Brit. J. Phil. Sci. **51** (2000), 381–410

(Dis-)Solving the Puzzle of the Arrow of Radiation

Mathias Frisch

ABSTRACT

I criticize two accounts of the temporal asymmetry of electromagnetic radiation—that of Huw Price, whose account centrally involves a reinterpretation of Wheeler and Feynman's infinite absorber theory, and that of Dieter Zeh. I then offer some reasons for thinking that the purported puzzle of the arrow of radiation does not present a genuine puzzle in need of a solution.

1 Introduction

An electromagnetic field associated with a source of radiation consists of coherently diverging, outgoing waves that appear after the source has been turned on. The corresponding time-reversed phenomena involving converging waves that appear before the source with which they are associated is turned on do not seem to occur. Thus, electromagnetic radiation appears to exhibit a temporal asymmetry: diverging fields are the temporal inverse of converging fields (imagine a film that depicts a diverging wave run backwards), but only the former, but not the latter fields seem to be associated with radiating sources in nature. This apparent asymmetry might not strike anyone as particularly puzzling were it not for the fact that the laws that we use to describe these phenomena are invariant under time reversal. The Maxwell equations, the equations at the heart of classical electrodynamics, are time-symmetric and allow for both converging and diverging fields to be associated with sources of radiation. If the underlying laws are time-symmetric, then where does the temporal asymmetry come from? Why does 'nature choose' one solution to the underlying equations rather than another? In this paper I will discuss three different proposals that are meant to account for the apparent temporal asymmetry of phenomena involving electromagnetic radiation, the proposals due to Huw Price and Dieter Zeh respectively, and a proposal suggested by the account familiar from textbooks in classical electrodynamics.

Huw Price's account, which he presented in a series of papers (Price [1991a], [1991b]; see also [1994]) and more recently in his book *Time's Arrow & Archimedes' Point* ([1996]), centrally involves a reinterpretation of the infinite absorber theory of John Wheeler and Richard Feynman ([1945]). Price argues that what he calls 'the mathematical core' of the Wheeler–Feynman theory can be taken to show that electromagnetic radiation is symmetric on the micro-level. Thus, for Price the apparent asymmetry of radiation is purely a macroscopic phenomenon that is due to cosmological initial conditions and has the same origin as the thermodynamic asymmetry. However, as I will show, Price's reinterpretation of the Wheeler–Feynman theory faces serious problems and his argument for the micro-symmetry of radiation fails. In his book *The Physical Basis of The Direction of Time* ([1989], [1999]), the physicist Dieter Zeh argues that the radiative asymmetry in electrodynamics—the 'radiative arrow'—can be derived from the thermodynamic arrow. But Zeh's argument presupposes what it is meant to show, namely that all actual electromagnetic radiation fields are outgoing, or retarded. Moreover, even if Zeh's account were successful, it would be far too limited in scope, since there is a large class of physically reasonable circumstances to which it does not appear to apply. Thus, we are left with what I will call the 'textbook account'.

This account simply postulates a general physical or, as it is sometimes called, 'causal' constraint—the so-called 'retardation condition'—which imposes a restriction on the set of solutions to the Maxwell equations that can represent physically possible situations. The textbook account does not provide much of an explanation for this constraint and has therefore struck some, including Price and Zeh, as *ad hoc* and unsatisfactory. In the last section of this paper I will try to dispel some of the unease one might feel about the type of answer that the textbook account provides. I will argue that the puzzle of the radiative asymmetry in the form in which it is usually presented arises only if one subscribes to the view that the Maxwell equations on their own delineate the range of what is physically possible. If we reject this view, as I believe we should, there is not much of a puzzle left and the textbook account, thin as it is, is perfectly adequate.

In this paper I will focus exclusively on electromagnetic radiation. Often discussions of the radiative asymmetry are motivated by appealing to non-electromagnetic cases as well, such as that of water waves spreading on a pond. Usually, however, the arguments that are then offered depend crucially on mathematical features of solutions to the Maxwell equations. This is the case for both Price's and Zeh's account. Thus, their discussions (and mine) directly extend to phenomena involving non-electromagnetic radiation only insofar as these phenomena can be modelled by equations that are formally equivalent to the Maxwell equations or the wave equation that can be derived from them.[1]

2 The mathematical background

I want to begin by summarizing some of the mathematical background necessary for understanding the asymmetry associated with electromagnetic radiation (see Jackson [1999], section 6.6; Rohrlich [1965], section 4.7; Zeh [1999], section 2.1). The starting point for constructing mathematical models of radiative phenomena that involve electric charges is the wave equation for the electromagnetic potential, which can be derived from the Maxwell equations. This wave equation is an *inhomogeneous* partial differential equation; this means that the right-hand side of the equation, which specifies the distribution of the electromagnetic charges, or sources, to which the field is coupled, is non-zero. Solutions to the inhomogeneous equation represent electromagnetic fields that are associated with electric charges. If the right hand side of the equation—that is, the source term—is zero, then the equation is what is known as a *homogeneous* partial differential equation, solutions to which represent source-free fields. One can obtain different solutions to the inhomogeneous equation by adding solutions to the homogenous equation to any one solution to the inhomogeneous equation. Differential equations are equations that specify how the values of certain variables change. In order to obtain a particular solution to a differential equation one needs to specify 'what the values change from', that is, one has to specify certain boundary conditions. It is an important fact that any solution to the inhomogeneous equation matching particular boundary conditions can be expressed in terms of any other solution to the inhomogeneous equation, if an appropriate solution to the homogenous equation is added.[2]

The problem of finding the field associated with an arbitrary source distribution is usually approached in terms of so-called 'Green's functions', which specify the field component associated with an infinitesimal point charge. The

[1] In personal correspondence Huw Price has suggested to me that his account should be understood as extending to non-electromagnetic cases of radiation as well. I have no objection to make here against this suggestion. Price's account might very well prove to be a useful way of thinking about non-electromagnetic radiation. My argument here is only that his account fails in the electromagnetic case.

[2] If both $F1$ and $F2$ are solutions to an inhomogeneous differential equation, then $(F1-F2)$ is a solution to the corresponding homogeneous equation. $F1$ can be expressed as $F2+(F1-F2)$, that is, as sum of an arbitrary solution to the inhomogeneous equation and a solution to the homogeneous equation.

total field associated with a given source distribution can be determined by integrating over all the infinitesimal contributions to that source distribution. One particular solution to the wave equation for the electromagnetic potential of a single point charge specifies the potential at a field point P in terms of the unique spacetime point Q at which the world line of the charge intersects the past light cone of P. This solution is known as the *retarded solution*. If we focus on the point Q on the world line of the charge that is picked out by the retarded solution, then all field points whose associated retarded potentials pick out Q lie on the future light cone of Q. Thus, at later and later times field points further and further away from Q depend on the charge at Q; that is, the retarded solution represents an electromagnetic disturbance concentrically diverging from Q into the future. Another solution specifies the potential in terms of the unique intersection of the world line of the charge with the future light cone of the field point P. This is the so-called *advanced solution*. The advanced solution represents an electromagnetic disturbance concentrically converging into a source point Q' from the past.

Since any solution to the wave equation can be represented as the sum of an arbitrary specific solution to the inhomogeneous equation and free fields, any solution can be represented as the sum of a retarded field and a free (incoming) field, $F_{ret} + F_{in}$, or similarly as the sum of an advanced field and a free (outgoing) field, $F_{adv} + F_{out}$. Moreover, since the wave equation is linear, any linear superposition of solutions will also be a solution. Thus, the most general solution to the wave equation can be written as

$$k(F_{ret} + F_{in}) + (1 - k)(F_{adv} + F_{out}) \tag{1}$$

The free field component of the retarded solution is called an *incoming field*, since, in the retarded case, the free field contribution to the value of the field at P in a certain region R of spacetime is given in terms of the value of the free field at the intersection of the *past* light cone of P with a space-like hyperplane which constitutes the past boundary of R. Similarly, the free field is called *outgoing* in the case of the advanced solution, because the relevant boundary conditions are now those on a hyperplane in the *future* of P.

Even though both retarded and advanced solutions (and any linear combination of the two) are allowed by the Maxwell equations, the retarded solution is that solution to the inhomogeneous equation which for a source configuration in the absence of external fields appears to represent the physical situation correctly. The field associated with a single charge satisfies what is known as the 'Sommerfeld radiation condition': the free incoming field F_{in} is equal to zero. Of course, this field can alternatively be represented as the sum of an advanced field and a source-free field, $F_{adv} + F_{out}$, but the two representations are not symmetric: the latter representation includes a source-free field, while the former does not.

