

ERWIN SCHRÖDINGER

The fundamental idea of wave mechanics

Nobel Lecture, December 12, 1933

On passing through an optical instrument, such as a telescope or a camera lens, a ray of light is subjected to a change in direction at each refracting or reflecting surface. The path of the rays can be constructed if we know the two simple laws which govern the changes in direction: the law of refraction which was discovered by Snellius a few hundred years ago, and the law of reflection with which Archimedes was familiar more than 2,000 years ago. As a simple example, Fig. 1 shows a ray A-B which is subjected to refraction at each of the four boundary surfaces of two lenses in accordance with the law of Snellius.

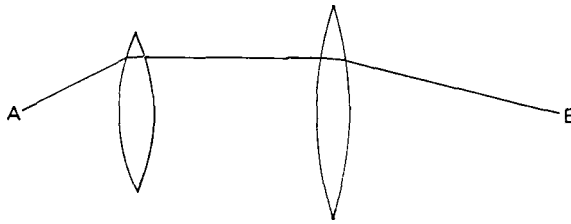


Fig. 1.

Fermat defined the total path of a ray of light from a much more general point of view. In different media, light propagates with different velocities, and the radiation path gives the appearance as if the light must arrive at its destination *as quickly as possible*. (Incidentally, it is permissible here to consider *any two* points along the ray as the starting- and end-points.) The least deviation from the path actually taken would mean a delay. This is the famous Fermat *principle of the shortest light time*, which in a marvellous manner determines the entire fate of a ray of light by a single statement and also includes the more general case, when the nature of the medium varies not suddenly at individual surfaces, but gradually from place to place. The atmosphere of the earth provides an example. The more deeply a ray of light penetrates into it from outside, the more slowly it progresses in an increasingly denser air. Although the differences in the speed of propagation are

infinitesimal, Fermat's principle in these circumstances demands that the light ray should curve earthward (see Fig. 2), so that it remains a little longer in the higher "faster" layers and reaches its destination more quickly than by the shorter straight path (broken line in the figure; disregard the square,

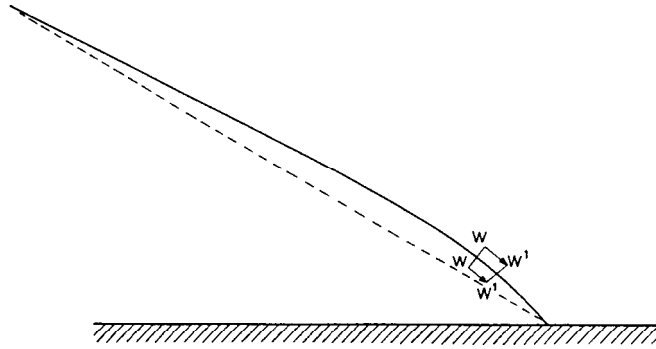


Fig. 2.

WW'W'' for the time being). I think, hardly any of you will have failed to observe that the sun when it is deep on the horizon appears to be not circular but flattened: its vertical diameter looks to be shortened. This is a result of the curvature of the rays.

According to the wave theory of light, the light rays, strictly speaking, have only fictitious significance. They are not the physical paths of some particles of light, but are a mathematical device, the so-called orthogonal trajectories of wave surfaces, imaginary guide lines as it were, which point in the direction normal to the wave surface in which the latter advances (cf. Fig. 3 which shows the simplest case of concentric spherical wave surfaces and accordingly rectilinear rays, whereas Fig. 4 illustrates the case of curved

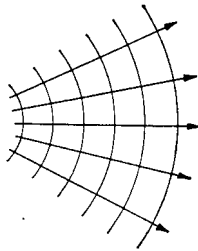


Fig. 3.

