Time in Cosmology

Chris Smeenk

Modern cosmology apparently answers many longstanding questions regarding the nature of time. For example, is the universe temporally finite or eternal, and is there a unique global sense of time? The Standard Model of cosmology renders the following verdicts. The universe is temporally finite, and approximately 13.7 billion years old. Events in the universe can be ordered according to a "cosmic time," which corresponds to time as measured by a particular class of fundamental observers since the "big bang." Many questions that remain unanswered seem to be answerable by empirical methods, if only in principle. Whether the universe will also be temporally finite to the future can be resolved, in principle, by measuring the actual matter and energy density in the universe. If it exceeds the so-called critical value, then the universe will be temporally finite to the future as well, collapsing back into a "big crunch."

Taking relativistic cosmology to provide direct answers to such questions is, however, misleading in (at least) two different senses. First, it neglects the extent to which the questions have not been so much answered as transformed into, or replaced with, new questions. A similar transformation occurred with regard to the central cosmological question of the 16-17th centuries: does the Sun or the Earth move? Posing this question sufficiently clearly for astronomical evidence to provide a decisive resolution of it required reformulating the concepts of space, time, and motion. Similarly, aspects of time presumed in posing the questions above pre-theoretically are significantly revised in relativistic cosmology. Second, even taking these conceptual changes into account, the answers to the questions are subtler than is usually acknowledged. The answers summarized above hold for a class of particularly simple, idealized cosmological models. But in what sense do these answers apply to the real universe? And in what sense do these answers address the initial questions regarding the nature of time?

This essay aims to provide a self-contained introduction to time in relativistic cosmology that clarifies both how questions about the nature of time should be posed in this setting and the extent to which they have been or can be answered empirically. The first section below recounts the loss of Newtonian absolute time with the advent of special and general relativity, and the partial recovery of absolute time in the form of cosmic time in some cosmological models. Section II considers the beginning and end of time in a broader class of models in which there is not an analog of Newtonian absolute time. As we will see, reasonable physical assumptions imply that the universe is finite to the past, and Section III turns to consideration of the "beginning" itself. We critically review conventional wisdom that a "singularity" reveals flaws in general relativity and briefly assess ways of avoiding the singularity.

Overall the first three sections describe how to translate questions about the nature of time into the language of the spacetime geometry of cosmological models. This approach to understanding the content of cosmological models is not controversial, although the implications for philosophical disputes are more contentious. The final two sections turn to two recent debates, with no pretense of an entirely even-handed survey. First, the propriety of attributing geometrical properties to spacetime itself, independently of the dynamical behavior of matter within spacetime, has been challenged. If there is nothing that can function as a clock, what does it mean to attribute a temporal length to a curve? We will consider the ramifications of this debate for cosmology in Section IV. The final section takes up the source of the asymmetry in our experience of time. Here I will critically review the popular idea that the pervasive temporal asymmetry of our experience can be traced to a cosmological asymmetry, in the form of the stipulation that the universe "began" in a far-from-equilibrium state.

I. Absolute Time: Lost and (partially) Regained

Reflecting on the implications of Einstein's special theory of relativity, Minkowski famously declared in 1909 that "Henceforth, space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality" (Lorentz et al. 1952, p. 75). This shift to a spacetime perspective is driven by the fact that the relation of simultaneity is observer-dependent rather than absolute, as in earlier theories. As a result, there is no longer an invariant, observer-independent way to decompose Minkowski spacetime, as it is now known, into space and time. In the first act of our tale of the evolving concept of time, this spacetime perspective emerges from a previous, more intuitive conception of absolute time.

To set the stage, consider how to introduce an appropriate structure to characterize the intuitive properties of time. The structure ought to capture features of time such as measurements of the duration of processes and their relative ordering. We will begin by defining the structure holding amongst idealized intervals without duration, "points of time." Suppose that we introduce a temporal metric, representing "time elapsed" between any two such points. This gives the "time line" a structure slightly weaker than that of the real number line.[1] The duration and temporal order of finite temporal intervals are then defined in terms of the temporal metric.

Next, consider how to assign time to spatially distant regions. It is intuitively quite compelling to suppose that the collection of events (localized spatially as well as temporally) can be partitioned into sets of simultaneous events, "instants." This will be possible if the relation of simultaneity is an equivalence relation (that is, symmetric, reflexive, and transitive), as it seems natural to assume. For a given event, the equivalence class under the simultaneity relation includes all the events happening at the same time at different spatial locations throughout the universe. The time line is then the structure defined on the collection of such "instants." Note that we have not done justice to one of the most striking features of our experience of time, namely the sense of time's cruel passing as our minutes hasten to their end. We have not distinguished any "instant" from any other, nor have we introduced a partition into past, present, and future.

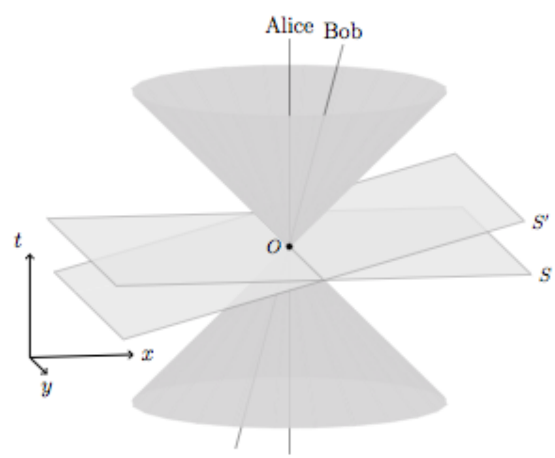
Newton gave a clear explication of time consistent with this intuitive account in the *Principia*. But the crucial aspect of Newtonian mechanics that insured compatibility between physical theory and this account is precisely what Einstein was forced to modify. One of Newton's achievements was the reconciliation between two apparently conflicting ideas in 17th century physics, namely inertia and Galilean relativity. With the introduction of inertia, motion was treated as a state of a body that is changed by a non-zero net force. The fundamental distinction between a state of motion that is changing versus one that is not (that is, between accelerated and inertial motion) should hold for all observers. Yet Galilean relativity introduces a set of observers regarded as physically equivalent, called inertial observers, and these observers disagree about quantities like spatial position and velocity.[2] Evading the conflict requires recognizing that although some quantities vary, others do not --- in particular, acceleration is absolute in the sense of being observer-independent whereas velocity is not. In different language, acceleration is invariant under the transformations between observers allowed by Galilean relativity, the Galilean transformations. This suffices to draw the appropriate physical distinction between inertial and accelerated motion.

