ASPECTS OF DETERMINISM IN MODERN PHYSICS

John Earman

1 INTRODUCTION

The aims of this chapter are to review some aspects of determinism that are familiar to physicists but are little discussed in the philosophical literature and to show how these aspects connect determinism to issues about symmetries in physics, the structure and ontological status of spacetime, predictability, and computability.[1] It will emerge that in some respects determinism is a robust doctrine and is quite hard to kill, while in other respects it is fragile and requires various enabling assumptions to give it a fighting chance. It will also be seen that determinism is far from a dead issue. Whether or not ordinary non-relativistic quantum mechanics (QM) admits a viable deterministic underpinning is still a matter of debate. Less well known is the fact that in some cases QM turns out to be more deterministic than its classical counterpart. Quantum field theory (QFT) assumes determinism, at least at the classical level, in order to construct the field algebra of quantum observables. Determinism is at the heart of the cosmic censorship hypothesis, the most important unsolved issue in classical general relativity theory (GTR). And issues about the nature and status of determinism lie at the heart of key foundation issues in the search for a theory of quantum gravity.

2 PRELIMINARIES

2.1 The metaphysics of determinism

The proposal is to begin by getting a grip on the doctrine of determinism as it was understood pre-GTR and pre-QM, and then subsequently to try to understand how the doctrine has to be adjusted to accommodate these theories. In pre-GTR physics, spacetime serves as a fixed background against which the drama of physics is enacted. In pre-QM physics it was also assumed that there is a set $\mathcal{O}$ of genuine physical magnitudes (a.k.a. "observables") each of which takes a determinate

[1]Recent surveys of determinism are found in Butterfield [1998], Earman [2004a], and Hoefer [2003]. A collection of articles on various aspects of determinism is found in Atmanspacher and Bishop [2002].

Handbook of the Philosophy of Science. Volume 2: Philosophy of Physics
Volume editors: John Earman and Jeremy Butterfield. Handbook editors: Dov M. Gabbay, Paul Thagard and John Woods.

value at every moment of time; call these the *occurrent* magnitudes. Other physical magnitudes may be dispositional in character and may take on determinate values only in appropriate contexts; but it was assumed that these dispositional magnitudes supervene on the nondispositional magnitudes.[2] A *history* H is a map from $\mathbb{R}$ to tuples of values of the basic magnitudes, where for any $t \in \mathbb{R}$ the *state* $H(t)$ gives a snapshot of behavior of the basic magnitudes at time t. The world is *Laplacian deterministic* with respect to $\mathcal{O}$ just in case for any pair of histories H_1, H_2 satisfying the laws of physics, if $H_1(t) = H_2(t)$ for some t, then $H_1(t) = H_2(t)$ for all t.

Several remarks are in order. First, the 't' which appears in the above definition is supposed to be a *global time function*. This notion can be defined in a manner that applies to classical, special relativistic, and general relativistic spacetimes: a global time function is a smooth map $t : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is the spacetime manifold, such that for any $p, q \in \mathcal{M}$, $t(p) < t(q)$ just in case there is a future directed timelike curve from p to q.[3] In classical spacetimes, all of which possess an absolute (or observer independent) notion of simultaneity, a timelike curve is one which is oblique to the planes of absolute simultaneity. And the levels $t = const$ of a global time function must coincide with the planes of simultaneity; thus, in the classical setting t is determined up to a transformation of the form $t \to t' = t'(t)$. In the relativistic setting a timelike curve is one whose tangent at any point lies inside the light cone at that point. In causally pathological general relativistic spacetimes (e.g. Gödel spacetime — see Section 6.1) there can be no global time function, and the global sense of Laplacian determinism as defined above makes no sense.[4] But if one global time function exists for a relativistic spacetime, then many exist. A poor choice of global time function can lead to the failure of Laplacian determinism on the above definition. Thus, in the relativistic setting, the definition of determinism must be construed as applying to a suitable choice of time function, the nature of which will be clarified below.

Second, the above formulation of determinism assumes a distinction between laws of nature on one hand and initial/boundary conditions on the other. Where this distinction becomes mushy, so does the doctrine of determinism. There is a

[2]The general idea of supervenience is that X supervenes on Y iff within the range of possible cases, there is no difference in X without a difference in Y. The strength and type of supervenience depends on what are counted as possible cases. Here the concern is mainly with physical supervenience where the possible cases are those compatible with the laws of physics.

[3]This definition presupposes that the spacetime is temporally orientable and that one of the orientations has been singled out as giving the future direction of time. The first presupposition is satisfied for classical and special relativistic spacetimes. A general relativistic spacetime (see [Malament, this vol.]) may not be temporally orientable, but a covering spacetime always is since temporal orientability fails only if the spacetime is not simply connected. The second presupposition implies that some solution to the problem of the direction of time has been found (see [Uffink, this vol.]).

