Studies in History and Philosophy of Modern Physics 37 (2006) 542–558

Studies in History and Philosophy of Modern Physics

www.elsevier.com/locate/shpsb

A tale of two arrows ☆

Mathias Frisch

Skinner Building, University of Maryland, College Park, MD 20742, USA

Abstract

In this paper I propose a reasonably sharp formulation of the temporal asymmetry of radiation. I criticize accounts that propose to derive the asymmetry from a low-entropy assumption characterizing the state of the early universe and argue that these accounts fail, since they presuppose the very asymmetry they are intended to derive.

Keywords: Radiation; Classical electrodynamics; Thermodynamics; Past-hypothesis; Entropy; Time asymmetry

1. Introduction

When electric charges accelerate coherently, for example in an antenna, we observe a radiation field coherently diverging from the source. The time-reversed phenomenon—that is, radiation waves coherently converging into an accelerating source—is not something we observe. How can we explain this asymmetry? The fundamental equations governing radiation phenomena, the Maxwell equations and the wave equation that can be derived from them, are time symmetric. Why, then, are there coherently diverging waves in nature but no coherently converging waves? One persisting idea has been that the temporal

☆ This paper grew out of many and long conversations on the nature of the arrow of radiation between Huw Price, David Atkinson and me. Our original plan was to write a joint paper on this issue. In the end, too many points of disagreement remained to make writing a joint paper feasible. Nevertheless, our discussions significantly affected my own thinking on the issue, and I want to thank both for our discussions. I want to thank Huw Price especially for inviting me to visit the Centre of Time at Sydney University and for his hospitality during my stay. I also want to thank two anonymous referees for this journal for their very detailed comments and criticisms on an earlier version of this paper and Jill North for a helpful correspondence concerning her entropy account of radiation.

E-mail address: mfrisch@umd.edu.

1355-2198/$ - see front matter
doi:10.1016/j.shpsb.2005.03.004

'arrow' of radiation can be explained by appealing to another temporal asymmetry: the entropy asymmetry exhibited by thermodynamic systems. Often those who argue that such an explanation is possible see that explanation as part of a bigger project that reduces various temporal asymmetries to a single 'master arrow' providing a unified account of all asymmetries *in* time and perhaps even of the asymmetry *of* time.

Most accounts that see the radiation asymmetry as arising from thermodynamic or statistical considerations are descendents of John Wheeler and Richard Feynman's infinite absorber theory of radiation (Wheeler & Feynman, 1945). Radiation itself, according to these accounts, is inherently time symmetric, yet the presence of absorbing media introduces an additional constraint that results in a time asymmetry in the fields. All the accounts in this tradition of which I am aware fail—often by committing what Huw Price has called "the temporal double standard fallacy" by illegitimately introducing time-asymmetric considerations (see Price, 1996). I criticize absorber accounts in detail in (Frisch, 2005) and will not further discuss them here. Yet in recent years several philosophers of science have argued that there is a much simpler and much more direct thermodynamic argument for the asymmetry of entropy (see Arntzenius, 1994; Callender, 2001; North, 2003), and it is this type of argument that I want to evaluate critically here.

I will proceed as follows. In the next Section I will offer a more precise formulation of the radiation asymmetry, emphasizing that the asymmetry characterizing total fields is an asymmetry between prevailing temporal boundary conditions. This claim in itself is not novel and ought not to be controversial. Yet while it is generally accepted in the literature on classical electrodynamics that the asymmetry characterizing total fields consists in an asymmetry between initial and final conditions, there are certain difficulties with standard formulations of the asymmetry for which I will suggest a solution. In Section 3, I will provide a brief sketch of a standard account of the thermodynamic asymmetry. In Section 4, I will present what one might call "the simple entropy account" of radiation and will argue that it fails. I end with a brief conclusion.

2. The asymmetry of radiation

I said above that there are coherently diverging but not coherently converging waves in nature, and this is in fact how the asymmetry of radiation is often characterized. In one important sense this claim is false, since it is a mathematical fact that *every* radiation field can both be represented as involving diverging waves and as involving converging waves. Radiation fields are governed by the wave equation that can be derived from the Maxwell equations. Commonly this equation is solved by setting up a modified initial value problem. The total field in a given region of spacetime is given by the fields on some initial value surface in the past together with the contribution of the field sources in that region with known trajectories.[1] In this case the contributions of the sources are diverging or *retarded* fields. That is, in an initial value problem the total field is represented as a combination of source-free incoming fields and retarded fields.

[1] The initial value problem is 'modified' since one treats the trajectories of the sources as independently given and does not also set up an initial value problem for the sources. The reason for this is that it is not clear that there is a well-defined initial value problem for systems of fields and particle-like sources. I discuss some of the problems faced in trying to arrive at a satisfactory classical particle-field theory in Frisch (2005).

But equally the total field can be represented in terms of a final value problem. In that case the contributions of the sources appear as converging or *advanced* fields.[2] The same total field, thus, can also be represented as a combination of source-free outgoing and advanced fields:

$$F_{\text{total}} = F_{\text{ret}} + F_{\text{in}} = F_{\text{adv}} + F_{\text{out}}. \tag{1}$$

Both representations are representations of one and the same field. Every radiation field can be represented either as a combination of *retarded* and source-free *incoming* fields or as a combination of *advanced* and source-free *outgoing* fields. What is more, the field can even be represented as a linear combination of retarded and advanced fields together with appropriate source-free fields.

How a given total field is carved up into a component field associated with the sources present and a source-free field, depends on the particular representation chosen; there is no unique way to carve up the total field. From a formal standpoint, neither a purely retarded nor a purely advanced field representation appears privileged. The difference in representations is solely due to whether we are choosing to represent the field in terms of a modified initial or final value problem. If we choose an initial value problem, then any fields at the initial time *before* the sources turn on appear as source-free fields and the sources contribute retarded fields *after* the sources turn on. If we choose a final value problem, then any fields at the final time *after* the sources 'turn on' appear as source-free fields and the sources contribute advanced fields *before* the sources turn on. Just as there is no unique field that is formally associated with the sources in a given problem, there is no unique source-free field. Just as the question as to what component of the total field is mathematically associated with the field sources depends on our choice of initial or final value problem, so does the question as to what the source-free (or 'background') field is. There is no more *the* source-free field as there is *the* field mathematically associated with a given configuration of sources.

