The arrow of time

The subject of this lecture is probably the most slippery question in the whole of the foundations of physics (at least as we know it today). The distinction between past and future seems at least as deeply built into our instinctive, common sense thinking as for example the assumption of the existence of unobserved objects, and if one were a Kantian it would almost certainly count as an important piece of synthetic a priori knowledge about the world. When one is dealing with a new theory which the behavior of objects like atoms and electrons, very far in scale from the everyday world has forced upon us, such as quantum mechanics, the questions and dilemmas which pose themselves often seem obvious and it is finding the answers which is difficult. By contrast, in a problem that arises as it were in our own intellectual back yard the difficulty may rather lie in finding sensible questions to pose in the first place. The problem of the "arrow", or direction, of time is just such a problem, and I rather suspect that the reason for the spectacular lack of progress on it is largely due to the fact that we are somehow asking the wrong questions.

Let's start by reviewing a few things on which just about everyone interested in the problem agrees. First, it is generally agreed that with one minor and probably not very important exception (see below), the laws of physics at the *microscopic* level are invariant under time reversal. This is particularly easy to see in the case of Newtonian mechanics. The first and third laws clearly do not involve the sense of time, so consider the second, "force = mass × acceleration". What happens if we reverse the reckoning of time, i.e. set $t \to -t$? The "force" on the LHS can usually be expressed as the gradient of a potential energy (though see below), and thus does not care about the direction in which time is measured. On the RHS, the "mass" clearly is invariant: as to acceleration, this is the rate of change of velocity with time. When time is reversed, the velocity is also, so the acceleration is unchanged. Thus, Newton's laws, and therefore the whole of Newtonian mechanics, is invariant under time reversal. What this means, operationally, is that if we were to show an astronomer from a planet of some distant star a speeded-up movie of the motion of the planets of our own solar system, he would be unable to tell, just from a knowledge of Newtonian mechanics, whether it was being run forwards or backwards. Similarly, if we imagine an idealized billiard table which is so smooth (and the balls so elastic) that dissipation both in the motion of the balls over the cloth and in their collisions with one another is totally negligible, again we would be unable to tell whether a movie of the processes going on it is being run in the right direction: the time reverse of every process is also a possible process! (In real life, of course, dissipation is not negligible and we can usually tell after a few seconds).

At this point let me dispose of an obvious objection which may occur to people familiar with the motion of a charged particle in a magnetic field. Such a particle feels a transverse force proportional to its velocity and therefore spirals around the direction of the field, clockwise if the field is in the positive direction, anticlockwise if it is negative. Now if we reverse the direction of time, the direction of circulation is reversed: so doesn't that mean that we have a means of knowing whether the movie is being run forward or backward? The answer is no, because the magnetic field is itself produced by electric currents (or phenomena related to currents) and these currents change sign when time is reversed. Thus, the magnetic field itself reverses, and invariance of the equations of motion is restored. (The situation is essentially the same as with respect to invariance under inversion, i.e. transformation from a righthanded to a left-handed coordinate system: compare lecture 7).

So much for classical mechanics. Is quantum mechanics also invariant under time reversal? This question is slightly trickier. If one simply follows the standard presentation of the subject given in most textbooks, it looks at first sight as if the answer is no. For in this presentation one assigns at some *initial* time t_i an initial state vector to an ensemble on the basis of its preparation, then follows its (deterministic) development according to the laws of QM (Schrödinger's equation) and then at some later time t_f makes measurements on the systems of the ensemble and compares the resulting statistics with the QM predictions. At first sight, this procedure does not look at all invariant under time reversal. However, this apparent asymmetry is illusory, at least at the formal level. In a famous paper in 1964, Aharonov, Bergman and Lebowitz showed that instead of the standard textbook formulation of QM, which they call "predictive", one can equally well introduce a "retrodictive" version, in which we assign systems to an ensemble on the basis of the behavior observed at final time t_i, thus obtaining the state vector at t_f: we then follow the completely deterministic) development of the state vector *backwards* in time, and eventually are in a position to make a "retrodiction" of the outcomes of measurements at the initial time t_i. This procedure is almost exactly analogous to solving Newton's equations of motion subject to specified final conditions rather than the usual procedure of using initial conditions: the only difference is that in QM one obtains only statistical predictions or retrodictions instead of the exact ones of Newtonian physics. Aharonov and co-workers further show that there is a third possible formulation, which they call "time-symmetric"; in this formulation one imposes *both* initial *and* final conditions on the ensemble (e.g. we consider only those photons which are known not only to have had polarization say up at initial time t_i, but also to have had polarization sideways at the final time t_f). We then assume that some measurement was actually carried out at a time intermediate between t_i and t_f and ask for the probability that a given one of the possible results was obtained. QM gives a perfectly definite prescription for answering the question, which has in fact to some extent been checked in recent experiments; though it is interesting that the answer *cannot* be expressed in terms of an assumed wave function (state vector) at the immediate time in question - a result which might perhaps lead us to reconsider the question whether the state vectors (probability amplitudes) of the "standard" formulation are actually quite as fundamental as they seem. It is clear that the "time-symmetric" version of QM also has an analog in Newtonian physics, namely a calculation in which we fix the position (only!) of the cannonball at the initial and final times and use those conditions to calculate its trajectory in between*.