[4]A necessary and sufficient condition for the existence of a global time function for a relativistic spacetime is *stable causality* which (roughly speaking) requires that there exists a widening of the null cones that does not result in closed timelike curves; for a precise definition, see [Wald, 1984, 198-199]. Not only does Gödel spacetime not admit a global time function, it does not admit any global time slices (i.e. spacelike hypersurfaces without edges).

huge philosophical literature on laws of nature.[5] Since most of it is unilluminating when it comes to understanding the nature and function of laws in the practice of physics, it will be largely ignored here. For present purposes I will simply stipulate that an acceptable account of laws must satisfy the empiricist constraint that the laws supervene on the totality of non-modal, particular facts.[6] Philosophers like to speculate about non-empiricist laws; but such entities, should they exist, would seem to be beyond the ken of science, and as such they are irrelevant for present purposes. I prefer David Lewis' [1973, 72-77] way of fulfilling the empiricist constraint since it connects the account of laws to the practice of physics: the laws of physics are the axioms or postulates that appear in the ideal theory of physics, where the ideal theory is the one that, within the class of true theories, achieves the best balance between simplicity and information content. All of the historical examples we have of candidates for deterministic laws involve a relatively small subset $\mathcal{B} \subset \mathcal{O}$ of basic occurrent magnitudes, the assumption being that the remaining ones supervene on those of $\mathcal{B}$.[7] This is hardly surprising if, as has been claimed, simplicity is a crucial feature of physical laws. Hermann Weyl shared the conviction that simplicity must figure into an account of laws, but he noted that "this circumstance is apt to weaken the metaphysical power of determinism, since it makes the meaning of natural law depend on the fluctuating distinction between simple and complicated functions or classes of functions" [1932, 42]. This is, I think, a consequence that has to be swallowed and digested. Philosophers who are answerable only to their armchairs are free to think otherwise.

Third, it is conceptually possible that the world could be partially deterministic, i.e. deterministic with respect to partial histories defined by the values of magnitudes in some proper subset $\mathcal{D} \subset \mathcal{O}$ of the occurrent physical magnitudes but non-deterministic with respect to partial histories defined by the values of magnitudes in some other proper subset $\mathcal{N} \subset \mathcal{O}$. But it is hard to imagine a scenario in which this could happen if both $\mathcal{D}$ and $\mathcal{N}$ are basic magnitudes. For in order that the non-deterministic evolution of the elements $\mathcal{N}$ not upset deterministic evolution for $\mathcal{D}$, the magnitudes in $\mathcal{N}$ must not interact with those in $\mathcal{D}$, or else there would have to be a conspiracy in which the upsetting effects of the $\mathcal{N}$ magnitudes on $\mathcal{D}$ cancel out, which is operationally the same. However, this plausibility consideration fails to operate when the $\mathcal{N}$ are non-basic magnitudes; in particular, as discussed below, stochastic processes on one level can supervene on deterministic processes at a lower level (see [Butterfield, 1998]). This fact makes the inference from observed stochastic behavior to indeterminism fraught with peril.

Fourth, the laws of physics typically take the form of differential equations, in which case the issue of Laplacian determinism translates into the question of whether the equations admit an *initial value formulation*, i.e. whether for ar-

[5]For an overview of different accounts of laws of nature, see [Carroll, 2004].

[6]This is what David Lewis has termed "Humean supervenience" with regards to laws of nature; for a defense, see [Earman and Roberts, 2006].

[7]For example, in classical particle mechanics the elements of $\mathcal{B}$ are the positions and momenta of the particles, and it is assumed that any other genuine mechanical magnitude can be expressed as a functional of these basic magnitudes.

bitrary initial data there exists a unique solution agreeing with the given initial data.[8] What counts as initial data depends on the details of the case, but typically it consists of the instantaneous values of the independent variables in the equations, together with the instantaneous values of a finite number of time derivatives of these variables. "Arbitrary" initial data might be thought to include any kinematically possible values of the relevant variable — as with the initial values of particle positions and velocities in Newtonian mechanics — but "arbitrary" must be taken to mean arbitrary within the bounds of compatibility with the equations of motion, which may impose non-trivial constraints on the initial data. This leads to the next remark.

Fifth, in the relativistic setting, field equations often factor into constraint equations, which place restrictions on the initial data, and the evolution equations, which govern how initial data satisfying the constraint equations evolve over time — Maxwell's equations for electromagnetism and Einstein's gravitational field equations being prime examples. In these instances the evolution equations guarantee that once the constraint equations are satisfied they continue to be satisfied over time. This should be a feature of deterministic equations, for if the data at some future time in the unique solution picked out by the initial data do not satisfy the constraints, then the laws are self-undermining. It could be argued that a basic feature of time in relativistic worlds — perhaps the key feature that separates the time dimension from the space dimensions — lies precisely in this separation of evolution and constraint equations.[9]

Sixth, while there is no *a priori* guarantee that the laws of the ideal theory of physics will be deterministic, the history of physics shows that determinism is taken to be what might be termed a 'defeasible methodological imperative': start by assuming that determinism is true; if the candidate laws discovered so far are not deterministic, then presume that there are other laws to be discovered, or that the ones so far discovered are only approximations to the correct laws; only after long and repeated failure may we entertain the hypothesis that the failure to find deterministic laws does not represent a lack of imagination or diligence on our part but reflects the fact that Nature is non-deterministic. An expression of this sentiment can be found in the work of Max Planck, one of the founders of quantum physics: determinism (a.k.a. the law of causality), he wrote, is a "heuristic principle, a signpost and in my opinion the most valuable signpost we possess, to guide us through the motley disorder of events and to indicate the direction in which scientific inquiry should proceed in order to attain fruitful results" [1932, 26; my translation].[10]

[8]And as will be discussed below, there are further issues, such as whether the solution depends continuously on the initial data.

