The Arrow of Time

Kieran Hobden

Abstract

Throughout our lives we gain an intuitive understanding of the direction of time. Fires burn out, never to be re-lit; memories are lost, never to be remembered; and youth fades away, with no chance of return. But how can this apparent direction be formalised and understood? Through a look at time reversibility in thermodynamics, and the Newtonian and quantum pictures of the world, this essay tries to answer our question whilst also posing new questions. An understanding of entropy and the laws of thermodynamics, and basic quantum theory is a pre-requisite.

Contents

1 Introduction

This essay aims to provide the reader with a deeper understanding of the arrow of time converting what is often an intuitive understanding into one that is formalised and able to be developed further. Of course, the essay is limited as there is no agreed upon interpretation of quantum mechanics, which governs our world, and it is still an active topic of research so we can provide no definitive conclusion. However, the essay shall attempt to come to the defence of time reversibility in the Newtonian picture of the world and of determinism with a distinct arrow of time in the thermodynamic picture.

Section 2 aims to justify time reversibility in the Newtonian picture of the world whilst Section 3 defends the thermodynamic arrow of time by eliminating the possibility of an isolated system with the effect of being overwhelmingly likely to reduce the entropy of the system. Section 4 discusses different interpretations of quantum mechanics and their involvement in the determinism of the universe.

This essay is largely based off my research of David Z. Albert's book: Time and Chance[1] and as such, only this citation of his book shall appear in the essay to avoid littering the text with citations. His work was crucial in the development of the essay and any errors shall be due to my misinterpretation of his work.

2 Newtonian Mechanics and Time Reversibility

The Newtonian picture of the world, although known to not be an entirely correct view of the world, is a good place to start when considering the symmetries of time due to its simplicity. All processes in this picture are seen to be time-reversal symmetric: a ball, for example, thrown upwards such that its trajectory is a parabolic arc could be filmed and the video, run in reverse, would appear to be as likely to have occurred as the original footage. This is a common starting point for discussing time-reversal symmetries. We start by saying that a process is time-reversal symmetric if the process run in reverse (under the transformation $T : t \longrightarrow -t$) is a legitimate process according to the laws of the governing theory. In this way, all processes in the Newtonian picture of the world seem to be time-reversal symmetric. A video of a billiard ball on a frictionless table which knocks into another, setting both balls on new trajectories could be reversed and the situation would look as believable as if it had occurred running forwards in time. This is due to the microscopic reversibility of Newtonian mechanics. Unfortunately, by discussing the motion of billiard balls and other macroscopic processes, we have taken a foray into the one of the simplest violators of time reversibility: friction. Imagine a book sliding along a surface. The book will be seen to have some initial velocity which is reduced parabolically until the object is stationary. If this process were filmed and viewed backwards, it would appear non-sensical that an object might go from being stationary to having its velocity increase parabolically. This clearly violates Newton's First Law and therefore shows no time-reversal symmetry.

Before we get ahead of ourselves, we must take a more exact description of the necessary conditions for a process to be time-reversal symmetric. If we suppose that there is some fundamental theory of the world, T, any process can then be described as a sequence of instantaneous events, $S_i, \ldots, S_f$. Then the requirement for the process to happen backwards is simply that the sequence $S_f, \ldots, S_i$ must also occur in T.

But how can we know the sequence of instantaneous conditions? Suppose we had a system which was sufficiently small such that we could gather and store all the information we needed to fully define the system (remembering that absolute positions and velocities are possible as we are still in a Newtonian picture and quantum effects are not present). The causal nature of Newtonian mechanics would allow us to determine the later instantaneous conditions in the sequence $S_i, \ldots, S_f$ thus, with the initial conditions, we would be able to determine the full sequence.

