Time's Arrows Today

Recent physical and philosophical work on the direction of time

Edited by *Steven F. Savitt*
University of British Columbia

156

CAMBRIDGE
UNIVERSITY PRESS

Introduction to "Time's Arrow Today"

STEVEN F. SAVITT

1 Eddington and the running-down of the universe

The phrase 'time's arrow' seems to have entered the discussion of time in Sir Arthur Eddington's Gifford Lectures, which were published in 1928.[1] An 'arrow' of time is a physical process or phenomenon that has (or, at least, seems to have) a definite direction in time. The time reverse of such a process does not (or, at least, does not seem to) occur. Eddington thought he had found such an arrow in the increase of entropy in isolated systems. He wrote:

> The law that entropy always increases – the second law of thermodynamics – holds, I think, the supreme position among the laws of Nature.[2]

Since he held the universe to be an isolated system, he thought that its entropy, which he called its 'random element', must ineluctably increase until it reached thermodynamic equilibrium (until it is 'completely shuffled'), by which point all life, and even time's arrow itself, must have disappeared. He called this process 'the running-down of the universe'. This vision of the universe is stark, compelling, and by no means hopelessly dated. P. W. Atkins recently wrote:

> We have looked through the window on to the world provided by the Second Law, and have seen the naked purposelessness of nature. The deep structure of change is decay; the spring of change in all its forms is the corruption of the quality of energy as it spreads chaotically, irreversibly, and purposelessly in time. All change, and time's arrow, point in the direction of corruption. The experience of time is the gearing of the electrochemical processes in our brains to the purposeless drift into chaos as we sink into equilibrium and the grave.[3]

[1] Eddington (1928), 68.
[2] Eddington (1928), 74.

Eddington's view was not so grim.

> At present we can see no way in which an attack on the second law of thermodynamics could possibly succeed, and I confess that personally I have no great desire that it should succeed in averting the final running-down of the universe. I am no Phoenix worshipper. This is a topic on which science is silent, and all that one can say is prejudice. But since prejudice in favour of a never-ending cycle of rebirth of matter and worlds is often vocal, I may perhaps give voice to the opposite prejudice. I would feel more content that the universe should accomplish some great scheme of evolution and, having achieved whatever may be achieved, lapse back into chaotic changelessness, than that its purpose should be banalised by continual repetition. I am an Evolutionist, not a Multiplicationist. It seems rather stupid to keep doing the same thing over and over again.[4]

Most of this volume is given over to fairly technical discussions of issues in the foundations of physics. It is worth reminding ourselves, before we become immersed in detail, that these issues are linked in many ways to the pictures we have of the fate of the universe and our part in its story. These connections explain, in part, the perennial appeal of problems concerning the existence and nature of time's arrows.

2 G. N. Lewis and 'two-way time'

Although Eddington wrote in 1928 that he could 'see no way in which an attack on the second law of thermodynamics could possibly succeed,' the brilliant chemist G. N. Lewis had mounted just such an attack in his Silliman Lectures, published two years earlier.[5] According to Lewis, the second law is: 'Any system left to itself approaches in a single direction a definite state of equilibrium, or, in other words, its entropy increases steadily toward a maximum.'[6] Of this law, Lewis wrote:

> We . . . find not even a shred of truth left in the statement that an isolated system moves toward the state of equilibrium. It will move toward it and move away from it, and in the long run as often in one direction as in the other. It is only when we start far away from the state of equilibrium, that is, when we start with some state of unusual distinction, and when we follow the system a little way along its path, that we can state that it will, as a rule, proceed toward more *nondescript* states.[7]

4 Eddington (1928), 86.
5 Lewis (1926), chapter 6.
6 Lewis (1926), 143.
7 Lewis (1926), 153.

Lewis, of course, was analysing the behaviour of isolated systems in accordance with the statistical mechanics of Boltzmann and of Gibbs. The laws of statistical mechanics are time symmetric, and hence the behaviour of isolated systems cannot distinguish future from past. On the micro-level in the long run such systems will depart from equilibrium just as often as they approach it. Appearances to the contrary on the macro-level result from the existence of far more nondescript than interesting macro-states. 'It is not true that things left to themselves approach a constant state,' wrote Lewis, 'but only that they approach a state which ordinarily appears constant to us because of the dullness of our perceptions.'[8]

Lewis employed elegant examples to help overcome what he saw as an illusion engendered by coarse-graining.