It is worth reiterating the central point: It is tempting to ask whether, for a given total field, the sources are formally associated with retarded or with advanced radiation; yet this question simply is not well-defined. The fields associated with sources are only retarded or advanced *relative to a particular representation* of the field. Without specifying a particular representation, the question as to whether sources are associated with retarded or advanced radiation has no answer, but—and this is important as well—once we are given the representation, there is nothing else we need to know in order to determine whether fields are retarded or advanced: If we represent the total field in terms of an initial value problem, then sources 'contribute' retarded fields; and if we represent the field in terms of a final value problem, then sources 'contribute' advanced radiation.

In what sense, then, is radiation asymmetric? The answer that is usually given is that the asymmetry consists in the fact that the free incoming but not the free outgoing fields are approximately equal to zero. If incoming fields are equal to zero, then the total field can be represented as a fully retarded field; while if outgoing fields are appreciably different from zero, the total field cannot, of course, be represented as being fully advanced. Moreover,

[2]That a final value problem involves converging waves can most easily seem by imagining a spacetime diagram of a diverging wave in an initial value problem and then flipping the diagram on its head (which amounts to reversing the direction of time). The initial value surface then becomes a final value surface and the diverging wave turns into a converging wave.

the free fields are governed by the homogenous wave equation and will be identically equal to zero in a region of spacetime exactly if the fields and their normal derivatives are zero on an appropriate initial or final value surface. This leads H. Dieter Zeh to express the puzzle of the arrow of radiation as follows: "Why does the Sommerfeld radiation condition $F^{\mu}_{\text{in}} = 0$ (in contrast to $F^{\mu}_{\text{out}} = 0$) approximately apply in most situations?" (Zeh, 2001, p. 21) The Sommerfeld radiation condition is a temporal boundary condition—an initial value condition. Thus the asymmetry of the total radiation fields can be expressed as an asymmetry concerning prevailing temporal boundary conditions and, hence, as an asymmetry between *instantaneous* states of the field: The fields (and their derivatives) on a spacelike hypersurface functioning as an initial value surface are approximately equal to zero, while the fields on a final value surface are generally not equal to zero.

Zeh's formulation of the asymmetry makes clear that the asymmetry in the total fields amounts to one between prevailing temporal boundary conditions. Yet, as it stands this formulation cannot be entirely correct, since *every* initial value surface can, in principle, also be used as a final value surface for representing the fields in its past. Thus it is not clear in what sense the condition $F^{\mu}_{\text{in}} = 0$ applies in more situations than the condition $F^{\mu}_{\text{out}} = 0$. Since every hyperplane that can be used as an initial value surface can be used as a final value surface as well, there seem to be just as many purely advanced field representations as there are purely retarded field representations. Therefore, Zeh's condition needs to be made more precise. To capture fully an asymmetry characterizing the total radiation fields, we also need to include a reference to the field sources involved—a point that in fact is implicit in Zeh's discussion of the asymmetry, even if not in his 'official' characterization of it. Intuitively, the recent pasts of regions with outgoing fields approximately equal to zero contain either no sources or a very large number of sources that do not radiate coherently and that act as absorbers of any fields in their pasts. What we do not find is that the outgoing field is approximately equal to zero in a region with a small number of radiating sources or with a large number of coherently radiating sources in its recent past. By contrast, the temporal inverse of such situations—approximately zero incoming fields in the past of a small number of radiating sources or of a large coherent source—is something with which we are well familiar.

A free field is approximately zero if it is much smaller than the retarded or advanced fields associated with the field sources in a given representation. Thus, the difference between the retarded and advanced representations is that the free fields in an initial value representation are often negligibly small, while the free fields in a final value representation are of a magnitude roughly equal to that of the advanced fields.

One might worry that the claim that there are strictly *no* purely advanced fields in the presence of coherently radiating sources is false. For should it not be possible to set such a situation up, even if this might be difficult? And if this is possible, perhaps someone has in fact done so. But allowing for this possibility does not remove the asymmetry. If there indeed are fully converging fields, they clearly are much more rare than the diverging fields with which we are familiar. One might also worry that there also are many situations where the free incoming fields are not equal to zero, even in the presence of a single coherent source (for example when I turn on the light in a room on a sunny day). This last worry can be addressed by not demanding that *every* field representation be fully retarded, but by expressing the asymmetry in terms of an existential claim: there are many situations where the field can be represented as fully retarded, while the temporal inverse of such situations is much, much more rare. Putting all this together, I want to propose the following as a reasonably precise formulation of an asymmetry characterizing total radiation fields in the presence of sources: (R)

> There are many situations in which an initial value surface with $F_{in} = 0$ can be chosen and the total field can be represented as being approximately equal to the sum of the retarded fields associated with a small number of field sources or with a large coherent source, and there are almost no situations in which a final value surface with $F_{out} = 0$ can be chosen and the total field can be represented as being approximately equal to the sum of the advanced fields associated with a small number of sources or with a large coherent source.

Of course, as we have seen above, each total field in the presence of sources can be represented both as involving retarded fields and as involving advanced fields associated with the field sources. But, according to (R) there is an asymmetry between the two types of representations: In the former case the free fields will be approximately equal to zero, while in the latter case they will not be approximately equal to zero. That is, in Eq. (1) the total field is approximately equal to the retarded fields associated with the field sources, while the total field is not approximately equal to the sum of the advanced fields associated with the field sources.