[9]See [Callender, 2005] and [Skow, 2005] for defenses of related views on the difference between space and time.

[10]For a history of the debates about the status of determinism among the founding fathers of QM, see [Cushing, 1994] and [Stöltzner, 2003].

2.2 Varieties of determinism

There is a tendency in the philosophical literature to fixate on the Laplacian variety of determinism. But other kinds of determinism crop up in physics. For example, some processes are described by delay-differential equations for which instantaneous initial data may not suffice to single out a unique solution. A simple example is given by the first order ordinary differential equation (ode) $\dot{x}(t) = x(t - C)$ with a constant delay $C > 0$. Laplacian determinism fails since given initial data $x(0)$ is compatible with multiple solutions. However, a near cousin of Laplacian determinism holds since a specification of $x(t)$ for the interval of time $t \in [-C, 0]$ fixes a unique solution.[11] If the constant delay C replaced by a function $\tau(t)$ of t which is unbounded, or if the delay-differential equation has a more complicated form than in the simple example, then even the weakened forms of Laplacian determinism can fail. An illustration of the latter is given by the equation of motion is $\dot{x}(t) = f(t)x(t-1)$ where $f(t)$ is a continuous function that vanishes outside of $[0, 1]$ and satisfies $\int f(t)dt = -1$. Raju [1994, 120ff] gives an example of an f such that unless $x(t)$ is identically 0 for all $t \geq 1$, the equation of motion admits no solutions for $t < 0$; whereas if $x(t)$ is identically zero for $t \geq 1$, then the equation of motion admits an infinity of solutions for $t < 0$. Changing the delay term $x(t-1)$ in this example to an advance term $x(t+1)$ produces an example where an entire past history fails to fix a unique future solution. Very little is known about the initial value problem for what is probably the most important physical application of delay/advance differential equations; namely, charged particles moving under their mutual retarded/advanced interactions.[12]

For sake of definiteness, fix on the Laplacian variety of determinism. Within this variety there is a distinction between future and past determinism. *Past Laplacian determinism* means that for any pair of histories H_1, H_2 satisfying the laws of physics, if $H_1(t) = H_2(t)$ for some t, then $H_1(t') = H_2(t')$ for all $t' > t$. *Future Laplacian determinism* is defined analogously. In principle, Laplacian determinism can hold in one direction of time but not in the other. However, if the laws of motion are time reversal invariant, then future and past determinism stand or fall together. Time reversal invariance is the property that if H is a history satisfying the laws, then so is the 'time reverse' history H^T, where $H^T(t) := {}^R H(-t)$ and where 'R' is the reversal operation that is defined on a case-by-case basis, usually by analogy with classical particle mechanics where $H(t) = (\mathbf{x}(t), \mathbf{p}(t))$, with $\mathbf{x}(t)$ and $\mathbf{p}(t)$ being specifications respectively of the particle positions and momenta at t, and ${}^R H(t) = (\mathbf{x}(t), -\mathbf{p}(t))$.[13] Since all of the plausible candidates for fundamental laws of physics, save those for the weak interactions

[11] See [Driver, 1977] for relevant results concerning delay-differential equations.

[12] Driver [1979] studied the special case of identically charged particles confined to move symmetrically on the x-axis under half-retarded and half-advanced interactions. He showed that, provided the particles are sufficiently far apart when they come to rest, a unique solution is determined by their positions when they come to rest.

[13] A different account of time reversal invariance is given in [Albert, 2000, Ch. 1]; but see [Earman, 2002] and [Malament, 2004].

of elementary particles, are time reversal invariant, the distinction between past and future determinism is often ignored.

This is the first hint that there are interesting connections between determinism and symmetry properties.[14] Many other examples will be encountered below, starting with the following section.

2.3 Determinism and symmetries: Curie's Principle

The statement of what is now called 'Curie's Principle' was announced in 1894 by Pierre Curie:

> (CP) When certain effects show a certain asymmetry, this asymmetry must be found in the causes which gave rise to it. [Curie 1894, 401]

Some commentators see in this Principle profound truth, while others see only falsity, and still others see triviality (compare [Chalmers, 1970]; [Radicati, 1987]; [van Fraassen, 1991, 23–24], and [Ismael, 1997]). My reading of (CP) makes it a necessary truth. It takes (CP) to assert a conditional:

