12.

IRREVERSIBILITY.

By E. SCHRÖDINGER.

(From the Dublin Institute for Advanced Studies.)

[Read 23 May, 1949. Published 28 August, 1950.]

1.

It may seem an audacity if one undertakes to proffer new arguments in respect of a question about which there has been for more than eighty years so much passionate controversy, some of the most eminent physicists and mathematicians siding differently or favouring opposite solutions — Boltzmann, Loschmidt, Zermelo, H. Poincaré, Ehrenfest, Einstein, J. von Neumann, Max Born, to name only those who come to me instantly. But, to my mind, in this case, as in a few others, the "new doctrine" which sprang up in 1925/26 has obscured minds more than it has enlightened them. It is sometimes believed that only quantum mechanics, or some processes of thought borrowed from it, give the final clue to the problem. I wish to show here that this is wrong and that the solution given previously can be defended against the last objection that continues to be raised again and again.

This objection, in short, is this: a proof that a reversible model shows an irreversible behaviour, i.e. that it "nearly always" exhibits a temporal succession of observable states which it "almost never" passes through in the reverse order of time—such a proof needs must be at fault somewhere. This consideration seems so absolutely irrefutable that it obtrudes itself to the most sagacious minds again and again as an irremovable stumbling block. Of late[1] a new way out has been sought—it is the one I had in mind when I spoke of the borrowing of thought from the new doctrines.

The following is known and is universally agreed upon: the overwhelming majority of all those micro-states that would impress our crude senses as the same observable (= macro-) state do lead to identical, moreover to the actually observed *consequences*. That seems fine. What ails us is only, that we can equally well scan the *antecedents*. And they are—again for an overwhelming majority—entirely wrong, inasmuch as

[1] Max Born, Natural Philosophy of Cause and Chance, Oxford, at the Clarendon Press, 1949.

the antecedents are the mirror image in time of the aforesaid consequences; it would thus appear that the system has reached its momentary state by an "anticipation" of its actual future history in reversed order. It is true that our two "overwhelming majorities" do not exactly coincide. In particular, among the first lot—that with correct consequences—there is a small subset which has correct antecedents as well. But the micro-states of this subset are so rare, even a little rarer than those we are inclined to neglect for having not the right consequences. It is therefore hard to see how the overwhelming majority concerns us at all and what benefit for understanding the observed phenomena we could draw from its, as it were, 50% correct behaviour.

From this awkward situation Born, *l.c.*, if I understand him aright, proposes the following rescue. Since we do not know the actual micro-state of the system, we must—and that is where the philosophical loan from quantum mechanics comes in—refrain from drawing inferences from it. We must draw conclusions by averaging over all the micro-states that may equally well be at the back of the observed macrostate. That looks splendid. For, after what has been said before and is agreed upon by everybody, we thus arrive at a correct prediction of the system's future behaviour. But it would seem to me a rather crude way of killing off the undesirable inference with respect to the opposite direction in time, if one prohibited any conclusions concerning the past by saying that our observation of the system in that particular moment is in itself an irreversible process which cannot be used for drawing any inferences concerning the past. Yet I can see no other way of freeing the conclusions drawn, after averaging, from the unfortunate symmetry with respect to time that has bothered us before. Surely the system continues to exist and to behave, to undergo irreversible changes and to increase its entropy in the interval between two observations. The observations we might have made in between cannot be essential in determining its course. Moreover, the very spirit of quantum mechanics, combined with that of thermodynamics, forbids us even to *think* of such observations taking place, if—as has often to be assumed—the system is isolated from the rest of the world in the interval between the two observations.

2.

The problem before us here is *not* actually to derive irreversibility—say, the increase of entropy with time—from any kind of general or special reversible model. Not from a general one: for it is hardly possible to devise a model general enough not only to comprise all kinds of physical events but also to anticipate all changes the reversible theories of physics may undergo in future, and to be inviolable to any such change; I mean changes as we have experienced them when Newton's absolute notions of space and time had to yield to the Theory of Relativity or classical mechanics to quantum mechanics. Still less would it serve our purpose if we were only to refute the objections raised against some special model, e.g. Boltzmann's model of a gas, purporting to picture certain irreversible happenings.

Our scope is in some respect narrower, in other respect wider. I do not wish to derive irreversibility at all. I wish to reformulate the laws of phenomenological irreversibility, thus certain statements of thermodynamics, in such a way, that the logical contradiction *any* derivation of these laws from reversible models seems to involve is removed once and for ever.

