1

A Brief History of Time Reversal

Précis. *The symmetries of time can be understood through the symmetries of motion, both in a sense that is familiar to philosophers and in the history of physics.*

Can time be accurately described in an undirected way, like a great eternal string with no preference for one direction over the other? Or, is it directed like an arrow, with two distinct ends? Philosophers often point out that human experience is vividly directed: we remember the past and not the future; we age towards the future and not the past. But, does time have a direction beyond such facts about human psychology and physiology? This chapter will introduce the main thesis of this book, that the answer is yes: time really is directed like an arrow, in a sense given by what physicists call 'time reversal' asymmetry. In particular, this asymmetry can be detected empirically through our experience of the motion of matter-energy. This asymmetry will be familiar to philosophers, but the evidence for it was developed over the course of two centuries in the history of physics. In this chapter, I will explain both the philosophy and the history behind these claims.

The majority of this book will be cast in the language of physics, which is best-suited to capturing our empirical evidence about the structure of time. However, I would also like to point out a connection between this evidence and the broader philosophy of time. So, Section 1.1 connects my argument to the asymmetries of time that are perhaps most familiar to philosophers, known as the 'A series' and the 'B series' of John McTaggart. The remaining sections then show how the symmetries of time have played a prominent role in two centuries of physics. Section 1.2 points out that the origins of time reversal can be traced to Carnot's theory of engines. Section 1.3 reviews its role in the famous reversibility paradox of statistical physics. Section 1.4 describes how time reversal invariance rose to prominence in the first half of the twentieth century, and Section 1.5 recounts the great shock that physicists felt when they discovered the first evidence of time asymmetry in electroweak interactions.

1.1 On the A Series and the B Series

John McTaggart, an eccentric Cambridge philosopher of Trinity College who was known to salute cats as he met them[1], gave an account of time's arrow that has been influential amongst philosophers: call an undirected description of time a *C series*, and a directed description a *B series* (we will shortly have an A series too). The C series provides language to say whether or not an event falls between two others, or a 'betweenness' relation, while the B series adds the language of an ordering relation. The ordering relation allows one to say something that goes beyond the C series: that an event stands in a before-after relation with respect to others, and (ordinarily[2]) not vice versa. In this language, our question "Is time directed?" becomes "Beginning with a C series description, is there reason to think that time is accurately described by a B series?"

McTaggart believed that it would take a special sort of process to produce a B series from a C series description. Inspired by Hegel's categories, he took this process to involve causality. He also proposed a candidate: the characteristics of being past, present, and future, which he called *A series* descriptions, seem to "pass along" the C series as the future becomes present, and the present becomes past. This 'passage' would determine the kind of ordering required for the B series: say that one event occurs 'after' another event if and only if it happens, during passage – that one is in the future while the other is in the past (or present), but not the reverse. The schema is illustrated in Figure 1.1. Unfortunately, McTaggart himself found it hard to make sense of his A series notion of 'change' from future to present to past, and he ultimately rejected it, as well as the reality of time more generally, as incoherent.[3]

[1] As reported by Dickinson (1931, p.68).
[2] Following Lewis (1979), one might make an exception for closed timelike curves and the cyclic histories of Nietzsche (1974). But, as we will see in Chapter 2, this is no barrier to defining temporal asymmetry.
[3] Mellor (1998, Chapter 7) provides a classic discussion, and Ingthorsson (2016) at book-length.

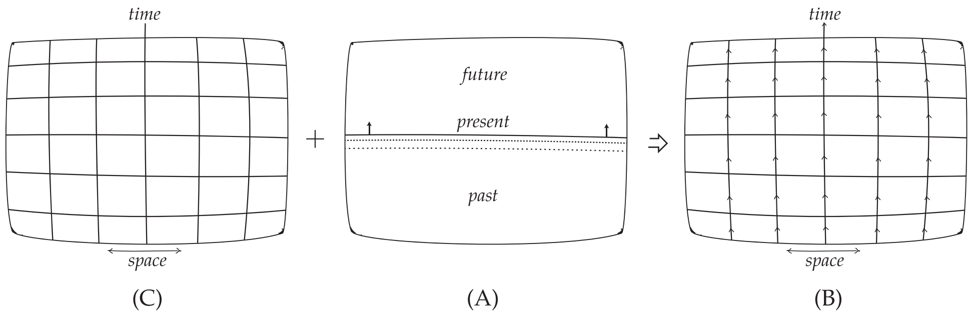

Figure 1.1 McTaggart took his C series plus A series to determine a B series.

McTaggart inspired a voluminous metaphysics of time literature that I'm afraid I won't breach. My aim here is rather to bring that metaphysics a little closer to the physics: McTaggart's A, B, and C series each have a natural expression in physics, so long as we are willing to replace his notions of time and change with more modern ones. For example, Earman (2002a) construes McTaggart's B series as a spacetime with a temporal orientation.[4] The C series is then just a spacetime without a temporal orientation. But, according to McTaggart, the A series is supposed to be linked to the B series and the C series, through what metaphysicians after Broad (1923) now variously interpret as 'passage' or 'becoming'. Many philosophers of physics have despaired of finding an A series in modern physics.[5] Others, such as Maudlin (2002a, 2007), are more optimistic.