> Let us consider a box with a one-gram weight resting on its floor. Let us place this box in a bath maintained at an extremely constant temperature, we will say 65 °F, and let the whole be protected by the most perfect mechanism that we can think of to shield it from external jars. Let us, in other words, shut it off from all external influences, leaving only a small hole through which we may observe the weight. We may look into the box millions of times and always find the weight upon the floor, and we then state this to be a law of nature. But the time will come when we look in and find the weight some distance from the floor. This chance is so very small that I cannot express it in any ordinary way. We state chances as fractions, but to denote this chance I should have to put down a decimal point and zero after zero, and would spend my whole lifetime before I could write down a number not a zero. But the calculation is none the less exact.
>
> The chance becomes larger if I consider smaller weights and lesser heights from the floor. Let the height be one hundred million times as small and the weight also one hundred million times as small, and then the calculation shows that if we look in every second we shall find the weight as far off the floor as this 6.32 times in every million years. If you bet five to one on the appearance of this phenomenon, in a million years you might lose at first but would come out ahead in the long run.[9]

From this and other examples Lewis concluded that the concept of time must be split. There is a 'one-way time' that we find in consciousness, memory, and apparently irreversible phenomena, and there is a symmetric or 'two-way time' which is the time of physics. Lewis, then, sharply distinguished between a psychological arrow and the thermodynamic arrow of time, and he denied

8 Lewis (1926), 147.
9 Lewis (1926), 145–6.

that the thermodynamic arrow, the one that Eddington regarded as fundamental, was an arrow at all.[10] This clash epitomizes the philosophical problems raised by the emergence of kinetic theory in the nineteenth century and which lie at the heart of the tangle of problems known as *the problem of the direction of time*.

3 Time's arrows

Eddington had in mind, as we saw, one particular fundamental time-asymmetric process, but by 1979 Roger Penrose was concerned with seven possible 'arrows'.[11]

(1) The decay of the neutral K meson (K^0) seems to be governed by a time-asymmetric law, the only such law in particle physics.[12] The inference to this law is somewhat delicate, however, and the behaviour of K^0 seems to be unrelated to the other more salient time-asymmetric phenomena.[13]

(2) The process of measurement in quantum mechanics, along with its attendant 'collapse of the wave function', is often supposed to be a time-asymmetric phenomenon. William Unruh describes measurement in more detail in section 2 of chapter 1 below, where he, like Penrose in 1979, subscribes to a time-symmetric account. Unruh's paper deals with very broad features of time in twentieth century physical theories and the ways these broad features may need to be modified if we are to develop a successful theory of quantum gravity.

(3) The second law of thermodynamics asserts that in irreversible processes the entropy of isolated systems will increase until it reaches a maximum value. 'After that,' write the authors of a recent physics text, 'nothing else happens.'[14] An analogue of the second law can be derived in statistical mechanics.[15] The relation between the second law and its statistical mechanical analogue, and between both of these and our sense of the direction of time, is puzzling, as we have already seen in Lewis's objections to Eddington.

Barrett and Sober attempt, in a rigorous way, to apply the statistical mechanical formalism to biological examples. In so doing, they find that the

[10] Lewis, of course, did not know of the expansion of the universe when he wrote *The Anatomy of Science*. He did, however, react to the radiative asymmetry (to be introduced in section 3 below) by developing a time-symmetric theory of electromagnetism that was a precursor of the absorber theory. See footnote 10 of Wheeler & Feynman (1945).

[11] Penrose, R. (1979), 582.

[12] The decay violates T-invariance or, in the distinctions to be developed in section 6, it is not time reversal invariant$_1$.

[13] For further details see Horwich (1987), 54–7, or Sachs (1987), chapter 9.

[14] Olenick, Apostol, & Goodstein (1985), 328.

[15] Some elementary properties of entropy in statistical mechanics are derived in appendices A and B of chapter 9 by Barrett & Sober in this volume.

explanation of and conditions for the increase of entropy as given by Khinchin[16] generalize to the biological examples more readily than the classical explanation of entropy increase, which they explain in section 2 of their chapter.

(4) We see radiation emitted from (point) sources in spherical shells expanding in the future time direction. We do not see such shells converging to a point, even though the equations of classical electromagnetism permit such 'advanced' as well as the usual 'retarded' solutions. In section 2 of chapter 4 Philip Stamp outlines what he calls the 'orthodox standpoint' with respect to the various arrows of time, which holds that the radiative arrow is determined by the thermodynamic arrow. The radiative arrow receives no independent consideration in this volume.

(5) The fifth arrow is the direction of psychological time.

The arrow most difficult to comprehend is, ironically, that which is most immediate to our experiences, namely the feeling of relentless forward temporal progression, according to which potentialities seem to be transformed into actualities.[17]

This 'arrow' is rather a grab-bag, for under this heading Penrose mentions our feeling (a) that the future is mutable but the past is fixed, (b) that we know more about the past than the future, and (c) that causation acts toward the future only.

The feeling of 'relentless forward temporal progression' is discussed below in section 4 on 'Becoming'. The future-directedness of causation seems incompatible with the possibility of time travel into the past, a possibility that is defended by Horwich and Earman in the final two chapters in this book.