The task is clearly outlined. No such derivation can avoid introducing right at the outset a time variable t. If the model—whether it be a visualizable model of the old style or just a system of equations and prescriptions as is nowadays favoured in some quarters—I say, if the model is reversible, any general behaviour you rightfully infer for increasing t, must also hold for decreasing t. In other words it must be an invariant of the transformation $t' = -t$. Hence our task is to formulate all statements about irreversibility in such a fashion that they are invariant to the said transformation. At first sight it would seem that phenomenological time can have nothing to do with the variable t. It could not be defined by t. And it could not be defined by $-t$. This is true. And if you unite these statements and say it can be defined neither as t, nor as $-t$, that is also true. We shall see however that it can be defined as "either t or $-t$".

The most usual way of enouncing the Second Law is to say that a system perfectly isolated from the rest of the world never decreases its entropy, and, apart from the exceptional case that it happens to be already in thermodynamical equilibrium, increases its entropy until thermodynamical equilibrium is reached.

No model that has ever been conceived behaves in this way. Left to itself for a sufficient time it will take on all possible states again and again. Its entropy decreases as often as it increases. What is true is that only in an infinitesimal fraction of all the time you will encounter the model in a state *not* perceptibly corresponding to thermodynamic equilibrium. Moreover *if* you encounter it in such a non-equilibrium state then—supposing you have made a record of its history—you will find that it has left a state close to equilibrium comparatively not long before and that it returns to one not long after. During the time intervals of ascent to the non-equilibrium state and of the return from it, the quantity corresponding to entropy in the model behaves, apart from imperceptible fluctuations, monotonically, it decreases during the first and increases during the second of these intervals.

All this is well known. Perhaps less well known is the following. If you know that during the period of ascent or during the period of return your system has separated into two systems isolated from each other (as may happen), these two systems will also have their entropy changing monotonically (apart from small fluctuations), decreasing or increasing, as the case may be, but both in the same direction of time. While everybody will be prepared to grant this for the period of return to equilibrium some may be loath to accept my statement concerning the period of ascent, thus of decreasing entropy. To them I need only answer, that the two enouncements stand and fall together, since the model is supposed to be reversible. But perhaps it is well to tell the reason also in the customary jargon which calls the period of ascent an "infinitely improbable" one, the period of return one "following the ordinary laws of nature," necessary and unavoidable once the system has had the audacity to escape into this "frightfully improbable" state. (Actually, of course, since you know the system has escaped—and we had *chosen* such a rare moment—there is no longer anything improbable about it, it is just certain.) Generally speaking the periods of escape and the periods of return are exact time-mirror-images of each other. If you consider the period of escape as an extremely improbable one, and include the case of splitting mentioned above, well then you must tell yourself this: I *know* the system to have reached this very abnormally low value of the entropy (in fact I have waited until it did!). It is so infinitely less probable that the system should reach this state otherwise than by a monotonical decrease of the entropy during all the period in all its parts—even when separated—in a word otherwise than by a direct time-mirror-image of normal behaviour, that it is next to certain that it did follow this way. ("Infinitely less probable" has here the clear-cut meaning: only in an infinitesimal fraction of all the cases when a certain low value of the entropy has been reached would it have been reached in another fashion.)

3.

It is now quite obvious in what manner you have to reformulate the law of entropy—or for that matter all other irreversible statements—so that they be capable of being derived from reversible models. You must not speak of one isolated system but at least of two, which you may for the moment consider isolated from the rest of the world, but not always from each other. Envisage them for a time that ought not to be *too* long (but if it does not substantially exceed the time the universe has existed in its present form there will be no trouble). Let them be isolated from each other between the "moments" t_A and $t_B > t_A$ of that time variable t of which we spoke above, but in contact for $t < t_A$ and for $t > t_B$. Labelling the two systems by 1 and 2 and calling S_{1A} the entropy of system 1 at t_A, etc., the formulation of the entropy law I propose is

$$(S_{1B} - S_{1A})(S_{2B} - S_{2A}) \geqslant 0,$$

with the corrollary that, whenever an entropy difference is different from zero the change is (apart from imperceptible fluctuations) monotonical. If at least one of the differences is positive, t is the time, if at least one is negative, $-t$ is the time, if they are both zero, this experiment has not succeeded in deciding the issue.