One aspect of the asymmetry is especially important for our purposes here. The asymmetry is an asymmetry between initial and final value representations and not an asymmetry in the dynamical evolutions of electromagnetic systems. I began by examining the question 'Why are there coherently diverging but not coherently converging waves in nature?' If this question is meant to capture the asymmetry of the *total* observable waves or radiation fields that we observe in nature, then it ought to be replaced by the more carefully formulated question 'Why can the *total* fields in the presence of a small number of sources be represented as approximately fully retarded but not as approximately fully advanced?' But that a field is approximately fully retarded is strictly equivalent to the free incoming field's being approximately equal to zero. That the incoming field is zero is simply *what it means* for the total field to be fully retarded.[3]

(R) captures an asymmetry of the observed *total* fields in the presence of field sources. And it captures the asymmetry in a 'metaphysically austere' way. That is, the condition provides us with a way of capturing an asymmetry between mathematically equivalent representations of one and the same total field that does not invoke notions such as that of sources *producing* a field, which some might find objectionable. This, I take it, is an advantage of the condition. The asymmetry captured by (R) provides a starting point on which everyone interested in the puzzle of the arrow of radiation should be able to agree

[3]Is it really correct to claim that the fact that the incoming field is approximately equal to zero is simply what it means for the total field to be fully retarded? Are there not worlds, one might object, in which the assumption that the source-free field is small results in sources being associated with advanced waves? And does not this show that my claim that a total field's being fully retarded is equivalent to the incoming source-free field's being zero is false? But this objection confuses the claim that there is a negligibly small source-free field on *some* representation of the total field with the claim that the free field is negligibly small *in an initial value representation*. And it is the latter claim with which I am concerned with here. Simple inspection of Eq. (1) shows that, in the case of an initial value representation, the total field is approximately fully retarded *exactly if* the source-free incoming field is approximately equal to zero. Since $F_{total} = F_{ret} + F_{in}$, $F_{total} = F_{ret}$ exactly if $F_{in} = 0$. Of course, there are possible 'worlds' in which the field is zero on some final value surface and hence in which the total field is fully advanced. This fact does nothing to impugn the claim I am making here—that the total field is approximately fully retarded *if and only if* the source-free *incoming* field is approximately equal to zero.

and any potential solution to the puzzle should at least be able to account for this asymmetry. It is worth repeating that (R) is not meant to offer a radically new way of thinking about the asymmetry of radiation. I understand the condition to be nothing more but a somewhat more precise characterization of the asymmetry as it is usually expressed in the literature, where it is stated as the claim that the initial fields in an initial value representation of the total field can be chosen to be zero but the final fields in a final value representation cannot.[4]

Yet one might think that my formulation of the asymmetry also leaves something out and that there are asymmetries going beyond those captured by (R). There are two different types of worries one might have. First, one might argue that the asymmetry exhibited by *total* radiation fields goes beyond the types of situations included in (R). As I said above, there are clearly cases where incoming fields are appreciably different from zero in the presence of a coherently radiating source from which diverging radiation appears to emanating. By contrast, the time-reverse of these situations is not something we observe. But we can grant this point, and still maintain that (R) captures an important *subset* of those situations in which total radiation fields exhibit a temporal asymmetry. Hence any explanation of the asymmetry characterizing total fields at the very least has to be able to account for the asymmetry captured by (R). And as we will see below, the simple entropy account falls short in this respect. Thus, for my purposes here it is unimportant whether total fields exhibit an asymmetry also in cases where incoming fields are appreciably different from zero. The asymmetry is, as it were, most cleanly exhibited in situations where at minimum has to account for the asymmetry in these cases.

The second worry one might have about my discussion so far is that I have not discussed the intuitive asymmetry that field sources *produce* retarded rather than advanced fields. I focused on an asymmetry exhibited by the *total* radiation fields and have not said anything about a putative asymmetry concerning the fields *produced by* or *physically associated with* a source. One might think that we ought to distinguish the question as to whether total radiation fields in the presence of sources are asymmetric from that of whether the field produced by an individual source is asymmetric. While (R) may capture at least an important aspect of the former asymmetry, it does not address the latter.

But what does it mean for a source to produce a field? Some, like Zeh, maintain that there is no legitimate sense to be made of the notion of the field produced by a source independently of a given initial or final value problem. As Zeh argues, "in field theory, no (part of the) field 'belongs to' a certain source [...]. Rather, sources in a bounded spacetime volume determine only the *difference* between outgoing and incoming fields." (Zeh, 2001, p. 20) On Zeh's view, we happen to employ causal talk, including the notions of production or creation, when we consider initial value representations of the total field. That is, what we mean by the field produced by a source is the field associated with the source in an initial value representation of the total field. Yet this does not, on Zeh's view, point to a deep physical or even metaphysical difference between initial and final value representations. Physically, initial and final value representations are on a par, and there is no asymmetry of production beyond that of the total fields.

[4]All attempts in the literature to solve the puzzle of the arrow of radiation with which I am aware take as at least one central *explanandum* the asymmetry of the total fields that is captured in Zeh's formulation above and, I hope, more precisely, in my condition (R). I am here interested in asking whether a 'simple' entropy account can adequately explain this asymmetry of the total fields.

By contrast, I argue in Frisch (2005) that there is a notion of the field physically associated with or produced by a source that is independent of any given mathematical representation of the total field. I argue that this is a causal or counterfactual notion that plays a crucial explanatory role in accounting for the asymmetry of the total field expressed by (R). The total fields exhibit the asymmetry captured by (R), because sources produce retarded rather than advanced radiation. On my account, then, what it is for a source to produce a retarded field ought to be spelled out causally: Each individual source causes retarded field excitations. While (R) concerns an asymmetry of the *total* fields, the notion of a source producing retarded radiation postulates an asymmetry of the field *associated with* a charge. The purpose of the present paper, is not, however, to defend this notion or its adequacy in accounting for the asymmetry in the total fields.

A somewhat different sense of production is suggested in North (2003). According to North's view, what it means for sources to *produce* retarded radiation is for the *total* field to be fully retarded relative to the most natural background field. If I understand North's proposal correctly, then she proposes that we find the field produced by a charge by comparing all possible ways of carving up a total field into a field associated with sources and a source-free field and then picking that representations that contains the 'most natural' source-free field. Sources are taken to produce whatever field is mathematically associated with them in this particular representation. If there is a unique-most natural background field, then the field produced by the sources is whatever field needs to be added to this background field to obtain the total field.

