based on the behaviour of objects and phenomena we encounter in it in everyday life, and perceive with our senses. For both these reasons, the information thus gathered is restricted to a rather narrow range of variation (or window) of the values of physical quantities such as mass, length, time, velocity, momentum, force, and so on. This is the so-called ‘world of middle dimensions’. It informs our beliefs and our expectations (“intuition”) regarding the nature of the physical world. Broadly speaking, comprehension of the world of middle dimensions at a certain level, and the ability to deal with it, has been hard-wired into our brains for reasons based on survival in the evolutionary sense. However, there is absolutely no reason to expect that nature itself would be restricted to this narrow range of values of physical quantities, and indeed it is not — in a most spectacular way. It turns out that our own windows of mass, length and time only encompass about seven or eight orders of magnitude, while our present state of knowledge shows that nature and natural phenomena extend over at least *ten times that number of orders of magnitude*. It is hardly surprising, then, that our naive, essentially hard-wired, intuition and common sense are totally inadequate to comprehend the universe unaided. We need the assistance of better, finer and more powerful tools to do so. These are provided by our instruments (to probe nature) and by our mathematics (to analyze our findings).

absolute and separate space and time, as Newton had viewed them, as synthetic *a priori* principles; so also the uniform flow of time and the Euclidean geometry of space.⁴

Kant claimed that all these ideas are already present in our minds before we have contact with, and experience of, nature; and that they must be necessarily valid because science would be impossible without them. The modern understanding of this matter comes, surprisingly enough, from evolutionary biology. It is largely due to the work of the zoologist and ethologist Konrad Lorenz (1903–89; Nobel Prize, 1973) around 1940, further developed and described by the molecular biologist Max Delbrück (1906–81; Nobel Prize, 1969). The basic idea is

---

⁴As it happens, he was wrong on both counts (for reasons he could not have foreseen). A profound lesson about nature is to be learnt here. See Box 3.
that as species evolve biologically over long periods of time, governed by natural selection, they acquire and retain those capabilities which are most useful for survival. Among these are the abilities to recognize the most important physical features of the world around us, corresponding to our own scales of size and time. These include the properties of space and time as stated by Newton, features of nature that are valid (to a high degree of approximation) at our own level. Thus, these abilities are the result of slow experience of nature by the species over immense stretches of time, but to the individual member of the species they are available at and soon after birth. This is why they seem to be given a priori, available before any actual experiences during life. From the present perspective, the main point is that all this relates only to our world of normal everyday experiences in life. It does not apply either to the sub-microscopic world of atoms or, at the other end, to the cosmos of galaxies, since these remote parts of nature do not appear to be immediately relevant to our day-to-day existence.

We referred earlier to our intuitive ideas about space and time. Now this intuition, and the way it is built up, also has a fascinating explanation. It is not something available ready-made at birth. Rather, it is acquired in the early years of infancy out of experience, using the abilities and apparatus of sensory perception. This is what biological evolution gives to each of us at birth – not knowledge about the important features of nature around us, but the capacity to learn about them.

The experiments of the child psychologist Jean Piaget (1896–1980) in the 1950’s led to remarkable insights into the learning process in infancy and childhood, in particular into learning about space and time. It seems that the regularity of events occurring soon after one another in time is recognized and remembered very early on; while the idea of space comes a little later. Several years later come the concepts of universal space and time common to, and containing, all things and events. Piaget suggested that the order in which children acquire concepts of space is opposite to what is taught at school and college. Broad concepts that are topological in character – the idea of nearness, of the deformability of shapes into one another, and so on – are acquired first; next, concepts of projective geometry such as direction and perspective; and only at the end, the underlying concepts of distance, shape, size, congruence, etc., on which Euclidean geometry is based. But in formal teaching the sequence followed is exactly the opposite. Euclidean geometry is always taught first, at an elementary level. This is followed, at a more mature level of instruction in mathematics (or applications of mathematics), by the concepts of projective geometry. Topology is introduced at an even more advanced level of instruction!

To summarize: We are born with the capacity to acquire knowledge about nature, not the knowledge itself. Natural selection has tuned these capacities to the actual properties of the world of those dimensions (or sizes, or magnitudes) that are directly and immediately relevant to our survival in the natural environment. But we retain no later memories of this learning process that we go through in our infancy; and when we reach adulthood we imagine that we were born with an ‘intuitive’ knowledge of nature in this domain. Returning to Lorenz, his ideas are well captured in these remarks by Delbrück:
“... there can be little doubt that our spatial concepts develop in childhood by way of an adaptation to the world in which we live ... the cognitive capacity that permits man to analyze space in geometric terms must, to a large extent, be evolutionarily derived.”

In an appendix titled ‘Physics and Perception’ in his book Special Relativity, the physicist-philosopher David Bohm (1917–94) describes how a child gradually builds concepts of space and time outside of herself, and the role of memory in the construction of the ideas of past, present and future time. Only by the age of ten or more does a child develop the mental capacity to conceive of a universal space and a universal time common to all objects and events that are perceived. This just goes to show how deep and genuinely difficult these concepts are, and that we are not born with them.

4. Beyond Euclidean Geometry

After this brief excursion into biology, let us come back to mathematics. Euclid’s geometry has been admired and studied for over two thousand years. As we have already said, the approach starts with a very small number of fundamental postulates, and builds upon them by logical argument. But from very early times people were puzzled by the status of the fifth postulate of Euclid, also known as the ‘parallel postulate’. Was it independent of the first four postulates, or could it be derived from them (in which case there would be no need for an additional fifth postulate)? Eminent thinkers over the centuries pondered over this problem, from the astronomer and geographer Claudius Ptolemy (circa 85–165 AD) to the famous mathematician Adrian-Marie Legendre (1752–1833). Many equivalent reformulations of this postulate were found, but the unchanging belief was that the geometry of space had to be Euclidean, for the sake of consistency. Along the way, René Descartes (1596–1650) invented the method of coordinates to represent points in space, so that the powerful mathematics of algebra became available for tackling problems in geometry. As we have seen, Kant subsequently expressed the view that physical space had necessarily to obey the laws of Euclidean geometry, and that there was no other option.

A few decades after Kant, in the early 1800’s, the final resolution of this long-standing problem occurred. The fifth postulate was indeed independent of the others, since it could be replaced in a self-consistent manner by other assumptions, and one could think of alternatives to Euclid’s geometry. This breakthrough was achieved independently by three mathematicians—the incomparable Carl Friedrich Gauss (1777–1855), around 1824, building on his theory of

---

5 Let $L$ be a straight line of infinite extent at both ends, and let $P$ be a point that does not lie on $L$. The parallel postulate asserts that there is one- and only one- straight line passing through $P$ that is parallel to $L$. See C R Pranesachar, *Euclid and 'The Elements'*, Resonance, Vol. 12, No. 4, 2007.

6 Like so many fundamental advances, the idea of coordinate geometry first appeared in the appendix to a book! The title of Descartes’ book, published in 1637, translates to *Discourse on the Method of Properly Conducting One’s Reason and of Seeking the Truth in the Sciences*. Nearly four hundred years later, the titles of scientific books and papers today are somewhat more specialized, if less imposing!
curved surfaces; Nikolai Ivanovich Lobachevsky (1792–1856), in 1829; and János Bolyai (1802–60) in 1832. Thus was born the subject of non-Euclidean geometry.\(^7\)

Around the middle of the 19th century, Gauss’ gifted student and one of the greatest of mathematicians, Bernhard Riemann (1826–66), took the next big step: the creation of differential geometry. He presented his ideas in his famous probationary lecture titled *On the hypotheses that lie at the foundations of geometry*, given at Göttingen on June 10, 1854. The fundamental insight was to determine the geometry of a space starting from the definition of (the square of) the interval or distance between nearby points in the space. From this, the concepts of tensors, parallel transport, intrinsic differentiation, connection, curvature, etc., could all be developed. This was a stupendous achievement, and many great contributions from a galaxy of geometers followed.\(^8\) In particular, Riemann’s methods were powerful enough to deal with spaces of any number of dimensions.

### 5. Maxwell and the Road to Special Relativity

We now come back to physics. In 1865, not long after Riemann developed differential geometry, one of the greatest scientists of all time, the brilliant physicist James Clerk Maxwell (1831–79), put together his system of coupled partial differential equations for time-dependent electric and magnetic fields—the basis for all electromagnetic phenomena. He then discovered the possibility of freely propagating electromagnetic waves, identified them with light, and thus unified three fields of classical physics—electricity, magnetism and optics. Maxwell assumed, tacitly, the Newtonian picture of separate and absolute space and time, and also believed in the existence of a medium, the ‘ether’, to carry electromagnetic waves.\(^9\)

However, it soon became clear that there was a deep-seated contradiction between Newton’s mechanics and the set of equations encapsulating this new understanding of electromagnetism. The equations of mechanics were unchanged in form (or ‘form-invariant’) under a set of transformations of the space and time coordinates called *Galilean transformations*. These transformations allow for the arbitrariness in the origin of the coordinate axes and of time, in the alignment or orientation of the coordinate axes, and in the choice of the inertial frame of

---

\(^7\)There is evidence that Gauss had conceived the idea of non-Euclidean geometry already in 1792, as a precocious teenager. But his first mention of the subject is in a letter written in 1824. Lobachevsky made his discovery in 1826, and published his work in 1829. This was the first formal publication of non-Euclidean geometry. Bolyai arrived at the idea in 1823, but the first publication of his results was dated 1832. All three mathematicians dealt with what we now call hyperbolic geometry, applicable to spaces of negative curvature. Subsequently, Riemann completed the picture with spherical geometry, applicable to spaces of positive curvature.

\(^8\)The list includes Bruno Christoffel (1829–1900), Gregorio Ricci-Curbastro (1853–1925), his son-in-law Tullio Levi-Civita (1873–1941), and Luigi Bianchi (1856–1928), among others. Like the Soviet tradition in the theory of probability, and the later Indian tradition in statistics, Italy has had a fine tradition in geometry.

\(^9\)For a brief but excellent overview of Maxwell’s ideas and discoveries, see the article titled *James Clerk Maxwell: a force for physics* by F Everitt, at the website http://physicsworld.com/cws/article/print/26527. It is interesting to note that the terms *field* and *relativity*, both of which are so basic to contemporary physics, were first introduced in their current senses by Maxwell.
reference used to analyze the mechanical system. Time, of course, is assumed to ‘flow inexorably’ at exactly the same rate in all mutually inertial frames of reference. Astonishingly enough, Maxwell’s equations did not share the property of form-invariance—they did not remain unchanged in form under Galilean transformations! This, in turn, implied that the ether had a special feature that could be established experimentally: Maxwell proposed electromagnetic experiments which could detect the motion of the earth through the ether, the medium that he believed to be at, and in fact defined, absolute rest. (Remember that, in Galilean–Newtonian mechanics, there is no such thing as absolute rest.) However, the famous and crucial experiments of Michelson and Morley, done in 1887, failed to detect the expected effects, and this led to a crisis. Many great figures in physics and mathematics, including Hendrik Antoon Lorentz (1853–1928; Nobel Prize, 1902) and Henri Poincaré (1854–1912) worked on attempts to settle this serious issue.

The final resolution of the problem came in 1905, with Einstein’s Special Theory of Relativity. It turned out that Newton’s mechanics had to be modified to fall in line with Maxwell’s electromagnetism! It was the latter that led to a new and more correct view of space and time, and mechanics had to be made consistent with this view. It became clear that the equations of both mechanics and electromagnetism would have to be form-invariant under a common set of transformations of the space and time coordinates, comprising the so-called Poincaré group (also known as the inhomogeneous Lorentz group). Newton’s separate and individually absolute space and time were replaced by a new unified space-time which alone was absolute, the same for all observers. But each observer separates space-time into her own space and her own time in her own way. The earlier Galilean transformations of space and time gave way to new transformations named after Lorentz, and these form the basis of special relativity. Newton had posited that time was universal, the same for everyone, flowing inexorably from the past into the future. If two events occurred simultaneously, all observers would agree on it. But now, with the special theory of relativity, that was shown to be untrue. Simultaneity is not absolute, but depends on the frame of reference. Two events that occur simultaneously according to one observer need not do so for another observer in a different frame of reference. According to the Lorentz transformations, not only the coordinates of points in space, but also the times of events change in a specific manner from one reference frame to another. As a consequence, the distances between objects or events, as well as the time intervals between events, also vary from one frame of reference to another. Even though space and time remain profoundly different in their physical characteristics, the new rules of transformation unite them much more closely than is the case in the Galilean–Newtonian worldview.

Even in special relativity, though, combined space-time remains absolute. Its quantitative characterization by different observers (using metre rods, clocks and their equivalents) is encoded in the Lorentz transformation rules. It serves as the stage for all physical phenomena, but is unaffected by them.

---

10 Albert Abraham Michelson (1852–1931; Nobel Prize, 1907) and Edward Morley (1838–1923).
6. General Relativity: Dynamical Space-Time

The essentially final step (within the framework of classical physics) in the development of our understanding of space and time occurred with Einstein’s General Theory of Relativity, which he completed in 1915. This is a theory of superlative beauty. In it, finally, space and time no longer remain a stage on which physical processes take place, but themselves become actors or active participants in the proceedings. The spirit behind this idea is beautifully expressed by Einstein:

“It is contrary to the mode of thinking in science to conceive of a thing ... which acts itself, but which cannot be acted upon.”

The very geometry of space-time, which Einstein identified with gravitation, now becomes changeable and dynamical. Geometry acts upon matter and electromagnetism, and in turn is acted upon by them. In the expressive words of Wheeler: “Space-time tells matter how to move, matter tells space-time how to curve.” In general relativity, Einstein used extensively the mathematics initiated by Riemann and developed further by Christoffel, Ricci, Levi-Civita and others. But the fundamental physical sense in which he went beyond Riemann was that, from the geometrical point of view, time was treated on the same footing as space, in spite of the profound differences between them. This was something Riemann had not foreseen. It became possible only because special relativity had already been developed (by Einstein himself). A more detailed discussion of general relativity requires mathematics, and we do not go into this here. However, we describe some of the relevant salient features of general relativity in Box 4.

The well-known physicist Chen-Ning Yang (1922–; Nobel Prize, 1957) was once asked by Dirac what he (Yang) regarded as Einstein’s greatest discovery. When Yang chose general relativity, Dirac disagreed and opined that it had to be special relativity, because it showed for the first time that space and time had to be treated and transformed together.

General relativity provided for the first time a dependable language for discussing problems of cosmology – the behaviour of the universe as a whole. It also led to the study of exotic objects like black holes. On the practical side, the Global Positioning System in current use takes into account the effects of both special and general relativity. If this is not done, the GPS would go haywire within about an hour. In the last few years, a spectacular confirmation of general relativistic effects has been found in a double-pulsar system about 2000 light years away from us. The pulsars, separated by a distance of a million kilometres, orbit around each other with a time period of 2.4 hours. This time period is affected by the strong gravitational fields present (about 100,000 times stronger than that due to the sun). A comparison of the observed variations in the interval between the pulses arriving here from the system with the...
There were three unique features about general relativity that Einstein found appealing. They also represented part of the reason for his lifelong opposition to quantum mechanics as an approach to understanding nature. He believed that quantum mechanics was indeed successful in describing nature within its domain of applicability, and that a future unified field theory would have to reproduce the results of quantum mechanics, perhaps as a linear approximation to a deeper nonlinear theory. If true, this would be similar to how the relativistic gravitational field of general relativity (with a finite propagation speed of the gravitational force) led to Newton's law of gravitation (with its action-at-a-distance force) in the non-relativistic limit. But Einstein was convinced that quantum mechanics was not the correct approach to deducing the fundamental laws of physics. The three important features of general relativity were as follows.

- With the equations of general relativity, Einstein found that space-time was no longer just a passive stage on which particles and fields performed their acts, but was an active participant in the performance. Thus, the geometric structure of space-time was determined by the matter in it, and of course the matter responded to this geometry and was constrained by its structure.
- General relativity was the first theory in physics that was inherently nonlinear. An important consequence of this was that the equation of motion of an infinitesimal test particle in the gravitational field of a massive object was contained in the field equations themselves. This equation specified the geodesic path followed by the particle in the curved space-time in which it moved. In other words, one did not have separate equations for the gravitational interactions between the matter present, and for the response of the matter to these interactions. Everything was contained in one set of field equations. In contrast, a linear theory like Maxwell's equations for the electromagnetic field could only describe the (electromagnetic) field manifestation of (charged) matter, while the response of the matter (or inertial manifestation) was contained in separate 'equations of motion'. Indeed, this is true of any linear theory – hence Einstein's opposition to regarding the linear formalism of quantum mechanics as the final or underlying theory.
- For the first time in physics, a theory predicted that the inertia of a system depended on its surroundings. In keeping with Mach's idea that inertia is a consequence of a body's interactions with the rest of the universe, the equations of general relativity showed that the inertia of a system increases when it is placed in the vicinity of other ponderable masses. Inertia was no longer some inherent, given property of a system, but was at least partly determined by the environment. Einstein's unfulfilled dream was to find a fully unified field theory which would show that all of the inertia (and not just part of it) was due to the interactions with the environment.

calculated values shows that the predictions of general relativity are borne out to an astonishing accuracy of 99.95%.

Soon after Einstein presented general relativity, Hermann Weyl\textsuperscript{14} in 1917–18 suggested an extension of Riemannian geometry for space-time, involving what he called the gauge principle, which he thought would unify gravitation and Maxwell's classical electromagnetism. A few years later, Theodore Kaluza (1885–1954) in 1921 and Oskar Klein (1894–1977) in 1926 independently published their ideas on another way to unify the two forces, based on

\textsuperscript{14}Hermann Weyl (1885–1955) was one of the great mathematicians – today, we would also rank him as a preeminent mathematical physicist - of the 20th century. His book \textit{Raum-Zeit-Materie (Space-Time-Matter)}, first published in 1918, was highly influential for several decades in spreading the ideas of relativity among physicists and mathematicians. Weyl also played a prominent role in applying the ideas of symmetry and group theory to physics, including quantum mechanics.
increasing the number of dimensions of space-time from four to five – namely, four spatial dimensions and one time dimension. Klein, in fact, pioneered the idea that the extra space dimension was actually ‘curled up’ into a tiny circle of radius of the order of the Planck length, \( l_P = \sqrt{\frac{Gh}{c^3}} \sim 10^{-35} \) metres. He also advanced the idea that the ‘quantization’ of electric charge could be related to a topological feature such as the curling up of a dimension into a circle. These ideas were extremely original and beautiful, but did not succeed or lead to any progress at that time. Many decades later, however, they have become important, in a modified and extended form, in the context of string theory.

### 7. Space and Time in Quantum Physics

All that we have discussed so far has been in the realm of classical physics. We must now give at least a brief account of the surprising results that emerge in quantum theory vis-à-vis space and time.

The initial form of quantum mechanics, completed during the period 1925–27, assumed that space and time were Newtonian, i.e., non-relativistic, as we would say today. This suffices for most of chemistry, atomic, molecular and even nuclear physics. Soon after, in 1928, Dirac found a description of the electron which combined the (then) new quantum mechanics with special relativity and was spectacularly successful. From the 1930’s to the 1970’s and beyond, this led to the impressive development of quantum field theory in which, too, space-time is treated according to special relativity–so that once again it (space-time) remains a stage for phenomena. The theories of the strong, electromagnetic and weak interactions are all of this form. Along the way, in 1956, Tsung-Dao Lee (1926–; Nobel Prize, 1957) and Yang found that in the weak processes, such as those responsible for radioactive decays, space reflection or parity is not a valid symmetry. That is, nature does distinguish between left-handedness and right-handedness at a fundamental level, contrary to what one might expect ‘intuitively’.

Analogous to space reflection, we have time reflection or time reversal. The situation here is quite subtle. The inexorable march of time gives us the notion of an *arrow of time* that always points to the future. A drop of ink that falls into a glass of water disperses spontaneously in the water, till it becomes virtually invisible. It does not re-assemble into the original drop, no matter how long we wait. All of us age and grow older; we never evolve ‘backwards’ from old age to infancy. And yet, the laws governing almost all kinds of time evolution, both classical and quantum mechanical, are – *at the fundamental level* – almost always time-reversal invariant. That is, they remain unchanged when the sign of the time variable is changed. The deep question, then, is how this *microscopic reversibility* in time leads, nevertheless, to *macroscopic irreversibility*. The answer is a subject in itself, and we shall not go into it here. It is related, as you might guess, to the statistical nature of the second law of thermodynamics and the mechanism by which entropy increases spontaneously. It is also related to dynamical

---

15 With the important proviso that the effects of spin are to be incorporated. The origin of spin is often believed to lie in relativistic quantum physics, but is actually a logical and rigorous consequence of non-relativistic quantum mechanics.
Vasant Natarajan, V Balakrishnan and N Mukunda

chaos (or exponential sensitivity to initial conditions), and the impossibility of specifying these conditions with infinite precision. We must also mention that a complete answer that is fully within the purview of quantum mechanics is not yet known, and that the matter is the subject of on-going research.

But there is yet another twist to the tale. There is now indirect evidence that there is a certain violation of time-reversal symmetry in the fundamental physical laws themselves. This is based on the discovery in 1964 of what is called CP-violation in certain elementary particle decays, for which James W Cronin (1931–) and Val L Fitch (1923–) were awarded the Nobel Prize in 1980. A consequence of the violation of time-reversal invariance is that a particle such as an electron, or a neutron, or a free atom will have an intrinsic electric dipole moment.16 The search for such an electric dipole moment is an important endeavour in experimental physics, as it amounts to a test of the validity of a fundamental principle of nature. Several laboratories around the world are currently engaged in this pursuit (e.g. at Harvard, Yale, Colorado, Washington, and Princeton in the US, and Imperial College and KVI Groningen in Europe), including that of one of the authors (VN) at the Indian Institute of Science. The violations of parity invariance and time-reversal invariance mentioned above suggest that nature has some asymmetry built into it: elementary particle interactions do not exhibit the degree of symmetry with regard to space-time itself that one might expect a priori. These asymmetries, in turn, have implications in cosmology and the manner in which the universe has evolved from the Big Bang. As you can see, all these aspects go to show how deeply intertwined are the apparently diverse parts of the subject of physics–clearly, nature is above these artificial distinctions.

8. What Does the Future Hold?

Finally, we come to the problem of combining general relativity and quantum mechanics. This problem has been attacked since the early 1930’s, but it has proved to be uncommonly difficult. In a very real sense, it may be regarded as an ‘ultimate problem’ of sorts, at least as far as the physics built up over the last four centuries or so is concerned. As the jury is still out on this question what we go on to say now must be regarded as tentative and speculative.

In recent decades, some promising lines of progress may have appeared in the attempts to tackle the problem. They involve quite novel ideas and, along with them, complicated mathematics to implement these ideas in a consistent manner. In essence, it appears that the ingredients or features necessary for a consistent and satisfactory amalgamation of gravity and quantum mechanics are likely to be, at the very least, the following: (i) the replacement of zero-dimensional point-like particles as the ultimate fundamental objects in the universe by extended objects, specifically, one-dimensional strings; (ii) a space-time with a dimensionality higher than four–perhaps a space-time with nine spatial dimensions and one time dimension,

---

16 As you may be aware, protons and neutrons (via the quarks of which they are composed), as well as electrons, have intrinsic magnetic dipole moments related to the spins of these particles. This does not violate time-reversal invariance. We are now speaking of intrinsic electric dipole moments, whose existence would indicate such a violation.
six of the spatial dimensions being curled up (or ‘compactified’) to a size within the Planck length, but in a far more (topologically) intricate and complicated manner than Klein anticipated; and (iii) an underlying symmetry between bosons and fermions called *supersymmetry* that is spontaneously broken in the present-day universe. Taken together, they constitute the basic features of the so-called superstring theory.

What is slowly emerging is even more interesting, and cuts even deeper. Over and above the features listed above, there appear to be theoretical indications that the resolution of the problem will entail even more drastic revisions of our understanding of both quantum mechanics and space-time itself. For instance: this universe may be only one of a vast (or even infinite) array of ever-branching-out universes in a ‘multiverse’; the dimensionality of space-time may emerge as a dynamical property of this universe; there may be an ultimate underlying granularity or discreteness in the very notions of length and time durations; and space-time coordinates (even including the possible higher-dimensional ones) may have to be augmented by other, ‘internal’ variables of a different kind in order to provide a complete description of nature at the most fundamental level.

These are but theoretical and mathematical speculations, as yet. Time will tell (!) whether they are correct, and to what extent. While there has been a great deal of progress on the theoretical side over the past twenty-five years or so, experimental advances in elementary particle physics have lagged behind. This phase lag is rather an anomaly, and the experience of the past four hundred years suggests strongly that true progress will most likely be made only when the gap between experiment and theory is reduced considerably. This is one of the leading reasons why physicists are eagerly awaiting the first results from the Large Hadron Collider (LHC), which is expected to go on stream at CERN, Geneva, in the very near future.

But it is good to conclude this account here, seeing how far we have come from the Egyptian farmers toiling on the banks of the Nile, from whose mud the subject of geometry was born – aided by the upward look into the sky, and the challenge posed by the splendour of the cosmos stretching out before the eyes of humankind.
Symmetries and Conservation Laws in Classical and Quantum Mechanics: Classical Mechanics

K S Mallesh, S Chaturvedi, V Balakrishnan, R Simon and N Mukunda

We describe the connection between continuous symmetries and conservation laws in classical mechanics. This is done at successively more sophisticated levels, bringing out important features at each level: the Newtonian\(^1\), the Euler–Lagrange\(^2\), and the Hamiltonian phase-space forms of mechanics. The role of the Action Principle is emphasised, and many examples are given.

1. Introduction

It is generally well known that the description and consequences of symmetry are important and beautiful components of both classical mechanics (CM) and quantum mechanics (QM). The connection between the ten basic ‘Galilean’ conservation laws in Newtonian mechanics and fundamental space-time symmetries was first shown by G Hamel in 1904. A few years later, in 1918, Emmy Noether\(^4\) brought out the connection between the invariances of variational principles under groups of continuous transformations and conservation theorems. Depending on the age of the reader, these may seem to be fairly recent advances, or else ancient knowledge.

In this two-part article, we shall review these topics using the notations appropriate for systems with a finite number of degrees of freedom; the generalization to (classical) field theory is quite elementary, in principle. In Part 1, we consider CM in successively more sophisticated versions, and explore in each the definition, description and consequences of symmetry. In Part 2, we shall consider the transition to QM. We emphasize the general theory, and look at both similarities and differences between classical and quantum mechanics as far as symmetry is concerned.

2. Symmetry in Classical Mechanics

In recalling familiar material, we make a step-by-step progression from elementary considerations to a comprehensive formalism. Boldface symbols will denote vectors, as usual.


\(^1\)See Isaac Newton, Resonance, Vol. 11, No. 12, 2006.
\(^3\)See Joseph Louise Lagrange, Resonance, Vol. 11, No. 4, 2006.
Newton’s First Law of Motion states that an isolated material body (of sufficiently small size) maintains a state of rest or of uniform motion in a straight line. Here, the concept of inertial frames of reference, the validity of Euclidean geometry for physical three-dimensional space, and the uniform flow of absolute time are all assumed. We can see that in the absence of external forces, space is homogeneous, and the momentum of the body is a constant of the motion (COM).

Next, moving on to a system of two bodies or ‘particles’, we have Newton’s Second Law, the equation of motion (EOM), together with his Third Law relating the forces the particles exert on each other:

\[ p_1 = m_1 \dot{r}_1, \quad p_2 = m_2 \dot{r}_2. \]
\[ \dot{p}_1 = F_{12}, \quad \dot{p}_2 = F_{21}. \]
\[ F_{12} = -F_{21} \Rightarrow p_1 + p_2 \text{ is conserved (it is a COM).} \quad (1) \]

This is the first nontrivial instance of a conservation theorem, in which the crucial role of the Third Law (entirely Newton’s contribution) should be appreciated.

If, in (ii) above, the two forces arise from an inter-particle potential, we have

\[ F_{12} = -\nabla_1 V(r_1 - r_2), \quad F_{21} = -\nabla_2 V(r_1 - r_2). \quad (2) \]

The Third Law then holds because of the translation invariance of the potential. We see that momentum conservation is connected to a symmetry, namely, translation invariance. The proper understanding and appreciation of angular momentum came somewhat later, in Euler’s work. (Recall, though, that Kepler’s\(^5\) Second Law already amounts to the conservation of angular momentum.)

Let us consider, next, the Lagrangian form of classical mechanics. We have a system with \( n \) degrees of freedom described by generalized coordinates \( q_r (r = 1, 2, \ldots, n) \), and are given a Lagrangian \( L(q, \dot{q}) \). Here, and in all that follows, we use \( q \) and \( \dot{q} \) as short-hand for the sets of variables \( \{q_r\} \) and \( \{\dot{q}_r\} \), where \( r \) runs from 1 to \( n \). Explicit time dependence in \( L \) can be included. We further assume that \( L \) is a nonsingular Lagrangian, so that no constraints arise in passing to the Hamiltonian formulation. (This means that we can invert the defining equations \( p_r = \partial L/\partial \dot{q}_r \) to find the \( \dot{q} \)'s as functions of the \( q \)'s and \( p \)'s, and thus eliminate them to write the Hamiltonian as a function of the \( q \)'s and \( p \)'s. In technical terms, this requires that the \((n \times n)\) matrix whose \((ij)\)th element is \( \partial^2 L/\partial \dot{q}_i \partial q_j \) be a non-singular matrix.) The Euler–Lagrange EOM are equivalent to the simplest form of the Action Principle:

\[ \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_r} \right) - \frac{\partial L}{\partial q_r} = 0 \quad (r = 1, 2, \ldots, n) \]

Symmetries and Conservation Laws in Classical and Quantum Mechanics:

\[
\delta \int_{t_1}^{t_2} L(q, \dot{q}) dt = 0, \quad \delta q_r(t_1) = \delta q_r(t_2) = 0, \\
(r = 1, 2, \cdots, n). \tag{3}
\]

In Figure 1 we depict the kind of variations in configuration space trajectories involved in this form of the Action Principle. In this framework, a COM is any function \( f(q, \dot{q}, t) \) whose total time derivative vanishes identically:

\[
\text{EOM} \Rightarrow \frac{d}{dt} f(q, \dot{q}, t) = 0. \tag{4}
\]

Thus, along any solution of the EOM – a physical trajectory in configuration space – the COM has an unchanging numerical value (see Figure 2). There is as yet no link to any symmetry, but we move in that direction via two simple examples. This will set the stage for a general formulation.

(v) Suppose the Lagrangian \( L(q, \dot{q}) \) has no dependence on a particular generalized coordinate \( q_k \) (but does, of course, depend on \( \dot{q}_k \)), for some \( k \). This coordinate is then said to be ‘cyclic’. The corresponding EOM reads:

\[
\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) = 0 \quad \Leftrightarrow \quad p_k = \frac{\partial L}{\partial \dot{q}_k} \quad \text{is a COM.} \tag{5}
\]
We can express this as an invariance property of the Lagrangian. The cyclic nature of $q_k$ means (trivially) that $L(q, \dot{q})$ is invariant under (infinitesimal) translations or shifts in $q_k$ which are time-independent. We can therefore express (5) as follows:

$$\text{If } \delta q_r = \varepsilon \delta_{rk} \Rightarrow \delta L = 0, \text{ then } p_k = \text{COM.}$$  

(6)

The Kronecker delta ensures that only the $k$th generalized coordinate is shifted. $\varepsilon$ is a small (infinitesimal) time-independent parameter, so that $\delta \dot{q}_r = 0$.

(vi) We have said that an explicit time dependence in the Lagrangian can always be permitted. In case there is none, we can say ‘time is cyclic’, even though time is not treated as a generalized coordinate. The Lagrangian then enjoys invariance under translation in time, and this leads to a COM, which is just the Hamiltonian. Using the summation convention for repeated indices (here and in all that follows), we have

$$\frac{\partial L}{\partial t} = 0 \Rightarrow \frac{d}{dt} L(q, \dot{q}) = \frac{\partial L}{\partial q_r} \dot{q}_r + \frac{\partial L}{\partial \dot{q}_r} \ddot{q}_r$$

$$= \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_r} \right) \dot{q}_r + \frac{\partial L}{\partial \dot{q}_r} \ddot{q}_r \quad (\text{using the EOM})$$

$$= \frac{d}{dt} (p_r \dot{q}_r) \quad (\text{recall that } p_r \equiv \partial L/\partial \dot{q}_r).$$

(7)

Hence

$$\frac{\partial L}{\partial t} = 0 \text{ together with the EOM } \Rightarrow p_r \dot{q}_r - L(q, \dot{q}) \text{ is a COM.}$$

(8)

As you know, this is the Hamiltonian of the system. Its numerical value is just the total energy of the system, in the familiar situations.
We now move towards a general treatment of the connection between infinitesimal (⇒ continuous) symmetries and conservation laws in CM, covering it in two steps. An infinitesimal point transformation in Lagrangian mechanics is a small change made in each generalized coordinate at each time, of the form

$$\delta q_r = \varepsilon \phi_r(q), \quad r = 1, 2, \ldots, n.$$ (9)

Here $\varepsilon$ is an infinitesimal parameter, and in all calculations we need only to keep terms up to first order in $\varepsilon$. The functions $\phi_r$ depend on the $q$'s alone (hence the name ‘point transformation’), and they specify the transformation: we view it as a geometrical point-to-point mapping in configuration space. By the rules of Lagrangian mechanics, as in the working out of the Action Principle, we have

$$\delta \dot{q}_r = \frac{d(\delta q_r)}{dt} = \varepsilon \frac{d\phi_r(q)}{dq_r} = \varepsilon \frac{\partial \phi_r(q)}{\partial q_l} \dot{q}_l.$$ (10)

Hence these increments depend on the $q$'s, and linearly on the $\dot{q}$'s. If the Lagrangian happens to be unchanged under this transformation, then, combining this property with the EOM leads to a COM that is linear in the canonical momenta, as follows:

$$\delta L = \frac{\partial L}{\partial q_r} \delta q_r + \frac{\partial L}{\partial \dot{q}_r} \delta \dot{q}_r = \frac{\partial L}{\partial q_r} \delta q_r + p_r \frac{d(\delta q_r)}{dt} = 0$$

$$\Rightarrow \frac{dp_r}{dt} \delta q_r + p_r \frac{d(\delta q_r)}{dt} = 0 \quad \text{using the EOM}$$

$$\Rightarrow G(q, p) = p_r \phi_r(q) = \text{COM}.$$ (11)

We have introduced the symbol $G$ for a generic COM, regarded as a function in phase space. Thus,

- the invariance of the Lagrangian under an infinitesimal point transformation leads, when combined with the EOM, to a COM linear in the momenta.

This conclusion can also be arrived at in an elegant manner from the extended form of the Action Principle, in which we allow variations that do not vanish at the initial and final times. For brevity, let us write (in honour of Euler and Lagrange!)

$$(\text{EL})_r \equiv \frac{\partial L}{\partial q_r} - \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_r} \right).$$ (12)

We develop the argument starting with the change in the action under general infinitesimal changes in the $q$'s at each instant of time as well as in the initial and final times $t_1$ and $t_2$. 

$$\delta \int_{t_1}^{t_2} L(q, \dot{q}) dt = [L\delta t]_{t_1}^{t_2} + \int_{t_1}^{t_2} \delta L(q, \dot{q}) dt$$
for any $\delta t_1$, $\delta t_2$ and $\delta q_r(t)$. In the first term on the right-hand side, we may express $\delta q_r(t_1)$ and $\delta q_r(t_2)$ in terms of $\Delta q_r(t_1)$ and $\Delta q_r(t_2)$, the total variations at the end points, using

$$\Delta q_r(t_i) = \delta q_r(t_i) + \dot{q}_r(t_i)\delta t_i$$

or

$$\delta q_r(t_i) = \Delta q_r(t_i) - \dot{q}_r(t_i)\delta t_i,$$

where $i = 1, 2$. Further, putting $H = p_r \dot{q}_r - L$, we get

$$\delta \int_{t_1}^{t_2} L(q, \dot{q})dt = \left[p_r \Delta q_r - H\delta t\right]_{t_1}^{t_2} + \int_{t_1}^{t_2} (\text{EL})_r \delta q_r(t) dt,$$

for any $\delta t_1$, $\delta t_2$ and $\delta q_r(t)$. Now using the EOM (i.e., $(\text{EL})_r = 0$ for each $r$), this reduces to

$$\delta \int_{t_1}^{t_2} L(q, \dot{q})dt = \left[p_r \Delta q_r - H\delta t\right]_{t_1}^{t_2},$$

for any $\delta t_1$, $\delta t_2$ and $\delta q_r(t)$.

Finally, using $\delta q = \epsilon \phi(q)$, $\delta L = 0$ and setting $\delta t_1 = \delta t_2 = 0$, we have

$$0 = \left[\epsilon \ p_r \phi_r(q)\right]_{t_1}^{t_2} \Rightarrow G(q, p) = p_r \phi_r(q) = \text{COM}.$$

We can view this as a derivation of the result (11) from a global rather than local point of view in time. For this limited purpose, it is adequate to use the extended form of the Action Principle but with $\delta t_1 = \delta t_2 = 0$. The meaning of the two kinds of variations $\delta q(t)$ and $\Delta q(t)$ that appear in the discussion above is illustrated schematically in Figure 3.

(viii) When we translate the result above into the Hamiltonian or phase space language, new and crucial aspects enter the picture. Let us recall some important quantities and expressions in phase space mechanics. Given two functions $f(q, p)$ and $g(q, p)$, their Poisson Bracket (PB) is a third function given by

$$\{f, g\}(q, p) = \frac{\partial f}{\partial q_r} \frac{\partial g}{\partial p_r} - \frac{\partial f}{\partial p_r} \frac{\partial g}{\partial q_r}. $$

The canonical PBs satisfied by the $q$'s and $p$'s are of course

$$\{q_r, q_s\} = 0, \ \{p_r, p_s\} = 0, \ \{q_r, p_s\} = \delta_{rs}. $$
Symmetries and Conservation Laws in Classical and Quantum Mechanics:

PBs satisfy the familiar properties of antisymmetry, bilinearity and the Jacobi identity. Next, a canonical transformation (CT) is a phase space mapping

\[ q, \ p \rightarrow Q(q, p), \ P(q, p) \]  \hspace{1cm} (20)

such that all PB relations are maintained. Lastly, an infinitesimal CT is one in which \( Q, \ P \) differ infinitesimally from \( q, \ p \):

\[ Q_r = q_r + \delta q_r, \quad P_r = p_r + \delta p_r, \]  \hspace{1cm} (21)

where \( \delta q_r \) and \( \delta p_r \) are functions of \( q, \ p \). The assertion is that such a CT is always of the form

\[ \delta q_r = \varepsilon [q_r, G(q, p)], \quad \delta p_r = \varepsilon [p_r, G(q, p)], \]  \hspace{1cm} (22)

where \( \varepsilon \) is a small parameter, and \( G(q, p) \) (determined up to an additive constant) is the \textit{generator} of the transformation. This statement is valid whether or not the infinitesimal CT arises from a symmetry. We now show that an infinitesimal point transformation \textit{symmetry}, when regarded as a phase space transformation in the natural manner, turns out to be an infinitesimal CT with the COM as its generator, in the sense of equation (22).