The problem, then, of the asymmetry of electromagnetic radiation is this: given that the Maxwell equations are symmetric in time and do not by themselves distinguish in any way between retarded and advanced solutions, why is it that the radiation field associated with a charge can be represented as being fully retarded (but not as being fully advanced)? Physics textbooks (see Jackson [1999]; Rohrlich [1965]) offer answers like the following to this question: electromagnetic disturbances propagate at a finite velocity (which in vacuum is the speed of light *c*) into the future. Thus, one should expect the field at a time *t* some distance away from an electric charge to depend not on the motion of the charge at *t*, but rather on the motion of the charge at a time t_R *earlier* than *t* and at a position *R* which will in general be different from the charge's position at *t*, where t_R is determined by the time it takes for the disturbance to travel from the retarded point *R* to the observation point. The temporal direction in which electromagnetic disturbances propagate, therefore, imposes a 'causal' (Jackson [1999], p. 245; Rohrlich [1965], p. 77) or 'physical' (Rohrlich [1965], p. 79) constraint on possible solutions to the wave equation, a constraint which is satisfied only by the retarded solution. The advanced solution, by contrast, does not satisfy the constraint; it requires that an electromagnetic disturbance associated with a charge be present before the relevant motion of the charge—charges would have to radiate 'backwards in time'.

One basic desideratum for a theory that describes the interaction of charges with electromagnetic fields is that the theory should allow us to derive an equation of motion for a charge in a field. However, it has proven to be surprisingly difficult to derive such an equation in classical electrodynamics in a coherent and satisfactory way. The most widely accepted proposal for an equation of motion is due to Dirac. Since Wheeler and Feynman's infinite absorber theory, which features prominently in Price's account of the radiative asymmetry, has to be understood as an attempt to overcome some of the difficulties associated with Dirac's derivation, I want to mention Dirac's theory briefly.

Dirac ([1938]) presents a derivation of what has come to be known as the 'Lorentz–Dirac equation', a classical equation of motion for a point charge in an electromagnetic field. Unlike the Lorentz law, which is familiar from introductory physics courses, the Lorentz–Dirac equation takes into account the radiative reaction of the charge—that is, the effect of the charge's own field on the motion of the charge. The difficulty one faces in trying to include the radiative reaction is that the field of a charge is infinite at the location of the charge.[3] In order to overcome this difficulty, Dirac does the following. He assumes that all actual

[3] This infinity is familiar from elementary electrodynamics. The Coulomb field of a point charge in the rest frame of the charge varies as $1/r^2$, which diverges as r goes towards zero.

fields associated with charges are fully retarded, but then rewrites the retarded field of a charge a as

$$F^a{}_{ret} = 1/2(F^a{}_{ret} + F^a{}_{adv}) + 1/2(F^a{}_{ret} - F^a{}_{adv}) \tag{2.1}$$

The first term on the right-hand side is the problematic infinite term. Since this term formally acts like a mass term in the equation of motion, Dirac proposes that one should simply absorb it into the total finite mass of the charge (which effectively means that one has to postulate a negative diverging 'bare' mass to make up for the diverging electromagnetic mass). This procedure, which is known as 'renormalization', has become a standard procedure in quantum field theories. The second term, which is finite at the location of the charge, represents the radiative reaction. This term can be calculated explicitly and turns out to depend not only on the acceleration of the charge, but also on the derivative of the acceleration. The total field force acting on a charge is given by the sum of this radiative reaction and the retarded fields of all other charges, plus any free incoming fields:

$$\sum_{(k \neq a)} F^k{}_{ret} + F_{in} + 1/2(F^a{}_{ret} - F^a{}_{adv}) \tag{2.2}$$

The explicit form of the Lorentz–Dirac equation need not concern us here.

Wheeler and Feynman propose their infinite absorber theory as an alternative way of deriving Dirac's equation of motion that avoids what they take to be serious conceptual problems with Dirac's approach. The first problem they see concerns the very issue of temporal asymmetries. Wheeler and Feynman take it to be unsatisfactory that Dirac's derivation, according to which all actual fields are fully retarded, is based on a temporally asymmetric interaction between charges and fields, even though the underlying basic equations, the Maxwell equations, are time-symmetric. By contrast, the Wheeler–Feynman theory assumes that the interaction is symmetric in time, half retarded plus half advanced. The second problem stems from Dirac's treatment of the self field of a charge. The procedure of mass renormalization might strike one as conceptually or mathematically problematic. Wheeler and Feynman's theory avoids any problems associated with renormalization by assuming that the electromagnetic force on a charge is due only to the fields of all other charges. According to the theory, there is no self-interaction and hence no infinities need to be 'swept under the rug'. The two putative advantages of the Wheeler–Feynman theory, however, come at a price. In order to arrive at the same equation of motion as Dirac and, in particular, in order to generate the radiative reaction term in a theory without self fields, Wheeler and Feynman have to postulate that the charge in question is surrounded by an infinite absorber. I will return to the Wheeler–Feynman theory below when I criticize Price's reinterpretation of that theory.

3 Price's argument for the micro-symmetry of radiation

Zeh, along with textbook authors such as Jackson and Rohrlich, takes the radiative asymmetry to be characterized by the claim that

> (3.0) All accelerated charges (or sources) can be associated with fully retarded (but not with fully advanced) radiation fields.

These physicists also agree that the microscopic fields associated with individual charges exhibit the asymmetry. By contrast, Price argues that the apparent asymmetry of radiation arises only for fields of macroscopic collections of charges and maintains that the asymmetry should be characterized by the claim that

> (3.1) Organized waves get emitted, but only disorganized waves get absorbed.[4]

Price makes much of the distinction between emitters and absorbers in his account. An emitter is a charge or collection of charges that emits electromagnetic energy, while an absorber is a charge that absorbs energy. Since Price seems to hold that only emitters, but not absorbers of radiation are associated with retarded waves, he contrasts (3.1) not with (3.0) but rather with

> (3.2) All emitters produce *retarded* rather than *advanced* wave fronts.

The difference between (3.0) and (3.2) is that according to (3.0) *all* electric charges, independently of whether they act as emitters or absorbers of energy, are associated with retarded fields, while (3.2) makes a claim only about charges that act as emitters.[5]

Price argues, first, that the truth of (3.2) does not imply that radiation is asymmetric and, hence, that the putative asymmetry of radiation is best captured by (3.1), and not by (3.2). He argues, second, that (3.1) is false on the micro-level, where radiation is fully symmetric. In his argument for the micro-symmetry of radiation Price appeals to a reinterpretation of Wheeler and Feynman's infinite absorber theory, which Price takes to show that, for any given configuration of emitters and absorbers, the retarded fields associated with the emitters can equivalently be represented as a superposition of coherent

[4] Claims (3.1.), (3.2), and (3.3) are the claims Price distinguishes in ([1996], pp. 60–61). My numbering here is identical to Price's there.

[5] The way I use the term, a field *source* is any object that couples to the field in question, be it an emitter or an absorber. Thus, I use the terms 'source' and 'charge' interchangeably. Price, on the other hand, uses 'source' to refer only to net emitters of radiation, while he refers to absorbers also as 'sinks'. In his earlier account ([1991a]) Price uses Roman numerals to distinguish various claims. Claim **II** in ([1991a]) plays the role of (3.2) in ([1996]), but contains the term 'source' instead of 'emitter': 'All sources of coherent radiation produce *retarded* rather than *advanced* wavefronts' ([1991a], p. 962). Price is explicit about the fact, however, that he means 'emitter' by 'source': '**II** focuses our attention on emitters' ([1991a], p. 963). Thus, despite the appearance, **II** is equivalent to (3.2), and not to (3.0). I would like to thank an anonymous referee for drawing my attention to this difference in formulation in Price's two versions of his account.

converging (advanced) waves centered on the absorber particles. From this Price concludes that radiative processes are symmetric on the micro-level in the sense that both emissions and absorptions can be associated with organized waves. That is, on the micro-level (3.1) is false and the following holds:

> (3.3) Both emitters and absorbers are centered on coherent wave fronts (these being outgoing in the first case and incoming in the second).

Price believes that the puzzle of the arrow of radiation presents a genuine puzzle; he believes that *if* electromagnetic radiation is asymmetric, then this calls out for an explanation. But he argues that the antecedent of this conditional is false on the micro-level—radiation is symmetric on that level—and that therefore only the apparent macro-asymmetry of radiation is in need of an explanation.[6] The apparent asymmetry of radiation on the macro-level, according to Price, is due to the fact that, because of cosmological initial conditions, there are large, macroscopic coherent emitters but no macroscopic *coherent* absorbers.

Against Price's account I now want to argue the following. First, Price's reinterpretation of the Wheeler–Feynman theory is unacceptable, since it is inconsistent both with the central assumption of that theory and with classical electrodynamics in general. Second, (3.3), properly understood, follows straightforwardly from the Maxwell equations, and is, contrary to what Price seems to think, not in need of support from a reinterpretation of Wheeler and Feynman's theory. Yet, third, the truth of (3.3) does not establish Price's conclusion that radiation is symmetric on the micro-level, since the asymmetry of radiation is correctly captured by (3.0) and not by (3.1); and (3.3) does not imply that (3.0) is false. Thus, Price has not shown that electromagnetic radiation is symmetric on the micro-level and has therefore not 'solved' the puzzle of the arrow of radiation.