If

> (CP1) the laws of motion governing the system are deterministic; and (CP2) the laws of motion governing the system are invariant under a symmetry transformation; and (CP3) the initial state of the system is invariant under said symmetry

then

> (CP4) the final state of the system is also invariant under said symmetry

When the first clause (CP1) in the antecedent holds, the second clause (CP2) can be understood as follows: if an initial state is evolved for a period Δt and then the said symmetry is applied to the (unique) evolved state, the result is the same as first applying the symmetry to the initial state and evolving the resulting state for a period Δt. With this understanding, the reader can easily derive (CP4) from (CP1)-(CP3). Concrete instantiations of Curie's principle at work in classical and relativistic physics can be found in [Earman, 2004b]. An instantiation for GTR is mentioned in Section 6.3 below.[15]

Although (CP) is a necessary truth, it is far from a triviality since it helps to guide the search for a causal explanation of an asymmetry in what is regarded as the final state of system: either the asymmetry is already present in the initial state; or else the initial state is symmetric and the asymmetry creeps in over time,

[14] See [Brading and Castellani, this vol.] for a discussion of symmetries and invariances in modern physics.

[15] For additional remarks on Curie's principle, see [Brading and Castellani, this vol.].

either because the laws that govern the evolution of the system do not respect the symmetry or because they are non-deterministic. If, as is often the case, the latter two possibilities are ruled out, then the asymmetry in the final state must be traceable to an asymmetry in the initial state. It is also worth noting that the use of (CP) has ramifications for the never ending debate over scientific realism; for the asymmetry in the initial state may be imperceptible not only to the naked eye but to any macroscopic means of detection.[16]

3 DETERMINISM AND INDETERMINISM IN CLASSICAL PHYSICS

3.1 The hard road to determinism in classical physics

Classical physics is widely assumed to provide a friendly environment for determinism. In fact, determinism must overcome a number of obstacles in order to achieve success in this setting. First, classical spacetime structure may not be sufficiently rich to support Laplacian determinism for particle motions. Second, even if the spacetime structure is rich, uniqueness can fail in the initial value problem for Newtonian equations of motion if the force function does not satisfy suitable continuity conditions. Third, the equations of motion that typically arise for classical particles plus classical fields, or for classical fields alone, do not admit an initial value formulation unless supplementary conditions are imposed. Fourth, even in cases where local (in time) uniqueness holds for the initial value problem, solutions can break down after a finite time.

The following subsection takes up the first of these topics — the connection between determinism and the structure and ontology of classical spacetimes. The others are taken up in due course.

3.2 Determinism, spacetime structure, and spacetime ontology

Here is the (naive) reason for thinking that neither Laplacian determinism nor any of its cousins stands a chance unless supported by enough spacetime structure of the right kind. Assume that the (fixed) classical spacetime background is characterized by a differentiable manifold $\mathcal{M}$ and various geometric object fields $O_1, O_2, ..., O_M$ on $\mathcal{M}$. And assume that the laws of physics take the form of equations whose variables are the O_i's and additional object fields $P_1, P_2, ..., P_N$ describing the physical contents of the spacetime. (For the sake of concreteness, the reader might want to think of the case where the P_j's are vector fields whose integral curves are supposed to be the world lines of particles.) A symmetry of the spacetime is a diffeomorphism d of $\mathcal{M}$ onto itself which preserves the background structure given by the O_i's — symbolically, $d^*O_i = O_i$ for all values if i, where

[16]For a more detailed discussion of Curie's Principle and its connection to spontaneous symmetry breaking in quantum field theory see [Earman, 2004b]; for spontaneous symmetry breaking in quantum statistical physics, see [Emch, this vol.].

d^* denotes the drag along by d.[17] By the assumption on the form of the laws, a spacetime symmetry d must also be a symmetry of the laws of motion in the sense that if $\langle \mathcal{M}, O_1, O_2, ..., O_M, P_1, P_2, ..., P_N \rangle$ satisfies the laws of motion, then so does $\langle \mathcal{M}, O_1, O_2, ..., O_M, d^*P_1, d^*P_2, ..., d^*P_N \rangle$.[18]

Now the poorer the structure of the background spacetime, the richer the spacetime symmetries. And if the spacetime symmetry group is sufficiently rich, it will contain elements that are the identity map on the portion of spacetime on or below some time slice $t = const$ but non-identity above. We can call such a map a 'determinism killing symmetry' because when applied to any solution of the equations of motion, it produces another solution that is the same as the first for all past times but is different from the first at future times, which is a violation of even the weakest version of future Laplacian determinism.

As an example, take *Leibnizian spacetime*,[19] whose structure consists of all and only the following: a notion of absolute or observer-independent simultaneity; a temporal metric (giving the lapse of time between non-simultaneous events); and a Euclidean spatial metric (giving the spatial distance between events lying on a given plane of absolute simultaneity). In a coordinate system (x^α, t), $\alpha = 1, 2, 3$ adapted to this structure, the spacetime symmetries are

$$\text{(1)} \quad \begin{array}{lcll} x^\alpha & \rightarrow & x'^\alpha = R^\alpha_\beta(t)x^\beta + a^\alpha(t) & \qquad \alpha, \beta = 1, 2, 3 \\ t & \rightarrow & t' = t + const & \end{array}$$

where $R^\alpha_\beta(t)$ is an orthogonal time dependent matrix and the $a^\alpha(t)$ are arbitrary smooth functions of t. Clearly, the symmetries (1) contain determinism killing symmetries.