It is not my purpose to take a position on this debate here. However, I would like to draw out a different aspect of McTaggart's picture that I think helps to maintain good, clear thinking about the nature of time. Namely, we should begin with a clean, clear separation of the concepts of 'time' and 'change'. Of course, these concepts must be intimately linked, as McTaggart suggests. But, let us not tether a concept as rich as time to just one conceptual framework. Like McTaggart, I would like to 'pull apart' two concepts of time, in order to examine their relationship.

I will pull these concepts apart in a way that is natural in the practice of physics. In physics, we sometimes analyse time using spacetime structure, as when we describe a relativistic spacetime in special or general relativity.

[4] See Section 2.5.3 for a more detailed discussion about temporal orientations.
[5] Callender (2017) and Earman (2002a) both identify the A series 'Becoming' as an aspect of the Manifest Image rather than the Scientific Image – adopting the nomenclature of Sellars (1962) – which led Earman to call for metaphysicians of Becoming to "remain locked in their mutual embrace of Becoming and sink from view into the metaphysical mire" (Earman 2002a, p.2). Maudlin (2002b) responded with a defence of the concept of change in modern physics.

https://doi.org/10.1017/9781009122139.002 Published online by Cambridge University Press

Other times we analyse a concept that is perhaps more appropriately called ‘change’, when we imagine the replacement of one state of the world with another. The latter can be described using a structure commonly called state space, or configuration space, or phase space, as in classical or quantum mechanics. When change is described this way in physics, it is often referred to as a *dynamical system*, whose selection of possible changes is called a *law of motion*. So, let me make the distinction in this way: ‘time itself’ will refer to spacetime structure, while ‘change’ will refer to the changing state in a state space.

The overarching idea that will be carried through every chapter in this book can be put in these terms: that time and change are linked in a way that allows one to learn about the structure of time by studying the structure of change. In particular, in order to learn whether time has an asymmetry or ‘arrow’, one can study the asymmetries of change in the material world.

In Chapter 2, I will show how to make this idea precise, beginning with a concept called *time reversal:* we can understand an ‘undirected’ description of time to mean that the structure of time does not change when it is ‘reversed’. I will then show how this concept can be used to determine whether time itself has an arrow. Disclaimer: my aim with this proposal is not to reanimate McTaggart, nor to argue that he would endorse any such view.[6] If one likes, it may be possible to associate the B series with spacetime structure and the A series with change in dynamical systems. Indeed, if one does so, then there are certain kinds of change that provide evidence for a direction of time: not all change, but just a special kind of change that is called ‘time reversal violating’, and which is discussed in Chapter 7.

The framework threading through this book finds its origins in the pioneering work of Wigner (1939) on the representation theory of relativistic quantum mechanics. It can be distilled down into two postulates:

1. If changing states are interpreted as occurring in spacetime, then those changes must share a common structure with spacetime.
2. Given this, the asymmetries of spacetime can be inferred from asymmetries of those changing states.

The mathematical tool that Wigner used to describe the ‘common structure’ in the first postulate is called a *representation:* roughly speaking, it is a structure-preserving map, from a spacetime structure to a dynamical

[6] As it happens, the great ‘Space and Time’ address of Hermann Minkowski (1908) was given in the same year that McTaggart (1908) published his famous article, but I know of no evidence that either one knew of the other’s work at the time.

system. So, to keep that clearly in mind, I will refer to the first postulate as the ‘Representation View’. This view will be motivated and developed in detail in Chapter 2. A special case will be of particular interest to me: that if states are described as changing *with respect to time*, then that change must share some common structure with time itself.

Wigner used the first postulate to determine the possible dynamical systems of quantum theory, given that they are formulated in the context of Minkowski spacetime. My proposal throughout this book will reverse this thinking and instead use the structure of dynamical change to draw inferences about the structure of spacetime. This leads to the second postulate: by drawing on our observations of change in dynamical systems, I will argue that one can determine whether time has an arrow – and indeed, that there is extremely strong evidence that it does.

The way that this inference works can be illustrated using a toy theory. Suppose the changing state of an animal is described by the metamorphosis of a caterpillar into a butterfly. There is an asymmetry in this theory of change, which is that the reverse metamorphosis cannot occur. In other words, the ‘time reversed’ description is impossible, as illustrated in Figure 1.2. This is an asymmetry in a description of change. However, if time shares the symmetries of this particular change, then it might provide evidence that time itself has an asymmetry too.

This toy theory takes place at a level that omits a great deal of information about change. For example, the interaction of the animal with its environment is completely ignored. Once that hidden information is restored, it is not so clear that the change being described really is asymmetric. I call such erroneous inferences ‘misfiring’ arrows of time and discuss them in detail in Chapter 5. However, a first step in avoiding them is to move from theories of biology to theories of fundamental physics. If we describe motion

Figure 1.2 Time asymmetry: a possible description (left) whose time reverse is not possible (right).

https://doi.org/10.1017/9781009122139.002 Published online by Cambridge University Press

on a fundamental level, by drilling down to the most basic description of change that we can find in the nature of matter and energy, then we might manage to avoid misfiring arrows and identify a true asymmetry in time. In Chapter 7, I will argue that we have evidence for time asymmetry in this sense.

The situation is perhaps similar to a claim of McTaggart (1908, p.464), that, "[i]t is only when the A series, which gives change and direction, is combined with the C series, which gives permanence, that the B series can arise". If the A series is a description of change in a dynamical system, and if that description shares the symmetries of time, then an asymmetry in time itself can arise, which one might interpret as the B series. This helps to dispel a well-known concern about how the laws of motion can be used to make inferences about the direction of time itself, rather than just motion.[7] In this book I will cleanly separate time and change. But, like Wigner, I will argue that the two are linked through a representation. It is this link that allows one to make inferences about the nature of time on the basis of observations about motion.