For our purposes, the main point is that the Galilean transformations have no impact on the relation of simultaneity: inertial observers agree on how to partition spacetime into instants. Newton's laws of motion presume that the local time parameters appearing in the description of spatially separated systems can be combined into a single universal time. Rather than being mere shadows of spacetime, space and time thus have their own independent reality in the sense that there is a unique, universal foliation.

This appealing account of time conflicts with ideas introduced in the theory of electromagnetism. Einstein recognized this conflict and argued that Newton's account was not well founded empirically as a result. In a nutshell, Newton's approach assumed that absolute simultaneity of two events could be established by suitable measurements. But in fact measurements of the time elapsed between, for example, two astronomical events rely on light signals, with a finite signal speed. The concept of simultaneity appropriate for a theory with a finite, observer-independent maximum signal speed is not absolute as in Newton's theory. Einstein argued that mechanics should be reformulated in terms of the Lorentz transformations discovered in the study of electromagnetism rather than the Galilean transformations. These transformations can be derived by requiring that all inertial observers measure the same value for the speed of light.

One consequence of the form of these transformations is that there is no longer a unique foliation of spacetime into instants valid for all inertial observers. For an inertial observer Alice, for example, spacetime is foliated into a sequence of instants orthogonal to her worldline.[3] The instants according to a second inertial observer, Bob, moving at some velocity with respect to Alice, are tilted (see figure 1). The Lorentz transformation from Alice to Bob maps one of Alice's instants into a distinct set of events that intersects but does not overlap it. As a result, this set of points deemed to be simultaneous according to Bob cannot be an equivalence class of Alice's simultaneity relation. Rather than an absolute or universal simultaneity relation we thus have simultaneity relative to an observer; it is only relative to a chosen observer that we recover the foliation of spacetime into instants. More generally, Alice and Bob disagree about the spatial and temporal distances between any two events while agreeing on the spacetime distance between them. The universal quantity all observers agree upon is thus spacetime distance rather than temporal distance, as in Newtonian physics. We are left with spatial and temporal quantities as observer-dependent, pale reflections of universal spacetime quantities --- and Minkowksi's emphasis on the primacy of *spacetime*.

INSERT FIGURE 1 HERE (SmeenkSimultaneity)



CAPTION:
Figure 1: Alice judges events in the hyperplane S to be simultaneous with O, whereas Bob's surface of simultaneity is S'. Alice and Bob agree, however, regarding the light cone: the set of events that can be connected to O via light signals. As is conventional in spacetime diagrams, one spatial dimension is suppressed and time goes up along the vertical axis.

In the second act of this oft-rehearsed tale of the loss of absolute time there is an unexpected reversal of fortune. Einstein's theory of gravity, general relativity (GR), differs from both Newtonian theory and special relativity in describing spacetime geometry as variable and dynamical. In GR, the "gravitational field" is incorporated into the spacetime geometry, uniting inertial structure and gravitation in a single inertio-gravitational field. What Newton would have described as spatial variation of the gravitational field within a fixed background spactime geometry is represented instead as a change in the geometry itself. Just as the gravitational field depends on mass in Newton's theory, in GR the spacetime curvature within a region is related to the stress-energy present via Einstein's field equations.[4] The trajectory of a body deflected from an inertial path by gravity alone, in the language of Newtonian theory, is the straightest path, or geodesic, within this curved geometry.[5] Rather than having an immutable background spacetime fixed *a priori*, in GR the spacetime geometry varies from one solution of the field equations to another and there are few constraints on allowed geometries. Indeed, there are solutions of the field equations with exotic spacetime geometry quite different than that of Minkowski spacetime.

GR adds a further obstacle to the analysis of time as a foliation of spacetime into instants. Consider again Alice and Bob, now in a general relativistic spacetime. If their trajectories cross at a point and they have non-zero relative velocity, then --- just as in special relativity --- they will decompose spacetime distance into spatial and temporal distances differently, and their "instants" will be tilted with respect to each other. But suppose we respond to the relativity of simultaneity by choosing one observer, say Alice, and privileging her simultaneity relation. This would suffice to re-establish absolute time in special relativity, if there were legitimate grounds for granting the privilege. In GR, however, there is a further question: is it possible to extend Alice's instants beyond her "local neighborhood," effecting a global division of spacetime into space and time? The question can be framed more precisely by considering the foliation into instants with respect to a family of observers, represented by a congruence of curves. When is it possible to define a global foliation of spacetime into instants everywhere orthogonal to a given congruence?

It is possible in the models that are now part of the standard model of cosmology, the FLRW (Friedman-Lemaître-Robertson-Walker) models. These models follow from requiring that the spacetime geometry is both *homogeneous* and *isotropic*.[6] Imposing these symmetries implies that the models can be foliated by (globally extended) instants, leading to a well-defined notion of "cosmic time." There is also a naturally privileged congruence of observers: the "fundamental observers" who remain at rest with respect to the matter in the model. The worldlines of the fundamental observers are orthogonal to the instants, and a watch worn by a fundamental observer measures cosmic time elapsed. Einstein's field equations reduce to a pair of equations governing the scale factor $R(t)$, which represents the changing spatial distance between the fundamental observers as a function of cosmic time. Cosmic time in the FLRW models resembles Newtonian absolute time: for any two events in spacetime, the cosmic time elapsed between the events has a definite, observer-independent value. All of these properties lead to the conclusion that spacetime in the FLRW models breaks down into space and time much as in Newtonian spacetime, as Jeans and Eddington emphasized soon after the models were discovered.

The existence of such a global foliation is, however, a contingent feature that is not guaranteed to hold by the dynamical laws. This is vividly illustrated by a solution to Einstein's field equations discovered by Gödel (1949). This spacetime represents a "rotating universe," in which matter is in a state of uniform rigid rotation. Due to this rotation it is not possible to define global "instants" as in the FLRW models. To see why this is so, consider cutting through a collection of threads with a knife. If the threads are parallel to each other, it is possible to cut through with the knife orthogonal to each and every thread, but this is impossible if the threads are twisted into a rope. Analogously, the construction of global "instants" described above for the FLRW models can only be carried out if there is no "twist" or rotation of the congruence under consideration.[7]

Furthermore, spacetimes that admit a global foliation into instants typically admit *many* different foliations, with no physical grounds to choose one as the "true" foliation. A unique foliation can be chosen in the FLRW models by exploiting their extreme symmetry. Rather than having a unique "cosmic time" as a replacement for Newtonian absolute time in a cosmological setting, we then have a variety of different possible notions of cosmic time and associated foliations. These foliations lead to striking differences regarding, for example, whether the universe is finite or infinite in spatial extent. Again it seems natural to regard spacetime as the primary entity, which can be broken down into spatial sections evolving in a global cosmic time in a variety of inequivalent ways.[8]

Cosmology thus recovers an analog of absolute time in specific models with a high degree of symmetry, a contingent feature. Gödel, for one, argued that this is unsatisfactory:

> … if someone asserts that this absolute time [in the FLRW models] is lapsing, he accepts as a consequence that, whether or not an objective lapse of time exists … depends on the particular way in which matter and its motion are arranged in the world. This is not a straightforward contradiction; nevertheless, a philosophical view leading to such consequences can hardly be considered as satisfactory. (Gödel 1949, 562)

This dissatisfaction is hard to square with the main innovation of GR, namely that various (presumably) objective features of spacetime geometry depend on "how matter and its motion are arranged."