On this view the claim that sources produce retarded radiation turns out to be equivalent to the asymmetry expressed by (R). If we assume that a field approximately equal to zero is the most natural background field, then it follows from (R) and the definition of production that sources *produce* retarded radiation. That is, (R) implies that sources produce retarded radiation. Conversely, if sources produce retarded radiation, then it follows from the definition of production that the field associated with the sources is a retarded field relative to the most natural background field. But the field formally associated with the sources is retarded exactly in an initial value representation. Hence it follows that the most 'natural' fields are the source-free incoming fields in an initial value representation, and thus, since the most natural background field is one that is approximately equal to zero, the source-free fields in an initial value representation are approximately equal to zero. That is, the claim that sources produce retarded radiation, according to North's proposal, implies (R).

The moral of this discussion is that we have to be extremely careful in how we use talk of the field associated with a source or of sources producing a field. In the literature on classical electrodynamics and the radiation asymmetry, we can find many examples where talk of the field produced by a source is not carefully distinguished from discussions of the equivalence of different representations of the total field. Thus, J. D. Jackson in his influential book on classical electrodynamics stresses that the choice between retarded and advanced representations of the total field depends on the choice of '*boundary conditions in time* that specify the physical problem' (Jackson, 1975, p. 224, italics in original) and, hence, that the asymmetry characterizing the total fields is an asymmetry between initial and final condition. Yet Jackson also says that the retarded representation exhibits the causal behavior associated with a wave disturbance. (p. 225) Similarly, Fritz Rohrlich refers to the fields 'produced' by a charge as reason for singling out the retarded representation, while he also, of course, notes that this representation of the total fields is not unique and depends on our choice of initial- or final-value surface (Rohrlich, 1990, 77ff.).

Given the fact that different representations of the total field are possible and that there is no unique way mathematically to associate a given source distribution with a field, one may have legitimate doubts about the appropriateness of any notion of *the* field produced by a source. By contrast one may hold, as I argue in Frisch (2005), that there is also a genuine sense in which sources individually *cause*, *contribute*, or *produce* retarded fields. And in fact, I believe that such a view fits the discussions in both Jackson (1975) and Rohrlich (1990) rather well. Yet the aim of the present paper is not to defend the need, in explaining the asymmetry of the total fields, for invoking a rich causal notion of a source being physically associated with retarded radiation, but rather to discuss a rival explanation of this asymmetry. Whether or not one thinks that there is a legitimate role for the notion of a source producing a field, total radiation fields exhibit the asymmetry expressed by (R), and any solution to the puzzle of the arrow of radiation ought to account for this asymmetry.

3. Thermodynamic preliminaries

At the core of the account I want to discuss here lies the suggestion that both the thermodynamic asymmetry and the arrow of radiation can ultimately be accounted for by appealing to the hypothesis that the early universe was in a state of extremely low entropy. Here, then, is a brief sketch of how assumptions about the state of the early universe enter into a Boltzmannian account of the thermodynamic asymmetry (see Albert, 2000 for what is perhaps the most emphatic recent defense of this account).

The second law of thermodynamics, according to one formulation, states that the entropy of a closed system never decreases. The temporal asymmetry in this case is an asymmetry characterizing macroscopic systems, consisting of many particles. As a concrete example we can think of the behavior of a gas that is initially confined to one half of a container by means of a partition. If the partition is removed, the gas will spread to fill the entire container. The time-reversed phenomenon of the gas 'spontaneously' contracting into only part of the container does not occur. The final entropy of the gas is higher than its initial entropy. Entropy, roughly, is a measure of the disorder of the state of the gas; and if the gas occupies only one half of the container, it is in a more highly ordered state than if it is spreads throughout the container approximately evenly.

Now, the basic strategy is to try to account for this asymmetry by appealing to the dynamics governing the micro-state of the system and a probability postulate. The overwhelming majority of micro-states compatible with the complete macro-state of a system not in non-equilibrium (such as the gas immediately after removing the partition), one argues, evolve into micro-states which correspond to macro-states with higher entropy. From this one concludes, assuming some intuitively plausible probability measure over micro-states, that it is overwhelmingly likely that the system is in a micro-state which leads to an increase in entropy.

Famously, one of the problems for an explanation of the thermodynamic asymmetry along these lines is that the micro-dynamics is taken to be time-symmetric. Thus, one also ought to be able to conclude from the fact that a system at present is in a relatively low-entropy macro-state far from equilibrium that its entropy increases *toward the past*—i.e. that the present state of the system was reached (in the normal time sense) through a *decrease* in entropy in the past. This is Loschmidt's reversibility objection. The empirically unacceptable application of the argument to retrodict a system's past can be blocked if we conditionalize the distribution of micro-states not only on the present macro-state of the system, but also on the (thermodynamically normal) past of the system—that is, if we postulate a 'past hypothesis' which says that the universe began its life in an extremely low-entropy state and that it is overwhelmingly probable that the initial micro-state of the universe was one that is 'typical' given this macro-constraint. Since the reversibility objection seems to apply at any point in the past, we seem ultimately to be driven to conditionalize the distribution of micro-states on a low-entropy state of the entire early universe. That is, we are led to assume that the universe currently is in a micro-state that is compatible *both* with its current macro-state *and* an extremely low-entropy initial state. This assumption is time-asymmetric (since we are not similarly postulating a low-entropy final state) and implies, according to the standard argument, that it is overwhelmingly probable that the entropy of the universe increases not only in the future but also monotonically increased in the past. Given that the universe was in an extremely low-entropy state in its distant past, it is much more probable that the universe evolved into its present state through a monotonic increase in entropy than that the current state evolved from a higher entropy past. Also, the account implies that *as far as predictions into the future are concerned* the 'naïve' account works: conditionalizing micro-states on a low-entropy past in addition to the present macro-state almost always yields the same result for the future evolution as conditionalizing on the present macro-state alone.