(6) The expansion of the universe has been invoked to explain the thermodynamic arrow.[18] Would the thermodynamic arrow, then, reverse in a contracting universe? Would our sense of time reverse as well? These questions, amongst others, are taken up by Huw Price in chapter 2.

(7) According to the general theory of relativity (GTR), the gravitational collapse of a sufficiently massive star results in a *black hole*. After the collapse (and neglecting quantum-mechanical effects), 'the hole settles down and remains unchanging until the end of time.'[19] The time reverse of this process, which is in principle permitted by the equations of GTR, would be a singularity, known as a *white hole*, that sits for some indeterminate amount of time from the beginning of the universe and then erupts in a shower of ordinary matter. Black holes may well exist; we have no evidence that white holes do. The 'orthodox standpoint', as presented by Stamp, holds that we do not yet know

[16] Khinchin (1949).

[17] Penrose, R. (1979), 591.

[18] For instance, in Layzer (1975).

[19] Penrose, R. (1979), 600–1.

enough about the nature and existence of such singularities to evaluate this arrow and its relation(s) to the other arrows.

The existence of *any* of these arrows, if they do exist, is puzzling, because all basic theories of physics seem to be time symmetric or time reversal invariant.[20] If these theories do govern the fundamental physical processes in our universe, then it is unclear how processes which are directed in time or are not time symmetric arise. A thorough examination of a number of recent attempts to explain the origin of the thermodynamic arrow, for instance, will be found in Lawrence Sklar's chapter, 'The elusive object of desire'. This chapter is necessarily condensed,[21] but in general he finds that each attempt fails because either it introduces a time-asymmetric postulate which, overtly or covertly, begs the question at issue or because it fails to recognize that its rationale for the increase of entropy of a system in the future time direction would also serve as a rationale for increase of entropy in the past time direction (the 'parity of reasoning' problem). Sklar sees time symmetry at the micro-level, time asymmetry at the macro-level, and no fully compelling connection between the two.

In chapter 2 Huw Price looks at modern cosmology from a standpoint that is similar to Sklar's. Price calls failure to meet the parity of reasoning problem applying a 'double standard', and he shows the double standard at work via a detailed examination of recent cosmological theories of Davies, Penrose, and Hawking.

Difficult questions can be raised about the nature and existence of other arrows as well. Does the wave function really 'collapse', for instance, and what is the nature of this collapse? Must causes precede their effects? Other kinds of problems concern the order of dependence or explanatory priority of pairs of the arrows. Is one of them fundamental, are there clusters of dependencies, or are they unrelated to one other? Does the expansion of the universe, for example, explain the thermodynamic asymmetry (or perhaps *vice versa*)? These problems, lumped under the label *the problem of the direction of time*, straddle the border between physics and philosophy.[22]

The authors in this volume tackle more precise or more narrowly defined versions of these or of closely related questions that may crop up in various

20 Time reversal invariance will be discussed in section 6 below.
21 He examines these issues in more detail in Sklar (1993).
22 In the physics literature extended discussions are to be found in Davies (1974), Sachs (1987), and Zeh (1989). Recent important philosophical works on this subject include Reichenbach (1956), Earman (1974), Horwich (1987), and Sklar (1993).

branches of physics (and, in one case, biology); but before turning to their contributions, a brief examination of a *metaphysical* 'arrow of time' may prove helpful in dispelling some confusions and providing some distinctions that are necessary background for the chapters to follow.

4 Becoming

The problem of the direction of time, as raised by Eddington and Lewis, might seem to have an obvious solution. It seems manifest in our experience that time flows – from the past, to the present moment, and into the future. Newton wrote that absolute time

> of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration.[23]

According to P. T. Landsberg 'the time variable is rather like a straight line on which a point marked "The Now" moves uniformly and inexorably.'[24] Santayana, somewhat more poetically, wrote, 'The essence of nowness runs like fire along the fuse of time.'[25] The arrow *of* time, one could say, points in the direction in which time flows, moves, or runs. The other arrows are arrows *in* time.

Natural as the view of flowing time is, there are a number of philosophical and physical objections to it. The first type of philosophical objection is rooted in the difficulty of giving a literal explanation of the notion of temporal flow. Here is the metaphysician Richard Taylor:[26]

> Of course there is a temptation to say that the present moves in some sense, since the expression 'the present' never designates the same time twice over; a moment no sooner emerges from the future and becomes present than it lapses forever into an ever receding past. But this kind of statement, gravely asserted, says only that the word 'now' never, i.e., at no time, designates more than one time, viz., the time of its utterance. To which we can add that the word 'here' likewise nowhere, i.e., at no place, designates more than one place, viz., the place of its utterance.

