TIME-REVERSAL INVARIANCE

1. THE NEWTONIAN PICTURE OF THE WORLD

What I want to talk about here is a certain tension between fundamental microscopic physical theory and everyday macroscopic human experience, a tension that comes up (more particularly) in connection with the question of precisely how the past is different from the future.

And the fundamental theory in which it will work best to start that talk out, the fundamental theory (that is) in which this tension is at its purest and most straightforward, is the mechanics of Newton. Never mind (for the moment) that the mechanics of Newton turns out not to be the mechanics of the actual *world*.[1] We'll talk about that later.

▲▲▲ According to Newtonian mechanics, or at any rate according to the particularly clean and simple version of it that I want to start off with here, the physical furniture of the universe consists entirely of point particles. The only *dynamical variables* of such particles—the only physical attributes of such particles that can *change with time*—are (on this theory) their *positions*; and (consequently) a list of what particles exist, and of what *sorts* of particles they are,[2] and of what their positions are at all times, is a list of absolutely everything there is to say about the physical history of the world.[3]

1. What the actual *world* turns out to be (insofar as we can tell at present) is *quantum* mechanical, or quantum field theoretic, or quantum string theoretic, or something like that.

2. That is, of their *non*-dynamical properties: their *masses* and their *charges* and so forth.

3. This is certainly not to deny that there are such things in the world as *extended objects*; the idea is just that all the facts about objects like that (facts, say, about where the tables and chairs are, and about who punched whom, and about who said what, and so on) are determined, in principle, by the facts about the *particles* of which those objects are *composed*.

And Newtonian mechanics is *deterministic.* Given a list of the positions of all the particles in the world at any particular time, and of how those positions are *changing,* at that time, as time flows forward, and of what *sorts* of particles they are, the universe's entire history, in every detail, from that time on, can in principle be calculated (if this theory is true) with certainty.

▲▲▲ All this will be worth spelling out in some detail.

Let's start slow.

The rate at which some particular particle's position is changing, at some particular time, as time flows forward, is called its *velocity* at that time. And the rate at which such a particle's *velocity* is changing, at some particular time, as time flows forward, is called its *acceleration* at that time.

And what Newtonian mechanics has to say about the motions of particles, the *entirety* of what it has to say about the motions of particles, is that a certain breathtakingly simple mathematical relation—$F = ma$—invariably holds between the force on any particle at any particular instant, and its acceleration at that instant, and its mass.

Let's say a bit about where forces come from.

What happens in the most familiar cases (think, for example, of gravitational attraction, or of electrostatic repulsion) is that forces arise exclusively *between pairs of particles,* and (moreover) that the forces which any two particles are exerting on each other at any particular instant depend only on *what sorts of particles they are* and on their relative *positions.*

And the third and final fundamental principle of the Newtonian picture of the world (the first is that the world consists entirely of particles, and the second is the relation between F and m and a) is that as a matter of fact all the forces there *are* are like that.

And so (on this picture) a specification of the positions of all the particles in the world at some particular time, and of what sorts of particles they are, amounts (at least insofar as these familiar sorts of forces are concerned) to a specification of what the forces on each of those particles are at that time as well.

Good. Let's see how all this results in precisely the sort of determinism I said it did above.

Call the "initial" time, the time we will want to calculate forward from, $t = 0$.

And suppose that what we're given at the outset are the positions of all the particles in the world (or in some isolated subsystem of the world) at $t = 0$ (call those x_0^i), and their velocities at $t = 0$ (call those v_0^i), and their masses (m^i), and their electric charges (c^i), and all their other intrinsic properties.

And let's say that what we would like to calculate is the positions of all these particles at $t = T$.

The most illuminating way of doing that, I think, will be by means of a succession of progressively better and better *approximations.*

The first goes like this: calculate the positions of all the particles at $t = T$ by supposing that their velocities are constant—and equal to their above-mentioned values (v_0^i) at $t = 0$—throughout the interval between $t = 0$ and $t = T$.

This calculation will place particle i at $x_0^i + v_0^i T$ at $t = T$; but it hardly needs saying that this calculation is not a particularly accurate one, because (unless it happens that no forces are at work on any of the particles here) the velocities of these particles will in fact *not* remain constant throughout that interval.

Here's a somewhat better calculation:

Divide the time-interval in question into two, one extending from $t = 0$ to $t = T/2$ and the other extending from $t = T/2$ to $t = T$. Then calculate the positions of all the particles at $T/2$ by supposing that their velocities are constant—and equal to their values at $t = 0$—throughout the interval between $t = 0$ and $t = T/2$ (this will place particle i at $x_0^i + v_0^i(T/2)$ at $T/2$).

Then calculate the *forces* acting on each of the particles at $t = 0$ (what those forces are, remember, will follow from the *positions* of those particles at $t = 0$ together with their masses and their charges and their other internal properties—all of which we are given at the outset).

Then calculate each particle's *velocity* at $T/2$ by plugging those forces into the above-mentioned law of motion (plugging them, that is, into $F = ma$), and assuming that the particles' accelerations are constant throughout the interval from $t = 0$ to $t = T/2$—and are equal to their values at $t = 0$ (this will put the velocity of particle i at $v_0^i + a_0^i(T/2)$, where a_0^i is equal to the force on particle i at $t = 0$ divided by particle i's mass).