*A little thought shows that the analogy is not quite exact - the QM case is more closely analogous to "over-determining" the classical problem by specifying both position and velocity at both t_i and t_f.

The most interesting observation in the work of Aharonov and coworkers, however, is that one can start with the time-symmetric formulation of QM and then do one of two things: either

(a) one can completely neglect, or "scramble" the final conditions, in which case we get back to the standard "predictive" QM of the textbooks. Or

(b) we can neglect the information contained in the initial conditions (but keep that in the final ones), in which case we reduce to the "retrodictive" formulation. (Of course, in any specific physical situation it is not clear that we *have* to do either of these things: we might be able to stick with the original time-symmetric version!).

Thus the question of why we in practice prefer to work with the "predictive" version of QM rather than the "retrodictive" one reduces to the question of why we regard information about the initial state of a system as more relevant or "determining" than information about its final state. As we shall see, this question is actually a very generic one and in no way peculiar to QM.

For completeness we must note the one known instance of failure of time-reversal invariance at the microscopic level. The general scheme which most people currently believe describes "elementary" (or not-so-elementary!) particles and their interactions is quantum field theory, which is essentially quantum mechanics generalized so as to be relativistically covariant and to allow the possibility of processes in which particles are created or destroyed . This theory makes on quite general grounds, a number of remarkable predictions, in particular that the laws of nature should be invariant under the *combination* of three operations: change of sign of all charges, or more generally conversion of every particle into its "antiparticle" (C), space inversion (P) and time reversal (T). This so-called "CPT theorem" seems to be a necessary consequence of quite generic properties of ("local") quantum field theory, irrespective of the details of the interactions etc., and is therefore strongly believed by most physicists; in the few cases when it has been experimentally tested it has indeed seemed to hold. Now it has been known since 1964 that the laws of nature at the elementary-particle level are nearly but not quite invariant under the combination of C and P: this lack of complete invariance is in some sense the "weakest" effect known in physics (at the laboratory level-but see below) but it is nevertheless firmly established. Consequently, if one believes the CPT theorem one must conclude that time reversal (T) is also not an exact symmetry. Thus, if one could make a sufficiently fine-scale movie of the behavior of elementary particle of the appropriate sort, one would be able to tell whether or not it was being run backwards–always assuming the people in white coats in the background are right handed humans rather than left-handed antihumans! I.e. in the actual world we inhabit (where "matter" dominates over "antimatter" by a very wide margin), the laws of nature at the microscopic level are not exactly invariant with respect to time reversal.

Does this feature of elementary particle physics have any relevance to the problem of the arrow of time? Most people think not, if only because of the extreme weakness of the effect (so weak that to date it has never been observed directly and is only inferred as described above). This argument could be a little dangerous, because it not infrequently turns out in particle physics that an interaction which is small at the laboratory energy scale (such as the so-called "weak interaction" responsible for most radioactive decays) nevertheless becomes strong and even dominant at a higher energy scale. Indeed, there are plausible scenarios which use the effect of the CP- (or T-) violation at an early stage in the evolution of the Universe to account for the observed dominance of matter over antimatter (c.f. lecture 26). A more telling objection, however, against the relevance of this T-violation effect to the problem of the "arrow of time" is that prima facie it just does not seem, intuitively speaking, to introduce the right *kind* of asymmetry. This may become clearer below.