McTaggart (1908, p.474) himself asks, near the end of his article, whether events in the C series might have some quality that gives them order, writing, "[w]hat is that quality, and is it a greater amount of it which determines things to appear as later, and a lesser amount which determines them to appear as earlier, or is the reverse true?" One way to understand the argument I will make over the course of this book is that time does have a quality somewhat like this. It is not a quality of any one event but rather of the structure of time as a whole: its symmetries are linked to the symmetries of dynamical change in a way that establishes an asymmetry. As to which direction is truly 'later' and which is 'earlier', my account say very little. The arrow of time is as Wittgenstein (1958, §454) described the drawing, '$\rightarrowtail$': "[t]he arrow points only in the application that a living being makes of it". In my view, this makes it no less remarkable that time in our world has an arrow.

The remaining chapters will develop the argument for this view, through an analysis of temporal symmetry under the time reversal transformation. Time reversal is a thoroughly modern concept, and so I will analyse its meaning using the language of modern physics. However, I would also like to convey the charming way that temporal symmetry came to be so important, through an easy-going history of time reversal. That history begins, in the next section, with engines.

7 A version of this concern can be found in Black (1959), with more sophisticated statements found in Earman (1974), Gołosz (2017), and Sklar (1974, §F).

1.2 Ingenuity and Engines

In the summer of 1816, the French physicist Jean-Baptiste Biot convinced the owners of a former church to let him use its boiler to study the polarisation of light passing through turpentine vapour. Not something to be left unattended near an open flame, the experiment detonated in a great explosion that sent the boiler's cover flying and set the roof of the church on fire. Undeterred, Biot advised anyone repeating his experiment to place the boiler behind an impenetrable wall, since

> "the explosion of the vapor, its ignition and that of the liquid, could cause miserable death, and in the most inevitable and cruel manner, to people located at quite a distance."[8]

Explosions aren't always an inconvenience: that flying boiler cover might have more helpfully been used to push an object along a track, like a train. It is really most useful when it can be repeated in a controlled manner to keep the train going, as had been achieved by British inventors like Newcomen and Watt in the eighteenth century.[9] Indeed, soon after the boiler incident, Biot (1817) published a textbook describing a burgeoning class of machines that were powered by vapour explosions. What held these ingenious machines or 'engines' back was a lack of understanding as to what distinguishes a useless explosion from an optimally useful one.

Answering this question made use of a proto-concept of time reversal, introduced by Sadi Carnot in his 1824 *Reflections on the Motive Power of Fire*. Writing while on duty in the French army, Carnot stumbled on a crucial observation, that a useful engine would have to cycle back to its initial state so that the explosive motion could be repeated. This was the ingenuity that ultimately led to modern engines: that all processes that produce motion from heat "can be executed in a reverse sense and in a reverse order" (Carnot 1824, p.19). Carnot's 'reverse sense' and 'reverse order' introduced the concept of time reversal for the first time but applied in a way that is subtly different from its modern usage. Let me review it in a little more detail, using Carnot's most famous example.

Carnot began with the Carnot cycle, which he describes in terms of a stunningly simple example of a gas in a cylinder that expands and contracts, and so can be used to force a piston and drive motion. A model of such a gas is illustrated in the pressure–volume diagram of Figure 1.3. This model is well-known to physicists: the cylinder is initially in contact with a source

8 My translation of Biot (1819, p.133); this curious article provided one of the first studies of optical rotation in turpentine. Happily, no one was injured in the accident.
9 A classic history of this development is Dickinson (1939).

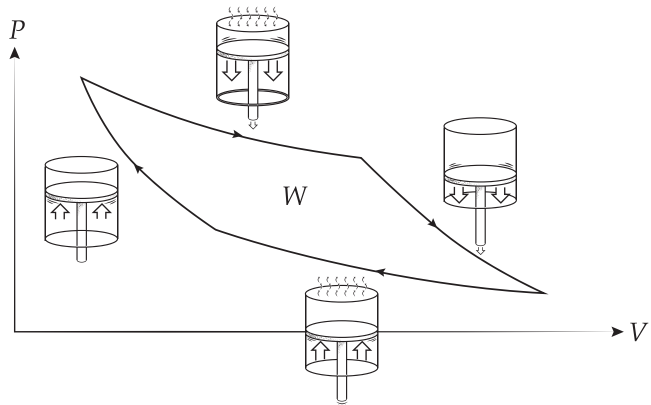

Figure 1.3 Carnot's heat engine: isothermal expansion (top), adiabatic expansion (right), isothermal compression (bottom), and adiabatic expansion (left).

of heat, which allows it to expand while retaining constant temperature (isothermally) along the top path in the diagram. It then continues to expand in isolation from any heat exchange (adiabatically) along the right path, resulting in a drop in temperature. A reverse process then follows: the pressure on the piston is increased to drive the volume back down, maintaining a constant temperature by losing the same amount of heat to a cold source, as along the bottom path. The compression then continues adiabatically until the temperature is elevated back to its initial value along the left path. As a result, the pressure, volume, and heat of the system are all restored to their original values, and the process can be repeated. There has also been a total amount of work done by the engine, $W := \int P dV$, which is equal to the area of the shape traced out by the curves.