Even for those who do not share Gödel's dissatisfaction, there is a further notable consequence. Questions like those mentioned in the introduction – "how old is the universe?" – are only well-posed within a remarkably restricted class of spacetime models. The questions are moored to the structure of the models fixed by strong symmetry assumptions, so to speak. But in more realistic models describing a lumpy, bumpy "almost-FLRW" universe, these questions are cast adrift. Below we will consider ways of formulating questions regarding the nature of time more broadly, so that they apply to a subset of spacetimes including the FLRW models but extending beyond them. There are a number of questions that can be posed in terms of invariant features of the spacetime geometry, although it is generally not possible to make a clean separation between spatial and temporal structure.

II. The Beginning and End of Time

Is the universe finite to the past? In the FLRW models this question is well-defined due to the unique foliation of spacetime into instants labeled by cosmic time. Tracing backward from a particular instant taken to represent the universe "now," the clock carried by any fundamental observer measures the cosmic time elapsed. As one goes backward, the density of matter $\rho(t)$ increases without bound and the scale factor $R(t)$ decreases towards zero. This behavior reflects the basic fact that, in outmoded Newtonian language, gravity is a force of attraction. Since $\rho(t)$ is positive the spacetime curvature is positive, leading to convergence of timelike geodesics. Running backwards from "now," the worldlines of fundamental observers converge and $R(t)$ decreases; in turn, the density $\rho(t)$ increases as the matter and energy is compressed into a smaller volume. In the FLRW models, $\rho(t) \to \infty$ and $R(t) \to 0$ within a finite time, T. A fundamental observer would regard this T as the "age of the universe," or time elapsed since the Big Bang as measured by her clock.

There is no "first instant" of time in this picture any more than there is a "first point" in an open interval of the real line. Although the density increases without bound within a finite interval, there is no time in the FLRW models at which the density is infinite. The curves representing fundamental observers do not "run into the Big Bang," although they are incomplete in the sense of having finite length. Talk of the "Big Bang" can be misleading as it suggests that there is a "singular region" *within* spacetime. It would be inconsistent to treat singular points as points in spacetime, although one can study singularities by adding boundary points. (In this case, for example, one could add a boundary consisting of the point $-T$ on each geodesic, much like adding boundary points to make an open interval closed. But the resulting spacetime *with* boundary is not taken to represent physical spacetime directly.) Singularities are perhaps best regarded as global properties of the spacetime reflected in features such as the existence of incomplete geodesics.[9] In any case, as Torretti (1999) remarks, modern cosmology evades Kant's first antinomy through mathematical subtlety: there is no "first instant" or beginning in time, yet the universe does not have an infinite past.

An alert reader will by now be wondering how much of this generalizes beyond the FLRW models. In some solutions, such as Gödel's described above, there is no way to meaningfully speak of cosmic time because there is no cosmic time function or foliation into instants. There is, however, a clearly defined subspace of solutions for which cosmic time functions exist --- those that are "globally hyperbolic."[10] These solutions are topologically the product of space at an instant and the cosmic time, $\Sigma \times \Re$, although in general there is no way to privilege one of the many ways to decompose the spacetime.

The conclusion that time is finite to the past holds, surprisingly, for spacetimes satisfying relatively weak assumptions. In the early days of relativistic cosmology Einstein and others regarded the singularity in the FLRW models as an artifact of the idealizations in these models. Surely, they thought, giving a more realistic distribution of matter would undercut the conclusion that the universe has a finite age. The singularity theorems established by Hawking, Penrose, and Geroch in the 60s show that they were wrong. Gravity leads to convergence of nearby geodesics, as described above, in the presence of "normal" matter. For spacetimes satisfying a short list of physically motivated conditions, much weaker than the symmetries assumed in the FLRW models, this focusing property is sufficient to establish that there is a finite lower bound on time to the past.[11] This statement exploits the fact that these models admit a global time function, and the universe is finite to the past if the range of this function is bounded below. Unlike the FLRW models, there are many possible choices of a global time function and no grounds to distinguish one. Some questions that can be answered directly in the FLRW models will be trickier in the general case. For example, the finite lower bound on cosmic time in the FLRW models has a fairly direct interpretation: a clock ticking along the worldline of a fundamental observer would measure approximately 13.7 billion years elapsed from the big bang until now. But in the more general class of models, a given global time function need not correspond directly to any similarly physically meaningful quantity.

Turning to face the future, the recent discovery of dark energy has dramatically altered the conventional wisdom regarding the fate of the universe. Textbooks often say that "geometry is destiny" in the FLRW models, because the spatial geometry of the universe at a given instant determines the fate of the universe --- that is, the behavior of $R(t)$ as $t \to \infty$. The FLRW models divide into three classes of solutions depending on how the energy density compares to the "critical density," defined as the density required to counteract exactly the initial expansion (such that $\dot{R}(t) \to 0$ as $t \to \infty$). The "critical" case has spatial sections with flat (Euclidean) spatial geometry and expands forever, as does any FLRW model with density lower than the critical density. In the latter case the sections have hyperbolic geometry and the sign of $\dot{R}(t)$ does not change. But for an FLRW model with greater than critical density, the gravitational attraction is sufficient to reverse the expansion, leading to a change in the sign of $\dot{R}(t)$ and a "Big Crunch." The spatial sections of these models have spherical geometry. Hence the spatial geometry determines the eventual fate of the universe. Several observational programs have aimed to determine which of the three models best describes the actual universe by measuring the matter density.[12]

But all of this is based on assumptions regarding the matter and energy contents of the universe that now seem untenable. "Normal" matter slows down the initial expansion of the universe, as one would expect of matter with positive energy density given the attractive nature of gravitation. One can also introduce types of matter, however, that have the opposite effect --- that is, matter with negative energy density, which leads to repulsion and an acceleration of the expansion rate. Two teams discovered that the expansion rate is accelerating in studies of high-redshift supernovae initially announced in 1998, for which they received the 2011 Nobel Prize. This result is in agreement with several other lines of evidence in cosmology indicating that the best current models need to include a substantial amount of what is now called "dark energy."