Then, finally, calculate the position of particle i at $t = T$ (which is what we're after here) by supposing that this particle maintains this *new* velocity throughout the interval between $t = T/2$ and $t = T$.

This calculation isn't going to be perfect either, but (since the intervals during which the velocities of the particles are erroneously presumed to be constant are shorter here than in the previous calculation) it amounts to a clear improvement.

And of course this improvement can *itself* be improved upon by dividing the interval further, into *four* intervals. That calculation will proceed as follows.

To begin with, the approximate positions of all the particles at $t = T/4$ can be calculated (just as we did above) from the positions and velocities at $t = 0$ alone. Moreover, the forces on all the particles at $t = 0$ can now be read off (as we did above) from their intrinsic properties and their positions at $t = 0$, and thus (with the aid of $F = ma$) the approximate *velocities* of all the particles at $t = T/4$ can be deduced as well. And so what we now have in hand is a list of approximate positions and approximate velocities and particle-types at $T/4$, and of course those approximate positions and particle-types can *now* be used to determine the approximate *forces* on all the particles at that time as well, and *that* will in turn allow us to determine the positions and velocities and forces at $T/2$, and so on.

And then we can go on to eight intervals, and then to sixteen. And as the number of intervals approaches *infinity*, the calculation of the particles' positions at $t = T$ patently approaches perfection. And it happens that there is a trick (and the name of that trick is *the calculus*) whereby—given a simple enough specification of the dependence of the forces to which the particles are subjected on their relative positions—that perfect calculation can actually and straightforwardly be carried out.

And of course T can be chosen to have any positive value we like. And so the positions of all the particles in the system in question at any time between $t = 0$ and $t =$ infinity (and with that the *velocities* of all those particles between those times, and their energies, and their angular momenta, and everything else about them) can in principle be calculated, exactly and with certainty, from the positions and velocities and intrinsic properties of all those particles at $t = 0$.[4]

2. TIME-REVERSAL INVARIANCE IN THE NEWTONIAN PICTURE

Newtonian mechanics has a number of what are referred to in the literature as *fundamental symmetries*; and what that means is that in Newtonian mechanics there are certain sorts of facts about the world which—as a matter of absolutely general principle—*don't make any dynamical difference.*

Suppose, for example, that we are given the present positions and velocities of all the particles in the world, and that we are told what *sorts* of particles they are, and that we would like to calculate their positions and velocities (say) two hours from now. It is an extremely straightforward consequence of

4. Note, by the way, that this *overall* determinism of the evolution of a universe of classical particles (whereby all the present positions and velocities determine all the *future* ones) can invariably be *taken apart* (as it were) into *separate* determinisms running in *parallel.*

Consider, for example, a universe consisting of a single classical particle, which (being all alone in the world) is never subjected to force.

Exactly two logically independent pieces of information—two *numbers*—are required in order to specify fully the present physical situation of such a particle, and to nail down (by means of the classical laws of motion) all its future situations. The pair of numbers we're *usually* presented with in contexts like this is the particle's present *position* (x_0) and its present *velocity* (v_0); but there are, of course, an infinite collection of *other, mathematically equivalent,* such pairs ($x_0 + v_0$ and $x_0 - v_0$, for example, or $5x_0 + 14v_0$ and $36x_0 - 7v_0$, or x_0 and $x_0 + 23v_0$, or what have you) which will patently do just as well.

Imagine, then, that we are informed of the values of the quantities v_0 and $x_0 + Tv_0$ (where T is some number), and that we would like to deduce, by means of the laws of classical mechanics, the values of the velocity and the position of the particle at some later time $t = T$. The calculations involved here are trivial, of course, but what I want to draw the reader's attention to here is that the outcome of the first of those calculations (which is: $v_T = v_0$) will depend *exclusively* on the value of v_0 and *not at all* on the value of $x_0 + Tv_0$, and that the outcome of the *second* of those calculations (which is: $x_T = x_0 + Tv_0$) will depend exclusively on the value of $x_0 + Tv_0$ and not at all on the value of x_0. And so if we were informed only of the value of v_0, and were left entirely in the dark about $x_0 + Tv_0$, we could nonetheless deduce, with certainty, from the laws of classical mechanics alone, the value of v_T; and if we were informed only of the value of $x_0 + Tv_0$, and were left entirely in the dark about v_0, we could nonetheless deduce, with certainty, from the laws of classical mechanics alone, the value of x_T.

And this turns out to be an absolutely general phenomenon, which applies to classical worlds consisting of any number of particles, interacting with one another in any way you like: the velocity (say) of particle number 789 at $t = 6$ years will necessarily be equal to some definite function of the positions and velocities of all the particles in the world at $t = 0$, and the position of particle number 3 at that time will necessarily be some *other* such function, and the values of those two functions will necessarily be logically *independent* of each other.

the Newtonian picture of the world I described above that that calculation can be carried through in perfect ignorance of what time "now" is. If the classical laws of motion entail that a certain set of positions and velocities at 4:02 evolves into a certain *other* set of positions and velocities at 6:02, then those laws will necessarily *also* entail that the first set at 4:07 will evolve into the second set at 6:07, and that the first set at 12:23 will evolve into the second at 2:23, and so on. Any sequence of position and velocity values for every particle in an isolated collection which is in accord with classical mechanics and which begins at time t would necessarily (to put all this slightly differently) *also* be in accord with classical mechanics if it were to begin instead at t'. And in virtue of all that, Newtonian mechanics is said to have *time-translation-symmetry*, that is, it is said to be *invariant* under translations like that.