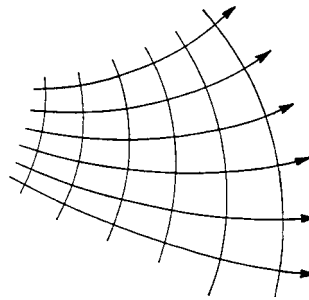


Fig. 4.

rays). It is surprising that a general principle as important as Fermat's relates directly to these mathematical guide lines, and not to the wave surfaces, and one might be inclined for this reason to consider it a mere mathematical curiosity. Far from it. It becomes properly understandable only from the point of view of wave theory and ceases to be a divine miracle. From the wave point of view, the so-called *curvature* of the light ray is far more readily understandable as a *swerving* of the wave surface, which must obviously occur when neighbouring parts of a wave surface advance at different speeds; in exactly the same manner as a company of soldiers marching forward will carry out the order "right incline" by the men taking steps of varying lengths, the right-wing man the smallest, and the left-wing man the longest. In atmospheric refraction of radiation for example (Fig. 2) the section of wave surface WW must necessarily swerve to the right towards W^1W^1 because its left half is located in slightly higher, thinner air and thus advances more rapidly than the right part at lower point. (In passing, I wish to refer to one point at which the *Snellius'* view fails. A horizontally emitted light ray should remain horizontal because the refraction index does not vary in the horizontal direction. In truth, a horizontal ray curves more strongly than any other, which is an obvious consequence of the theory of a swerving wave front.) On detailed examination the Fermat principle is found to be completely *tantamount* to the trivial and obvious statement that - given local distribution of light velocities - the wave front must swerve in the manner indicated. I cannot prove this here, but shall attempt to make it plausible. I would again ask you to visualize a rank of soldiers marching forward. To ensure that the line remains dressed, let the men be connected by a long rod which each holds firmly in his hand. No orders as to direction are given; the only order is: let each man march or run as fast as he can. If the nature of the ground varies slowly from place to place, it will be now the right wing, now the left that advances more quickly, and changes in direction will occur spontaneously. After some time has elapsed, it will be seen that the entire path travelled is not rectilinear, but somehow curved. That this curved path is exactly that by which the destination attained at any moment could be attained *most rapidly* according to the nature of the terrain, is at least quite plausible, since each of the men did his best. It will also be seen that the swerving also occurs invariably in the direction in which the terrain is worse, so that it will come to look in the end as if the men had intentionally "bypassed" a place where they would advance slowly.

The Fermat principle thus appears to be the trivial quintessence of the wave

theory. It was therefore a memorable occasion when Hamilton made the discovery that the true movement of mass points in a field of forces (e.g. of a planet on its orbit around the sun or of a stone thrown in the gravitational field of the earth) is also governed by a very similar general principle, which carries and has made famous the name of its discoverer since then. Admittedly, the Hamilton principle does not say exactly that the mass point chooses the quickest way, but it does say something so similar - the analogy with the principle of the shortest travelling time of light is so close, that one was faced with a puzzle. It seemed as if Nature had realized one and the same law twice by entirely different means: first in the case of light, by means of a fairly obvious play of rays; and again in the case of the mass points, which was anything but obvious, unless somehow wave nature were to be attributed to them also. And this, it seemed impossible to do. Because the "mass points" on which the laws of mechanics had really been confirmed experimentally at that time were only the large, visible, sometimes very large bodies, the planets, for which a thing like "wave nature" appeared to be out of the question.

The smallest, elementary components of matter which we today, much more specifically, call "mass points", were purely hypothetical at the time. It was only after the discovery of radioactivity that constant refinements of methods of measurement permitted the properties of these particles to be studied in detail, and now permit the paths of such particles to be photographed and to be measured very exactly (stereophotogrammetrically) by the brilliant method of C. T. R. Wilson. As far as the measurements extend they confirm that the same mechanical laws are valid for particles as for large bodies, planets, etc. However, it was found that neither the molecule nor the individual atom can be considered as the "ultimate component": but even the atom is a system of highly complex structure. Images are formed in our minds of the structure of atoms *consisting of* particles, images which seem to have a certain similarity with the planetary system. It was only natural that the attempt should at first be made to consider as valid the same laws of motion that had proved themselves so amazingly satisfactory on a large scale. In other words, Hamilton's mechanics, which, as I said above, culminates in the Hamilton principle, were applied also to the "inner life" of the atom. That there is a very close analogy between Hamilton's principle and Fermat's optical principle had meanwhile become all but forgotten. If it was remembered, it was considered to be nothing more than a curious trait of the mathematical theory.