Dark energy enters into one of equations for $R(t)$ with the opposite sign as "normal" matter, leading to accelerated expansion, and it also does not dilute with expansion like normal matter. Dark energy is thus expected to dominate the evolution of the universe as ordinary matter is diluted away by the expansion. This undermines the usefulness of measuring the spatial geometry or current matter density for predicting the eventual fate of the universe, as the fate depends instead on dark energy and its density compared to that of normal matter. Currently there are several different models of dark energy, which differ in treating it as a true "cosmological constant" or as a fluid with an equation of state that varies with cosmic time.

These models lead to different scenarios for the far future of the universe, replacements for the 19th century "heat death." Lord Kelvin and his contemporaries argued that a Newtonian universe will eventually reach a uniform temperature, erasing the temperature differences needed for heat engines to extract work. The inclusion of dark energy in contemporary cosmological models leads to other possibilities, such as the "Big Chill," so called because the temperature decreases with the expansion, leading to a modern version of heat death with an ever-decreasing, uniform temperature, and the "Big Rip," in which $R(t) \rightarrow \infty$ ("ripping" spacetime apart) within a finite time.

III. Through the Big Bang?

What do our current theories imply regarding the beginning of time? One influential line of thought aims to avoid a true beginning of time, by eliminating the big bang or extending through it. First we will assess whether the grounds for abhorrence of singularities are compelling, before turning to the possibility of extending spacetime through a singularity.

The singularity theorems establish that singularities occur generically in a class of physically reasonable spacetimes, including but more general than the FLRW models. Many physicists have followed Einstein's lead in regarding the existence of singularities as a deep flaw of GR. The singularity theorems make it harder to brush singularities under the rug, by treating them as artifacts of unrealistic assumptions. Rather than seeking other ways of avoiding singularities, perhaps we should reconsider the alleged problems that arise in spacetimes with a singularity (following, in particular, Earman 1995)?

Assessing one apparent problem requires disambiguation of different kinds of singularities. Popular discussions of singularities misleadingly imply that their existence insures a failure of determinism; one pictures the kitchen sink, Ray Kurzweil, and everything in between emerging at random from a lawless region of spacetime. The initial singularities in the FLRW models are, however, compatible with determinism in the sense that the laws and appropriate initial data fully fix the solution throughout spacetime. There are different kinds of singularities (called "naked") that do, however, pose a threat to determinism. Penrose has conjectured that such singularities do not arise under physically reasonable conditions in GR, but this so-called "cosmic censorship" hypothesis remains unproven. A proof would show that GR only allows singularities that do not lead to a conflict with determinism, with the "cosmic censor" effectively hiding the naked singularities from view. Even without such a proof, it is clear that some singularities – such as the initial singularity in the FLRW models --- cannot be rejected on these grounds.

A second line of argument takes singularities to be a sign of the incompleteness of GR. The theory proves its own undoing, so this argument goes, in the sense that it predicts singularities where the laws of GR themselves break down. This argument is also problematic. It is misleading to think of singularities as if they are localized regions of the spacetime --- "the region where curvature is infinite" --- and, relatedly, it is misleading to think of other regions as "close to the singularity." There is then not a clear way of giving content to the rough idea that GR breaks down at, or as one approaches, a singularity. While it is true that quantities appearing in the relevant equations would no longer be well-defined "at the singularity," this does not establish incompleteness of GR. By hypothesis the spacetime itself does not include such singular regions, and one can still maintain that the laws of GR hold throughout spacetime.

Although this discussion falls far short of a full assessment, Earman (1995)'s challenge to the conventional wisdom regarding the implications of singularities reveals that it is on shaky footing. Regardless of one's view about their other implications, singularities certainly spell the end of the story in classical GR. Revisiting a debate from the early days of relativistic cosmology will help to make this point clear. The debate concerned whether De Sitter spacetime, introduced by De Sitter in correspondence with Einstein, harbored a singularity. The debate was resolved by showing that Einstein and De Sitter had been concerned with only a part of the full spacetime. There is no difficulty with extending the spacetime through what they had mistakenly called a singularity (and is now called an "event horizon"). Singularities such as that in the FLRW models cannot be handled in the same way, however. The scale factor and the density evolve such that no extension can preserve mathematical conditions typically imposed to insure that the field equations are well defined.[13] Speculation regarding what happened "before the big bang" has to be based on something other than just classical GR.

There are reasons other than a hope to indulge in such speculations for taking GR to be incomplete. Since it is a theory of gravity and sets aside the other fundamental forces it is incomplete, and it is incompatible with the theory describing the other fundamental forces, quantum field theory (QFT) (cf. Callender and Huggett 2001). The incompatibility of QFT and GR is not a pressing problem for most of the applications of each theory, due to the different length scales at which the strength of the different forces is relevant. However, the world is not cleanly divided into separate domains of applicability for QFT and GR, and neither theory offers a complete account of phenomena even within their intended domains. The early universe provides one example of such overlapping domains. Developing a full account of physical processes in the early universe requires drawing on aspects of both GR and QFT. The need for such an account does not derive from the existence of singularities *per se.* However, it is natural to expect that quantum effects will be important in the evolution of a system that would classically lead to a singularity.

Claims regarding the fate of singularities in a successor theory to GR are, at this stage, not on entirely firm footing. One approach to assessing the impact of quantum effects is to reconsider the status of the basic assumptions of the singularity theorems. These theorems assume that the presence of matter leads to the convergence of neighboring geodesics. In QFT, however, fields can have negative energy densities, which lead to a divergence of neighboring geodesics. This opens up the possibility of exotic states of matter that would avoid full collapse to a singularity with a "bounce" due to this repulsive effect.

But this approach takes GR to set the terms of the discussion, and perhaps the classical spacetime description fails more dramatically with respect to singularities. Recent work drawing on different approaches to quantum gravity has led to very different accounts of the initial singularity. Bojowald has argued, based on applying loop quantum gravity to cosmology, that spacetime will reach minimum finite size rather than reaching a singularity. Yet GR will fail to apply even approximately at this "bounce," so in a sense there is still a "beginning" of classical spacetime on this account. Turok and Steinhardt have pursued a program called ekpyrotic or cyclic cosmology, based on string theory, which exploits the possibility of extending through the initial singularity in a higher-dimensional spacetime.[14] On either approach, questions regarding the beginning of time and extensions through the initial singularity must be addressed based on a successor to GR.