And it has a number of other significant invariances as well: absolute *positions* don't play any role in Newtonian mechanics (although the positions of particles *relative to one another* certainly do), and neither do absolute *directions in space*, and neither do absolute *velocities.*[5]

▲▲▲ And neither does *the direction of time.*

Let's stop and talk about that some.

Imagine, to begin with, watching a film of a baseball which is thrown directly upward, and which is subject to the influence of the gravitational force of the earth; and then imagine watching the same film *in reverse.* The film run forward will depict the baseball moving more and more slowly upward; and the film run in reverse will depict the baseball moving more and more quickly downward. What *both* films will depict, though, is a baseball which (whatever its *velocity*) is accelerating, constantly, at the rate of 32 feet per second per second, *in the direction of the ground.*

And this is (of course) an absolutely general phenomenon: the apparent velocity of any particular material particle at any particular frame of any film of any classical physical process run forward will be equal and opposite to the apparent velocity of that particle at that frame of that film run in reverse; but the apparent *acceleration* of any particular particle at any particular frame of the film run forward will be *identical, both* in magnitude *and* in direction, to the apparent acceleration of that particle at that frame of the film run in reverse.[6]

5. And all of these invariances are (by the way) also invariances of Maxwellian electrodynamics, and of relativistic quantum string theories, and of every other fundamental theory in the canon too.

Now, the Newtonian law of motion (which is, remember, the entirety of what the Newtonian picture of the world has to say about the motions of particles) is that a certain mathematical relation holds, at every instant, between mass and force and acceleration. And of course the mass of any particular particle at any particular frame of the sort of movie we've been talking about depends on nothing other than *what* particular particle it *is;* and the *force* on any particular particle at any particular frame of the sort of movie we've been talking about depends on nothing other than what particular *set* of particles happens to exist, and what those particles' spatial distances from one another at that frame happen to be; and what we've just seen is that the *acceleration* of any particular particle, at any particular frame of such a movie, will be entirely independent of the direction in which the film is run. And so if a certain film, run forward, depicts a process which is in accord with Newtonian mechanics, then, necessarily, the same film run in *reverse* will depict a process which is in accord with Newtonian mechanics as well.[7]

6. The proof is trivial. Let $x(t)$ represent the apparent trajectory (that is, the apparent position as a function of time) of the particle depicted in the film run forward, and let $v(t)$ represent the apparent *velocity* (that is, the apparent *derivative* of the position with *respect* to time) of that particle, and let $a(t)$ represent the apparent *acceleration* (that is, the apparent derivative of the *velocity* with respect to time) of that particle. Then the apparent trajectory of the particle depicted in the film run in *reverse* will be $x(-t)$. And of course the apparent *velocity* of a particle whose apparent trajectory is $x(-t)$ is (by definition) $dx(-t)/dt$, which is equal (by the chain rule) to $-v(-t)$, which is (of course) the negative of the velocity of the particle (at the frame in question) depicted in the film run forward. By contrast, the apparent *acceleration* of a particle whose apparent trajectory is $x(-t)$ is (by definition) $(d/dt)(dx(-t)/dt)$, which is equal (by the upshot of the previous sentence) to $(d/dt)(-v(-t))$, which is equal (by the chain rule) to $-(-a(-t))$, which is equal (because (-1) times (-1) is $(+1)$) to $a(-t)$, which is (of course) the *same* as the acceleration of the particle (at the frame in question) depicted in the film run forward.

7. Let's put this a bit more formally. Consider a history $\{x_1(t) \ldots x_N(t)\}$ of some isolated collection of N particles. What's just been shown is that if

$$d^2x_i(t)/dt^2 = F_i(x_1(t) \ldots x_N(t))$$

for all i (where x_i is the position of particle i, and F_i is the force *on* particle i), then necessarily

$$d^2x_i(-t)/dt^2 = F_i(x_1(-t) \ldots x_N(-t));$$

which is to say that if $\{x_1(t) \ldots x_N(t)\}$ is a solution to the Newtonian equations of motion, then necessarily

$$\{x_1(-t) \ldots x_N(-t)\}$$

is too.

And so it is a consequence of Newtonian mechanics that nothing in the laws of nature can be of any help whatsoever in deciding *which way* any film is ever being run. And so it is a consequence of Newtonian mechanics that whatever can happen can just as easily, just as *naturally*, happen *backward.*[8]

And so the Newtonian-mechanical instructions for calculating future physical situations of the world from its present physical situation turn out to be identical to the Newtonian-mechanical instructions for calculating *past* physical situations of the world from its present physical situation. The instructions for calculating (say) the positions of all the particles in the world ten minutes from now are to plug the present positions of all those particles, and the rates at which those positions are changing as time flows forward, into a certain algorithm (the sort of algorithm we explicitly went through above); and the instructions for calculating the positions of those particles ten minutes *ago* are to plug their present positions, and the rates at which those positions are changing as time flows *backward*, into precisely the *same* algorithm.

And so if we are told the positions of all the particles in the world at present, and if we are told the rates at which those positions are changing as time flows toward some other moment *M*, and if we are told the size of the time-interval that separates *M* from the present, then we can in principle calculate the positions of all the particles in the world at *M*, with certainty, without *ever* having been told (and also without ever *learning*, as the calculation *proceeds*) whether *M* happens to lie *after* the present or *before* it.