Consider a point transformation symmetry of the Lagrangian leading to the COM \( G(q, p) \) in (11). Since this is linear in the canonical momenta, the canonical PBs (19) immediately yield (recall (9))

\[ \delta q_r = \varepsilon \phi_r(q) = \varepsilon [q_r, G(q, p)]. \]  \hspace{1cm} (23)

Note that this is a kinematic fact, independent of the EOM. We thus have the ‘first half’ of an infinitesimal CT in phase space, the first of equations (22). This result can now be
extended to the momenta, too, by defining and calculating their changes as follows:

$$
\delta p_r = \delta \left( \frac{\partial L(q, \dot{q})}{\partial q_r} \right) = \frac{\partial^2 L}{\partial q_r \partial q_s} \delta q_s + \frac{\partial^2 L}{\partial q_r \partial \dot{q}_s} \delta \dot{q}_s
$$

$$
= \varepsilon \frac{\partial^2 L}{\partial q_r \partial q_s} \phi_s(q) + \varepsilon \frac{\partial^2 L}{\partial q_r \partial \dot{q}_s} \frac{\partial \phi_s(q)}{\partial \dot{q}_l} \dot{q}_l,
$$

(24)

using (9) and (10) for $\delta q_s$ and $\delta \dot{q}_s$. The expression on the right-hand side involves only the $q$’s and $\dot{q}$’s (and no higher derivatives), and hence is kinematically expressible in terms of the $q$’s and $p$’s. Further, since the point transformation is taken to be a symmetry of the Lagrangian, we have:

$$
\delta L = \frac{\partial L}{\partial q_s} \delta q_s + \frac{\partial L}{\partial \dot{q}_s} \delta \dot{q}_s = 0
$$

$$
\Rightarrow \frac{\partial L}{\partial q_s} \phi_s(q) + \frac{\partial L}{\partial \dot{q}_s} \frac{\partial \phi_s(q)}{\partial \dot{q}_l} \dot{q}_l = 0.
$$

(25)

Differentiating this equation with respect to $\dot{q}_r$ and using the result in (24), we get (after a little bit of algebra, which you will find instructive to work out)

$$
\delta p_r = -\varepsilon p_r \frac{\partial \phi_s}{\partial q_r} = \varepsilon \{p_r, G(q)\}.
$$

(26)

This is the ‘second half’ of an infinitesimal CT, the second of equations (22). Thus, we have the important result:

- An infinitesimal point transformation symmetry of the Lagrangian appears in phase space as an infinitesimal CT, with the associated COM as its generator. This is a kinematical fact that does not require the use of the EOM.

- The EOM are needed only to show that the generator of the CT is a COM.

(ix) We come, now, to the last natural generalization of the symmetry↔COM connection in the Lagrangian formalism. We extend (9) and consider an infinitesimal change in the $q$’s of the form

$$
\delta q_r = \varepsilon \phi_r(q, \dot{q}),
$$

(27)

permitting a dependence on the generalized velocities $\dot{q}$ as well. We then define the changes in the generalized velocities $\{\dot{q}_r\}$, in the Lagrangian spirit, as

$$
\delta \dot{q}_r = \frac{d}{dt} \delta q_r = \varepsilon \frac{d}{dt} \phi_r(q, \dot{q}).
$$

(28)

Hence $\delta \dot{q}_r$ may now involve the accelerations $\{\ddot{q}_l\}$ as well, in general. If, under these changes, the Lagrangian changes by a total time derivative,

$$
\delta L(q, \dot{q}) = \varepsilon \frac{d}{dt} F(q, \dot{q}),
$$

(29)
where \( F(q, \dot{q}) \) is a function of the indicated variables that is ‘local in time’, then we have a dynamical symmetry of the Lagrangian. The following three consequences can be established quite easily, either by using the differential EOM directly, or (more elegantly) via the extended Action Principle:

\[
\begin{align*}
(a) \quad G(q, p) &= p_r \phi_r(q, \dot{q}) - F(q, \dot{q}) = \text{COM}, \\
& \quad \text{not necessarily linear in } p, \quad (30a) \\
(b) \quad \delta q_r &= \varepsilon \{q_r, G(q, p)\}, \quad \text{without using the EOM} \quad (30b) \\
(c) \quad \delta p_r &= \varepsilon \{p_r, G(q, p)\}, \quad \text{using the EOM, in general.} \quad (30c)
\end{align*}
\]

Thus:

- The most general infinitesimal symmetry of a Lagrangian is a dynamical symmetry, characterized by equations (27)–(29).
- It leads to a COM, given by the function \( G \) of (30a).
- When recast in phase space language, the symmetry appears as an infinitesimal CT, with the COM as the generator.

In all the quantities above, namely, \( \phi_r, L, F \) and \( G \), we can permit explicit time dependence without any difficulty.

We invite the reader to convince herself that the argument leading from an infinitesimal symmetry to a COM can now be expressed in a very economical or compact form, as follows. Given a Lagrangian \( L(q, \dot{q}, t) \), the following implications hold good:

(a) The Euler–Lagrange EOM are obeyed \( \Leftrightarrow \) the variation in the Lagrangian, for arbitrary infinitesimal variations \( \delta q_r(t) \) at each instant \( t \), is of the form

\[
\delta L = \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_r} \delta q_r \right). \tag{31}\n\]

(b) An infinitesimal transformation \( \delta q_r = \varepsilon \phi_r(q, \dot{q}, t) \) of specified functional form is a (dynamical) symmetry transformation \( \Leftrightarrow \) there exists a function \( F(q, \dot{q}, t) \) such that

\[
\delta L = \varepsilon \frac{d}{dt} F(q, \dot{q}, t). \tag{32}\n\]

(c) Using the \( \delta q_r \) of a symmetry transformation in statement (a), we conclude that

\[
\text{EOM plus infinitesimal symmetry } \Rightarrow G = p_r \phi_r - F
\]

is a COM. \( \tag{33} \)

The simplicity and naturalness of the argument should now be evident.
Just as the evolution in time of a classical Hamiltonian system can be pictured as the continuous unfolding of a family of CTs, so also every symmetry in the sense described above is a continuous family of CTs. Moreover, every such symmetry preserves the EOM, as is most easily seen from the Action Principle. Therefore it acts as a mapping of the set of all solutions of the EOM onto itself.

We remark that the case of no explicit time-dependence in the Lagrangian, considered in (vi) above, is itself an example of dynamical symmetry. In that case equations (27)–(29) become, respectively,

\[ \delta q_r = \epsilon \dot{q}_r, \quad \delta \dot{q}_r = \epsilon \ddot{q}_r, \]

so that

\[ \delta L = \epsilon \frac{dL}{dt}. \]  (34)

Then (30a) directly leads to (8), namely, the conservation of the quantity \( p_r \dot{q}_r - L(q, \dot{q}) \).

It is important to appreciate the depth of these results. We have defined an infinitesimal symmetry of a Lagrangian as a transformation (27) that changes the Lagrangian at most by a total time derivative (or preserves the action functional up to end-point terms). Such a transformation will preserve the EOM and will lead to a COM. But then the COM \( G(q, p) \) plays a new role: it generates the symmetry transformation in phase space as a CT. Thus, there are two roles for each COM:

(a) \( G(q, p) \) is constant, in numerical value, along each solution of the EOM.

(b) \( G(q, p) \) generates a symmetry as an infinitesimal CT, mapping each solution of EOM onto another, generally different, solution.

In the second role above, it is the algebraic form rather than the numerical value that is important. By EOM, we mean here the complete Hamiltonian system of equations

\[ \dot{q}_r = \{q_r, H(q, p)\} = \frac{\partial H}{\partial p_r}, \]
\[ \dot{p}_r = \{p_r, H(q, p)\} = -\frac{\partial H}{\partial q_r}. \]  (35)

Figure 4. Roles of COM in Hamiltonian mechanics.
Symmetries and Conservation Laws in Classical and Quantum Mechanics:...

In the spirit of Figure 2, we can depict the situation as in Figure 4, except that the trajectories shown now are phase space trajectories. We need only add that in the case of a single transformation involving just one COM \(G(q, p)\), the numerical values of \(G(q, p)\) are the same on both the original and the transformed solutions of the EOM.

Table 1. Consequences and properties of infinitesimal symmetry transformations.

<table>
<thead>
<tr>
<th>Form of dynamics</th>
<th>Maps solutions of EOM to solutions</th>
<th>Leads to COM</th>
<th>Symmetry is a CT</th>
<th>COM generates Symmetry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newtonian</td>
<td>√</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Lagrangian</td>
<td>√</td>
<td>√</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Hamiltonian</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

At this point, it is helpful to summarise what we have learnt about symmetry in classical mechanics in the form of two tables. In Table 1, a cross indicates ‘not necessarily’ or ‘not meaningful’. Clearly, it is only in the Hamiltonian formalism that we have a complete and well-rounded picture. Table 2 describes the transformations, their COMs, etc., in more detail. Of the two rows in this table, the first is a simple special case, while the second is the most general case in the framework of classical mechanics.

Table 2. Structure of symmetry transformations and their COMs.

<table>
<thead>
<tr>
<th>Type of Symmetry</th>
<th>Behaviour of Lagrangian</th>
<th>COM</th>
<th>Symmetry as CT</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\delta q_r = \varepsilon \phi_r(q))</td>
<td>(\delta L = 0)</td>
<td>(G(q, p) = p_r \phi_r(q))</td>
<td>Kinematic; Geometric; point transformation; COM linear in (p)</td>
<td></td>
</tr>
<tr>
<td>(\delta q_r = \varepsilon \phi_r(q, \dot{q}, t))</td>
<td>(\delta L = \varepsilon dF(q, \dot{q}, t)/dt)</td>
<td>(G(q, p) = p_r \phi_r(q, \dot{q}, t))</td>
<td>dynamic; may need symmetry; (G) nonlinear in (p)</td>
<td></td>
</tr>
</tbody>
</table>

157
3. Examples

We turn now to some examples of systems in which the generalized coordinates undergo infinitesimal variations under symmetry transformations. These variations can be of different types, as mentioned in Section 2. The aim is to identify the COMs in each case.

3.1 System of Particles Moving in a Potential due to Newtonian Forces

We begin with a system of \( n \) point masses moving in a potential that depends only on the inter-particle distances – for instance, the case of particles interacting among themselves via two-body conservative Newtonian forces. The Lagrangian for such a system is

\[
L = \frac{1}{2} \sum_{j=1}^{n} m_j \dot{r}_j^2 - V(|r_{jk}|), \quad r_{jk} = |r_j - r_k| \\
(j, k = 1, 2, \ldots, n). \tag{36}
\]

(a) Consider the variation \( \delta r_j = \epsilon a \), where \( \epsilon \) is infinitesimal in magnitude and \( a \) is a constant vector. It is obvious that \( r_{jk} \) and \( \dot{r}_j \), and hence \( L \), remain unchanged. From (11), we then have \( G(r, p) = a \cdot \sum_j p_j = \text{COM} \). Since \( a \) is constant and arbitrary, the total linear momentum \( P = \sum_j p_j \) is a COM.

(b) Next, consider the variation \( \delta r_j = \epsilon (n \times r_j) \), where \( n \) is any constant unit vector. This is just the change in \( r_j \) under an infinitesimal rotation of the coordinate axes about the direction of the unit vector \( n \). Comparing this with the general form \( \delta r = \epsilon \phi(r) \), we have \( \phi(r_j) = n \times r_j \). The variation now depends on the generalized coordinates \( r_j \). Under this infinitesimal change,

\[
\delta \dot{r}_j = \frac{d}{dt}(\delta r_j) = \epsilon (n \times \dot{r}_j) \quad \text{and} \quad \delta r_{jk} = 0. \tag{37}
\]

Hence \( r_j \cdot \delta r_j = 0 \) and \( \dot{r}_j \cdot \delta \dot{r}_j = 0 \), so that \( L \) remains unchanged up to first order in \( \epsilon \), i.e.,

\[
L(r_j + \delta r_j, \dot{r}_j + \delta \dot{r}_j) \approx L(r_j, \dot{r}_j) \quad \text{or} \quad \delta L = 0. \tag{38}
\]

The COM is therefore

\[
G(r, p) = \sum_j p_j \cdot \phi(r_j) = \sum_j p_j \cdot (n \times r_j)
\]

\[
= \sum_j (n \cdot (r_j \times p_j)) = n \cdot \sum_j (r_j \times p_j). \tag{39}
\]

Since \( n \) is constant and arbitrary, the total orbital angular momentum \( L = \sum_j (r_j \times p_j) \) is a COM.
Symmetries and Conservation Laws in Classical and Quantum Mechanics:

(c) Third, consider the explicitly time-dependent variation \( \delta \mathbf{r}_j = -\varepsilon \mathbf{v} t \), where \( \mathbf{v} \) is a constant vector. (Obviously, this variation simulates the effect of shifting to a mutually inertial frame moving with a uniform velocity \( \varepsilon \mathbf{v} \) with respect to the original frame of reference.) As before, using the rule given by (10), we have

\[
\frac{d}{dt} \delta \mathbf{r}_j = \frac{d}{dt} \mathbf{r}_j = \delta \mathbf{r}_j = -\varepsilon \mathbf{v}.
\]

The change in the Lagrangian up to first order in \( \varepsilon \) turns out to be

\[
L(\mathbf{r}_j + \delta \mathbf{r}_j, \dot{\mathbf{r}}_j + \delta \dot{\mathbf{r}}_j) \approx L(\mathbf{r}_j, \dot{\mathbf{r}}_j) - \varepsilon \frac{d}{dt} \sum_j m_j \mathbf{v} \cdot \mathbf{r}_j,
\]

so that

\[
\delta L = -\varepsilon \frac{d}{dt} \sum_j m_j \mathbf{v} \cdot \mathbf{r}_j.
\]

The COM can now be written down by looking at the general form in (22) and deducing \( G \), or else by using (29) and (30a). We get

\[
G(\mathbf{r}, \mathbf{p}) = -t \mathbf{v} \cdot \sum_j \mathbf{p}_j + \mathbf{v} \cdot \sum_j m_j \mathbf{r}_j
\]

\[
= \mathbf{v} \cdot \left[ \sum_j m_j \mathbf{r}_j - \mathbf{P} t \right].
\]

Since \( \mathbf{v} \) is constant and arbitrary, \( \sum_j m_j \mathbf{r}_j - \mathbf{P} t \) is a COM. If \( M \) is the total mass and \( \mathbf{R} \) is the position vector of the centre of mass of the system of particles, the COM can also be expressed, apart from the overall constant factor \( M \), as \( \mathbf{R} - \mathbf{P} t / M \). You will recognise that this is nothing but the initial position vector of the centre of mass. The numerical values of the three components of this COM are therefore given by the coordinates of the centre of mass at \( t = 0 \).

A final remark is in order here. The Lagrangian \( L \) in (36) does not have any explicit \( t \)-dependence. Hence the corresponding Hamiltonian \( H = \sum_1^N \mathbf{p}_j \cdot \dot{\mathbf{r}}_j - L \) is a COM. Together with the three components each of \( \mathbf{P}, \mathbf{L} \) and \( \mathbf{R} - \mathbf{P} t / M \), this makes a total of 10 independent COMs. These are called the Galilean invariants. For a general potential \( V(\{r_{jk}\}) \), these are the only independent COMs (in the form of smooth functions of the dynamical variables and time) that the system possesses. In the language of the theory of dynamical systems, the system is, in general, non-integrable, and the dynamics is generically chaotic.

3.2 Relativistic Free Particle

Consider a relativistic free particle of rest mass \( m \) moving with a velocity \( \mathbf{v}(t) \) relative to an inertial frame \( S \). Its Lagrangian is

\[
L = -m c^2 \sqrt{1 - \mathbf{v}^2(t)/c^2}.
\]
The time argument of the instantaneous velocity $\dot{r}$ has been indicated explicitly with good reason, as will become clear shortly. Under an infinitesimal Lorentz boost $\varepsilon \mathbf{v}$ to a frame $S'$, the space-time coordinates $(r, t)$ of the particle transform to $(r', t')$, where

$$ r'(t') \approx r(t) - \varepsilon \mathbf{v} t, \quad t' \approx t - \varepsilon \mathbf{v} \cdot r(t) / c^2. \quad (45) $$

The crucial point is that we can regard the difference $\delta r(t) \equiv r'(t) - r(t)$ (note especially the time argument of $r'$ in this definition!) as a variation of the coordinate $r$ in the frame $S$. The variations $\delta r(t)$ and $\delta \dot{r}(t)$ are then given, to first order in $\varepsilon$, by

$$ \delta r(t) = -\varepsilon \left[ \mathbf{v} t - (\mathbf{v} \cdot \mathbf{r}) \dot{r} / c^2 \right], \quad \delta \dot{r}(t) = -\varepsilon \left[ \mathbf{v} - (\mathbf{v} \cdot \mathbf{r}) \mathbf{r} / c^2 - (\mathbf{v} \cdot \mathbf{r}) \ddot{r} / c^2 \right]. \quad (46) $$

After some algebra, the change $\delta L$ in $L$ works out to

$$ \delta L = -\varepsilon m \frac{d}{dr} \left[ (\mathbf{v} \cdot \mathbf{r}) \sqrt{1 - \dot{r}^2 / c^2} \right]. \quad (47) $$

Since this is in the form of the total derivative of a function with respect to the time, we have an associated COM. Using the well-known relations

$$ \mathbf{p} = m \dot{r} / \sqrt{1 - \dot{r}^2 / c^2} \quad \text{and} \quad \dot{r} = c \mathbf{p} / \sqrt{\mathbf{p}^2 + m^2 c^2}, \quad (48) $$

this COM is given by

$$ G = -\mathbf{p} \cdot \left[ \mathbf{v} t - (\mathbf{v} \cdot \mathbf{r}) \dot{r} / c^2 \right] + m (\mathbf{v} \cdot \mathbf{r}) \sqrt{1 - \dot{r}^2 / c^2} $$

$$ = \mathbf{v} \cdot \left[ -\mathbf{p} t + m \mathbf{r} / c^2 + m \mathbf{r} \sqrt{1 - \dot{r}^2 / c^2} \right] $$

$$ = \mathbf{v} \cdot [(E / c^2) \mathbf{r} - \mathbf{p} t], \quad (49) $$

where we have used the relation $E = (\mathbf{p}^2 c^2 + m^2 c^4)^{1/2}$ in the last line. As before, since $G$ is a COM for any constant vector $\mathbf{v}$, we may conclude that the combination $(E / c^2) \mathbf{r} - \mathbf{p} t$ is a COM. In the limit $c \to \infty$ (implying that $|\dot{r}| / c \to 0$ or $|\mathbf{p}| / (mc) \to 0$), this COM reduces to the Galilean invariant $(m \mathbf{r} - \mathbf{p} t)$ of a free nonrelativistic particle.

### 3.3 The Two-Dimensional Isotropic Harmonic Oscillator

Consider the two-dimensional isotropic harmonic oscillator, namely, a particle of mass $m$ moving in a plane in the circularly symmetric potential $\frac{1}{2} m \omega^2 r^2$. Let $q_\alpha (\alpha = 1, 2)$ be the two Cartesian coordinates in this plane. The Lagrangian is

$$ L = \frac{1}{2} m q_\alpha q_\alpha - \frac{1}{2} m \omega^2 q_\alpha q_\alpha, \quad (50) $$
where the repeated index $\alpha$ is to be summed over, as usual. This system has an interesting symmetry, with associated COMs. A comprehensive way of finding them is as follows.

Consider the group $U(2)$ of unitary $(2 \times 2)$ matrices, and consider an element $u$ of this group that is infinitesimally close to the identity matrix. Then $u$ is of the form $u \approx I + i \epsilon h$, where $h$ is a hermitian $(2 \times 2)$ matrix. In terms of matrix elements,

$$u_{\alpha \beta} = \delta_{\alpha \beta} + i \epsilon h_{\alpha \beta}, \quad \text{where} \quad h^*_{\alpha \beta} = h_{\beta \alpha}. \quad (51)$$

But $h_{\alpha \beta}$ itself can be written as the sum of a real symmetric part and a pure imaginary antisymmetric part, according to

$$h_{\alpha \beta} = \frac{(h_{\alpha \beta} + h^*_{\alpha \beta})}{2} + i \frac{(h_{\alpha \beta} - h^*_{\alpha \beta})}{2i} \equiv s_{\alpha \beta} + i a_{\alpha \beta}, \quad (52)$$

where $s_{\alpha \beta} = s_{\beta \alpha}$ and $a_{\alpha \beta} = -a_{\beta \alpha}$. (Hence $a_{11} = a_{22} = 0$.) Now, observe that the variations

$$\delta q_\alpha = \epsilon (a_{\alpha \beta} q_\beta - \omega^{-1} s_{\alpha \beta} \dot{q}_\beta), \quad \delta \dot{q}_\alpha = \epsilon (a_{\alpha \beta} \dot{q}_\beta - \omega^{-1} s_{\alpha \beta} q_\beta) \quad (53)$$

lead to a dynamical symmetry, because the corresponding change in $L$ is given by a total time derivative:

$$\delta L = \epsilon \frac{dF}{dt}, \quad \text{where} \quad F = \frac{m}{2} (\omega q_\alpha q_\beta - \omega^{-1} \dot{q}_\alpha \dot{q}_\beta) s_{\alpha \beta}. \quad (54)$$

The corresponding COM is $\epsilon G = p_\alpha \delta q_\alpha - \epsilon F$, where

$$G = p_\alpha (a_{\alpha \beta} q_\beta - \omega^{-1} s_{\alpha \beta} \dot{q}_\beta) - \frac{m}{2} (\omega q_\alpha q_\beta - \omega^{-1} \dot{q}_\alpha \dot{q}_\beta) s_{\alpha \beta}$$

$$= a_{12} (p_1 q_2 - p_2 q_1) - \frac{s_{\alpha \beta}}{2} \left( \frac{p_\alpha p_\beta}{m \omega} + m \omega q_\alpha q_\beta \right). \quad (55)$$

But the matrix element $a_{12}$ and the three matrix elements $s_{11}$, $s_{22}$ and $s_{12} (= s_{21})$ are independent and arbitrary real numbers. Therefore (55) implies that we have four COMs, namely,

$$\begin{align*}
c_1 &= (m \omega)^{-1} p_1^2 + m \omega q_1^2, \\
c_2 &= (m \omega)^{-1} p_2^2 + m \omega q_2^2, \\
c_3 &= p_1 q_2 - p_2 q_1, \\
c_4 &= (m \omega)^{-1} p_1 p_2 + m \omega q_1 q_2. \quad (56)
\end{align*}$$

The COM $c_3$ arises as the coefficient of $a_{12}$. But $\delta L$ has no $a_{12}$ term. Therefore $c_3$ is a kinematical COM. You will recognise it as the orbital angular momentum of the particle about the centre of force (the origin). The other three COMs represent dynamical symmetry. $c_1$ and $c_2$ may be interpreted (apart from a constant factor) as the individual energies of two independent harmonic oscillators, which is another way of looking at the two-dimensional oscillator.
We can choose specific combinations of the COMs above to express them in a more compact form. Define a complex column vector \( a \) and its Hermitian conjugate \( a^\dagger \) according to
\[
a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad a^\dagger = \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix},
\]
where \( a_\alpha = \frac{p_\alpha - im_\omega q_\alpha}{\sqrt{m_\omega}}, \) (\( \alpha = 1, 2 \)).

The COMs can then be re-defined as
\[
s_0 = c_1 + c_2 = a^\dagger a, \quad s_1 = c_1 - c_2 = a^\dagger \sigma_3 a, \quad s_2 = 2c_4 = a^\dagger \sigma_1 a, \quad s_3 = -2c_3 = a^\dagger \sigma_2 a,
\]
in terms of the Pauli matrices \( \sigma_i, \) \( i = 1, 2, 3. \)

3.4 The Kepler Problem and the Laplace–Runge–Lenz Vector

The Lagrangian of a particle of mass \( m \) moving in the Coulomb potential \( V(r) = -\kappa/r \) is
\[
L = \frac{1}{2} m \dot{r}^2 + \frac{\kappa}{r},
\]
where the constant \( \kappa \geq 0 \) according as the inverse-square force is attractive or repulsive. The Hamiltonian \( H = p^2/(2m) - \kappa/r \) is a COM because \( L \) does not have explicit \( t \)-dependence. The three components of the orbital angular momentum \( L = r \times p \) are COMs because the potential \( V(r) \), and hence \( L \), are rotationally invariant. Over and above these, this system has another vector-valued COM, the well-known Laplace–Runge–Lenz vector. We may deduce this as follows.

Let \( a \) be an arbitrary vector of infinitesimal magnitude (\( |a| \ll 1 \)). Consider the virtual displacement in \( r \) (see, e.g., Saletan and Cromer 1971) given by
\[
\delta \dot{r} = a \times (\dot{r} \times r) + r \times (\dot{r} \times a) = 2(a \cdot r) \dot{r} - (a \cdot \dot{r}) r - (r \cdot \dot{r}) a.
\]

Hence the variation in the velocity, \( \delta \dot{r} = d(\delta \dot{r})/dr \), is
\[
\delta \dot{r} = (a \cdot \dot{r}) \dot{r} - \dot{r}^2 a + 2(a \cdot r) \dot{r} - (a \cdot \dot{r}) r - (r \cdot \dot{r}) a.
\]

The incremental change in the Lagrangian turns out to be the total time derivative of a function \( F(r, \dot{r}) \), namely,
\[
\delta L = \frac{dF}{dt} = \frac{d}{dt} \left[ m \left( (a \cdot \dot{r}) \dot{r}^2 - (a \cdot \dot{r}) (r \cdot \dot{r}) \right) + (a \cdot r) \frac{\kappa}{r^2} \right].
\]
The associated (dynamical) COM is then

\[
G = \mathbf{p} \cdot \delta \mathbf{r} - F \\
= \mathbf{p} \cdot \delta \mathbf{r} - m \left[ (\mathbf{a} \cdot \mathbf{r}) \dot{\mathbf{r}}^2 - (\mathbf{a} \cdot \dot{\mathbf{r}}) (\mathbf{r} \cdot \dot{\mathbf{r}}) \right] - (\mathbf{a} \cdot \mathbf{r}) \frac{k}{r} \\
= m^{-1} \mathbf{a} \cdot \left[ \mathbf{p} \times (\mathbf{r} \times \mathbf{p}) - \kappa m \mathbf{e}_r \right] \\
= m^{-1} \mathbf{a} \cdot \left[ (\mathbf{p} \times \mathbf{L}) - \kappa m \mathbf{e}_r \right],
\]

(63)

where \( \mathbf{e}_r = \mathbf{r}/r \) is the unit vector in the radial direction. Since \( \mathbf{a} \) is an arbitrary constant vector, the conserved quantity is \( (\mathbf{p} \times \mathbf{L}) - \kappa m \mathbf{e}_r \), the Laplace–Runge–Lenz vector. Note that it is a COM for either sign of the constant \( \kappa \). In the more familiar case of planetary motion (in an attractive inverse-square force), the orbits in physical space are ellipses. It is not hard to see that the Laplace–Runge–Lenz vector is in the direction of the semi-major axis of the ellipse in each case. Its conservation implies that there is no precession of the orbits in a pure inverse-square force field.

In Part 2, we shall extend the foregoing discussion to the case of QM, and also comment further on some group-theoretical and algebraic aspects of COMs vis-à-vis the generation of symmetry transformations.

**Suggested Reading**


Symmetries and Conservation Laws in Classical and Quantum Mechanics: Quantum Mechanics

K S Mallesh, S Chaturvedi, V Balakrishnan, R Simon and N Mukunda

In Part 1 of this two-part article we have spelt out, in some detail, the link between symmetries and conservation principles in the Lagrangian and Hamiltonian formulations of classical mechanics (CM). In this second part, we turn our attention to the corresponding question in quantum mechanics (QM). The generalization we embark upon will proceed in two directions: from the classical formulation to the quantum mechanical one, and from a single (infinitesimal) symmetry to a multi-dimensional Lie group of symmetries. Of course, we always have some definite physical system in mind. We also assume that the reader is familiar with the elements of quantum mechanics at the level of a standard first course on the subject. Operators will be denoted with an overhead caret, e.g., \( \hat{A}, \hat{G}, \hat{U}, \) etc., while \([\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}\) is the commutator of \( \hat{A} \) and \( \hat{B} \).

1. Symmetry in Quantum Mechanics

The treatment of symmetry and invariance in QM is closely modelled on the Hamiltonian formalism in CM. As is well known, we have the formal replacements

- Poisson bracket (PB) in CM \( \rightarrow \) commutator/(\(i\hbar\)) in QM.
- Canonical transformation (CT) in phase space \( \rightarrow \) unitary transformation (UT) on Hilbert space.

These statements will be qualified and elaborated upon, subsequently. Corresponding to any continuous symmetry of a quantum system, we have a constant of the motion (COM) that is now a hermitian operator \( \hat{G} \). A finite symmetry transformation, as opposed to an infinitesimal one, is represented by a UT built up from a succession of infinitesimal transformations. It has the general form

\[
\hat{U}(\alpha) = e^{-i\alpha \hat{G}/\hbar}.
\]  

(1)

Here, \( \hat{G} \) is the infinitesimal generator (often abbreviated to simply ‘the generator’) of the transformation, and \( \alpha \) is the (real) parameter characterising the transformation. The constant \( \hbar \) has been introduced explicitly in the exponent in (1) for convenience – this is the form in which unitary transformations commonly occur in QM. Note that the product \( (\alpha \hat{G}) \) has the same
physical dimensions as $\hbar$, i.e., those of angular momentum, or (length) $\times$ (linear momentum). The effects of the transformation on state vectors (or wave functions) and on dynamical variables, respectively, are given by

$$\rho \rightarrow \rho' = \hat{U}(\alpha) \rho \hat{U}^{-1}(\alpha).$$

Since $\hat{G}$ is hermitian and $\hat{U}(\alpha)$ is unitary, we have

$$\hat{U}^{-1}(\alpha) = \hat{U}^\dagger(\alpha) = e^{i\alpha \hat{G}/\hbar}.$$ (3)

A combination like $\hat{U} \hat{A} \hat{U}^{-1}$ is called the conjugation of $\hat{A}$ by $\hat{U}$.

The equation of motion (EOM) in quantum mechanics is the Schrödinger equation for the state vector of a system, namely,

$$i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H}(t) |\Psi(t)\rangle,$$ (4)

where $\hat{H}(t)$ is the Hamiltonian of the system. For generality, we have allowed for a possible explicit time-dependence in $\hat{H}$. Now, a dynamical variable $\hat{G}(t)$ with possible explicit time dependence is a COM if and only if

$$\frac{d}{dt} [\hat{G}(t), \hat{H}(t)] = 0,$$ i.e.,

$$\frac{d}{dt} \hat{G}(t) = \frac{i}{\hbar} [\hat{G}(t), \hat{H}(t)].$$ (5)

Observe that an explicitly time-dependent COM cannot commute with the Hamiltonian. The noteworthy point is that, even in the general case in which the COM has an explicit time dependence, the corresponding unitary transformation preserves the EOM (i.e., the Schrödinger equation). That is, under the UT

$$|\Psi(t)\rangle \rightarrow |\Psi'(t)\rangle = e^{-i\alpha \hat{G}(t)/\hbar} |\Psi(t)\rangle,$$ (6)

where $\alpha$ is time-independent, we continue to have

$$i\hbar \frac{d}{dt} |\Psi'(t)\rangle = \hat{H}(t) |\Psi'(t)\rangle.$$ (7)

It is a simple exercise to establish (7), and we invite the reader to do so. First set $|\Psi(t)\rangle = \hat{U}^{-1}(\alpha) |\Psi'(t)\rangle = e^{i\alpha \hat{G}(t)/\hbar} |\Psi'(t)\rangle$ in (4). Note that it is the explicit $t$-dependence in $\hat{G}(t)$ that is to be differentiated with respect to $t$. We need the quantity $(\partial / \partial t) e^{i\alpha \hat{G}(t)/\hbar}$. Expanding the exponential in its power series, we need $\partial \hat{G}^n / \partial t$ for all positive integer values of $n$. Care must be exercised, because $\hat{G}$ and $\partial \hat{G} / \partial t$ do not commute with each other. Hence

$$\frac{\partial \hat{G}^2}{\partial t} = \frac{\partial \hat{G}}{\partial t} \hat{G} + \hat{G} \frac{\partial \hat{G}}{\partial t} = i \hbar [\hat{G}^2(t), \hat{H}(t)].$$ (8)
using (5). Proceeding in this manner, we get
\[
\frac{\partial}{\partial t} \hat{G}^n(t) = \frac{i}{\hbar} [\hat{G}^n(t), \hat{H}(t)],
\]
and hence
\[
\frac{\partial}{\partial t} e^{i\alpha \hat{G}(t)/\hbar} = \frac{i}{\hbar} [e^{i\alpha \hat{G}(t)/\hbar}, \hat{H}(t)].
\]
(9)
Equation (7) then follows easily.

Note that it is the original Hamiltonian \(\hat{H}(t)\) that appears on the right-hand side of (7). This means that solutions of the Schrödinger equation with a Hamiltonian \(\hat{H}(t)\) are mapped onto other, generally different, solutions of the same equation by a unitary symmetry. The existence of state vectors in QM, on which UTs can act directly, is a new feature without a parallel in CM, as we shall see. This requires a deeper comparison of the two schemes.

Consider, first, the situation in CM. Recall from Part 1 that an infinitesimal CT with parameter \(\epsilon\) and generator \(G(q, p)\) implies that
\[
q_r \rightarrow q_r + \epsilon \{q_r, G(q, p)\}, \quad p_r \rightarrow p_r + \epsilon \{p_r, G(q, p)\}.
\]  
(10)
A succession of \(n\) such infinitesimal CTs amounts, in the limit \(n \rightarrow \infty, \epsilon \rightarrow 0\) such that
\[
\lim_{n \rightarrow \infty} \epsilon \rightarrow \alpha,
\]
where
\[
f(q, p) \xrightarrow{C(\alpha)} f'(q, p) = f(q, p) + \alpha \sum \{G(q, p), f(q, p)\}
\]
(11)
The series on the right-hand side can be written in a compact form by defining the differential operator
\[
\mathcal{G}(q, p) = \frac{\partial G(q, p)}{\partial q_r} \frac{\partial}{\partial p_r} - \frac{\partial G(q, p)}{\partial p_r} \frac{\partial}{\partial q_r}.
\]
(12)
(As always, summation over a repeated index is to be understood.) Then
\[
\frac{\alpha^2}{2!} \mathcal{G}(q, p) \mathcal{G}(q, p) f(q, p) + \cdots
\]
Thus, from the phase space function \(G(q, p)\) we construct the first-order linear partial differential operator \(\mathcal{G}(q, p)\) on phase space functions; and by exponentiating this operator, we get the finite classical CT \(C(\alpha)\). Symbolically,
\[
C(\alpha) f(q, p) = e^{\alpha \mathcal{G}(q, p)} f(q, p).
\]
(14)
There is an analogue of the foregoing in QM, at the level of the action of UTs by conjugation on dynamical variables or operators. From (2b) we have

$$\hat{A} \xrightarrow{U(\alpha)} \hat{A}' = \hat{U}(\alpha)\hat{A}\hat{U}^{-1}(\alpha) = e^{-i\alpha \hat{G}/\hbar} \hat{A} e^{i\alpha \hat{G}/\hbar}$$

$$= \hat{A} - \frac{(i\alpha/\hbar)}{1!} [\hat{G}, \hat{A}] + \frac{(i\alpha/\hbar)^2}{2!} [\hat{G}, [\hat{G}, \hat{A}]] + \ldots$$

(15)

In writing the last equation above, we have used Hadamard’s operator identity for an expression of the form $e^{-\hat{B}} \hat{A} e^{\hat{B}}$, namely:

$$e^{-\hat{B}} \hat{A} e^{\hat{B}} = \hat{A} - [\hat{B}, \hat{A}] + \frac{1}{2!} [[\hat{B}, [\hat{B}, \hat{A}]] + \ldots .$$

(16)

As before, the nested commutators can be written in a compact form by defining an operator $\hat{G}$ that acts on operators according to

$$\hat{G}(\cdot) = (i\hbar)^{-1} [\hat{G}, \cdot].$$

(17)

Then

$$\hat{A}' = \hat{A} + \frac{\alpha}{1!} \hat{G}\hat{A} + \frac{\alpha^2}{2!} \hat{G}\hat{G}\hat{A} + \cdots = e^{\alpha \hat{G}} \hat{A}.$$ 

(18)

Here, the action of taking the commutator of $\hat{G}/(i\hbar)$ with a general operator has been promoted to an operator $\hat{G}$ that acts on operators themselves. Such an object is known as a super-operator or dynamical map. We thus have the parallel situations in classical and quantum mechanics:

- In CM: Phase space function $G(q, p) \rightarrow G(q, p)$: $G f = \{G, f\}$, any $f$.
- In QM: Operator $\hat{G}$ on Hilbert space $\rightarrow \hat{G}$: $\hat{G}\hat{A} = (i\hbar)^{-1}[\hat{G}, \hat{A}]$, any $\hat{A}$.

$\hat{G}$ arises from $G$, but the two have different mathematical natures. Similarly, $\hat{G}$ arises from $\hat{G}$, and again they are different kinds of mathematical objects. The PB in $G$ is replaced by the commutator (divided by $i\hbar$) in $\hat{G}$. But the special feature of QM is the existence of state vectors $|\Psi\rangle$ on which $\hat{U}(\alpha)$ and $\hat{G}$ can act directly, in contrast to the conjugation on operators. We may say:

(i) In CM: $G$, the dynamical variable $\neq \hat{G}$, the transformation generator.

(ii) In QM: $\hat{G}$, the dynamical variable $=\hat{G}$, the transformation generator, in its action on state vectors.

We can appreciate this difference in yet another way: In CM there is no significant role, and indeed generally no meaning because of dimensional reasons, for the numerical exponential expression $e^{\alpha G(q, p)}$; while in QM the analogous operator expression in (1), $\hat{U}(\alpha) = e^{-i\alpha \hat{G}/\hbar}$, is meaningful and important.
As mentioned already, functions of the dynamical variables of a classical Hamiltonian system are replaced by operators when we go over to quantum mechanics. Unlike classical variables, these operators do not commute with each other, in general. As a result, the ‘mapping’ from the space of functions $f(q, p)$ to the space of operators $\hat{f}(\hat{q}, \hat{p})$ is not unique, except in the simplest cases. When we attempt to replace the classical Hamiltonian $H(q, p)$ by the quantum mechanical Hamiltonian $\hat{H}(\hat{q}, \hat{p})$, the order in which the operators appear in various terms of $\hat{H}$ is not uniquely specified, and this leads to ambiguities. Up to the level of quadratic functions of the $q$’s and $p$’s, it is quite straightforward to make the classical variable → quantum operator correspondence, by the direct replacements $q \mapsto \hat{q}$, $p \mapsto \hat{p}$. The only non-trivial correspondence rule at this level is the replacement of the classical product $qp (= pq)$ by the symmetric operator combination $\frac{1}{2}(\hat{q}\hat{p} + \hat{p}\hat{q})$. But the uniqueness of the correspondence breaks down even for polynomials of order higher than the second, let alone other functions of $(q, p)$.