However, a complete, instantaneous description of the world is not sufficient in the Newtonian picture as the positions and velocities of all particles in the universe (along with non-dynamical properties such as mass and charge) are necessary. The positions and velocities of all the particles in the world at any one instant are not independent of their positions and velocities at other instants therefore this description of the world is not that of an instantaneous state. Consider now a sequence of dynamical states $D_i, \ldots, D_f$ which fully describe the Newtonian picture of the world. If this sequence describes a particle moving to the left, then its reverse should logically describe the particle moving to the right if the theory is time-reversal symmetric. However, the sequence $D_f, \ldots, D_i$ (the reverse of the original) would describe the particle moving to the left with its velocity vector pointing to the right which is non-sensical. This suggests that something more complex is required to reverse time than simply the events run in reverse. In this example we must invert the dynamical conditions such that the velocity vector points in the opposite direction. This idea is specific but can be generalised. A sequence of dynamical conditions $D_i, \ldots, D_f$ must be translated to a sequence of instantaneous states such that they can be reversed. If the Newtonian picture simply requires that the velocity direction is reversed then the commonsensical sequence of instantaneous states $S_i, \ldots, S_f$ that we originally used, in reverse, amount to the same as the set of transformed dynamical conditions in reverse such that the more simple sequence can be used (although slightly incorrectly) to describe the Newtonian picture of the world.

So, can this time-reversal symmetry ever be violated in the Newtonian picture? Suppose we have a video recording of a ball that is rolled up the side of a dome with just enough energy such that the ball reaches the pinnacle of the dome but rather than overshooting or rolling back down, the ball remains stationary. Now, if we watch the film in reverse, the ball appears to go from being stationary to accelerating for an extended period and seemingly violating Newton's First Law of Motion. Naturally, due to quantum fluctuations the ball will inevitably roll back down the dome eventually but in the Newtonian picture of the world this is not relevant, and we expect the ball to remain in position. This highlights our inability to use physical intuition in this scenario. John Norton[2] proposed that this apparent contradiction to time-reversal symmetry is not an issue for several reasons. Firstly, that a ball placed on top of a dome cannot show a preference as to whether it remains or rolls down the side of the dome. It is not just a lack of preference but furthermore, the system produces no probabilities to distinguish which outcome is favourable.

It is often suggested that this is because it is impossible for the ball to roll back down due to Newton's First Law of Motion which brings us on to Norton's next argument. He argued that the law: "In the absence of a net external force, a body remains at rest or in a state of uniform motion in a straight line." Should be altered to "In the absence of a net external force, a body is unaccelerated." Norton then provides the equation for the acceleration of the ball and shows that at time $t = T$ (when the ball appears to first move), the acceleration is zero which satisfies Newton's First Law. My personal objection arises due to Norton's re-statement of the law. I have previously stated that a complete description of the Newtonian picture of the world requires the dynamical conditions: position and velocity. However, Norton requires the law to be instantaneous before using it to describe dynamical conditions. Another objection might arise from the fact that the ball never actually reaches the pinnacle of the dome. It simply takes a progressively longer time to make progressively smaller movements up the dome such that it never actually reaches an unstable equilibrium where it is stationary. This means that the acceleration on the ball is never zero and therefore the time-reversal of the motion would appear perfectly legitimate in the Newtonian picture. However, if we suppose that the ball could be placed exactly at the point of unstable equilibrium then might it spontaneously slide down the dome? Unfortunately, we can gain no further insight into this approach as the time-reverse is simply the ball being picked just as it was placed beforehand. It should be noted that this situation is in no way special and the analysis can be applied to other systems whereby an object reaches a point of unstable equilibrium with no local stable equilibrium.

If we assume however that Norton was correct and that the ball, in an unstable equilibrium, would roll back down the slope (note this is not verifiable from the real world as it is non-Newtonian), then we enter the possibility that the dynamical world might require an instantaneous description under observation similar to that of quantum mechanics. Perhaps quantum mechanics provides the mapping for the dynamical conditions of the world to an instantaneous set such that time-reversal is possible?

3 Entropy and the Arrow of Time

It is well known that time has a specific direction; the past is unchangeable and we are left with only memories, whilst the future is not yet determined and we can have no understanding of the future until it arises. In this section, we discuss this intuitive arrow of time through the topic of thermodynamics. Our natural thermodynamical, intuitive understanding of the arrow of time can be seen through our observation that a fire, once started, will burn until there is no fuel left and then will go out with no chance of re-ignition. Unfortunately, to make significant progress in this essay, we must assume that the reader has a basic background in thermodynamics and has already encountered entropy as an arrow of time.