And so (if the laws of Newtonian mechanics are all the fundamental natural laws there are) there can be no lawlike asymmetries whatsoever between past and future.

▲▲▲ And the thing is that all this is wildly at odds with our everyday experience.

To begin with, every corner of the world is positively *swarming* with ordinary physical processes that don't, or don't regularly, or don't naturally, or don't familiarly, happen backward (the melting of ice, say, or the cooling of soup, or the breaking of glass, or the passing of youth; whatever).[9]

And (on top of that) there's an asymmetry of *epistemic access:* our capacities to know what happened yesterday, and our methods of *finding out* what happened yesterday, are as a general matter very different from our capacities to know and our methods of finding out what will happen *tomorrow.*[10]

And (on top of *that*) there's what you might call an asymmetry of *intervention:* it seems to us that we can bring it about that certain things occur—or that they don't—in the future, but we feel absolutely incapable of doing anything at all about the *past.*

▲▲▲ And that's the tension I mentioned before. And that's more or less what I want to talk about in this book. Or at any rate that's the *Newtonian* version.

The next thing to do is to generalize it some.

3. TIME-REVERSAL INVARIANCE IN GENERAL

Let's start by thinking through what it means to give a complete description of the physical situation of the world at an instant.

There would seem to be two things you want from a description like that:

a. that it be genuinely *instantaneous* (which is to say that descriptions of the world at different times have the appropriate sort of logical or conceptual or metaphysical *independence* of one another, that a perfectly explicit and intelligible *sense* can be attached to *any temporal sequence whatever* of the sorts of descriptions we have in mind here—whether the sequence happens to be in accord with the dy-

8. Maybe a few of the standard illustrations are in order here. Think, then, of watching films, run forward and run in reverse, of a single particle, alone in the universe, moving (say) to the right; or of two billiard balls colliding; or of a rock moving downward, and accelerating downward, in the gravitational field of the earth.

9. Maybe this is worth belaboring a bit further. Take soup. It isn't that soup never *heats up;* it's (rather) that occasions when soup *does* heat up never look anything at all like mere *temporal inversions,* like *films watched backward* of occasions when soup *cools off.* The former occasions are always *different,* somehow. They involve fires or electrical currents or parabolic mirrors or something like that. And *that's* the point here: that you can bet your life that a tepid pot of soup, in (say) an otherwise empty, cold, closed, insulated room, is invariably and ineluctably in the process of getting *colder.*

10. Note that this is no less a physical business than the stuff about the cooling of soup: this too is about the sequences in which the *states of physical systems* occur; this is about the fact that (say) detailed and accurate depictions of freak accidents (photographic depictions, or tape-recorded ones, or written ones, or ones stored in the physical states of human brains, or whatever) almost never *precede* those accidents *themselves.*

namical laws or not, that any such sequence whatever is *readable—against the background* or *within the context* or *relative to the framework* of the best or last or canonical metaphysical interpretation of whatever complete theory of the world is under discussion—as a *story of the physical world*); and

b. that it be *complete* (which is to say, that *all* the physical facts about the world can be read off from the full temporal set of its descriptions).

Good. Let's call whatever satisfies (a) and (b) an instantaneous physical *state* of the world.

What satisfies (a) and (b) in the *Newtonian* picture (for example) is a specification of the *positions*, at the time in question, of all the particles in the world: no specification of the positions of those particles at any *one* time, or at any *collection* of times, logically entails anything whatsoever about their positions at any *other* time; and given such specifications for *all* times, *everything* about the history of the world can straightforwardly be read off.

What typically gets referred to in the *physical literature* as an "instantaneous state" of a Newtonian-mechanical universe, of course, is a specification of the positions *and the velocities* of all the particles in the world at the time in question. But the trouble with that is just that specifications of the positions *and the velocities* of all the particles in the world at one time are *not* conceptually *independent* of specifications of the positions and velocities of all the particles in the world at all *other* times.[11] The trouble (to put it slightly differently) is that a specification of the positions and the velocities of all the particles in the world at some particular instant is *not* a specification of the physical situation of the world at that instant *alone;* it is *not* a

11. And what I mean here (and maybe this deserves to be emphasized) is *all* other times. The positions and the velocities of any set of particles at any one time are *indeed* perfectly logically and conceptually and metaphysically independent of the positions and the velocities of those particles at any particular *other* time. But suppose that I is some time-*interval* within which some particular time t happens to *fall.* Then the positions and the velocities of those particles at t will certainly *not* be logically or conceptually independent of their positions and velocities throughout the *complement* of t in I.

Think (for example) of a single particle moving uniformly to the right throughout the interval ($t = -1$ minute) to ($t = +1$ minute). And now replace $x(t)$ and $v(t)$ with $x(-t)$ and $v(-t)$, respectively. And note that this maneuver leaves x and v at $t = 0$ (but not at any other time) unchanged. And note that what this maneuver results in is a straightforward logical contradiction.

specification of the physical situation of the world at that instant *as opposed to all others,* at all!

And so the Newtonian laws of motion turn out *not* (exactly) to amount to a deterministic connection between all the states of the world at all times and any single *one* of them. What those laws amount to (if you want to be careful) is a deterministic connection between all the states of the world at all times and all the states of the world throughout any arbitrarily small time-*interval.*[12]

▲▲▲ What is it, then, for something to happen *backward?*

Simple. Suppose that the true and complete fundamental physical theory of the world is something called T. Then any physical process is necessarily just some infinite sequence $S_I \ldots S_F$ of instantaneous *states* of T. And what it is for that process to happen *backward* is just for the sequence $S_F \ldots S_I$ to occur.