A general procedure for the classical-to-quantum correspondence was first given by Weyl, and further generalized and developed by Moyal, Stratonovich, Wigner, and others. The Moyal–Stratonovich–Weyl quantization procedure enables us to start with a classical Hamiltonian system and obtain consistently the operators corresponding to arbitrary functions of the classical $q$’s and $p$’s in such a manner that (i) the Poisson bracket → (commutator)/$i\hbar$ requirement is satisfied, up to the leading order in $\hbar$ (see below); and (ii) a symmetry transformation acting on a classical phase space function $f(q, p)$ is implemented by a unitary transformation acting on the corresponding operator $\hat{f}(\hat{q}, \hat{p})$. For our present purposes, these are the relevant properties for extending the consideration of symmetry generators in CM to the case of QM.

We mention that the quantization procedure still leaves open the question of the ordering of operators in the quantum mechanical counterpart of a classical system. More than one ordering prescription is possible. There are at least three different standard prescriptions, each leading to a specific ‘quasi-probability distribution’ in phase space that can be used to reproduce the quantum mechanical expectation values of any operator $\hat{f}(\hat{q}, \hat{p})$ as the statistical average of the corresponding phase space function $f(q, p)$ weighted by the distribution.

Finally, for the sake of completeness, we make a comment on the inverse problem of expressing the Hilbert space formulation of quantum mechanics, involving operators and expectation values, in terms of a purely statistical description involving distribution functions in phase space. It turns out that, in order to do this consistently, and to take into account fully the non-commutativity of quantum mechanical operators, the correct replacement of the commutator (divided by $i\hbar$) is the so-called Moyal bracket, which is the Poisson bracket plus an infinite series of ‘correction’ terms involving higher derivatives with respect to the $q$’s and $p$’s, multiplied by powers of $\hbar$ starting with $\hbar^2$.

Reading the statement in (ii) above ‘backwards’, we see that in QM, operators that generate or represent transformations on state vectors are themselves observable quantities. That is why, in quantum mechanics, things like the parity transformation (or space inversion) and the operator corresponding to the permutation of identical particles are observables or dynamical variables. They obey the Heisenberg EOM, have eigenvectors and measurable eigenvalues, and so on. But this is not so at all classically: there, parity and permutation are solely rules of transformation, and are not themselves dynamical variables.
2. Lie Groups of Symmetries

Some of the foregoing comments are meant to be a comparison of the structural features of classical versus quantum mechanics. Clearly, they are not limited to our main theme, which is the connection between symmetry principles and conservation laws.

We revert now to that theme in the multi-dimensional case. A Lie group is a well-defined mathematical object, and it could be relevant in both CM and QM as a group of symmetries of some physical system or class of systems. Examples connected with space-time are the three-dimensional proper rotation group $SO(3)$, the Euclidean group $E(3)$, the Galilei group relevant to Newtonian mechanics, the proper orthochronous homogeneous Lorentz group $SO(3,1)$ of special relativity, and the Poincaré group of inhomogeneous Lorentz transformations.

Table 1. Pattern of classical and quantum realizations of symmetry groups.

<table>
<thead>
<tr>
<th>$\mathcal{G}$</th>
<th>CM</th>
<th>QM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(\alpha) = \exp(\alpha_j e_j)$, $g(\alpha)g(\beta) = g(f(\alpha, \beta))$</td>
<td>$C(\alpha) = e^{\alpha_j G_j}$, $C(\alpha)C(\beta) = C(f(\alpha, \beta))$</td>
<td>$\hat{U}(\alpha) = e^{-i\alpha_j \hat{G}_j/h}$, $\hat{U}(\alpha)\hat{U}(\beta) = e^{i\omega(\alpha, \beta)} \hat{U}(f(\alpha, \beta))$</td>
</tr>
<tr>
<td>$[G_j, G_k] = C_{jk}^l G_l$</td>
<td>$[\hat{G}_j, \hat{G}<em>k] = C</em>{jk}^l \hat{G}_l$</td>
<td></td>
</tr>
<tr>
<td>$[G_j, G_k] = C_{jk}^l G_l + d_{jk}$</td>
<td>$(ih)^{-1}[\hat{G}_j, \hat{G}<em>k] = C</em>{jk}^l \hat{G}<em>l + d</em>{jk}$</td>
<td></td>
</tr>
</tbody>
</table>

We stress that one and the same group may be relevant, and be represented or expressed, in one form of mechanics or the other. Hence the group structure must go with the appropriate mechanical formalism. The situation is best conveyed by the flow chart in Table 1, displaying how a given group $\mathcal{G}$ of symmetries is represented classically and quantum mechanically. Let us explain the symbols and notation. Elements of the group $\mathcal{G}$ are written as $g(\alpha)$, where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ are the independent real parameters or coordinates that label the elements continuously and vary within their specified ranges. Here $n$ is the order or dimension of $\mathcal{G}$, and the $\alpha_j$’s chosen are called canonical coordinates of the first kind. In the case of $SO(3)$, for instance, these are the so-called axis-angle parameters. (The Euler angles frequently used to specify rotations in three-dimensional space are not canonical coordinates of the first kind.) Group composition in $\mathcal{G}$ is expressed in coordinates by the $n$ real functions $\{f_j(\alpha, \beta), j = 1, 2, \ldots, n\} \equiv f(\alpha, \beta)$ of $2n$ independent real arguments each. The functions $f(\alpha, \beta)$ are restricted by the laws
of group structure, namely: associativity, the existence of the identity element, and the existence of the inverse of every element. \{e_j\} is the set of basis vectors in the Lie algebra. This is a real linear vector space whose elements describe elements in \(G\) that lie very close to the identity – what we have referred to (in Part 1) as ‘infinitesimal transformations’. The expression of a finite group element \(g(\alpha)\) in exponential form is in a sense symbolic – reminiscent of (14) and (18). The set of Lie bracket relations among the basis vectors, \([e_j, e_k] = C_{jk}^l e_l\), is a local expression of the group multiplication law. The coefficients \(C_{jk}^l\) are called the structure constants of the Lie algebra. The quantities \(d_{jk}\) that appear in the last lines of each column are called neutral elements, and we shall say a little more about them shortly.

Moving down along the two arms of Table 1, finite group elements are realized as CTs or UTs, respectively. They are ‘honest’ exponentials of \(\alpha_j G_j\) and \(\alpha_j \hat{G}_j/(ih)\), which represent \(\alpha_j e_j\) in CM and in QM, respectively. Thanks to the commutation relations satisfied by \(G_j\) and \(\hat{G}_j\) in the two cases, the composition laws among CTs and UTs hold as they do in the group, apart from some phase factors in the case of UTs. (And apart, also, from certain issues related to the ‘global’ properties of the parameter space of the group. We do not discuss these here.) These commutation relations are concrete realizations of the abstract Lie bracket relations among the \(e_j\)'s, referred to earlier. In both arms, the associative law is automatic, and does not require separate demonstration or verification.

The important distinctive feature in QM at the level of the composition law for the UT \(\hat{U}(\alpha)\) is the appearance of the extra phase factor, \(e^{i\omega(\alpha, \beta)}\). Such a factor is permitted by the structure of QM, since the overall phase in a state vector \(|\Psi\rangle\) is physically unobservable. It turns out that, in general, \(\hbar \omega(\alpha, \beta)\) is an infinite series in the \(\alpha_j\)'s, \(\beta_j\)'s and the structure constants \(C_{jk}^l\), but linear in the neutral elements \(d_{jk}\). It is straightforward to derive this result, but we shall not do so here.

When we finally descend to the last line in the two columns, we obtain PB relations among the classical COMs \(G_j\), and commutation relations among the quantum COMs \(\hat{G}_j\). In order to go from the penultimate line to the final equations in the two columns, we may either write out all the Poisson brackets (respectively, commutators) involved, and simplify the resulting expressions; or else, and less tediously, we may use the algebraic relations in (20) and (21) below, which are expressions of the Jacobi identities for Poisson brackets and commutators, respectively. Let \(f, g\) and \(h\) be phase space functions such that \(\{f, g\} = h\). As in (12), define the corresponding operators

\[
\hat{f} = \frac{\partial f}{\partial q_r} \frac{\partial}{\partial p_r} - \frac{\partial f}{\partial p_r} \frac{\partial}{\partial q_r},
\]

and similarly \(g\) and \(h\). Then the Jacobi identity for Poisson brackets translates to the commutation relation

\[
[f, g] = h
\]
K S Mallesh, S Chaturvedi, V Balakrishnan, R Simon and N Mukunda
for the operators. Similarly, let \( \hat{A}, \hat{B} \) be operators on a Hilbert space, and let \( [\hat{A}, \hat{B}] = \hat{C} \). Further, as in the definition (17), let \( \hat{A} = (i\hbar)^{-1}\{\hat{A}, \cdot\} \) and similarly \( \hat{B}, \hat{C} \) be the corresponding superoperators. Then the Jacobi identity for commutators of operators translates to the relation

\[
[\hat{A}, \hat{B}] = (i\hbar)^{-1}\hat{C}
\]  

(21)
at the level of the superoperators.

It must be noted that the final equations in both columns are stated in terms of dynamical variables. Classically, the neutral elements \( d_{jk} \) are seen only at this concluding stage, while in QM they are ‘infinitesimal forms’ of the phases \( \omega(\alpha, \beta) \) that have appeared earlier. (This feature again reinforces the fact that the set \( \{\hat{U}(\alpha)\} \) of UTs is unique to QM, with no direct classical analogue.) There are, of course, certain functional relations these phases should satisfy, leading to conditions on the \( d_{jk} \)'s. Classically, they have vanishing PBs with all phase space functions, while quantum mechanically their commutators with all dynamical variables vanish.

**Box 2. A Basic Difference in the Role of Symmetry in CM and QM.**

There is an important difference between CM and QM in the way a symmetry manifests itself. It is helpful to understand this difference with the help of a simple and familiar example.

Consider the Kepler problem, in which a classical particle moves in a closed orbit around a centre of force in an attractive Coulomb potential \( V(r) = -\kappa/r \), where \( \kappa \) is a positive constant. (Recall that, in Part 1, we have discussed the COMs in this instance, including also the Laplace–Runge–Lenz vector.) The potential, and hence the Hamiltonian of the particle, \( H = \frac{p^2}{2m} + V(r) \), are rotationally invariant – that is, \( H \) does not depend on the orientation of the coordinate axes. The particle is acted upon by a central force, so that there is no torque on it. Therefore its orbital angular momentum \( L = r \times p \) is a COM. This means that both the magnitude and the direction of \( L \) (and hence the plane of the orbit) remain unchanged in time as the particle traverses any orbit. But what determines \( L \) for any given orbit? Clearly, the initial conditions, i.e., the initial position \( r(0) \) and initial momentum \( p(0) \), do so.

Consider, for simplicity, the set of circular orbits corresponding to the set of initial conditions for which the magnitudes \( r(0) \) and \( p(0) \) are the same, but the vectors \( r(0) \) and \( p(0) \) point in all possible directions. The orbits in this set are circles of the same radius, but lying in all possible planes. However, a rotation of the coordinate axes will take us from any of these orbits to any other. Any particular orbit breaks the rotational invariance of the Hamiltonian, but spherical symmetry is restored in the full set of orbits.

In contrast, the situation in QM is quite different. All possible solutions of the Schrödinger equation that are related to each other by a symmetry transformation are superposed to constitute the state vector (or wave function) of the system. The quantum mechanical analogue of the classical example considered above is provided by the ground state of the electron in a hydrogen atom. This state has a spherically symmetric wave function, proportional to \( e^{-r/a_0} \), where \( a_0 \) is the Bohr radius. Thus the ground state retains the property of rotational invariance that the Hamiltonian enjoys. (This feature is quite general for quantum mechanical systems, and may be termed the Weyl–Wigner realization of symmetry, after Hermann Weyl and Eugene Wigner.) On the other hand, note that the most probable value of \( r \) in the ground state is not zero, but rather, the Bohr radius \( a_0 \); and yet the orbital angular momentum quantum number of the state is \( l = 0 \). This is only possible because, in a sense, all possible classical orbits are superposed to produce the single wave function corresponding to the ground state.
At this point, we can say more precisely which quantum quantities are analogues of which classical ones, following up on our opening remarks in Section 1. To the classical CT $C(\alpha)$ there corresponds the operation $\hat{U}(\alpha) \cdot \hat{U}^{-1}(\alpha)$ of conjugation of operators in QM, rather than $\hat{U}(\alpha)$ itself. The correspondences $G \leftrightarrow \hat{G}$ and $\hat{G} \leftrightarrow \hat{\hat{G}}$ then follow. We can think of the individual factor $\hat{U}(\alpha)$ in the conjugation operation as a ‘square-root’ of that operation in some sense, just as the state vector $|\Psi\rangle$ is a (complex) ‘square-root’ of the probability. A single factor $\hat{U}(\alpha)$ can act on $|\Psi\rangle$. As we have reiterated, this is a uniquely quantum mechanical feature. Some remarks on the physical meaning of the COMs $G_j$ and $\hat{G}_j$ are in order. When the Lie group $G$ is related to space-time, these COMs are important mechanical quantities. If $G$ is the 3-parameter group $SO(3)$ of spatial rotations, the generators are the components of the total angular momentum of the physical system. When spatial translations are included and $G$ becomes the 6-parameter Euclidean group $E(3)$, the additional generators are the components of the total linear momentum. Next, extension to the nonrelativistic Galilei group or to the relativistic Poincaré group, both of which are 10-parameter groups, brings in time translations and the three generators of pure velocity transformations (or boosts). The generator corresponding to time translations is the Hamiltonian. If the system has time-translation invariance, i.e., if the Hamiltonian $H$ has no explicit time dependence, then $H$ is a COM. (Its numerical value is the total energy of the system, in the usual situations.) The generators of boosts are somewhat non-trivial. The corresponding COMs are related to the motion of the centre of mass and the centre of energy, respectively, in the nonrelativistic and relativistic cases. Recall that, in Part 1, we have already illustrated how the 10 Galilean invariants of the motion arise explicitly in a system of particles interacting via 2-body potentials that depend only on the distances between pairs of particles. Each of the space-time groups listed above is in fact relevant for whole classes of physical systems.

A comment on the neutral elements $d_{jk}$: It so happens that, out of all the Lie groups relevant in physical contexts, the Galilei group is the only one where a nontrivial neutral element can be present, and in that case it is the total mass of the system. In all other cases -- $SO(3)$, $E(3)$, and the Poincaré group -- they turn out to be trivial, and can be ignored. Interestingly, in the case of Galilean quantum mechanics, the fundamental Heisenberg canonical commutation relations turn out to be a consequence of the Lie algebra relations pertaining to the Galilei group.

Finally, let us make a few comments on the dynamical symmetry group of any given system, in order to put this aspect in perspective, and also to summarise very briefly a few key points. Discovering the COMs of a dynamical system is not a trivial task, either in CM or in QM. In CM, we need to find all phase space functions $G(q, p, t)$ such that $dG/dt \equiv 0$. In QM, we need to find all quantum observables $\hat{G}$ such that $d\hat{G}/dt \equiv 0$. All these COMs will generate symmetries of the dynamical system. It helps to note that the PB of two COMs is again a COM, even if the two COMs have explicit time dependence. (It is an instructive exercise to establish this statement, which is called Poisson’s Theorem.) Likewise, in QM, the commutator of two COMs is again a COM. In each case, we must discover the largest possible group of such symmetries. This will be some subgroup of the group of CTs (in CM) or UTs (in QM) of the system.
No discussion of symmetry can be complete without a special mention of the great mathematician Emmy Noether, whose seminal work on the relationship between symmetry principles and conservation laws is a central theme in modern physics. The examples we have used as illustrations in this set of articles are special cases of her fundamental theorems, which find their full expression in systems with continuously infinite numbers of degrees of freedom – namely, classical and quantum fields. The Wikipedia article on Emmy Noether gives a succinct account of the life and achievements of this extraordinary intellect. Scientists and mathematicians of the calibre of Einstein, Hilbert and Weyl have acclaimed her as the greatest woman mathematician in recorded history. Among other achievements, an entire field of modern mathematics, abstract algebra, may be said to be her creation.

The existence of COMs and symmetry generators is closely linked to the integrability or otherwise of the dynamical system. This is a subject by itself, and we shall not discuss it here. However, some remarks about a specific example might be useful to the reader. This is the classical Kepler problem of a nonrelativistic particle moving in an inverse-square field of force, discussed in Part 1. We saw there that the COMs generating symmetries were the Hamiltonian $H$ (time-translation invariance), the angular momentum $L$ (rotation invariance) and the Laplace–Runge–Lenz vector. The first two are somewhat obvious symmetries, because $H$ is not explicitly time-dependent, and is clearly rotationally invariant. The existence of the Laplace–Runge–Lenz vector, however, shows that the Kepler problem truly has a further hidden symmetry. The (dynamical) symmetry group in this case is not just $SO(3)$, as it would be for a particle moving in a general spherically symmetric potential. It turns out to be a larger sub-group of the group of CTs of the system – it is $SO(4)$ in the case of an attractive $r^{-1}$ potential. The implications of this fact are equally interesting in the quantum-mechanical counterpart of the problem, namely, the (nonrelativistic) electron in a hydrogen atom. Symmetry, in general, implies degeneracy of the spectrum of the Hamiltonian, i.e., of the energy levels. For a particle in a general spherically symmetric potential, rotational invariance implies that the energy levels cannot depend on the ‘magnetic’ quantum number $m$. In the case of an attractive $r^{-1}$ potential, they do not depend on the orbital angular momentum quantum number $l$, either. The energy levels $E_n$ of the electron in a hydrogen atom depend only on the principal quantum number $n$. This absence of any dependence of $E_n$ on $l$ is familiar to you as accidental degeneracy. Its origin actually lies in the extra dynamical symmetry of this system.

3. Broken Symmetry

Our pedagogical account of the way continuous symmetries of dynamical systems are connected to conservation laws will hopefully convince the reader that this is a very important and beautiful component of both classical mechanics and quantum mechanics. The ideas of symmetry transformations, action principles, generators and COMs mesh together so intricately in the classical Hamiltonian and the quantum operator formalisms. One aspect of CM which should also be looked at in this context is the Hamilton–Jacobi formulation which stands, in some sense, at the same level as the state vector or wave function in QM. But this formulation
must be given a character that is globally well-defined in the phase space concerned. Proper implementation of this task requires a combination of ideas from symplectic and differential geometry. We have therefore not ventured into it here.

There is, however, another very important aspect of symmetry that is so fundamental in nature that it deserves at least a mention in this article. As mentioned at the outset in Part 1, our discussion, though restricted to systems with a finite number of degrees of freedom (NDF), can be generalized to systems with an infinite NDF. Two types of physical systems are of interest in this regard: (i) Macroscopic collections of interacting particles, studied in statistical mechanics; in the so-called thermodynamic limit, the number $N$ of particles and the volume $V$ of the system both tend to infinity such that their ratio remains finite. (ii) Fields of different kinds, for which the NDF is a continuous infinity – one or more degrees of freedom being associated with each point in space. In the context of symmetry, a very interesting and important new feature arises in such systems. This feature is absent in the case of systems with a finite NDF.

As mentioned in Box 2, the standard manner in which symmetry is manifested in QM is the Weyl–Wigner realization. The ground state of a system possesses the symmetry of the Hamiltonian. In systems with an infinite NDF, however, there is another, more subtle realization of symmetry that may be termed the Nambu–Goldstone realization, after Yoichiro Nambu and Jeffrey Goldstone. In this case, the system has a whole set of possible ground states, such that we can go from one ground state to another by a symmetry transformation (a transformation that keeps the Hamiltonian unchanged). The system selects one of this set of ground states – either randomly, triggered by fluctuations, or as a result of the manner in which it is prepared. An individual ground state does not possess the symmetry of the Hamiltonian. On the other hand, the set of possible ground states does have this symmetry. Moreover, there are certain collective excitations of the system (called Goldstone bosons) that connect the different possible ground states, and thus help restore the symmetry broken by the choice of a specific ground state. This is the phenomenon known as the spontaneous breakdown of symmetry. This phrase is something of a misnomer, because there is really no breakdown of symmetry, merely a different and less obvious realization of it. The term secret symmetry is also used in this context, but ‘spontaneous breakdown of symmetry’ is the more common name for the phenomenon.

An example from condensed matter physics is helpful in understanding the issue. Among other effects, the spontaneous breakdown of symmetry plays a crucial role in certain phase transitions in condensed matter. Consider a ferromagnetic material. We are concerned with a system at a non-zero temperature. Hence it is the free energy (rather than the Hamiltonian itself), and the thermodynamic equilibrium state (rather than the ground state) of the system, whose symmetry we have to consider. The equilibrium state of the material is effectively represented by its remnant or ‘spontaneous’ magnetization $M_0$. (The subscript serves to remind us that this is the magnetization in the absence of a magnetic field.) At any temperature above the Curie temperature of the material, the different atomic magnetic dipole moments (equivalently, spins) are disordered, and the net magnetization $M_0 = 0$. The interaction between these spins is invariant under rotations of the coordinate axes. Thus both the free energy and the magnetization (the counterparts of $H$ and the ground state) share the property of rotational symmetry.
K S Mallesh, S Chaturvedi, V Balakrishnan, R Simon and N Mukunda

(The null vector is rotationally invariant!) This is the situation in the so-called disordered or paramagnetic phase of the material.

When the temperature of the material is lowered to a value below the Curie temperature, the material goes into its ordered or ferromagnetic phase. It acquires a non-zero magnetization $M_0$. The equilibrium state of the material no longer has the rotational symmetry of the free energy, since the direction of $M_0$ singles out a special direction in space. The actual direction in which $M_0$ points is arbitrary, and may be selected in practice by placing the material in a small applied magnetic field in any desired direction, cooling the sample to below its Curie temperature, and then switching off the field. States corresponding to different possible directions of $M_0$ are connected by collective excitations known as spin waves, which are the Goldstone bosons in this instance.

Where does the condition of an infinite NDF come in? It turns out that there can be no spontaneous breakdown of symmetry in a system with a finite NDF because of the internal or intrinsic fluctuations in the system. Now, there are two primary, distinct, sources of such fluctuations in nature: quantum fluctuations (owing to the fact that $\hbar \neq 0$), and thermal fluctuations (at any non-zero temperature). They have, in general, the effect of restoring the symmetry of the state of the system. A simple example is provided by a particle moving in one spatial dimension, under the influence of the symmetric double-well potential $V(x) = k(x^2 - a^2)^2$, where $k$ and $a$ are positive constants. The Hamiltonian of the particle is invariant under the reflection (or, in this case, parity) transformation $x \mapsto -x$. The points $x = \pm a$ are minima of the potential. A classical particle in this potential has two stable equilibrium states of minimum energy: $(x, p) = (a, 0)$ and $(x, p) = (-a, 0)$. When it is in either of these two states, it breaks the reflection symmetry of the Hamiltonian, although the pair of equilibrium states does have that symmetry. Now consider what happens in QM, where the particle can tunnel from one well of the potential to the other. The true ground state of the particle has a wave function that is symmetric under $x \mapsto -x$, with maxima of equal heights at $x = \pm a$ and a minimum at $x = 0$. Thus, quantum fluctuations (tunnelling being one manifestation of these) lead to a ground state that has the same symmetry property as the Hamiltonian, in this case. More generally, quantum fluctuations, or thermal fluctuations, or a combination of both, ensure that the ground state (or equilibrium state) of any system with a finite NDF has the same symmetry as its Hamiltonian.

Only when the NDF is infinite is there a possibility of genuine spontaneous breakdown of symmetry, under suitable conditions. These conditions must ensure that phenomena such as tunnelling, or other effects of the intrinsic fluctuations in the system, do not suffice to restore the symmetry of the ground state or thermal equilibrium state. Take, for instance, the double-well potential example above. Imagine a lattice of such potentials, and a collection of particles in them. The probability of coherent or simultaneous tunnelling of all the particles may well vanish in the thermodynamic limit, as it would involve (in the lowest approximation) a product of $N$ ‘Gamow factors’ or negative exponentials representing individual tunnelling probabilities, each less than unity. On the other hand, if the system is at a sufficiently high temperature, the symmetry might well be restored by thermal fluctuations. This is, of course, only a rough example, and we have not mentioned other important factors. In practice, the occurrence or
otherwise of the spontaneous breakdown of symmetry depends on a rather small number of basic features. These include the precise nature of the symmetry (e.g., discrete versus continuous symmetry), the spatial dimensionality of the system, the number of components of the physical quantity characterizing the system (such as the vector $\mathbf{M}_0$ in the magnet example above), etc.

The spontaneous breakdown of symmetry is the mechanism underlying a large number of phase transitions in a wide variety of physical circumstances, ranging from superconductors to the quark-gluon plasma in the early universe. The fact that only a few general features control the occurrence of the phenomenon leads to the famous universality properties associated with critical phenomena. Even at the absolute zero of temperature, when thermal fluctuations are absent, there can be spontaneous breakdown of symmetry in certain systems – quantum fluctuations alone cannot restore the symmetry in such cases. This is what happens in some of the so-called quantum phase transitions that are of much current interest in condensed matter physics.

**Suggested Reading**


States of Physical Systems in Classical and Quantum Mechanics*

K S Mallesh, S Chaturvedi, R Simon and N Mukunda

We discuss the descriptions of states of physical systems in classical and quantum mechanics. We show that while it is possible to evolve a terminology common to both, the differences in the underlying mathematical structures lead to significant points of departure between the two descriptions both at mathematical and conceptual levels. We analyse the state spaces associated with physical systems described by two and three dimensional complex Hilbert spaces in considerable detail to illustrate how subtle this question can in general be. We highlight the role the Bargmann invariants play in the passage from the Hilbert space to the ray space, the space of states in quantum mechanics, and also in the context of Wigner’s theorem on symmetries in quantum mechanics where they originally appeared.

1. Introduction

The basic mathematical structures encountered in classical and in quantum mechanics are very different from one another, so it is probably surprising that analogies appear at certain points of comparison. In both there is a separation between kinematical and dynamical aspects. The kinematical part in the classical case is rather simple. In quantum mechanics, on the other hand, it is extremely subtle and occupies a very substantial part of the whole theory. Indeed, aspects of quantum kinematics (of composite systems) are of intense current research interest. The discussion in this article will be predominantly kinematical, and restricted to ‘single systems’ as against composite systems.

2. States of Classical Systems

We describe a classical mechanical system in the Hamiltonian formalism. The basic mathematical object is the phase space $\Gamma$ of even dimension, say $2n$. A canonical coordinate system consists of $n$ generalised coordinates $\mathbf{q} = (q_1, q_2, \cdots, q_n)$ and their conjugate momenta $\mathbf{p} = (p_1, p_2, \cdots, p_n)$. A general dynamical variable representing some physical property is a real-valued function $A(\mathbf{q}, \mathbf{p})$ on $\Gamma$. The Hamiltonian $H(\mathbf{q}, \mathbf{p})$ determines the dynamics via the canonical equations of motion

$$\dot{q}_j = \frac{\partial H}{\partial p_j} = \{q_j, H\},$$

\[ \dot{p}_j = -\frac{\partial H}{\partial q_j} = \{p_j, H\}, \]  

(1)

where \( \{A, B\} \) is the familiar Poisson Bracket of \( A \) with \( B \):

\[ \{A, B\} = \sum_k \left\{ \frac{\partial A}{\partial q_k} \frac{\partial B}{\partial p_k} - \frac{\partial A}{\partial p_k} \frac{\partial B}{\partial q_k} \right\}. \]  

(2)

In this framework a *pure state* is represented by a single point \((q_0, p_0) \in \Gamma\). Every dynamical variable \( A \) has, in this pure state, a definite numerical value \( A(q_0, p_0) \) – no spread, no dispersion, and no need or role for probabilities at the level of pure states. Such states are cases of maximum possible information.

In contrast, a general or *mixed state* brings in probabilities. Such a state is defined completely by a probability density \( \rho \) on the phase space \( \Gamma \):

\[ \rho(q, p) \text{ real, } \geq 0; \]

\[ \int d^nq \, d^n p \, \rho(q, p) = 1. \]  

(3)

The meaning is that unlike in a pure state, the \( q \)'s and \( p \)'s do not now have definite values. Rather, different ranges of values are likely to occur with corresponding probabilities:

Probability that system ‘lies in’ small phase space region

\[ d^nq \, d^n p \text{ around } (q, p) = \rho(q, p) d^nq \, d^n p. \]  

(4)

A general dynamical variable \( A(q, p) \) has an expectation value \( \langle A \rangle \) and a spread or dispersion \( \langle (\Delta A)^2 \rangle \) computed as follows:

\[ \langle A \rangle = \int d^n q \, d^n p \, \rho(q, p) A(q, p), \]

\[ \langle (\Delta A)^2 \rangle = \langle (A(q, p) - \langle A \rangle)^2 \rangle \]

\[ = \langle A(q, p)^2 \rangle - \langle A \rangle^2 \geq 0. \]  

(5)

Using the Dirac delta function one can describe a pure state too in terms of a probability distribution,

\[ \rho_{q_0, p_0}(q, p) = \delta^n(q - q_0) \delta^n(p - p_0), \]  

(6)

which is an elementary statement. A mixed state can be expressed in an essentially *unique manner* as a *convex sum or mixture* of pure states,

\[ \rho(q, p) = \int d^n q_0 \, d^n p_0 \rho(q_0, p_0) \rho_{q_0, p_0}(q, p), \]  

(7)
which is again essentially trivial. It can also be described as a mixture of mixed states; clearly, such an ensemble would not be unique.

The set of all (mixed and pure) states is a convex set: if $\rho_j(q, p)$ for $j = 1, 2, \cdots$ are the distributions for any states, and \{P_j\} is any classical system of probabilities, then

$$\rho(q, p) = \sum_j P_j \rho_j(q, p), P_j > 0, \sum_j P_j = 1, \quad (8)$$

is a possible (mixed) state. Only pure states cannot be so decomposed in a nontrivial manner as convex sum of other states, so they are the extremal points (or extremal elements) in the space of all states.

A word about dynamics: the equation of motion for a general state of a Hamiltonian system is the Liouville equation

$$\frac{\partial}{\partial t} \rho(q, p; t) + \{\rho(q, p; t), H(q, p; t)\} = 0, \quad (9)$$

in which the Poisson Bracket of $\rho(q, p; t)$ with the Hamiltonian appears. The important difference in sign compared to (1) is to be noted. For a non-Hamiltonian system like an open system more general Liouville equations will be needed, though.

3. States of Quantum Systems

Given a quantum mechanical system, the basic mathematical object now is a Hilbert space $\mathcal{H}$, a complex linear vector space with hermitian nonnegative inner product. The dimension $N$ of $\mathcal{H}$ may be finite or infinite. Vectors in $\mathcal{H}$ are denoted by $\psi, \phi, \cdots, |\psi\rangle, |\phi\rangle, \cdots$; they are used to describe states of the system. Dynamical variables – physical properties of the system – are represented by hermitian operators $\hat{A}, \hat{B}, \cdots$ acting on $\mathcal{H}$. The structure, mathematical properties, and physical interpretation of states are all quite intricate, much more so than in the classical case.

3.1 Pure States

Each nonzero vector $|\psi\rangle \in \mathcal{H}$, assumed to be normalized, determines a corresponding pure state, a state with maximum possible information. But already for such states, probabilities of a quantum mechanical nature play an essential role:

$$\text{QM probability} \sim |\text{Complex QM probability amplitude}|^2. \quad (10)$$

Let $\hat{A}$ be some dynamical variable, with real eigenvalues $\{a_j\}$ and corresponding orthonormalized eigenvectors $\{|\phi_j\rangle\}$. Assume for simplicity that the eigenvalues are discrete and nondegenerate. Then we have the relations
\[ \hat{A} |\phi_j \rangle = a_j |\phi_j \rangle, \ a_j \text{ real}; \]
\[ \{ |\phi_j \rangle \} = \text{an orthonormal basis (ONB) for } \mathcal{H}, \]
\[ \langle \phi_j | \phi_k \rangle = \delta_{jk}, \ \sum_k |\phi_k \rangle \langle \phi_k | = \mathbb{I}; \]
\[ \hat{A} = \sum_j a_j |\phi_j \rangle \langle \phi_j | \text{ (spectral resolution)}. \quad (11) \]

All these are elementary consequences of hermiticity of \( \hat{A} \).

If a measurement of \( \hat{A} \) is carried out by an appropriate experimental arrangement the result will always be one of the eigenvalues \( a_j \), each with some probability. In a self-evident notation we have, the initial pure state being the one determined by a normalized \( |\psi \rangle \):
\[ \Pr (\hat{A} \rightarrow a_j; |\psi \rangle) = |\langle \phi_j | \psi \rangle|^2 = \langle \phi_j | \psi \rangle \langle \psi | \phi_j \rangle. \quad (12) \]

Here \( \Pr (\hat{A} \rightarrow a_j; |\psi \rangle) \) stands for the probability that an experiment set-up to measure the dynamical variable \( \hat{A} \) will return the value \( a_j \), given that the system is in state \( |\psi \rangle \). These are irreducible QM probabilities, not caused by inadequate knowledge of the condition of the system. So even in pure states we have, in general, expectation values and spreads or dispersions or uncertainties:
\[ \langle \hat{A} \rangle_{\psi} = \sum_j a_j \Pr (\hat{A} \rightarrow a_j; |\psi \rangle) = \langle |\psi \rangle |\hat{A} |\psi \rangle \]
\[ = \operatorname{Tr} (\hat{A} \hat{\rho}(\psi)), \ \hat{\rho}(\psi) = |\psi \rangle \langle \psi |; \]
\[ \langle (\Delta A)^2 \rangle_{\psi} = \langle (\hat{A} - \langle \hat{A} \rangle_{\psi})^2 \rangle_{\psi} = \langle \hat{A}^2 \rangle_{\psi} - \langle \hat{A} \rangle_{\psi}^2 \geq 0. \quad (13) \]

Here we have introduced the projection operator \( \hat{\rho}(\psi) \) on to the vector \( |\psi \rangle \), called the density operator (or density matrix) associated with the pure state \( |\psi \rangle \). So in general any \( \hat{A} \) has a spread \( \Delta A \) unless \( |\psi \rangle \) happens to be an eigenstate of \( \hat{A} \). Omitting the argument \( \psi \), pure state density matrices are fully characterized thus:
\[ \hat{\rho}^\dagger = \hat{\rho} \geq 0, \ \operatorname{Tr} \hat{\rho} = 1, \ \hat{\rho}^2 = \hat{\rho}. \quad (14) \]

Any such \( \hat{\rho} \) is \( \hat{\rho}(\psi) \) for some (normalized) \( |\psi \rangle \in \mathcal{H} \).

3.2 The Superposition Principle of QM

We have seen in (12) that even in a state of maximum possible information, probabilities of a QM nature come in for the description of measurement. As if in compensation for this, and in fact leading to a great simplification, we have a Principle of Superposition of states. We are dealing here of course with pure states. Quoting from Dirac’s discussion of this key principle of QM [1]:
“It requires us to assume that between these states there exist peculiar relationships such that whenever the system is definitely in one state we can consider it as being partly in each of two or more other states. The original state must be regarded as the result of a kind of superposition of the two or more new states, in a way that cannot be conceived on classical ideas. ··· any two or more states may be superposed to give a new state. ··· The intermediate character of the state formed by superposition (thus) expresses itself through the probability of a particular result for an observation being intermediate between the corresponding probabilities for the original states, not through the result itself being intermediate between the corresponding results for the original states.”

At the level of vectors in \( \mathcal{H} \): if \( \psi_1, \psi_2, \cdots \) are (nonzero, distinct) vectors in \( \mathcal{H} \), each determining after normalisation a corresponding pure state, and if \( c_1, c_2, \cdots \) are complex numbers, then

\[
\psi = c_1 \psi_1 + c_2 \psi_2 + \cdots \in \mathcal{H}
\]

is or determines another possible pure state. Note that the \( \psi_j \)'s need not be mutually orthogonal; we also assume there are no superselection rules which would limit the allowed superpositions. In general, given the linear vector level relationship (15),

\[
\hat{\rho}(\psi) \neq \text{any expression involving only } \hat{\rho}(\psi_1), \hat{\rho}(\psi_2), \cdots.
\]

There are very subtle features involved here, which we study in some depth later.

### 3.3 General Mixed States, Density Matrices

These arise from pure states by the same classical statistical mixing as in Section 2. Let \( \psi_1, \psi_2, \cdots \) be distinct normalized, but not necessarily mutually orthogonal, vectors; and let \( P_1, P_2, \cdots \) be any sequence of classical probabilities. Then the (most) general mixed state of the quantum system is given by

\[
\hat{\rho} = P_1 \hat{\rho}(\psi_1) + P_2 \hat{\rho}(\psi_2) + \cdots = \sum_j P_j \hat{\rho}(\psi_j) = \sum_j P_j |\psi_j\rangle\langle\psi_j|.
\]

The complete defining properties of such density matrices are:

\[
\hat{\rho} = \hat{\rho}^\dagger \geq 0, \quad \text{Tr}\hat{\rho} = 1,
\]

\[
\hat{\rho} - \hat{\rho}^2 \geq 0 : \begin{cases} = 0 & \text{pure case}, \\ \neq 0 & \text{mixed or impure case.} \end{cases}
\]
In such a state, for a dynamical variable $\hat{A}$:

$$\text{Pr} (\hat{A} \to a_j; \hat{\rho}) = \langle \phi_j | \hat{\rho} | \psi_k \rangle$$

$$= \sum_k P_k \langle \phi_j | \hat{\rho} | \psi_k \rangle = \sum_k P_k |\langle \phi_j | \psi_k \rangle|^2. \quad (19)$$

In each term in the last expression here, the first factor is a classical probability, the second factor a quantum mechanical one. For the expectation value we have

$$\langle \hat{A} \rangle = \sum_j a_j \text{Pr} (\hat{A} \to a_j; \hat{\rho})$$

$$= \sum_k P_k \langle \hat{A} \rangle_{\hat{\rho} | \psi_k} = \text{Tr} (\hat{\rho} \hat{A}), \quad (20)$$

and similarly for $\langle (\Delta A)^2 \rangle_{\hat{\rho}}$.

In the expression (17), the $P_j$’s and $\psi_j$’s are generally not the eigenvalues and eigenvectors of $\hat{\rho}$. Of course $\hat{\rho}$ does have its own spectrum of eigenvalues and associated eigenvectors, so they do lead to a corresponding representation (the spectral representation) like (17) for $\hat{\rho}$. Unlike the classical equation (7), however, a mixed state $\hat{\rho}$ has infinitely many decompositions of the type (17) in terms of pure states, one among which is given by its own eigenvalues and eigenvectors. This enormous multiplicity means that while the probabilities (19) have both classical and quantum components, there is no unique separation into these components. It is like saying that there is a boundary, but it is movable.