It is also worth noting that if the structure of spacetime becomes very minimal, no interesting laws of motion, deterministic or not, seem possible. For example, suppose that the time metric and the space metric are stripped from Leibnizian spacetime, leaving only the planes of absolute simultaneity. And suppose that the laws of physics specify that the world is filled with a plenum of constant mass dust particles and that the world lines of these particles are smooth curves that never cross. Then either every smooth, non-crossing motion of the dust is allowed by the laws of motion or none is, for any two such motions are connected by a symmetry of this minimal spacetime.

Two different strategies for saving determinism in the face of the above construction can be tried. They correspond to radically different attitudes towards

[17]A diffeomorphism d of the manifold $\mathcal{M}$ is a one-one mapping of $\mathcal{M}$ onto itself that preserves $\mathcal{M}$'s differentiable structure. For the sake of concreteness, assume that d is C^∞.

[18]For on the assumption that the laws are (say) differential equations relating the O_i and P_j, they cannot be sensitive to the "bare identity" of the points of $\mathcal{M}$ at which the O_i and P_j take some given values. This diffeomorphism invariance of the laws is one of the ingredients of what is called substantive general covariance (see section 6.2). One might contemplate breaking diffeomorphism invariance by introducing names for individual spacetime points; but the occurrence of such names would violate the "universal" character that laws are supposed to have.

[19]The details of various classical spacetime structures are to be found in [Earman, 1989].

the ontology of spacetime. The first strategy is to beef up the structure of the background spacetime. Adding a standard of rotation kills the time dependence in $R^{\alpha}_{\beta}(t)$, producing what is called *Maxwellian spacetime*. But since the $a^{\alpha}(t)$ are still arbitrary functions of t there remain determinism killing symmetries. Adding a standard of inertial or straight line motion linearizes the $a^{\alpha}(t)$ to $v^{\alpha}t + c^{\alpha}$, where the v^{α} and c^{α} are constants, producing *neo-Newtonian spacetime*[20] whose symmetries are given by the familiar Galilean transformations

$$(2) \quad \begin{aligned} x^{\alpha} &\rightarrow x'^{\alpha} = R^{\alpha}_{\beta}x^{\beta} + v^{\alpha}t + c^{\alpha} \qquad \alpha, \beta = 1, 2, 3. \\ t &\rightarrow t' = t + const \end{aligned}$$

The mappings indicated by (2) do not contain determinism killing symmetries since if such a map is the identity map for a finite stretch of time, no matter how short, then it is the identity map period. Note that this way of saving determinism carries with it an allegiance to "absolute" quantities of motion: in neo-Newtonian spacetime it makes good sense to ask whether an isolated particle is accelerating or whether an isolated extended body is rotating. To be sure, this absolute acceleration and rotation can be called 'relational' quantities, but the second place in the relation is provided by the structure of the spacetime — in particular, by the inertial structure — and not by other material bodies, as is contemplated by those who champion relational accounts of motion.

The second strategy for saving determinism proceeds not by beefing up the structure of the background spacetime but by attacking a hidden assumption of the above construction — the "container view" of spacetime. Picturesquely, this assumption amounts to thinking of spacetime as a medium in which particles and fields reside. More precisely, in terms of the above apparatus, it amounts to the assumption that $\langle \mathcal{M}, O_1, O_2, ..., O_M, P_1, P_2, ..., P_N \rangle$ and $\langle \mathcal{M}, O_1, O_2, ..., O_M, d^*P_1, d^*P_2, ..., d^*P_N \rangle$, where d is any diffeomorphism of $\mathcal{M}$ such that $d^*P_j \neq P_j$ for some j, describe different physical situations, even when d is a spacetime symmetry, i.e. $d^*O_i = O_i$ for all i. Rejecting the container view leads to (one form of) relationism about spacetime. A spacetime relationist will take the above construction to show that, on pain of abandoning the possibility of determinism, those who are relationists about motion should also be relationists about spacetime. Relationists about motion hold that talk of absolute motion is nonsensical and that all meaningful talk about motion must be construed as talk about the relative motions of material bodies. They are, thus, unable to avail themselves of the beef-up strategy for saving determinism; so, if they want determinism, they must grasp the lifeline of relationism about spacetime.

Relationalism about motion is a venerable position, but historically it has been characterized more by promises than performances. Newton produced a stunningly successful theory of the motions of terrestrial and celestial bodies. Newton's opponents promised that they could produce theories just as empirically adequate

[20]Full Newtonian spacetime adds a distinguished inertial frame — 'absolute space' — thus killing the velocity term in (2).

and as explanatorily powerful as his without resorting to the absolute quantities of motion he postulated. But mainly what they produced was bluster rather than workable theories.[21] Only in the twentieth century were such theories constructed (see [Barbour, 1974] and [Barbour and Bertotti, 1977]; and see [Barbour, 1999] for the historical antecedents of these theories), well after Einstein's GTR swept away the notion of a fixed background spacetime and radically altered the terms of the absolute vs. relational debate.