3.1 The infinite absorber theory and Price's reinterpretation

At the heart of Price's discussion of the radiative asymmetry lies his reinterpretation of Wheeler and Feynman's infinite absorber theory. According to the Wheeler–Feynman theory, the field associated with an electric charge is symmetric in time, half retarded plus half advanced, in contrast with the customary fully retarded field. Wheeler and Feynman ([1945]) show that such a time symmetric field leads to the same equations of motions for a (classical) charge in an electromagnetic field as a fully retarded field, provided one postulates that the charge is surrounded by an infinite absorber. According

[6] By contrast, the account I will advocate below says that radiation is asymmetric on the micro-level, but denies that this is a fact in need of an explanation. Thus, both Price and I believe that there is no microscopic puzzle of the arrow of radiation, albeit for very different reasons.

to Price, however, Wheeler and Feynman's mathematical derivations should not be read as supporting the idea that the radiation associated with a charge is half retarded plus half advanced; rather they should be taken to show that any outgoing, fully retarded field can equivalently be represented as a sum of incoming, advanced fields. Thus, Price agrees with Wheeler and Feynman that classical electrodynamics can be given a fully symmetric formulation, but he disagrees as to where the symmetry of electrodynamics should properly be located.

In their paper Wheeler and Feynman offer four different derivations of the equivalence of their theory with Dirac's classical theory of the electron, all four of which rely crucially on the infinite absorber assumption. The first three derivations rely on making certain special assumptions about the nature of the absorber to explicitly derive the absorber field. The fourth and most general derivation is also the simplest. Since this derivation will be useful in understanding my criticism of Price's reinterpretation, I want to give a brief summary of this derivation.

If a charge is surrounded by an infinite absorber, then all fields vanish outside the absorber. Thus,

$$\sum F_{ret} + \sum F_{adv} = 0 \text{ (outside the absorber).} \tag{4}$$

Since the retarded fields represent an outgoing wave and the advanced fields an incoming wave and complete destructive interference between two such waves is impossible, the two sums have to vanish individually:

$$\sum F_{ret} = 0 \text{ and } \sum F_{adv} = 0 \text{ (outside the absorber).} \tag{5}$$

Then the difference of the two fields vanishes as well:

$$\sum F_{ret} - \sum F_{adv} = 0 \text{ (outside the absorber).} \tag{6}$$

As Dirac has shown, the difference between the retarded and the advanced field is a source-free field, from which it follows that if this field vanishes somewhere, it has to vanish everywhere and not just outside the absorber:

$$\sum F_{ret} - \sum F_{adv} = 0 \text{ (everywhere).} \tag{7}$$

Now, the field exerting a force on a point charge a surrounded by an infinite absorber, which by assumption is given by the sum of the half retarded-half advanced fields of the absorber particles, can be rewritten as follows:

$$\begin{aligned} 1/2 \sum_{(k \neq a)} (F^k{}_{ret} + F^k{}_{adv}) = \sum_{(k \neq a)} F^k{}_{ret} + 1/2(F^a{}_{ret} - F^a{}_{adv}) \\ - 1/2 \sum_{(all\ k)} (F^k{}_{ret} - F^k{}_{adv}). \end{aligned} \tag{8}$$

Here the sum on the left-hand side and the first sum on the right-hand side are taken over all the absorber particles; that is, over all particles except the charge a. According to (7), the last sum on the right hand side vanishes, so that the field force acting on a is

$$\sum_{(k \neq a)} F^k{}_{ret} + 1/2(F^a{}_{ret} - F^a{}_{adv}). \tag{9}$$

The first term represents the familiar external retarded fields due to all other charges, while the second term is the radiative reaction term that Dirac had evaluated. (9) is equivalent to (2.2), the equation derived by Dirac, if one assumes, as do Wheeler and Feynman, that there are no free incoming fields.

Price motivates his reinterpretation of the infinite absorber theory by criticizing not this derivation but of one of Wheeler and Feynman's earlier explicit calculations of the absorber field, which relies on certain special assumptions concerning the distribution of the absorber particles. In that derivation Wheeler and Feynman argue that a 1/2 retarded plus 1/2 advanced field of a charge results in an advanced response field of the absorber equal to 1/2 the retarded minus 1/2 the advanced field of the charge. Adding the two contributions, one obtains a field equal to the fully retarded field of the source. Wheeler and Feynman arrive at their conclusion by discussing what effect the original 1/2 retarded field of a charge has on the absorber. I think Price is correct in claiming that the logic of that derivation is somewhat murky (see Price [1996], pp. 67–70). For example, Wheeler and Feynman's argument seems to require that we ignore that the charge is associated with an advanced field as well. By parity of reasoning, this advanced field should result in an earlier retarded field of the absorber, which should cancel out the retarded field due to the source, in apparent conflict to what Wheeler and Feynman intend to show.

Price also offers a more general criticism of Wheeler and Feynman's explanation of the apparent asymmetry of radiation. Since Wheeler and Feynman argue that electromagnetic radiation is symmetric, they need to explain how it is that radiation nevertheless appears to be asymmetric. Wheeler and Feynman's general derivation of the equivalence of their symmetric theory with a purely retarded field theory can be used equally to show the equivalence of their theory with a purely advanced field theory, as they themselves show. So why is it then, that electromagnetic radiation to us appears to be fully retarded? Appealing to thermodynamic considerations Wheeler and Feynman argue that in situations in which we are interested the retarded representation is the only one that can be applied in practice. The retarded representation of the total field acting on a source is given by (9), while the advanced representation of the field is

$$\sum_{(k \neq a)} F^k{}_{adv} - 1/2(F^a{}_{ret} - F^a{}_{adv}) \tag{10}$$

Wheeler and Feynman argue that before the source a turns on, the absorber particles will be in random motion or at rest so that the retarded absorber field in (9) can be taken to be equal to zero. The advanced absorber field in (10), on the other hand, will not be zero, since the radiation from the source will lead to correlated motions among the absorber particles and a coherent advanced response wave. Both (9) and (10) provide us with correct representations of the field at a, but (10) cannot be used to calculate the total fields in practice, since, according to Wheeler and Feynman, that would require the impossible task of keeping track of all the absorber fields.

Price argues, I think correctly, that Wheeler and Feynman's reasoning is guilty of what he calls the 'temporal double standard' fallacy ([1996], p. 68). If the radiation due to the source were in fact half retarded and half advanced, then the advanced component of that field should result in correlated motions of the absorber particles equivalent to those due to the retarded component of the field, and there should be a non-zero retarded 'response' wave as well. Thus, Wheeler and Feynman have not show how it is that all radiation produced by sources appears to be fully retarded rather than fully advanced.[7]

Despite these criticisms, Price thinks that what he calls the 'mathematical core' of the Wheeler–Feynman theory can be saved and incorporated into a successful reinterpretation of that theory. However, as I want to show now, Price's reinterpretation faces a number of serious problems of its own. First, it is unclear whether Price's reinterpretation leaves room for the radiative reaction term in the equation of motion. Second, Price's proposal conflicts with Wheeler and Feynman's central assumption of an infinite absorber and, thus, cannot constitute a mere reinterpretation of the mathematical core of the Wheeler–Feynman theory. Third, and to my mind most damaging to the proposal, it is incompatible with the Maxwell equations.

Price says that

> the real lesson of the Wheeler–Feynman argument is that the same radiation field may be described equivalently either as a coherent wave front diverging from [the charge a], or as the sum of coherent wave fronts converging on the absorber particles ([1996], p. 71).

where the diverging wave is 'a fully retarded wave' and the converging waves are 'fully advanced' (*ibid*., p. 70). Price, thus, claims that Wheeler and Feynman can be taken to have shown that the retarded field of the charge a is identical to the sum of the advanced fields of the absorber particles; the two fields, he says, '*are one and the same*' (*ibid*., p. 71, italics in original). That is, according to Price, Wheeler and Feynman can be taken to have established that

$$F^{a}{}_{ret} = \sum_{(k \neq a)} F^{k}{}_{adv} \tag{11}$$

[7] Davies ([1974], Ch. 5, esp. p. 144) seems to endorse an argument similar to that of Wheeler and Feynman in his exposition of the absorber theory and Price's criticism applies with equal force to that argument.

for the field associated with a charge a that is surrounded by an infinite absorber.[8]

A first problem for Price's reinterpretation is this. In reintroducing the fully retarded field associated with a charge his theory faces the following dilemma. Either, he assumes with the traditional view that charges interact with their own fields; then his theory needs to deal with the same infinities that Wheeler and Feynman tried to avoid. Or, with Wheeler and Feynman he assumes that the force on a charge is due only to the fields associated with all other charges. But then it is not clear how the radiative reaction term, $F^a{}_{ret} - F^a{}_{adv}$, can arise in Price's theory. This term, which is due to the self-interaction in Dirac's theory, can arise in the Wheeler–Feynman theory only because of the way in which the time-symmetric fields of the absorber interact with the time-symmetric fields of the source. But there appears to be no way to generate the term on Price's proposal, if it is not to include self fields. To see this, imagine what the field would be at the location of a second charge b some distance away from a. If b acted as part of the absorber, then it would follow from (11) that the following equation should hold:

$$F^a{}_{ret} = \sum_{(k \neq a,b)} F^k{}_{adv} + F^b{}_{adv}. \qquad (11')$$

Thus, according to Price, the total field at the location of b is equal in value to $F^a{}_{ret}$ and does not include the necessary radiative reaction term $1/2(F^b{}_{ret} - F^b{}_{adv})$. In general, b will experience a non-zero acceleration due to the field of the charge a. If b is accelerated, then, according to the Maxwell equations, there should be a radiation field associated with the charge and the radiative reaction term cannot be zero.