For any general (pure or mixed) states $\hat{\rho}_1, \hat{\rho}_2, \cdots$ and any classical probabilities $P'_1, P'_2, \cdots$, the convexity law says, as in the classical situation, that

$$\hat{\rho} = P'_1 \hat{\rho}_1 + P'_2 \hat{\rho}_2 + \cdots = \sum_j P'_j \hat{\rho}_j \quad (21)$$

is another possible mixed state. The great contrast between this and superposition (15) at the vector level is to be emphasised. So the picture of states in QM is:

Pure states $\hat{\rho}(\psi) = |\psi\rangle \langle \psi| \longrightarrow$

General mixed states

$$\hat{\rho} = \sum_j P_j \hat{\rho}(\psi_j), \text{highly nonunique} \longrightarrow$$

Convexity law

$$\hat{\rho} = \sum_j P'_j \hat{\rho}_j, \text{mixed state for any states } \hat{\rho}_j. \quad (22)$$
For a given closed system described by a Hamiltonian operator $\hat{H}(t)$, the equation of motion is the quantum Liouville equation
\[
\frac{d\hat{\rho}(t)}{dt} = \frac{i}{\hbar} [\hat{\rho}(t), \hat{H}(t)].
\] (23)

As with the classical case, an evolution equation of this kind holds only for Hamiltonian systems; open systems are described by Liouville equations of a more general kind, usually called ‘master equations’.

4. Two-Level and Three-Level Systems

Let us illustrate all this first in the case of $\dim \mathcal{H} = N = 2$. In this case a vector $\psi$ is a two-component column vector
\[
\psi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad \langle \psi | \psi \rangle = |\alpha|^2 + |\beta|^2.
\] (24)

Any operator on $\mathcal{H}$ can be expressed as a unique linear combination of the unit matrix $I_2$ and the three Pauli matrices $\sigma$:
\[
\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\] (25)

The $\sigma_i$’s are hermitian $\sigma_i^\dagger = \sigma_i$ and obey the commutation and anticommutation relations:
\[
[\sigma_r, \sigma_s] = 2i\epsilon_{rst}\sigma_t, \\
\{\sigma_r, \sigma_s\} = 2\delta_{rs}; \quad r, s, t = 1, 2, 3,
\] (26)

where $\epsilon_{rst}$ is the familiar Levi–Civita symbol.

In particular, for the density operator $\hat{\rho}$, we have
\[
\hat{\rho} = a_0 I_2 + \mathbf{a} \cdot \mathbf{\sigma} ; \\
\hat{\rho}^\dagger = \hat{\rho} \Rightarrow a_0, \mathbf{a} \text{ real} ; \\
\text{Tr } \hat{\rho} = 1 \Rightarrow a_0 = \frac{1}{2} ; \\
\hat{\rho} \geq 0 \Rightarrow \text{eigenvalues } a_0 \pm |\mathbf{a}| \geq 0 \Rightarrow |\mathbf{a}| \leq \frac{1}{2}.
\] (27)

Therefore in general
\[
\hat{\rho} = \frac{1}{2} (I_2 + \mathbf{n} \cdot \mathbf{\sigma}) , |\mathbf{n}| \leq 1.
\] (28)

Thus each mixed (or pure) state corresponds uniquely to one point $\mathbf{n}$ inside (or on) the unit sphere $S^2$ in a fictitious three-dimensional Euclidean space:
Pure case: \[ \hat{\rho} = \hat{\rho}^2 \Rightarrow |\hat{n}| = 1, \]
\[ \hat{\rho} = \frac{1}{2} (\mathbb{I}_2 + \hat{n} \cdot \hat{\sigma}), \hat{n} \in \mathbb{S}^2; \quad (29a) \]

Mixed case: \[ \hat{\rho} - \hat{\rho}^2 > 0 \Rightarrow |\hat{n}| < 1, \]
\[ \hat{\rho} = \frac{1}{2} (\mathbb{I}_2 + \hat{n} \cdot \hat{\sigma}), \hat{n} \in \text{Interior of } \mathbb{S}^2. \quad (29b) \]

The sphere \( \mathbb{S}^2 \) is the Poincaré or Bloch sphere. As examples of (29a) we have:

\[ \psi = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \hat{n} = (0, 0, 1), \text{ North Pole}; \]
\[ \psi = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \hat{n} = (0, 0, -1), \text{ South Pole}; \]
\[ \psi = \begin{pmatrix} \cos \theta/2 \\ e^{i\phi} \sin \theta/2 \end{pmatrix}; \hat{n} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \in \mathbb{S}^2. \quad (30) \]

The state corresponding to \( \hat{n} = 0 \), i.e., the one that sits at the centre of the Bloch sphere corresponds to
\[ \hat{\rho}_0 = \frac{1}{2} \mathbb{I}_2, \quad (31) \]

the maximally mixed state – the state corresponding to maximal ignorance about the system.

For any two states \( \hat{\rho}(\hat{n}), \hat{\rho}(\hat{n}') \) we have:
\[ \text{Tr}(\hat{\rho}(\hat{n})\hat{\rho}(\hat{n}')) = \frac{1}{2} (1 + \hat{n} \cdot \hat{n}'). \quad (32) \]

Clearly the RHS can vanish only when both \( \hat{\rho}(\hat{n}), \hat{\rho}(\hat{n}') \) describe pure states and in that case one finds that two mutually orthogonal pure states correspond to diametrically opposite – or antipodal – points on \( \mathbb{S}^2 \).

The nonuniqueness of the expansion (17) for a given mixed state is very easy to illustrate. The most dramatic case corresponds to \( \hat{\rho}_0 = \frac{1}{2} \mathbb{I}_2 \), for then:
\[ \hat{\rho}_0 = \frac{1}{2} \hat{\rho}(\hat{n}) + \frac{1}{2} \hat{\rho}(-\hat{n}), \text{ any } \hat{n} \in \mathbb{S}^2. \quad (33) \]
States of Physical Systems in Classical and Quantum Mechanics

Since a density operator is completely specified by the Bloch vector \( \mathbf{n} \), one may introduce a notion of the ‘distance’ \( d(\hat{\rho}(\mathbf{n}), \hat{\rho}(\mathbf{n}')) \) between two density operators as:

\[
d^2(\hat{\rho}(\mathbf{n}), \hat{\rho}(\mathbf{n}')) = \frac{1}{2} \text{Tr} \left( (\hat{\rho}(\mathbf{n}) - \hat{\rho}(\mathbf{n}'))^2 \right) \tag{34}
\]

\[
= \frac{1}{4} (\mathbf{n} - \mathbf{n}')^2 \tag{35}
\]

which directly relates to the Euclidean distance between the corresponding Bloch vectors. In particular, the distance between \( \hat{\rho}(\mathbf{n}) \) and \( \hat{\rho}_0 \), the ‘origin’ in the set of density operators is given by

\[
d^2(\hat{\rho}(\mathbf{n}), \hat{\rho}_0) = \frac{1}{4} \mathbf{n}^2 \tag{36}
\]

and can be expressed entirely in terms of the eigenvalues \( \lambda_r \) of \( \hat{\rho}(\mathbf{n}) \) as

\[
d^2(\hat{\rho}(\mathbf{n}), \hat{\rho}_0) = \frac{1}{2} \left( \sum_{r=1}^{2} \lambda_r^2 - \frac{1}{2} \right) \tag{37}
\]

These results for \( N = 2 \) have been known for more than a century, first in the realm of classical polarization optics and later in quantum mechanics. An explicit treatment of the case \( N = 3 \) has been developed much more recently, and we describe it at this point, both to show that it is not very complicated and to show what new features are involved.

In the case of \( N = 3 \), \( \mathcal{H} \) is a 3-dimensional complex space with a hermitian inner product and every pure state \( \psi \in \mathcal{H} \) can be written as

\[
\psi = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \quad \langle \psi | \psi \rangle = |\alpha|^2 + |\beta|^2 + |\gamma|^2. \tag{38}
\]

Further, the roles of SU(2) and the Pauli matrices are respectively played by the group SU(3) defined by

\[
SU(3) = \left\{ A = 3 \times 3 \text{ complex matrix} \left| A^\dagger A = 1, \det A = 1 \right\} \right.; \tag{39}
\]

and its eight Hermitian and trace orthogonal generators \( \lambda_i \):

\[
\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\]
The λ’s obey the characteristic commutation and anticommutation relations:

\[ [\lambda_r, \lambda_s] = 2if_{rst}\lambda_t, \quad \{\lambda_r, \lambda_s\} = \frac{4}{3}\delta_{rs} + 2d_{rst}\lambda_t, \]

where \( f_{ijk} \) are completely antisymmetric and \( d_{ijk} \) are completely symmetric. The independent and nonvanishing \( f \) and \( d \) symbols are:

\[
\begin{align*}
 f_{123} & = 1, f_{458} = f_{678} = \frac{\sqrt{3}}{2}, \\
 f_{147} & = f_{246} = f_{257} = f_{345} = f_{516} = f_{637} = \frac{1}{2}; \\
 d_{118} & = d_{228} = d_{338} = -d_{888} = \frac{1}{\sqrt{3}}, \\
 d_{448} & = d_{558} = d_{668} = d_{778} = -\frac{1}{2\sqrt{3}}, \\
 d_{146} & = d_{157} = -d_{247} = d_{256} = d_{344} = d_{355} = -d_{366} = -d_{377} = \frac{1}{2}.
\end{align*}
\]  

(41)

These symbols allow one to define two products \( a_\lambda b \) and \( a \star b \) among real vectors \( a, b \) in an eight-dimensional real Euclidean space \( \mathbb{R}^8 \):

\[
\begin{align*}
 (a_\lambda b)_r &= f_{rst} a_s b_t, \quad a_\lambda b = -b_\lambda a; \\
 (a \star b)_r &= \sqrt{3}d_{rst} a_s b_t, \quad a \star b = b \star a.
\end{align*}
\]  

(42)

The eight \( \lambda_r \)'s together with the unit matrix \( I_3 \) provide a basis in the space of linear operators on \( \mathcal{H} \) appropriate to a three-level system. Any density matrix for a three-level system, being hermitian and having unit trace, can therefore be uniquely expressed as a real linear combination of these as

\[
\hat{\rho} = \frac{1}{3} \left( I_3 + \sqrt{3}c \hat{n} \cdot \lambda \right),
\]  

(43)

where \( c \) is a scalar and \( \hat{n} \) is a unit vector in \( \mathbb{R}^8 \).
States of Physical Systems in Classical and Quantum Mechanics

It is easy to see that if $\rho$ is pure, i.e., $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle \in \mathcal{H}$, then,

$$\hat{\rho}^\dagger = \hat{\rho}^2 = \hat{\rho} \geq 0 \iff c = 1 \quad \text{and} \quad \hat{n} \cdot \hat{n} \star \hat{n} = 1,$$

i.e., $\hat{n} \star \hat{n} = \hat{n}$.

(44)

Every pure state for a three-level system thus corresponds in a one-to-one manner to a unit vector, $\hat{n} \in S^7$, the unit seven-sphere in $\mathbb{R}^8$, which in addition obeys $\hat{n} \star \hat{n} = \hat{n}$. The set $O$ of all such $\hat{n} \in \mathbb{R}^8$ is a 4-parameter region in $S^7$. This $O$ is the analogue of the Poincaré sphere for three-level systems.

It may also be worth mentioning here that if $\hat{\rho}$ and $\hat{\rho}'$ are two pure states and $\hat{n}, \hat{n}'$ are their corresponding 8-dimensional vectors, then,

$$\hat{\rho} = |\psi\rangle\langle\psi|, \quad \hat{\rho}' = |\psi'\rangle\langle\psi'| \quad \Rightarrow \quad \text{Tr}(\hat{\rho}'\hat{\rho}) = |\langle\psi'|\psi\rangle|^2$$

$$= \frac{1}{3}(1 + 2\hat{n}'\cdot\hat{n}),$$

$$0 \leq \text{Tr}(\hat{\rho}'\hat{\rho}) \leq 1 \iff 0 \leq \cos^{-1}(\hat{n}'\cdot\hat{n}) \leq \frac{2\pi}{3},$$

(45)

and so mutually orthogonal vectors in $\mathcal{H}$ do not correspond to diametrically opposite points on $O$. They correspond to points with a maximum opening angle of $\frac{2\pi}{3}$ radians. In the case of $N = 2$, $\hat{n} \in S^2 \Rightarrow -\hat{n}$ is the vector for the orthogonal state and so $-\hat{n} \in S^2$, while in case $N = 3$, if $\hat{n} \in O$, then $-\hat{n} \notin O$.

Moving to a general density operator, not necessarily describing a pure state, positivity of $\hat{\rho}$ constrains $\hat{n}$ and $c$ as follows [2]:

$$\hat{n} \cdot \hat{n} \star \hat{n} = -\sin 3\phi,$$

$$c \leq \frac{1}{2}\csc(\phi + \pi/3), \quad \phi \in [\pi/6, \pi/2].$$

(46)

(47)

The subset characterised by

$$c = \frac{1}{2}\csc(\phi + \pi/3),$$

(48)

consists of density operators for which at least one eigenvalue is zero and constitutes the boundary of the full set. The boundary set obviously contains both pure and mixed states. The pure states in the boundary set, i.e, those with only one nonzero eigenvalue, correspond to

$$\phi = \frac{\pi}{2}$$

(49)

in which case we have

$$c = 1, \quad \hat{n} \cdot \hat{n} \star \hat{n} = 1, \quad \text{i.e.,} \quad \hat{n} \star \hat{n} = \hat{n} \quad (50)$$

as noted earlier.
The other extreme value of $\phi$ viz $\pi/6$ corresponds to those density operators which have two equal nonzero eigenvalues and in this case we have

$$c = \frac{1}{2}, \quad \hat{n} \cdot \hat{n} \star \hat{n} = -1, \text{ i.e., } \hat{n} \star \hat{n} = -\hat{n}. \quad (51)$$

Again, as before, if one defines the ‘distance’ $d(\hat{\rho}(n), \hat{\rho}(n'))$ between two density operators as

$$d^2(\hat{\rho}(n), \hat{\rho}(n')) = \frac{1}{2} \text{Tr}\left((\hat{\rho}(n) - \hat{\rho}(n'))^2\right),$$

then, one finds that distance between $\hat{\rho}(n)$ and $\hat{\rho}_0 = I_3/3$, the ‘origin’ in the set of density operators is given by

$$d(\hat{\rho}(n), \hat{\rho}_0) = \frac{c}{\sqrt{3}} \quad (53)$$

and further that the states in the boundary set that are farthest from $\hat{\rho}_0 = I_3/3$ are the pure states, and those on the boundary set closest to $\hat{\rho}_0 = I_3/3$ are the ones that have only two nonzero and equal eigenvalues. The values of the distances in the two cases are $1/\sqrt{3}$ and $1/2 \sqrt{3}$ corresponding to $c = 1$ and $c = 1/2$ respectively.

One can extend the study of other concepts too for three-level systems and see several similarities as well as differences with the corresponding ones for two-level systems. As mentioned earlier, such an attempt has been done earlier while trying to obtain the Pancharatnam phase formula for three-level systems [3], [4]. For extensions of the considerations above to arbitrary $N$ and also for other notions of ‘distance’ between density operators, the reader may consult ref [5].

5. Further Features of Pure States

It is on account of the normalisation of total probability to unity that we use unit vectors $\psi \in \mathcal{H}$ to describe pure quantum states. But this is still a many to one relation since, as (13) shows, $\psi$ and $e^{i\alpha} \psi$ for any $\alpha$ always determine the same $\hat{\rho}(\psi)$. Let us first define

$$\mathcal{B} = \{ \psi \in \mathcal{H} \mid \langle \psi | \psi \rangle = 1 \} = \text{unit sphere } S^{2N-1} \subset \mathcal{H},$$

so we can say to begin with:

$$\psi \in \mathcal{B} \overset{\text{many-to-one}}{\longrightarrow} \text{definite pure state } \hat{\rho}(\psi). \quad (55)$$

Now if we wish to convert this to a one-to-one relation, it is useful to define rays and a ray space $\mathcal{R}$:

$$\psi \in \mathcal{B} \to \text{ray determined by } \psi : \text{ equivalence class}$$
States of Physical Systems in Classical and Quantum Mechanics

of unit vectors related by phases
\[ = \text{collection } \{ e^{i\alpha} \psi \mid \psi \text{ fixed}, \ 0 \leq \alpha < 2\pi \}. \] (56)

So each ray consists of a circle or $S^1$ worth of unit vectors.

Thus each $\psi \in B$ defines a definite ray to which it belongs, and we can say:

\[ R \equiv \text{set of all rays} \quad \overset{\text{one-to-one}}{\longrightarrow} \quad \text{set of all pure quantum states}. \] (57)

We thus have the sequence

\[ \mathcal{H} = \text{linear space of all vectors } \psi \longrightarrow \mathcal{B} = \text{unit sphere in } \mathcal{H} \longrightarrow \mathcal{R} = \{ \hat{\rho} = |\psi\rangle\langle\psi| \mid \psi \in B \}. \] (58)

Neither $B$ nor $R$ is a linear vector space; from $B$ to $R$ there is a well-defined projection map $\pi$:

\[ \pi : B \rightarrow R : \psi \in B \rightarrow \pi(\psi) = \hat{\rho}(\psi) = |\psi\rangle\langle\psi| \in R. \] (59)

These objects and relationships can be depicted (possibly inadequately) as in Figure 1.

For $N = 2$, the ray space $R$ is the Poincaré sphere $S^2$.

Now we introduce Bargmann Invariants: these are simple expressions involving inner products of $n$ vectors in $B$, but actually defined on $R$. With a single vector, $n = 1$, we have nothing interesting as $\langle\psi|\psi\rangle = 1$. Next, for the case of two vectors, we have the two-vertex

Figure 1. Picture of ray space, rays, unit vectors in $\mathcal{H}$, projection $\pi$. 

# 191
K S Mallesh, S Chaturvedi, R Simon and N Mukunda

Bargmann invariant:

\[ n = 2 : \quad \psi_1, \psi_2 \in \mathcal{B} : \quad \Delta_2(\psi_1, \psi_2) = (\psi_1, \psi_2)(\psi_2, \psi_1) = |\langle \psi_1 | \psi_2 \rangle|^2 = \text{Tr}(\hat{\rho}_1 \hat{\rho}_2) = \text{a quantum mechanical probability, real nonnegative}. \]  \hspace{1cm} (60)

Referring to Figure 1, we see that \( \Delta_2(\psi_1, \psi_2) \) is unchanged under \( \psi_1 \to e^{ia_1} \psi_1, \psi_2 \to e^{ia_2} \psi_2 \), i.e., under sliding \( \psi_1 \) and \( \psi_2 \) on their respective rays, always projecting onto \( \hat{\rho}_1, \hat{\rho}_2 \) in \( \mathcal{R} \). For three vectors we get something more interesting:

\[ n = 3 : \quad \psi_1, \psi_2, \psi_3 \in \mathcal{B} : \quad \Delta_3(\psi_1, \psi_2, \psi_3) = (\psi_1, \psi_2)(\psi_2, \psi_3)(\psi_3, \psi_1) = \text{Tr}(\hat{\rho}_1 \hat{\rho}_2 \hat{\rho}_3). \]  \hspace{1cm} (61)

This three-vertex Bargmann invariant is cyclically symmetric, invariant under independent phase changes and, most importantly, complex in general. For higher orders one has:

\[ n \geq 4 : \Delta_n(\psi_1, \psi_2, \cdots, \psi_n) = (\psi_1, \psi_2)(\psi_2, \psi_3)\cdots(\psi_{n-1}, \psi_n)(\psi_n, \psi_1) = \text{Tr}(\hat{\rho}_1 \hat{\rho}_2 \cdots \hat{\rho}_n). \]  \hspace{1cm} (62)

It is a simple exercise to show that in general \( \Delta_n \) for \( n \geq 4 \) can be reduced to \( \Delta_3 \)'s and \( \Delta_2 \)'s, so the basic Bargmann invariant capturing the presence of complex numbers in quantum mechanics is \( \Delta_3 \).

Now let us go back to Figure 1. Remember that \( \dim \mathcal{H} = N \). Given any \( n \) distinct nonzero vectors \( \psi_1, \psi_2, \cdots, \psi_n \in \mathcal{B} \), they are linearly dependent if for some nontrivial complex coefficients \( c_1, c_2, \cdots, c_n \) we have

\[ c_1 \psi_1 + c_2 \psi_2 + \cdots + c_n \psi_n = 0; \]  \hspace{1cm} (63)

otherwise they are linearly independent. For \( n \leq N \) they may be linearly dependent or independent, while if \( n > N \) they are definitely linearly dependent. In the independent case (so \( n \leq N \)), every nontrivial linear combination appearing on the left in (63) leads, after normalization, to a new pure state – a superposition of \( \psi_1, \psi_2, \cdots, \psi_n \). We ask: how is this seen or reflected at the ray space level?

Start with \( n \) distinct pure states \( \hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_n \in \mathcal{R} \). Choose any vectors \( \psi_1, \psi_2, \cdots, \psi_n \in \mathcal{B} \) in the respective rays, or equivalence classes of unit vectors, so that

\[ \hat{\rho}_1 = \hat{\rho}(\psi_1), \quad \hat{\rho}_2 = \hat{\rho}(\psi_2), \quad \cdots, \quad \hat{\rho}_n = \hat{\rho}(\psi_n). \]  \hspace{1cm} (64)
States of Physical Systems in Classical and Quantum Mechanics

Each of \(\psi_1, \psi_2, \ldots, \psi_n\) is fixed up to a phase. But whether they are linearly dependent or independent is a phase-change-independent or gauge-invariant property! To explore and show this, we form an \(n \times n\) matrix of scalar products

\[
M(\psi_1, \psi_2, \cdots, \psi_n) = \begin{pmatrix} (\psi_j, \psi_k) \end{pmatrix}.
\] (65)

That is, the \(jk\)th entry of the matrix equals the inner product \((\psi_j, \psi_k)\). If we make independent phase changes

\[
\psi_j \rightarrow \psi_j' = e^{i\alpha_j} \psi_j, \quad j = 1, 2, \cdots, n,
\] (66)

then clearly

\[
M(\psi_1', \psi_2', \cdots, \psi_n') = D^\dagger M(\psi_1, \psi_2, \cdots, \psi_n) D,
\]

\[
D = \text{diag} (e^{i\alpha_1}, e^{i\alpha_2}, \cdots, e^{i\alpha_n}).
\] (67)

For any \(n\), the matrix \(M(\psi_1, \psi_2, \cdots, \psi_n)\) is hermitian positive semidefinite. Clearly, the signal that the \(\psi\)'s are linearly dependent is the singularity of \(M(\psi_1, \psi_2, \cdots, \psi_n)\), the vanishing of \(\det M(\psi_1, \psi_2, \cdots, \psi_n)\). But from (67), this is a ray space criterion! We can check this for small values of \(n\):

\[
n = 2 : \quad \det M(\psi_1, \psi_2) = \begin{vmatrix} 1 & (\psi_1, \psi_2) \\ (\psi_2, \psi_1) & 1 \end{vmatrix} = 1 - \text{Tr}(\hat{\rho}_1 \hat{\rho}_2);
\]

\[
n = 3 : \quad \det M(\psi_1, \psi_2, \psi_3)
= \begin{vmatrix} 1 & (\psi_1, \psi_2) & (\psi_1, \psi_3) \\ (\psi_2, \psi_1) & 1 & (\psi_2, \psi_3) \\ (\psi_3, \psi_1) & (\psi_3, \psi_2) & 1 \end{vmatrix}
= 1 - \text{Tr}(\hat{\rho}_2 \hat{\rho}_3) - \text{Tr}(\hat{\rho}_3 \hat{\rho}_1) - \text{Tr}(\hat{\rho}_1 \hat{\rho}_2)
+ \Delta_3(\psi_1, \psi_2, \psi_3) + \Delta_3(\psi_1, \psi_3, \psi_2).
\] (68)
For any \( n \) we can see that

\[
\det M(\psi_1, \psi_2, \cdots, \psi_n) = \epsilon_{k_1 k_2 \cdots k_n} (\psi_1, \psi_{k_1})(\psi_2, \psi_{k_2}) \cdots (\psi_n, \psi_{k_n}) = n! \text{ terms, each a product of some Bargmann invariants (modulo signature) as each } \psi_j \text{ appears once as ket and once as bra in scalar products}
\]

\[
= \text{expressible in terms of } \Delta_m \text{'s, } m = 1, 2, \cdots, n. \tag{69}
\]

We can say, remembering Dirac’s words quoted earlier:

Vectors \( \psi_1, \psi_2, \cdots, \psi_n \) are linearly independent \( \iff \)

\[ \det M(\psi_1, \psi_2, \cdots, \psi_n) > 0 \iff \]

pure states \( \hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_n \) are ‘physically independent’, and necessarily \( n \leq N \); \( \tag{70a} \)

Vectors \( \psi_1, \psi_2, \cdots, \psi_n \) are linearly dependent \( \iff \)

Superposition principle is at work here \( \iff \)

\[ \det M(\psi_1, \psi_2, \cdots, \psi_n) = 0 \iff \]

pure states \( \hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_n \) are ‘physically dependent’, necessarily so if \( n > N \). \( \tag{70b} \)

The important points are that these criteria are entirely expressed via Bargmann invariants which are ray space objects; and in case (70b) there is a ‘physical dependence’ in the sense of the superposition principle even though (in general) none of the \( \hat{\rho} \)’s can be explicitly expressed in terms of the others. All this is evidence of the subtleties involved in the relationships among the quantum mechanical pure states.
6. The Theorem on Symmetries in QM

This theorem is a remarkable fact concerning pure states of quantum systems and is well worth a careful description even in the absence of a proof. It is a result of a kinematical analysis, with no particular Hamiltonian in mind. We have $\mathcal{H}$, $\mathcal{B}$, and $\mathcal{R}$ as indicated in (58). From (60):

$$\hat{\rho}_1, \hat{\rho}_2 \in \mathcal{R} \to \text{Tr}(\hat{\rho}_1 \hat{\rho}_2) =$$

a quantum mechanical probability. \hspace{1cm} (71)

Then, a Wigner Symmetry (WS) is defined to be a one-to-one onto map: $\mathcal{R} \to \mathcal{R}$ which preserves these probabilities, i.e., a pure state map leaving transition probabilities unchanged:

$$\hat{\rho} \in \mathcal{R} \to \Omega(\hat{\rho}) \in \mathcal{R} :$$

$$\text{Tr}(\Omega(\hat{\rho}_1) \Omega(\hat{\rho}_2)) = \text{Tr}(\hat{\rho}_1 \hat{\rho}_2), \text{all } \hat{\rho}_1, \hat{\rho}_2 \in \mathcal{R}. \hspace{1cm} (72)$$

Any unitary transformation $U$ on $\mathcal{H}$ leads via conjugation to a WS:

$$\mathcal{U}(\hat{\rho}) = U\hat{\rho}U^{-1} : \text{Tr}(\mathcal{U}(\hat{\rho}_1) \mathcal{U}(\hat{\rho}_2)) = \text{Tr}(U\hat{\rho}_1 U^{-1} U\hat{\rho}_2 U^{-1}) = \text{Tr}(\hat{\rho}_1 \hat{\rho}_2). \hspace{1cm} (73)$$

It is easy to show that the inverse $\Omega^{-1}$ of a WS $\Omega$ is another WS. So also the composition $\Omega' \circ \Omega$ of two WS’s is another WS.

Wigner’s theorem now says: any WS $\Omega : \mathcal{R} \to \mathcal{R}$ can be ‘lifted’ to the level of vectors in $\mathcal{H}$, acting there either as a linear unitary transformation or as an antilinear antiunitary transformation. In any case, given the action of $\Omega$ on $\mathcal{R}$, we can say:

$$\psi \in \mathcal{B} \to \hat{\rho}(\psi) \in \mathcal{R} \xrightarrow{\Omega} \hat{\rho}' = \Omega(\hat{\rho}(\psi)) = \hat{\rho}(\psi'),$$

$$\psi' \in \mathcal{B} \text{ determined up to a phase.} \hspace{1cm} (74)$$

The theorem claims that $\psi'$ is determined by $\psi$ basically in one of two possible ways, unitarily or antiunitarily. This may be pictorially depicted as in Figure 2.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{wigner_symmetry.png}
\caption{Lifting a Wigner symmetry from $\mathcal{R}$ to $\mathcal{B}$ and $\mathcal{H}$.}
\end{figure}
So the basic result is that $\Omega$ can always be ‘lifted’ to a map $\omega : \mathcal{B} \rightarrow \mathcal{B}$, which can be extended to act $\mathcal{H} \rightarrow \mathcal{H}$, and then $\omega$ is one of the two types mentioned above. Even though this $\omega$ is not unique, its nature is! The Bargmann invariants were originally introduced as a way to see immediately which alternative obtains in a particular case, since $\Delta_3(\psi_1, \psi_2, \psi_3)$ is invariant in the unitary case and simply goes into its complex conjugate in the antiunitary case.

The only physically interesting WS which ‘utilises’ the antiunitary choice is time reversal. A consequence is that in quantum mechanics parity is an observable, its eigenvalues can be measured, but time reversal is not!

7. Concluding Comments

The richness of the space of states of quantum systems is quite remarkable; so many new features appear compared to the classical scene which seems so mundane in comparison. For composite systems some specifically new features appear, such as entanglement, which have profound consequences for information theory and quantum computation. We hope to treat these too in a pedagogical fashion comparable to the above, on another occasion!

Suggested Reading

Reflections
Science and the Human Condition*

N Mukunda

Introduction

Modern science began around the middle of the 17th century, it is just about 350 years old, it is very young. We can express its youth by saying that it spans just about 18 human generations. It is commonly agreed that it was inaugurated by Galileo and Newton. They saw that mathematics was the language of nature, and this language could be exploited most effectively in the description and analysis of nature. Indeed Galileo declared that the book of nature was written in the language of mathematics, and if we do not learn this language, we cannot read a single word of nature. More importantly we can say that while the preceding heroic era of Copernicus and Kepler was still descriptive (and this is no disparagement at all to their genius and achievements), it was only with Galileo and Newton that there was a transition to the understanding and explanation of natural phenomena. In the case of the life sciences, we may say that the corresponding point of transition was the publication of Charles Darwin’s theory of natural selection. Galileo and Newton were concerned mainly with mechanics, astronomy, optics and of course mathematics as well. But they spelt out the agenda and showed the way for science to follow in succeeding centuries. One important and helpful factor in the philosophical background at that time was a liberation in the ways of thinking, a loosening of restrictions. As Max Born described it, “The distinctive quality of these great thinkers was their ability to free themselves from the metaphysical traditions of their time and to express the results of observations and experiments in a new mathematical language regardless of any philosophical preconceptions”.

It is this modern science that I have in mind in the title of this talk. At this point let me give you some idea of the ground I wish to cover, so that I can carry you along with me. At first I will describe the picture of nature achieved through modern science, through assemblies of carefully selected facts. The aim is to convey to you, admittedly rather briefly, what we now know at three levels: the World of the Very Large; the World of the Very Small; and in between, the World of Life. Though my own training is in physics, I will spend a lot of time on the third part; the reasons will emerge later. I want you to get some feeling for the scales of length and time in these different realms, to appreciate the vastness of the macrocosmos and equally of the microcosmos, and then of life. I will then go back a little bit in time and recall some episodes from the history of science, bringing out its human face. We will see some of the philosophical

and social repercussions of science at the end of the 18th century, then go on to progress made in different directions during the 19th. That will then bring us to the revolutionary advances of the present soon-to-end century.

At this point and against this background, I will turn to an examination of ourselves and ‘our condition’. I wish to say something about science as an organised collective human activity; then view ourselves against the canvas of science, and lastly against the vast canvas of nature. It is important to see where we are led if we take the message of science seriously to heart.

The World of the Very Large

*Table 1* gives you a very condensed survey of the World of the Very Large. We begin with the familiar earth-moon-sun system, and work our way steadily outwards. Here we are introduced to ever increasing scales of length and time (All figures in the various Tables are approximate). I have given distances both in terms of meters, roughly our size, and also equivalently in terms of light travel times. As we progress through the solar system, the stars and the Milky Way galaxy, clusters and superclusters of galaxies and finally reach the limits of the observable universe, we pass through some 27 orders of magnitude - the universe is $10^{27}$ times as large as we are, that is its immensity. Light takes about 8 minutes to reach us from the sun, but some 11 hours to go across the solar system. From one star to another one nearby it takes a few years; our nearest star after the sun is Alpha Centauri, about 4 light years away. Our galaxy – the Milky Way – contains some $10^{11}$ stars, and is about 100,000 light years across, a slowly rotating disc. Where in it are we located? About three quarters of the way out from the centre, hurtling along at 250 km per second.

Our galaxy and Andromeda are the two big ones in a local group of 30 galaxies; from one to the other light takes two million years to cross. Galaxies are grouped together into clusters and superclusters and so on to the entire universe. The universe contains about as many galaxies as each galaxy contains stars, so our glorious sun is but one among $10^{22}$ suns! The universe is about 15 billion years old, while the solar system – our address – is about 4.5 billion years old. And it takes light 10 billion years to go across the universe.

Staggering as these magnitudes are, the point to emphasize is that the universe is lawful and understandable all the way. We are able to observe, analyse and theorise about the life histories of stars, the formation of galaxies, and the structure and evolution of the entire universe. Many problems remain at every stage, but we have the key concepts and methods and the power to create new ones, to comprehend it all. The universe is very largely empty space, save for electromagnetic radiation and neutrinos. At the level of stars the major physical forces at play are nuclear, electromagnetic, weak and gravitational forces; beyond that, at the galactic and higher levels, it is gravity all the way that controls everything.

In passing I might mention that the sun, about 4.5 billion years old, has another 6 billion years life left. Then it will expand, become a red giant, reachupto Mars and heat up the earth till lead melts and the oceans boil. By then life as we presently know it will disappear, at least from the solar system. The reassuring thing is that we still have 6 billion years to go.
Table 1. World of the very large.
by = billion years = 10^9 years; 1 ly = 1 light year = 10^{16} m
Age of the Universe ~ 15 by; age of sun and planets ~ 4.5 by ~ (1/3 to 1/4) × age of universe

<table>
<thead>
<tr>
<th>Object/distance</th>
<th>Length in meters</th>
<th>Length in ly/light travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth radius</td>
<td>6.4 × 10^6</td>
<td>0.02 secs</td>
</tr>
<tr>
<td>Earth to moon</td>
<td>3.8 × 10^8</td>
<td>1.3 secs</td>
</tr>
<tr>
<td>Earth to sun</td>
<td>1.5 × 10^{11}</td>
<td>8.3 mins</td>
</tr>
<tr>
<td>Speed of earth around sun</td>
<td>~ 30 km/sec</td>
<td></td>
</tr>
<tr>
<td>Sun radius (typical star)</td>
<td>7 × 10^8</td>
<td>2.3 secs</td>
</tr>
<tr>
<td>Sun to Pluto (solar system size)</td>
<td>6 × 10^{12}</td>
<td>5.6 hrs</td>
</tr>
<tr>
<td>Distance between nearby stars</td>
<td>few ×10^{16}</td>
<td>few ly ~ few × 10^3 × size of solar system</td>
</tr>
<tr>
<td>No. of stars in typical galaxy</td>
<td>~ 10^{11}</td>
<td></td>
</tr>
<tr>
<td>Mass of typical galaxy</td>
<td>~ 10^{11} × M_{sun}</td>
<td></td>
</tr>
<tr>
<td>Diameter of Milky Way</td>
<td>8 × 10^{20} (~ 10^{21})</td>
<td>8 × 10^4 ly (~ 10^5 ly)</td>
</tr>
<tr>
<td>(typical galaxy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thickness of Milky Way</td>
<td>6 × 10^{19}</td>
<td>6 × 10^3 ly</td>
</tr>
<tr>
<td>Diameter of spherical halo</td>
<td>10^{21}</td>
<td>10^5 ly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solar system in Milky Way</td>
<td>~ 3 × 10^4 ly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>from centre, above galactic plane ~ 3/4 of the way out. Speed of solar system around galactic centre ~ 250 km/sec</td>
<td></td>
</tr>
<tr>
<td>Milky Way to Andromeda</td>
<td>2 × 10^{22}</td>
<td>2 × 10^6 ly ~ 20× diameter of galaxy</td>
</tr>
<tr>
<td>Distance between nearby galaxies</td>
<td>4 × 10^{22}</td>
<td>4 × 10^6 ly</td>
</tr>
<tr>
<td>Supercluster ~ 10^5 galaxies; No. of superclusters ~ 10^6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radius of supercluster</td>
<td>10^{24}</td>
<td>10^8 ly ~ 1000× Diameter of galaxy</td>
</tr>
<tr>
<td>Distance between nearby</td>
<td>10^{24}</td>
<td>10^8 ly</td>
</tr>
<tr>
<td>Superclusters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of galaxies in universe</td>
<td>~ 10^{11}</td>
<td></td>
</tr>
<tr>
<td>Radius of universe</td>
<td>10^{26}</td>
<td>10^{10} ly</td>
</tr>
</tbody>
</table>
The World of the Very Small

Now from this World of the Very Large let us swing all the way to the World of the Very Small, in Table 2. As mentioned earlier, the universe is practically matter-free; what matter there exists is made up of protons, neutrons and electrons. These are its normal constituents. The number of distinct chemical elements is about one hundred; and in mass they range from the hydrogen atom at one proton mass to the atoms of lead, uranium and the like at about 250 proton masses. Sizes in this realm are conveniently expressed in terms of the Angstrom, one Å being $10^{-8}$ cm. Molecules made up of a few atoms have sizes in the range of a few Å, may be 2 or 3 up to 6 or 8 Å. Individual atoms range in size from $1/2$ Å for hydrogen to again 2 Å or 3 Å at the heavy end. Chemistry deals with atoms and molecules, and the dominating force there is electromagnetism. Looking now at the interior of the atom, again it is mostly empty space. The nucleus at the heart of the atom is a few fermis in diameter, one fermi being $10^{-5}$ Å = $10^{-13}$ cm. Practically all the atom’s mass is in the nucleus, but the atom is 100,000 times as large as the nucleus! Far from the nucleus, the electrons roam around at the atom’s periphery; they are the principal players in chemistry. Electrons are point-like down to $10^{-16}$ cm i.e. a thousandth of a fermi. Within the nucleus all the forces - strong, electromagnetic and weak – have roles to play; what may be ignored in chemistry downwards is gravity. For all phenomena in the microscopic world and also involving radiation the language we need to use is the language of quantum mechanics. The most accurate theory in science, quantum electrodynamics or QED, belongs here. Beyond the electron, today we speak of quarks and leptons as the fundamental building blocks of matter. And bold theorists dream of phenomena at the Planck scale, lengths of order $10^{-33}$ cm, twenty orders of magnitude smaller than the nucleus.