Now, it is very difficult, without further going into details, to convey a proper conception of the success or failure of these classical-mechanical images of the atom. On the one hand, Hamilton's principle in particular proved to be the most faithful and reliable guide, which was simply indispensable; on the other hand one had to suffer, to do justice to the facts, the rough interference of entirely new incomprehensible postulates, of the so-called quantum conditions and quantum postulates. Strident disharmony in the symphony of classical mechanics - yet strangely familiar - played as it were on the same instrument. In mathematical terms we can formulate this as follows: whereas the Hamilton principle merely postulates that a given integral must be a minimum, without the numerical value of the minimum being established by this postulate, it is now demanded that the numerical value of the minimum should be restricted to integral multiples of a universal natural constant, Planck's quantum of action. This incidentally. The situation was fairly desperate. Had the old mechanics failed completely, it would not have been so bad. The way would then have been free to the development of a new system of mechanics. As it was, one was faced with the difficult task of saving the *soul* of the old system, whose inspiration clearly held sway in this microcosm, while at the same time flattering it as it were into accepting the quantum conditions not as gross interference but as issuing from its own innermost essence.

The way out lay just in the possibility, already indicated above, of attributing to the Hamilton principle, also, the operation of a wave mechanism on which the point-mechanical processes are essentially based, just as one had long become accustomed to doing in the case of phenomena relating to light and of the Fermat principle which governs them. Admittedly, the individual path of a mass point loses its proper physical significance and becomes as fictitious as the individual isolated ray of light. The essence of the theory, the minimum principle, however, remains not only intact, but reveals its true and simple meaning only under the wave-like aspect, as already explained. Strictly speaking, the new theory is in fact not *new*, it is a completely organic development, one might almost be tempted to say a more elaborate exposition, of the old theory.

How was it then that this new more "elaborate" exposition led to notably different results; what enabled it, when applied to the atom, to obviate difficulties which the old theory could not solve? What enabled it to render gross interference acceptable or even to make it its own?

Again, these matters can best be illustrated by analogy with optics. Quite

properly, indeed, I previously called the Fermat principle the quintessence of the wave theory of light: nevertheless, it cannot render dispensable a more exact study of the wave process itself. The so-called refraction and interference phenomena of light can only be understood if we trace the wave process in detail because what matters is not only the eventual destination of the wave, but also whether at a given moment it arrives there with a wave peak or a wave trough. In the older, coarser experimental arrangements, these phenomena occurred as small details only and escaped observation. Once they were noticed and were interpreted correctly, by means of waves, it was easy to devise experiments in which the wave nature of light finds expression not only in small details, but on a very large scale in the entire character of the phenomenon.

Allow me to illustrate this by two examples, first, the example of an optical instrument, such as telescope, microscope, etc. The object is to obtain a sharp image, i.e. it is desired that all rays issuing from a point should be reunited in a point, the so-called focus (cf. Fig. 5 a). It was at first believed that it was only geometrical-optical difficulties which prevented this: they are indeed considerable. Later it was found that even in the best designed instru-