Since the second horn of the dilemma is clearly unacceptable, because without the radiative reaction term the very phenomenon Price's theory is meant to account for—radiation—disappears from the theory, Price seems to be forced to accept the first horn. But then his proposal to associate charges with fully retarded fields begins to look strikingly similar to Dirac's theory, except, of course, for the equation (11), which appears to be the new insight gleaned from the Wheeler–Feynman theory.

However, (11) quickly leads into serious problems. From (11) it follows that

$$(F^a{}_{ret} + F^a{}_{adv}) = \sum_{(all\ k)} F^k{}_{adv}. \qquad (12)$$

[8] Price does not explicitly write down this equation, but it seems to be the only natural way of understanding the quotation above. The left-hand side represents a 'coherent wave front diverging from' a, while the right-hand side represents a 'sum of coherent wave fronts converging on the absorber particles'. Price claims that his proposal represents a reinterpretation of the mathematical core of the Wheeler–Feynman theory, yet unfortunately he includes no mathematical formulae in his presentation of the proposal. I think it would have been helpful had he presented at least some of the mathematics involved explicitly and shown how it lends itself to his reinterpretation.

The left-hand side of this equation will in general not be equal to zero far away from the charge. Thus, the right-hand side does in general not vanish everywhere outside of the system at issue, in contradiction to (4), which embodies Wheeler and Feynman's assumption that the charge is surrounded by a complete, or infinite, absorber. Therefore, Price's proposal is inconsistent with the central assumption of the absorber theory, and hence cannot be a mere reinterpretation of the mathematical core of that theory.

The point that Price's 'reinterpretation' is incompatible with the mathematical results of Wheeler and Feynman has, as far as I know, first been made by Leeds ([1994]).[9] Ridderbos ([1997], p. 484) makes an objection similar to that of Leeds.[10] Both Leeds and Ridderbos focus on the fact that according to Wheeler and Feynman's picture there is a non-zero advanced response field of the absorber (which is present even before the charge *a* accelerates) and that this field is absent in Price's reinterpretation. By contrast, I have here focused on Wheeler and Feynman's infinite absorber assumption to argue that Price's reinterpretation is mathematically inconsistent with this assumption. The advantage of presenting the incompatibility between Wheeler and Feynman's theory and Price's reinterpretation in this particular way is that one can see immediately that Price cannot avail himself of their mathematical reasoning, since the absorber assumption (4) is crucial to Wheeler and Feynman's derivations.[11]

Now the fact that Price's account is mathematically inconsistent with Wheeler and Feynman's infinite absorber theory does not provide sufficient grounds for rejecting the account, only for rejecting it as a reinterpretation of that theory. Even if Price's account is inconsistent with the absorber theory, the account might still be interesting in its own right as a way of establishing that electromagnetic radiation is symmetric on the micro-level. But unfortunately there is a more serious problem with Price's proposal: It violates Gauss's Law which is part of the Maxwell equations.

Half of the Maxwell equations can be written as

$$dF = 4\pi j, \tag{13}$$

which says that the four-dimensional divergence of the field F is related to the four-current j. Now let us focus on a region surrounding the charge a. The retarded field of the charge obviously has a source in this region, the advanced field of the absorber particles, however, does not. The retarded field associated

[9] See also Price's reply to Leeds (Price [1994]) and Leeds's reply to the reply (Leeds [1994]).

[10] I would like to thank an anonymous referee for drawing Ridderbos's paper to my attention.

[11] Thus, it is unclear why Price believes on the one hand, that his account is a reinterpretation of the mathematical core of Wheeler–Feynman, and on the other hand, that his reinterpretation has 'no need for a future absorber' ([1996], p. 73). If the universe was not completely opaque, then, as Wheeler and Feynman show, their theory would predict the existence of explicit advanced effects and the radiation field would not be equal to the fully retarded field of the source. See (Ridderbos [1997], p. 485) for essentially the same point.

with a is by definition a solution to the Maxwell equations if a is the only charge in the world. Similarly, the advanced field associated with the absorber is a solution to the Maxwell equations if the absorber particles are the only charges. Thus, at the location of the charge the divergence of the absorber field $\sum_{(k \neq a)} dF^{k}{}_{adv}$ is equal to zero, while the divergence of the field of the charge $dF^{a}{}_{ret}$ is not. Therefore, if the two fields are identical, then one of them cannot satisfy the Maxwell equations. Or, alternatively, it follows from the Maxwell equations that (11) is false.[12]

In my arguments I have assumed that (11), like Wheeler and Feynman's equation (9), is meant to hold everywhere in spacetime. Have I perhaps misconstrued Price in taking him to be committed to (11) read in this way? Price's reinterpretation, as I said above, is motivated by one of Wheeler and Feynman's earlier derivations of the equivalence between their proposal and Dirac's theory. In that derivation Wheeler and Feynman explicitly show that the total advanced field due to all the absorber particles is equal to the difference between the retarded and advanced fields of the source under the assumption that the retarded absorber field is equal to zero:

$$\sum_{(k \neq a)} F^{k}{}_{adv} = F^{a}{}_{ret} - F^{a}{}_{adv} \left(\text{in regions where } \sum_{(k \neq a)} F^{a}{}_{ret} = 0 \right). \qquad (14)$$

Thus, in spacetime regions where $F^{a}{}_{adv}$ is equal to zero as well (11) holds. For example, for a source radiating for some finite time the advanced field is zero long before the source has turned on and after the source has turned off.

In fact, Price explicitly says that Wheeler and Feynman show that the retarded and advanced fields are equal in value in a certain spacetime region—'in the region between [source *a*] and the receiver, after the initial acceleration of [*a*]' ([1996], p. 71).[13] So is he perhaps committed only to the unassailable claim that (11) holds in that region? Since Price wants to show that the retarded field of the source and the advanced field of the absorber '*are one and the same*' (*ibid*., p. 71, italics in original), the restricted equality derivable from (14) is clearly not enough for his purposes. All that (14) allows us to conclude is that there is a region of spacetime in which two representations of fields are equal in value. Price needs more than that, however, for his argument against Wheeler and Feynman's original theory to succeed. Price argues that what Wheeler and Feynman take to be two distinct fields are in fact

[12] One might want to think here of the static case in which the charge a is at rest and the field associated with a is a static electric field. Then Gauss's law in its integral form says that the flux of the electric field through any closed surface is equal to the total charge enclosed by the surface. Hence for any surface which encloses the charge a but none of the absorber particles the flux of the field associated with a is non-zero, while the total flux of the absorber field is zero; the two fields cannot be equal.

[13] This gloss on Wheeler and Feynman's result is not quite right. The fields are equal only after the charge *stops* accelerating and not in general after the initial acceleration.

two different representations of one and the same field. But one way in which one might show that Wheeler and Feynman's original interpretation is correct while that of Price is mistaken is by showing that the two representations agree only within a limited region of spacetime. If the two representations do not agree everywhere (which in fact they do not), then they cannot be representations of one and the same field.

3.2 What is the role of the Wheeler–Feynman theory in Price's argument?

Price proposes his reinterpretation of Wheeler and Feynman's infinite absorber theory as an argument for (3.3), the claim that '[b]oth emitters and absorbers are centered on coherent wave fronts (these being outgoing in the first case and incoming in the second)'. In the last section I showed that Price's reinterpretation is unacceptable since it is inconsistent with the Maxwell equations. In this section I want to argue that, nevertheless, (3.3) is true in classical electrodynamics. (3.3) follows directly from the Maxwell equations and is in no need of support from a controversial theory such as that of Wheeler and Feynman.

The precise structure of Price's argument is not entirely clear from what Price says. It appears that Price begins his argument by strictly identifying emission of radiation with retarded fields and absorption with advanced fields. Thus, after describing the set-up of the Wheeler–Feynman theory involving a radiating source and a collection of absorber particles Price says:

> Because [the absorber particles] are receivers, we expect that from their point of view the radiation associated with this field [i.e. the field of the source] is fully advanced, or incoming. However, let us now assume that *contrary to appearances*, this radiation is coherently centered on the absorber particles. In other words, we assume that each absorber particle is centered on what in the usual time sense looks like a converging wave front (*ibid*., pp. 70–71, italics in original).