Table 2. World of the very small

1 Angstrom = 1 Å = $10^{-8}$ cm, 1 fermi = 1 fm = $10^{-13}$ cm.

Universe mostly matter-free, only radiation and neutrinos. Normal matter made up of protons, neutrons, electrons. Number of chemical elements ~ 100

Range of sizes: hydrogen ~ 0.5 Å . . . . . . . . uranium, lead ~ 2 to 3 Å

Range of masses: hydrogen ~ one proton mass . . . . . . . heaviest elements ~ 250 proton masses

Molecules of a few atoms: CO₂, NH₃ ~ 2 to 3 Å . . . . . . . . . . . . . 10 atoms ~ 5 to 8 Å

Chemistry: world of atoms and molecules: main force electromagnetism

Inside the atom proton ~ 1 fm . . . . . . . Heavy nuclei ~ 5 to 6 fm

Nuclear size/atomic size ~ $10^{-5}$

Electron size < $10^{-16}$ cm = $10^{-3}$ fm

In nucleus: strong, electromagnetic, weak forces.

Basic building blocks of matter: quarks and leptons

Microscopic world: language of physics = Quantum mechanics

Planck length ~ $10^{-33}$ cm
All of this again is accessible to human understanding, and new laws of nature couched in the language of quantum mechanics hold sway. The presently accessible physico-chemical world thus spans some 44 orders of magnitude, and it is comprehensible all the way.

The World of Life
Let us now finally look at the World of Life, and our own place in it. Here it becomes necessary to focus successively on several different aspects, each important for our story. To begin with, we must realise that life differs from physics and chemistry in certain ways. The laws of physics and chemistry are the same always and everywhere in the universe, they have a character of universality. But – as far as we know – life has occurred and evolved just once, and here on planet earth. Therefore its study has a historical and teleological character, a subtle illusion of moving towards a goal, in ways which are inappropriate in physics and chemistry. Keeping these in mind, let me first briefly review the evolution of life, sketched in Table 3. As we saw, the earth is some \(4\frac{1}{2}\) billion years old, and it is interesting that life already appeared 700 million years later. In this context, that is quite fast; life has been around for one quarter of the age of the universe itself! Now we go down the different levels of steadily increasing sophistication and complexity of life forms – bacteria appeared 3.1 billion years ago, the cell and its structures ‘soon after’. The fishes are 500 million years old, mammals 200 million years old. And so we move on through primates, monkeys and apes to our own ancestors – the hominids – who appeared on the scene a mere 3 to 5 million years ago.

Table 3. The Evolution of Life
(by = billion years, my = million years)

<table>
<thead>
<tr>
<th>Emergence of</th>
<th>How long ago</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth</td>
<td>4.5 by</td>
</tr>
<tr>
<td>Nucleotides, aminoacids</td>
<td>4.0 by</td>
</tr>
<tr>
<td>Earliest life forms</td>
<td>3.8 by</td>
</tr>
<tr>
<td>Bacteria</td>
<td>3.1 by</td>
</tr>
<tr>
<td>Cell + Structures inside</td>
<td>2 to 3 by</td>
</tr>
<tr>
<td>Protists</td>
<td>1.2 by</td>
</tr>
<tr>
<td>Fish</td>
<td>500 my</td>
</tr>
<tr>
<td>Mammals</td>
<td>200 my</td>
</tr>
<tr>
<td>Primates (origin of monkeys, humanoids)</td>
<td>80 to 90 my</td>
</tr>
<tr>
<td>Monkeys</td>
<td>50 my</td>
</tr>
<tr>
<td>Apes (origin of chimps, gorillas, orangutans, hominids)</td>
<td>35 my</td>
</tr>
<tr>
<td>Hominids</td>
<td>3 to 5 my</td>
</tr>
</tbody>
</table>
Bacteria — 1 by → cell + nucleus — 1 by →
Multicellular organisms, — 1 by → mammals
Nerves, CNS

To keep things in focus let me remark: Evolution from bacteria to the primitive cell took one billion years; from there to multicellular organisms, the development of nerve pathways and a central nervous system (CNS) was another billion years; and the next major step, the appearance of mammals, took again a billion years.

It is interesting to trace the points of divergence in the evolutionary tree leading ultimately to ourselves, starting from mammals 200 million years ago; I give this in Table 4. Mammals themselves evolved out of reptiles. Then the primates took off 70 million years ago. The two groups of monkeys branched off next, one after the other, around 45 million years ago. The next stage was the appearance of the apes some 30 million years ago. Finally from out of these came the hominids just 5 million years ago. And we can date the advent of erect stature, leaving two limbs free to do other things, to about 3 million years ago.

<table>
<thead>
<tr>
<th>Event</th>
<th>How long ago</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reptiles to mammals</td>
<td>200 my</td>
</tr>
<tr>
<td>Arboreal mammals to primates</td>
<td>70 my</td>
</tr>
<tr>
<td>Primates to New World monkeys</td>
<td>&gt; 45 my</td>
</tr>
<tr>
<td>Primates to Old World monkeys</td>
<td>45 my</td>
</tr>
<tr>
<td>Old World monkeys to Apes</td>
<td>30 my</td>
</tr>
<tr>
<td>Chimps, gorillas to hominids</td>
<td>5 my</td>
</tr>
<tr>
<td>Erect stature in hominids</td>
<td>since 3 my</td>
</tr>
</tbody>
</table>

I am giving you a series of snapshots of shorter and shorter time spans. We look next, in Table 5, at just the Hominid line over the past 5 million years, ending with our immediate ancestors the *Homo sapiens sapiens*, indeed, ourselves. This is presented in the form of an (inverted) tree with branching off points. In biological terms the hominidae form a Family and they began some 3 to 5 millions years ago. Already they were upright walkers. They gave rise and place to the Genus Australopithecus (genus is a smaller unit than family) who flourished $1 \frac{1}{2}$ to 3 million years ago. Next, the species *Australopithecus afarensis* (species is a smaller unit than genus) to whom ‘Lucy’ belonged; that was around 3.8 million years ago (You see that these figures are very approximate!). Then there was a divergence: the species *Australopithecus africanus* followed by the species *Australopithecus robustus* on one branch, both now long extinct. The other branch, via *Homo habilis* and then *Homo erectus*, split again into two lines: the species of the Neanderthalers and the species *Homo sapiens*, ‘just’ around 200,000 years back. And no more than 100,000 years ago came our own modern human species, *Homo sapiens sapiens*. The Neanderthals were wiped out by ‘us’ just 30000 to 40000
years ago. Between *Australopithecus* and *Homo sapiens*, in the time span from 3 million years to 200000 years ago, the size of the brain increased by a factor of four. We have probably had the language faculty since almost 200000 years. And with the arrival of *Homo sapiens sapiens* as recently as 100000 years ago we have the emergence of language, art and culture – cultural evolution speeding up and overtaking genetic evolution.

At this point you may wonder - why is he telling us all this? For good reasons – it is fascinating to trace our origins; I later want to speak about ‘the human condition’ against this background; and remember that in many places in the most developed country in the world it is forbidden to teach such things!

Table 5. The Hominid line (*F* = family, *G* = genus, *S* = species)

mya = million years ago, kya = thousand years ago

<table>
<thead>
<tr>
<th>Hominidae (F): 3 to 5 mya</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Australopithecus</em> (G): 1.5 to 3 mya</td>
</tr>
<tr>
<td><em>A. afarensis</em> (S): 3 mya</td>
</tr>
<tr>
<td><em>A. africanus</em> (S): 2.8 mya</td>
</tr>
<tr>
<td><em>A. robustus</em> (S): 1.8 mya</td>
</tr>
<tr>
<td><em>H. habilis</em> (S): 1.5-2 mya</td>
</tr>
<tr>
<td><em>H. erectus</em> (S): 1.6 mya-80 kya</td>
</tr>
<tr>
<td><em>H. sapiens neanderthalensis</em> (S) 1.25kya - 35 kya</td>
</tr>
<tr>
<td><em>H. sapiens</em> (S): 200 kya</td>
</tr>
<tr>
<td><em>H. sapiens sapiens</em> (S): 100 kya</td>
</tr>
</tbody>
</table>

Four fold increase in brain size

We have been talking of vast expanses of time – billions and millions of years – and it is not easy to grasp these enormities. Just as a help to the imagination, I show you in Table 6 a few events since the arrival of *Homo sapiens sapiens* just 100000 years back, in two ways – in terms of number of years elapsed, and in terms of human generations elapsed. This may help to make things more tangible at least at that scale. As a species we are a modest 5000 generations old. Agriculture and civilizations arose 500 generations ago, and the earliest Pharaohs ruled 250 generations ago. They built the Pyramids at Giza some 225 generations in the past. The Miss Universe of those days, Queen Nefertiti, and her nephew King Tut, ruled nearly 170 generations ago. The Buddha and Christ date back respectively 125 and 100 generations; and modern science is a mere 18 human generations old.
Table 6. A few recent events

<table>
<thead>
<tr>
<th>Event</th>
<th>How long ago (years)</th>
<th>No. of generations ago</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Homo sapiens sapiens,</em> Language faculty</td>
<td>100,000</td>
<td>5000</td>
</tr>
<tr>
<td>Agriculture, civilizations</td>
<td>10,000</td>
<td>500</td>
</tr>
<tr>
<td>Early Pharaohs</td>
<td>5,000</td>
<td>250</td>
</tr>
<tr>
<td>Pyramids at Giza</td>
<td>4,500</td>
<td>225</td>
</tr>
<tr>
<td>Queen Nefertiti, King Tut</td>
<td>3,350</td>
<td>168</td>
</tr>
<tr>
<td>Buddha</td>
<td>2,500</td>
<td>125</td>
</tr>
<tr>
<td>Christ</td>
<td>2,000</td>
<td>100</td>
</tr>
<tr>
<td>Modern science (Galileo, Newton)</td>
<td>350</td>
<td>18</td>
</tr>
</tbody>
</table>

Continuing our journey through the World of Life, let us now briefly switch from evolutionary history to the mechanics of life, how it is organized and what makes for its unity. A convenient starting point is the cell (which you recall appeared in primitive form around 3 billion years ago); from here we can either go upwards towards the organism, or downwards into its microscopic constituents. From Table 7 you see that a typical cell is about $10^{-4}$ cm in diameter, and has a mass of about $5 \times 10^{-12}$ kg, which comes to some $10^{15}$ proton masses. An adult human has about $10^{14}$ cells in all; this number can be grasped as being the result of close to 50 cell divisions or cell generations, each division doubling the number of cells. The cell is a most incredibly efficient chemical factory. And some of the most fascinating problems in biology involve development, the route from the single fertilized egg to the adult organism.

Table 7. Cell and upwards

- Size of a typical cell $\sim 10^{-6}$ m = $10^{-4}$ cm
- Mass of a typical cell $\sim 5 \times 10^{-12}$ kg $\sim 10^{15}$ proton masses
- No. of cells in adult human $\sim 10^{14}$ = result of $\sim 50$ cell divisions
- Single fertilized egg $\rightarrow$ adult organism: domain of developmental biology

Table 8. The cell and some contents

DNA = basic genetic material constituting chromosomes
  = double stranded helix, each strand
  = string of nucleotides

Alphabet of DNA has 4 letters: A, T, G, C; pairing rule across strands: A $\leftrightarrow$ T, G $\leftrightarrow$ C
Word of 3 letters (3 nucleotides) $\rightarrow$ code for one amino acid
No. of amino acids = 20
Mass of amino acid $\sim 100$ proton masses
100 to 10000 (average 300) amino acids in sequence = one protein
Mass of protein $\sim 10^4$ to $10^6$ proton masses
Bacterium $\sim$ 3000 distinct proteins
Humans $\sim 10^5$ distinct proteins
Human genome $\sim$ 3.5 million base pairs

Some of the major contents of the cell, the players in the game of life, are shown in Table 8. From the bottom up: DNA - the basic genetic material making up the chromosomes – is a double stranded helix, built on a repertoire of 4 letters. The alphabet of life consists of the 4 nucleotides A, T, G and C as they are called, with the pairing rule in the helix that A and T always face each other, and similarly for G and C. A word of three letters – a string of 3 nucleotides – represents or codes for one amino acid, of which there are 20; and anywhere from 100 to 10000 amino acids (300 on the average) strung together form a protein. A gene is a stretch of DNA coding for one protein, hence on average it is some 1000 nucleotides in length. Thus we have the progression nucleotides $\rightarrow$ genes $\rightarrow$ proteins built up of some 300 amino acids each. In terms of proton masses amino acids have on average the mass of a hundred protons, while a protein has a mass of $10^4$ to $10^6$ protons. While a simple bacterium contains some 3000 distinct proteins, we humans have about 100,000 of them, and our genome – entire gene content – is about $3\frac{1}{2}$ million base pairs or nucleotides long. All this is accompanied by an intricate machinery for replication, translation, transcription, transport, catalysis and manufacture and synthesis and the like, involving many other actors like RNA, etc. All are results of billions of years of slow evolution. And essentially these structures pervade all of life – plants, insects, animals, us… without exception. That is the meaning of the unity of life.

I have shown you glimpses of all the three Worlds – the Very Large, the Very Small, and Life – all the results of $3\frac{1}{2}$ centuries of scientific effort. Now it is time to look a little bit at the history of that effort itself. And to answer a natural worry in your minds: after so much information, where is the knowledge; and further, where is the wisdom? Patience!

The Early Impact of Science

From the foundations laid by Galileo and Newton, right through to the end of the 18th century, physical science scored triumph after triumph in explaining natural phenomena. These were in celestial mechanics or astronomy, in mechanics, in fluid dynamics, and towards the end in electricity and magnetism as well. The approach based on Galilean Newtonian principles was so successful that in the late 1700’s it led to three very interesting claims or developments which I would now like to mention. One was the statement of the mathematican-physicist Pierre Simon Laplace that the entire universe was like a clockwork mechanism – to a supremely intelligent mind with all-seeing eyes and unlimited calculating powers, on the basis of Newton’s equations and knowledge of all the forces in nature, the present would completely determine the future, and for that matter the past as well. This was Laplace’s doctrine of complete determinism. Another was the impact on the thinkers of the Enlightenment – the success of natural science
N Mukunda

led them to claim that human societies too could be perfectly organized and social progress was inevitable. In the words of the Marquis de Condorcet as paraphrased by Edward Wilson – “...culture is governed by laws as exact as those of physics. We need to only understand them ...to keep humanity on its predestined course to a more perfect social order ruled by science and secular philosophy. These laws ... can be adduced from a study of past history”.

Today we can say that both these views were results of being blinded by the successes of science of that time, and of rash extrapolations both within science and outside it. Science is much more modest in its attitudes today – the physical universe is much too complex and rich for us to believe literally in Laplacian determinism; and as for society there are far too many historical and cultural factors and variations to naively believe in the inevitable universal progress of social institutions.

In some ways the most profound of the effects of science on thinking at that time is the third point to which I now turn. The great Immanuel Kant, the philosopher of the Enlightenment, set out to explain why Galilean–Newtonian physics was so successful. At that time there were two contrasting philosophies concerning our knowledge of nature – the rationalist which believed that nature was subordinate to and had to obey reason; and the empiricist which held that all knowledge had to come from experience. Kant tried to combine both in a new way and said that while knowledge is certainly drawn from experience, we already possess at birth certain “synthetic a priori categories of thought” that constrain the behaviour of nature. In other words our understanding of nature is in terms of certain principles which are inborn in us prior to experience of nature. In his view, some of the empirical facts about nature – such as the properties of space and time, the law of cause and effect, simultaneity, the conservation of matter and in the end even some of Newton’s Laws – were not subject to verification or denial by experiment; they had necessarily to be true as they were the basis of all understanding of experience and could never be contradicted by it. Briefly put, we know certain things about the behaviour of nature even before we have experience of nature. So something was explained by saying that it was inevitable, it could never be otherwise. The question that arises for us today is: if we know something about the world even before we see it, what is the source of this knowledge?

I will tell you later about the resolution of this problem. But for the moment let me just remark that the Enlightenment program, especially in its social aspects, failed and led in the early 19th century to a romantic reaction in philosophy, in poetry and in the expression of our relation to nature. For Science however this was a good thing - it was essentially left alone to make progress by itself based on experiment and mathematical analysis. In all fields there was great progress and I can mention only a few highlights.

**Science in the 19th and 20th Centuries**

In physics, based on landmark experiments, the concept of the electromagnetic field was created by Michael Faraday. In Maxwell’s hands, this led later to the unification of electricity, magnetism and optics; light was seen to be an electromagnetic wave. The science of ther-
modynamics grew, while celestial mechanics continued to prosper. Chemistry progressed as a combination of descriptive and quantitative aspects, and the pattern of the elements as systematised by Mendeleev's table was established. In the life sciences the Darwinian theory of evolution by natural selection was propounded, though at that time the material underpinnings of heredity and its variation and transmission were not yet known. The basic principles of heredity were discovered by Gregor Mendel, laid forgotten, and had to be rediscovered in the beginning of this century. At least as far as physics was concerned, the basic view of natural phenomena was a mechanical one – in fact for a long time Maxwell tried to interpret electric and magnetic fields in terms of gears and wheels, not as basic irreducible entities in their own right.

The physics of this century has shown nature to be far more subtle, abstract and rich than was ever previously imagined. This has been so with all the three major advances – special relativity, general relativity and then quantum mechanics. Each of these has deep and characteristic mathematical structures, aesthetic features leading to unparalleled beauty. The role of mathematics in understanding the physical universe has grown beyond any reasonable earlier expectations. Eugene Wigner spoke of “The Unreasonable Effectiveness of Mathematics in the Natural Sciences”; and Paul Dirac wrote of “…some mathematical quality in nature, a quality which the casual observer of nature would not suspect, but which nevertheless plays an important role in nature’s scheme”.

Today we know that it is through quantum mechanics that we reach true understanding and explanation in chemistry. In a sense, chemistry and the Mendeleev table before quantum mechanics are like Kepler's Laws before Newton's Laws of Motion and of universal gravitation. The parallel in the life sciences would be the efforts of Carolus Linnaeus (at least as far as natural history is concerned) before the Darwinian revolution. In turn chemistry and biochemistry have provided the foundations for understanding life processes, and through them we see disclosed the amazing unity of life in all its forms. If we look at physics, chemistry and the life sciences in that sequence, we see their relationships in a new light: they are mutually consistent, each dovetails into the next and provides its foundations, but each also preserves its autonomy too and works with concepts suited to its domain. It is just too hard for us to trace all the phenomena of chemistry back to the starting principles of physics, and similarly from life to chemistry. It is all a fine combination of hierarchical interdependence, consistency and autonomy because what is calculable in principle is not so easy in practice.

Ourselves Viewed Against Science

Time now to view ourselves against science. It is important to realise that science is an organised human activity, a collective process; and its results are part of our cultural heritage. In Heisenberg's words: “Nature is prior to man, and man is prior to natural science”. For all its shortcomings, and remembering that scientists are no angels, it is the one activity that has a built-in self-correcting mechanism and leads to dependable truth, a body of critically tested knowledge we can trust. It grows cumulatively, and while individuals make discoveries and
N Mukunda

contribute to the body of verified scientific knowledge, once discovered it becomes part of an objective whole, in a sense detached from the particular individual. There is no sense of value within scientific knowledge itself, no meaning to good or bad in a moral sense; though we do judge some discoveries as being deeper or more fundamental than others. This was very succinctly expressed by Bertrand Russell: “Science, by itself, cannot supply us with an ethic. It can show us how to achieve a given end, and it may show us that some ends cannot be achieved”. Perhaps most important, while science is a never-ending exploration, is the lesson that nature is understandable, and at no stage is there reason for fear of as yet unanswered questions, no need to look for makeshift pseudo explanations.

We have seen that we are the products of a very long evolutionary process in which over the past 100,000 years the genetic component has been dominated by the cultural component. We come equipped with certain senses, but they are limited as they have been fashioned by evolution to function within a certain biological niche. It is through science that we transcend these limitations and come to know of the great range and variety of natural phenomena inaccessible to our unaided senses. As science advances, previously understood phenomena become instruments for further exploration into nature. In the words of Julian Schwinger: “It is remarkable how nature aids mankind’s groping towards an understanding of the universe. As we raise the level of our scientific skills and sharpen our artificial senses, fascinating new phenomena continue to appear, testing and challenging our growing comprehension of nature’s grand design”. The evolutionary process is also a learning process – slow progress via natural selection retains and perfects those capacities in us that are best able to recognize the aspects of the world which are important for life. Thus the species ‘learns’ as it evolves; but to the individual member of the species those evolved abilities appear inborn, given right at the start at birth and perfectly suited to handle later experience. This is the explanation given by Konrad Lorenz to the origin of the Kantian a priori. Both experience and learning operate at two levels - at the species level as it evolves, and also between birth and death in the life span of each individual. A fascinating ‘play within a play’ situation, an insight based on Darwinian evolution not available in Kant’s time. As Delbrück expressed it, “What is a priori for individuals is a posteriori for the species”.

The brain is the most complex form of organised matter known to us anywhere in the universe. It too has been through a very long evolutionary process. In Table 9, I present some salient features of this past. The evolution of the brain is a 400 million year old saga, passing from fish through frog through reptiles to mammals, and then on to primates to the hominid line. While the human brain reached its present condition some 100,000 years ago when Homo sapiens sapiens appeared, it contains relics of this long evolutionary growth. Many patterns or predispositions for thought and action are of genetic origin, howsoever we may wish otherwise. In a real sense we are on a genetic leash, we are not as free as we think or like to imagine we are. Instinct is much older than reason; and even while we are the only species to deliberately harm and kill one another, the ethical sense and even altruism are much older than religion.
Table 9. Evolution of the brain.

History of brain = 400 million years saga: fish → frog → reptiles → mammals → primates → hominid line
3 million years ago to 200000 years ago: 4-fold increase in brain size, mainly neo cortex → higher functions, language, culture.

Ancient instinct much older than reason
Ethical sense much older than religion

As for the complexity of the human brain, the bare figures in Table 10 speak for themselves.

Table 10. Complexity of human brain

Brain \( \sim 10^{11} \) nerve cells, each \( 10^{-4} \text{cm} \) diameter, each connected to 100–1000 others.

Total number of connections \( \sim 10^{14} \)

Volume \( \sim 1.5 \times 10^3 \text{cc} \), mass \( \sim 1.3\text{kg} \)

Brain structure controlled by at least 3200 genes (out of 100000)

Highly folded cortex \( \sim 75 \text{ cm} \times 75 \text{ cm square sheet} \)

Language faculty 100000 years old; \( \sim 10000 \) languages in all

Capacity to simulate experience, then choose action.

Religious feelings – genetic roots

Preconscious processing – sensory inputs, decisions, illusion of free will

What next about free will and the nature of the mind? Here we learn that a great deal of preconscious processing and activity goes on all the time in the brain; this is particularly well known in the handling of sensory inputs such as vision. What emerges into the conscious realm is the end product of all this prior activity. With free will too it is similar. Patterns of decisions and choices among alternatives are genetically influenced in the form of predispositions, they follow epigenetic rules; and only at a late stage do they surface into consciousness and create the illusory feeling of having consciously made a free choice. But even as an illusion this feel-
ing is important for survival! As for the mind, the evolutionist tells us that it has developed to help with biological survival, not to understand itself; and in Delbrück’s words: “In the context of evolution, the mind of the adult human, the object of so many centuries of philosophical studies, ceases to be a mysterious phenomenon, a thing unto itself. Rather, mind is seen to be an adaptive response to selective pressures, just as is nearly everything else in the living world”.

**Ourselves Viewed Against Nature**

Let me draw to the end and come to the question - how do we view ourselves against nature? In more than one way we see that we are very lonely. We have seen the vastness of nature and - starting with the Copernican revolution – it is difficult indeed to believe that we are central to nature. Feynman said it beautifully: “... I can’t believe these special stories that have been made up about our relationship to the universe at large because they seem to be too simple, too connected, too local, too provincial. The earth, he came to the earth, one of the aspects of God came to the earth, mind you, ... and look at what’s out there. How can He... it isn’t in proportion”.

Then there are yet other ways in which loneliness – as a species and as individuals – bears down upon us. The emergence of life on earth is the result of a chance occurrence; before the event, the probability of its happening was essentially zero. And in Monod’s powerful phrase taken from Democritus, the processes of life combine ‘Chance and Necessity’ in a profound way. The genetic roulette upon which natural selection plays is a roulette of pure blind chance, with no direction or purpose. It is the intensely conservative nature of the replication processes of life - the preservation of genetic structure – which allows for testing competing alternatives in the biological world and weeding out the less adapted. This too speaks of loneliness. And at the level of the individual, there are deep unbridgeable limits to communication with other individuals. We believe we all ‘see’ the same colours, ‘hear’ the same sounds, and think similar thoughts; but never will we ever know that this is so. For in nature itself there is neither ‘colour’ nor ‘sound’. In Erwin Schrödinger’s eloquent words: “…the scientific world view contains of itself no ethical values, no aesthetical values, not a word about our own ultimate scope or destination, and no God, if you please ...Science cannot tell us a word about why music delights us, of why and how an old song can move us to tears”. It is our belief in our common biological heritage that encourages us to think we perceive and react to nature in the same ways, but of this there can be no direct confirmation. In another context R K Narayan wrote: “A profound and unmitigated loneliness is the only truth of life”.

Science has shown to us our location in nature, and also the fact of our loneliness - as a species, as individuals. There are many different ways in which one may react to this. My own is along these lines.

To come to terms with individual loneliness is very difficult, it can be frightening. To realise and accept that there is no one ‘out there’ specially taking care of us, that the world was not created for us or with us in mind, *that* needs extraordinary courage. And yet we have evolved as social beings, we crave companionship – may be a subconscious reaction to realising the
limits of communication? Einstein once said that “Man is, at one and the same time, a solitary being and a social being”. All that can sustain us then are knowledge, and compassionate love. Knowledge that there is a ‘central order’ in nature accessible to our understanding, even if it is insensitive to values, and which we can approach with wonder rather than fear. And compassion for the other whose loneliness is as profound as one’s own. From these alone can spring ethical principles, aesthetic values and the pursuit of quality in all striving – all of which are vital and real to us as a species, even if absent in Nature – and all are ultimately attempts to bridge the unbridgeable.

Suggested Reading

Einstein’s Life and Legacy*

N Mukunda

Introduction

Albert Einstein is the most luminous scientist of the past century, and ranks with Isaac Newton as one among the greatest physicists of all time. There is an enormous amount of material to choose from in talking about Einstein. He is without doubt also the most written about scientist of the past century, may be of all time. The Einstein Archives contain about 43,000 documents, and so far as I know the “Collected Papers of Albert Einstein” have only come upto 1917 with Volume 8 in English translation; another 32 volumes remain to be produced. In the face of all this, this account must be severely selective, and coherent as well.

Einstein’s life was incredibly rich and intense in the intellectual sense. This will become clear as I go along. In any case let me begin by presenting in Box 1 a brief listing of a few important dates in his life, howsoever inadequate it may be.

He was scientifically active essentially from 1902 upto 1935 at the highest imaginable levels, thus for more than three decades.

The Miraculous Year

Now let us turn to technical matters. First, a brief mention of his creative outburst of 1905, whose centenary we are celebrating this year. There were four fundamental papers, and the doctoral thesis, all in the six months from March to September. The first paper on the light quantum concept and explanation of the photo electric effect was submitted to Annalen der Physik in March; the second on Brownian Motion in May; and the third setting out the Special Theory of Relativity in June. His doctoral thesis, just 18 pages long and titled “A new determination of molecular dimensions”, was submitted on July 20th; and the mass-energy equivalence paper in September. Of these four papers, all but the first were within the framework of classical, i.e., prequantum, physics. And his own assessment of them was expressed in a letter to Konrad Habicht on May 18th or 25th, 1905:

“...I promise you four papers in return, the first of which I might send you soon, since I will soon get the complimentary reprints. The paper deals with radiation and the energy properties of light and is very revolutionary, as you will see ...”.

While his work on Special Relativity, leading to a new view of space and time, has caught the public imagination for decades, Einstein himself would say years later that it was no comparison at all to the struggles he faced with both quantum theory and General Relativity. He

abstracted the light quantum concept from an incisive analysis of the high frequency or non-classical Wien limit of Planck’s radiation law, combining thermodynamic and probability arguments. His conclusion was expressed in these words (in modern notation):

\[ \text{Box 1. Some Important Dates in Einstein’s Life} \]

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 March 1879</td>
<td>Born in Ulm, Germany, to Pauline Koch and Hermann Einstein.</td>
</tr>
<tr>
<td>1886–1895</td>
<td>Catholic primary School, then Luitpold Gymnasium, Munich.</td>
</tr>
<tr>
<td>1895</td>
<td>Year spent in Pavia, Italy.</td>
</tr>
<tr>
<td>1895–1896</td>
<td>Cantonal School, Aarau, Switzerland; gives up German citizenship.</td>
</tr>
<tr>
<td>1896–1900</td>
<td>Student at ETH, Zurich for Diploma to teach in high school.</td>
</tr>
<tr>
<td>1901</td>
<td>Gets Swiss Citizenship.</td>
</tr>
<tr>
<td>1903</td>
<td>Marries Mileva Maric.</td>
</tr>
<tr>
<td>1905</td>
<td>Annus Mirabilis; PhD from University of Zurich.</td>
</tr>
<tr>
<td>1909–1911</td>
<td>Enters academia, Associate Professor at University of Zurich.</td>
</tr>
<tr>
<td>1911–1912</td>
<td>Professor, University of Prague.</td>
</tr>
<tr>
<td>1912–1914</td>
<td>Professor, ETH, Zurich.</td>
</tr>
<tr>
<td>1914–1932</td>
<td>Professor, University of Berlin, no teaching duties, under Prussian Academy of Sciences.</td>
</tr>
<tr>
<td>1917</td>
<td>Director, Kaiser Wilhelm Institute, Berlin.</td>
</tr>
<tr>
<td>1919</td>
<td>Divorces Mileva, marries cousin Elsa Einstein Lowenthal.</td>
</tr>
<tr>
<td>1921</td>
<td>First trip to USA; Nobel Prize in Physics.</td>
</tr>
<tr>
<td>1933</td>
<td>Resigns from Prussian Academy, leaves Europe for USA.</td>
</tr>
<tr>
<td>1933–1955</td>
<td>Institute for Advanced Study, Princeton.</td>
</tr>
<tr>
<td>1940</td>
<td>Becomes US citizen (remains Swiss Citizen).</td>
</tr>
<tr>
<td>18 April 1955</td>
<td>Dies at Princeton at 76.</td>
</tr>
</tbody>
</table>

“...monochromatic radiation of low density (within the range of validity of Wien’s radiation formula) behaves thermodynamically as if it consisted of mutually independent energy quanta of magnitude $h\nu$.”

He then applied this ‘heuristic viewpoint’ as he called it to explain Lenard’s results on the photoelectric effect. As for his profound analysis of space and time and the physical meaning of Special Relativity, let me only quote from a later 1907 review to show you how he disposes of the ether hypothesis in such authoritative fashion:

“...electromagnetic forces appear here not as states of some substance, but rather as independently existing things that are similar to ponderable matter and share with it the feature of inertia”.

216
Later work - the Quantum Theory Track

Now I would like to convey to you some idea of the magnitude and grandeur of his later work. For this it is useful to follow two tracks – track one on quantum theory, track two on the relativity theories. Each will be brief. Following the light quantum paper of 1905, the major milestones on the quantum theory track are given in Box 2.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1907</td>
<td>Application of quantum theory to the specific heat problem.</td>
</tr>
<tr>
<td>1916</td>
<td>New derivation of Planck’s Law using Bohr’s idea of stationary states of atoms, spontaneous and stimulated emission and absorption of radiation by matter.</td>
</tr>
<tr>
<td>1924</td>
<td>Theory of the ideal quantum (Bose) gas, number or density fluctuations, wave particle duality for matter, prediction of Bose–Einstein condensation.</td>
</tr>
<tr>
<td>1927–1930</td>
<td>Debates with Bohr – attempts to show inconsistency of quantum mechanics.</td>
</tr>
</tbody>
</table>

Now to some comments. The light quantum idea, conceived in 1905, was not accepted by the physics community in general for a very long time. Some of the principal opponents were Planck, Millikan and Bohr – they all believed that on account of interference and diffraction phenomena, one had to retain the classical Maxwell description of light at all costs, and limit quantum effects to matter and its interaction with radiation. Indeed Planck said in 1909:

“I believe one should first try to move the whole difficulty of the quantum theory to the domain of the interaction between matter and radiation”.

As an amusing interlude, it should be mentioned that in their recommendation of Einstein for membership in the Prussian Academy in 1913, Planck and others wrote:

“That he may sometimes have missed the target in his speculations, as, for example, in his hypothesis of light quanta, cannot really be held too much against him, for it is not possible to introduce really new ideas even in the most exact sciences without sometimes taking a risk”.

But Einstein, while always stressing the provisional nature of his hypothesis, steadily refined it over the years. While in 1905 he referred to light quanta as “mutually independent” and that light energy

“...consists of a finite number of energy quanta that are localized in points in space, move without dividing, and can be absorbed or generated only as a whole”.

in 1909 he referred to them as “independently moving point like quanta with energy $hv$.” In all these statements the reference was to the Wien limit of Planck radiation. In his 1916 work for the first time he explicitly attributed a momentum as well as energy to light quanta. He then wrote in two letters to his friend Michele Besso:

“With that, (the existence of) light quanta is practically certain”, and

“I do not doubt anymore the reality of radiation quanta, although I still stand quite alone in this conviction”.

217
The final acceptance of the light quantum came as late as 1923, after the discovery of the Compton effect, and the name ‘photon’ was invented by the chemist G N Lewis in 1926. But here is Einstein’s own expression of his struggle for comprehension, from a letter to Besso as late as December 1951:

“All the fifty years of conscious brooding have brought me no closer to the answer to the question, ‘What are light quanta?’ Of course today every rascal thinks he knows the answer, but he is deluding himself”.

Till the very end, Einstein never accepted the traditional interpretation of quantum mechanics, and the claim of its finality or completeness. He never gave up his belief in the existence of a reality independent of observation, and his rejection of indeterminism at the fundamental level. After Heisenberg, Dirac and Schrödinger had created quantum mechanics in the two years 1925–1927, at first Einstein tried to show that it was inconsistent. He constructed ingenious thought experiments which he felt would circumvent the Uncertainty Principles, but each time Bohr was able to point out the flaw in the argument. Then in 1935, along with Podolsky and Rosen, he presented an argument purporting to show that quantum mechanics was incomplete. The key idea was to involve the classically valid concepts of locality and realism, and demand that any complete theory contain them. However later developments have shown that their arguments are untenable. For all that, his advice and comments proved crucial for both Heisenberg and Schrödinger in working out their respective forms of quantum mechanics. When Heisenberg told Einstein that in setting up matrix mechanics he had taken inspiration from Special Relativity and its emphasis on observables, Einstein replied:

“...on principle, it is quite wrong to try founding a theory on observable magnitudes alone.... It is the theory which decides what we can observe”.

And in Schrödinger’s case what proved crucial were Einstein’s comments on de Broglie’s thesis of matter waves – “has lifted a corner of the great veil” – and the concept of wave particle duality for matter coming from the density fluctuation formula for the ideal quantum gas.

The Relativity Track

Turn now to track two: the major signposts on the long and tortuous road from Special Relativity in 1905 to the field equations for gravitation in November 1915 are given in Box 3.

It would be completely out of place for me to attempt to discuss these events in any detail, nor am I qualified to do so. I can only offer some comments and quotations to convey the magnitude of what was achieved. One aspect of Einstein’s greatness was his recognition that in reconciling Special Relativity and Newtonian gravity, he had to go beyond both of them. He was willing to give up his own Special Relativity in favour of a more encompassing theory. His unerring instinct based on the Equivalence Principle led him along. Students of General Relativity know how enchantingly beautiful that theory is, but it was the end result of a truly superhuman struggle. Einstein himself said:
Einstein’s Life and Legacy

**Box 3. The Road from Special to General Relativity**

<table>
<thead>
<tr>
<th>Year</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1907</td>
<td>First statement of Equivalence Principle for constant gravitational fields; gravity bends light rays; gravitational redshift.</td>
</tr>
<tr>
<td>1911–1912</td>
<td>Speed of light varies in gravitational field; attempt to use local light speed as scalar gravitational field.</td>
</tr>
<tr>
<td>1912</td>
<td>Analysis of rotating coordinate system; recognition of non-Euclidean geometry of space-time with gravitation present; metric tensor as gravitational field.</td>
</tr>
<tr>
<td>1913–1915</td>
<td>Rejection of general covariance based on meta-argument: metric “should be” uniquely determined by sources and boundary conditions.</td>
</tr>
<tr>
<td>1915</td>
<td>Return to general covariance earlier given up “only with a heavy heart”; the final field equations of general relativity.</td>
</tr>
</tbody>
</table>

“Compared with this problem, the original relativity is child’s play”.

Why did it take so many years to create General Relativity? Here in his own words are the reasons:

“...The postulate of relativity in its most general formulation ... makes the spacetime coordinates into physically meaningless parameters”.

“...Why were a further seven years required for setting up the general theory of relativity? The principal reason is that one does not free oneself so easily from the conception that an immediate physical significance must be attributed to the coordinates”.

After returning to the principle of general covariance, and his discovery of the final field equations for gravitation on 25 November 1915, Einstein wrote in a postcard to Arnold Sommerfeld in February 1916:

“You will be convinced of the general theory of relativity when you have studied it. That is why I am not mentioning a word in its defense”.

He is at his most eloquent in this passage from a 1933 lecture in Glasgow:

“The years of searching in the dark for a truth that one feels but cannot express, the intense desire and the alternations of confidence and misgiving until one breaks through to clarity and understanding are known only to him who has himself experienced them.”

Today General Relativity is a wonderfully mature subject. Ninety years have passed since it left Einstein’s hands to become everyone’s possession. The relationship between Special Relativity and General Relativity is extraordinarily delicate and beautiful. Echoing Landau and Lifshitz – general relativity “represents probably the most beautiful of all existing physical theories”. And, notwithstanding some ambiguities, each special relativistic theory passes over in a natural manner into one consistent with general relativity:

“Every physical theory compatible with special relativity can be aligned into the system of general relativity by means of the absolute differential calculus, without (general relativity) supplying any criterion for the acceptability of that theory”.

Einstein’s deepest insights came from the use of invariance or covariance principles on the one hand, and the study of statistical fluctuations and thermodynamic arguments on the other. He viewed Special Relativity as a restrictive principle, not as a specific physical model:
“The principle of relativity is a principle that narrows the possibilities; it is not a model, just as the second law of thermodynamics is not a model”.

**So far ahead**

In ever so many instances he was far ahead of all others in his understanding of Nature. Leaving aside General Relativity which was his own creation, we can cite several examples. In 1905, as we saw, he had ‘teased out’ the light quantum concept from the Wien limit of Planck’s law; he then realised that in the context of the complete law these quanta could not be regarded as mutually independent but would influence one another in some way. He even posed this years later as a problem for Bose to look at. In Stachel’s view, this was an early glimpse of quantum mechanical entanglement, brought out fully into the open in the 1935 EPR paper. Then already in 1909 he grasped wave particle duality for light:

“It is my opinion that the next phase in the development of theoretical physics will bring us a theory of light that can be interpreted as a kind of fusion of the wave and the emission theory”.

After the work in 1916–17 on spontaneous and stimulated emission he saw very clearly that the Maxwell field had to be quantized:

“The properties of elementary processes . . . make it seem almost inevitable to formulate a truly quantized theory of radiation”.