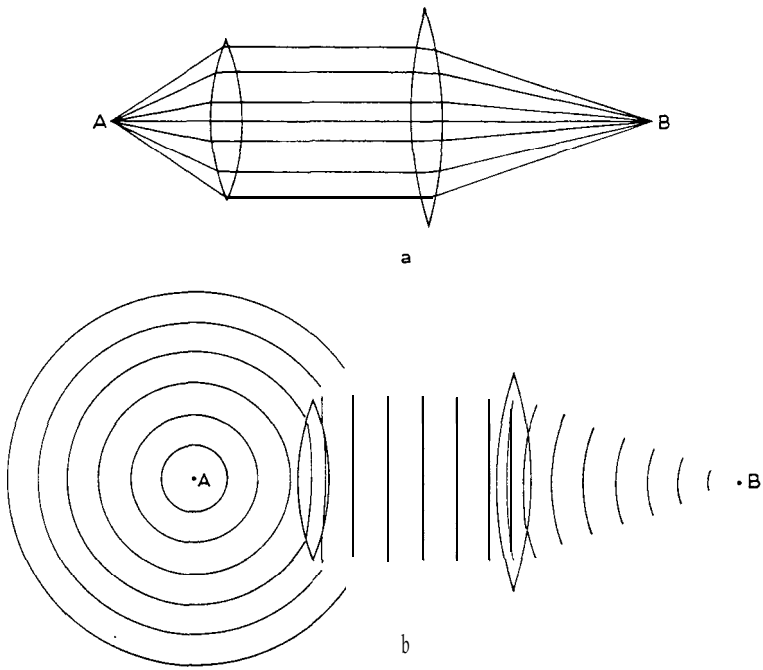


Fig. 5.

ments focussing of the rays was considerably inferior than would be expected if each ray exactly obeyed the Fermat principle independently of the neighbouring rays. The light which issues from a point and is received by the instrument is reunited behind the instrument not in a single point any more, but is distributed over a small circular area, a so-called diffraction disc, which, otherwise, is in most cases a circle only because the apertures and lens contours are generally circular. For, the cause of the phenomenon which we call *diffraction* is that not all the spherical waves issuing from the object point can be accommodated by the instrument. The lens edges and any apertures merely cut out a part of the wave surfaces (cf. Fig. 5b) and - if you will permit me to use a more suggestive expression - the injured margins resist rigid unification in a point and produce the somewhat blurred or vague image. The degree of blurring is closely associated with the wavelength of the light and is completely inevitable because of this deep-seated theoretical relationship. Hardly noticed at first, it governs and restricts the performance of the modern microscope which has mastered all other errors of reproduction. The images obtained of structures not much coarser or even still finer than the wavelengths of light are only remotely or not at all similar to the original.

A second, even simpler example is the shadow of an opaque object cast on a screen by a small point light source. In order to construct the shape of the shadow, each light ray must be traced and it must be established whether or not the opaque object prevents it from reaching the screen. The *margin* of the shadow is formed by those light rays which only just brush past the edge of the body. Experience has shown that the shadow margin is not absolutely sharp even with a point-shaped light source and a sharply defined shadow-casting object. The reason for this is the same as in the first example. The wave front is as it were bisected by the body (cf. Fig. 6) and the traces of this injury result in blurring of the margin of the shadow which would be incomprehensible if the individual light rays were independent entities advancing independently of one another without reference to their neighbours.

This phenomenon - which is also called diffraction - is not as a rule very noticeable with large bodies. But if the shadow-casting body is very small at least in one dimension, diffraction finds expression firstly in that no proper shadow is formed at all, and secondly - much more strikingly - in that the small body itself becomes as it were its own source of light and radiates light in all directions (preferentially to be sure, at small angles relative to the inci-

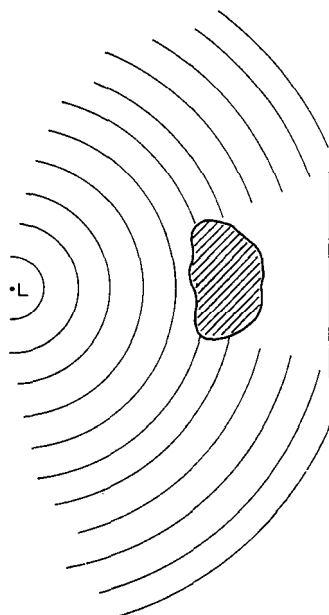


Fig. 6.