Price then goes on to argue that the 'mathematical core' of the Wheeler–Feynman theory can be used to show that the assumption that each absorber particle is the center of an incoming wave is in fact compatible with the appearances.

This passage is rather puzzling. First, it is not clear what contrast Price wants to draw here since, as I explained above, for an absorber particle to be associated with an advanced field *is* for this particle to be associated with a coherent converging wave front. If, according to Price, it is 'contrary to appearances' for the absorber particles to be associated with an advanced field why does he think that this is what 'we expect' to be the case?[14] Second, once we assume that absorbers are associated with advanced waves, (3.3)

[14] Perhaps by 'advanced' Price simply means 'incoming' here (contrary to common usage), despite the fact that he introduces the term 'advanced' in the standard way earlier (see [1996], p. 50). I discuss the consequences of this alternative reading in the following paragraphs.

follows immediately and it is not clear what work is left to do for Price's reinterpretation of the absorber theory. If every absorber particle is associated with an advanced field, then it is associated with a converging coherent wave centered on the absorber particle and it follows directly that (3.1) is false on the micro-level. Any appeal to the Wheeler-Feynman theory is superfluous.

Perhaps, then, *pace* what Price says in the above passage, the Wheeler–Feynman theory is not meant *to presuppose* but rather *to establish* that absorbers can be associated with coherent, advanced wave-fronts. But once again it is not clear why Price thinks that he needs to appeal to the Wheeler–Feynman theory in order to establish that.

There are two claims implicit in (3.3). The first is that all sources, whether they act as emitters or absorbers of radiative energy, are centered on coherent wave fronts. The second claim is that emitters are associated with outgoing waves, while absorbers are associated with incoming waves. The second claim can be read as a conditional: If all sources are centered on coherent wave fronts, then emitters are associated with outgoing wave fronts and absorbers are associated with incoming wave fronts. I want to focus on this second claim first. Price says that absorbers *are* associated with advanced waves and not merely that they *can be* associated with advanced waves. Now if this claim is meant to imply that the fields associated with absorbers somehow are truly advanced but not retarded, then the claim is false. It is not the case that absorbers can only be associated with advanced fields and cannot be associated with retarded fields as well.

The strict identification of absorption processes with advanced waves and of emissions with retarded waves has some intuitive appeal. Given a specific temporal orientation, the retarded solution to the wave equation describes a disturbance that originates at the source at a time t_0 and travels outwards for times $t > t_0$, while the advanced solution describes a disturbance that converges into the source at times $t < t_0$ and 'disappears' at the source at time t_0. Intuitively, the former solution seems to characterize an emission process and the latter solution an absorption process. However, an absorber can be associated with a retarded field as well, as Zeh ([1999], p. 16) argues. Zeh considers the case of an incoming field that interacts with a source that in turn emits a purely retarded field, where the retarded field interferes destructively with the incoming field. Energy then flows from the field into the source, which, therefore, acts as an absorber, even though its contribution to the total field is represented in terms of a retarded field. Similarly, the emission of energy can be associated with a purely advanced field. Of course one *could* have represented the absorption process in terms of advanced fields. The point here is that this representation is not the only one possible; both emission and absorption processes can be represented in terms of either retarded or advanced waves by including appropriate free fields.

However that may be, presumably it would be enough for Price's purposes to establish the weaker claim that if an absorber is centered on a coherent wave front, then it *can* be associated with an advanced field; and that claim is true. As the example in the previous paragraph illustrates and as I explained in Section **2** above, *any* field centered on a field source can be represented either in terms of retarded or in terms of advanced fields, but this point has nothing to do with the Wheeler–Feynman theory. So the second part of (3.3) is either false or, if construed as making a claim about what representations are possible, can be established without appealing to the Wheeler–Feynman theory.

What about the first part? The first part is the claim that all sources, whether they act as emitters or absorbers of radiative energy, are centered on coherent wave fronts. Again this claim is an immediate consequence of the Maxwell equations, according to which *every* charge, be it a net absorber or a net emitter of energy, contributes a component to the total field, which is centered at the source (where that field component can be either retarded or advanced depending on the particular representation chosen). Thus, Price is right in saying that on the micro-level (3.3) is true (if correctly understood), if not for the reasons he himself thinks. (3.3) is true, but does that mean that electromagnetic radiation is symmetric on the micro-level? This is the question to which I will turn next.

3.3 In what sense is electromagnetic radiation asymmetric?

In this section I want to argue that the traditional problem of the temporal asymmetry of electromagnetic radiation involves (3.0), the claim that all accelerated charges are associated with fully retarded (rather than fully advanced) radiation fields, and that even though (3.3) is true, this does not imply that radiation is symmetric.

Here is how Price himself initially characterizes the puzzle of the temporal asymmetry in classical electrodynamics:

> Maxwell's theory clearly permits both kinds of solutions [i.e. retarded and advanced solutions], but nature appears to choose only one. In nature it seems that radiation is always retarded rather than advanced. Why should this be so? ([1996], p. 50)

We have already seen that the sense in which radiation seems always retarded in nature is that the radiation field associated with a source of radiation can be represented as being *fully* retarded, but not as being *fully* advanced. Thus, given a careful spelling out of Price's formulation of the puzzle it seems to follow that (3.0) captures the sense in which radiation is asymmetric; Price's own question seems to be the question as to why (3.0) is true.[15]

[15] Zeh characterizes the asymmetry in essentially the same way: 'Why does the Sommerfeld radiation condition [of zero incoming fields] approximately hold in most situations?' ([1989], p. 17). Since the Sommerfeld radiation condition holds exactly if a radiation field can be represented as fully retarded, this formulation is equivalent to (3.0).

Even though Price initially characterizes the radiative asymmetry in a way similar to (3.0), he later seems to argue that the fact that emissions can be represented in terms of purely retarded fields does not imply that radiation is asymmetric in any interesting sense. I take it that his banking analogy is meant to provide an argument to that effect (see *ibid*., pp. 58–61). According to Price, since deposits into a bank account are temporal inverses of withdrawals, there is nothing temporally asymmetric about banking as a whole. Even if *deposits* were in some sense temporally asymmetric, *banking as a whole* is temporally symmetric: Transactions that look like deposits in one temporal direction, turn into withdrawals if the direction of time is reversed, while withdrawals turn into deposits. Similarly, Price holds, electromagnetic absorptions can be construed as temporal inverses of emissions and, thus, radiation processes as a whole are not temporally asymmetric.

But are absorptions really temporal inverses of emissions? For this to be true, it has to be the case that we can represent absorptions in terms of fully advanced fields. A simple model of a microscopic absorption process is the absorption of radiation by a harmonically bound charge (see Jackson [1999], section 16.8). In response to an incident radiation field the charge begins to accelerate and oscillate. The field has to do work against the binding force and, thus, part of the energy of the incident field is removed from the field and converted into mechanical motion of the oscillating charge. Since the charge accelerates, it not only absorbs energy, but also radiates off energy. Therefore, the effect of a microscopic absorber is partly to absorb energy and partly to re-radiate and scatter the incident field. If such an absorption process is to be the temporal inverse of an emission process, then it has to be possible to represent any contribution to the total field due to the presence of the bound charge in terms of a fully advanced field. However, this is in general not possible. Since any microscopic absorber re-radiates energy, the field associated with the absorber has a component along the forward light cone of the charge and, therefore, cannot be represented as a fully advanced field. There are emissions without absorptions, but no absorptions without re-emissions. Thus, unlike banking transactions, radiation processes *can* be distinguished from their temporal inverses: The temporal inverse of the radiation phenomena we observe are situations in which there are no emissions without absorptions, but absorptions without re-emissions. There is an important difference, then, between electromagnetic radiation and banking; and the banking analogy cannot establish that (3.0) fails to capture what is asymmetric about electromagnetic radiation.

Price himself discusses what he takes to be the consequences of the fact that there are no absorptions without re-emissions for his argument involving the banking analogy. He imagines a banking system analogous to the electromagnetic case where withdrawals always are accompanied by re-deposits (and which he somewhat confusingly calls 'nonfrictionless banking' ([1996], p. 60)).[16] These 'impure' withdrawals, according to Price, can be described as a mixture of 'pure' withdrawals and deposits. Analogously, he suggests, in the case of electromagnetic radiation the 'complete interaction [of a re-emitting absorber] with the field comprises a mixture of advanced and retarded solutions' (*ibid*., p. 60). And this is meant to be enough to show that 'radiation is intrinsically symmetric, in the sense that the advanced solutions do actually occur in nature' (*ibid*., p. 60). However, these considerations do not show that the asymmetry of radiation is not characterized by (3.0). In fact, Price seems to concede that absorptions are not the temporal inverse of emissions with respect to the property invoked in (3.0): emissions can be represented by *fully* retarded fields, while the fields associated with absorptions need to be represented by a 'mixture of advanced and retarded solutions'.