Thus, the key ingredients in this account of the thermodynamic asymmetry are this. There is an asymmetry at one 'level': The entropy of macroscopic closed systems never decreases. This asymmetry is a putative law at the macro-level, constraining the dynamic evolution of a quantity characterizing a system's macro-state. The micro-statistical account does not allow us to derive the macro-regularity as a strict law, as a consequence of the dynamical laws on the micro-level. Instead the micro-account changes the character of the macro-regularity in two important respects. First, despite the fact that the asymmetry may have looked like a *strict* regularity, it turns out to be only probabilistic: It is overwhelmingly probable for each isolated macro-system that it is in a micro-state such that its entropy does not decrease, yet there are micro-states compatible with the macro-state (and the past hypothesis) that result in anti-thermodynamic behavior. Second, the asymmetry fundamentally turns out not to be a matter of an asymmetry in the dynamical laws at all, but rather an asymmetry between initial and final conditions. Since the dynamics governing the evolution of the micro-states is time-symmetric, the asymmetry is accounted for by appealing to the fact that the initial state of the system (but not its 'final' state anytime in the near future of the universe) is one of extremely low entropy.

This is only a brief sketch of an account of the thermodynamic asymmetry in the Boltzmannian tradition. The account is far from unproblematic and at various stages needs to appeal to certain plausibility arguments, but this need not concern us here. What I am interested in investigating is whether, to the extent that the account is successful in the case of the entropy asymmetry, it can be extended to explain the asymmetry of radiation as well.

4. Entropy accounts of the radiation asymmetry

Here is a survey of how supporters of an entropy account propose to account for the arrow of radiation in a manner closely parallel to the entropy asymmetry. Frank Arntzenius says:

> Our own universe seems to be in the situation where the temperature of large chunks of matter is much hotter than that of the surrounding radiation. Therefore stars emit 'retarded' [i.e. coherently diverging] radiation. This behavior is the likely evolution towards the elusive goal of equilibrium. (Arntzenius, 1994)

While Craig Callender maintains that

> the crucial point to see is that the usual retarded radiation is a kind of improbable-to-probable transition. A concentrated source is improbable, but given its existence, a system will evolve toward more probable regions of the phase space, i.e., the waves will spread. Advanced radiation is likewise a species of improbable-to-probable transitions. Given an improbable source in the past, it will spread backwards in time to more probable regions of phase space too … whether one or both happens depends on the boundary conditions. (Callender, 2001)

And, most recently, Jill North expresses this view as follows:

> The temporally symmetric laws say that both advanced and retarded radiation could be emitted. However, given the universe's thermal disequilibrium, the charges are overwhelmingly likely to radiate towards the future, as part of the overwhelmingly likely progression towards equilibrium in that temporal direction. They are overwhelmingly unlikely to radiate towards the past because the universe was at thermal equilibrium in that direction. (North, 2003, p. 1095)

Thus, as in the case of entropy, according to these proposals, what appeared to be a strict law turns out to be a probabilistic feature of the world. And as in the case of entropy, the asymmetry is taken ultimately to be explained by the fact that the early universe was in a state of extremely low entropy in which hot sources were surrounded by a cool background field of approximately zero field strength.

While none of these quotes provide a fully fleshed out account of how entropy considerations are meant to explain the radiation asymmetry, the intuitions behind these suggestions seem powerful. The crucial feature of the early universe for a thermodynamic account appears to be that stars are hot objects surrounded by a much cooler radiation field—the thermalized 2.7 K cosmic background radiation. Given the extreme temperature gradient, thermodynamic considerations do seem to ensure that it is overwhelmingly more probable that the star be a net emitter of radiation rather than an absorber, which in turn implies (given that the background field is negligible compared with the retarded or advanced fields of the star) that it is overwhelmingly more probable that the radiation field associated with the star is retarded rather than advanced. At least in situations where the total field is very nearly fully retarded or fully advanced, retarded waves represent emission of energy and advanced waves represent absorption. Thus, it seems that hot sources in a cool surrounding are thermodynamically expected to be associated with retarded radiation, while cool sources in a hot radiation field should be associated with advanced radiation.

How in more detail is an entropy account of the radiation to proceed? Before discussing two possible suggestions for trying to flesh out such an account, I want to draw attention to several differences between the case of radiation and that of thermodynamics. Recall the *explanandum* in the case of the asymmetry characterizing total radiation fields. The asymmetry is that the total field can be represented as being fully retarded but not as being fully advanced; or, equivalently, that in setting up a Cauchy problem for the fields in the presence of coherent sources, we can pick initial-value surfaces with incoming fields approximately equal to zero but not final-value surfaces with outgoing fields approximately equal to zero. Thus, the first important difference between the *explanandum* in the radiation case and that in the thermodynamic case is that the latter concerns a *dynamic* constraint while the former does not. In the thermodynamic case, the *explanandum* is why *changes* in entropy are generally greater than zero. By contrast, the fact that fields can be represented as fully retarded is not a fact about the change of anything: If a field in a finite region of space can be represented as fully retarded at one time and if the fields on the temporal boundaries are zero, then this field will evolve into a field that at later times, in that particular representation, is fully retarded. Similarly, if for an infinite spatial volume we can find a Cauchy surface with zero incoming fields, then fields at all times in the future of that surface can be represented as fully retarded.[5] The *explanandum* in the radiation case concerns a non-dynamic property of fields at all times (which is equivalent to an asymmetry between certain *instantaneous* states—an asymmetry between the fields on initial as opposed to final value surfaces), while the thermodynamic asymmetry is an asymmetry concerning a dynamic constraint on changes in entropy.[6]

A second important difference between the two cases concerns the relation between the micro- and macro-levels in each case. So far in my discussion I have not distinguished between microscopic and macroscopic electrodynamic systems. This distinction has not been necessary, since the dynamical equations governing micro and macro phenomena in classical electrodynamics are formally equivalent and the very same asymmetry applies to phenomena in both domains. Following Lorentz's work on his theory of the electron, one can derive the macroscopic Maxwell equations from microscopic field equations through a statistical averaging procedure. The crucial step in the derivation is an argument to the effect that space and time differentiations and statistical averaging commute, which allows us to derive the macroscopic field equations as statistical averages from their microscopic analogues. The macro-equations look formally not much different from the microscopic field equations, aside from the fact that instead of the two microscopic fields—the electric and magnetic field vectors—the macro-equations involve four different fields, including the two 'displacement vectors' **D** and **H**. Both micro- and macro-fields are governed by a wave equation, and macroscopic retarded and advanced field representations can be derived from their microscopic analogues. In particular, that incoming macro-fields are

[5]Of course, once we find such a surface, then fields in its past can be represented as fully advanced. This is why a careful formulation of the asymmetry needs to take into account the number of sources in the problem. Fields are zero in the future of absorbers. Hence the fully advanced fields in the past of a Cauchy surface with approximately zero fields will contain a very large number of incoherent sources.