To get back to the ordinary formulation you may take S_1 to refer to the system under consideration, S_2 to the rest of the world. There is no danger of contradictory time definitions ever resulting from various experiments, since every system is in contact with the rest of the world when YOU observe it.

Once time—time's arrow—is settled in this manner it is no longer extraordinary to find that with respect to it friction, diffusion, viscosity and whatnot act in the fashion they do and not in the opposite fashion. In itself the latter is equally possible and even in a way equally probable with reversible models, since in every display of such phenomena the bodies involved find themselves at every moment in a microstate the time-mirror-image of which would lead to the whole process running backwards through all its previous stages in opposite order of time. But all these processes imply an entropy change and will therefore fit into our picture with the correct arrow, not spoil it by the wrong one.

It is hardly necessary to mention that our inequality is to be understood in the same approximate sense as the customary statement about the increase of entropy. If one of the entropy differences does not appreciably exceed the normal thermodynamic fluctuation of the system in question, the product is to be considered "practically zero." This might give rise to various objections, particularly when one of the two systems is so very big as "the rest of the world". If one feels uneasy about it, one may use a more cautious formulation. Instead of stating, as we did, that the product is greater than or equal to zero, it is quite sufficient to maintain that it is positive whenever both its factors appreciably exceed the normal thermodynamic fluctuations of the systems to which they respectively refer.

4.

I beg to be allowed to enhance this paper by a brief dialogue in which a physicist who refuses to use reversible models for representing irreversible events is likened to a prisoner, who is afraid of drawing a fairly safe conclusion, thereby missing the opportunity for ending his detention. The numbers on the counters represent observed entropy values. The numbers *of* counters (81, 9, 1) and their ratios ought to be astronomical. This is the story:

James was in prison with no hope of being released. One day the gaol-keeper came to his cell and said: I am in the position of offering you a chance for freedom. Would you accept it, if, in doing so, you ran the risk of losing your life, though only with odds of 1 in 10 against you?

— Certainly I would, said James, who is not a coward.

= Would you also accept it if the odds were equal for your either being executed or getting free?

— I would not, said James.

= Very well, said the gaol-keeper, you need not decide until the very last moment. What I am ordered to propose to you is a sort of gambling, for your life or for freedom. I have here an urn. I am putting into it 81 counters, each with a **3** on one side and a **4** on the other side. I am adding 9 counters with a **3** on one side and a **2** on the other side. And finally one counter with a **2** on one side and **1** on the other side. Shuffle them.

— I have. Now tell me what the game is going to be. You make me curious.

= Listen. You will draw one of the counters out of the urn at random and, without looking at it, toss it up into the air, then look at the side it shows. You are to guess what is on the other side. If you guess right, you are free, if you guess wrong, you'll be put to death—but you may refuse to guess, then you remain in gaol, with no danger to your life.

After thinking a while, James exclaimed:

— Of course I accept.

= Be careful, said the gaol-keeper, reserve your decision until after tossing. You might be very unlucky in the counter you draw.

— How should that be, said James. If I see a **4** or a **1**, I am saved anyhow. If I see a **3**, I can pretty safely guess **4**, and, with a **2**, pretty safely guess **3**. That would be just the odds of 1 in 10 of which you spoke. I shall run that risk.

= Well, I hope you'll be lucky. But, mind you, you may still refuse to guess after having tossed. That is the inalienable rule of the game.

James drew and tossed and up came a **2**. He was on the point of opening his mouth to say: three—when the gaol-keeper violently put his hand on James' lips and said: Think, before you decide.

James was angry and worried. But soon the following occurred to him:

Well, I believe this counter to be one of the nine 2/3 counters (rather than the 2/1 counter, of which there is only one) and all my hope is set upon the correctness of this guess. But *if* it is correct, then, in the moment of tossing, there was an equal chance for the counter coming down the other way, showing me a **3**. Then the same mathematical principles which I am on the point of using with confidence would ruin me.

In fact, James realized that to a person who followed these mathematical principles the mere incident of drawing a 2/3 counter involved a 50% death-danger. This made him lose the courage to use those principles when they definitely seemed to point to such a dangerous counter. But he knew of no better principles and therefore definitely refused to guess and thus remained in prison.

I don't know was it mercy or cruelty that made the gaol-keeper snatch the counter at this moment and throw it back into the urn, so that nobody ever knew what was on the other side.

What do *you* believe?

Poor James went almost mad about it. When he realized that he had been fooled, he wrote into his diary:

Never be afraid of dangers that have gone by! It's those ahead that matter.