This is to be particularly contrasted with the views of Planck and Bohr, mentioned earlier, that radiation should continue to be treated classically. The quantum theory of the Maxwell field was formulated ten years later, in 1927, by Paul Dirac. Again in that same body of work he glimpsed the role of chance in quantum phenomena but regarded it as a deficiency, for he said it is

“a weakness of the theory . . . that it leaves time and direction of elementary processes to chance; nevertheless I have full confidence in the reliability of the way entered upon”.

So much of later quantum mechanics sensed so early!

**Language, God, Music . . .**

Let us now turn briefly from his science to other aspects. Einstein had a remarkable gift with language, and wrote with great eloquence. He showed wonderful clarity, independence and boldness in thought and expression, and greatly inspired many of his own and later generations. Here are his tributes to some whom he greatly respected. After a careful critique of the foundations and inadequacies of Newton’s mechanics, in the Autobiographical Notes, he says:

“Enough of this. Newton, forgive me; you found the only way which, in your age, was just about possible for a man of highest thought and creative power. The concepts, which you created, are even today still guiding our thinking in physics, although we now know that they will have to be replaced by others farther removed from the sphere of immediate experience, if we aim at a profounder understanding of relationships”.

220
Einstein’s Life and Legacy

About Lorentz, without whose preparatory work Einstein said he could never have discovered Special Relativity, he wrote:

“He meant more to me personally than anybody else I have met in my life time”.

And soon after his first meeting with Niels Bohr in 1920, he said in a letter:

“Not often in life has a human being caused me such joy by his mere presence as you did”.

About his work, some other gifted writers had this to say:

Hermann Minkowski on Special Relativity in 1908: Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.

Hermann Weyl on General Relativity in 1918: ‘Einstein’s Theory of Relativity has advanced our ideas of the structure of the cosmos a step further. It is as if a wall which separated us from Truth has collapsed.’

Paul Dirac in 1979: ‘The Einstein theory of gravitation has a character of excellence of its own. Anyone who appreciates the fundamental harmony connecting the way nature runs and general mathematical principles must feel that a theory with the beauty and elegance of Einstein’s theory has to be substantially correct’.

And here is a comment in a lighter vein: when Chaim Weizmann and Einstein reached New York after a long ship journey from Europe in 1921, Weizmann told the waiting reporters:

“Einstein explained his theory to me every day, and on my arrival I was fully convinced that he understood it”.

Quite early in life, at age 12, Einstein lost faith in formal religion. He studied Kantian philosophy in his later student years. Often in scientific matters he referred to God in what may be termed a playful manner:

“Subtle is the Lord, but malicious He is not”.

And about quantum mechanics:

“The theory yields much, but it hardly brings us close to the secrets of the Old One. In any case, I am convinced He does not play dice”.

As he once explained, the God he had in mind was the God of Spinoza: Nature.

An oft-quoted pithy sentence is:

“Science without religion is lame; religion without science is blind”.

Deep though it is, there is an austere quality about it too. Probably inspired by it, Victor Weisskopf expressed the essence of the idea in more human terms:

“Human existence is based upon two pillars: compassion and knowledge. Compassion without knowledge is ineffective; knowledge without compassion is inhuman”.

Music was a vital component of Einstein’s life, having learnt to play the violin at a young age. The most memorable incident in this connection is his reaction to Yehudi Menuhin’s debut in 1929, at age 13, with the Berlin Philharmonic, playing the Bach, Beethoven and Brahms concertos in one programme. Einstein was so moved that he took Menuhin in his arms and exclaimed: “Now I know that there is a God in heaven!”

Through his clarity, courage and beauty of imagination, Einstein continues to inspire those of us who have chosen to pursue science, physics in particular. We are fortunate to be relatively
close to him in time – born 126 years ago, worked from 100 to 70 years ago, died just 50 years ago. Much closer to us than figures like Kepler, Galileo and Newton. We constantly wonder – how could one person have accomplished so much? We physicists must consider ourselves fortunate – we are able to try and understand what he did. Just as someone who knows Sanskrit can savour Kalidasa, those at home with English can delight in Shakespeare, and those with some musical training can lose themselves in the classics. So for us with a physics training the highest achievements of human thought in science are accessible. Others too can admire him for his human qualities.

Einstein realised full well that others would go beyond his work, however great it is. He said:

“In my opinion there is the correct path and . . . it is in our power to find it”.

When he died, his intellectual heir Wolfgang Pauli declared:

“Einstein’s life ended . . . . with a demand on us for synthesis”.

How then to conclude this talk? I can think of no better way than to say of Einstein what he said about Gandhi in 1939:

“Generations to come, it may be, will scarce believe that such a one as this in flesh and blood walked upon this earth”.

Suggested Reading


The Philosophy of the Physical Sciences*

N Mukunda

The various philosophical traditions of the world form an important part of the intellectual and cultural achievements of the civilizations which produced them. Typically, their roots go back thousand of years – as in the cases of India and Greece. There is in them much poetic imagery and logical and deep thinking, as well as a sizeable speculative component. In contrast, modern science as we know it developed barely four hundred years ago, in the seventeenth century, arising in the main out of the combined efforts of Copernicus, Kepler, Galileo, and Newton. It was only then that the importance of controlled experiments and careful and systematic quantitative study of natural phenomena was clearly recognized. However, in spite of these great differences in age, at least in the Western tradition the interactions between modern physical science and philosophy have been deep and profound.

I am not a professional philosopher. I have only been attracted to some philosophical questions, and been impressed by certain philosophical systems, as a result of a study of physics. Thus the content of this article may sometimes reveal a sense of naivety as regards formal philosophical matters, schools of thought, traditions, and the like. Nevertheless, I hope that what follows will be of interest to the readers of this journal, most of whom may not be professional scientists but would still have a lively interest in these matters.

It may not be out of place to mention here some contrasting attitudes to the possible roles and value of philosophical thinking that are evident in the developments in physics over the past century. As a consequence of the European, in particular the continental tradition, the general writings of the two discoverers of quantum mechanics – Werner Heisenberg from Germany and Erwin Schrödinger from Austria – show great familiarity with and interest in various philosophical systems of thought, from the Greeks onwards. While the writings of Niels Bohr and Albert Einstein also often have a philosophical bent, their references to formal systems of philosophy tend to be fewer, but nevertheless important. In contrast, when the focus of work in the new physics shifted from Europe to the US around the middle of the twentieth century, this regard for general philosophical thinking among the leading professional physicists does seem to have weakened. Typical statements of Richard Feynman and Steven Weinberg, for instance, display a certain degree of disdain, or certainly a lack of sympathy, for the value of philosophical thinking in the physical sciences.

In any case, in the present account I assume that there is value in looking at the growth of modern physical science from a ‘philosophical point of view’, though it may require some degree of maturity as well as sympathy to adopt this attitude.

N Mukunda

We may say for our present purposes that philosophy of science is generally concerned with the nature of knowledge, the way we acquire it, the meaning of understanding, and the evolution of concepts, all in the context of the physical sciences. It may in addition be ultimately concerned with an appreciation of our place in nature. Philosophy of science deals with the understanding of natural phenomena and how this understanding is achieved, with the general features common to the various branches of science, and with the interdependence of these branches. It is more interested in the overall pattern of natural laws than in the details of any particular area of science.

Our aim will be to come up to the modern era in physics, and to see what it has taught us with regard to questions of a philosophical nature. Along the way we shall briefly review some historical developments and ways of thinking or schools of thought, both in philosophy and in physical science. We will consider how concepts are created, how they grow, and how they have sometimes to be greatly modified or even abandoned. Naturally, developments in physics will be covered in slightly greater detail than those in formal philosophy.

Rationalism and Empiricism

In our account of the beginnings of science and philosophical thinking we go back to Greek times. The major creative period, lasting about four hundred years, began with Thales of Miletus (c.624–c. 546 BCE) and included, among many renowned thinkers, Pythagoras (c.580–c.500 BCE), Anaximander (610–c.545 BCE), Democritus (c.460–370 BCE), Leucippus (fl.5th cent.BCE), Plato (427–347 BCE), Aristotle (384–322 BCE), and Euclid (fl.c.300 BCE). In the early period, with Thales, there was a strong impulse towards, as Benjamin Farrington puts it, ‘a new commonsense way of looking at the world of things... the whole point of which is that it gathers together into a coherent picture a number of observed facts without letting Marduk [the Babylonian Creator] in’. The attempt was to deal with nature on its own, not bringing in mystical or mythical leanings. To quote from Heisenberg: ‘The strongest impulse had come from the immediate reality of the world in which we live and which we perceive by our senses. This reality was full of life, and there was no good reason to stress the distinction between matter and mind or between body and soul.’

Thales was familiar with the knowledge of geometry developed by the Egyptians, the basic facts of static electricity, and the magnetic properties of lodestone. Later, Democritus and Leucippus propounded the atomic concept of matter, not in a casual manner but based on careful reasoning. However, it goes without saying that philosophical thinking in these early times had a considerable speculative content, and there were others such as Plato and Aristotle who later strongly opposed the atomic hypothesis. This should come as no surprise at all, since as late as the end of the nineteenth century there were influential figures – Ernst Mach and Wilhelm Ostwald – who were still opposed to the idea of atoms. This idea finally triumphed only thanks to the heroic efforts of Ludwig Boltzmann, and Einstein’s work on Brownian movement.

The knowledge of geometry brought by the Greeks from Egypt was perfected and presented in an axiomatic form by Euclid of Alexandria around 300 BCE. The fact that this subject could
The Philosophy of the Physical Sciences

be presented as a deductive system – a large number of consequences or theorems following logically from a very few ‘self-evident’ axioms or ‘obvious’ truths – must have made a deep impression on the Greek mind. It led in course of time to the idea that the behaviour and laws of nature could be derived from pure reason, without the help of direct inputs from experience. This was the so-called rationalist philosophy of science, which lay in stark contrast to the initial empiricist approach of Thales and Democritus. Plato held that ‘knowledge of Nature does not require observation and is attainable through reason alone’. Before Plato, Pythagoras too espoused this point of view, other illustrious followers being Aristotle and, in much later times, René Descartes, Wilhelm Leibniz, and Benedict de Spinoza. One may say that this rationalist philosophy accords a privileged position to human beings in the scheme of things.

The opposite – empiricist – point of view holds that knowledge comes ultimately from experience of phenomena and not from reason As we saw, this was the attitude of both Thales and Democritus; and in later centuries it was revived by Francis Bacon and carried forward by John Locke, George Berkeley, and David Hume as a reaction to the rationalist view on the European continent. We shall return to some of these contrasting philosophies later, only noting now that empiricism goes with a more modest attitude towards our place in nature.

From Galileo and Newton to Kantian Philosophy

Modern science emerged in Europe during the Renaissance – the reawakening of classical ideals in arts, literature, and philosophy during the fourteenth to seventeenth centuries, brought about by a combination of social, political and religious factors. This is not the place to go into this crucial advance in any detail, but we note that it occurred against the background of a liberating intellectual and philosophical atmosphere to which many – including Descartes, Leibniz, and Spinoza – contributed.

Empirical Advances: Nicolaus Copernicus initiated the movement away from a human-centred view of nature with his heliocentric model of the solar system, and Francis Bacon showed the way to freedom from reason alone as the source of all knowledge. Indeed, Bacon said of Aristotle: ‘He did not consult experience as he should have done... but having first determined the question according to his will, he then resorted to experience, and... led her about like a captive in a procession.’ Copernicus’s work, as well as Kepler’s discovery of the three laws of planetary motion during the years 1609–19, was but preparation for what was to come in the work of Galileo and Newton.

Galileo, rightly regarded as the founder of modern science, not only discovered the law of inertia in mechanics, the kinematic description of motion, and the law of free fall, but also stressed the importance of performing controlled experiments, of quantitative measurement, and of the use of mathematics in expressing experimental results. He stated this last point with particular emphasis, saying about the ‘book of nature’: ‘It cannot be read until we have learned the language and become familiar with the characters in which it is written. It is written in mathematical language.’
It was Isaac Newton, born the year Galileo died (at least by one calendar), who completed the work initiated by Galileo and his other illustrious predecessors, and paved the way for the systematic scientific investigation of physical phenomena over the succeeding centuries. We can say that without Newton’s crowning achievements, this tradition – the Galilean–Newtonian worldview – would not have been securely established. Speaking of the importance of what Galileo and Newton achieved, Max Born says: ‘The distinctive quality of these great thinkers was their ability to free themselves from the metaphysical traditions of their time and to express the results of observations and experiments in a new mathematical language regardless of any philosophical preconceptions.’

**Scientific Method:** Newton expressed clearly his views on the independent and absolute natures of space and time, stated his three laws of motion for material bodies as axioms, enunciated his law of universal gravitation, and established mechanics as a deductive system. His whole approach and accomplishments made explicit and clear all the steps in the chain of scientific work: observation and experimental data → analysis using mathematics → discovery and enunciation of fundamental laws → further mathematical deduction → predictions to be tested by new experiments. As he put it: ‘To derive two or three general Principles of Motion from Phaenomena, and afterwards to tell us how the Properties and Actions of all corporeal Things follow from those manifest Principles, would be a very great step in Philosophy, though the causes of those Principles were not yet discover’d’.

**Absolute Space and Time:** For the purpose of developing mechanics, Newton invented the calculus. In his presentation he adopted the Greek attitude to geometry and the style of Euclid. Thus he converted knowledge obtained inductively from (refined!) experience – extension from the particular to the general – into a deductive style of presentation. From his laws of motion and universal gravitation, all the empirical laws of Kepler and Galileo followed as logical mathematical consequences. His clear statements about the natures of space and time were of critical importance at this juncture. They mark an important phase in our understanding of these key components of nature, and as we emphasize later, this understanding is never final but develops continually ‘in time’ as we gather more and more experience. Who better than Einstein to express all this: ‘It required a severe struggle [for Newton] to arrive at the concept of independent and absolute space, indispensable for the development of theory. Newton’s decision was, in the contemporary state of science, the only possible one, and particularly the only fruitful one. But the subsequent development of the problems, proceeding in a roundabout way which no one could then possibly foresee, has shown that the resistance of Leibniz and Huygens, intuitively well-founded but supported by inadequate arguments, was actually justified. ...It has required no less strenuous exertions subsequently to overcome this concept [of absolute space].’

**Theory and Experiment:** In Newton’s work we see a confluence of the inductive and deductive methods, each playing its due role. There was a unification of celestial and terrestrial gravitational phenomena, and many previously intractable problems became amenable to analysis and understanding. At one point he went so far as to claim that he made no hypotheses – ‘Hy-
The Philosophy of the Physical Sciences

\textit{potheses non fingo}\textsuperscript{1} – hinting at pure empiricism; but this actually shows that modern science was still young. As Einstein aptly said: ‘The more primitive the status of science is the more readily can the scientist live under the illusion that he is a pure empiricist.’ Today the level of sophistication of the physical sciences is such that every worthwhile experiment is heavily dependent on previous and current theory for its motivations, goals, methods, and analysis.

Over the course of the eighteenth century, the Galilean-Newtonian approach to physical science was amazingly successful. It was applied to problems of celestial mechanics or astronomy, fluid dynamics, and elastic media among others. A distinguished line of mathematical physicists – Leonhard Euler, Joseph Lagrange, Pierre Simon de Laplace, and many others – took part in this endeavour. At one point Lagrange complained that, after Newton, there was nothing left to be discovered! Towards the end of the century, the laws of static electricity and magnetism also fell into the Galilean-Newtonian pattern.

**Thought as a Synthetic A Priori:** Around this time, the philosopher of the Enlightenment, Immanuel Kant, was so impressed by these successes of the Galilean-Newtonian approach that he created a philosophical system to explain or justify them. We mentioned earlier the contrasting rationalist and empiricist schools of philosophy. Kant tried to bring them together and offered an explanation of the triumphs of Galilean–Newtonian science along the following lines. He distinguished between \textit{a priori} and \textit{a posteriori} forms of knowledge – respectively in advance of, and as a result of, experience of nature – and between two kinds of statements: \textit{the analytic}, which are empty (such as definitions and statements of a logical nature); and \textit{the synthetic}, which had nontrivial content and could in principle be false. He saw two paths to knowledge about nature – that which is a priori, and that which results from experience. Some of the basic physical ideas underlying Galilean–Newtonian physics, which were actually the results of long human experience and experiment, were regarded by him as synthetic a priori principles. Thus they were claimed to be available to us innately – as a result, one might say, of pure reason – and were necessarily valid and obeyed by natural phenomena. Some of these synthetic a priori principles were the separate and absolute natures of space and time, as expressed by Newton; the validity of Euclidean geometry for space; the law of causality; and later on even the permanence of matter, and the law of conservation of mass. In effect, Kant took the knowledge of physical phenomena available in his time and made some of it necessarily and inevitably true and binding on nature. These synthetic a priori principles were present in our minds before any experience of nature; they were thought of as preconditions for, rather than results of, science.

Kant’s attempt was made about two centuries ago, and today it is clear that it was tied to his age and to the science of his time. Schrödinger characterizes well the impulse that lay behind Kant’s attempt: ‘One is very easily deceived into regarding an acquired habit of thought as a peremptory postulate imposed by our mind on any theory of the physical world.’ We will shortly look at some of the ways in which physical science has gone beyond Kant’s framework, and will describe a fascinating new way of understanding the origin of synthetic a priori principles of thought.
Fields as Distinct from Matter: At the start of the nineteenth century the fields of optics, electricity, and magnetism were separate from one another and from mechanics. Chemistry was a distinct discipline. But over the century many advances were made, which we can only briefly describe here. An early step forward was in the understanding of the nature of light. Thomas Young’s experiments on interference brought the wave theory of light back into favour, as against Newton’s corpuscular ideas. This was carried forward and firmly established by Augustin Fresnel. Then, as a result of fundamental experimental discoveries by Hans Oersted, André Amperè, and Michael Faraday, the concepts of time-dependent electric and magnetic fields came into being. There were things in nature in addition to and distinct from matter.

Meanwhile, celestial mechanics continued to record stunning successes. Perhaps the most striking example was the prediction by both John Adams and Urbain Le Verrier, based on Newtonian mechanics and gravitation, of the existence of a new planet, Neptune, to account for the observed discrepancies in the motion of Uranus. In 1846 it was found exactly where the astronomers were told to look. (However, a later similar attempt to trace discrepancies in the motion of Mercury to a perturbing planet Vulcan was unsuccessful. The answer came from an entirely unexpected direction – general relativity.)

Electromagnetism and Light: After Faraday’s powerful intuition had led to the idea of electric and magnetic fields, James Maxwell put all the known laws in the subject of electricity and magnetism into a coherent mathematical form. He then found an important discrepancy, saw the way to correct it, and was thus led to his comprehensive classical unified theory of electromagnetic phenomena. A prediction of this theory was the possibility of self-supporting electromagnetic waves whose speed when calculated turned out to be exactly the known speed of light. Then Maxwell identified light with these waves, and optics became a part of electromagnetism. During this period, following Fresnel’s work, it was believed that the propagation of light needed a material medium, the so-called luminiferous ether, and this concept was taken over by Maxwell as well.

Non-Euclidean Geometry: In the area of mathematics, the subject of geometry witnessed a major advance. We saw that Kant in his philosophy had made Euclidean geometry an inevitable or inescapable property of physical space – it was a synthetic a priori principle. Within mathematics, for centuries the status of one of Euclid’s postulates – the fifth one, the parallel postulate (that there is exactly one parallel to a given line through a given point) – had been repeatedly studied: was it logically independent of the other postulates or a consequence of them? During the first half of the nineteenth century, three mathematicians – Karl Gauss, Nikolai Lobachevsky, and János Bolyai – independently showed that it was a logically independent statement. It could be altered, allowing one to create logically consistent alternatives to Euclidean geometry. Thus was born within mathematics the concept of non-Euclidean geometry, which, as we will soon see, was to enter physical science just under a century later.
Statistical Physics: Over the latter half of the nineteenth century, statistical physics and statistical mechanics became established as foundations of thermodynamics. Thus by the century’s end the principal components of the physicist’s view of the world were Newton’s mechanics, Maxwell’s electromagnetism, and statistical ideas and thermodynamics.

Relativity: The important departures from the Kantian picture of physical science – from the framework Kant developed to justify the successes of Galilean–Newtonian ideas – came one by one with the revolutionary theories of twentieth-century physics. First came the special theory of relativity, the resolution of a clash between Newton’s mechanics and Maxwell’s electromagnetism. It turned out that Newton’s views of separate and absolute space and time, and the Galilean transformations that go with them, were incompatible with Maxwell’s electromagnetic equations. These equations led to a profoundly different view of the properties of space and time. What special relativity achieved was to make clear these properties, show that there was no need for ether as a carrier of electromagnetic waves, and then amend Newton’s mechanics of material particles to make it consistent with electromagnetism. The earlier separateness and individual absoluteness of space and time – included among Kant’s synthetic a priori principles – gave way to a unified view in which only a combined space-time was common to and shared by all observers of natural phenomena. However, each observer could choose how he or she would split space-time in a physically meaningful way into separate space and time. The earlier absoluteness of the concept of simultaneity was lost, and now varied from observer to observer. For each observer, though, space continued to obey the laws of Euclidean geometry. Special relativity took one more step beyond the Kantian framework – now only a combined law of conservation of matter and energy was valid, not separate laws for matter and for energy.

The other major twentieth-century development in physics was the discovery of the quantum nature of phenomena and the formulation of quantum theory. In many ways, quantum theory is more profound in its implications than the relativity theories. Quantum theory arose out of a clash between Maxwell’s electromagnetism and the principles of statistical physics, which, as we saw, provide the foundation for thermodynamics. We can only try to convey why quantum theory has had such a profound influence on the philosophy of science, and cannot venture into much technical detail. The view of the nature of light has swung back towards Newton’s corpuscular conception – with important and subtle differences – expressed in the concept of the photon. As for the mechanics of matter, the Galilean-Newtonian picture and description of motion has given way to a much more mathematically elaborate and subtle complex of ideas, which goes by the name of quantum mechanics. Material particles no longer travel along well-defined paths or trajectories in space in the course of time. Their evolution in time can only be given in the language of probability – that is, all the predictive statements of quantum mechanics are probabilistic in nature. The quantitative description of physical properties of systems undergoes two important changes in quantum mechanics: on the one hand, many physical variables show a quantization of the values they can possess – thus, typically, energies are restricted to a discrete set of values rather than a continuum. On the other hand, the physical variables of a given system have such mathematical properties, or are of such nature,
that we cannot imagine that each of them always possesses some numerical value which, if we so wish, can be revealed by a measurement. According to Bohr, we can never speak of a quantum system as having such and such a value for such and such a physical property on its own, independent of our measurement of it. And with a pair of so-called incompatible properties, an effort to measure one of them automatically precludes any effort to simultaneously measure the other as well.

We have to learn to use language with much more caution or circumspection when speaking of quantum phenomena than was the case earlier. Many classically meaningful and answerable questions become devoid of meaning in the quantum domain. The kind of ‘visualizability’ of physical systems in complete detail which was possible in classical physics is denied by quantum mechanics.

From the perspective of Kantian thinking, quantum mechanics has made us give up strict determinism, substituting a kind of statistical causality for it. On the other hand, it has supplied the basic theoretical concepts for all of chemistry, for atomic, molecular, nuclear, and elementary particle phenomena, and for all processes involving radiation. The old law of the permanence of matter has gone, as it can be converted to radiation, and vice versa. Up to the present time, the agreement of quantum mechanics with experiments has been outstanding – nature does seem to behave, in many situations, in classically unreasonable ways.

The Reinterpretation of Kantian Ideas

It is understandable that when physics advanced into new territories involving the very fast, the very large, and the very small – as judged by everyday standards and experience – some of the Kantian synthetic a priori principles had to be given up. As we said, Kant’s ideas were rooted in the physical science and Galilean-Newtonian tradition of his time; he could not have foreseen the revolutionary developments that were to come later. This much is natural. However, what is remarkable is that the ‘problem’ with his philosophical basis for physical science has been illumined during the mid-twentieth century from a rather unexpected direction – namely, biology and the theory of evolution by natural selection. One might wonder if, apart from having to give up particular synthetic a priori principles as a result of advances in physical science, the very concept of such principles has also to be given up. After all, one might ask how principles supposedly known in advance of experience could necessarily constrain our later experiences. The answer to this question involves a subtle reinterpretation of Kant’s notions, using ideas not available to him. This fascinating development – the work of Konrad Lorenz – leads to a better understanding of the entire situation, and has been eloquently presented by Max Delbrück.

The basic contrast is between the slow evolution of species governed by the force of natural selection, involving innumerable generations and enormous stretches of time; and the relatively short life span of an individual member of the species. In the former process – phylogenesis – those abilities thrown up by random genetic changes which are beneficial to biological survival are retained and refined. The others are discarded. Those retained include the ability to recognize the most important physical features of the world around us at our own scales of length and time, because it is just these scales that are relevant for biological evolution. Thus,
gradual evolution of species governed by natural selection develops these useful capacities, and then endows each individual with them at birth. From the point of view of the individual’s development over a single lifetime – ontogenesis – the capacities in question seem to be given ready-made at birth, in advance of experience; they seem to be a priori. But this argument shows that from a longer time perspective there is nothing a priori about them, as they are the fruits of experience of the species. In Delbrück’s words:

> It appears therefore that two kinds of learning are involved in our dealing with the world. One is phylogenetic learning, in the sense that during evolution we have evolved very sophisticated machinery for perceiving and making inferences about a real world.... In other words, whereas in the light of modern understanding of evolutionary processes, we can say the individual approaches perception a priori, this is by no means true when we consider the history of mankind as a whole. What is a priori for individuals is a posteriori for the species. The second kind of learning involved in dealing with the world is ontogenetic learning, namely the lifelong acquisition of cultural, linguistic, and scientific knowledge.

The one added subtle point is that species evolution endows each individual with the capacity to acquire knowledge about the world outside, but not the knowledge itself. This has to be acquired through the experiences of infancy and childhood, and indeed is a lifelong endeavour. The difference between capacity and content is profound. In this way Kant’s conceptions acquire new meaning. We also learn that the biologically evolved Kantian a priors can only be expected to work for a limited range of natural phenomena, and our ‘sense of intuition’ is based on this range alone. We should therefore not be surprised if Galilean-Newtonian principles do not extend beyond this limited world to the world of the very fast, very large, or very small. But the truth is that our intuition is so much a part of us that it is very difficult to escape from or transcend it.

**Some Important Features of Physical Science**

Returning to physical science, there are several important features it has acquired, some more recently than others, with significant philosophical implications.

The descriptions and understanding of natural phenomena given by physical science are always developing or evolving, always provisional and never final. Since this is so very important, let me cite several examples which lead one to this sobering point of view. There have been occasions in the past – with Lagrange in the eighteenth century, and William Thompson (Lord Kelvin) at the end of the nineteenth century – when the feeling was expressed that all the laws of physics had been found and nothing remained to be discovered. Our experiences since then have made us much more modest in our claims. We both recognize the existence of limits of validity for every physical theory or body of laws, even for those yet to be discovered; and admit that future experience can always lead to unexpected surprises. In this important sense Nature is inexhaustible: we will always be learning from her. The lack of finality of every
physical theory in this sense means that we can only continually increase the accuracy of our
description of the phenomena of ‘the real world out there’ but can never say we have been able
to describe them exactly as they are, or have reached true reality.

Our first example to drive these points home is connected with the Newtonian description
of universal gravitation as an instantaneous attraction between any two mass points governed
by an inverse square law. Before Newton, the prevailing idea was Descartes’ theory of vortices – all physical actions or influences were by contact alone. Newton’s law was a major
change, giving rise to the concept of action at a distance. Privately, Newton himself expressed
uneasiness at what seemed an unreasonable aspect of his law – how could material bodies in-
fluence one another instantaneously across intervening empty space? But his law worked, its
quantitative predictions agreed with experience (at that time!), and with the passage of time
the idea of action at a distance became gradually accepted. Even the initial laws of electricity
and magnetism – in the static limit – were expressed in such a framework. The return to action
by contact via an intervening field came about in the case of gravitation only in 1915 with
Einstein’s theory of general relativity.

The next example concerns the nature of light. As we have discussed earlier, the cor-
puscular viewpoint championed by Newton was replaced by the wave concept after Young’s
experiments on interference. After Maxwell’s classical electromagnetism arrived, light was
identified with the propagating waves of Maxwell’s theory: now one ‘knew’ what the waves
were made of. But when Einstein developed the photon concept in 1905, our understanding
moved once more in the direction of the corpuscular viewpoint, involving a subtle combination
of wave and particle concepts which can be properly expressed only in the language and im-
agery of quantum mechanics. At none of the above stages of development could one claim that
one had finally understood the real nature and properties of light. It was always a movement
towards improved understanding.

Our third example concerns the explanation of the spectrum of the simplest atom in Nature,
hydrogen. Bohr’s 1913 theory was the first breakthrough; it gave the vital clue to the wealth of
data in the field of spectroscopy. Spectral lines corresponded to transitions of electrons between
atomic states with various discrete energies. His model for the hydrogen atom was able to
explain the spectral lines of the so-called Balmer series, and also several other series. This
vital first step fell within the framework of the old quantum theory. A few years later, Arnold
Sommerfeld introduced special relativistic corrections to the Bohr model, and was thus able to
explain the so-called fine structure in the spectrum. This was then regarded as a triumph of the
existing theoretical framework. But after the advent of quantum mechanics in 1925–26, the
‘correct’ understanding of the spectrum of hydrogen was supplied by the Schrödinger equation
and its solutions. The framework of physical ideas was completely different from Bohr’s, but
the data explained was the same. Then in 1928, after Dirac had found the relativistic wave
equation for the electron, the fine structure came out as a straightforward consequence. After
this, the Sommerfeld explanation became a fortuitous coincidence, not to be taken seriously
anymore. Almost two decades later, as improved experimental techniques and measurements
revealed new and finer details of the hydrogen spectrum – the so-called Lamb shift – one had
to go beyond the Dirac equation and appeal to the theory of quantum electrodynamics (QED) for an explanation. This turned out to be one of the triumphs of that theory. Clearly at no stage could we have said that we had understood the origin of the lines of the spectrum of hydrogen in complete detail, or that we had the complete and real truth in our possession.

Turning from physics to mathematics, in the field of geometry we have seen a similar evolution, though over a much longer period of time. As we mentioned earlier, only after almost two millennia was it realized that Euclid’s geometry is not the only logically possible system of geometry for space; other non-Euclidean geometries are certainly conceivable and consistent. And after general relativity, the changeable geometry of space-time has become an ingredient in the laws of physics, specifically of gravitation. Today there is talk of the quantum features of geometry, one more step in the continuing effort to understand the natures of space and time.

These examples, and many others, teach us that the problem of what is physically real is a time-dependent one: it always depends on what is known at each epoch in the growth of physical science, and can see dramatic changes at certain points. Concepts like phlogiston and ether seemed essential at certain stages in the history of physics, but were later given up in the light of improved understanding.

The accuracy of observations and measurements and the sophistication of the instruments available for experimental investigation also continually increase, so they too contribute to the transitoriness of physical theories. But it should also be pointed out that at any given time we have trust in certain tested and successful ideas and theories, and keep working with them until we are compelled by new experience to go beyond them; then we modify them or in some cases even abandon them. Thus at the present time, we have full confidence that within their respective domains of validity, Newton’s mechanics, Maxwell’s electromagnetism and the nonrelativistic quantum mechanics and its later developments can certainly be used.

Mathematics: The Language of Nature

Next we turn to the important role of mathematics in physical science. Galileo’s remark about mathematics being the language of Nature has turned out to be true, at least in physical science, to a degree far beyond what anyone might have imagined. In the eighteenth and much of the nineteenth century, as the concepts about the physical universe grew in complexity and subtlety, so did the mathematics used to describe them. The same gifted individuals contributed to both disciplines in these periods – Euler, Lagrange, Laplace, Fourier, Gauss, Hamilton, and Jacobi, to name a few. Thereafter, there was to some extent a parting of ways. The relativity and quantum revolutions in the twentieth century exploited mathematical ideas previously and independently developed purely within mathematics. In any event, there has been a steadily increasing role for mathematical ideas in physical science. In one sense this is connected to the reinterpretation of Kantian ideas sketched in the preceding section. As we move away from the domain of normal daily experience and into unfamiliar realms, it is understandable that our intuition often fails us, and then we depend increasingly on the mathematical structure of
physical theory for guidance. Furthermore, the accuracy with which effects can be predicted by modern physical theories, and then checked by experiments, is truly staggering. In Eugene Wigner’s view, there seems to be no rational explanation for this to be so.

There are some who regard the body of mathematical truths as an independently existing ‘continent out there’ and the process of mathematical discovery as the result of continual exploration of this continent. However, it is likely that this is a psychological response from some gifted individuals who have made really deep discoveries in mathematics based on a variety of motivations. A more modest and less problematic attitude is to regard mathematics as a human invention, similar to but far more compact and rigorous than language, given that in the first place evolution has equipped us with the capacity to create it. But then the extraordinary degree of detail and verification of physical theories via their predictions – this is what seems difficult to explain, and what Wigner terms a miracle. In Dirac’s view, the reason why the method of mathematical reasoning works so well in physical science is along these lines: ‘This must be ascribed to some mathematical quality in Nature, a quality which the casual observer of Nature would not suspect, but which nevertheless plays an important role in Nature’s scheme’.

Another related point stressed by Dirac should also be mentioned. It turns out that in the long run, the deductive method is not suitable for physical science. One cannot base one’s ideas on a fixed, initially stated, and unchanging set of axioms, and then rely on logic to obtain all possible physical consequences. One may adopt this strategy – inspired by Euclid – to a limited extent to grasp the logical structure of a particular set of ideas in a compact way, but one is bound sooner or later to transcend the confines of such a structure. This has been the case, for instance, with Newton’s axiomatic approach to mechanics – witness the changes wrought by special relativity on the one hand, and quantum theory on the other. Such may well be the case with the present highly successful quantum mechanics as well. Turning to Dirac:

The steady progress of physics requires for its theoretical formulation a mathematics that gets continually more advanced. This is only natural and to be expected. What, however, was not expected... was the particular form that the line of advancement of the mathematics would take, namely, it was expected that the mathematics would get more and more complicated, but would rest on a permanent basis of axioms and definitions, while actually the modern physical developments have required a mathematics that continually shifts its foundations and gets more abstract. ... It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalization of the axioms at the base of the mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation.

Looking Back Philosophically

Philosophical insights into and speculations about Nature go far back in time; modern science in comparison is very recent. We have followed the growth of physical science from its modern
The Philosophy of the Physical Sciences

David Bohm and Renee Weber on Physics and Maths

Weber The modern physicist is more like the materialist.

Bohm Basically; except for this tremendous emphasis on mathematics, which is like saying that God is a mathematician. If you emphasize mathematics as much as scientists now do, without any physical picture of matter, you are tacitly saying that the essence of the world is something abstract and almost spiritual, if you really think about it.

Weber Mathematics is pure thought.

Bohm That’s right. You won’t find it anywhere in matter.

Weber You are saying that even today’s physicists who might be least inclined towards anything spiritual are practically forced to assume that it is beyond the material.

Bohm Tacitly, anyway. Physicists may not accept this, but they are attributing qualities to matter that are beyond those usually considered to be material. They are more like spiritual qualities in so far as we say there is this mathematical order which prevails, which has no picture in material terms that we can correlate with it.

Weber Is it an aesthetic principle of something deeper still that makes them hold out for one rather than for three or four ultimate laws? Is it a spiritual drive, without their realizing it?

Bohm It probably is a universal human drive, the same one which drives people to mysticism or to religion or art....

Weber Feynman said that those who don’t understand mathematics don’t realize the beauty in the universe. Beauty keeps coming up, together with order and simplicity and other Pythagorean and Platonic categories.

Bohm Order and simplicity and unity, and something behind all that which we can’t describe.

— Dialogues with Scientists and Sages: The Search for Unity

beginnings at the hands of Galileo and Newton, and the impact it had on philosophy in that period. We saw how classical physics seemed to have achieved a kind of completeness at the end of the nineteenth century, after which the relativity and quantum revolutions occurred.

In discussing or evaluating ancient philosophical ideas in the light of knowledge attained much later, a great sense of balance is needed. Such comparisons can easily be misunderstood. On this point, Heisenberg explains:

‘It may seem at first sight that the Greek philosophers have by some kind of ingenious intuition come to the same or very similar conclusions as we have in modern times only after several centuries of hard labour with experiments and mathematics. This interpretation of our comparison would, however, be a complete misunderstanding. There is an enormous difference between modern science and Greek philosophy, and that is just the empiricist attitude of modern science.... This possibility of checking the correctness of a statement experimentally with very high precision and in any number of details gives an enormous weight to the statement that could not be attached to the statements of early Greek philosophy. All the same, some statements of ancient philosophy are rather near to those of modern science.’
It is important to stress, as Bohr particularly did, that science is a social human activity crucially dependent on communication among individuals. Each scientific theory is properly viewed as a human creation. Here is Yakov Zeldovich’s expression of this aspect: ‘Fundamental science is ...needed, among other things, because it satisfies man’s spiritual requirements. Scientific endeavour is a remarkable manifestation of human intellect. It perfects human intelligence and ennobles the soul’.

We have seen how difficult it is to give precise definitions of what is physically real; any statement reflects the state of knowledge at the time it is made, and may have to be revised later. From a philosophical stance, the importance of mathematics in physical science, and the changing ways in which it is used, are noteworthy. In the discussions about quantum mechanics we see the extreme care required in the use of language (not to mean, of course, that we can be careless in other realms!).

Again, from a philosophical standpoint, we see that pure empiricism and a purely deductive approach are both limited in scope. We need to combine caution, flexibility, and rigour – all at the same time. Nature is inexhaustible, and only experience hand in hand with reason can guide us to dependable knowledge. These seem to be the characteristics of a philosophy useful for the physical sciences.

**Suggested Reading**

Brief Life Sketches
Abdus Salam

N Mukunda

In his celebrated three-volume Lectures on Physics, Richard Feynman introduces the subject of electromagnetic radiation with these words:

“The most dramatic moments in the development of physics are those in which great syntheses take place, when phenomena which previously had appeared to be different are suddenly discovered to be but different aspects of the same thing. The history of physics is the history of such syntheses, and the basis of the success of physical science is mainly that we are able to synthesize.”

The first such synthesis – the most dramatic and the greatest leap of imagination ever taken by the human intellect – occurred in 1665 when Isaac Newton discovered his law of universal gravitation, and showed that terrestrial and celestial motions are governed by the same laws. Two centuries later, around 1865, James Clerk Maxwell succeeded in uniting electric and magnetic phenomena into a comprehensive electromagnetic theory, and deduced that light consists of electromagnetic waves. And then, another century later during the 1960’s, three physicists – Sheldon Lee Glashow, Abdus Salam and Steven Weinberg – achieved the third great synthesis, that of electromagnetism with the so-called weak nuclear force responsible for beta decay. They were awarded the Nobel Prize for physics in 1979 “for their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, including inter alia the prediction of the weak neutral current.”