[6]Echoing one of the objections I considered above, one might ask whether there could not be worlds in which the incoming fields are zero but in which at some later time a source 'produces' advanced radiation. But of course such a world is not nomically possible, since any advanced field associated with a source in the future of the initial value surface would have to 'show up' on that surface, contradicting our assumption that incoming fields are equal to zero.

approximately equal to zero (and, hence, that macro-fields are approximately fully retarded), is obtained from the corresponding claim for micro fields by statistical averaging. Thus, in the radiation case the dynamics on *both* the micro- and the macro-levels are time-symmetric, and *both* levels exhibit analogous asymmetries in initial conditions. By contrast, the thermodynamic asymmetry concerns a quantity, i.e. entropy, that is not the average of a microscopic quantity. The asymmetry on the macro-level involves a dynamic constraint on that level, while the micro-dynamics is time-symmetric.[7]

How, then, do these differences in the *explananda* affect the prospects for an entropy account of the radiation asymmetry? I want to propose two different ways of construing an entropy account. On the first construal, an explanation of the radiation asymmetry is meant to proceed strictly analogously to that of the thermodynamic asymmetry. That is, the account postulates what the initial low-entropy macro-state of the system is. In particular, the initial macro-state is assumed to be one characterized by approximately zero, fully thermalized fields and large coherent sources. And it is then argued that it follows from these assumptions about the initial macro-state, a probability postulate, and the micro-dynamics that radiation is fully retarded.

On this first construal, however, the entropy account consists in little more than a restatement of the *explanandum*. As we have seen, the puzzle concerning the asymmetry of total radiation fields is this. Why is it, given that radiation fields are governed by a time-symmetric dynamical equation, that initial value surfaces can be chosen with approximately zero fields in the presence of coherent sources, while fields on final value surfaces are generally non-zero? Thus to specify, in accordance with the present suggestion, that initial fields are approximately zero in the presence of coherent sources merely restates the puzzle that we are trying to explain. Moreover, contrary to what advocates of an entropy account suggest, whether retarded radiation results in an evolution toward an equilibrium state simply plays no role in deriving that the total fields are fully retarded from the initial conditions. Once we specify that the initial fields are approximately equal to zero, it simply follows from the wave equation for the fields that the total fields in the future of the initial value surface are fully retarded.

On the present construal, the 'entropy' account amounts to this. It restates that initial fields are approximately equal to zero and it allows us to mimic a 'Lorentz-style' derivation of the field equations for macro-fields from their microscopic analogues. Given the initial conditions and the macroscopic wave equation, it simply follows mathematically that the total fields will be fully retarded. (Recall Eq. (1) above.) Thus, we have made no progress in accounting for the asymmetry between initial and final conditions that is commonly taken to constitute the puzzle of the arrow characterizing total radiation fields.

One might object that an entropy account consists in more than the mere postulate of zero incoming fields. North, for example, argues that the radiation asymmetry can be explained along the following lines. The universe began its life in a state of evenly distributed matter and fields—a state of thermal equilibrium but of extremely low gravitational entropy. As matter clumped— corresponding to states of higher gravitational entropy— it heated up, while at the same time the universe expanded and the field cooled, leading to a thermal disequilibrium. At this point entropy considerations are supposed to

[7]The relation between the microscopic and macroscopic Maxwell equations is discussed briefly in (Jackson, 1975, section 6.7). For a much more detailed derivation of the macroscopic theory from the microscopic equation, including a fully covariant derivation beginning from fully retarded micro-fields, see (Groot, 1969).

enter and North invokes the thermal disequilibrium between the hot stars and the cool field to explain the presence of retarded radiation.

But once we specify that the initial fields are isotropic, homogenous, and approximately equal to zero, it strictly follows that total fields will be fully retarded. Thus, the 'game is over,' as it were, once the early universe is posited as involving weak fully thermalized fields, and before entropy considerations can make its contribution. If we specify that initial fields are weak and fully thermalized, there is nothing left for the entropy account to explain.

Perhaps, then, contrary to what the quotes from North and Arntzenius suggest, we should try to locate the real contribution of the entropy account elsewhere—in an explanation of why the initial fields are fully thermalized. We have seen that an entropy account of the radiation asymmetry strictly paralleling an account of the thermodynamic asymmetry fails. It fails, since a thermodynamic account of the asymmetric dynamic constraint on thermodynamic systems captured by the second law crucially involves postulating asymmetric initial conditions; while in the radiation case it is precisely an asymmetry between temporal boundary conditions that is taken to be in need of an explanation. Thus, I now want to consider a second possible construal of an entropy account, according to which the fact that the initial field is fully thermalized—that is, that there initially is no coherent radiation—is meant to be the core *explanandum* of the entropy account, rather than to be assumed as part of the *explanans*. That is, while in the thermodynamic case certain assumptions about the initial macro-state are meant to account for a *dynamic* constraint *on the future evolution* of thermodynamic systems, the account I am considering now aims to explain why *a certain more fine-grained* constraint *on the initial macro-state*—i.e., the absence of coherent incoming radiation—*follows from very broad, more coarse-grained* constraints *on that state*—i.e., the assumption of low-temperature fields.

Now, one might think that despite the difference in explanatory structure the thermodynamic account does provide the resources for such an explanation. For it is a crucial assumption of the thermodynamic account that the initial state it posits is 'typical' relative to the constraints imposed on that state. This is usually put in terms of a distinction between micro- and macro-states: we assume that the microscopic initial conditions are typical relative to the initial macro-state. But presumably an analogous assumption can be made concerning fine-grained constraints on the macro-state relative to broader, more general constraints. Applied to our case, in assuming that the incoming field has a certain temperature we also assume that the incoming state of the field is one that is 'typical' for a field at that temperature, which presumably means that the field is fully thermalized. Since intuitively there are many fewer ways for an incoming field to contain large coherent radiation than for the field to be fully thermalized, the former is less 'typical' than the latter.