▲▲▲ What is it, then, for a fundamental theory of the world to fail to distinguish between past and future?

I mentioned two ideas about that, some pages back, and talked about them as if they amounted to more or less the same thing. One was that the theory entails that whatever can happen can also happen backward, and the other was that the theory offers identical algorithms for inferring toward the future and the past. And these are actually, logically, altogether different propositions. And I want to say something about precisely what their relation is.

Both of them, of course, turn out to be true of Newtonian mechanics; and

12. The laws (that is) turn out to amount to a deterministic connection not between the positions of all the particles in the world at one time and their positions at any other time, but (rather) between the positions of all the particles in the world at one time, and the *rates* at which those positions are *changing* in the immediate vicinity of that time, and the positions at any other time.

And those rates are emphatically *not* features of the physical situation of the world at any particular *instant.* And so (on this way of looking at things) the Newtonian-mechanical laws of motion turn out emphatically *not* to be anything along the lines of a set of rules whereby the world decides, on the basis of its physical situation exactly *now,* what to do exactly *next.* But (come to think of it) they *couldn't* have been. The temporal instants (after all) form a *continuum;* there is *no such thing* as the instant *immediately after* (say) five o'clock.

it goes without saying that one can imagine *other* theories of which *neither* is true; and it turns out (more interestingly) that theories can be imagined of which one of them is true and the other isn't.

There are (to begin with) two entirely distinct ways in which a theory might *fail* to offer us identical algorithms for inferring toward the future and the past.

One—the obvious one—is for the theory in question to offer us an algorithm for calculating toward the past and an algorithm for calculating toward the future and for those two algorithms to be *different.* Here's a theory like that: somewhere in space there is a fixed blue dot. And there are particles. And the law of motion is that each of those particles invariably proceeds toward that dot as time flows forward, and that the *speed* with which any particular particle proceeds toward that dot at any particular time measured in feet per second is equal to the distance between them at that time measured in feet.

But something else can happen too.

Consider (for example) a theory like this: there are, at a *number* of points in space, fixed blue dots. And there are particles. And all the particles invariably move in perfect accord with the Newtonian laws of motion, except that at noon, on a certain particular day, as the clock strikes, each particle jumps, instantaneously, to the nearest blue dot, and thereafter proceeds onward, with its pre-jump velocity, once again in accord with the Newtonian laws, forever after.

The algorithm for determining the future positions of all the particles in the world from their present ones, and from their rates of change as time flows forward, will be perfectly deterministic on this theory. But note that this theory will yield *no algorithms whatsoever* for inferring from times *after* the noon in question to times *before* it.

And note that the two very fanciful theories we've just been talking about will both flatly *deny*—unlike Newtonian mechanics—that whatever can happen to a collection of particles can also happen backward. And as a matter of fact, it turns out that *any* deterministic theory—that deterministic theories *in general*—can allow that whatever can happen forward can also happen backward *only* if the theory offers us identical algorithms for inferring toward the future and the past, and (equivalently) it turns out that deterministic theories can *deny* that whatever can happen forward can also happen backward only if they *fail* to offer us identical algorithms for inferring toward the future and the past.[13]

Indeterministic theories are a bit more complicated. Theories with probabilistic algorithms for inferring toward the future (that is, theories whose laws stipulate the probability that such-and-such goes on at time 2 given that such-and-such goes on at time 1) generally entail nothing about the business of inferring toward the past—and yet many such theories allow that whatever can happen forward can also happen backward.

Consider, for example, a system consisting of a single particle, which can be located in either of two boxes. And suppose that the full theory of the dynamical evolution of this system, so long as it is isolated from outside influences, is that the particle is as likely as not, over any one-second interval, to switch boxes (that is, the full theory of the free dynamical evolution of this system is that the particle's probability of now being in box 1, given that it was in box 1 one second ago, is 1/2; and the particle's probability of now being in box 2, given that it was in box 1 one second ago, is 1/2; and the particle's probability of now being in box 2, given that it was in box 2 one second ago, is 1/2; and the particle's probability of now being in box 1, given that it was in box 2 one second ago, is 1/2).

This theory will entail that the time-reverse of any physically possible free trajectory is *another* physically possible free trajectory—it will entail (as a matter of fact) that the *probability* of any physically possible free trajectory (given its initial state) will be *equal* to the probability of that trajectory's time-reverse (given *its* initial state). And yet this theory will tell us *nothing whatsoever* about the probability that (say) the particle was in box 2 one second ago given that it is currently in box 1—just as the proposition that the probability of a certain die landing on 4, given that it is a fair die, is 1/6,

13. Here's why: if whatever can happen forward can happen backward, then there is a one-to-one mapping—the mapping that takes any trajectory into its *time-inversion*—between physically possible trajectories proceeding from any given present state toward the future and physically possible trajectories proceeding from that same given present state toward the past. Thus, if there is an algorithm whereby the present state plus (say) present-to-future rates of change invariably pick out a *single* possible future trajectory, then there must be only a single possible *past* trajectory with that same present state and the equivalent present-to-*past* rates of change; and that present state, and those present-to-past rates of change, must necessarily pick that past trajectory out in accord with *precisely the same algorithm.*

entails nothing whatsoever about the probability that the die is fair given that it *does* land on 4.[14]

Of course, if any theory whatsoever offers us *both* predictive *and* retrodictive algorithms, and if those two algorithms happen to be *identical*, and if the theory in question entails that a certain process can happen forward, then it will necessarily also entail that the process can happen backward. *That's* what I'll mean, then, from here on, when I speak of a theory as being invariant under time-reversal.