From an engineering perspective, the Carnot cycle aims to do two things: to do as much work as possible and to return back to where it started so that the process can repeat. These are both achieved with the help of what Carnot took to be his central conceptual insight, the pairing of two processes with two 'inverse' processes:

> The operations that we have just described could have been done in the inverse order and sense. . . . In our first operations, there was at the same time a production of motive power and a transport of caloric [heat] from body A to body B; in the inverse operations, there was at the same time an expense of motive power and a return of caloric from body B to body A. (Carnot 1824, pp.10–11)[10]

[10] My translation. Carnot's successful use of 'caloric' here, a chemical element postulated to characterise heat before the kinetic theory that was assumed to be conserved, has been the subject of much debate in the philosophy of science (cf. Chang 2003; Laudan 1981; Myrvold 2020a; Psillos

https://doi.org/10.1017/9781009122139.002 Published online by Cambridge University Press

Table 1.1. *An expansion in the Carnot cycle is the time reverse of some compression, though these specific pairings are not the time reverse of each other.*

Process		Inverse Process
Isothermal expansion	$\leftrightarrow$	Isothermal compression
Adiabatic expansion	$\leftrightarrow$	Adiabatic compression

These pairings, shown in Table 1.1, relate each process to some 'time reversed' process, in that each compression corresponds to an expansion described in the reverse time direction. In particular, an isothermal compression with heat flowing in is the time reverse of some isothermal expansion with heat flowing out; and, an adiabatic expansion is the time reverse of some adiabatic compression.

This is a subtle variation on typical modern usage of time reversal: strictly speaking, Carnot's pairings are not the time reverse of each other, since they take place at entirely different pressures and volumes. In fact, if one were to carry out the 'strict' time reversal of the first two parts of the cycle (the top and right paths in Figure 1.3), one would just trace back along the same lines to the original state. This produces a cycle with zero area, and which thus does zero work. How then does Carnot choose the right compression process to follow the expansion?

It is the natural choice of an engineer: choose the inverse processes that are 'optimal', in the sense of maximising the amount of work done by the engine. After following the top and right paths in the diagram, there are various ways of zig-zagging back to restore the original amounts of pressure, volume, and heat. But, since the work done is given by the area inscribed by the paths, these will always be less than or equal to the work done in Carnot's cycle. Assuming that the first two paths in the cycle achieve the engine's maximum and minimum temperatures, the unique work-maximising cycle is the Carnot cycle. That is how Carnot selects the 'inverse' operations: he does not pair expansions with their strict time reverses but rather chooses those 'inverse operations' that produce the best possible engine.[11]

1999). The subsequent development of equilibrium thermodynamics discussed in Chapter 6 is often viewed as a response to the discovery that no such chemical element exists!

[11] A reading of Carnot along these lines is set out in much more careful detail by Uffink (2001, §4), who duly cautions that Carnot himself does not make any explicit connection between 'inverse operations' and 'time inversion'.

This is the proto-version of time reversal that appeared in Carnot's theory of heat and work, on the road to identifying the behaviour of an optimal engine. Unfortunately, all of this discussion took place with a rather rough idea of what 'time reversal' actually means. That was a side-effect of the limited language of thermodynamics that was available at the time of Carnot. Fortunately, more precise thinking about time reversal would become available in the next episode in our story, the development of statistical mechanics.

1.3 Well, You Just Try to Reverse Them!

The appearance of time asymmetry is commonly associated with the phenomena of classical thermodynamics, like an exploding boiler or a realistic mechanical engine that dissipates heat. We tend to experience these processes as unfolding in one way but not the other: the boiler explodes but does not 'un-explode'; the engine dissipates the heat it generates but does not spontaneously heat up. That sort of time asymmetry is often said to be a consequence of the second law of thermodynamics, that in at least some contexts, entropy does not decrease. In Chapter 6, I will argue that the situation is more subtle. But, for this story, the more important difficulty is that classical thermodynamics makes no mention of a system's underlying constituents. With growing interest in the nature of the material that makes up a gas or an engine, the natural next step was to use a theory of fundamental matter to try to explain thermodynamic behaviour.

One prominent perspective on fundamental matter in the nineteenth century was the atomist one, commonly attributed to Democritus, Boyle, and Bošković. On this view, all physical phenomena can be reduced to "the particular sizes, shapes, and situations of the extremely little bodies that cause them" (Boyle 1772, p.680). The possible motions of these phenomena would then be described by the laws of a dynamical theory, in the sense of Section 1.1.

What does it mean to 'time reverse' these structureless little bodies? We could imagine a film of the particles played back in reverse. One would at least expect to see their positions occur in the reverse time-order and with velocities in the opposite directions. This provides a rough, preliminary way to think about the time reversal transformation, which will be clarified in Chapters 2–3. For now, following the discussion of Section 1.1, we can take time reversal symmetry in a dynamical system to mean that there is a possible trajectory of particle positions and velocities such that, if we consider the trajectory in the reverse time-order and with reversed velocities, then the resulting curve is a possible trajectory as well. Time *asymmetry* would then be a dynamical system that is not time symmetric. In a time asymmetric system, the time reverse of at least one trajectory describes motion that is impossible according to the laws.

This concept had a dramatic effect on the work of Ludwig Boltzmann, who proposed to reduce all of thermodynamics to the statistics of huge numbers of little particles (Boltzmann 1872). Boltzmann's explanation of their apparent time asymmetry was given in his famous 'H-theorem', where he seems to conclude that generic classical mechanical systems are likely to approach a 'stationary state', meaning one that does not change over time and which attains the maximum possible entropy.[12] The reverse 'entropy-decreasing' process, according to Boltzmann, would be highly unlikely.