It is important to notice that of the two components of the past hypothesis—that the universe began its life in a state of extremely low entropy and that the initial micro-state was one that is typical relative to the macro-state—the second component is doing all the explanatory work. And contrary to what the quotes from Arntzenius or North suggest, the fact that the sources are hotter than the radiation field plays no role in the account. Indeed, there is no mention of sources at all. It seems to follow simply from the assumption that the radiation field is one that is typical for a field of a given temperature that there is no coherent incoming radiation. That is, whatever the state of any sources present, the most probable initial state for a field is assumed to be one where the field is fully thermalized, which implies, in the case that the temperature of the incoming field is very low, that the field is approximately equal to zero.

How successful is this explanation? Could we, for example, apply these considerations to the history of the solar system to argue that the radiation in the presence of Sun has to be fully retarded?

While this construal of how to determine the shape of a typical field appears to give the correct answer when applied to incoming fields, ignoring the sources leads to problems in the case of boundary values past the initial time and after the sources have turned on. What is the typical state for a field at some later time? The naïve answer to this question would simply be to use the same reasoning as in the case of initial fields: At any time the typical state of the field is one that is the most probable given the field's temperature, which will always be a fully thermalized field. Now, according to the account of the thermodynamic asymmetry that I sketched out above, if we use the state of a system at one time to predict the system's future evolution, the 'naïve' answer in almost all cases gives the correct result: We can treat the macro state of the system at any time past its 'birth' as the system's initial state and ask what the most probable micro states relative to that macro state are in order to predict the system's future evolution. Thus, if the correct procedure for determining the most probable shape of the field at one time is simply to determine what shape of the field is typical for a field of that temperature independently of any sources present, we get the false result that statistical considerations seem to imply that the total field should *never* be anything but fully thermalized.

The underlying problem can perhaps be brought into sharper focus in the following manner. While the naïve account for determining the typical state of a system ought to work as far as *prediction* is concerned, Loschmidt's reversibility objection has taught us that this procedure will give the wrong result, when we attempt to *retrodict* the past evolution of a thermodynamic system, since the most likely past evolution of such a system, given nothing but constraints on its present macro state, will be one that is anti-thermodynamic. Instead then, the standard account of the thermodynamic asymmetry asks us to conditionalize micro-states at a final time not only on the macro-constraints at that time but also on the macro-constraints at the initial time. Applied to our case, the procedure ought to be, it seems, that we have to conditionalize possible 'shapes' of the final field not only on the final temperature of the field T_f on a final value surface but also on the fact that the field evolved from one with initial temperature T_i. Yet it is unclear how this added constraint could affect the shape of the final field. No matter what the temperature difference between initial and final fields is, it seems, the 'typical' final state of the field ought to be one that is fully thermalized.[8]

But now it is easy to see that our procedure for determining 'typical' fields is incompatible with the field equations. In the presence of sources it will not be possible in general to find a solution for which *both* incoming *and* outgoing fields contain no coherent radiation. This is perhaps most easily seen in the particular case in which we are in fact interested—the case of a low-entropy background radiation with a small enough number of sources such that the emitted radiation does not appreciably affect the temperature of the background field. In this case the proposed procedure for determining typical initial

[8] Similarly for a gas in a container undergoing a temperature change, the gas in its final state will be spread evenly throughout the container, independently of its initial temperature.

and final fields seems to yield the result that both initial and final fields are approximately equal to zero. But this is incompatible with the fact that there are large coherent sources that have to be associated with either retarded or advanced radiation (or a linear combination of the two). According to the field equations, if the incoming fields are approximately equal to zero, then the final fields will not be; and if the final fields are approximately equal to zero, the initial fields will not be.

The problem with the proposal for determining the typical shape of a field we have been considering is—and this might have been obvious from the beginning—that in the presence of sources fields are *not* what they 'typically' would be in the absence of sources. What the most probable shape of a field with a certain temperature is, depends on the state of the sources present. But what, then, is the typical field in the presence of sources? Is it typical for the field to be fully thermalized in the *past* of field sources and to contain coherent diverging radiation in the *future* of sources? Or is it the other way around? Or, does the field perhaps contain coherent radiation in both directions, as Wheeler and Feynman's symmetric action-at-a-distance theory assumes? The correct answer apparently is that initial fields in the past of coherent field sources typically are fully thermalized, while final fields in the future of coherent sources contain coherent retarded radiation. But why is there this asymmetry between typical initial and final fields in the presence of coherent sources? What can explain this asymmetry? In trying to determine the correct shape of a field of a given temperature in the presence of sources we are faced with the very same puzzle with which we began our investigation.

The puzzle of the arrow of radiation is why in the presence of large coherent sources there is coherent outgoing radiation but no coherent incoming radiation. This puzzle is not solved by reasoning that incoming fields are fully thermalized because this is their typical state in the *absence* of sources. For this line of argument owes us an explanation for why it is appropriate to ignore sources in determining the typical field at initial times but not at final times. But neither can we solve the puzzle by simply positing that the presence of sources affects the shape of the total field in their future but not in their past, since this is precisely the asymmetry that we took to be in need of an explanation in the beginning.

Finally, and perhaps most damagingly, an entropy account of the radiation asymmetry is not general enough. At most the account could explain an asymmetry associated with thermal sources. Yet, as I discussed above, the radiation asymmetry applies equally to sources whose micro-state we take to be fully determined—such as microscopic sources in particle accelerators. Standard treatments of classical electrodynamics, such as (Jackson, 1975), assume that fields associated with micro-sources exhibit the radiation asymmetry as well, and the entropy of the field associated with a micro-source of a given determinate trajectory is zero (since the instantaneous state of the source and the field occupy only a point in phase space). An accelerating source of determinate trajectory does not, therefore, change the entropy of the field, even when it is possible to represent the total field in the presence of such a source as fully retarded. The simple entropy account cannot explain the asymmetry of radiation in this case, even if—contrary to what I argued above—the account would be able to explain the asymmetry in the case of thermal sources.