3.3 Determinism and gauge symmetries

When philosophers hear the word "gauge" they think of elementary particle physics, Yang-Mills theories, etc. This is a myopic view. Examples of non-trivial gauge freedom arise even in classical physics — in fact, we just encountered an example in the preceding subsection. The gauge notion arises for a theory where there is "surplus structure" (to use Michael Redhead's phrase) in the sense that the state descriptions provided by the theory correspond many-one to physical states. For such a theory a gauge transformation is, by definition, a transformation that connects those descriptions that correspond to the same physical state.

The history of physics shows that the primary reason for seeing gauge freedom at work is to maintain determinism. This thesis has solid support for the class of cases of most relevance to modern physics, viz. where the equations of motion/field equations are derivable from an action principle and, thus, the equations of motion are in the form of Euler-Lagrange equations.[22] When the Lagrangian is non-singular, the appropriate initial data picks out a unique solution of the Euler-Lagrange equations and Laplacian determinism holds.[23] If, however, the action admits as variational symmetries a Lie group whose parameters are arbitrary functions of the independent variables, then we have a case of underdetermination because Noether's second theorem tells us that the Euler-Lagrange equations have to satisfy a set of mathematical identities.[24] When these independent variables include time, arbitrary functions of time will show up in solutions to the Euler-Lagrange equations, apparently wrecking determinism.

The point can be illustrated with the help of a humble example of particle mechanics constructed within the Maxwellian spacetime introduced in the preceding subsection. An appropriate Lagrangian invariant under the symmetries of this spacetime is given by

$$(3) \quad L = \sum\sum\nolimits_{j<k} \frac{m_j m_k}{2M}(\dot{\mathbf{x}}_j - \dot{\mathbf{x}}_k)^2 - V(|\mathbf{x}_j - \mathbf{x}_k|), \qquad M := \sum_i m_i.$$

[21]Newton's opponents were correct in one respect: Newton's postulation of absolute space, in the sense of a distinguished inertial frame was not needed to support his laws of motion.

[22]See [Butterfield, this vol.] and [Belot, this vol.] for accounts of the Lagrangian and Hamiltonian formalisms.

[23]At least if the continuity assumptions discussed in Section 3.5 below are imposed.

[24]For an account of the Noether theorems, see [Brading and Brown, 2003] and [Brading and Castellani, this vol.].

This Lagrangian is singular in the sense that Hessian matrix $\partial^2 L/\partial\dot{\mathbf{x}}_i\partial\dot{\mathbf{x}}_j$ does not have an inverse. The Euler-Lagrange equations are

$$(4)\qquad \frac{d}{dt}\left(m_i(\dot{\mathbf{x}}_j - \frac{1}{M}\sum_k m_k\dot{\mathbf{x}}_k)\right) = \frac{\partial V}{\partial\dot{\mathbf{x}}_i}.$$

These equations do not determine the evolution of the particle positions uniquely: if $\mathbf{x}_i(t)$ is a solution, so is $\mathbf{x}'_i(t) = \mathbf{x}_i(t) + \mathbf{f}(t)$, for arbitrary $\mathbf{f}(t)$, confirming the intuitive argument given above for the apparent breakdown of determinism. Determinism can be restored by taking the transformation $\mathbf{x}_i(t) \to \mathbf{x}_i(t) + \mathbf{f}(t)$ as a gauge transformation.

The systematic development of this approach to gauge was carried out by P. A. M. Dirac in the context of the Hamiltonian formalism.[25] A singular Lagrangian system corresponds to a constrained Hamiltonian system. The *primary constraints* appear as a result of the definition of the canonical momenta. (In the simple case of a first-order Lagrangian $L(q, \dot{q}, t)$, where q stands for the configuration variables and $\dot{q} := dq/dt$, the canonical momentum is $p := \partial L/\partial\dot{q}$.) The *secondary constraints* arise as a consequence of the demand that the primary constraints be preserved by the motion. The total set of constraints picks out the *constraint surface* $\mathcal{C}(q,p)$ of the Hamiltonian phase space $\Gamma(q,p)$. The *first class constraints* are those that commute on $\mathcal{C}(q,p)$ with all of the constraints. It is these first class constraints that are taken as the generators of the gauge transformations. The gauge invariant quantities (a.k.a. "observables") are then the phase function $F : \Gamma(q,p) \to \mathbb{R}$ that are constant along the gauge orbits.

Applying the formalism to our toy case of particle mechanics in Maxwellian spacetime, the canonical momenta are:

$$(5)\qquad \mathbf{p}_i := \frac{\partial L}{\partial\dot{\mathbf{x}}_i} = \frac{m_i}{M}\sum_k m_k(\dot{\mathbf{x}}_i - \dot{\mathbf{x}}_k) = m_i\dot{\mathbf{x}}_i - \frac{m_i}{M}\sum_k m_k\dot{\mathbf{x}}_k.$$

These momenta are not independent but must satisfy three primary constraints, which require the vanishing of the x, y, and z-components of the total momentum:

$$(6)\qquad \phi_\alpha = \sum_i p_i^\alpha = 0, \quad \alpha = 1, 2, 3.$$

These primary constraints are the only constraints — there are no secondary constraints — and they are all first class. These constraints generate in each configuration variable $\mathbf{x}_i$ the same gauge freedom; namely, a Euclidean shift given by the same arbitrary function of time. The gauge invariant variables, such relative particle positions and relative particle momenta, do evolve deterministically.