A simple way to understand Boltzmann's thinking is in terms of a *counting argument:* roughly, that high-entropy states can happen in such an enormous variety of ways that they occupy the 'greatest volume' of possibilities. It is like imagining a house with a thousand blue rooms and one room that is red.[13] Think of the blue rooms as analogous to high-entropy states. Now, suppose that you were to leave a red room and enter an arbitrary new room (with a uniform probability of entering a given room); it is overwhelmingly likely that the new room will be blue. Moreover, it is likely that if you continue to repeat this process, your room colour will (with high probability) be unchanging or 'stationary' over time.[14]

This sort of counting argument has the potential to explain many asymmetries: an exploded boiler and a dissipated gas are both descriptions that belong to the overwhelming majority of possible states. So, it is natural that we should expect a system to end up in such a state. Unfortunately, this is not enough to explain the time-asymmetric behaviour of such systems. Returning to the house analogy, suppose we find a person in a red room and ask what colour room they are most likely to have come from? The very same volume argument concludes: a blue one. That is, the counting argument by itself provides equally good evidence for a high entropy state to the future and to the past.

12 There is some debate about whether this is actually what Boltzmann meant to conclude by this theorem and about whether the resulting state is really 'stationary'; see Klein (1973, p.73) vs. Von Plato (1994, p.81), and Uffink (2007, §4) for a convincing clarification.

13 In fact, given around 10^{26} molecules of air in your kitchen, the analogy there would require a house with around $10^{10^{26}}$ rooms, with all but one coloured blue.

14 Here, the supposition that you enter an 'arbitrary' room with uniform probability is contentious: there is little agreement among philosophers of physics on the extent to which it works (cf. Uffink 2001).

Figure 1.4 Reversing the velocities of a thermodynamic process (left) produces an antithermodynamic process (right).

One might hope that the apparent time asymmetry of thermodynamics could be explained by some other aspect of the laws of classical particle mechanics. This hope was famously dashed by the Austrian chemist Josef Loschmidt,[15] using the concept of time reversal:

Indeed, if in the above case, after a time τ which is long enough to obtain the stationary state, one suddenly assumes that the velocities of all atoms are reversed, we would obtain an initial state that would appear to have the same character as the stationary state. For a fairly long time this would be appropriate, but gradually the stationary state would deteriorate, and after passage of the time τ we would inevitably return to our initial state. (Loschmidt 1876, p.139)

Loschmidt's observation was that all known possible motions in classical particle mechanics correspond to a time-reversed motion that is also possible, with particle positions occurring in the reverse time-order and with their velocities reversed. Boltzmann had described a physical system in terms of a time-symmetric system of classical particles. So, if such a system evolves classically from a low-entropy state to a high-entropy one, then there is always a counterpart system that evolves 'anti-thermodynamically', from a stationary state to a non-stationary one, as in Figure 1.4. This dramatically contradicts Boltzmann's conclusion that a system of molecules generically evolves towards a stationary state.

In discussion with Loschmidt about this reversibility paradox, Boltzmann is rumoured to have replied, "Well, you just try to reverse them!" (Brush

[15] This 'reversibility paradox' was anticipated by Maxwell, Tait, and Thomson between 1867 and 1870, and revived by Ehrenfest and Ehrenfest-Afanassjewa (1907); see Brush (1976a, pp.82–3) and Brush (1976b, pp.602–5).

1976a, p.605). However, he took the problem very seriously and developed a collection of creative responses to Loschmidt's reversibility paradox.[16] One popular textbook response today is to restrict Boltzmann's argument to those descriptions that begin in a special low-entropy initial state. It is then sometimes claimed that, as an economical way of explaining all time asymmetry at once, one can postulate that the universe as a whole began with very low entropy. The meaning and status of this postulate, dubbed the 'Past Hypothesis' by Albert (2000), is a matter of ongoing debate.[17] Price (1996, 2004) has argued that such arguments do not provide evidence for time asymemtry. I agree, and will give this argument from the perspective of my account in Chapter 5.

However, all is not lost for the arrow of time: developments after Boltzmann led to an entirely new kind of time asymmetry. To appreciate how this happened, it helps to first review how time reversal came to play a more centre-stage role in physics.

1.4 The Rise of Time Reversal

Physicists of the nineteenth century, like many of us, were fascinated by cats. Witness the French mathematician Jules Richard:

> The problem of the cat who falls back on its paws has preoccupied scholars for several years. Here is what the problem amounts to. A cat launched into the air always falls back on its paws; how can this be done? It seems that the cat launched in the air and with no point on which to press could not modify its motion in any way.[18]

The problem was widely studied throughout the nineteenth and twentieth centuries, including by James Clerk Maxwell, who in a letter to his wife described a widespread rumour about his time spent in Cambridge: "There is a tradition in Trinity that when I was here I disovered a method of throwing a cat so as not to light on its feet, and that I used to throw cats out of windows".[19]

The origin of this problem, according to Paul Painlevé (1904, p.1171–2), is an assumption of time reversal symmetry: when a system begins at rest, the equations of motion "do not change, and neither do the initial conditions,

16 See Uffink (2001) for an overview.
17 Cf. Earman (2006); Wallace (2010, 2017).
18 My translation of Richard (1903, p.183).
19 Reported in Maxwell's biography by Campbell and Garnett (1882, p.499). The present author recommends that you be nice to your cat.