Taylor's challenge is to find a literal meaning for the assertion that The Now moves that distinguishes The Now from The Here, the temporal from the

23 Newton (1686), 6.
24 Landsberg (1982), 2.
25 Quoted in Williams (1951).
26 Taylor (1955), 388–9.

spatial, a challenge that has proved surprisingly difficult to meet. An object moves if it occupies different spatial positions at different times. Does The Now move by occupying different temporal positions at different times? If one thinks of The Now's motion in this way (i.e., by analogy with change of spatial position), then The Now's motion, which is supposed to *be* the passage of time, must take place *in* time, so we are covertly postulating a second temporal dimension in which The Now moves. An infinite hierarchy of temporal dimensions beckons. One might try to avoid the spatial analogy by claiming, as did C. D. Broad, that the movement of The Now, which he called 'absolute becoming', is conceptually basic or primitive and hence cannot be explained or conceived in other terms without loss or distortion.[27] Is this more than to point to an unsolved problem?

One might hope to deal with the moving Now in the manner of Austin Dobson.

> Time goes, you say? Ah no!
> Alas, Time stays, *we* go . . . [28]

Dobson is contrasting our passage *through* time with the passage *of* time, but there is less to this contrast than meets the eye. Recall Landsberg's description of dynamic time: 'the time variable is rather like a straight line on which a point marked "The Now" moves uniformly and inexorably.' Suppose, as in a classical spacetime, that all events can be put into exhaustive and mutually exclusive equivalence classes under the relation '*x* is simultaneous with *y*' and that these equivalence classes can be linearly ordered by the relation '*x* is later than *y*' (or, equivalently, '*x* is earlier than *y*'). If we imagine that The Now is a tiny, intense, *static* light and that sets of simultaneous events flow past The Now in the direction from future to past, being illuminated for an instant, we have Dobson's 'Time stays, *we* go . . .' The relativity of motion, however, tells us that this view of time is equivalent to Landsberg's original picture with the light gliding along the static sets of events.

A second philosophical objection is J. M. E. McTaggart's famous argument that the dynamic concept of time is self-contradictory.[29] A moving present, he said, and its cognate notions of past and future (which he called A-determinations, in contrast to the unchanging B-determinations '*x* is later than *y*', '*x* is earlier than *y*', and '*x* is simultaneous with *y*') involve a

[27] Broad (1938), 281

contradiction because, on the one hand, every event can either be present or have but one definite degree of pastness or futurity while, on the other hand, every event must be past, present, and future. This argument might sound like an obvious sophism, but thoughtful philosophers like Michael Dummett, D. H. Mellor, and Paul Horwich[30] have contended that it is either a serious or a conclusive objection to the possibility of a moving Now.

These abstract arguments run counter to everyday experience, which seems to involve at the very deepest level and in the most direct manner the passage of time or, as Eddington called it, 'Becoming'.

> [I]f there is any experience in which this mystery of mental recognition can be interpreted as *insight* rather than as *image-building*, it would be the experience of 'becoming'; because in this case the elaborate nerve mechanism does not intervene. That which consciousness is reading off when it feels the passing moments lies just outside its door.[31]

Recognition of Becoming leads, however, to an uncomfortable dilemma that is delineated by Lawrence Sklar (following Eddington) in chapter 8 of this volume, 'Time in experience and in theoretical description of the world'. Many distinguished thinkers, especially Boltzmann and Reichenbach, held that all the various time-asymmetric features of the world were 'reducible to' or somehow 'grounded in' the entropic asymmetry. But, as Sklar points out, straightforward identification of experienced succession with increased entropy seems wildly implausible.

> We know from perception what it is for one state of a system to be temporally after some other state. And we know what it is for one state to have a more dispersed order structure than another state. We also know that these two relations are not the same.

On the other hand, once we sunder the two, we have no reason to suppose that the 'time' of the physical world is anything like or has any feature like experienced succession. Since the time of Galileo various ostensible sensory qualities or properties of things, like warmth and colour, have been held by reflective scientists and philosophers to be *secondary qualities*, mind-dependent reflections of whatever real properties or structures in the world give rise to them in our consciousness. If succession joins the ranks of the secondary

[30] Dummett (1960), Mellor (1981), chapter 6, and Horwich (1987), chapter 2. I argue in Savitt (1991) that not all the premises in Horwich's version of McTaggart's argument can be simultaneously true.

The rest of this chapter is optional.

qualities as the subjective reflection of entropy increase, then, Sklar argues, we would thereby be deprived of the last insight we have into the real or intrinsic qualities of the world around us.

> Put most crudely, the problem is that at this point the veil of perception has become totally opaque, and we no longer have any grasp at all upon the nature of the physical world itself. We are left with merely the 'instrumental' understanding of theory in that posits about nature bring with them predicted structural constraints upon the known world of experience.

The unpalatability of either horn of this dilemma might lead one to re-think the privileged epistemological status accorded to our knowledge of becoming by both Eddington and Sklar.