▲▲▲ Good. Let's talk some (with all this now under our belts) about the candidates for a fundamental physical theory that have been taken seriously *since* Newton.

Look, for example, at the classical theory of a universe made up of electrically charged particles and electromagnetic fields. What counts as an instantaneous state of the world, according to classical electrodynamics (which is what that theory is called), is a specification of the positions of all the particles and of the magnitudes and directions of the electric and magnetic fields at every point in space. And it turns out *not* to be the case that for any sequence of such states $S_I \ldots S_F$ which is in accord with the dynamical laws of this theory, $S_F \ldots S_I$ is too. And so this theory is *not* invariant under time-reversal. Period.

And neither (it turns out) is quantum mechanics, and neither is relativistic quantum field theory, and neither is general relativity, and neither is supergravity, and neither is supersymmetric quantum string theory, and neither (for that matter) are any of the candidates for a fundamental theory that anybody has taken seriously since Newton. And everything everybody has always said to the contrary (of which more later) is wrong.

14. Perhaps this is worth spelling out in more detail. The point is just this: if some large collection of particles is known, at time *t*, to be in (say) box 1, and if it is known that none of these particles will be disturbed from the outside over the course of the next second, then it can be reliably inferred that about half of them will be in box 1, and the other half in box 2, at *t* plus one second. But if some large collection of particles is known at time *t* to be in box 1, and if it is known that none of these particles *was* disturbed from the outside over the course of the *past* second, it can *by no means* be inferred that about half of those particles were in box 1 at *t minus* one second, and the other half in box 2. Suppose, for example, that the second collection of particles is itself half of some much *larger* one, all of which were *placed*, at *t* minus one second, in box 1. Nothing about the situations of those particles at *t*, needless to say, will have any bearing whatsoever on the likelihood of *that!*

There is, though, a curious *vestige* of time-reversal invariance in all these theories. There's something about all these theories that *isn't* time-reversal invariance but nonetheless somehow *recalls* time-reversal invariance or *suggests* time-reversal invariance or *smacks* of time-reversal invariance or is capable of *masquerading*, for certain purposes, as time-reversal invariance.

Let's talk about classical electrodynamics again.

It turns out that for every sequence of instantaneous states $S_I \ldots S_F$ which is in accord with the laws of classical electrodynamics, a sequence of the form $\tilde{S}_F \ldots \tilde{S}_I$ will necessarily be in accord with them too, where the only differences between any $\tilde{S}_K$ and its corresponding S_K have to do with where the magnetic fields are pointing. And so classical electrodynamics *does* have what you might call a *partial* time-reversal invariance, a time-reversal invariance insofar as the *positions of the particles* are concerned: classical electrodynamics *does* entail that whatever motions *particles* can execute, they can also (though under *other circumstances*, with differently directed magnetic fields around) execute backward.

And so the unbreaking of glass can be no less in accord with the laws of Maxwellian electrodynamics than the breaking of glass is, and the spontaneous heating of soup can be no less in accord with Maxwellian electrodynamics than its spontaneous cooling is, and the coming of youth can be no less in accord with Maxwellian electrodynamics than its passing is, since (when you come right down to it) what it is for glass to break or for soup to cool or for people to get older-looking is just for the particles that make them up to assume certain particular sequences of positions. And so classical electrodynamics (even though it is decidedly *not* invariant under time-reversal) is every bit as much at odds with the time-directedness of our everyday macroscopic experience as Newtonian mechanics is.

And the broad outlines of all this have remained more or less in place, or at any rate they have suffered only two further complications (of which more in a minute), ever since.

None of the fundamental physical theories that anybody has taken seriously throughout the past century and a half is (as I mentioned above) invariant under time-reversal.

Most of them *are* invariant under time-reversal, though, *insofar as the positions of particles are concerned.* For most of them (more particularly) there is some fairly straightforward transformation linking every state S_K with another state $\tilde{S}_K$—a transformation which varies from theory to theory, but which in every case has the property that it leaves the positions of particles unaffected—such that if $S_I \ldots S_F$ is in accord with the theory, then $\tilde{S}_F \ldots \tilde{S}_I$ is too. And (once again) since our everyday macroscopic experience is first and foremost an experience of *the positions of material bodies,* those theories are all at odds with the time-directedness of that experience in much the same way as Newtonian mechanics is.

And there are two curious pieces of contemporary physical theory that appear *not* to be invariant under time-reversal, even in the limited sense we have just been talking about.

One concerns the decays of certain sub-atomic particles. And those decays (insofar as anybody has yet been able to imagine) have nothing whatsoever to do with the time-directedness of our everyday experience.

The other is more interesting. There is—very briefly—a problem at the foundations of quantum mechanics. And there are a variety of proposals around for modifying quantum mechanics in such a way as to make that problem go away. And it happens that *some* of those proposals (though not *all* of them, and not even *most* of them) involve violations of partial time-reversal invariance too. And *those* violations (if there turn out to *be* any, if the proposals in question turn out to be *right*) might well have something to do with the time-directedness of our everyday experience. And we will be talking a great deal about all that in the last chapter of this book.