IV. Operationalism

We have above adopted a way of speaking common in relativity textbooks, taking spacetime geometry to be a mathematical structure defined over a collection of events. Brown (2005), however, has argued that this reflects a philosophical mistake; on his view, spacetime geometry should be regarded instead as representing salient properties of the dynamics governing matter. Although the label is not entirely appropriate for Brown's position, an operationalist would likewise seek an observable manifestation of the geometrical quantities ascribed to spacetime. We will not tackle the general dispute directly here, instead considering implications of cosmology. What does time mean in the very early universe, or in the far future, where there may not be physical systems that could possibly function as clocks (cf. Rugh and Zinkernagel 2010)?

Suppose that the matter filling spacetime is governed by *conformally invariant* dynamics. A theory is conformally invariant if solutions can be generated from a given solution by transformations that "re-scale" spatial and temporal distances. Under such a transformation the light cone structure remains invariant --- i.e., if two points p,q can be connected by a light ray in the given solution, this will also be true in solutions generated by a conformal transformation. Spatial and temporal distances, however, generally vary from one solution to another. A theory with dynamical equations that are conformally invariant treats these solutions as equivalent. There is then no basis to claim that the proper time elapsed along a given worldline is given by, say, 1 second rather than 1.3 hours. (Similarly, there will be no basis for measurements of length.) In such a situation the spacetime metric does not have the full physical significance usually attributed to it --- the metric says *more* about the spacetime geometry than can be revealed through measurements.

Conventional accounts of the early universe describe it as going through a series of symmetry-breaking phase transitions that lead to the physical distinctions between the strong, weak, and electromagnetic forces, and the generation of particle masses. On some accounts, prior to these phase transitions the laws are approximately conformally invariant. Penrose (2010) has also argued that in the far future the universe may be governed by conformally invariant dynamics: all matter will eventually decay, black holes will evaporate, and so on, leading to a universe filled with only electromagnetic and gravitational radiation.[15] (This leads to his speculative conformal cyclic cosmology, based on identifying the initial singularity with the future singularity, leading to a cyclic model.) If there are cosmological regimes described by conformally invariant dynamics, then GR gives a richer description of spacetime geometry than the matter theory demands. In such regimes GR would have the flaw of introducing asymmetries that are not inherent in the phenomenon.

Regardless of one's philosophical approach, this ought to be recognizable as a problem. On a spacetime approach, a match between the dynamical symmetries of the matter theory and the spacetime symmetries is typically treated as a condition of adequacy. On Brown's approach, the spacetime symmetries *just are* the dynamical symmetries (misleadingly characterized), so in a case like this the spacetime geometry would not have its usual physical significance. Considerations of conformally invariant dynamics may force physicists to revise the spacetime geometry ascribed to the universe in different regimes. Yet it is hard to see how these cosmological considerations could force a resolution of philosophical debates regarding the status of spacetime geometry.

V. *Time's Arrow*

At various points above we assumed a distinction between the past and future, even though there is no basis for it in the austere conception of time outlined in § I. Finding a physical basis for this familiar aspect of our phenomenological experience of time has been the subject of a large literature.[16] Boltzmann's original proposals to solve the "problem of time's arrow," as it is now called, invoked speculative, obscure cosmological ideas, and in the ensuing 150 years of debate cosmology has continued to play a central role. After first formulating the problem, we will briefly consider the consensus view regarding how cosmology may contribute to a solution. We will then turn to the implications of cosmology for the arrows of time associated with electromagnetism and gravitation.

The problem arises due to an apparent conflict between the time-reversal invariance of fundamental laws of physics and the temporal asymmetry of observed phenomena. Consider the *history* of a particular system, as described by some physical theory, to be an assignment of an "instantaneous state" of the system over some time interval (that is, $t \mapsto D(t)$ where $D(t)$ is the state at time t) that is compatible with the laws of that theory. For example, a history of Bobby Thomson's 1951 home run to clinch the pennant would specify the state of the baseball --- in the simplest case, the position and velocity of the center of mass --- over the interval from, say, the ball leaving Branca's hand to its destination in the left-field stands. A theory is time reversal invariant (TRI) if the "time-reversed history" ($t \mapsto {}^{T}D(t)$) is also allowed. Here T operates on the history in two ways: it reverses the time order of instantaneous states and also "flips" the sign of some quantities in the instantaneous states. (Representing the "flip" as a reversal operator R, ${}^{T}D(t) = {}^{R}D(-t)$.) Thomson's home run in reverse would be a history with states in reverse sequence, where the velocity in each state has the opposite sign. The ball would fly towards Thomson's bat rather than toward the stands.[17] This trajectory would also be a possible solution of the equations governing the motion of projectiles, given that the relevant theory is, like nearly all fundamental theories, TRI.[18] Yet obviously we never see a baseball fly out of the stands, ricochet off the bat, and, with elegantly arranged speed and spin, shoot back into the pitcher's hand.

Time-asymmetric patterns of succession are a pervasive part of our experience. We can all readily distinguish histories that represent the kind of progression of states we ordinarily experience from their time-reversed counterparts, which are usually as odd as a home-run-in-reverse. This leads to a problem if we assume that every aspect of our phenomenological experience of time must be grounded in the physical laws. For the laws of a TRI theory are apparently too permissive, providing no physical basis for drawing this distinction. Note, however, that this way of formulating the problem presumes that the physical theory provides a *complete* account of the nature of time, such that its absence from the laws undercuts this aspect of our phenomenal experience. Furthermore, as Earman (1974) emphasized, generally one does not expect a particular history to reflect the symmetries of the laws; it is no surprise that a particular history can be time-asymmetric even though the laws are TRI. There needs to be further argument that the temporal asymmetry of experience should be explained in some deeper sense --- for example, because this aspect of experience is too fundamental to be treated as a contingent feature of a particular history.