Finally, does it follow from (3.3) that radiation is symmetric in the sense that is invoked in (3.0)? The answer is 'no'. According to (3.3), all field sources, whether they act as emitters or absorbers of radiation, can be associated with either retarded or advanced fields. But clearly this does neither imply that the contribution of a radiating source to the total field *cannot* be represented as being *fully retarded* nor does it imply that it *can* be represented as *fully advanced*. Nothing whatever about the truth of (3.0) follows from (3.3).

This concludes my criticism of Price's account of the temporal asymmetry of electromagnetic radiation. I want to summarize briefly what I have argued. Price's proposed reinterpretation of Wheeler and Feynman's absorber theory is problematic, since it is inconsistent both with the central assumption of the absorber theory and, more seriously, with the Maxwell equations. Price invokes his reinterpretation of Wheeler–Feynman to argue for the claim that both emitters of electromagnetic radiation and absorbers can be associated with coherent wave fronts. I have argued that this claim is an immediate consequence of the Maxwell equations and is not in need of support from the absorber theory or Price's reinterpretation. It does not, however, follow from the truth of this claim that electromagnetic radiation is symmetric on the micro-level. Thus, Price's attempt to show that radiative phenomena are symmetric on the micro-level does not succeed. But if radiation is asymmetric on the micro-level, then Price's account of the macro-asymmetry of radiation can provide at best a partial solution the puzzle of the arrow of radiation.

4 Zeh's appeal to asymmetric boundary conditions

Zeh ([1989], [1999]) argues that the apparent asymmetry of radiative phenomena is due to prevailing physical boundary conditions—in particular the

[16] I say 'confusingly', since Price's 'frictionless' ([1996], p. 60) case of pure withdrawals is meant to be analogous to the case of 'pure' electromagnetic absorption without re-emission, that is, the electrodynamic case of *infinite* friction.

presence of ideal absorbers in the past of spacetime regions in which we are interested—which ensure that there are no source-free incoming fields. If the incoming fields are zero, then the total field can be expressed as a purely retarded field, but not as a purely advanced field, since the outgoing fields will in such cases generally not be equal to zero. Zeh defines an absorber in terms of its thermodynamic properties (expressions in parentheses refer to the ideal case at a temperature of absolute zero): 'A spacetime region is called ''(ideally) absorbing'' if any radiation propagating in it (immediately) reaches thermodynamical equilibrium at the absorber temperature $T\,(=0)$' ([1999], p. 22). This definition implies, according to Zeh, 'that no radiation can propagate within ideal absorbers, and in particular that no radiation may *leave* the absorbing region (along forward light cones)' (*ibid*., p. 23, italics in original). Hence, Zeh argues, the free incoming fields must be zero at the absorber boundary and therefore the total field in spacetime regions that are bounded by an absorber in the past can be represented as fully retarded but not as fully advanced. Zeh maintains that laboratories prior to an experiment closely approximate an ideal absorber and also offers some cosmological reasons for why we might think that the past of the universe constitutes an ideal absorber. On Zeh's account, then, the temporal asymmetry of radiation can be derived from the thermodynamic asymmetry.

In reply to Zeh's account I want to argue three points. First, since classical electrodynamics (as it is understood today) is fundamentally a microscopic theory, Zeh's account, which relies crucially on macroscopic thermodynamic properties of the absorber, can at best provide a partial or preliminary answer to the problem of the radiative asymmetry. Second, even if Zeh's account were successful in explaining why the retardation condition is satisfied in spacetime regions that are bounded by an ideal absorber in the past, the account is too limited in its scope. If Zeh were right, classical electrodynamics would leave radically underdetermined what, given a certain configuration of charges, the electromagnetic fields are in spacetime regions not bounded by an ideal absorber in their past. Third, assuming that a material is an ideal absorber in Zeh's sense is not sufficient to fix uniquely the value of the field at the absorber boundary. In assuming that the physical boundary conditions uniquely fix the mathematical boundary conditions such that the incoming fields are zero, Zeh needs to assume implicitly what he is trying to show, namely that all fields associated with sources are fully retarded.

Zeh acknowledges explicitly that his account, based on macroscopic 'thermodynamical reasons', cannot be given a micro-physical foundation, since 'statistical reasons are insufficient for deriving the thermodynamic arrow' ([1999], p. 22). A general discussion of the relation between thermodynamics and statistical physics would take us too far afield, but one can see why statistical arguments fail in the present case. If the field associated with a charge in the spacetime region of interest were fully advanced, the field would be non-zero at the absorber boundary. But, one might try to argue, the field at the absorber boundary would consist of electromagnetic disturbances which arise coherently at different places on the surface of the absorber, and such correlated behavior is overwhelmingly improbable. Thus, one might conclude, advanced fields are not found in nature. This argument, however, is guilty of what Price calls the 'temporal double standard' fallacy: If one allows for explicit advanced effects then the coherent disturbances on the past boundary are no more improbable than correlations which exist on a future boundary in connection with a fully retarded field. In the advanced case the correlations have a common 'cause' in their future—the motion of the charge—just as future correlations in the retarded case have their common 'cause' in the past.[17]

The problem, however, with having to rely on a macroscopic argument to derive the radiative arrow is that classical electrodynamics is fundamentally a microscopic theory. Macroscopic electrodynamics is usually presented as being derivable from microscopic electrodynamics. But it seems to follow from Zeh's account that no such derivation is possible, since macroscopic fields exhibit a temporal asymmetry that cannot be derived from micro-physical considerations. Moreover, Zeh's account leaves us without an answer to the question as to what the *purely* microscopic fields associated with *purely* microscopic configurations of charges are.[18]

For my second and third points it will be useful to first look at the standard situation of a fully retarded field. A fully retarded field, as we have seen, can alternatively be represented as the sum of an advanced and a source-free outgoing field. One might ask where the source-free outgoing field can come from, if there is no source-free incoming field. The answer is that the charge (or charges) of the problem are responsible for that field. For a spacetime point on the world-line of a charge at a time t_0, the associated retarded field is zero for times $t < t_0$, while the advanced field is zero for times $t > t_0$. Thus, if a field equal to a fully retarded field is to be represented in terms of advanced fields, this representation has to include an outgoing field that represents the field after the advanced field has 'turned off.' If we take the fully retarded field to be associated with the charges and their motions, then one should likewise take the partly advanced partly free field to be associated with the charges. After all, the field is one and the same and only its mathematical representation has changed.

What is it for a field to be associated with a certain charge configuration? The notion of a source's causing a certain field might be suspicious to some,

[17] See Zeh ([1999], p. 16) and Price ([1996], pp. 56–57) for essentially the same point.

[18] At best, Zeh's account seems to allow for a hybrid treatment that uses a macroscopic characterization of the absorber as input for a calculation of the microscopic fields associated with a charge configuration in the spacetime region of interest.

but the notion of a field's being associated with a source should not be. The field component associated with a source is simply that component of the total field that would be absent, if the source were absent. If a fully retarded field is associated with a certain source, then that very field represented as advanced plus free outgoing field is also associated with the source. Thus, a particular representation of a field might be somewhat misleading in that even components of a field that do not have a source in a certain spacetime region can be associated with a source in that region. There is a certain danger in reading claims about the independent existence of fields into a particular mathematical representation of these fields. It does not follow from the fact that we can rewrite a fully retarded field associated with a source as the sum of an advanced field and a source-free field that there is a source-free field that exists independently of the source. If the source were absent, so would be the field.

Now let us imagine a region of spacetime (that is not bounded by an absorber in the past) with no charges but with an arbitrary non-zero electromagnetic free field. If the region is source-free, then the incoming field F_{in} will be equal to the outgoing field F_{out}. We can then ask how the total electromagnetic field would change if a charge were introduced into the region. From a physical standpoint, the situation appears to be completely determined: initially the total field is a source-free field (which we can assume to be known). Then a charge (with a known trajectory) is introduced. The resulting total field should be given by the sum of the source-free external field and the field associated with the charge and its motion. What, then is the field associated with the charge? The Maxwell equations alone do not allow us uniquely to determine this field, since they permit both retarded and advanced solutions (and any linear combination of the two). Which solution represents the field correctly? The correct solution, one might say, is determined by the boundary conditions. However, an appeal to boundary conditions is of no help here, because what the correct boundary conditions are is precisely what is at issue. If the field associated with the charge were fully retarded, then the field on the past boundary would be given by F_{in}, the incoming field with no charge present, while the field on the future boundary would have a contribution due to the charge in addition to that given by the field F_{out} with no charge present. If, on the other hand, the field associated with the charge were fully advanced, then the field on the future boundary would be given by F_{out}, while the field on the past boundary would have an additional contribution to that given by F_{in}. Thus, without knowing already what the field associated with a charge is we cannot know how to choose the boundary conditions. Since in the case we are imagining there is no additional physical 'stuff' present such as Zeh's absorber, which one might hope could fix the value of the field at the boundary, the only way to single out a unique solution to the equations seems to be by imposing an additional *general* constraint such as the one invoked in physics textbooks, according to which the purely retarded solution correctly represents the field associated with a charge. Without such a general constraint situations like the one we are imagining are radically underdetermined. If Zeh were right—that is, if only the presence of an ideal absorber could ensure that the fields associated with charges are fully retarded—then there would be many physically reasonable situations in which classical electrodynamics does not allow us to determine what the electromagnetic fields are.