While Newton’s and Maxwell’s achievements belong to the world of classical physics, this most recent synthesis is totally within the quantum domain.

Salam was born in January 1926 in Jhang, a small town in the erstwhile West Punjab of undivided India, into a family of modest means. His father rose from being a school teacher to head clerkship in a Government Office; and the entire family of parents, seven sons and two daughters lived in a one-room tenement. At school in Jhang, Salam was taught that gravitation and capillarity were the fundamental forces in nature and while electricity could be found in Lahore, the nuclear force was only in far-away Europe! (At this point one cannot help recalling that Glashow and Weinberg were contemporaries in the Bronx High School in New York, an institution with a good reputation in those days).

Starting from such humble beginnings, and fulfilling his father’s hopes and dreams, Salam went on to college in Jhang, and then obtained his M.A in mathematics at Punjab University in Lahore, performing brilliantly at every stage. In 1946, just before partition, he won a scholarship to study at Cambridge in England. Here he came into close contact with the legendary Paul Dirac, who became Salam’s (and so many others’) hero for the rest of his life. His thesis adviser was Nicholas Kemmer, and his collaborator of that period – Paul Mathews – became

his best friend in life. For his PhD, awarded in 1952, Salam solved an outstanding and very technical problem in the renormalization theory of quantum electrodynamics, that of handling the so-called 'overlapping divergences'. This work was of such a class that it already gained him an international reputation.

Salam returned to Pakistan in 1951 as Head of the Mathematics Department of his old college in Lahore. However there was no encouragement to pursue creative research and no peers to talk to. Then, in 1953, the anti-Ahmadiyyah riots – Salam was a devout Muslim and belonged to this sect – took place. This, and the indifference of the bureaucracy, compelled Salam to leave Pakistan for good in 1954; he spent the rest of his creative life in the West. For three years he stayed at Cambridge. Then in 1957 he moved to Imperial College in London as Professor of Physics, and put together an outstanding group working on problems of quantum field theory and elementary particle physics. In 1964 he founded the International Centre for Theoretical Physics at Trieste in Italy, and then divided his time between England and Italy. He retained links with Pakistan for some years, as a member of the Atomic Energy Commission and as Chief Scientific Adviser to the President.

After his landmark contribution to the renormalization theory of quantum electrodynamics, Salam extended these methods to the meson theory of nuclear forces. At the time of the parity revolution around 1956–57, he was among the first to conceive of a natural explanation for the lack of mirror reflection symmetry in nature, namely that it was due to the neutrino being a left handed or chiral particle. (The other genius to hit upon this particular idea was Lev Landau). However the senior physicist Wolfgang Pauli criticised Salam’s ideas so strongly that the latter withdrew his paper before publication! Prior to this, in 1954, Salam’s student Ronald Shaw at Cambridge had discovered the basic structure of non-abelian gauge field theories, a highly nontrivial generalization of Maxwell’s electromagnetism. Independently and at the same time, Yang and Mills also constructed this class of theories, now known by their names. As we know today, this is the leitmotif of all the fundamental interactions in nature. In the early 1960’s Yuval Ne’eman, another student of Salam, discovered the SU(3) symmetry of elementary particles at the same time as Murray Gell-Mann. All these instances show Salam as a world leader in elementary particle physics at that time.

In his work on the electro-weak unification, Salam brought together two key ingredients – nonabelian gauge symmetry and spontaneous symmetry breaking – to each of which he had earlier made crucial contributions. Later he turned his attention to the notoriously difficult problem of including gravity, and also built up the basic mathematical language for dealing with supersymmetry. Till his last years when a crippling illness overtook and finally claimed him in November 1996, Salam remained a source of inspiration and creativity to many.

Salam visited India several times. His Centre at Trieste was created to save Third World scientists from the isolation and deprivation that he had faced in his youth. And he understood deeply and identified with the problems of growing science in the developing world. For all this, as much as for his great achievements in physics, Salam will always be remembered with admiration and gratitude.
About Wolfgang Pauli*

N Mukunda

Among the handful of names that appears across the canvas of twentieth century theoretical physics, that of Wolfgang Pauli is one of the most prominent. The century had begun with Max Planck’s discovery of his radiation law and quantum of action in October 1900; it was also the year of Pauli’s birth. The adjustment of the fabric of physics to accommodate the new constant of nature occupied the next quarter century. It resulted in the new quantum mechanics which, unlike the relativity theories, was the work of many hands. The years 1901 and 1902 saw the birth of many of those destined to give shape to the new physics: Heisenberg and Fermi in 1901, Wigner and Dirac in 1902. The other principal characters – Einstein (1879), Max Born (1882), Niels Bohr (1885), Erwin Schrödinger (1887) and Louis de Broglie (1892) – all belonged more or less to an earlier generation.

Wolfgang Pauli (jr) was born on April 25, 1900 in Vienna. His father Wolfgang Josef Pauli, a medical doctor and later a University Professor of biology, had converted from the Jewish to the Catholic faith. His mother Berta Camilla Schutz was a Catholic, his sister Hertha was nine years younger. He was baptized in the Catholic church, his godfather being the physicist – philosopher Ernst Mach. Pauli learnt an incredible amount of physics and mathematics in the school years, graduating from the Dobling Gymnasium in Vienna in 1918.

He then joined Arnold Sommerfeld as a student in Munich, and obtained his Ph.D in 1921 for a thesis showing the failure of Bohr’s atomic model for the ionized hydrogen molecule. On the way, barely 21, at Sommerfeld’s invitation he completed a masterly encyclopedia article on relativity, to replace one that Einstein had agreed to but did not write.

Pauli spent winter 1921–22 with Max Born at Göttingen, and spring 1922 with Wilhelm Lenz at Hamburg. In summer 1922 he met Niels Bohr at the latter’s lectures at Göttingen; this led to his spending 1922–23 at Copenhagen. In early 1924 he became an Assistant Professor at Hamburg. The period 1924–27 saw many extended visits to Copenhagen – this was when quantum mechanics was discovered and its orthodox or Copenhagen interpretation created by Bohr, Heisenberg and Pauli. In April 1928 Pauli moved as Professor to the Swiss Federal Polytechnic Institute (ETH) in Zurich. He stayed there for the rest of his career, except for the war years 1940–46 spent at the Institute for Advanced Study at Princeton. The 1945 Physics Nobel Prize was awarded to Pauli ‘for the discovery of the Exclusion Principle, also called the Pauli Principle’.

In 1929 Pauli renounced the Catholic faith. His first marriage in 1930 failed after just a year, leading to deep depression during 1931–32. He then married Franciska Bertram on April 4, 1934.

N Mukunda

Pauli’s landmark achievements in physics are many. Both the exclusion principle and the discovery of a new two-valued variable for the electron date from late 1924 – early 1925. The latter came from a penetrating analysis of the anomalous Zeeman effect. Years later he recalled:

“A colleague who met me strolling rather aimlessly in the beautiful streets of Copenhagen said to me in a friendly manner, ‘You look very unhappy’; where-upon I answered fiercely ‘How can one look happy when he is thinking about the anomalous Zeeman effect?’

His description of the new variable associated with the electron reads:

“According to this point of view the doublet structure of alkali spectra as well as the deviation from Larmor’s theorem is due to a particular two-valuedness of the quantum theoretic properties of the electron, which cannot be described from the classical point of view.”

In late 1925 he applied Heisenberg’s new matrix mechanics to the hydrogen atom, beating Dirac by a few days, and showed that it worked. In 1927 he showed how to extend the (non-relativistic) quantum mechanics to accommodate the spin of the electron. Along with Heisenberg, and following upon Dirac’s earlier work, he created quantum field theory in 1928. In 1930 he conceived of the neutrino as a new hitherto unknown particle in nature, as a means of saving the principle of energy conservation in beta decay. His famous letter containing this idea appears elsewhere in this issue. As Leon Lederman said decades later: “Obviously, a particle that reacted with nothing could never be detected. It would be a fiction. The neutrino is just barely a fact.”

Pauli’s 1927 work on the paramagnetism of an electron gas inaugurated solid state physics, though he said: “I don’t like this solid state physics ... I initiated it though”. In 1930 with Weisskopf he showed how to make sense of the Klein–Gordon equation in quantum field theory, and in 1940 came the spin-statistics theorem. In the 1940’s at Princeton he worked on the strong coupling theory, and in 1955 came the CPT theorem.

Pauli was a severe critic, not least of himself, and an exceedingly sharp mind. Quoting Heisenberg: “Pauli had a very strong influence on me. I mean Pauli was simply a very strong personality ... He was extremely critical. I don’t know how frequently he told me, ‘You are a complete fool’, and so on. That helped a lot.” His sharp criticisms stopped many from pursuing or publishing speculative ideas – Kronig on electron spin, de Broglie on alternative interpretations of quantum mechanics, Salam on the two-component neutrino, to mention a few. He once said about a manuscript: “It is not even wrong”. He was however completely honest when he said: “I have indeed mistakenly held something right to be wrong, but never considered something wrong to be right”.

In time, Pauli became the “conscience of physics”, playing successor to Einstein in this role. In a 1960 tribute, no less a person than Niels Bohr wrote: “At the same time as the anecdotes about his personality grew into a veritable legend, he more and more became the
About Wolfgang Pauli

very conscience of the community of theoretical physicists”. His critical nature helped him foresee difficulties exceedingly quickly, and probably held back his own creativity. Tomonaga has characterised the situation thus: “... Pauli’s perfectionism was phenomenal. He not only wanted to be perfect himself, but was well known for applying this standard to other people, sometimes arguing harshly against other people’s work. For this reason when people asked his opinion of their work and he gave his assent, they called it Pauli’s sanction”.

Pauli ventured to express his philosophical views on physics – a long standing and deep interest – late in life, in an extensive correspondence with Markus Fierz. These letters, kept at CERN in Geneva, are even now being studied and assessed. In the interpretation of quantum mechanics he was Bohr’s closest pupil, and he had the courage to carry this interpretation to its logical end.

During a lecture at Zurich in early December 1958 he suddenly took ill due to intensive pains, and died ten days later on 15 December 1958.
Paul Adrien Maurice Dirac – An Appreciation*

N Mukunda

Of the two major revolutions in physics in the twentieth century – the relativity theories and quantum theory – in many ways the latter was the more profound. Its discovery was the work of many hands, an essentially European effort, and the entire development lasted a quarter of a century. It consisted of two distinct phases – the period of the Old Quantum Theory, from 1900 up to about 1923–24, was the first phase, with the major contributions coming from Planck, Einstein, Bohr and Sommerfeld. The second phase was inaugurated by Heisenberg’s discovery of matrix mechanics in summer 1925 and culminated in the completion of quantum mechanics more or less as we know it today by 1926–27. In this phase the dominating figures were Heisenberg, Dirac, Schrödinger, Pauli, de Broglie and Max Born. In particular the actual creation of quantum mechanics was the result of the work of the trio – Heisenberg, Dirac and Schrödinger. Thus Dirac belongs to the period of the construction and consolidation of the new quantum mechanics.

August 2004 issue of Resonance celebrates the life and work of Dirac, through this life-sketch and several articles describing his many remarkable achievements.

Paul Adrien Maurice Dirac – hereafter Dirac – was born on 8 August 1902 in Bristol in England. His father Charles Adrien Ladislas Dirac was originally from Monthey in Switzerland, and had come to Bristol in the early 1890’s to work as a French tutor, later becoming French lecturer at the Merchant Venturers Technical College. His mother Florence Hannah (née Holten) came from Liskeard in Cornwall in England. Dirac had two siblings, an elder brother Reginald Charles Felix born in 1900 and a younger sister Beatrice Isabelle Marguerite born in 1906.

Dirac attended elementary and secondary schools in Bristol and in 1918 became a student of electrical engineering at the University. In 1921 he obtained the BSc degree, and also competed for a scholarship to enter St.Johns College in Cambridge. He won an Open Exhibition but due to lack of finances could not take it up. He could not find a job as an electrical engineer, but was permitted by the Mathematics Department at Bristol University to attend lectures there without payment of fees. Finally two years later in 1923 he was awarded a research studentship which enabled him to go to St.Johns College in Cambridge. The association with St.John’s continued for the rest of his life.

At Cambridge Dirac worked under the guidance of R H Fowler (son-in-law of Ernest Rutherford), initially on problems of relativistic statistical mechanics and the old quantum theory. (Incidentally Fowler was PhD guide for Homi Bhabha and Subrahmanyan Chandrasekhar later on). His major break came in August 1925. Heisenberg had visited Cambridge in July

1925 and told Fowler about his work earlier that summer. After his return to Göttingen, Heisenberg sent Fowler a proof copy of his work on matrix mechanics, which Fowler passed on to Dirac. While initially unimpressed by Heisenberg’s work, after a few weeks Dirac realized that Heisenberg had actually solved the problem of the construction of a new mechanics. In particular Dirac saw that the failure of commutativity of multiplication, which Heisenberg had regarded as a possible flaw in his theory, was in fact vitally important and a key feature of the new theory. Ignited by this spark, Dirac went on to create his own version of the new quantum mechanics with many original and essential ideas. His discovery of the link between commutators in quantum mechanics and Poisson brackets in classical mechanics is a beautiful expression of Bohr’s Correspondence Principle and gave particular pleasure to Dirac. His explorations into the structure of the new quantum mechanics led to the concepts of $q$-numbers and $c$-numbers, the development of the transformation theory, the invention and use of the delta function, the analysis of symmetry properties of wave functions for identical particles and the link to their Bose or Fermi statistics, and somewhat later on the invention of the famous bra and ket notation. All these and more have become part and parcel of every physicist’s training since several decades.

I will return to a survey of some of Dirac’s many original ideas and achievements later, after a brief look at important events in his life.

Dirac’s 1926 PhD thesis of about 140 pages was titled ‘Quantum Mechanics’ and consisted of his own remarkable contributions to its foundations. He was elected Fellow of St.Johns College in 1927. In 1932 he succeeded Joseph Larmor as Lucasian Professor of Mathematics, a position he held till 1969. He was elected Fellow of the Royal Society in 1930. In 1933 Dirac shared the Physics Nobel Prize with Schrödinger, being cited for his ‘discovery of new fertile forms of the theory of atoms and for its applications’. Soon after completing the PhD, Dirac made important visits to the Bohr Institute in Copenhagen, as well as to Max Born’s school in Göttingen. In the ensuing years he made many trips to the erstwhile Soviet Union, to the USA and to Japan. In 1954–55 he spent several months at the Tata Institute of Fundamental Research in Bombay at the invitation of Homi Bhabha.

In 1937 Dirac married Margit Balasz (née Wigner), sister of the well-known physicist Eugene Paul Wigner. They had two daughters, Mary Elizabeth born in 1940 and Florence Monica born in 1942. After retirement from Cambridge in 1969, he moved to Florida, USA becoming Research Professor at Florida State University, Tallahassee in 1971. He was awarded the first Oppenheimer Prize in 1969, and the Order of Merit in 1973.

Dirac’s relationship with his father was not a happy one. At home Charles strictly enforced the rule that the children should only speak in French or not at all. Dirac took the option of speaking very little - “so I became very silent at that time, that started very early” – a trait he retained throughout his life. He was always a man of very few but well-chosen words. His brother Reginald committed suicide in 1924. And as his biographer Kragh records: “He could not forget the traumatic experiences of his childhood, for which he blamed his father, with whom he came to want to have as little contact as possible”. When his father died in 1936 Dirac wrote to Margit “I feel much freer now”.

N Mukunda
Paul Adrien Maurice Dirac – An Appreciation

Coming back to his everlasting contributions to physics, he inaugurated quantum field theory in 1927 by performing the quantisation of the electromagnetic field. This brought to a conclusion the glorious chapter of modern physics starting with Planck’s discovery of the radiation law in 1900, Einstein’s photon concept of 1905, Einstein’s A and B coefficients and the ideas of spontaneous and stimulated emission and absorption of radiation by matter from 1916, and finally Bose’s discovery of the new quantum statistics in 1924. In the next year, 1928, Dirac discovered the relativistic wave equation for the electron. The key idea was the introduction of spinors into physics, and it led to the first successful union of special relativity with quantum mechanics. Tomonaga records that the basic ideas came to Dirac while he was staring at a fire place at St.John’s College! This discovery led to the explanation of many puzzling features of the electron – its intrinsic spin, its magnetic moment, the fine structure of the spectrum of hydrogen, and finally the prediction of the positron and the general concept of antimatter. In 1930 he completed his masterly exposition of the new mechanics titled *Principles of Quantum Mechanics* which has become a classic. It appears that Einstein always carried around a copy, and he spoke admiringly of “Dirac to whom in my opinion we owe the most logically perfect presentation of quantum mechanics”. This book was revised four times, with Dirac making important improvements of notation in successive editions.

His work on the prediction of the positron and the remarkable properties of magnetic monopoles in the framework of quantum mechanics dates from 1931. In 1933 he examined the role of the Lagrangian of classical mechanics in the quantum context. This seminal paper of Dirac inspired Feynman a few years later to create his own Path Integral formulation of quantum mechanics, a third form comparable in significance to Heisenberg’s original matrix mechanics and Schrödinger’s later wave mechanics. In the mid 30’s he did some work on wave equations for particles of higher spin. The paper on the Large Numbers hypothesis dates from 1937, and the work on the classical theory of the electron from 1938. In 1942 he suggested the use of an indefinite metric in Hilbert space of quantum mechanics and in 1945 he inaugurated the field of unitary representations of noncompact Lie groups by working out such representations for the Lorentz group. This was carried forward by his student Harish-Chandra from India.

Another major contribution of Dirac is to the classical mechanics of so-called singular systems. These are cases where the usual textbook recipes for passing from the Lagrangian to a Hamiltonian formulation, as a prelude to quantisation, fail. To handle such systems Dirac created concepts and methods of great elegance and depth. He applied them himself to the case of Einstein’s general relativity. These methods were first presented in a series of lectures at the Canadian Mathematical Congress in Vancouver in 1949, and later appeared as a beautiful book of lectures given at Yeshiva University in 1964.

There are several other important ideas of Dirac one might recount, but it is time to say something about his personality. Dirac was a shy and genuinely modest person who readily acknowledged his debt to others. Bohr said of him: “of all physicists, Dirac has the purest soul”. While introducing Heisenberg at a lecture in 1972 he said that they were both research
students at the same time working on the same problem, and Heisenberg succeeded where he himself failed.

Dirac was a great believer in beauty in the mathematical formulation of physical laws. In a 1939 lecture he said: “Mathematical beauty cannot be defined any more than beauty in art can be defined, but which people who study mathematics usually have no difficulty in appreciating”. And on another occasion: “It seems that if one is working from the point of view of getting beauty in one’s equations, and if one has a really sound insight, one is on a sure line of progress”.

In his younger days Dirac was quite atheistic in his thinking. During a conversation in 1927 involving Dirac, Heisenberg, Pauli and others, Dirac is reported to have said: “we must admit that religion is a jumble of false assertions, with no basis in reality. The very idea of God is a product of the human imagination”. To which Pauli retorted: “There is no God and Dirac is his prophet”. But he presumably changed his attitude somewhat in later years, going by a letter from Margit to Abraham Pais, in which she said: “Paul was no atheist. Many times did we kneel side by side in chapel, praying. We all know he was no hypocrite”. Dirac’s lifestyle was ascetic, he was indifferent to discomfort or food, and neither smoked nor touched alcohol – making N F Mott compare him to Gandhi.

Dirac had a great insight into nature’s laws, and an unequalled combination of mathematical inventiveness and elegance in expression. On the occasion of Dirac’s 80th birthday Harish-Chandra wrote: “I have often pondered over the roles of knowledge or experience on the one hand and imagination or intuition on the other, in the process of discovery. I believe that there is a certain fundamental conflict between the two, and knowledge, by advocating caution, tends to inhibit the flight of imagination. Therefore, a certain naïveté, unburdened by conventional wisdom, can sometimes be a positive asset. I regard Dirac’s discovery of the relativistic equation of the electron as a shining example of such a case.... Within a few years, nature conferred her approval on Dirac’s marvelous insight by the discovery of the positron”.

Dirac died in Florida on 20 October 1984. That was the passing of the last of the founders of quantum mechanics, who in his life had shown, again in Harish-Chandra’s words, “... profound originality of thought and purity and gentleness of spirit”.

N Mukunda
It is said, only partly in jest, that while the French regard Joseph Louis Lagrange as one of their greatest figures, the Italians claim Giuseppe Luigi Lagrangia for their own. Similarly, both mathematicians and physicists think of Lagrange as a towering figure in their disciplines. He belonged to an age of such eminent personalities – some others of comparable distinction being Euler, Gauss, Hamilton and much later Poincaré.

Lagrange’s most important contributions to physics are in the area of (classical non-relativistic) mechanics. The fundamental physical concepts and laws were developed over the 17th century by Galileo followed by Newton - the principle of inertia, the clear understanding of acceleration and force, and the putting together – by Newton’s genius – of the three laws of motion. Over the succeeding century and more, great advances were made by Euler, Lagrange and others in expressing the physics of Galileo and Newton in successively more powerful mathematical forms. In the process, while the infinitesimal calculus of Newton and Leibnitz was greatly developed as the perfect language of mechanics, the extension to the partial differential calculus took place, and between them Euler and Lagrange created the beautiful calculus of variations.

The developments in mechanics most intimately linked with Lagrange are the introduction of generalized coordinates $q$, the role of a single function of coordinates and velocities – the Lagrangian $L(q, \dot{q})$ – in determining all the equations of motion as it were at one stroke, and the principle of least action which reached its most convenient and flexible form in Hamilton’s work. So one speaks of the Newtonian, Lagrangian and Hamiltonian forms of mechanics – always the same physical ideas but expressed with increasing mathematical sophistication and power. The Euler-Lagrange form of the equations of motion have the appearance

$$\frac{d}{dt} \frac{\partial L(q, \dot{q})}{\partial \dot{q}} - \frac{\partial L(q, \dot{q})}{\partial q} = 0.$$

The importance of Lagrange’s method can be appreciated in the following manner. Suppose one starts with the Newtonian equations of motion for a system of interacting particles written, say, in Cartesian position vectors, velocities and accelerations. If one wishes to re-express these equations using some other more convenient configuration space variables (such as spherical polar coordinates for each particle), the brute force method would be to express each Cartesian variable in terms of the new ones and their time derivatives, and then substitute in the equations of motion. But this is invariably very cumbersome. If the equations of motion are derivable from a single Lagrange function, a vastly more efficient and economical procedure is to simply express the Lagrangian in terms of the new variables, and then proceed as

---

before. Quite generally the process of passing from one system of generalized coordinates to another is greatly simplified as indicated in this ‘commutative diagram’:

\[
L(q, \dot{q}) \rightarrow \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0
\]

\[
\downarrow \quad \downarrow
\]

\[
L'(q', \dot{q}') \rightarrow \frac{d}{dt} \frac{\partial L'}{\partial \dot{q}'} - \frac{\partial L'}{\partial q'} = 0
\]

Step (1) and followed by Step (2) is much easier than the direct step (3)!

The Lagrange point of view gives such a unified and concise overall view of a physical system – its constitution, description and dynamics – that in fundamental physics one always refers to any specific theory or system by speaking of its Lagrangian. Thus one speaks of the Lagrangian for conservative multi particle systems, for the Maxwell electromagnetic field, for particle mechanics in special relativity, for general relativity, etc. It works for both discrete particle systems and for continuous fields. It also makes the discussion of continuous symmetries very direct and elegant: symmetries of the Lagrangian imply invariances of the equations of motion, and in turn laws of conservation via Noether’s Theorem.

In his work on perturbation theory in mechanics, Lagrange introduced an expression later called the Lagrange Bracket. In this and in other ways, he anticipated many formal features of the later canonical formalism of mechanics due to Hamilton and Jacobi.

A few words should be added concerning Lagrange and quantum mechanics! As students know, quantum mechanics initially grew out of the canonical phase space form – the Hamiltonian form – of classical mechanics. The key classical notions of Poisson brackets and canonical transformations found their counterparts in commutation relations and unitary transformations. As we saw earlier, the Hamiltonian form of classical mechanics had grown out of the preceding Lagrangian form. In 1933 Dirac asked the profound question: does the Lagrangian have any direct role to play in the formulation of quantum mechanics? He did show that there were formal analogies between certain classical expressions involving the Lagrangian, and quantities describing unitary transformations in quantum mechanics. Some years later Feynman carried Dirac’s idea to completion and created the Path Integral formulation of quantum mechanics in which the Lagrangian – more precisely its time integral (the action) – is the principal ingredient.
Max Planck – Founder of Quantum Theory*

N Mukunda

Max Karl Ernst Ludwig Planck – in short, Planck – is a towering figure of modern physics, the father of the quantum theory. He was born a century and a half ago, on 23 April, 1858, in Kiel in Germany where his father was Professor of Constitutional Law at the University. At age 17, he entered the University of Munich to study Physics, even though advised that all important work in the subject had been completed. In 1878, he moved to the University of Berlin as a student of Gustav Kirchoff, also attending lectures by Hermann von Helmholtz and Karl Weierstrass. In view of later developments, Kirchoff is aptly called the grandfather of the quantum theory.

Planck obtained his doctorate degree - ‘summa cum laude’ – from the University of Munich in 1879, the year of birth of Albert Einstein. He then continued as Privatdozent there. In 1885 he became Professor at Kiel, and in 1889 Professor Extraordinarius at Berlin. Finally in 1892, he succeeded his teacher Kirchoff as Full Professor at Berlin. In 1912 he became Permanent Secretary of the Prussian Academy of Sciences, and was awarded the Physics Nobel Prize in 1918. In 1926 he retired from Berlin, becoming Professor Emeritus. He was succeeded by Erwin Schrödinger, just fresh from his creation of wave mechanics while in Zurich.

The first Max Planck medal of the German Physical Society was awarded jointly to Planck and Einstein in 1929. In 1930, Planck became the President of the Kaiser Wilhelm Society for Advancement of Science, the highest academic position in Germany. Many years later, over 1946–48, this was transformed into the Max Planck Society.

Planck died in 1947, close to ninety.

February 2008 issue of Resonance carries two excellent general articles, by G S Ranganath and M Harbola, describing in pedagogical fashion the background and the details of Planck’s monumental work of 1900, its consequences and later developments in the hands of Einstein, Niels Bohr and others. Just a few added comments will suffice here. Planck took up the study of black body radiation – the search for Kirchoff’s universal function – in 1894. For some time, he evidently believed that a formula guessed by Wilhelm Wien in 1896 was exact. Then, on Sunday 7th October, 1900 (incidentally the fifteenth birthday of Bohr), the experimental physicist Heinrich Rubens and his wife visited the Plancks for tea. Rubens told Planck about the failure of the Wien formula for long wavelengths, where another classical formula of Lord Rayleigh and James Jeans worked better. After the Rubens left, Planck set to work to find a formula interpolating between the Wien and the Rayleigh–Jeans expressions, and then arrived at his famous radiation law.

Thus quantum theory was born in the space of a few hours on a Sunday evening. Ranganath mentions that Planck sent a postcard to Rubens with his result the same evening, and it was received the next morning – the postal system must have been remarkably efficient then. This result was presented at a meeting of the German Physical Society on Tuesday, 9th October, 1900. Then as he recalled later: “On the very day when I formulated the radiation law, I began to devote myself to the task of investing it with true physical meaning”. Two months later, at another German Physical Society meeting on Friday 14th, December 1900, Planck was able to present a derivation incorporating his revolutionary idea of quantization of energy. In his own words, this was “an act of desperation… I had to obtain a positive result, under any circumstances and at whatever cost.” As Abraham Pais wrote: “His reasoning was mad, but his madness has that divine quality that only the greatest transitional figures can bring to science”. Bohr expressed it in this way: “Scarcely any other discovery in the history of science has produced such extraordinary results within the short span of our generation as those which have directly arisen from Max Planck’s discovery of the elementary quantum of action”.

Planck has been called a ‘reluctant revolutionary’. He was a transitional figure heralding the passage from a classical past to the modern era. How happy that this happened in the closing months of the 19th century! In a letter to a colleague he said: “By nature I am peacefully inclined and reject all doubtful adventures…”. While he was an enthusiastic supporter of Einstein’s Special Theory of Relativity right from the beginning, he was for long very skeptical of Einstein’s light quantum concept. Indeed he wrote to Einstein in 1907: “I do not search for the meaning of the elementary quantum of action in the vacuum but at the points of absorption and emission, and I assume that processes in the vacuum are exactly represented by the Maxwell equations”. Going even further, when in 1913 he – along with Nernst, Rubens and Warburg – recommended Einstein for election to the Prussian Academy of Sciences, he wrote : “That he may sometimes have missed the target in his speculations, as, for example, in his hypothesis of light quanta, cannot really be held too much against him, for it is not possible to introduce really new ideas even in the most exact sciences without sometimes taking a risk”.

Quoting from a brief essay by James Murphy, Planck regarded “…physical science as part of human culture, forming an integral part with the other branches of human learning and exercising its influence on the destiny of humanity not merely in a material way but even more deeply in a spiritual way”.

In his personal life Planck faced repeated tragedies. His first wife Marie died in 1909. All their children died in tragic circumstances within Planck’s life time: the elder son Karl of wounds in World War I; the younger son Erwin executed by the Nazis in 1945; and their twin daughters Emma and Grete both in childbirth. His son Hermann by his second wife Marga was ‘mentally challenged’. Planck lived through two World Wars and the Nazi nightmare; in William Cropper’s words “He deplored everything the Nazis did, but chose to remain in Germany, with the hope that he could help pick up the pieces after it was all over”. No wonder that late in life he wrote these sadly profound yet inspiring words:
Max Planck – Founder of Quantum Theory

“For no man is born with a legal claim to happiness, success and prosperity in life. We must therefore accept every favourable decision of Providence, each single hour of happiness, as an unearned gift, one that imposes an obligation. The only thing that we may claim for our own with absolute assurance, the greatest good that no power in the world can take from us, and that can give us more permanent happiness than anything else, is integrity of soul, which manifests itself in a conscientious performance of ones duty.”
Israel Moiseevich Gelfand – An Appreciation from a Physicist’s Viewpoint

N Mukunda

The erstwhile Soviet Union – which collapsed around 1990 – had by all accounts set up an excellent framework for spotting and training really talented young people in the sciences and mathematics from all parts of its enormous territory. As part of this system, many leading scientists and mathematicians also were gifted writers of textbooks and monographs at advanced levels. In particular, many ‘pure’ mathematicians wrote books in an extremely lucid manner, always keeping in mind readers from the physics and engineering communities. This style of exposition of mathematics, consciously making it accessible to users who may not themselves have been professional mathematicians, seems to have been carefully cultivated. Many classic texts of that era were available at very affordable prices, often brought out by Mir Publishers.

Israel Moiseevich Gelfand was a leading figure in Soviet mathematics in the 20th century, and contributed enormously in the manner described above. The vast range of his work is explained in V S Sunder’s article in the February 2011 issue. One aspect of it is worth mentioning as it has in a sense an ‘Indian connection’, and that is his work with M A Naimark on the unitary representations of the Lorentz group.

The homogeneous Lorentz group $SO(3,1)$ is fundamental to special relativity; it describes the transformations between the inertial frames of physics, familiar to all students of physics. It is made up of special Lorentz transformations, rotations in three-dimensional space, and their combinations. This group is basic to both classical relativistic physics, and relativistic quantum mechanics.

Many of the familiar relativistic physical quantities belong to various finite dimensional representations of the Lorentz group, and these are necessarily non-unitary. (Unitarity is a concept which every student of quantum mechanics learns quite early on). Examples are space-time position four-vectors of point particles, as well as their energy-momentum four-vectors; the Maxwell electromagnetic field strengths, the relativistic energy-momentum tensor of general field theories; etc.

The famous Dirac relativistic wave equation for the electron, discovered in 1928, also involves quantities belonging to another finite (in fact four) dimensional representation of the Lorentz group, namely the Dirac spinor representation.

In 1945, in a fundamental paper, Dirac constructed some new infinite-dimensional unitary representations of the Lorentz group, and suggested that they may be important for physics. This idea itself was new both within physics and within mathematics. Just around that time,

N Mukunda

Harish-Chandra joined him as a PhD student in Cambridge. Harish-Chandra had completed his MSc in physics from the University of Allahabad in 1943 and then moved to the Indian Institute of Science in Bangalore as a research student. In the period 1943–1945 he did a considerable amount of work in collaboration with Homi Bhabha. His being accepted as a PhD student by Dirac was facilitated by strong recommendations by Bhabha and K S Krishnan who had taught Harish-Chandra in Allahabad.

Dirac proposed to Harish-Chandra that he study the problem of constructing all the unitary (irreducible) representations of the Lorentz group. This challenging problem was solved by Harish-Chandra during 1945–1947, leading to two landmark papers in 1947. As it happened, this same problem was also solved essentially simultaneously and independently by Gelfand and M A Naimark in the Soviet Union, and published by them also in 1947.

Later developments in physics have not made use of these representations of the Lorentz group to any significant extent, but in any case the credit for constructing these goes both to Harish-Chandra and Gelfand and Naimark. All this - starting with Dirac’s 1945 work – is justifiably regarded as the origin of a beautiful chapter in modern mathematics.

Going back to the opening paragraph of this article, for the benefit of younger readers it may be useful to mention some of the outstanding books by Gelfand and his collaborators, and by others belonging to the old Soviet School in mathematics as well as in physics. Here is a partial list:

**Suggested Reading**

1. I M Gelfand, R A Minlos and Z Ya Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications*.
5. L D Landau and E M Lifshitz, *Course of Theoretical Physics*, Volumes I – VII.

Each of these, and many others, will attest to the high pedagogical skills and clarity achieved by the authors of that era.
Richard Phillips Feynman – Physicist and Teacher Extraordinary*

N Mukunda

The first three decades of the twentieth century have been among the most momentous in the history of physics. The first saw the appearance of special relativity and the birth of quantum theory; the second the creation of general relativity. And in the third, quantum mechanics proper was discovered. These developments shaped the progress of fundamental physics for the rest of the century and beyond. While the two relativity theories were largely the creation of Albert Einstein, the quantum revolution took much more time and involved about a dozen of the most creative minds of a couple of generations.

Of all those who contributed to the consolidation and extension of the quantum ideas in later decades – now from the USA as much as from Europe and elsewhere – it is generally agreed that Richard Phillips Feynman was the most gifted, brilliant and intuitive genius out of many extremely gifted physicists. Here are descriptions of him by leading physicists of his own, and older as well as younger generations:

“*He is a second Dirac, only this time more human.*” – Eugene Wigner

…Feynman was not an ordinary genius but a magician, that is one “who does things that nobody else could ever do and that seem completely unexpected.” – Hans Bethe

“…an honest man, the outstanding intuitionist of our age and a prime example of what may lie in store for anyone who dares to follow the beat of a different drum.” – Julian Schwinger

“…the most original mind of his generation.” – Freeman Dyson

Richard Feynman was born on 11 May 1918 in Far Rockaway near New York to Jewish parents Lucille Phillips and Melville Feynman. His physics teacher at High School was Abram Bader, who saw Feynman’s exceptional talents and in a special lecture explained to him the beautiful principle of least action in classical mechanics (see chapter 19, Vol. 2, *Feynman Lectures on Physics*). After graduating from MIT Feynman joined Princeton University in 1939 as a graduate student under the guidance of John Archibald Wheeler. They initially worked on extending a 1938 paper of Paul Dirac on the classical theory of radiating electrons in an action-at-a-distance framework. Dirac had always been Feynman’s hero, and for his thesis Feynman turned again to Dirac. It was a conversation with Herbert Jehle in 1941 that led Feynman to a seminal 1933 paper of Dirac in which he had drawn attention to an analogy between the classical action principle, and certain basic expressions in quantum mechanics. In Feynman’s hands this analogy was sharpened to an equality, thus leading to his own new path integral approach to quantum mechanics. This, by the way, is yet to achieve an acceptable mathematical form.


257
The war years interrupted the efforts of both Feynman and Schwinger to tackle the divergence problems in quantum electrodynamics, another of Dirac’s pioneering creations from 1927. In 1965 the Physics Nobel Prize was shared by the two of them and Sin-Ichiro Tomonaga from Japan for their development of renormalization theory. Here is a fine passage from Feynman’s prize acceptance speech:

“\textit{Theories of the known, which are described by different physical ideas, may be equivalent in all their predictions and are hence scientifically indistinguishable. However, they are not psychologically identical when trying to move from that base into the unknown.”}

Feynman also did pioneering work in elementary particle physics and in the quantum theory of gravitation. He was a brilliant teacher and lectured extensively to both scientific and general audiences. While he was certainly flamboyant and charismatic, he was not a little above wanting and seeking attention to himself. For the benefit of our younger readers here is a list of many of his books on specialised as well as other topics, several published after his passing:

a) \textit{Quantum Mechanics and Path Integrals} (with A R Hibbs) – 1965.
b) \textit{The Character of Physical Law} – 1965.
e) \textit{Surely You’re Joking, Mr. Feynman} - 1988.
f) \textit{What Do You Care What Other People Think?} – 1988.

To these should be added his lecture notes on Statistical Mechanics, Photon Hadron Interactions, QED, and Theory of Fundamental Processes. His scientific research papers, it may be noted, number not much more than about fifty! In fact in 1945 at an early stage in his career he said to Schwinger: “I haven’t done anything, but you’ve already got your name on something”.

Feynman was married thrice, and had a son and a daughter. He had a serious interest in art, learning painting and sketching from a professional artist friend. After suffering for many years, he finally succumbed to cancer on 15 February 1988. Two extensive biographies of him have been written: \textit{Genius – Richard Feynman and Modern Physics} by James Gleick (1992) and \textit{The Beat of a Different Drum – The Life and Science of Richard Feynman} by Jagdish Mehra (1994).
Arnold Sommerfeld – Physicist and Teacher Beyond Compare*

N Mukunda

Introduction - Gymnasium and University in Königsberg

The German tradition in physics – as in music, philosophy and literature – has been an exceptional one over a very long period of time. Recall the likes of Gustav Kirchhoff, Hermann von Helmholtz, Heinrich Hertz and Wilhelm Röntgen during the 19th century; to be followed by Max Planck, Wilhelm Wien, Albert Einstein, Max von Laue into the 20th. But even in such exalted company Arnold Sommerfeld occupies a very special place, as the father of theoretical physics in Germany, and arguably the greatest teacher the world of physics has ever seen. He combined excellence in research, teaching, mentoring and scholarship to a degree never equalled by any other. To do justice to such a personality in a few pages is not at all easy.

Arnold Johannes Wilhelm Sommerfeld – hereafter Sommerfeld – was born on December 5, 1868, to Cäcile Mathias and Franz Sommerfeld, a medical doctor, in Königsberg, then in Prussia. This was the city where the philosopher Immanuel Kant (1724–1804) was born and lived all his life; and also where the great mathematician David Hilbert (1862–1943), who straddled both the 19th and 20th centuries, was born. In 1946, after World War II, this city became Kaliningrad in the Soviet Union, now Russia. It is good to remember the years of birth of some other physicists and mathematicians of those and slightly later times with whom Sommerfeld interacted so closely over an entire life time: Felix Klein in 1849, Hendrik Lorentz in 1853, Max Planck in 1858, Hermann Minkowski and Wilhelm Wien in 1864, Albert Einstein in 1879, Max Born in 1882, Niels Bohr and Hermann Weyl in 1885, and Erwin Schrödinger in 1887.