The modern debate regarding time's arrow in statistical mechanics began with Boltzmann's claim to have derived time-asymmetric behavior from TRI laws. Boltzmann argued that a system initially in a non-equilibrium state would evolve towards equilibrium at later times. Consider, for example, a thermally isolated box at room temperature enclosing a warm cup of coffee. Supposing that at 9:00 the cup is warm, Boltzmann's argument implies that it should be cooler when retrieved at 9:30, as the system evolves towards equilibrium, in which the cup and air in the box have the same temperature. Given the TRI of the laws used in the derivation, however, it follows from the same argument that evolving *backwards* from the chosen state should also lead toward an equilibrium state. But this is incompatible with experience: it implies that the warm cup of coffee at 9:00 evolved *from* a cooler state at 8:30. Boltzmann claimed to have ruled out the second type of evolution but it was initially puzzling how the trick was done. After objections from Boltzmann's contemporaries, it eventually became clear that the derivation depended on smuggling in a subtle asymmetry in the boundary conditions.[19]

In defense of asymmetric boundary conditions Boltzmann shifted from consideration of isolated systems to the state of the universe as a whole. The second problematic evolution described above can be ruled out by assuming that the box and coffee cup were in a state further from equilibrium at 8:30 than at 9:00. But what about the state of the cup at 8:15? Pursuing this line of thought leads to ever-earlier states, and eventually to a postulate regarding the initial state. Boltzmann suggested two possible justifications of the choice of a far-from-equilibrium "initial state":[20] (1) on global scales the universe is in an equilibrium state, but the box (and everything else in the observable universe) is the result of an enormously improbable fluctuation to a far-from-equilibrium state, or (2) the universe began in a far-from-equilibrium state. Either case yields the desired conclusion: evolution since the initial state would be towards equilibrium, allowing for home runs but not their time reversed counterparts. Although Boltzmann preferred the first option, modern cosmology apparently lends support to the second proposal --- with the far-from-equilibrium initial state taken to hold "at the Big Bang."

This proposal faces an immediate objection: the state of the early universe seems *close* to an equilibrium state, rather than far from it, given (for example) the uniform geometry at early times revealed by the uniform temperature of the cosmic background radiation. The relevant question, however, is whether the state with all the relevant degrees of freedom --- including matter, radiation, *and the gravitational field* --- is close to an equilibrium state. Penrose, in particular, has argued that when gravity is included the early universe should be regarded as very far from equilibrium. Since gravity is a force of attraction, systems closer to equilibrium should be clumpier, with the equilibrium state given by "maximal clumpiness" – namely, a black hole. Hence the uniformity of the early universe is compatible with treating it as far-from-equilibrium. This line of argument qualifies as something like conventional wisdom, sometimes called the "Past Hypothesis" (following Albert 2000). The PH holds that modern cosmology provides exactly the sort of far-from-equilibrium initial state required for Boltzmann's solution to the problem of time's arrow.

We only have space here to highlight two critical points regarding this proposal, which is currently the subject of vigorous debate in foundations of physics. First, the line of thought described above regarding equilibrium states for gravitational systems is heuristic, at best. Difficulties with applying statistical mechanics to gravity arise even in Newtonian gravity and are amplified in GR (see Callender 2008). And in assessing ever-earlier states, the appropriate physical theory to employ to determine the equilibrium states --- namely, quantum gravity --- has not yet been formulated. It is unclear yet whether the open problems in applying statistical mechanics to gravity undercut the conventional wisdom regarding the PH.

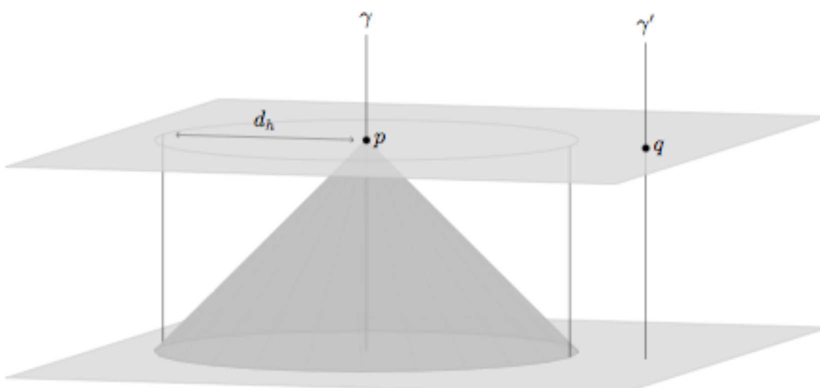
Second, even if we assume that solutions to the open problems will buttress the conventional wisdom, it is not clear that the PH resolves the original problem. Shifting from specifying local boundary conditions on subsystems to the boundary condition of the whole shebang only solves the original problem if the PH appropriately constrains the boundary conditions of the subsystems. Does the PH imply that the *initial* states of *subsystems* of the universe, such as our coffee cup, will (almost always) be far from equilibrium, as it must to account for the asymmetries of experience? Consider whether a "global" system consisting of, say, the observed universe can evolve towards equilibrium while isolated subsystems display anti-thermodynamic behavior. There is a lot of work to be done in ruling out anti-thermodynamic behavior of subsystems, and it is doubtful whether the PH itself provides sufficient grounds for doing so.[21]

Cosmology also has implications for the temporal asymmetry observed in electromagnetic phenomena. Rather than Thomson's home run, consider the antenna broadcasting news of the "shot heard round the world" and consider the radiation emitted from the antenna --- some of which is absorbed, and some of which escapes into the vast reaches of space. The time reverse of this process describes radiation coming in from the absorbers and space, elegantly contrived to converge on the antenna and excite coordinated oscillations among its electrons. Insofar as the radio broadcast involves collective phenomena there may not be a clear contrast between an "electromagnetic" and "thermodynamic" arrow. But restricting attention to a single electron and the electromagnetic field may suffice to isolate an electromagnetic arrow, similar in nature to "arrows" associated with other kinds of wave propagation. Just as we observe stones producing surface waves in water but not the time reverse, we observe sources such as an electron decelerating and emitting radiation, but not the time reverse.

The posit comparable to the PH in this case is called the Sommerfeld radiation condition, which prohibits "source-free radiation" such as that involved in the "anti-broadcast." Any given solution of Maxwell's equation can be decomposed, according to the Kirchoff representation theorem, into terms representing the contribution from sources within a given volume of spacetime and radiation registered on the boundary of that region. Consider the retarded representation, which takes the field at a point to be fixed by the sources where they intersect with the past light cone of the point and a surface term representing incoming radiation where the light cone is cut off at some specified earlier time.[22] The Sommerfeld radiation condition requires, in the limit as the volume of the spacetime region goes to infinity, that there is no incoming radiation registered on the boundary of the region. This is usually justified as a way of implementing the idea that radiation must have sources, as it rules out source-free incoming radiation.