Since it should be possible to set up situations like the one we are imagining in a laboratory (*pace* Zeh), Zeh's account should be open to empirical tests. We could experimentally compare the field in a spacetime region without sources with the field in a spacetime region that in addition to a non-zero free incoming field contains a radiating source. If the contribution to the total field due to the source were fully retarded—in particular, if the field at the past boundary of the spacetime region containing the charge were equal in value to the field at the boundary of the source-free region—then the radiative asymmetry could not merely be due to the presence of an absorber, simply because we set things up in a such a way that the relevant spacetime regions are not bounded by an absorber in the past and the fields at the past boundaries are non-zero.

So far I have argued that even if Zeh could successfully show that it follows from what it is for a spacetime region to be an absorber that all fields in the future of the absorber are fully retarded, his account does not offer a complete solution to the puzzle of the arrow of radiation. Now I want to argue that the presence of an ideal absorber in the past of a given spacetime region is not sufficient to ensure that the Sommerfeld radiation condition of zero incoming fields is satisfied. Let us assume that, contrary to what Zeh says, the total field in a region in the future of an ideal absorber is fully advanced and that this fully advanced field is associated with the charges and their motions in that region. In other words, we are assuming that if there are no charges, then the region will be field-free; but if there are charges, then the total field in the region will be fully advanced. If this assumption is compatible with the presence of an ideal absorber, then Zeh's explanation for why the Sommerfeld radiation condition is satisfied is unsuccessful.

It might appear that a fully advanced field is incompatible with the presence of an ideal absorber, because a retarded representation of a fully advanced field would have to include a non-zero incoming field at the boundary of the absorber. And since no field can propagate through the absorber in the direction of the forward light cone, the incoming field, it seems, has to be zero. Thus, one might think that a non-zero field at the boundary conflicts with Zeh's absorber assumption. This line of thought, however, presupposes that any free incoming field would have to propagate through the absorber and could not be a field that arises at the boundary of the absorber due to charges in the future of that boundary. This presupposition is not warranted. In the case of a fully retarded field one has to posit a free outgoing field that is associated with the relevant charges, if one switches to an advanced field representation. Analogously, the free incoming field which one has to introduce, if one wants to represent a fully advanced field in terms of retarded fields (and which represents the field at spacetime points before the retarded fields turn on), is associated with the charges. Thus, if we assume that the fields associated with charges are fully advanced, then a non-zero field at the absorber boundary would be due to the charges inside the spacetime region of interest and would not have to propagate through the absorber. Zeh's definition of an ideal absorber only rules out that any non-zero field at the boundary of the absorber could propagate through the absorber, but does not disallow that such fields could be the result of advanced effects associated with a charge in the spacetime region to the future of the absorber.[19]

This situation is analogous to that of an ideal absorber in the future of a spacetime region with fully retarded fields. Since the presence of the absorber should not affect the retarded field in spacetime regions in the past of the absorber, it should not affect the free outgoing field that is part of the advanced representation of that field. Thus, a free outgoing field is compatible with a future absorber. Once the field reaches the absorber (that is, once the absorber is 'turned on'), this field will be damped out immediately, but the field will be non-zero directly at the spacetime boundary of the absorber. In both the case of a future absorber and that of a past absorber, then, non-zero fields at the absorber boundary do not propagate through the absorber but are associated with charges within the spacetime region of interest. Thus, the situation Zeh imagines—the physical 'boundary condition' of an ideal absorber—is not enough to fix the mathematical boundary condition of zero incoming fields. Without already assuming that there are no explicit advanced effects Zeh cannot ensure that the Sommerfeld radiation condition holds.

5 The textbook account

My main aim in this paper is to argue that something like the textbook account provides the best account we have for characterizing the temporal asymmetry of electromagnetic radiation. So far, my argument has been purely negative: neither Price's nor Zeh's account presents a viable alternative. In this last section I want to try to dispel some of the dissatisfaction one might feel with the

[19] One might think that non-zero fields at the absorber boundary are ruled out because that would require the field to be discontinuous across the boundary. But as far as I can tell the field would not have to be continuous across the boundary, for the following two reasons. First, at the absorber boundary the physical properties of the medium change abruptly: a region of space is absorbing in the past of the boundary and transparent in its future. Second, there are charges present at the absorber boundary, since an absorber operates by converting radiation into mechanical energy associated with charged particles.

textbook account. Compared with Price's or Zeh's attempts to associate the radiative asymmetry with the thermodynamic asymmetry, the account that I am calling 'the textbook account' is rather thin. However, I take this to be one of the account's virtues, since there are reasons for thinking that the purported puzzle of the asymmetry of electromagnetic radiation should not be understood as presenting a genuine puzzle at all. While the textbook account is not much of an account, this should not worry us, since there is not much of a puzzle in need of a solution.

What exactly is the textbook account? According to the textbook account, electromagnetic fields have to meet an additional general constraint besides those imposed by the Maxwell equations. The account says that not all solutions to the Maxwell equations but only those that satisfy the retardation condition, according to which electromagnetic fields associated with a charge Q propagate along the future light cone of Q, can represent physically possible situations. In physics textbooks the retardation condition is sometimes presented as a causal constraint (even though the term 'causal' is occasionally put in scare quotes). So one might understand by 'the textbook account' an account that justifies or explains the retardation condition by appealing to a principle of causality and a temporal asymmetry that is supposed to be implied by that principle. This is not, however, the kind of account I want to advocate here. Rather, the account I wish to advocate simply stipulates that, in addition to the Maxwell equations, electromagnetic fields associated with electric charges satisfy the retardation condition *without offering any explanation as to why this condition should hold*.[20] Of course, anyone who thinks that the puzzle of the asymmetry of radiation presents a genuine puzzle will not be satisfied with an 'account' that offers little more than the statement that the asymmetry does in fact hold. Thus, in what follows I want to provide some reasons for thinking that there is indeed no real puzzle to be solved.

Both Price and Zeh put the puzzle concerning radiative phenomena this way: the underlying laws of nature that govern radiative phenomena (that is, the Maxwell equations) are time reversal invariant, yet in nature the fields associated with charges always appear to be fully retarded. The puzzle then is this: why do not all of the solutions to the Maxwell equations represent actually occurring physical phenomena?[21]

Now, why do we think that this is a genuine puzzle in need of a solution? Why does the fact that there is an additional constraint on models of electromagnetic

[20] I chose the name 'textbook account' as a convenient label for the position I am advocating here. The position is suggested by what textbook authors, such as Jackson, say when they introduce the retardation condition, but I do not wish to imply that the position is in fact the position held by any particular author.

[21] Here is how Price introduces the puzzle: 'If ingoing and outcoming waves are equally compatible with the underlying laws of physics, why does nature show such a preference for one case rather than the other? This is the puzzle of the arrow of radiation' ([1996], p. 50).

phenomena, beyond those given by the Maxwell equations alone, call for an explanation? We might compare the puzzle of the asymmetry of radiation with the following request for an explanation: why is it that the fields associated with radiating charges not only are fully retarded but also satisfy the additional constraints given by the Maxwell equations? We would, I believe, simply reject this request for an explanation. The fact that electromagnetic fields satisfy the retardation condition, but not the fact that they satisfy the Maxwell equations, seems to call for an explanation, because the Maxwell equations (unlike the retardation condition) are afforded a special role within classical electrodynamics: they are understood to be the fundamental equations or laws of the theory and as such are taken to delineate the range of what, within the domain of classical electrodynamics, is physically possible. What (most fundamentally) is physically possible need not, or perhaps cannot, be explained. But when the range of actual phenomena encompasses less than what is physically possible, a puzzle arises: if both retarded and advanced fields are physically possible, why is it that all actual radiative phenomena involve only retarded fields? A satisfactory answer to this question would have to invoke, it would seem, some kind of mechanism or physical process that restricts actual radiation fields to retarded fields and, thus, the textbook account, which does not provide such a mechanism or process, seems inadequate.

The apparent puzzle to which the textbook account provides no satisfactory answer arises, if one thinks of the Maxwell equations on their own as delineating what is physically possible. That is, the puzzle arises if one accepts the conjunction of the following two claims:

(a) The Maxwell equations alone are the fundamental laws of classical electrodynamics.
(b) The fundamental laws of a theory delimit what is physically possible within the domain of that theory.

Claim (a) is a claim about the role of the Maxwell equations within classical electrodynamics and their relationship to other parts of the theory; claim (b) concerns more general philosophical commitments. If one rejects either (a) or (b), the problem posed by Price and Zeh need not arise.