What options might be left open for a defender of an entropy account of the arrow of radiation in light of my discussion? I have argued that there is simply no explanatory gap, as it were, between the assumption that incoming fields are weak and fully thermalized and the claim that fields are approximately fully retarded. Thus, a revised entropy account would have to be careful not simply to assume weak incoming fields but rather to take them to be the central *explanandum* of the account. And the account would then have to show that the fact that incoming fields are weak and fully thermalized follows from the fact that such a state is typical for fields even in the presence of radiating sources—and even in the presence of microscopic sources with determinate trajectories—without smuggling in the assumption that sources contribute retarded radiation.

At this point one might be tempted to suggest that there is a different strategy by which thermodynamic consideration could help in deriving the initial condition of approximately zero incoming fields independently of the source distribution in the future of the initial boundary. Instead of focusing on the thermodynamics of the *sources* and the *surrounding fields*, one could try to argue that the absence of coherent incoming radiation is ensured by the presence of *absorbing media* in the past of initial value surfaces. This is the strategy pursued by absorber accounts of radiation, the most promising of which, I believe, is that in (Zeh, 2001). The advantage of absorber accounts, such as Zeh's, is that they promise to be general enough to cover both the case of large cosmological thermal sources (with the early universe functioning as the relevant absorber) and the case of micro-sources in a laboratory setting (where, according to Zeh, the walls of the laboratory function as absorbers). But pursuing Zeh's strategy means giving up on the 'simple' entropy account I examined here. I discuss and criticize several versions of absorber accounts of the radiation asymmetry in (Frisch, 2005), and I do not want to repeat this discussion here.

5. Conclusion

I want to sum up what I have argued in this paper. Total electromagnetic fields in the presence of sources of radiation exhibit a temporal asymmetry. This asymmetry can be expressed in terms of an asymmetry between prevailing initial and final conditions. This claim, obviously, is not new and is a direct consequence of the relevant mathematical formalism. I did, however, propose a formulation of this asymmetry characterizing temporal boundary conditions that is somewhat more precise and careful than those found in the literature to date.

I criticized accounts that argue that the radiation asymmetry is a direct consequence of micro-statistical considerations and that the asymmetry can be explained as an evolution toward thermodynamic equilibrium by considering two reconstructions of such an account. First, in close analogy to accounts of the thermodynamic asymmetry, the account might assume a low-entropy initial state of hot sources and very weak, fully thermalized fields as premise and then try to derive from this that radiation fields are approximately fully retarded. That initial fields are very weak and fully thermalized, means that (compared to the fields associated with the sources) incoming fields are approximately equal to zero, but this just is the fact that is in need of an explanation—the fact that incoming fields are approximately equal to zero (and hence the total fields are fully retarded), while outgoing fields are not. If, as the quote from North above suggests, an entropy account were to assume as premise that incoming fields are approximately equal to zero, then there would be no work to be done by statistical considerations, and the account would amount to nothing more than a restatement of the *explanandum*.

Secondly, I considered the proposal that, by contrast with accounts of the thermodynamic asymmetry, statistical considerations in the radiation case are meant to explain merely how certain more fine-grained constraints on the initial state of the field—i.e., the absence of coherent radiation—followed from more coarse-grained constraints—

i.e., that the field at the initial time has a certain temperature. The problem for this proposal is how to incorporate the fact that radiation fields are coupled to sources of radiation. If we simply ignore the coupling, we get the right answer for initial fields, but also get the obviously mistaken result that the total field *at all times* will be fully thermalized. If, on the other hand, we want to include the coupling, with the consequence that there will be coherent radiation fields present *somewhere*, the question arises whether the 'typical' field in the presence of sources contains coherent radiation before or after the sources turn on. This question just is the question we are trying to answer in the first place.

There is, of course, a well-known answer to the question why *macro-fields* in the presence of *macro-sources* are approximately fully retarded. This answer appeals to the way in which, following Lorentz, macro-fields are derived as averages from their microscopic analogues and argues that macro-fields are approximately fully retarded since micro-fields are. The standard derivation of the macro-field equations from the micro-equations in fact gives an answer appealing to micro-statistical considerations to the question as to why there is coherent radiation after but not before a macro-source turns on. The answer is that the initial micro-fields are approximately equal to zero and hence that the micro-fields are approximately fully retarded (see, e.g., Groot, 1969). This answer falls short of a solution to the puzzle of the arrow of radiation (and is not intended as such in the physics literature), since it simply replaces the puzzle regarding macro-fields with an equivalent puzzle concerning micro-fields. And as I argued, the simple entropy account cannot explain the asymmetry in the case of sources with microscopically determined trajectories.

In light, both of the problems faced by existing absorber theories of radiation, which I discuss in Frisch (2005), and of the results of the present discussion, it appears highly doubtful that thermodynamic considerations on their own can account for the asymmetry of radiation and that the arrow of radiation is reducible to that of thermodynamics.

References

Albert, D. Z. (2000). *Time and chance*. Cambridge, MA: Harvard University Press.

Arntzenius, F. (1994). The classical failure to account for electromagnetic arrows of time. In: T. Horowitz, A. I. Janis (Eds.), Scientific failure. Boston: Roman & Littlefield Publishers.

Callender, C. (2001). Thermodynamic asymmetry in time. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Frisch, M. (2005). *Inconsistency, asymmetry and non-locality: A philosophical investigation of classical electrodynamics*. New York: Oxford University Press.

Groot, S. R. de. (1969). *The Maxwell equations. Studies in statistical mechanics*, Vol. IV. Amsterdam: North-Holland Publishing Company.

Jackson, J. D. (1975). *Classical electrodynamics* (2nd ed.). New York: John Wiley & Sons.

North, J. (2003). Understanding the time-asymmetry of radiation. *Philosophy of Science*, *70*, 1086–1097.

Price, H. (1996). *Times arrow & archimedes point*. Oxford: Oxford University Press.

Rohrlich, F. (1990). *Classical charged particles*. Reading, MA: Perseus Books.

Wheeler, J. A., & Feynman, R. P. (1945). Interaction with the absorber as the mechanism of radiation. *Reviews of Modern Physics*, *17*, 157–181.

Zeh, H. D. (2001). *The physical basis of the direction of time* (4th ed.). Berlin, New York: Springer.