But let's leave it aside for the time being. There might well turn out (after all) *not* to be any such violations. And if there aren't, we're going to need to figure out how to alleviate this tension, or how to *live* with this tension, or what to *make* of this tension, *without* them. And the right place to start would seem to be Newtonian mechanics, where the tension is particularly simple and stark. Whatever we manage to discover there will presumably apply in the more up-to-date cases too.

4. TIME-REVERSAL INVARIANCE IN THE PHYSICAL LITERATURE

Before we get to that, though, it ought to be acknowledged in somewhat more detail that the thrust of what was reported in the previous section is quite radically at odds with what it says in the textbooks.

To begin with, what the books say it is to specify the world's complete physical situation at a certain instant is to specify what you might call its complete *dynamical conditions* at that instant, to specify (that is) all the information about the instant in question—or all the information which can in one way or another be uniquely *attached* to the instant in question—which is required in order to bring the full predictive resources of the *dynamical laws of physics* to bear. And the trouble with that (the trouble with it—that is—*as a conception of the situation of the world at an instant*) is that dynamical conditions of the world at *different* instants can turn out, as I have repeatedly emphasized, not to be logically or conceptually or metaphysically *independent* of one another.

Take Newtonian mechanics. The dynamical conditions of a Newtonian universe at any instant are not the *positions* of all the particles in the world at that instant but the positions *and the velocities* of all the particles in the world at that instant (together, as usual, with a specification of what sorts of particles they are). And the positions and velocities of all the particles in the world at some particular instant are patently *not* logically independent of their positions and velocities at other instants; and so a *specification* of those positions and velocities at some particular instant is *not* a description of the world at that instant *alone*—it is not a description of the world at that instant *as opposed to all others,* at all![15]

Talking about *going backward* in the language of dynamical conditions can consequently be a messy business. If (for example) $D_I \ldots D_F$ is an infinite sequence of Newtonian dynamical conditions corresponding to a single free particle moving to the right, then $D_F \ldots D_I$ will correspond not to

15. Maybe it ought to be stressed here that there is nothing wrong or misleading or incoherent about Newtonian dynamical conditions *per se,* and (moreover) that such conditions can perfectly well be *uniquely attached* to particular times. Those sorts of attachments (after all) are precisely the business of differential calculus. What needs to be kept in mind is just that there is all the difference in the world between being uniquely attachable *to* some particular time and being a component of the *instantaneous physical situation of the world* at that time!

There's nothing wrong with propositions like "the velocity of particle 5 at $t = 7$ is 12 miles per hour, in the *x*-direction." What a proposition like that is *about,* though, is not the instantaneous situation of particle 5 at $t = 7$ *itself,* but the *rate of change* of the position of that particle in the *immediate temporal vicinity* of $t = 7$. What a proposition like that is about (to put it a bit more technically) is the limit of the rate of change of the position of particle 5 over an interval *centered* on $t = 7$, as the length of that interval goes to zero.

a particle like that moving to the *left* (which is *what it is,* after all, for a process like that to happen backward) but to nothing whatsoever, to gibberish, to a contradiction.[16] And so if what counts for you as an instantaneous physical situation of the world is (somehow) a *dynamical condition,* then turning something around (at least in certain cases) must involve something *other,* something *more,* than a mere commonsensical *inversion* of the *temporal sequence* of those situations.

The books all tell it like this: for every possible dynamical condition of the world, there is such a thing as that condition "going backward." And here we are starting to get right up to our necks in paradox. What can it possibly mean for a single instantaneous physical situation to be happening *"backward"*? Never mind. Press on. For every such condition D, there is (whatever it means) some unique condition D^* which is D's so-called time-reverse. And what it is for a process $D_I \ldots D_F$ to happen backward is *not* for the inverted sequence of dynamical conditions $D_F \ldots D_I$ to happen (which will as often as not be illogical gibberish), but for the inverted sequence of *inverted* dynamical conditions—$D_F^* \ldots D_I^*$—to happen.

And what the books have to say on the question of the precise mathematical procedure for *obtaining* D_K^* from D_K is (1) that in the case of Newtonian mechanics the procedure is "obviously" to reverse the velocities of all the particles, and to leave everything else untouched; and (2) that the question needs to be approached afresh (but with the Newtonian case always somehow in the back of one's head) in each new theory one comes across; and (3) that what it is *in all generality* for one physical situation to be the time-reverse of another is (not surprisingly!) an obscure and difficult business.