5 Physical arguments against Becoming

Some physicists, like Lewis, are sceptical of Becoming as a physical concept. David Park, for example, wrote:

> No formula of mathematical physics involving time implies that time passes, and indeed I have not been able to think of a way in which the idea could be expressed, even if one wanted to, without introducing hypotheses incapable of verification.[32]

In the same vein P. C. W. Davies wrote that 'present day physics makes no provision whatever for a flowing time, or for a moving present moment.'[33] From time to time a stronger claim, that some portion of physics is actually *incompatible* with a dynamic conception of time, is made. Two important arguments of this type appear, conveniently, in one source, Kurt Gödel's 'A remark about the relationship between relativity theory and idealistic philosophy'.[34]

The first argument does not originate with Gödel, and he does not endorse it. A moving Now is supposed to have some special metaphysical significance, to be connected with existence or with the objective lapse of time. In the special theory of relativity (STR), however, the principle of the relativity of simultaneity assures us that there are a nondenumerable infinity of now's, and the standard symmetries assure us that no one of them can have special significance. Becoming does not fit easily into Minkowski spacetime.

[32] Park (1972), 112.
[33] Davies (1977), 3.
[34] Gödel (1949b).

Gödel was concerned, however, that in the general theory of relativity (GTR) it seemed possible to single out certain privileged frames and hence to re-establish something like an 'objective lapse of time'. But GTR, as he had recently discovered, also permitted the existence of a spacetime that contained closed timelike curves (CTCs). Gödel argued that in the model spacetime that he discovered (a) there could not be an objectively lapsing time, and (b) there could therefore be no objective lapse of time in our world either. Gödel's paper is terse, and the exact formulation of his argument is controversial.[35]

Storrs McCall has long tried to show that the argument from STR above lacks force by exhibiting what he calls 'a model of the universe' that is consistent with STR and also has a dynamic time. It is important to note that dynamic time has two metaphysical variants. If one focuses on the present moment, conceiving of 'The Now' as moving smoothly along an ordered series of events from past to future, one could hold that only the present exists – the future not yet having become, the past having become but also having vanished. As an alternative to this *presentism*, one might hold that although future events have indeed not yet come to be out of the myriads of unrealized possibilities, the past as well as the present are fully real. This *probabilist* picture or tree model of time (future possible paths being represented as branches on a tree-structure) most likely traces back to Aristotle,[36] and is the one that McCall attempts to reconcile with the multitude of now's to be found in STR.

Roy Douglas, motivated by the belief that the indeterministic foundation of quantum mechanics indicates an incomplete theory, looks at branching temporal structures from the standpoint of a topologist. In a nearly self-contained presentation, Douglas constructs (though he does not necessarily advocate) a mathematical model of a branching, deterministic spacetime which is specified in terms of local properties only. Roger Penrose concluded an earlier discussion of branching temporal structures by remarking, 'I must . . . return firmly to sanity by saying to myself three times: "spacetime is a *Hausdorff* differentiable manifold; spacetime is a *Hausdorff* . . . " '[37] Douglas's conclusion is, however:

> The (global) Hausdorff property certainly simplifies the mathematics of our models; unfortunately, it is also an entirely inappropriate restriction for models of spacetime, precisely because . . . it is a strictly global constraint . . .[38]

[35] See Yourgrau (1991) and Savitt (1994) for somewhat different views of it.
[36] *De Interpretatione*, chapter 9.
[37] Penrose, R. (1979), 595.
[38] [illegible]

He supports this claim with a detailed analysis of the global/local distinction.

Although Gödel's argument clearly aimed to undermine the notion of an 'objectively elapsing time', most commentators on it were struck by the fact that it raised the possibility of a certain kind of time travel. In fact, his model raised the novel possibility of travelling into one's own past by travelling around a CTC. But what sort of 'possibility' is this? John Earman explores in detail the idea that the global causal structure of spacetime (e.g., the existence of CTCs) may constrain the notion of physical possibility beyond the local constraints of GTR. These constraints, in turn, serve to illuminate the concept of physical law.

The possibility of travelling into one's own past, whether in a spacetime with CTCs or in some other manner, raises the possibility of the existence of closed causal loops. It is frequently alleged that purely conceptual arguments suffice to rule out the existence of closed causal loops, since it is thought that they must permit arrangements with contradictory outcomes. Paul Horwich rejects this conclusion and tries to show the precise senses and circumstances in which closed causal loops are possible.

6 Time reversal invariance

A question that it is natural to ask is: when does a physical theory pick out a preferred direction in time? In fact, the more usual way of approaching this question is by asking: when is it that a theory does *not* pick out a preferred direction in time? When is it, that is, that a theory is *time reversal invariant*?