It is worth taking a moment to refine the common understanding of the definition of entropy as the logarithm of the number of microstates of a system. Firstly, the number of microstates is invariably infinite as the phase space is not discrete but continuous hence we must instead refer to volumes of phase (or mu-) space. Secondly, from Liouville's theorem (Appendix 5.1) we see that the density in phase space evolves as an incompressible fluid hence the phase space volume is constant. We must therefore consider only the volume occupied by particles corresponding to the given macroscopic state. Hence, it should be clarified that the entropy of a system is the measure of the number of possible microscopic states in thermodynamic equilibrium, consistent with its macroscopic thermodynamic properties where the word measure refers to the phase space volume.

Since the introduction of the laws of thermodynamics there have been many well-reasoned objections. For example, many have wondered how an embryo can develop into a fetus and later into an adult. It seems as though a complex system is produced by the time evolution of a simple system which would oppose the second law of thermodynamics. This confusion arises as the surrounding environment is often disregarded and the system is viewed as isolated. The pregnancy stage of growth is notably an open system; energy and material transfers occur between the mother and her child rendering the womb an open not isolated system. Now, considering the whole system, we see that it does not violate the laws of thermodynamics as there are innumerable processes that raise the entropy of the surroundings through heat-production and the expulsion of waste products.

Other objections are significantly harder to refute. The Feynman-Smoluchowski Ratchet (or Brownian Ratchet), proposed by Gabriel Lippman in 1900, consisted of a freely rotating ratchet and pawl connected via an axle to a paddle wheel located in a fluid of molecules. The fluid has a temperature T_1 such that the molecules undergo Brownian motion. This defines the mean kinetic energy of the molecules. In absence of the ratchet and pawl, random collisions of the molecules with the paddle wheel turn the axle such that there is no net motion and no net work is produced. However, as the ratchet and pawl restrict the axle to turn in only one dimension, the system produces a net work through the rotation of the axle. It should be obvious that a net work is produced as a massive object attached to the axle with a string will be raised as the rotation causes the string to coil around the axle. This violates the Kelvin statement of the second law of thermodynamics which states that "it is impossible to devise a cyclically operating thermal engine, the sole effect of which is to absorb energy in the form of heat from a single thermal reservoir and to deliver an equivalent amount of work" [3].

Richard Feynman demostrated that if the ratchet and pawl, and fluid are at the same temperature T_1 then the random fluctuations in the gas that drive the axle also act in the pawl causing it to randomly release from the ratchet allowing the axle to return to it's original position. Another effect is that the spring that returns the pawl to the ratchet, exerts a sideways force on the tooth of the ratchet. A detailed and rigorous proof that no net motion occurs due to these two effects and for any shaped teeth is given by Magnasco. Another possible explanation arises from the lack of friction in the system. Suppose the ratchet were manufactured such that the it would produce the greatest net work i.e. that the ratchet teeth were smooth. When the pawl slides over a tooth of the ratchet, it drops to the bottom of the next tooth. However, as momentum must be conserved, the pawl bounces back up onto the previous tooth and thus the ratchet must remain stationary producing no net work. In order to keep the pawl at the bottom of the tooth, we must introduce friction to the side of the tooth. As the pawl slides over the (now rough) surface of the tooth and produces heat which dissipates, increasing the entropy of the surroundings. Hence this system, although complex, does not violate the laws of thermodynamics.

Our final objection to the theory (which will be useful starting information for the following section) comes in the form of Maxwell's Demon. Consider an isolated system S which contains two thermally-insulated, connected boxes filled with a gas separated by a shutter operated by a demon. The demon is able to very quickly measure the positions and velocities of incoming particles and operates by allowing faster-moving particles to transfer from the lower to higher temperature box and allowing slower-moving particles to transfer from the high to low temperature box. This process moves heat from the cooler gas to the warmer gas without any energy being supplied to the system which is prohibited by Clausius' formulation of the second law of thermodynamics (and hence all other formulations of the law). We must therefore be able to show that it is impossible for the demon to exist.