It isn't, really. If you just keep your eye on the ball (which is to say, if you're careful to represent instantaneous physical situations of the world *correctly,* if you're careful to represent them in accord with the requirements of instantaneity and completeness, if you're careful to represent them by means of the sorts of things I decided, a few pages back, to call *states*) then everything is perfectly straightforward. The way to figure out what it is for any sequence of dynamical conditions $D_I \ldots D_F$ to happen backward is to translate that sequence into a sequence of instantaneous states,[17] and then write that latter sequence down in reverse order, and then translate that *inverted* sequence back into the language of dynamical conditions;[18] and whatever you end up with, when all that's done, is (by definition) $D_F^* \ldots D_I^*$. The thing is that if you've fallen under the spell of the books, if the language of states is *unavailable* to you, if you've gotten it into your head that what counts as a complete description of the physical situation of the world at a pure indivisible structureless temporal instant is (per impossible!) a *dynamical condition,* then the above analysis can never even become an object of your attention.[19]

▲▲▲ Insofar as *Newtonian mechanics* is concerned, none of this ends up causing any actual trouble. The velocities of particles, after all, are nothing but the rates of change of their positions. And so if a certain sequence of genuinely instantaneous Newtonian states $S_I \ldots S_F$ corresponds to the sequence $D_I \ldots D_F$ of Newtonian dynamical conditions, and if the prescription for obtaining D_K^* from D_K is just to turn all the velocities around, then the commonsensically backward sequence $S_F \ldots S_I$ will necessarily correspond to the backward sequence $D_F^* \ldots D_I^*$ of the textbooks. And so the

16. What $D_F \ldots D_I$ will correspond to, in this case, will be a particle whose position is constantly being displaced toward the *left,* and whose *velocity* (which is by definition nothing other than the *rate of change* of that position) is constantly pointing *to the right.*

17. In all the candidates for a fundamental physical theory that anybody has taken seriously since the Renaissance, a complete history of the world's *dynamical conditions* is also a complete history of the *world,* and so a complete history of the world's dynamical conditions will correspond to exactly *one* complete history of its *states,* and so the translation we are talking about here will be completely *unique.*

But there can perfectly well (in principle) be theories in which it isn't; there can perfectly well be theories in which a given complete history of the world's dynamical conditions corresponds to *more* than one complete history of its states. In cases like that, *any one* of those latter complete histories will do.

18. That *this* translation is always unique follows from the fact that *states,* by *definition,* are *complete.*

19. Dynamical conditions aren't *necessarily* distinct from states, of course. On the two-state probabilistic theory discussed above, for example, states and dynamical conditions are identical. But on any theory which is deterministic, and which is time-reversal symmetric, and which is (in a sense that will presently be clear) *non-trivial,* they *can't* be.

To see why, think of a deterministic theory T on which the state of the world can evolve (over the course of a second, say) from S_A to S_B, and then (over the course of the *next* second) from S_B to S_C. And suppose that S_A is not the same state as S_C. And suppose that this is a theory on which whatever can happen can also happen backward. Then T must entail that there are at least *two* different states (S_C and S_A) into which the state S_B can lawfully evolve, over the course of the subsequent second. And so T must entail that *not* every state-specification is also a specification of a complete set of dynamical conditions.

textbook idea of what it is to go backward is cooked up in such a way as to amount to precisely the same thing as the commonsensical idea. And so it turns out to be a consequence of the Newtonian laws of motion, on *all* accounts, that any physical process that can happen forward can happen backward too.

But in *other* theories, and as a matter of fact in *all* the fundamental theories that anybody has taken seriously *since* Newton, the plot is a good bit thicker.

Take classical electrodynamics again. What counts as an instantaneous state of the world according to classical electrodynamics is (as I said before) a specification of the positions of all the particles and of the magnitudes and directions of the electric and magnetic fields at every point in space. And it isn't the case that for any sequence of such states $S_I \ldots S_F$ which is in accord with the dynamical laws of classical electrodynamics, $S_F \ldots S_I$ is too. And so classical electrodynamics is *not* invariant under time-reversal.

But the books tell it very differently. What the *books* count as a physical situation of the world at an instant (once again) is not an instantaneous physical state but a dynamical condition. And what counts according to classical electrodynamics as a dynamical condition is a specification of the positions and *velocities* of all the particles in the world, and the magnitudes and directions of the electric and magnetic fields at every point in space. And of course a simple inversion of any sequence of *those* which is in accord with the classical electrodynamical equations of motion gives you illogic. But there turns out to be a way of *transforming* those dynamical conditions (to wit: reverse all the velocities, and reverse all the magnetic fields, and leave everything else as it was) such that if a certain sequence of those conditions is in accord with the classical electrodynamical equations of motion, then the inverted sequence of *transformed* conditions necessarily is too. And it happens (or rather, it will come as no surprise) that the books identify precisely that transformation as the transformation of "time-reversal." And so, according to the books, classical electrodynamics is no less invariant under time-reversal than Newtonian mechanics is.

The thing is that this identification is *wrong*. Magnetic fields are *not* the sorts of things that any proper time-reversal transformation can possibly turn around. Magnetic fields are not—either logically or conceptually—the *rates of change* of anything. If $S_I \ldots S_F$ is a sequence of instantaneous states of a

classical electrodynamical world, and if the sequence of dynamical conditions corresponding to $S_I \ldots S_F$ is $D_I \ldots D_F$, and if we write the sequence of dynamical conditions corresponding to $S_F \ldots S_I$ as $D_F^* \ldots D_I^*$, then the transformation from D_K to D_K^* can involve nothing whatsoever other than reversing the *velocities of the particles*. And if *that's* the case, and if $D_I \ldots D_F$ is in accord with the classical electrodynamical laws of motion, then, in general, $D_F^* \ldots D_I^*$ will *not* be.

▲▲▲ And so (notwithstanding what all the books say) there have been dynamical distinctions between past and future written into the fundamental laws of physics for a century and a half now.

And nonetheless (and on this score the books are right), those laws are all very curiously at odds with the time-directedness of our everyday experience. And that (as I said before) is the tension I mentioned at the outset. And that's what the next couple of hundred pages will be about.