Suppose that some set of laws, L, are at least a significant component of a scientific theory, T. If these laws involve a time parameter, t, then one can define a *time reversal transform* as the mapping $t: t \rightarrow -t$. If the laws, L, are differential equations, then one can say that T is *time reversal invariant$_1$* if and only if every solution of L is mapped to a (not necessarily distinct) solution of L under the time reversal transform, t. One can find this characterization of time reversal invariance explicitly in P. C. W. Davies' *The Physics of Time Asymmetry*.[39]

Immediately following his official characterization of time reversal invariance, Davies remarks

> The references to 'time reversal' are purely mathematical statements, and have nothing to do with a return to the past. It is to be identified physically with process or velocity reversal.[40]

It is, in fact, in terms of this latter notion that Davies discusses three specific examples which cast some *prima facie* doubt on the time reversal invariance of classical physical theories. His second example, for instance, is the familiar one of the damped motion of a particle moving through a viscous medium or across a surface that exerts friction on it. If the viscous drag or frictional force is assumed to be proportional to $\dot{r}$, where $r(t)$ is the position of the particle at time t and $\dot{r}$ is the total derivative of $r(t)$, then the equation describing the motion of the particle is

$$m\ddot{r} = f(r) - \alpha\dot{r}, \tag{1}$$

where α is a positive constant and m is the particle's mass. If no additional force is acting on the particle, that is, if $f(r) = 0$, then the solutions of (1) that describe the motion of the particle have the form

$$|\dot{r}| \propto e^{[-(\alpha/m)]t}, \tag{2}$$

where $|\dot{r}| \rightarrow 0$ as $t \rightarrow \infty$. Under the time reversal, t, (2) becomes

$$|\dot{r}| \propto e^{[(\alpha/m)t]}. \tag{3}$$

Solutions of the form (3), according to Davies, 'are clearly unphysical as they correspond to a body being spontaneously accelerated to an infinite velocity as a result of its contact with a viscous medium or frictional surface.'[41] Does Davies then conclude from this example that the failure of 'physicality' under the time reversal transform implies that classical mechanics is not time reversal invariant? By no means. He argues, rather, that one has not yet fully factored in the 'process or velocity reversal' mentioned above.

> In the second example of viscous or frictional damping, the motion of the body is slowed by the communication of kinetic energy to the medium atoms in the form of *heat*. It follows that if the motions of the individual atoms are also reversed then, because of the invariance of the laws of physics governing the atomic interactions, each collision will be reversed, causing a cooperative transfer of momentum to the large body, which would then become exponentially accelerated.[42]

Let us pin down the concept of invariance that underlies the above argument. Suppose that the laws, L, of theory T concern some set of properties (or parameters) of a system, S, such as the positions or momenta at some time t of the set of particles making up S. A *state* of a system, S_j (relative to theory T), is some specification of values for all parameters of the components of the system.[43] A sequence of states of a system is *dynamically possible* (relative to theory T) if the sequence of states $S_i \rightarrow S_f$ (indicating that S_i is before S_f) is consistent with the laws of T (is 'a permissible solution of the equations of motion' of T). Finally, let $(S_j)^R$ be the *time-reversed state* of the state S_j. How the time-reversed state is to be specified will, in general, depend upon the theory that is under consideration.[44] We can now say that a theory T is *time reversal invariant*$_2$ under the following circumstances: a sequence of states $S_i \rightarrow S_f$ is dynamically possible (relative to the laws of T, of course) if and only if $(S_f)^R \rightarrow (S_i)^R$ is dynamically possible (relative to T). Davies's argument above successfully, I believe, defends the thesis that classical mechanics is time reversal invariant$_2$ from a purported counter-example.

Unfortunately, but importantly, there is yet a third concept of time reversal invariance that is sometimes used and sometimes conflated with the time reversal invariance$_2$. Let us say that a theory, T, is *time reversal invariant*$_3$ if and only if should S_i evolve to S_f, according to T, then $(S_f)^R$ must evolve to $(S_i)^R$.[45] Time reversal invariance$_3$ captures the essential idea of what John Earman introduced, by way of contrast with what he thought of as genuine time reversal invariance, as the 'reversal of motion transformation – when one asks whether the laws governing a system of particles are "time reversible", one is simply asking whether the laws imply that if the three velocities of the particles were reversed, the particles would retrace their trajectories but with opposite three velocities.'[46] This third concept of time reversal invariance is different from and stronger than the second. No indeterministic theory, T, can be time reversal invariant$_3$, since even in circumstances in which $(S_f)^R$ might evolve to $(S_i)^R$ (and so T might be time reversal invariant$_2$), the laws of T will in general permit $(S_f)^R$ to evolve in other ways as well.

Although I might seem to be multiplying distinctions beyond necessity, teasing out (at least) these three senses of time reversal invariance can help to clarify some recent arguments. Three authors in this volume (Leggett,

43 See section 2.1 of Hughes (1989) for further specification of the state of a classical system.

44 It should, however, be true in general that $((S_j)^R)^R = S_j$. As Davies's examples show, specification of the time-reversed state can be a matter of some subtlety.