Our conclusion, then is that with the exception of this "small" T-violation effect at the level of elementary particles the microscopic laws of physics are per se invariant under time reversal: any apparent lack of symmetry can always be traced to the selection of initial rather than final conditions.

When we turn to the macroscopic world, things are manifestly quite different. Just about any movie we can take, of any reasonably extended part of this world, will show whether it is being run forward or backward - and this is true, at least at first sight, whether or not human beings or even any kind of life is involved. (For example, even if the movie is of a desolate mountain face, we would be surprised to see a rock rise out of the scree at the bottom, move upwards and attach itself to the mountain face!) Needless to say, this asymmetry with respect to the direction of time is embodied in the phenomenological laws of macroscopic physics. A simple example is the introduction of a frictional term into Newton's laws of motion for, say, a pendulum (or into the forces which come into it). Typically the relevant force is proportional to *velocity*. Thus, when we reverse the time, the force changes sign but the acceleration produced by it does not - the equation is no longer invariant under time reversal. In effect, the mechanical (macroscopic) energy of motion of the pendulum is gradually converted, via this term, into heat (the motion of the air molecules, etc.), and this process is manifestly *irreversible*. All such examples can be regarded as special cases of the second law of thermodynamics, namely that the entropy of a closed system can only increase in time or at best stay constant, never decrease:

$$\Delta S \geq 0.$$

We must now ask what is the basis for this statement, given that the microscopic laws of physics show no such asymmetry?

At first sight the answer is obvious and can be seen directly in the "toy model" we used to illustrate the general statistical-mechanical basis of the entropy concept in lecture 24: In that case, the entropy $S(M_1)$ as a function of the subsystem magnetization M_1 is just a measure of the number of configurations compatible with that value of M_1. If we start the system with a value of M_1 which corresponds to only a few configurations (low entropy) and "let it go", it will eventually explore all the configurations which it is allowed to and hence spend most of its time with those values of M_1 which correspond to the maximum number of configurations, i.e. maximum entropy. While, as we have seen in lecture 24, there are some technical problems associated with the possible failure of ergodicity, they in no way affect the *qualitative* conclusion† that entropy is bound to increase as a function of time (or at best, in an extreme case, stay constant, but certainly never decrease). One can go through a similar analysis for any number of other examples: for example, when a gas of N molecules is contained by a shutter in half the volume of a closed box, and the shutter is then removed, the molecules of the gas will of course rapidly spread out to fill the box as a whole, and any subsequent observation is astronomically unlikely to find all of them (or even a fraction appreciably different from 1/2) occupying the original sub volume. (Formally, in this example the role of M_1 is played by the number N_1 in the original subvolume or more precisely by $N_1 - \frac{N}{2}$). So at first sight the statement that quite generally entropy tends to increase in time is an immediate and natural consequence of its statistical-mechanical definition. Why should there be any problem?

To see why there is a problem, let's take a simple everyday analogy-the shuffling of a deck of cards. Suppose we start off with a very "special configuration, say one in which the first 13 cards are all spades (though not necessarily in any particular order), the next 13 hearts etc. We then shuffle them in the usual random way. We expect intuitively that the "disorder" of the deck will increase as we shuffle. In particular, it is extremely unlikely that any configuration of the above "special" type will subsequently recur. So the "disorder", which is the qualitative analog of entropy in this model, certainly seems to be increasing with time. So far, so good.

But now suppose that we are told by a friend whom we trust that he has been shuffling the deck in the usual random way for some time, without in any way "cheating". We inspect the deck and, lo and behold, we find an arrangement of the "special" type! No doubt our first, instinctive, reaction is to suspect our friend of dishonesty. But suppose he convinces us otherwise (e.g. by providing a videotape of the shuffling process). What then? Well, it is clear that the "inverse" of a shuffling operation (at least if properly conducted) is itself a shuffling operation, i.e. the "randomization" works just as well backwards in time as forwards. So we would be forced to the conclusion that the deck was more disordered in the past then it is now, as well as being likely, if the shuffling process goes on, to be more disordered in the future. The apparent basis for the distinction between past and future, in *the randomization process itself* has vanished!