In school (gymnasium) in Königsberg from 1875 to 1886, Sommerfeld excelled in both the sciences and humanities. He wrote much later:

“To my energetic and intellectually vigorous mother I owe an infinite debt.”

At the University in Königsberg from 1886 to 1891, he was taught by the mathematicians Hilbert, Hurwitz and Lindemann; and completed a PhD under Lindemann’s guidance on ‘The arbitrary functions in mathematical physics’. This was exposure to the finest mathematics of the time.

Göttingen – Clausthal – Aachen

After another year for a teaching diploma, and then a year of military service, he moved in 1893 to Göttingen, ‘the seat of high mathematical culture’. In 1894 he became assistant to Felix Klein, of whom he wrote:

“Overwhelming was the impression which I received, in lectures and discussions, from Felix Klein’s grand personality … I have always regarded Klein as my real teacher, not only in things mathematical, but also in mathematical physics and in connection with mechanics”.

It was in 1895 with Klein’s guidance that Sommerfeld did his first truly outstanding piece of research – the problem of diffraction of light by a straight edge. (This work is described in an article elsewhere in Resonance May 2015 issue.) He also assisted Klein during 1895–1896 in organizing the latter’s lectures on tops and gyroscopes – this resulted later in the 4-volume work *Theorie der Kreisels* by the two of them completed in 1909–1910. At Klein’s request, Sommerfeld became editor of Volume V of the ‘Encyclopaedie der mathematischen Wissenschaften’ from 1898 to 1926.

From Göttingen, Sommerfeld moved in 1897 to Clausthal-Zellerfeld as Professor of Mathematics at the Bergakademie (mining academy). He was to stay at Clausthal till 1900. In this period he married Johanna Hopfner; they had three sons and a daughter.

The years 1900–1906 were spent at the Technical University at Aachen, as Professor of Applied Mechanics. In these years he took up work in hydrodynamics, the field in which much later his students Ludwig Hopf and Werner Heisenberg would both write their doctoral theses.

The Munich Years – Teacher Par Excellence

Finally in 1906 he came to the University of Munich as Professor of Theoretical Physics, a position for which he was selected by Röntgen and Lorentz. Here he stayed for the rest of his working life, as Director of his own Institute, until retirement in 1938. It is incredible to read of his weekly load at Munich: four hours of basic courses, two hours of specialized classes, a seminar and a Colloquium. As Sommerfeld said later:

“I used to organize my lectures in such a way that they were too easy for advanced students and too difficult for beginners.”

On the research front, quoting from Peter Ewald, Sommerfeld

“…was one of the central figures in achieving the transformation through which physics passed in the two decades from 1910 to 1930.”

He was mentor to a galaxy of greats in the ‘European half’ of 20th century physics – Peter Debye (in Aachen in 1908), Wolfgang Pauli (1921), Werner Heisenberg (1923), Hans Bethe (1928), all Nobel Laureates; then Walter Heitler, Rudolf Peierls, Peter Ewald, Otto Laporte, E C G Stueckelberg, Gregor Wentzel, Alfred Lande and Leon Brillouin – all names familiar to every serious student of physics. As Einstein put it so aptly:

“What I especially admire about you is that you have, as it were, pounded out of the soil such a large number of young talents.”
He gave the first ever courses on relativity and quantum theory as he was contributing to them himself. As his Institute in Munich grew, it became one of the three great European centres of that time, each expressing a characteristic flavour or attitude to theoretical physics – close to experiment in Munich, more mathematical with Born in Göttingen, and more philosophical with Bohr in Copenhagen. Heisenberg expressed it this way:

“I learned optimism from Sommerfeld, mathematics in Göttingen, and physics from Bohr.”

**Sommerfeld’s Major Contributions to Physics**

Turning to Sommerfeld’s own accomplishments in physics, there was work on the classical theory of electromagnetism, in particular a solution to the problem of a radiating hertzian dipole over a conducting earth, which has been widely applied. Then there were the many extensions of the Bohr model of the atom in the framework of the Old Quantum Theory: the generalization from circular to elliptical orbits for the electron in the atom; in 1914–1915, the introduction of the magnetic quantum number $m$ in addition to angular momentum $l$; in 1916, the extension of Bohr’s quantum condition to the case of multiply periodic systems by him and independently by Wilson; also in 1916, the relativistic correction to the Bohr energy level formula leading to the fine structure of the hydrogen spectrum and the introduction into physics of the ‘Sommerfeld fine-structure constant’ \( \alpha = \frac{e^2}{\hbar c} \approx \frac{1}{137} \) (see article by Biman Nath, *Resonance*, Vol. 20, No. 5, p. 383, 2015) In 1920, he introduced a ‘4th quantum number’ traditionally written as $j$.

Of this period and all this work, it is said that Sommerfeld believed in the Bohr theory more strongly that did Bohr himself! In fact, after his first meeting and discussion with Bohr in 1922, Heisenberg later wrote:

“…I was at once impressed by the difference in his way of seeing quantum theory from Sommerfeld’s way.”

As for the Sommerfeld fine structure formula, it turned out to be a ‘lucky accident’, the now accepted derivation being via the relativistic Dirac wave equation for the electron found in 1928. Quoting from Abraham Pais, while Sommerfeld’s “fine structure formula is quite correct, . . . , his derivation is wide off the mark, turning it into ‘perhaps the most remarkable numerical coincidence in the history of physics’ (Kronig)”.

Sommerfeld’s 4th quantum number would lead in time to the Pauli exclusion principle, and then on to the profound idea of electron spin. After the coming of wave mechanics, which he readily accepted, in 1927 he used the new Fermi-Dirac statistics to present his electron theory of metals, a great advance over the earlier Drude theory.

In addition to teaching and research, Sommerfeld was prolific in the writing of books as well. The *Theorie der Kreisels* has already been mentioned. Between 1919 and 1946 appeared the six editions of his ‘*Atombau und Spektrallinien*’ – veritably the bible for physicists spanning several generations. To this he added in 1928 a special volume on *Wave Mechanics*. Then, after retirement, he turned to the task of writing down his courses of lectures in Munich given regularly from 1906 to 1938. This occupied him from 1942 to 1951, with some of the later volumes completed by others, and resulted in the six volumes of *Lectures in Theoretical Physics*. Again a classic of the literature for physics teaching, recently reprinted by Levant Books, Kolkata.
The fact that Sommerfeld was held in the highest esteem by the physics community worldwide needs no repetition. As this year, 2015, marks the centenary of Einstein’s creation of the General Theory of Relativity, it is revealing to see how he conveyed his deepest feelings on the progress of his work to Sommerfeld. Those were the days when post cards were in regular use, and that was how they communicated with one another. Here are sentences from two postcards Einstein wrote to Sommerfeld in November 1915 and February 1916:

November 28, 1915:
“Dear Sommerfeld,
You must not be cross with me that I am answering your kind and interesting letter only today. But in the last month I had one of the most stimulating, exhausting times of my life, indeed also one of the most successful. I could not think of writing . . . .”

February 8, 1916:
“Dear Sommerfeld,

. . . You will be convinced of the general theory of rel. when you have studied it. That is why I am not mentioning a word in its defense.
Hearty congratulations on your fine discovery and best regards, yours,

Einstein”

The records of the Nobel Foundation show that Sommerfeld was nominated an enormous number of times for the Prize in physics, in fact every year from 1917 to 1937 (except one). As it happened, he never won this high recognition. It is tempting to agree with Pais when he says:
“I belong to those who regret (even more after recent studies) that Sommerfeld’s work was never sufficiently recognized by the Nobel Committee”.

Connections with India

Sommerfeld had regard for and warm relationships with Indian physicists of about a century ago, including of course C V Raman. He visited India during October 1928, and gave lectures at Calcutta, Madras and Bangalore. He had said:
“. . . it was in this ancient land of civilization that, during the last years, strong shoots of modern physics had grown, by which India suddenly emerged in the competition of research as an equal partner with her European and American sisters”.

He also visited Rabindranath Tagore at Santiniketan. After his talk at Presidency College, Madras, in conversation with the eighteen year old Subrahmanyan Chandrasekhar, Sommerfeld told him about the new Fermi–Dirac statistics, and gave him the proof sheets of his own article.
on the electron theory of metals. What then followed is of course history. When Sommerfeld left India, he expressed

“... deepest affection for the highly gifted, unhappy nation ... with sincere gratitude for the many acts of friendliness and honourings.”


The Later Years

Sommerfeld lived through both World Wars, and witnessed the rise and demise of Nazism in Germany. Quoting from Morris Kline:

“Sommerfeld’s life was saddened toward the end of his career by events in Germany. Anti-semitism ... became virulent in the Hitler period ... Sommerfeld was obliged to witness the emigration of famous colleagues, including Einstein”.

Already around 1934 Sommerfeld had written to Einstein:

“... I would now be willing to see Germany disappear as a power and merge into a pacified Europe.” On his part Einstein issued a strong statement in which Sommerfeld appears. Borrowing from Pais’ biography of Einstein:

“... I will have nothing further to do with Germans ... It is otherwise with those few who remained firm within the range of the possible.’ To him those few included Otto Hahn, Max von Laue, Max Planck, and Arnold Sommerfeld.”

On account of the loss of scientific strength in Germany due to World War II, Sommerfeld continued to teach even up to the age of 79. Three of his students – Pauli, Heisenberg and Heitler – had gone on after their doctoral degrees to work as assistants to Born in Göttingen. Here is Born’s appreciation:

“Theoretical physics is a subject which attracts youngsters with a philosophical mind who speculate about the highest principles without sufficient foundations. It was just this type of beginner that he knew how to handle, leading them step by step to a realization of their lack of actual knowledge and providing them with the skill necessary for fertile research .... He had the rare ability to have time to spare for his pupils, in spite of his duties and scientific work ...”.

Kline’s tribute supplements this:

“(He) was at the forefront of the work in electromagnetic theory, relativity and quantum theory and he was the great systematizer and teacher who inspired many of the most creative physicists in the first thirty years of this century.”

Sommerfeld died in Munich on April 26, 1951 at 82 as a result of injuries sustained in a road accident while walking with his grandchildren.

Truly he was one of the most respected and illustrious figures belonging to an incredibly rich scientific tradition, combining the roles of teacher and physicist in a manner never equalled.
After Hideki Yukawa won the Physics Nobel Prize in 1949, his class-fellow Sin-Itiro Tomonaga was the next Japanese theoretical physicist to be so honoured. He shared the 1965 Prize with Richard Feynman and Julian Schwinger for their renormalisation theory of quantum electrodynamics. Apart from being a master theoretical physicist, Tomonaga was also a highly gifted writer (and culturally very refined and sensitive in a deeply Japanese way); his two books on quantum mechanics are masterpieces deserving much greater attention than they seem to have received.

In The Story of Spin, Tomonaga describes many key developments in microscopic physics in the crucial period from the early 1920’s to about 1940 generally revolving around the concept of spin. It is based on a series of twelve lectures delivered over several months in the early 1970’s, and put together in the form of a book in Japanese in 1974. Tomonaga begins with the struggles to understand spectroscopic data and the detailed ways in which spectral lines split in magnetic fields, and how the idea of a new degree of freedom for the electron – its spin – slowly evolved. Many now forgotten historical details are carefully recalled. Around this time

quantum mechanics too was about to appear on the scene, and the stories get intertwined. We read about the state vector concept in quantum mechanics, Dirac’s ‘majestic transformation theory’, the Pauli description of spin and his nonrelativistic equation, and then – like a bolt from the blue – the emergence of the Dirac equation.

Tomonaga spent two years in Leipzig in the late thirties working with Heisenberg. It is understandable (and in any case it must be so!) that Heisenberg’s work in many areas is very carefully discussed, giving a lot of attention to his motivations and method of thinking. We learn about his work on ferromagnetism, the exchange interaction, the basic ideas of the nuclear force, isotopic spin and so on. The Dirac idea of second quantisation, and the Pauli theorem of 1940 on the relation between spin and statistics, get especially thorough treatment. The concept of spinors, and the way they behave under rotations, is also beautifully explained.

Tomonaga traces the deep connections and motivations linking important events and discoveries which we today tend to view as isolated advances – Bohr’s idea of nonconservation of energy in beta decay and inapplicability of quantum mechanics within the nucleus; Pauli’s neutrino idea and Fermi’s use of it; Heisenberg’s theory of the exchange force between protons and neutrons and the later Yukawa theory of the meson, and so on. Only a master in the subject can achieve such a perspective and write about it so well. There is also a good deal of mathematical treatment of the Lorentz group, in connection especially with the work of L H Thomas.

Probably the best chapters are the third, on the Pauli and Dirac equations for the electron; the eighth, sketching Pauli’s derivation of the spin-statistics connection; the ninth, on the amazing discoveries in 1932 and how they influenced the future of physics; and the tenth, largely concerned with Heisenberg’s inauguration of nuclear physics within the basic quantum mechanical formalism. Tomonaga’s explanations are very leisurely and lucid; and when mathematical he stops at just the right place! Some of the material is quite demanding, calling for a high degree of maturity from the reader. His assessments of the characters of the giants of the field and their attitudes towards one another are charming, a sample being: “Dirac’s acrobatics, Pauli’s frontal assault, and Heisenberg’s analogising: each is uniquely characteristic of its practitioner so that we are never bored following their work”. Anecdotes and vignettes abound.

This is a very sophisticated, highly personal account of major developments in physics, and those who brought them about, largely to do with the spin of elementary particles but covering a lot of ground in related areas. It bears much and careful re-reading to fully extract all the meanings and linkages, that only a gifted raconteur like Tomonaga can show us.
Profile of a Polymath*

N Mukunda

Among the many geniuses who carried physics through the quantum revolution, none possessed a more complex personality than Erwin Schrödinger. He was truly also a polymath - "one who knows many arts and sciences." His greatest achievement, the discovery of wave mechanics in 1925–26, occurred, as Weyl said, "during a late erotic outburst in his life." Considering that he was then 38 years old, an age at which, generally, theoretical physicists of such calibre are past their prime, we see the truth in Feynman’s assessment that “Schrödinger rose to the occasion in meeting the challenge of developing his version of quantum mechanics.”

Walter Moore’s Schrödinger: Life and Thought, published in 1989, was soon recognised as a deftly handled portrait of an extraordinarily gifted, yet complex genius. His A Life of Erwin Schrödinger is an abridged version, retaining all the charm and readability of the original.

Erwin Schrödinger was born into a relatively affluent family in Vienna in 1887. Being an only child, and in an environment of doting aunts and nurses, he grew up with definite attitudes.

N Mukunda
towards women and a need for attention. His feelings about formal religion and morality were formed early, in a Vienna with rich cultural, literary and scientific traditions and yet a brittle brilliance. He became the finest product of the Viennese school, with a deep feeling for statistical principles, and a taste for elegance in scientific work.

Schrodinger’s reading of Schopenhauer brought him very early in touch with the Upanishads and Vedantic philosophy; at one stage he seriously considered making philosophy rather than physics his vocation. He saw active duty during World War I. His early scientific work was shaped by the interests of the Vienna school which, in spite of the legacy of Boltzmann, was somewhat away from the mainstream.

Throughout his career, Schrödinger moved continuously from one university to another – first the game of musical chairs among German universities, then longer stays at Zurich and Berlin. His longest tenure, from 1939 to 1956, was in Dublin. It was at Zurich that he caught up with current developments and, at a resort in the company of a Viennese girlfriend, rather like “the dark lady of the sonnets,” created wave mechanics. No other work of his came near the magnitude of this one. His own contribution to the interpretation of quantum mechanics consisted largely of a critique of the conventional Bohr–Heisenberg view. In 1933 he shared the physics Nobel Prize with Dirac. As a rule he dealt only with basic issues and avoided applications, and did not collaborate much with others.

Women – indeed, many of them – played an important role throughout Schrödinger’s life. His condescending, male supremacist attitude towards them may have been the result of early unrequited love. His marriage to Anne-Marie Bertel in 1920 soon became a mutually agreed cover for affairs on both sides. His own, all passionate and most of them brief, even included one with the wife of an obliging junior colleague.

He had no regard for conventional social norms or morality, and on occasion displayed not a little conceit. His yearning for a son was never fulfilled, and each of his three daughters was born of a different affair.

All this contrasts sharply with his Vedantic world view and his deep belief in the unity of minds and consciousness. In some ways this philosophical stance, reinforced by his own wave mechanics, influenced his attitude to science in general. Yet his intellectual pursuits remained so, a world apart from his romantic affairs and unintegrated with actions and relationships. His eloquence and command of language, the persuasive quality of his writing and the ability to distil the essence of a subject into a few pages, remain unsurpassed. His range of interests, revealed in his writings – Nature and the Greeks, My View of the World, Space Time Structure, Mind and Matter, What is Life? Statistical Thermodynamics, among others – each one a jewel of exposition, is astonishing.

Moore treads a delicate path, and combines tact, sympathy and honesty in his account. His delineation of Schrödinger’s philosophical views is accurate and a pleasure to read. In assessing and reacting to this unique personality, a good guide is Max Born’s statement: “His private life seemed strange to bourgeois people like ourselves. But all this does not matter. He was a most lovable person, independent, amusing, temperamental, kind and generous, and he had a most perfect and efficient brain”. And the biblical injunction: “Judge not, that ye be not judged.”
Erwin Schrödinger, “What is Life? The Physical Aspect of the Living Cell”*

N Mukunda


Exactly fifty six years ago, during February 1943, the theoretical physicist Erwin Schrödinger gave a set of three lectures at Trinity College, Dublin under the auspices of the Dublin Institute for Advanced Studies (DIAS) on the subject “What is Life?” Coincidentally Schrödinger himself was fifty-six then, and the subject of quantum mechanics, whose wave mechanical form he had discovered in 1926, was in its teens. These lectures were published in 1944, and have long since become a classic of popular science. They have been reprinted a number of times, and have been reviewed, praised and criticised on scores of occasions. In 1993 a meeting was organised at Trinity College to celebrate the fiftieth anniversary of the original lectures. Quite a record for a tract of barely a hundred pages.

The background to these lectures is interesting. Schrödinger fled his native Austria for neutral Ireland (via Oxford) in 1938, and was to remain there as the Professor of Theoretical Physics at the DIAS till 1956. In the early 1940’s the theoretical physicist P P Ewald brought

to Schrödinger’s attention a 1935 paper by Timoféeff-Ressovsky, Zimmer and Delbrück titled “The Nature of Genetic Mutations and the Structure of the Gene”. There was also an obligation for the DIAS to arrange an annual public lecture. Schrödinger was so fascinated by the 1935 paper and his own thinking and reading on the subject that he gave the three lectures “to an audience of about four hundred, which did not substantially dwindle”. He further explains that he was inspired by the need to preserve the universality of knowledge. In this age of great specialisation, once in a while someone well versed in one area should attempt to build bridges to others, even at the risk of being accused of having inadequate background.

The book consists of seven short chapters, and an epilogue on Schrödinger’s Vedantic philosophical views. Right at the start he poses his basic question - “How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?” – and immediately gives his conclusion: “The obvious inability of present-day physics and chemistry to account for such events is no reason at all for doubting that they can be accounted for by those sciences”. In developing this answer the discussion keeps hovering between physics and biology.

The aim of the first Chapter is to show how a physicist, trained in the Boltzmann–Gibbs tradition of classical statistical mechanics, would approach this problem. Faced with the ubiquitous presence of heat noise or random atomic motions, reliable and stable patterns of behaviour appear in a statistical manner only in systems composed of very large numbers of atoms – the $\sqrt{n}$ law of fluctuations is illustrated via several examples (paramagnetism, diffusion, Brownian motion . . . ), and is also invoked to explain why we are so much bigger than individual atoms. Chapter 2 then immediately shows that such classical arguments are invalid in the biological realm. The basic mechanisms of heredity – cells, their sizes and chromosomal content, the processes of mitotic and meiotic division with cross-over - are briefly and lucidly explained, and the sizes of genes estimated. The upshot is that, far from a value for $n$ of the order of $10^{23}$, biology creates and uses molecular structures with no more than about a million atoms, already overcoming heat noise and displaying stable and orderly behaviours. The following Chapter 3 devoted to mutations, both natural and artificially induced, further reduces the size estimate of the gene to the range of about a thousand atoms. So the great puzzle is: how can an object consisting of just a thousand atoms, may be even less, show such great permanence and regularity of behaviour overcoming heat noise?

For this, Schrödinger turns in Chapter 4 to the quantum theory, with its basic non-classical features of discrete states, energy gaps and quantum jumps. By now he is able to declare - “In the light of present knowledge, the mechanism of heredity is closely related to, nay, founded on, the very basis of quantum theory” – a striking statement which could have come only from one of the creators of quantum mechanics! The basic recognition is that the nature of the chemical bond is intrinsically and unambiguously a result of quantum mechanics, and that holds the key. The picture emerges of the gene as an “aperiodic crystal”, a concept and phrase brought in earlier. Delbrück’s contribution to the 1935 three-man paper mentioned earlier was his model of the gene as a large molecule governed by quantum mechanics and possessing many isomeric states, with transitions between them – mutations! – determined by energy
differences and temperature. This model is taken up in Chapter 5 and an effort is made to show that its main features explain available data on mutation rates, both natural and induced, quite satisfactorily.

Having approached the problems of gene stability, mutations and orderliness in this way, the question that remains for Chapter 6 is: how do biological processes ‘evoke’ the Second Law of thermodynamics, and retain and reproduce existing order? Again Schrödinger points to a consequence of quantum mechanics, the Nernst heat theorem, which shows that the entropy of any system vanishes at the absolute zero of temperature. He contrasts two ways of creating order and stability: from disordered heat motion to order by the classical statistical mechanical “√n route”; and from order to order by retention as exemplified by life processes, thanks to quantum mechanical principles. Indeed he concludes by referring to chromosomes as “cogs of the organic machine”, and as for the gene: “the single cog is not of coarse human make, but is the finest masterpiece ever achieved along the lines of the Lord’s quantum mechanics”.

Schrödinger’s account is many decades old, and enormous progress has taken place since then in the areas and questions he touched. He has often been criticised for major omissions: his knowledge of genetics was from a very narrow set of sources; he ignored the roles of enzymes, catalysts and complementary structures; and “what was true in his book was not original, and most of what was original was known not to be true even when it was written” (Perutz). Much later, in his own ‘Tanner Lectures’ on ‘Origins of Life’, Freeman Dyson appreciatively points out that Schrödinger restricted himself mainly to the problem of replication, and avoided linking it up to metabolism, or even asking the question (wisely at that juncture) of how life arose in the first place.

But Schrödinger spoke (and wrote) to convey his fascination for the subject, and to express what seemed important to him at that time as a physicist. Two of his original ideas have become – after the discovery of the structure of DNA in 1953 – parts of modern biology: that the gene is best viewed as an aperiodic polymer, and that its message for development and reproduction is expressed in some code. And while the degree of influence may vary from person to person, he seems to have inspired many (physicists included) to turn to problems of modern biology: Maurice Wilkins, James Watson, Joshua Lederberg, Seymour Benzer, Francis Crick... Compared to Niels Bohr’s expectation that the Complementarity Principle of quantum mechanics would become indispensable for the understanding of life (which thought turned Max Delbrück from physics to biology), Schrödinger’s general ideas seem to have come much closer to modern biology, and had a more tangible impact. He succeeded in showing that some of the basic questions of biology could be phrased in the language of physics – and so eloquently too – and this was invaluable.

Schrödinger has also been criticised for having treated his interest in this subject as a passing fancy – he got deeply involved in it for a while, expressed his views to a general audience, but did not seriously follow up what he had begun and instead turned to other interests. But then, such is the nature of a polymath functioning to preserve the universality of knowledge. Who today would wish to say that ‘What is Life?’ should not have been written?
Max Delbrück began his scientific career as a theoretical physicist. However, inspired by a 1932 lecture by Niels Bohr titled “Light and Life”, where Bohr suggested the importance of the Complementarity Principle of quantum mechanics for the understanding of life, he decided to devote the rest of his scientific career to molecular biology. He became one of those who created this new field, and grew to attain an almost mythical status in it. And he reflected upon his own experiences in a 1949 essay “A Physicist looks at Biology”, and again in 1969 in “A Physicist’s Renewed Look at Biology – Twenty years Later”.

During 1974–75 Delbrück gave an extempore twenty lecture course at the California Institute of Technology, with the broad aim of assessing whether Niels Bohr’s expectations had been fulfilled. The present book grew out of his notes for these lectures, organized and edited by his colleagues after his death in 1981.

The range of topics covered is breathtakingly vast, all the way from our present understanding of the cosmic evolution of the universe to the emergence of the quantum mechanical understanding of microscopic phenomena; and against this canvas, the formation of the solar system and the planets, and the emergence of life on the earth. Each relatively short chapter gives an incisive account of one or another aspect of this enormous picture.

The main grand question before Delbrück is: can we understand how mind arose from an initially mindless ‘material’ universe? In building up to his answer, he provides masterly surveys of the growth of mathematics and of physics from their beginnings down to modern times, including the intricacies of the Gödel theorems on the one hand and of quantum mechanics on the other. Mechanics, electromagnetism, statistical physics and the relativity theories are covered on the way. Major philosophical developments such as the Cartesian cut and the later Kantian notion of a priori categories of thought are also brought into the discussion.

Initially Delbrück adopts the naïve realist standpoint, which takes the world as existing on its own ‘out there’, while our senses present a faithful picture of it to our minds. This standpoint suffices for an account of the initial occurrence of life on earth some three billion years ago; the appearance of photosynthesis; and then the spreading out of the many branches of the ‘tree of life’ guided by natural selection. Delbrück emphasizes both the unity and the continuity of the myriad forms of life – in the material biochemical sense, and in the psychic sense of organisms sensing their surroundings and reacting to them. But by the time he comes to describing the way the senses and the mind actually work, the limitations of the naïve realist point of view become clear. The intricate – and not yet fully mapped out – pathways of the various senses in the brain, and the way in which the brain – mind? – creates for itself a specifically and selectively treated image of the external world, are astounding. The human observer is seen to be far from passive.

As part of this account, Delbrück brings in the ideas of Konrad Lorenz on the one hand, and the results of Jean Piaget on development and child psychology on the other. Lorenz’s explanation of the relative roles of phylogeny and ontogeny – the development of the species over enormous periods of evolutionary time versus the experience of one individual member of the species during one lifetime – as an explanation of Kant’s ideas is simply fascinating. This helps us understand the basis for the Kantian a priori categories of thought, based on Darwinian evolution, in a way not available to Kant in his own lifetime. The description of Piaget’s findings – acknowledging that according to professional psychologists there has been much progress since – illustrates the ways in which the innate potentialities in every individual, thanks to slow evolutionary development, are realised and put to use based on individual experience of and interaction with the world. The unique role of the language faculty in humans is also covered in the discussion.

Delbrück’s final answer to his main question is that the mind is not mysterious at all – it is a capacity resulting from selective pressures in the world of life. Quoting him: “The point of view of the evolutionist forces us to view mind in the context of other aspects of evolution. . . . In the context of evolution, the mind of the adult human. . . , ceases to be a mysterious phenomenon, a thing unto itself. Rather, mind is seen to be an adaptive response to selective pressures, just
as is nearly everything else in the living world.” It also transpires that the ‘riddle of life’ has been solved, thanks to Watson and Crick, in much simpler and merely mechanical ways than Bohr had expected. In that sense, the mysterious features of quantum mechanics lie beyond an explanation of the phenomenon of life! However it is good to remind oneself that another very distinguished thinker, Erwin Schrödinger, had in “What is Life?” traced the very possibility of life – specifically, the gene – to the principles of quantum mechanics.

This book is full of treasures, and needs a mature mind to absorb the points it makes. And it needs to be read more than once. The sweep, the grandeur of the canvas are stunning. One may prefer one’s own conclusions at the end, but one would have been infinitely better informed by reading Delbrück than otherwise.
Jules Henri Poincaré was one of the most profound thinkers of the late 19th and very early 20th century. Apart from his path-breaking work in many areas of mathematics, he was also deeply involved with the physics of his day, and made lasting contributions to celestial mechanics, optics and electrodynamics. His collection of essays titled “La science et l’hypothèse” was published in French in 1902, and its English translation appeared in 1905. It has long been a classic of the literature of science, covering some areas of mathematics and of physics of that period.

Poincaré divides his thirteen chapters into four major parts respectively titled ‘Number and Magnitude’, ‘Space’, ‘Force’ and ‘Nature’. While the first two parts devoted mainly to arithmetic and geometry retain their value to this day, the last two parts dealing with physics are understandably out of date on account of the great progress that has taken place over the past century. In his penetrating logical analysis of the arithmetical operations and the meaning of magnitude, Poincaré brings out the key role of the method of proof by induction. He expresses beautifully, as only a master can, the true spirit behind mathematical thinking: it is the form,
not the matter, that is of the essence; and relations between objects, not the objects themselves, are of the greatest concern. His extremely close touch with physics seems reflected in the statement: “The mind only uses its creative faculty when experiment requires it”.

When he turns to geometry the situation is rather different. As against the essentially unique structures in arithmetic, with geometry there are several distinct possibilities depending on the axioms one chooses. Poincaré’s deep philosophical insight and grasp keep surfacing all the time in short profound statements such as: “…the word ‘existence’ has not the same meaning when it refers to a mathematical entity as when it refers to a natural object”; and “The geometrical axioms are therefore neither synthetic a priori intuitions nor experimental facts. They are conventions”. He traces the development of the non Euclidean geometries of Bolyai and Lobachevsky; and in explaining our intuitive understanding of the properties of space he lays great stress on the interplay of the senses and of muscular effort in reaching out to and apprehending material objects. One sees how deep the unconscious processes of grasping space really are – we are actually totally unaware of them. Even the three-dimensional nature of space is traced to muscular experience. Poincaré’s view that the geometry we attribute to physical space is basically a convention would appear hard to accept for a physicist familiar with the general theory of relativity – indeed one wonders how Poincaré would have assessed that theory.

Once Poincaré turns to an analysis of the foundations of physics, in particular mechanics, two points become especially clear – the great distinction he draws between the French and the English ways of thinking in science; and his systematic conclusion that so many of the laws held sacred within physics evaporate into thin air and reduce to mere definitions or conventions! Poincaré sees absolute space, time and geometry as part of, and not prior to, mechanics. While he does explore whether the laws of mechanics might have been otherwise than they are, he concludes that Newton’s 1st Law can never be decisively tested, and even the 2nd and 3rd Laws become in his eyes (interdependent) conventions. Concerning the Galilean Principle of Inertia and of Relativity there is a seeming contradiction: at one point he asks “Is this a truth imposed on the mind a priori?” and soon concludes it is not; but somewhat later he declares “…it is imposed upon us for two reasons…. the consideration of the contrary hypothesis is singularly repugnant to the mind”.

The later chapters on energy, principle of least action, optics, electricity and electrodynamics are clearly dated. The general discussion of the role of mathematics in physics is beautifully perceptive. Poincaré sometimes seems to affirm the need for an ether, and also to doubt the reality of molecules. There is much pride in and admiration for the ‘immortal’ work of Ampère, and of Fresnel and Carnot; and at the same time quite strong criticism of Maxwell’s style in expressing his ideas and theories. The meaning of mechanical explanation of physical phenomena – the ideal in 19th century physics – is especially clearly outlined.

It needs a rather mature mind to read and understand Poincaré’s writings. And then one cannot help wondering how he might have reacted to the revolutions in physics of the 20th century.
Books by and about Werner Heisenberg*

N Mukunda

The three principal creators of quantum mechanics made their landmark contributions to the subject in very quick succession between July 1925 and January 1926 in the sequence: Werner Heisenberg, Paul Dirac and Erwin Schrödinger. They received their Nobel Prizes together in December 1933 – Heisenberg for 1932, Dirac and Schrödinger for 1933. The February 1999 and August 2003 issues of Resonance were devoted respectively to Schrödinger and Dirac. The August 2004 one celebrates Heisenberg.

The purpose of this article is to present to a younger generation of readers (students and teachers alike) a very brief account of some of Heisenberg’s general writings of a historical and philosophical nature. These are accessible both to scientists, even in areas other than physics, and to educated and interested nonscientists. In reading this article, at relevant places one may look at Resonance, August 2004, Virendra Singh’s Article-in-a-Box on page 3, and Jochen Heisenberg’s Personal Reflections on page 90.

Both Heisenberg and Schrödinger, from Germany and Austria respectively, wrote extensively on general philosophical matters apart from their technical scientific work. In contrast, Dirac has not written much along these lines. The following books by Schrödinger should be mentioned: Nature and the Greeks; Mind and Matter; My View of the World; Science, Theory and Man (Science and the Human Temperament); and Science and Humanism. Several of these are currently available either in the Canto Series of Cambridge University Press, or published by Ox Bow Press. Heisenberg’s major general writings, more or less in chronological sequence, are listed on pages 273–274 and numbered α, β, γ, δ and ε.

Schrödinger’s writings have rarely been matched for their clarity, eloquence of expression and persuasive power. ‘Heisenberg’s nontechnical writings on physics have always been notable for a strong sense of the history of the subject, and for a clear perception of the impact of modern discoveries on time-honored philosophical disputes’. Among his technical books at least two deserve mention: the classic 1929 Chicago lectures on The Physical Principles of the Quantum Theory (Dover 1949); and from much later in his life, Introduction to the Unified Field Theory of Elementary Particles (Interscience 1966).

The definitive biography of Heisenberg is David C Cassidy’s Uncertainty – The Life and Science of Werner Heisenberg, W H Freeman 1992. Also of great interest is Inner Exile - Recollections of a Life with Werner Heisenberg by his wife Elisabeth, published by Birkhauser in 1984. Michael Frayn’s play ‘Copenhagen’, Anchor Books 2000, has been briefly referred to by Virendra Singh, and is dealt with more extensively by Jochen Heisenberg.

As one would expect, Heisenberg has written frequently about the origins, historical development and difficulties of interpretation of quantum mechanics, the major revolution in thought

which he himself initiated. These are extensively dealt with in \( \beta \) as well as in the partly autobiographical \( \delta \). His discovery of matrix mechanics in May-June 1925 has been recounted in Virendra Singh’s article. Here is Heisenberg’s own description of his intense and dramatic feelings at that time, taken from \( \delta \):

“I had the feeling that, through the surface of atomic phenomena, I was looking at a strangely beautiful interior, and felt almost giddy at the thought that I now had to probe this wealth of mathematical structures nature had so generously spread out before me”.

Elsewhere in \( \delta \) he describes his Youth Movement days, his introduction to physics and the atomic concept, his first meeting with Niels Bohr in 1922, and the important conversation with Einstein in 1926 recalled in Virendra Singh’s article. (This last appears also in \( \epsilon \)). The chapters in \( \delta \) are extremely carefully reconstructed conversations with many different people on a great variety of scientific, philosophical and political topics ranging in time from 1919 up to 1965. They help us see the new physical concepts gradually unfolding with all the attendant suspense, excitement and drama. The evolution of the uncertainty and complementarity principles and of the standard Copenhagen interpretation of quantum mechanics during 1926–27 are all recounted in the first person singular – what more authentic source could one ask for? All these developments are also dealt with in great depth in \( \beta \).

Heisenberg’s accounts of the growth of ideas over the centuries, starting from early Greek speculations about the nature of matter and natural law, are deeply revealing. In essays in \( \alpha \) and \( \gamma \) titled ‘Ideas of the Natural Philosophy of Ancient Times in Modern Physics’ and ‘Natural Law and the Structure of Matter’ there are splendid recapitulations of the ideas of Leucippus, Democritus and Platonic philosophy; and when the comparison to modern day thinking has to be made he says:

“It was an unbelievable achievement of the ancient philosophers to have asked the right questions. But, lacking all knowledge of the empirical details, we could not have expected them to find answers that were correct in detail as well”.

Such a finely balanced and refreshingly honest assessment!

At more than one place in these books, in \( \delta \) and \( \epsilon \) for example, Heisenberg traces the changes that have taken place over time in the concept of elementary particles of matter out of which all matter is composed, and of course also offers his own feelings about how things will develop in the future. In \( \delta \) he says:

“The elementary particle, like the stationary state of an atom, is determined by its symmetry.... This is the natural consequence of the fact that symmetry is rooted in nature itself”.

And in \( \epsilon \) he suggests that we have reached the end of the usefulness of thinking that one thing consists of others:

“Here the question has obviously been asked, ‘What do protons consist of?’ But it has been forgotten in the process, that the term consists of only has a half way clear meaning if we are able to dissect the particle in question, with a small expenditure of energy, into constituents whose rest mass is very much greater than this energy-cost; otherwise, the term consist of has lost its meaning. And that is the situation with protons”.

282
Books by and about Werner Heisenberg

At a couple of places, in (γ) and in (δ), there are discussions of the relation between science and religion. In the chapter titled ‘Science and Religion’ in (δ), Heisenberg reconstructs superbly from memory what Bohr had said to him on this matter:

“... our attitude to religious questions cannot be separated from our attitude to society... religion helps to make social life more harmonious; its most important task is to remind us, in the language of pictures and parables, of the wider framework within which our life is set”.

His accounts of the significance of the works of Albert Einstein, Max Planck and Wolfgang Pauli are contained in three separate essays in (γ).

Virendra Singh refers briefly to Heisenberg’s visit to Bohr in Copenhagen in September 1941, when he failed in his attempt to ‘reach’ Bohr and to effect a reconciliation; and to the controversies surrounding his involvement in the German atomic bomb project. Jochen Heisenberg’s feelings on seeing Michael Frayn’s play ‘Copenhagen’ dealing with this episode are expressed in his article in August 2004 issue. In this context only two quotations will be recalled here. The first is from Elisabeth Heisenberg’s book:

“But they were not communicating; they did not understand each other. The two men who had been such close friends parted deeply disappointed, and there were no further attempts to contact each other. Ultimately, the cause of this misunderstanding must be seen in both of them”.

The second is from Victor Weisskopf’s Preface to this book:

“Whatever has happened, nobody has the right to reproach any body for having avoided a deadly risk. Who can stand up and say he would have taken the risk in the same situation with the same responsibilities? Those who have never been in such situations must be grateful to their fate of being spared such decisions”.

And he concludes by saying that this book

“will help to improve the bond between those who suffer and those who have been spared”.

Heisenberg’s writings on many profound subjects are so beautiful that they appear deceptively simple. One enjoys reading him many times over to truly appreciate his thinking. Some of the books described above may be rather difficult to get, but the serious reader must make the effort to find them. She will be very greatly rewarded.

α

Philosophical Problems of Quantum Physics:

essays and lectures from 1932 to 1948

N Mukunda

\(\beta\) Physics and Philosophy – the Revolution in Modern Science: the 1956 Gifford Lectures at the University of St. Andrews.

\(\gamma\) Across the Frontiers: essays and lectures from 1948 to 1973.

\(\delta\) Physics and Beyond – Encounters and Conversations.

\(\epsilon\) Encounters with Einstein And Other Essays on People, Places, and Particles:
essays and lectures from 1972 to 1976.