A common argument against for the possibility of the demon arises from a misconception in the definition of entropy. It is argued that if the evolution of a system is isolated and unobserved and if the laws that govern the system are deterministic and time-reversible then, regardless of the nature of the system, the total number of microconditions cannot possibly decrease thus the entropy of the system must be constant and hence allowing the possibility of the demon. However, as previously mentioned, we must consider the number of microconditions corresponding to a given macrocondition rather than the total number of microconditions.

A famous response, posed by Leó Szilárd, to the question argued that the act of acquiring information about the particles requires an expenditure of energy. This would produce an increase in the entropy of the demon larger than the entropy lost in the process of the gas changing to a new macroscopic state. Rolf Landauer argued against this theory as he realised that a thermodynamically reversible process could be used to measure the molecules without raising the thermodynamic entropy. This lead to the introduction of information entropy whereby the entropy of the demon increases when he deletes the information that he has stored about the particles. This process is entirely necessary as the demon cannot have infinite memory as shown by Charles Bennett. As a result, Maxwell's Demon can lower the entropy of the system but only by retaining information. To proceed back to the full original macrocondition, the information must be deleted and hence the entropy of the system cannot decrease.

To study this system further and to generalise to other systems we must formalise our approach. Let's define the gaseous system as G and the Demon as D such that the entire system is defined as $D+G$. For Maxwell's Demon to exist, we must be able to conclude that the system D has at least one macrocondition M such that if D is in M and G is in A and if D and G are brought into suitable proximity, and left alone, then the macrocondition of G after a certain, particular time is overwhelmingly likely to be B, and the entropy of D at that particular time later is overwhelmingly likely not to have gone up.

Before starting, we must remind ourselves of certain properties of a phase space. Firstly, that for any thermodynamic systems S_1 and S_2 and any macrocondition $\{c_1\}$ of S_1 and any macrocondition $\{c_2\}$ of S_2, the volume of the macrocondition $\{S_1$ is in c_1 and S_2 is in $c_2\}$ in the phase space (S_1 and S_2) is just the volume of the macrocondition $\{c_1\}$ in the phase space of S_1 multiplied by the volume of the macrocondition $\{c_2\}$ in the phase space of S_2. Secondly, following Liouville's theorem, the volumes of any particular initial set of microconditions of any isolated system after any given time evolution is constant. Therefore the fraction of macroscopic phase space occupied by microscopic conditions is the multiple of the two fractions in phase spaces S_1 and S_2. As the multiple of two proper fractions is smaller than either fraction on its own, there cannot, in general, be a system D with a macrocondition M such that the overwhelming majority of phase space volume is taken up by microconditions which will move, over the time-interval in question, into regions of the phase space associated with G being in B and with D being in some particular macrocondition whose entropy is not higher than M.

It is crucial that we notice that the final macrocondition of the demon is not required to be the same on every run nor be independant of the initial microcondition of the two gasses. This means that we can set the initial macrocondition M of D such that almost the entirety of the phase space volume M is occupied by the microconditions. In other words, we shrink the volume of the macrocondition around the microconditions by allowing the demon to learn more about the system hence reducing the number of possible microconditions that correspond to a given macrocondition. The microconditions will now move over the time interval in question into regions of the phase space associated with G being in B and D being in one or another of some set (with cardinality of greater than one) of possible final macroconditions, each of which has a lower entropy than M.

It is worth noting that the contraction of the volume in the phase space volume of D and G need not be solely undertaken in D. For example, a Maxwellian demon might store information about the two gasses using some gigantic array of billiard balls (or any other movable macroscopic device) thus its entropy could conceivably not decrease at all during the experiment.

Suppose, however, that we were trying to create a demon who could reduce the thermodynamic entropy of some given system. This demon must be microscopically sensitive and able to tailor its macroscopic behaviours to the particular microcondition that it measures. As such, the final macrocondition of the system shall not be predictable just as the microconditions are not predictable. It is for these reasons that there is no possible way for the ratchet and pawl mechanism to reduce the thermodynamic entropy.