This way of formulating a temporal asymmetry for electromagnetism does not extend to all cosmological models.[23] This is a consequence of the existence of particle horizons in models with a spacelike singularity. Roughly speaking, the particle horizon delimits the region from which signals traveling at or below the speed of light can reach a given point.[24] The Sommerfeld radiation condition cannot be applied in this context because the field at a point still depends on charges lying beyond the horizon. Explicitly, consider a charged particle whose worldline γ' never intersects the past lightcone at a point p (see figure 2). Although γ' is beyond p's horizon, since two of Maxwell's field equations are elliptic equations constraining the value of the electromagnetic field on a spacelike surface, the presence of γ' will be registered in the field value at p.

INSERT FIGURE 2 HERE (SmeenkParticleHorizons)



CAPTION
Figure 2: The curve γ' is beyond the horizon (the distance d_h) for a particle traveling along the curve γ, at point p.

There is a natural analog of the Sommerfeld radiation condition in the case of gravitation: Penrose (1979)'s "Weyl curvature hypothesis," according to which the Weyl curvature goes to zero (in an way that can be precisely specified) as $t \rightarrow 0$. The Weyl curvature tensor represents the "free gravitational field," and this hypothesis requires that there is no source-free gravitational radiation. As with the Sommerfeld radiation condition, accepting the hypothesis explicitly breaks time symmetry.

VI. Conclusion

It is striking how far cosmology has come in clarifying the empirical implications of long-standing questions about the nature of time. One can acknowledge this as important progress without accepting a stronger claim, that the only meaningful questions about time are the ones that can be posed in the precise language of relativistic cosmology. Yet these developments within physics have shown that many seemingly clear questions rest on subtle conceptual confusions. As a way of summarizing some of the discussion above, let me emphasize two general themes regarding the transformation that our initial questions have undergone. First, questions apparently about "time" are really questions about spacetime in relativity, and spatial and temporal structure can be cleanly separated only in quite special circumstances. Since spacetime geometry depends on how matter and motion are arranged in GR, questions about the nature of time are transformed into questions about the matter filling the universe and how it moves. Second, cosmology brings out an interesting interplay between local and global features of time. A globally defined cosmic time function generally doesn't correspond to the time elapsed on an observer's wristwatch. More significantly, features of cosmological models such as singularities have resisted local analysis. And the assessment of the past hypothesis as an explanation of observed temporal asymmetries depends on establishing that a global claim regarding the initial state constrains appropriately the initial states of subsystems of the universe. Finally, it is clear that the third act in the ongoing story of the development of the concept of time, which will take up where the account of section I left off, will involve significant further transformations. Especially with regard to questions regarding the big bang and early universe, the concepts of spacetime geometry central to GR are not expected to apply. Philosophers should welcome the challenging work that will come with clarifying the nature of time in successor theories.

References

Albert, David (2000). *Time and Chance*. Cambridge: Harvard University Press.

Belot, Gordon (2001). "Geometry and Motion," *British Journal for the Philosophy of Science* (**51**): 561-595. Reprinted in P. Clark and K. Hawley (eds.), *Philosophy of Science Today.* Oxford University Press (2003), pp. 201-235.

Belot, Gordon (2005). "Dust, Time and Symmetry," *British Journal for the Philosophy of Science* (**56**): 255-291.

Bojowald, Martin (2011). *Once Before Time: A Whole Story of the Universe.* New York: Vintage. (Originally published in German, translated by the author.)

Callender, Craig (2008). "The Past Hypothesis Meets Gravity," in *Time, Chance, and Reduction*, ed. By Gerhard Ernst and Andreas Hütteman. Cambridge: Cambridge University Press.

Callender, Craig, (2011). "Thermodynamic Asymmetry in Time," The Stanford Encyclopedia of Philosophy (Fall 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2011/entries/time-thermo/>

Callender, Craig and Huggett, Nick (2001), editors. *Physics Meets Philosophy at the Planck Scale*. Cambridge: Cambridge University Press.

Curiel, Erik. (1999). "The Analysis of Singular Spacetimes," *Philosophy of Science* **66**: S119 - 145.

DiSalle, Robert. (2006). *Understanding Spacetime*. Cambridge: Cambridge University Press.

Earman, John (1974). "An Attempt to Add a Little Direction to 'The Problem of the Direction of Time'," *Philosophy of Science* (**41**): 15-47.

Earman, John (1995). *Bangs, Crunches, Whimpers and Shrieks*. Oxford: Oxford University Press.

Earman, John (2002). "What time reversal invariance is and why it matters," *International Studies in the Philosophy of Science* (**16**): 245-264.

Earman, John (2006). "The 'Past Hypothesis': Not even false," *Studies in the History and Philosophy of Modern Physics* (**37**): 399-430.

Ellis, G. F. R. and Sciama, Dennis (1972). "Global and Non-Global Problems in Cosmology," in *General Relativity: Papers in Honor of J. L. Synge*, edited by L. O'Raifertaigh. Oxford: Oxford University Press, pp. 35-59.

Gödel, Kurt (1949). "A remark about the relationship between relativity theory and idealistic philosophy," in P. A. Schilpp (ed.), *Albert Einstein: Philosopher Scientist.* Chicago: Open Court, pp. 557-562.

Janis, Al (2010). "Conventionality of Simultaneity," *The Stanford Encyclopedia of Philosophy* (Winter 2012 edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/spacetime-convensimul/>.

Lorentz, Hendrik A., Albert Einstein, Hermann Minkowski, and Hermann Weyl (1952). *The Principle of Relativity*. Translations by W. Perrett and G. B. Jeffrey. New York: Dover.

Malament, David (1995). "Introductory note to *1949b," in S. Feferman et al. (eds.), *Kurt Gödel: Collected Works, Volume III*. Oxford: Oxford University Press, pp. 261-269.

Malament, David (2004). "On the time reversal invariance of classical electromagnetic theory," *Studies in the History and Philosophy of Modern Physics* (**35**): 295-315.

Malament, David (2007). "Classical General Relativity," in Butterfield, J. and Earman, J. (eds.), *Handbook of the Philosophy of Science. Volume 2: Philosophy of Physics.* Amsterdam: Elsevier, pp. 229-274.

Penrose, Roger (1979). "Singularities and time-asymmetry," in Hawking, S. W. and Israel, W. (eds.), *General Relativity: An Einstein Centenary Survey*. Cambridge: Cambridge University Press, pp. 581-638.

Ellis, George F. R. and Rothman, Tony (1993). "Lost Horizons," *American Journal of Physics* **61**: 883- 893.

Sachs, Robert G. (1987). *The Physics of Time Reversal*. Chicago: University of Chicago Press.

Sklar, Lawrence (1993). *Physics and Chance*. Cambridge: Cambridge University Press.

Steinhard, Paul J. and Turok, Neil (2007). *Endless Universe: Beyond the Big Bang*. New York: Broadway Books.