https://doi.org/10.1017/9781009122139.002 Published online by Cambridge University Press

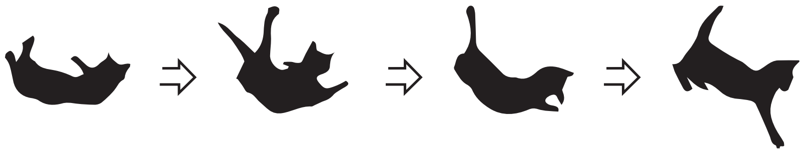

Figure 1.5 Non-rigid motion avoids the Bad News for Cats theorem, which was derived by Painlevé (1904) on the assumption of time reversal symmetry.

when one changes $(t - t_0)$ into $-(t - t_0)$". From this, Painlevé argued that a rigid cat cannot rotate as it falls. Earman (2002b) calls this the "Bad News for Cats Theorem". As far as I can tell, some further assumptions are needed to make Painlevé's argument work, but his insight that time reversal can give rise to an informative physical argument still broke new ground.[20] Fortunately for cats, there is a ready solution to the problem: the cat's motion need not be rigid (Figure 1.5). Simple models have even been proposed that clarify the non-rigid motion of a falling cat, and the problem remains of interest to modern mathematicians.[21]

Arguments underwritten by time reversal invariance became increasingly common after Wigner (1931) introduced the first precise definition in his celebrated book, *Group Theory and its Application to the Quantum Mechanics of the Atomic Spectra*. This book simultaneously introduced a central role for group theory in modern physics, as well as a central role for time reversal. I will examine Wigner's definition in precise detail in Chapter 3; but here is the passage in which time reversal made its debut, translated as 'time inversion' in Wigner's writing. After considering a system in which all the translation and velocity boost symmetries have been suppressed, Wigner writes:

The transformation $t \to -t$ remains an additional symmetry element. It transforms a state φ into the state $\theta\varphi$ in which all velocities (including the 'spinning' of the electrons) have opposite directions to those in φ. (Hence, 'reversal of the direction

[20] A Painlevé-like statement that *can* be proved is that the falling cat's motion is periodic. Let S be a set of states, and let $t \mapsto \varphi_t$ be a one-parameter group of bijections on S satisfying $\varphi_{t+t'} = \varphi_t\varphi_{t'}$ for all $t, t' \in \mathbb{R}$, representing the dynamics. Suppose τ is a representation of time reversal symmetry (see Chapter 4), so $\tau \circ \varphi_t = \varphi_{-t} \circ \tau$ for all $t \in \mathbb{R}$, which satisfies $\tau \circ R = R \circ \tau$ for some non-trivial bijection R, e.g., R could be a rotation. *Proposition:* a state $s \in S$ can only satisfy $\tau(s) = s$ (initially at rest) and $\varphi_t(s) = R(s)$ for some $t \neq 0$ (finally changed) if it is periodic, in that $\varphi_{t'}(s) = s$ for some $t' \neq t$. *Proof:* Being 'finally changed' is equivalent to $\varphi_{2t}s = \varphi_t \circ R(s)$. Our assumptions thus imply that $\varphi_{2t}(s) = \varphi_t \circ R(s) = \varphi_t \circ R \circ \tau(s) = \tau \circ \varphi_{-t} \circ R(s) = \tau(s) = s$. The conclusion then follows with $t' = 2t$.

[21] The first mechanical solution to the falling cat problem was given by Kane and Scher (1969); for later mathematical developments, see Montgomery (1993) and the references therein.

of motion' is perhaps a more felicitous, though longer, expression than 'time inversion'.) The relation between time inversion and the change which the lapse of time induces in a system is of great importance. (Wigner 1931, p.325)

Wigner's comment about the importance of time reversal symmetry was entirely correct.[22] In addition to Bad News for Falling Cats, time reversal symmetry was used in the development of low-temperature physics, through Wigner's analysis of energetic degeneracy (Wigner 1932). It was used to show why non-trivial superpositions of bosons and fermions are never observed in nature (Wick, Wightman, and Wigner 1952). Time reversal symmetry was even shown by Wald (1980) to be incompatible with the possibility of pure-to-mixed state transitions in quantum theories of gravity. Today, the concept of time reversal is a cornerstone of modern physics.

The idea that time is symmetric might even have been absorbed as a central axiom of modern physics, from which a great number of conclusions could be derived, had nature not conspired to have it otherwise. At the time of Wigner's writing, all known laws of elementary motion appeared to exhibit manifest time reversal symmetry. Shockingly, this pattern eventually failed: in 1964, time reversal symmetry was dramatically ejected from the axioms of physics, and a new origin for an arrow of time became available.

1.5 The Great Shock

In the first half of the twentieth century, the importance of time reversal in fundamental physics had come to be appreciated. However, few seemed to consider the possibility that it might fail to be a symmetry. The earliest exception that I know of is a comment by one of the founders of quantum mechanics, Paul Dirac. With characteristic foresight, Dirac wrote:

A transformation . . . may involve a reflection of the coordinate system in the three spacial dimensions and it may involve a time reflection, the direction . . . in space-time changing from the future to the past. I do not believe there is any need for physical laws to be invariant under these reflections, although all the exact laws of nature so far known do have this invariant. (Dirac 1949, p.393)

Dirac's comment reveals a remarkable puzzle for anyone interested in the status of time asymmetry: we are often in a situation of not knowing the correct laws of motion for a physical system. But, the definition of a symmetry of motion involves a statement of how the system can change over

[22] I have suggested in Section 1.1 that an asymmetry of motion allows us to infer an asymmetry of time. So, for now let me set aside Wigner's comment about motion; I will return to it in Chapter 2.

time. So, how can we possibly determine whether or not any transformation like time reversal is a symmetry without knowing the laws of motion?