45 In an unpublished manuscript, *Explaining Time's Arrow: a How-To Manual*, Craig Callender (Department of Philosophy, Rutgers University) suggests some reasonable names for some of the different sorts of time reversal invariance that I distinguish above. He calls my *time reversal invariance*$_1$ 'Formal TRI', *time reversal invariance*$_2$ 'Motive TRI', and *time reversal invariance*$_3$ 'Actual History TRI'.

46 Earman (1974), 25–6.

Stamp, and Unruh) defend the thesis that quantum mechanics (QM) is time reversal invariant. Leggett has shown that one can experimentally push QM surprisingly far into the 'classical' realm without any 'collapse' or 'reduction' of the state-vector, and in chapter 3 in this volume he incorporates the view that a macroscopic system cannot be in a superposition of states into a broader position that he calls *macrorealism* and then shows that macrorealism can in principle entail an experimental result that is in conflict with orthodox QM.

Stamp develops some of the philosophical implications of current research in the area of macroscopic quantum phenomena by using the idea of a 'quantum computer' as a model of a reversible measurement. This line of thought seems in conflict with a simple thought experiment by means of which Roger Penrose, in a recent, prominent book, *The Emperor's New Mind*,[47] argues that QM is time asymmetric – that it is, in some sense, not time reversal invariant. As an example of the utility of the above distinctions, I want to show that Penrose's argument establishes only the unsurprising conclusion that QM is not time reversal invariant$_3$ and does not contradict the view of Leggett, Stamp, and Unruh.

Penrose claims that a simple experiment shows that the measurement process, R, in QM is 'time-irreversible'. He imagines a source of photons, a lamp L, located at A and aimed precisely at a photon detector P, which is located at C.[48] In between is a half-silvered mirror, M, which is turned at a 45° angle to the path from L to P, permitting half the photons to travel from L to P and reflecting half to a point D on the laboratory wall. Emission and reception of photons is recorded at L and P, and by implication one can determine when photons emitted at L wind up at D.

The crux of Penrose's argument seems to lie in the comparison of a pair of questions regarding conditional probabilities:

(Q) Given that L registers, what is the probability that P registers? and the *time-reverse* question

(QR) Given that P registers, what is the probability that L registers?

I shall assume that the experiment envisaged by Penrose involves a run of photons emitted at L and (shortly thereafter) either detected at P or not detected at P and assumed to have travelled to D. I understand the question

164

47 Penrose (1989), 354–9. The relevant section is entitled 'Time-asymmetry in state-vector reduction'.

48 My discussion is [illegible]

(Q) to ask, given that L registers the emission of a photon, what is the probability that P registers its receipt? One-half, calculates Penrose. Of (QR) he then observes that

> the *correct* experimental answer to this question is not 'one-half' at all, but 'one'. If the photo-cell indeed registers, then it is virtually certain that the photon came from the *lamp* and not from the laboratory wall! In the case of our time-reversed question, the quantum-mechanical calculation has given us *completely the wrong answer*![49]

In light of this remark, it is clear that (QR) is intended to ask something like, given that P registers the receipt of a photon, what is the probability that the photon was emitted by the lamp (and not from anywhere else, like the wall for instance)? It is also clear that the answers provided by Penrose to (Q) and (QR) so understood are correct. What is not clear at all is under what understanding of time reversal invariance the fact that (Q) and (QR) have different answers would lead one to conclude that QM is not time reversal invariant.

Penrose in an earlier publication pointed out that the time evolution of the state of a quantum system according to the Schrödinger equation is time reversal invariant$_1$.[50] He there added that, although the process of measurement might seem to be time asymmetric, he explicitly endorsed a theory of measurement that is time reversal invariant$_1$.[51] Nothing in the present thought experiment *directly* addresses either of those two former positions.

I further claim that QM is time reversal invariant$_2$. As long as the half-silvered mirror is characterized in statistical terms, it is not actually inconsistent with the laws of QM (that is, it is dynamically possible according to the laws of QM) that *all* photons emitted at P wind up back at L. Of course, such a perverse reverse run will be extraordinarily rare, perhaps as rare as the exponential acceleration of the particle by friction in Davies's example if a sufficient number of photons are involved. But the *possibility* of such a run is all that is required for the preservation of time reversal invariance$_2$.

Penrose's thought experiment does illustrate the fact that QM is not time reversal invariant$_3$. The failure of this type of time reversal invariance, however, seems to be fully accounted for in this example by the indeterministic action of the half-silvered mirror and (in the absence of additional,

[49] Penrose (1989), 358.
[50] Penrose, R. (1979), 583.
[51] See Aharonov, Bergmann, & Lebowitz (1964).

controversial assumptions about metaphysical preconditions for indeterr ism) tells us little of interest about temporal asymmetry.