So what is left on which to base the distinction? Clearly, in this kind of case it has to do with our strong intuitive sense that when we come across a very "special " (hence a priori improbable) situation it is likely to be the result of the actions of some human agency. That is, we are very used to the situation in which we (or some other human being) "prepare" a statistically unlikely state of the system and thereafter refuse to intervene (in other than a random way - the fact that in the above example the random shuffling is actually carried out by a human is clearly irrelevant, we could as well replace

†Many discussions of this subject manage to convey the impression that the question of the arrow of time is somehow intimately connected with the ergodic problem. I believe this is a total misconception: ergodicity is relevant only to the question of how fast or how completely equilibrium is obtained, not to the direction in which it occurs.

him by a machine). To put it more formally, we are very familiar with situations in which the initial condition, and these alone, are set by some human agency. On the other hand we are quite unfamiliar with situations in which the final conditions (either alone or not) are set by human agency, (a process sometimes known as "retroparing" the state of the system): it somehow seems difficult even to imagine what this would be like. So, at least in this kind of context, our instinctive sense of the "arrow of time" seems to be connected with the fact that while we can "prepare" states of a physical system, we do not seem able to "retropare" them.

Yet this statement needs closer examination. At the level of simple Newtonian mechanics, for example, there doesn't seem to be any obvious difficulty about "retroparation". For example, let us suppose we are given an ensemble - for practical reasons it would have to be a "time" ensemble - such that the system in question is a billiard ball bouncing around on a frictionless billiard table. We can perfectly well select those systems and only those in which within a certain small stated error, the position at some specified "final" time t_f had the value $x(t_f)$ and the velocity a specified value $v(t_f)$; and we can then integrate Newton's equation of motion backwards in time to obtain the position and velocity at any earlier time. At first sight, at least, this is the exact time-reverse of the procedure we normally use in "predictive" Newtonian mechanics .

As we have already noted, the same state of affairs obtains, at least to an extent, in the quantum mechanics of simple systems: under appropriate conditions we can select a QM ensemble by its final state, and make probabilistic "retrodictions" about its behavior at an earlier time on the basis of this information. Although I do not know of specific experiments which have been done along these lines, there are certainly experiments in the literature on the related "time-symmetric" setup, and there seems no reason to doubt that the predictions (or rather retrodictions!) of QM would also hold good in the purely "retrodictive" scheme.

This, at first sight it seems that neither in classical nor in quantum mechanics is there any fundamental reason to prefer "preparation" to "retroparation". However, we now note a crucial point: In both cases, the assignation of a "state " to the ensemble in question has to be made on the basis of a *measurement* of the relevant variables, and moreover a measurement which is sufficiently "complete " in the context of the operation in question. (In the case of a non-chaotic ("integrable") classical system, we can tolerate a certain small error in the assignment of $x(t_f)$ and $v(t_f)$: in the QM case, (say for definiteness the case of photon polarization) we would need to specify exactly which polarizer the ensemble in question is specified as passing at t_f). But, as I have emphasized in an earlier lecture in the context of ordinary "predictive" quantum mechanics, the assignation of a QM state on the basis of a complete set of measurements (or even of complete "filtering") is actually very much the exception rather than the rule: moreover, even when dealing with classical systems, the moment we have to deal with anything much more complicated than a set of a few billiard-balls it is clear that the realization of a complete set of measurements becomes in practice highly unrealistic. It is actually helpful for pedagogic purposes to discuss the QM case first.