Progress towards a solution came in the form of a 'symmetry principle', originally formulated by Pierre Curie, which has been the subject of a great deal of study in the philosophy of physics (and which I will review in Chapter 6):

> When certain effects reveal a certain asymmetry, this asymmetry must be found in the causes that have given rise to it. (Curie 1894, p.401)

One interpretation of this principle takes a 'cause' to be a law of motion together with an initial state, an 'effect' to be a final state.[23] So, if a final state has an asymmetry that is not found in the initial state, then that asymmetry must somehow be in the dynamical law itself. This provides a way to evaluate certain symmetries of motion: if two successive states of a system do not share a unitary symmetry, *then that symmetry must be violated by the equations of motion*.

Curie's principle was first applied in particle physics to show that the parity transformation (or what Dirac called 'spatial reflection') is not a symmetry of the equations of motion, although the principle was not referred to by that name. This application arose out of a problem in particle physics known as the θ–τ puzzle. Two in-going particles, denoted θ and τ, were known to have the same masses and lifetimes but to decay into different outgoing states: the former into two pions (positive and neutral), and the latter into three pions (two positive and one negative). Since one of these decay products was invariant under the parity transformation and the other was not, parity invariance implies that they must have originated from different particles θ and τ, by an application of Curie's principle. So, given that θ and τ were not the same particle, the puzzle was to explain why nature appeared to conspire to give them the same masses, and indeed the same lifetimes in a decay interaction. Lee and Yang put the puzzle in a controversial way:

> One might even say that the present θ-τ puzzle may be taken as indication that parity conservation is violated in weak interactions. This argument is, however, not to be taken seriously because of the paucity of our present knowledge concerning the nature of the strange particles. (Lee and Yang 1956, p.254)

23 Cf. Belot (2003), Earman (2004), and Roberts (2013a).

The experimentalist Norman Ramsey was intrigued and asked Richard Feynman if he thought that it was worth designing an experiment to test for parity violation. Feynman said yes, but that he was almost sure that the result would be that the world is parity invariant. When Ramsey asked if Feynman was confident enough to bet 100 dollars to one, Feynman is said to have replied “No, but 50 dollars I will!” (Gardner 1991, p.91). Feynman soon lost his money when the experimentalists Chien-Shiung Wu, Ambler, et al. (1957) proved that the decay of the Cobalt-60 atom violates parity. In addition, her discovery immediately solved the θ–τ puzzle: the two particles were in fact one and the same – now known as a positive K^+ meson or kaon, which is subject to parity-violating interactions.

With parity symmetry violated, the physics community came to widely believe that another symmetry must hold instead. The transformation that exchanges ordinary matter with an exotic substance called ‘antimatter’, called ‘charge conjugation’ (denoted ‘C’), turns out to have been violated by the Wu experiment as well.[24] However, the combined transformation consisting of C together with parity P produces a transformation denoted CP, which was thought to remain a symmetry. And, by a very general theorem known as the CPT theorem, discussed more in Chapter 8, CP symmetry is equivalent to time reversal symmetry. In this way, the physics community hung on to the assumption of time reversal symmetry, in spite of the recent fate of parity.

Why did they assume this? James Cronin gave a colourful explanation some years later:

> It just seemed evident that CP symmetry should hold. People are very thick-skulled. We all are. Even though parity had been overthrown a few years before, one was quite confident about CP symmetry. (Cronin and Greenwood 1982, p.41)

In fact, it was not just the absurdity of the human condition that led physicists to replace parity symmetry with CP symmetry. After Wu’s experiment, a number of simple and powerful theoretical models were quickly developed to explain her result, including one developed by the young Stephen Weinberg (1958). These elegant models did a compelling job of describing the behaviour of these new interactions. As it happened, they also strongly suggested both CP invariance and time reversal invariance. So, *pace* Cronin,

[24] This was pointed out by Garwin, Lederman, and Weinrich (1957), who independently verified Wu’s experiment.

https://doi.org/10.1017/9781009122139.002 Published online by Cambridge University Press

this was not so much a matter of thick-skulled intuition but of the hard struggle to find any alternative.[25]

It is difficult to understate the great shock to the community when James Cronin and Val Fitch discovered that CP symmetry fails and therefore that time reversal symmetry fails too. The feeling was summarised here:

> It came as a great shock that microscopic T invariance is violated in nature, that 'nature makes a difference between past and future' even on the most fundamental level. We might feel that such a statement is sensationalist rather than scientific; yet there is indeed something very special about a violation of the invariance under T or CP. (Bigi and Sanda 2009, p.5)

This story is worth describing in some detail, although we will revisit it in Chapter 7.

Cronin identifies a deep paper by Gell-Mann and Pais (1955) as a central influence on his thinking, writing that, "you get shivers up and down your spine, especially when you find you understand it" (Cronin and Greenwood 1982, p.40). Gell-Mann and Pais had proposed a means of detecting the violation of C symmetry, but it was a means of detecting CP symmetry violation as well. Cronin and Fitch had an accelerator at the Brookhaven National Laboratory on Long Island, New York, which could set an accurate bound on both kinds of symmetry violation. So, to test CP symmetry, they designed an experiment firing a long-lived neutral kaon state K_L into a spark chamber and taking photographs of thousands of particle decay events to see what came out.