The technical elaboration of the constraint formalism is complicated, but one should not lose sight of the fact that the desire to save determinism is the motivation driving the enterprise. Here is a relevant passage from [Henneaux and

[25]The standard reference on these matters is [Henneaux and Teitelboim, 1992]. For a user friendly treatment of this formalism, see [Earman, 2003].

Teitelboim, 1992], one of the standard references on constrained Hamiltonian systems:

> The presence of arbitrary functions ... in the total Hamiltonian tells us that not all the q's and p's [the configuration variables and their canonical momenta] are observable [i.e. genuine physical magnitudes]. In other words, although the physical state is uniquely defined once a set of q's and p's is given, the converse is not true — i.e., there is more than one set of values of the canonical variables representing a given physical state. To see how this conclusion comes about, we note that if we are given an initial set of canonical variables at the time t_1 and thereby completely define the physical state at that time, we expect the equations of motion to *fully determine the physical state at other times*. Thus, by definition, any ambiguity in the value of the canonical variables at $t_2 \neq t_1$ should be a physically irrelevant ambiguity. [pp. 16–17]

As suggested by the quotation, the standard reaction to the apparent failure of determinism is to blame the appearance on the redundancy of the descriptive apparatus: the correspondence between the state descriptions in terms of the original variables — the q's and p's — and the physical state is many-to-one; when this descriptive redundancy is removed, the physical state is seen to evolve deterministically. There may be technical difficulties is carrying through this reaction. For example, attempting to produce a reduced phase space — whose state descriptions corresponding one-one to physical states — by quotienting out the gauge orbits can result in singularities. But when such technical obstructions are not met, normal (i.e. unconstrained) Hamiltonian dynamics applies to the reduced phase space, and the reduced phase space variables evolve deterministically.

In addition to this standard reaction to the apparent failure of determinism in the above examples, two others are possible. The first heterodoxy takes the apparent violation of determinism to be genuine. This amounts to (a) treating what the constraint formalism counts as gauge dependent quantities as genuine physical magnitudes, and (b) denying that these magnitudes are governed by laws which, when conjoined with the laws already in play, restore determinism. The second heterodoxy accepts the orthodox conclusion that the apparent failure of determinism is merely apparent; but it departs from orthodoxy by accepting (a), and it departs from the first heterodoxy by denying (b) and, accordingly, postulates the existence of additional laws that restore determinism. Instances that superficially conform to part (a) of the two heterodoxies are easy to construct from examples found in physics texts where the initial value problem is solved by supplementing the equations of motion, stated in terms of gauge-dependent variables, with a gauge condition that fixes a unique solution. For instance, Maxwell's equations written in terms of electromagnetic potentials do not determine a unique solution corresponding to the initial values of the potentials and their time derivatives. Imposing the Lorentz gauge condition converts Maxwell's equations to second order hyperbolic partial differential equations (pdes) that do admit an initial value formulation (see Section 4.2).[26] Similar examples can be concocted in general relativity theory where orthodoxy treats the metric potentials as gauge variables (see Section 6.2). In these examples orthodoxy is aiming to get at the values of the gauge independent variables via a choice of gauge. If this aim is not kept clearly in mind, the procedure creates the illusion that gauge-dependent variables have physical significance. It is exactly this illusion that the two heterodoxies take as real. The second heterodoxy amounts to taking the gauge conditions not as matters of calculational convenience but as additional physical laws. I know of no historical examples where this heterodoxy has led to fruitful developments in physics.

Since there is no *a priori* guarantee that determinism is true, the fact that the orthodox reading of the constraint formalism guarantees that the equations of motion admit an initial value formulation must mean that substantive assumptions that favor determinism are built into the formalism. That is indeed the case, for the Lagrangian/Hamiltonian formalism imposes a structure on the space of solutions: in the geometric language explained in Chapters 1 and 2 of this volume, the space of solutions has a symplectic or pre-symplectic structure. This formalism certainly is not guaranteed to be applicable to all of the equations of motion the Creator might have chosen as laws of motion; indeed, it is not even guaranteed to be applicable to all Newtonian type second order odes. In the 1880s Helmholtz found a set of necessary conditions for equations of this type to be derivable from an action principle; these conditions were later proved to be (locally) sufficient as well as necessary. After more than a century, the problem of finding necessary and sufficient conditions for more general types of equations of motion, whether in the form of odes or pdes, to be derivable from an action principle is still an active research topic.[27]

3.4 Determinism for fields and fluids in Newtonian physics

Newtonian gravitational theory can be construed as a field theory. The gravitational force is given by $\mathbf{F}_{grav} = -\nabla\varphi$, where the gravitational potential φ satisfies the Poisson equation

$$(7) \quad \nabla^2\varphi = \rho$$

with ρ being the mass density. If φ is a solution to Poisson's equation, then so is $\varphi' = \varphi + g(\mathbf{x})f(t)$ where $g(\mathbf{x})$ is a linear function of the spatial variables and $f(t)$

[26]Where $\mathbf{A}$ is the vector potential and Φ is the scalar potential, the Lorentz gauge requires that

$$\nabla \cdot \mathbf{A} + \frac{\partial \Phi}{\partial t} = 0$$

(with the velocity of light set to unity).