4 Quantum Mechanics and Determinism

From a basic understanding of quantum mechanics, no obvious description of time reversibility or even determinism is apparent. In fact, the world of quantum mechanics appears to be probabilistic and random with limited predictability. It is likely that we first discovered this learning about the measurement problem and Heisenberg's Uncertainty Principle. Simply put, the more precisely we determine the position of some particle, the less precisely we can determine its momentum. These principle can be expanded to contain any two incompatible observables.

In Newtonian mechanics, Newton's second law is used to determine the latter position of a particle given its initial position and velocity. This equation is a second order differential equation and its solution provides us with the position of the particle at later times. The analogue to this equation in quantum mechanics, is the Time Dependent Schrödinger Equation:

$$\widehat{H} \left|\Psi(t)\right\rangle = i\hbar \frac{d}{dt} \left|\Psi(t)\right\rangle \tag{1}$$

We must notice that this equation is first order with respect to time t hence only the position (and not the velocity as well) is required. This is a crucial point because it means we no longer need to convert between instantaneous states and dynamical conditions because our instantaneous states are the full dynamical conditions of the system. Furthermore, as the quantum-mechanical equations of motion have a partial invariance under time reversal (with regards to position) and these equations must be invariant under a translation in time, the state of the system must be unchanging with time. Clearly this is an issue as the theory seems to suggest that nothing can ever happen.

It is crucial that we base our theories of experiments as it is impossible to derive quantum mechanics from Newtonian mechanics. In doing so we observe that the standard, classical phase space is replaced by an analogous one over a set of possible quantum states and this is known as the statistical postulate. Fortunately, retaining the notion of a phase space allows us to make comparisons to the section on thermodynamics where the phase space of systems was considered in detail. We also observe that the wavefunction is able to evolve (deterministically) until it is measured when the observable becomes fixed. Perhaps we can consider the evolution of the wave function as the motion of microconditions through the phase space and the measurement as the shrinking of the macroscopic phase space volume onto the microconditions.

David Bohm argued that we cannot have two theories for the time evolution of the wave function (one for a wave function under measurement and another for an unobserved wave function). He proposed, with Louis de-Broglie, that there must be some hidden variable that allows for the deterministic time evolution of the wavefunction. The theory argues that the missing variable is the position of every particle that makes up the system. He argues that the wave function now acts in a similar way to a classical force-field which guide particles along their trajectories. However, we cannot know enough about our system (which is, in principle, the entire universe) therefore we must produce a probabilistic weighted average resolving the familiar measurement problem.

There are many other interpretations of quantum mechanics and it would be misleading to provide only the information of those which are time reversible. However, I shall mention one final theory known as the GRW theory after its discoverers G. Ghirardi, A. Rimini and T. Weber. They noticed that the older interpretations focussed too much on the process of measurement despite being unable to prove that it should be such a special act. They proposed that the wavefunction evolves deterministically according to the Schrödinger equation but randomly (about once every 10^9 years), the wavefunction is multiplied by a bell curve. This process localises the wavefunction and sets its value to zero everywhere except for one specific position. The theory therefore makes no attempt to discuss measurements or macroscopic and microscopic properties and is elegant in its simplicity. The theory has shown to be in very good accordance with the standard quantum mechanical probabilities.

5 Conclusion

Throughout this essay, we have explored the distinct direction of time and taken a tried to grasp the possibility of time-reversal and of time evolution and determinism. Our main success is the defence of the thermodynamic arrow of time in the face of general Maxwellian demons but also in the defence of time reversibility in the Newtonian picture. Unfortunately, we met issues when looking at the quantum picture of the world but it is very promising that the GRW theory avoids the measurement problem whilst providing the expected results and I hope that future theories will be created to supplement this theory.