Torretti, Roberto (2000). "On relativity, time reckoning, and the topology of time series," in Butterfield, J. (ed.), *Arguments of Time*. Oxford: Oxford University Press, pp. 65-82.

Torretti, Roberto (2007). "The Problem of Time's Arrow Historico-Critically Re-Examined," *Studies in the History and Philosophy of Modern Physics* (**38**): 732-756.

Uffink, Jos. "Compendium of the foundations of classical statistical physics," in Butterfield, J. and Earman, J. (eds.), *Handbook of the Philosophy of Science. Volume 2: Philosophy of Physics.* Amsterdam: Elsevier, pp. 923-1074.

Winsberg, Erik (2004). "Can Conditioning on the "Past Hypothesis" Militate Against the Reversibility Objections?" *Philosophy of Science* (**71**): 489-504.

---

[1] More precisely, the "time line" is an affine space: it lacks a preferred origin and other algebraic structures associated with $\Re$.

[2] The set includes inertial observers; an observer moving uniformly (at constant velocity) with respect to a given inertial observer will also be an inertial observer. See, e.g., Belot (2001) or DiSalle (2006) for further discussion of Galilean relativity, inertial frames, and the spacetime geometry appropriate for Newtonian theory.

[3] See Janis (2010) for a survey of the conventionality of simultaneity, a topic that we are setting aside.

[4] See, for example, Malament (2007) for a concise introduction to general relativity for philosophers.

[5] Note that from the standpoint of general relativity it is misleading to speak of "gravitational force" or "acceleration due to gravity": gravity is not a *force* that causes a body to accelerate and depart from an inertial trajectory. The effects of gravity are included in defining freely falling trajectories (the replacement for inertial trajectories), and non-zero net non-gravitational force yields acceleration and departure from such trajectories.

[6] Roughly speaking, homogeneity requires that at a given moment of cosmic time every spatial point "looks the same," and isotropy holds if there are no geometrically preferred spatial directions. Assuming that the models are simply connected, these requirements imply that the models are topologically $\Sigma \times \Re$.

[7] The analogy is borrowed from Malament (1995), which is an excellent source for further discussion of Gödel's arguments concerning the nature of time and the properties of the model he discovered. Gödel's model has the further property that there are no global slices, which follows from the fact that it is topologically $\Re^4$ and compact.

[8] This is not to be confused with the relativity of simultaneity familiar from special relativity; the breakdown isn't observer dependent, in that there is not a naturally preferred way to choose a foliation based on observer's state of motion. Furthermore, the hypersurfaces of simultaneity for an observer do not in general match the surfaces of constant cosmic time.

[9] This suggestion is elaborated and defended by Curiel (1999); which is in part a response to Earman (1995)'s discussion of singularities.

[10] A *slice* of a spacetime (an achronal, edgeless surface) is said to be a *Cauchy surface* if its domain of dependence is the entire spacetime. The domain of dependence of a slice consists of the points in spacetime such that all (past or future) inextendible causal curves through the point intersect the slice. In a globally hyperbolic spacetime, initial data for

the gravitational field specified on a Cauchy surface determine the full solution (up to diffeomorphism).

[11] Given that the bound can be formulated in terms of the length of a timelike geodesic going backward from a particular instant, this conclusion does not depend on singling out a unique cosmic time function.

[12] Add some detail here, e.g. Sandage's attempts to measure the deceleration parameter.

[13] A given spacetime is an extension of a second spacetime if the latter can be mapped into a subset of the former, while preserving relevant metrical structure. A spacetime is said to be maximal if there are no proper extensions of it. Both are relative to a choice of what mathematical conditions to impose; one can loosen these and allow more extensions.

[14] For popular level discussions of these two approaches, with references to pursue further, see Bojowald (2011) and Steinhardt and Turok (2007).

[15] Penrose assumes that the universe contains dark energy and asymptotically approaches a De Sitter solution in the far future. Electromagnetic and gravitational radiation are both massless fields governed by conformally invariant dynamics, but Penrose's proposal has to assume that electrons also decay, somehow, in order for the dynamics to be fully conformally invariant.

[16] See, in particular, Callender (2011) for a clear survey of these debates.

[17] Albert (2000) defends the heretical position that various quantities typically taken to be "flipped" by the time-reversal operation should not be, based on considerations regarding the "logical or conceptual" dependence of the quantities on time. Velocity is flipped because he takes it to be non-fundamental quantity that is defined as the time derivative of position. But he argues, for example that in electromagnetism the magnetic field should not change sign under time reversal, with the consequence that electromagnetism fails to be "Albert-TRI." Other physical theories judged to by TRI by everyone else are also not Albert-TRI. However, Albert's treatment of electromagnetism fails to acknowledge the fundamental contrast between electric and magnetic fields --- the former are polar vector fields, whereas the latter are *axial* vector fields, which change sign under change of parity or time reversal. See Earman (2002) and Malament (2004) for discussion and defense of the conventional definition of TRI.

[18] There is evidence from the decay of neutral K mesons, combined with the CPT theorem, for the failure of time reversal invariance in the theory describing the weak interaction. (See Sachs 1987 for a discussion.) The existence of non-TRI laws poses a challenge to the position that time is symmetric, but such laws are not relevant to the current debate if, as is quite plausible, the physical laws relevant to describing the baseball's trajectory are TRI.

[19] See Uffink (2007) for a masterful recent discussion of the debate among Boltzmann and his contemporaries, and Sklar (1993) for an earlier survey of issues in statistical mechanics. In the text I have avoided introducing the quantity "entropy" for ease of exposition; in terms of the Boltzmann entropy for a state, the debate regards why entropy increases with time for a thermally isolated system.

[20] "Initial state" is a misleading term as there is not a "first moment" at which the appropriate state is specified, but it suffices for this discussion to consider a state

specified after the initial fluctuation to a far-from equilibrium state (option 1) or in the early universe (option 2).

[21] See, in particular, Winsberg (2004) for an argument that ruling out the possibility of anti-thermodynamic subsystems requires an implausible principle in addition to the PH.

[22] One can instead consider the advanced representation, and reverse all temporal language in the rest of the paragraph (past goes to future, sources goes to sinks, etc.). Choosing the formulation of the Sommerfeld radiation condition described in the text as an explicit law of electromagnetism explicitly breaks TRI. (One can similarly formulate a "no sink-free outgoing radiation condition" that also breaks TRI.)

[23] See Ellis and Sciama (1972) for a clear discussion of this issue.

[24] See Ellis and Rothman (1993) for an introduction to horizons in relativistic cosmology.