What else besides the Maxwell equations, one might ask, should be a law of classical electrodynamics? The answer I want to propose is this: the retardation condition. If the retardation condition were a law just like the Maxwell equations, then the fact that radiative phenomena satisfy the condition would be no more in need of an explanation than the fact that they satisfy the Maxwell equations. We do not find it puzzling that Gauss's law alone does not determine the range of what is electrodynamically possible; the fact that there are four Maxwell equations (in a standard formulation), and not merely one, is not something we think is in need of an explanation. Similarly, the question Price and Zeh are puzzled by ought to present no real puzzle once we understand the retardation condition as having the same status in the theory as the Maxwell equations.

Are there convincing reasons against taking the retardation condition to be a law of classical electrodynamics? All actual electromagnetic phenomena seem to satisfy both the retardation condition and the Maxwell equations. Both need to be invoked in derivations of mathematical representations of radiative phenomena; both play a role in unifying electromagnetic phenomena. The retardation condition is extremely powerful in modelling electromagnetic phenomena, and it is simple—much simpler, in fact, than the Maxwell equations. So why should we take only the Maxwell equations to be law-like? To say that the Maxwell equations do not imply the retardation condition would beg the question. Of course, not all models of the Maxwell equations satisfy the retardation condition, but neither do all models of the retardation condition satisfy the Maxwell equations. It is not clear, then, on what grounds—except for a prior commitment to the Maxwell equations as the only laws of classical electrodynamics—one should take a violation of the retardation condition, but not of the Maxwell equations, to be electrodynamically possible and, hence, take the fact that the condition is obeyed to be in need of an explanation. Of course, the formulation of the Maxwell equations constitutes a far greater scientific achievement than that of the retardation condition, but this fact alone does not seem to provide a good enough reason to grant the status of a law only to the former and not to the latter.

There is no general, context-independent answer to the question whether something is in need of an explanation. Thus, I do not wish to suggest that there could be no context within which the radiative asymmetry does indeed present an interesting puzzle and that perhaps in such a context there could not be a fruitful explanation of the retardation condition. Moreover, I do not believe that just because a relation is law-like, it cannot or need not be explained. There are many contexts within which we can give explanations of particular laws. If, however, the retardation condition is law-like, then merely to point out that the condition is not implied by the Maxwell equations is not enough to create a genuine puzzle in need of an explanation; just as it is not enough to point out that three of the Maxwell equations do not imply the fourth in order to create a puzzle as to why that fourth equation holds.

The puzzle of the arrow of radiation arises if we think that there is a class of phenomena—those involving charges associated with fully advanced fields—that are physically possible but not actual. I have argued that there are no good reasons for thinking that advanced field phenomena are in fact physically possible. In the context of discussing whether the increase of entropy needs to be explained Price says: 'Roughly, things are more in need of an explanation the more they depart from their natural condition' ([1996], p. 39). One might also, then, put the question I am asking this way: why should we think that the fact that all radiative phenomena in nature satisfy the Maxwell equations is any more natural than the fact that they satisfy the retardation condition?

The puzzle of the radiative asymmetry derives its urgency partly from granting the Maxwell equations a status different from that of the retardation condition, but the puzzle also depends on much broader issues in the philosophy of science which I can only mention briefly here. For someone who does not believe that the laws or fundamental equations of a theory delimit what is physically possible—that is, for someone who rejects (b) above—the puzzle need not arise, even if the Maxwell equations alone were to be understood as the laws of electrodynamics. The general form of the putative puzzle is: why does only a subset of the possible solutions to a set of fundamental equations provide us with models of actually occurring phenomena? This kind of question represents a genuine puzzle only if we take it that *all* of the solutions of a theory's basic equations represent genuine physical possibilities. If put in this general way, however, this idea might strike one as quite unreasonable. It is not uncommon that the basic equations of a theory have more solutions than those that represent physically possible situations. For example, in solving quadratic equations we often discard solutions given by the negative square root as unphysical without puzzling why 'nature chooses' only the positive square root. The particular mathematical form that a theory's laws or basic equations take seems to be guided not only by the aim of arriving at equations that represent all and only physically possible situations in a given domain. At times a simpler set of equations, or one that unifies a wider range of phenomena, might be preferable, even if that set also has non-physical solutions that a more cumbersome formulation might be able to avoid.[22]

Whether or not one accepts claim (b) will depend on one's views on the nature of scientific theories. Many realists will probably believe that the fundamental equations of a theory delimit what is physically possible, while instrumentalists are likely to deny this. A detailed discussion of these issues would take us too far afield. Here I can do no more than point out that the problem of the arrow of radiation to some extent depends on certain general (and controversial!) philosophical assumptions that are usually not made explicit.

To sum up, Huw Price's and Dieter Zeh's puzzle is: why is it that, given that the Maxwell equations are time symmetric, all actual radiative phenomena satisfy the retardation condition? Both take it that advanced effects are physically possible and, thus, would occur if the physical 'stuff' were arranged

[22] There are also reasons specific to classical electrodynamics for rejecting the idea that the Maxwell equations in particular could be understood as delineating coherent electrodynamically possible worlds, since it is doubtful whether there is a consistent formulation of classical electrodynamics with a satisfactory equation of motion for microscopic charged particles.

differently in the actual world. Both then argue that the actual boundary or initial conditions are responsible for the absence of advanced effects (at least on the macro-level, in Price's case). I have tried to show that neither Price's nor Zeh's arguments are successful. The response I want to advocate in place of Price's and Zeh's solutions to the puzzle is that there is nothing surprising about the fact that not all the solutions to the Maxwell equations represent the phenomena. Thus, I propose the textbook account not as an answer to their puzzle but rather as a way to dissolve the puzzle. The Maxwell equations do not on their own delineate what is physically possible. I say this partly because I do not believe that it is the role of laws to delimit the range of what is physically possible, but more importantly because there is no obvious reason why the retardation condition should have a status in classical electrodynamics different from that of the Maxwell equations. Only models that satisfy both the constraints given by the Maxwell equations and those given by the retardation condition (and perhaps other, more 'local' constraints) represent physically possible phenomena. But if we do not think that advanced effects are physically possible, we no longer need a rich explanation that can show why such effects do not actually occur. Thus, the textbook account is the best solution to a puzzle that should not be viewed as much of a puzzle at all.

Acknowledgements

I would like to thank Craig Callender, Stephen Leeds, Huw Price, Steve Weinstein, Dieter Zeh, the members of the Chicago area philosophy of physics reading group, and the referees for this journal for extremely helpful and often very detailed comments and criticisms.

Department of Philosophy
Northwestern University
Evanston, IL 60208–1315
USA
m-frisch@nwu.edu

References

Davies, P. C. W. [1974]: *The Physics of Time Asymmetry*, Berkeley: University of California Press.

Dirac, P. A. M. [1938]: 'Classical Theories of Radiating Electrons', *Proc. Royal Soc.*, **A167**, pp. 148–68.

Jackson, J. D. [1999]: *Classical Electrodynamics*, New York: Wiley.

Leeds, S. [1994]: 'Price on the Wheeler–Feynman Theory', *British Journal for the Philosophy of Science*, **45**, pp. 288–94.

Leeds, S. [1995]: 'Wheeler–Feynman Again: A Reply to Price', *British Journal for the Philosophy of Science*, **46**, pp. 381–83.

Price, H. [1991a]: 'The Asymmetry of Radiation: Reinterpreting the Wheeler–Feynman Argument', *Foundations of Physics*, **21**, 8, pp. 959–75.

Price, H. [1991b]: 'Review of Denbigh, K. and Denbigh, J., *Entropy in Relation to Incomplete Knowledge*, and Zeh, H. D., *The Physical Basis of the Direction of Time*', *British Journal for the Philosophy of Science*, **42**, pp. 111–44.

Price, H. [1994]: 'Reinterpreting the Wheeler–Feynman Absorber Theory: Reply to Leeds', *British Journal for the Philosophy of Science*, **45**, pp. 1023–28.

Price, H. [1996]: *Time's Arrow & Archimedes' Point*, Oxford: Oxford University Press.

Ridderbos, T. M. [1997]: 'The Wheeler–Feynman Absorber Theory: A Reinterpretation?', *Foundations of Physics Letters*, **10**, 5, pp. 473–86.

Rohrlich, F. [1965]: Classical Charged Particles: Foundations of Their Theory, Reading, MA: Addison-Wesley.

Wheeler, J. A. and Feynman, R. P. [1945]: 'Interaction with the Absorber as the Mechanism of Radiation', *Reviews of Modern Physics*, **17**, pp. 157–81.

Zeh, H. D. [1989]: *The Physical Basis of The Direction of Time*, Berlin: Springer-Verlag.

Zeh, H. D. [1999]: *The Physical Basis of The Direction of Time*, 3rd edn, Berlin: Springer-Verlag.