So, how in practice do we prepare a QM ensemble in a way which enables us to assign it an initial state? (Here I recapitulate, for convenience, some considerations already discussed in lecture 21). Consider a typical example - we wish to measure the electrical polarizability of the He atom in its groundstate, and for this purpose we clearly need to prepare an ensemble of He atoms which are known to be in their (electronic) groundstate and whose electrons can hence be assigned a QM wave function (state vector) ψ_0 (The COM motion of the He atom is irrelevant for our purposes and can be ignored). What in practice we might do is to take as our ensemble a beam of He atoms issuing from a small aperture in the wall of an oven maintained at some reasonable temperature T, say $\sim$2000 K. Why are we so sure that the electrons in the atoms of such a beam are in fact in their groundstate? Essentially for thermodynamic or statistical-mechanical reasons: the energy needed to excite the first excited state of He is $\sim 10^6$ K, and if the system is in thermal equilibrium, then according to the standard results of statistical mechanics the probability of it being excited at a temperature of 2000 K is so tiny as to be completely negligible. How does this state of affairs come about? Suppose that initially the electron in a given atom was not in the groundstate but in some excited states (so that its wave function is *not* ψ_0). As a result of interaction with the radiation field in the oven, that electron is likely eventually to drop down to the groundstate, emitting one or more photons in the process. If the oven is indeed at a "reasonable" temperature, the possibility of the reverse process (re-excitation with absorption of a photon) is completely negligible. It is clear that we have in this way introduced a time-asymmetric aspect, *so long as we consider only the electron (atom)* and not the whole atom-radiation field system: various different initial states *of the atom* can lead to the same final state, but the converse in not true. It is now clear why we cannot "time-reverse" the situation, e.g. by choosing as our ensemble a beam of atoms which is directed *into* the oven: the fact that, eventually, the atoms come to thermal equilibrium and hence the electrons are indeed in their groundstate gives no information on what state they were in when they entered the oven!

It is now clear that the situation with preparation and retroparation in QM (under most realistic conditions) is really just a special case of a much more generic state of affairs: Generally, the kind of knowledge we are likely to be able to acquire in real life may enable us to specify (some aspects of) the present and future behavior of the system but it will not, without further a priori knowledge, allow us to specify its past. How is this compatible with the time-reversibility of the microscopic laws of physics? Essentially because we have not been able to obtain *complete* information. In the case of the atomic electron, which starts from an excited state and drops down to the groundstate, emitting a photon, it is of course quite true that the reverse process (excitation with absorption of an incoming photon) is equally compatible with the microscopic laws of QM. However, to get it to "go" one would need to start with a very exact configuration of the photon field, which in practice would be almost impossible to measure, even if it were not anyway exceedingly improbable (at temperatures $\sim$2000K) from the point of view of the Principle of Indifference. (This asymmetry can be regarded as a special case of the "electromagnetic arrow" of time, which is discussed briefly below).

We can make a similar point with respect to our toy model of the magnetic system:

We can, most certainly, prepare it in a microscopic state corresponding to a highly nonequilibrium value of M_1, and then leave it to itself, one which it will evolve to a macroscopic state with the equilibrium value (zero) of M_1. On the other hand, it is impossible in practice to timereverse the process, i.e. to prepare it in a state which, although it has $M_1 = 0$, is guaranteed to evolve to a highly nonequilibrium value of M_1. Why not? Simply because the vast majority of configurations with $M_1 = 0$ will not behave in this way, and although there is a small subset which will, picking them out would require a delicacy of information which in practice it is impossible for us realistically to acquire. By the same token, we cannot realistically make sufficiently delicate measurements on the final ("equilibrium") state to infer that it has evolved from one particular nonequilibrium state rather than a variety states.

So far, so good: I do not think many people would quarrel with the above statements. However, in the context of "retroparation" it is not at all clear that they answer our original question. All that the above arguments show is that it is impossible in practice to make sufficiently delicate final-state measurements to "retrodict" to an earlier much-Iower-entropy state. We have also sketched, in the specific QM case examined, why it would be impossible to "retropare" the QM ensemble by *one* obvious means which does not involve explicit measurement. However, that is a far cry from proving that "retroparation" of highly nonequilibrium states is in general a priori impossible, particularly in macroscopic (thermodynamic) systems. If one thinks about the actual operations involved e.g. in preparing the nonequilibrium magnetic state (move magnet into place: switch magnet on, wait a few minutes or hours, switch magnet off. . . etc.), it seems that our reluctance to envisage that we could "retropare" it in the same way has to do with our gut feeling than human action at a given time t can affect the physical state of affairs at times $(t' \geq t)$ but not at earlier times $(t' \leq t)$, or more generally that we "remember the past and affect the future" and not vice versa. It seems to me probable that this "psychological" arrow of time has in the last resort to do with the above considerations concerning the impossibility of fixing very delicate final conditions; and hence with the more general thermodynamic problem; but until the question is analyzed in a great deal more specific detail than it has usually happened in the literature to date, I do not think one can be at all sure of this (and in particular I think that Hawking's claim (p. 152) to have "shown" that the psychological and thermodynamic arrows are essentially the same, on the basis of an argument which is at best suggestive, is a bit too glib.)