dent light). All of you are undoubtedly familiar with the so-called "motes of dust" in a light beam falling into a dark room. Fine blades of grass and spiders' webs on the crest of a hill with the sun behind it, or the errant locks of hair of a man standing with the sun behind often light up mysteriously by diffracted light, and the visibility of smoke and mist is based on it. It comes not really from the body itself, but from its immediate surroundings, an area in which it causes considerable interference with the incident wave fronts. It is interesting, and important for what follows, to observe that the area of interference always and in every direction has at least the extent of one or a few wavelengths, no matter how small the disturbing particle may be. Once again, therefore, we observe a close relationship between the phenomenon of diffraction and wavelength. This is perhaps best illustrated by reference to another wave process, i.e. sound. Because of the much greater wavelength, which is of the order of centimetres and metres, shadow formation recedes in the case of sound, and diffraction plays a major, and practically important, part: we can easily *hear* a man calling from behind a high wall or around the corner of a solid house, even if we cannot see him.

Let us return from optics to mechanics and explore the analogy to its fullest extent. In optics the old system of mechanics corresponds to intellec-

tually operating with isolated mutually independent light rays. The new undulatory mechanics corresponds to the wave theory of light. What is gained by changing from the old view to the new is that the diffraction phenomena can be accommodated or, better expressed, what is gained is something that is strictly analogous to the diffraction phenomena of light and which on the whole must be very unimportant, otherwise the old view of mechanics would not have given full satisfaction so long. It is, however, easy to surmise that the neglected phenomenon may in some circumstances make itself very much felt, will entirely dominate the mechanical process, and will face the old system with insoluble riddles, if *the entire mechanical system is comparable in extent with the wavelengths of the "waves of matter" which play the same part in mechanical processes as that played by the light waves in optical processes.*

This is the reason why in these minute systems, the atoms, the old view was bound to fail, which though remaining intact as a close approximation for gross mechanical processes, but is no longer adequate for the delicate interplay in areas of the order of magnitude of one or a few wavelengths. It was astounding to observe the manner in which all those strange additional requirements developed spontaneously from the new undulatory view, whereas they had to be forced upon the old view to adapt them to the inner life of the atom and to provide some explanation of the observed facts.

Thus, the salient point of the whole matter is that the diameters of the atoms and the wavelength of the hypothetical material waves are of approximately the same order of magnitude. And now you are bound to ask whether it must be considered mere chance that in our continued analysis of the structure of matter we should come upon the order of magnitude of the wavelength at this of all points, or whether this is to some extent comprehensible. Further, you may ask, how we know that this is so, since the material waves are an entirely new requirement of this theory, unknown anywhere else. Or is it simply that this is an *assumption* which had to be made?

The agreement between the orders of magnitude is no mere chance, nor is any special assumption about it necessary; it follows automatically from the theory in the following remarkable manner. That the heavy *nucleus* of the atom is very much smaller than the atom and may therefore be considered as a point centre of attraction in the argument which follows may be considered as experimentally established by the experiments on the scattering

of alpha rays done by Rutherford and Chadwick. Instead of the *electrons* we introduce hypothetical waves, whose wavelengths are left entirely open, because we know nothing about them yet. This leaves a letter, say *a*, indicating a still unknown figure, in our calculation. We are, however, used to this in such calculations and it does not prevent us from calculating that the nucleus of the atom must produce a kind of diffraction phenomenon in these waves, similarly as a minute dust particle does in light waves. Analogously, it follows that there is a close relationship between the extent of the area of interference with which the nucleus surrounds itself and the wavelength, and that the two are of the same order of magnitude. What this is, we have had to leave open; but the most important step now follows: we *identify the area of interference, the diffraction halo, with the atom; we assert that the atom in reality is merely the diffraction phenomenon of an electron wave captured as it were by the nucleus of the atom*. It is no longer a matter of chance that the size of the atom and the wavelength are of the same order of magnitude: it is a matter of course. We know the numerical value of neither, because we still have in our calculation the *one* unknown constant, which we called *a*. There are two possible ways of determining it, which provide a mutual check on one another. First, we can so select it that the manifestations of life of the atom, above all the spectrum lines emitted, come out correctly quantitatively; these can after all be measured very accurately. Secondly, we can select *a* in a manner such that the diffraction halo acquires the size required for the atom. These two determinations of *a* (of which the second is admittedly far more imprecise because "size of the atom" is no clearly defined term) *are in complete agreement with one another*. Thirdly, and lastly, we can remark that the constant remaining unknown, physically speaking, does not in fact have the dimension of a length, but of an action, i.e. energy x time. It is then an obvious step to substitute for it the numerical value of Planck's universal quantum of action, which is accurately known from the laws of heat radiation. It will be seen that *we return, with the full, now considerable accuracy, to the first (most accurate) determination*.