At the time, neutral kaons were characterised by their decay into three pions, one neutral and two of opposite charges. This decay is compatible with CP invariance, by another application of Curie's principle: the neutral kaon K_L and the three-pion state are reversed by CP. In contrast, a two-pion state is left unchanged by CP, and so a decay into just two pions would imply CP violation (Figure 1.6). Cronin and Fitch set out to check whether they could show that, to a high degree of accuracy, no CP violating two-pion decay events could be found.

After a long analysis of all the photographs, they found that to the contrary, a small but unmistakable number of long-lived neutral kaons decayed into two pions, violating CP and time reversal symmetry. They immediately began checking their result and discussing it with colleagues at Brookhaven. After explaining it to their colleague Abraham Pais over coffee,

[25] Weinberg describes one phase relation in his model: "This phase relation would follow from CP invariance, but is difficult to understand on any other basis" (Weinberg 1958, p.783).

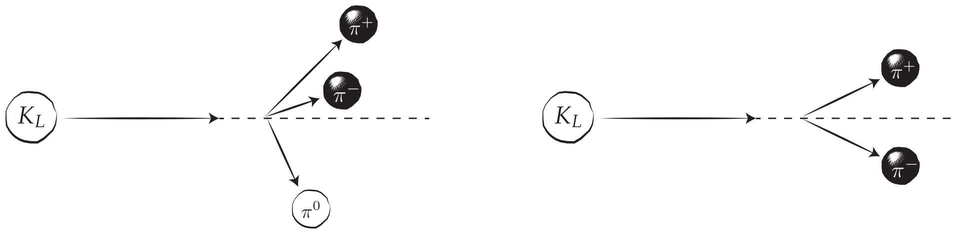

Figure 1.6 Neutral kaon decay into three pions (left) and two pions (right). The neutral pion is invisible to the spark chamber, but its trajectory is calculated by conservation of momentum. The two-pion decay implies CP-violation.

Pais reported that, "[a]fter they left I had another coffee. I was shaken by the news" (Pais 1990).

Cronin and Fitch were awarded the 1980 Nobel Prize for their discovery.[26] In the award ceremony speech, Gösta Ekspong described how the physics community viewed the implications for the direction of time in poetic terms:

> The laws of physics resemble a canon by Bach. They are symmetric in space and time. They do not distinguish between left and right, nor between forward and backward movements. For a long time everyone thought it had to be like that.... [Cronin and Fitch's] discovery ... implied consequences for time reflection. At least one theme is played more slowly backwards than forwards by Nature. (Ekspong 1980)

Theoreticians caught up with the detection of CP and time asymmetry over the next decade. In order to describe these interactions in terms of the developing theory of non-abelian gauge fields due to Yang and Mills (1954), the previous three-quark gauge theory had to be adjusted. This led to the modern six-quark theory of flavour mixing and paved the way for the unified quantum theory of strong and electroweak interactions known as the Standard Model.

As I will argue in Chapter 7, there are subtleties in Curie's principle that prevent its use in the detection of time asymmetry without appeal to CPT symmetry. However, new symmetry principles were soon developed that have been less discussed by philosophers but which enabled the more direct detection of time asymmetry. These principles were successfully applied to produce the first evidence of time reversal symmetry violation without appeal to CPT symmetry, by the CPLEAR Collaboration (1998) at CERN.

[26] Cronin describes this history in a delightful University of Chicago lecture transcribed by Margaret Greenwood (Cronin and Greenwood 1982).

This was followed by a number of creative tests of CP symmetry and time reversal symmetry violation in other sectors, including the B-meson and, quite recently, in the lepton sector through muon–electron neutrino oscillation.[27] This provided evidence that time reversal symmetry violation is here to stay, and in increasing quantities.

1.6 Summary

The dramatic discovery of time reversal symmetry violation has not yet been widely embraced by philosophers, at least as evidence for an arrow of time. Some have found it helpful to set aside the philosophical analysis of time asymmetry in particle physics, focusing their analysis on other important issues,[28] while others have expressed flat-out scepticism about its existence (Horwich 1989, p.56). Maudlin (2007, p.118) has suggested that this latter response has "a certain air of desperation" to it, insofar as Nobel prizes have already been awarded. But, I think there is also a conceptual issue to be overcome: these discoveries may establish an asymmetry in the way particle states change, but what reason is there to think that this establishes an arrow of time itself?

Answering this question requires some philosophical analysis. McTaggart's separation of time into 'A series' and 'B series' components is a start. However, in this book, I will make the separation using the language of asymmetries in dynamical systems on the one hand, and asymmetries in spacetime structure on the other. My central postulate will be that these two kinds of asymmetries are intimately linked. However, this link must be spelled out and motivated in more detail, if we are to have a convincing account of what time reversal in particle physics has to do with the reflection of time itself. The project of the next chapter is to develop one such account.

27 CP-violation by B_0-mesons was detected independently by the BaBar Collaboration (2001) and the Belle Collaboration (2001); time reversal symmetry violation was later detected by the BaBar Collaboration (2012). New evidence for CP violation to a much larger extent was recently discovered in the lepton sector by the T2K Collaboration (2020).

28 Cf. Callender (2000, p.249) and Price (1996, p.18).