Let's assume for the sake of argument that the connection, or equivalence, between the psychological and thermodynamic "arrows" is established. This then settles the question of why, given that thermodynamics is asymmetric in time, entropy increases in the direction of the "future" (which we can define to be the direction in which according to our common-sense conceptions, we can affect events but cannot remember them). However, we still have not answered the question why there is a thermodynamic arrow of time at all, let alone whether it makes sense to ask whether it can be related to some other even more profound asymmetry. (Incidentally, the considerations advanced here (with the exception of course of the QM one) are not at all new : they go back at least to Boltzmann, one of the founders of modern statistical mechanics, in the mid-nineteenth century).

At this point it is helpful to review explicitly the traditional five "arrows of time".[‡] The first two, the thermodynamic and the psychological, we have already introduced; the fifth is related to the weak T-violating interactions of particle physics and is at present generally believed not to have much to do with the other few. The remaining two are, first, the electromagnetic arrow; Although the equations of propagation of electromagnetic waves, and also of their emission and absorption by radiation sources (e.g. a radio antenna) are symmetric with respect to time, we normally impose so-called boundary conditions on the solutions in such a way that antennae, etc., *radiate* EM waves "to infinity" rather than sucking them in from infinity. The rationale usually given for this choice is that to achieve a situation in which the waves were focused in on the antennae in just such a way as to be absorbed would require a very delicate initial condition "at ∞". It is clear that the argument here is very similar to that used in connection with the thermodynamic arrow (it was in fact implicitly used in our discussion of the He atom problem), and it is tempting to regard the EM arrow as in some sense a special case of the thermodynamic one. It is interesting, however, that just as in the case of QM a "time-symmetric" electrodynamics which treats both directions on an equal footing has been constructed: so perhaps in some sense one can say that the status of EM in the context of time-reversibility is similar to that of QM, in that the asymmetry is not characteristic of the formalism itself but rather of the kind of information which has to be used in selecting one type of solution rather than the other as "physically relevant". The final "arrow" usually defined is the cosmological one: according to our current picture (see lecture 26) it is an experimental fact that the Universe as a whole is expanding rather than contracting. Presumably associated with this phenomenon is the fact that it also seems to be getting "clumpier" as galaxies and eventually individual stars form, and that the stars are burning nuclear fuel and using it to emit light "to ∞" rather than sucking in light "from ∞" and using it to dissociate e.g. He into H. However, any attempt to make this argument rigorous runs into quite serious conceptual difficulties concerning the description of the so-called hot big bang (see lecture 27).

Summing up, we can say that there is a picture of the relationship of the various arrows of time which seems at least free of any obvious internal inconsistencies, as follows: The fundamental "arrow" is the cosmological one; the electromagnetic arrow is then determined by it (i.e. it is not an accident that the stars radiate light to infinity rather than sucking it in); because radiation is essential to life, this then uniquely determines the direction of biological differentiation in time, and hence our psychological sense of "past" and "future"; and finally the thermodynamic arrow is connected in inanimate nature with the electromagnetic one, and in the laboratory with the psychological one as discussed above. If this is correct, then the fact that the Universe is expanding rather than contracting is as it were not a contingent one! (I think Kant would have been delighted with this conclusion). However, while it easy to think, at each step, of plausible reasons why the required connection *might* hold, it is probably fair to say that

[‡] Or six, if one prefers to add a "biological" arrow which is different from the "psychological" one.

in no case has a connection been established with anything approaching rigor, and I for one certainly would not want to stake my life on the assertion, for example, that conscious life must be impossible in a contracting Universe.