There is yet another sense of time reversal invariance (let us refer to it as *i reversal invariance$_4$*) that Penrose may have in mind. As Lawrence Sklar pointed out, '[I]n the quantum-theoretic context, it is transition probabili between reversed states which must equal the probabilities of the unrever states taken in opposite temporal order for the laws to be time-reve invariant.'[52] Since the transition probabilities (presumably represented this case by (Q) and (QR)) are unequal, it seems that one should concl that the relevant laws of QM are not time reversal invariant$_4$.[53]

Whatever the importance of this last claim might be, I believe that Penro argument is not sufficient to establish it. Two matters must be settled before status of his claim is clear. First of all, what does Penrose consider to be the ti reverse of his proposed thought experiment? One possible reading of wl Penrose has in mind is that the time reverse of the experiment he descril is that sequence of events that would be shown by playing a videotape of t original experiment backward. Another suggestion is, however, that a ti reverse of a given experiment is one in which the 'initial' state is the ti reverse of the final state of the original experiment, and then the time revers experiment evolves from its 'initial' state according to the laws of the thec whose time reversal invariance is in question. For an indeterministic thec like QM, a time reverse of a given sequence of events in the second sen need not coincide with the time reverse in the first sense.

Second, one can ask how the transition probabilities are to be calculate One might challenge Penrose's assertion that the answer to (Q) is 'one-ha while the answer to (QR) is 'one' by questioning whether, even if it is tr that in the course of the time reverse (on the tape-played-backward unde standing) of Penrose's proposed experiment all the photons that register at will register at L, the transition probabilities are to be established by countir the results of that one run. Let us, for convenience, call this approach the *narro frequentist* way of determining the transition probabilities. In opposition to tl narrow frequentist approach, one might suggest that the transition prob bilities are to be determined by averaging over the set of *all* physically possib runs having the same 'initial' conditions. In this latter way of determining th

[52] Sklar (1974), 368.
[53] This might constitute an indirect challenge to the claim that the laws of QM are *in toto* tim reversal invariant$_1$.

transition probabilities, which seems unavoidable if one uses the second sense of the time reverse of a sequence of events, both (Q) and (QR) receive the same answer, 'one-half'. The question of how to determine the transition probabilities is independent of, though it may become confused with, the question of choosing between the two senses described above of the time reverse of the original experiment.

The success of Penrose's argument seems to depend upon choosing the videotape-played-backward version of the time reverse of his proposed thought experiment and the narrow frequentist way of determining the answers to (QR). I see no positive reason to look at the matter in this fashion, and I see two reasons against it. First of all, the answer to (Q) is determined by broad theoretical considerations, rather than by counting the results of one run. How could one justify determining the answer to (QR) differently without begging questions about temporal asymmetry?

Second, consider an experimental set-up much like the one Penrose proposed, except that L shoots out at varying intervals little metal pellets or ball bearings aimed directly at P. Suppose also that in place of the half-silvered mirror there is an aperture controlled by a device that opens and closes it such that (1) the aperture is open exactly half the time during any experimental run and (2) there is no correlation or connection between the device that controls the aperture and the firing mechanism for the ball bearings. It seems that here we have described a classical experimental set-up about which one can ask exact analogs of the questions (Q) and (QR) and give precisely the same answers that Penrose gave in the QM case above, using the videotape-played-backward time reverse or the original experiment and the narrow frequentist method of determining probabilities. If the fact that (Q) and (QR) receive different answers in Penrose's experimental set-up implies that QM is not time reversal invariant, then the fact that the analogues of (Q) and (QR) receive different answers in the variant of Penrose's experiment that I propose should *mutatis mutandis* imply that classical mechanics (CM) is not time reversal invariant in that same sense. I think that Penrose would find this result uninteresting for CM, but to be consistent he should find it uninteresting for QM as well.

There are, I have shown, several assertions that might be intended by the claim that QM is time asymmetric. Despite Penrose's argument, I believe that no stronger claim has been justified than the claim that QM is not time

7 Conclusion

Penrose's seemingly simple thought experiment has called forth a host (distinctions, illustrating a point made by John Earman some years ago.

> [V]ery little progress has been made on the fundamental issues involved in 'th problem of the direction of time.' By itself, this would not be especially surprising since the issues are deep and difficult ones. What is curious, however, is that despite all the spilled ink, the controversy, and the emotion, little progress has been made towards clarifying the issues. Indeed, it seems not a very great exaggeration to say that the main problem with 'the problem of the direction of time' is to figure out exactly what the problem is supposed to be![54]

I hope that the papers contained in this volume will help to clarify anc perhaps even to resolve some of the problems of the direction of time.

[54] Earman (1974), 15.