Quantitatively speaking, the theory therefore manages with a minimum of new assumptions. It contains a single available constant, to which a numerical value familiar from the older quantum theory must be given, first to attribute to the diffraction halos the right size so that they can be reasonably identified with the atoms, and secondly, to evaluate quantitatively and correctly all the manifestations of life of the atom, the light radiated by it, the ionization energy, etc.

I have tried to place before you the fundamental idea of the wave theory of matter in the simplest possible form. I must admit now that in my desire not to tangle the ideas from the very beginning, I have painted the lily. Not as regards the high degree to which all sufficiently, carefully drawn conclusions are confirmed by experience, but with regard to the conceptual ease and simplicity with which the conclusions are reached. I am not speaking here of the mathematical difficulties, which always turn out to be trivial in the end, but of the conceptual difficulties. It is, of course, easy to say that we turn from the concept of a *curved path* to a system of wave surfaces normal to it. The wave surfaces, however, even if we consider only small parts of them (see Fig. 7) include at least a narrow *bundle* of possible curved paths,



Fig. 7.

to all of which they stand in the same relationship. According to the old view, but not according to the new, one of them in each concrete individual case is distinguished from all the others which are "only possible", as that "really travelled". We are faced here with the full force of the logical opposition between an

either - or (point mechanics)

and a

both - and (wave mechanics)

This would not matter much, if the old system were to be dropped entirely and to be *replaced* by the new. Unfortunately, this is not the case. From the

point of view of wave mechanics, the infinite array of possible point paths would be merely fictitious, none of them would have the prerogative over the others of being that really travelled in an individual case. I have, however, already mentioned that we have yet really observed such individual particle paths in some cases. The wave theory can represent this, either not at all or only very imperfectly. We find it confoundedly difficult to interpret the traces we see as nothing more than narrow bundles of equally possible paths between which the wave surfaces establish cross-connections. Yet, these cross-connections are necessary for an understanding of the diffraction and interference phenomena which can be demonstrated for the same particle with the same plausibility - and that on a large scale, not just as a consequence of the theoretical ideas about the interior of the atom, which we mentioned earlier. Conditions are admittedly such that we can always manage to make do in each concrete individual case without the two different aspects leading to different expectations as to the result of certain experiments. We cannot, however, manage to make do with such old, familiar, and seemingly indispensable terms as "real" or "only possible"; we are never in a position to say what really *is* or what really *happens*, but we can only say what will be *observed* in any concrete individual case. Will we have to be permanently satisfied with this. . . ? On principle, yes. On principle, there is nothing new in the postulate that in the end exact science should aim at nothing more than the description of what can really be observed. The question is only whether from now on we shall have to refrain from tying description to a clear hypothesis about the real nature of the world. There are many who wish to pronounce such abdication even today. But I believe that this means making things a little too easy for oneself.

I would define the present state of our knowledge as follows. The ray or the particle path corresponds to a *longitudinal* relationship of the propagation process (i.e. *in the direction* of propagation), the wave surface on the other hand to a *transversal* relationship (i.e. *normal* to it). *Both* relationships are without doubt real; one is proved by photographed particle paths, the other by interference experiments. To combine both in a uniform system has proved impossible so far. Only in extreme cases does either the transversal, shell-shaped or the radial, longitudinal relationship predominate to such an extent that we *think* we can make do with the wave theory alone or with the particle theory alone.