We must, of course, look back and comment further on the limitations of this essay. For example, we cannot provide a decisive, overarching answer to whether the world (as governed by the laws of quantum not Newtonian mechanics) is deterministic or evolves under a single description of time evolution as this is still a topic of active research and has there is no known completely accurate and sufficient theory. Nor is it easy to experimentally verify such a theory. Deterministic theories require a complete description of some given isolated system which, even for the most simple systems, is still beyond our experimental capabilities. Many even require a knowledge of the boundary conditions of the universe which we may, perhaps, never know. Similarly, it is impossible to experimentally verify whether time reversibility is possible in the Newtonian picture of the world because we do not have an available, fully Newtonian world to experiment with.

We have, however, been able to provide a very detailed and thorough understanding of thermodynamics and thus the thermodynamic or entropic arrow of time. We have eliminated many of the common misconceptions about the laws of thermodynamics and commented on many of the main opposing arguments. A main goal of this section was to prove the possibility of a Maxwellian demon, not just in the given example but, in a more general approach. As such, it was shown that it is possible to define a system A that will reduce the thermodynamic entropy of a system $A + B$ by increasing the information entropy thus ensuring the entropy of the universe increases providing a consistent backing for the thermodynamic arrow of time. This will lead to the reader to deeper topics in Information Theory that we were unable to discuss in the course of this essay.

As a final remark it should be mentioned that much was left out of this essay that could potentially provide the answer to our questions. Most notably, I believe the modal interpretation of quantum mechanics and its involvement with retrocausality could provide more clues as to the nature of the universe.

6 Appendix

6.1 Liouville's Theorem[4]

If we define a coordinate in phase space as that spanned by the canonical coordinates and conjugate momenta, e.g (x, y, z, p_x, p_y, p_z) for a single particle and $\{q_1, q_2, ..., q_n, p_1, p_2, ..., p_n\}$ for a system of n particles. As these points fully define the dynamical state of the particles, they are called representative points. These representative points move with velocities v where:

$$\boldsymbol{v} = \{\dot{q_1}, \dot{q_2}, ..., \dot{q_n}, \dot{p_1}, \dot{p_2}, ..., \dot{p_n}\} \tag{2}$$

Using Hamilton's equation, this can be re-written as:

$$\boldsymbol{v} = \{\frac{\partial H}{\partial p_1}, \frac{\partial H}{\partial p_2}, ..., \frac{\partial H}{\partial p_n}, ..., -\frac{\partial H}{\partial q_1}, -\frac{\partial H}{\partial q_2}, ..., -\frac{\partial H}{\partial q_n}\} \tag{3}$$

For the proof we must use the n-dimensional divergence theorem:

$$\int_{\text{volume}} \nabla \cdot \boldsymbol{V} d\tau = \int_{\text{surface}} \boldsymbol{V} \cdot d\boldsymbol{S} \tag{4}$$

Suppose the representative points are confined to some (n-dimensional) volume V with surface S. The points move with a velocity $\boldsymbol{v}$ thus the volume occupied by the points changes at a rate $dV = \boldsymbol{v} \cdot d\boldsymbol{S}$. Hence:

$$\begin{aligned} \Delta V &= \int_{\text{surface}} \boldsymbol{v} \cdot d\boldsymbol{S} \\ &= \int_{\text{volume}} \nabla \cdot \boldsymbol{v} d\tau \end{aligned} \tag{5}$$

However,

$$\begin{aligned} \nabla \cdot \boldsymbol{v} &= \sum_i \frac{\partial}{\partial q_i}\dot{q_i} + \frac{\partial}{\partial p_i}\dot{p_i} \\ &= \sum_i \frac{\partial^2 H}{\partial q_i \partial p_i} - \frac{\partial^2 H}{\partial p_i \partial q_i} = 0 \end{aligned} \tag{6}$$

Therefore the volume, in phase space, occupied by the ensemble's representative points does not change.

Bibliography

[1] David Z. Albert *Time And Chance* Columbia University, 2003.

[2] John D. Norton, *Causation as Folk Science*, University of Pittsburgh, 2003.

[3] Y. V. C. Rao, *Chemical Engineering Thermodynamics*, University Press, 2004.

[4] C Castelnovo, B Gripaios, *Theoretical Physics 1 — Classical Field Theory*, University of Cambridge, 2017.