CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-12332-7 — Reversing the Arrow of Time
Bryan W. Roberts
Excerpt
More Information

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.

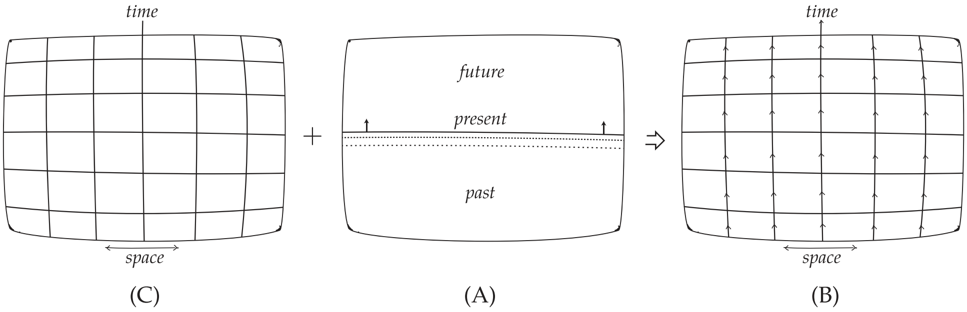

Figure 1.1 McTaggart took his C series plus A series to determine a B series.

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.

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

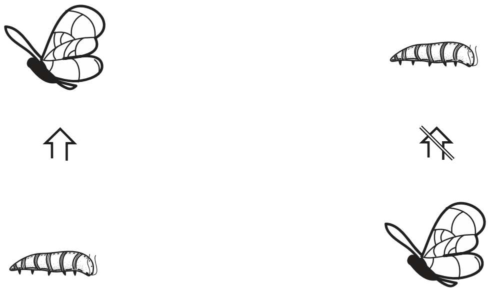

Figure 1.2 Time asymmetry: a possible description (left) whose time reverse is not possible (right).

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, '↣': "[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).

Cambridge University Press & Assessment
978-1-009-12332-7 — Reversing the Arrow of Time
Bryan W. Roberts
Excerpt
More Information

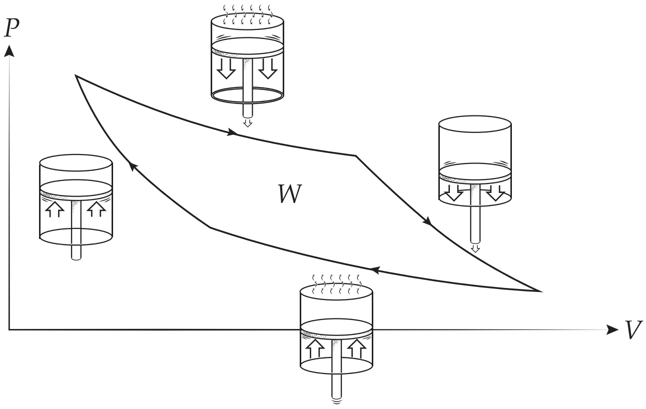

Figure 1.3 Carnot's heat engine: isothermal expansion (top), adiabatic expansion (right), isothermal compression (bottom), and adiabatic expansion (left).

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, $W := \int PdV$, 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

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	↔	Isothermal compression
Adiabatic expansion	↔	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'.

Cambridge University Press & Assessment
978-1-009-12332-7 — Reversing the Arrow of Time
Bryan W. Roberts
Excerpt
More Information

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