[27]Mathematicians discuss this issue under the heading of the "inverse problem." For precise formulations of the problem and surveys of results, see [Anderson and Thompson, 1992] and [Prince, 2000].

is an arbitrary function of t. Choose f so that $f(t) = 0$ for $t \leq 0$ but $f(t) > 0$ for $t > 0$. The extra gravitational force, proportional to $f(t)$, that a test particle experiences in the primed solution after $t = 0$ is undetermined by anything in the past.

The determinism wrecking solutions to (7) can be ruled out by demanding that gravitational forces be tied to sources. But to dismiss homogeneous solutions to the Poisson equation is to move in the direction of treating the Newtonian gravitational field as a mere mathematical device that is useful in describing gravitational interactions which, at base, are really direct particle interactions.[28] In this way determinism helps to settle the ontology of Newtonian physics: the insistence on determinism in Newtonian physics demotes fields to second-class status. In relativistic physics fields come into their own, and one of the reasons is that the relativistic spacetime structure supports field equations that guarantee deterministic evolution of the fields (see Section 4.2).

In the Newtonian setting the field equations that naturally arise are elliptic (e.g. the Poisson equation) or parabolic, and neither type supports determinism-without-crutches. An example of the latter type of equation is the classical heat equation

$$(8) \qquad \nabla^2 \Phi = \kappa \frac{\partial \Phi}{\partial t}$$

where Φ is the temperature variable and κ is the coefficient of heat conductivity.[29] Solutions to (8) can cease to exist after a finite time because the temperature "blows up." Uniqueness also fails since, using the fact that the heat equation propagates heat arbitrarily fast, it is possible to construct surprise solutions Φ_s with the properties that (i) Φ_s is infinitely differentiable, and (ii) $\Phi_s(\mathbf{x}, t) = 0$ for all $t \leq 0$ but $\Phi_s(\mathbf{x}, t) \neq 0$ for $t > 0$ (see [John, 1982, Sec. 7.1]). Because (8) is linear, if Φ is a solution then so is $\Phi' = \Phi + \Phi_s$. And since Φ and Φ' agree for all $t \leq 0$ but differ for $t > 0$, the existence of the surprise solutions completely wrecks determinism.

Uniqueness of solution to (8) can be restored by adding the requirement that $\Phi \geq 0$, as befits its intended interpretation of Φ as temperature; for Widder [1975, 157] has shown that if a solution of $\Phi(\mathbf{x}, t)$ of (8) vanishes at $t = 0$ and is non-negative for all $\mathbf{x}$ and all $t \geq 0$, then it must be identically zero. But one could have wished that, rather than having to use a stipulation of non-negativity to shore up determinism, determinism could be established and then used to show that if the temperature distribution at $t = 0$ is non-negative for all $\mathbf{x}$, then the uniquely determined evolution keeps the temperature non-negative. Alternatively, both uniqueness and existence of solutions of (8) can be obtained by limiting the

[28]This demotion of the status of the Newtonian gravitational field can also be supported by the fact that, unlike the fields that will be encountered in relativistic theories, it carries no energy or momentum.

[29]The fact that this equation is not Galilean invariant need cause no concern since Φ implicitly refers to the temperature of a medium whose rest frame is the preferred frame for describing heat diffusion.

growth of $|\Phi(\mathbf{x},t)|$ as $|\mathbf{x}| \to \infty$. But again one could have wished that such limits on growth could be derived as a consequence of the deterministic evolution rather than having to be stipulated as conditions that enable determinism.

Appearances of begging the question in favor of determinism could be avoided by providing at the outset a clear distinction between kinematics and dynamics, the former being a specification of the space of possible states. For example, a limit on the growth of quantum mechanical wave functions does not beg the question of determinism provided by the Schrödinger equation since the limit follows from the condition that the wave function is an element of a Hilbert space, which is part of the kinematical prescription of QM (see Section 5). Since this prescription is concocted to underwrite the probability interpretation of the wave function, we get the ironic result that the introduction of probabilities, which seems to doom determinism, also serves to support it. The example immediately above, as well as the examples of the preceding subsection and the one at the beginning of this subsection, indicate that in classical physics the kinematical/dynamical distinction can sometimes be relatively fluid and that considerations of determinism are used in deciding where to draw the line. The following example will reinforce this moral.[30]

The Navier-Stokes equations for an incompressible fluid moving in $\mathbb{R}^N$, $N = 2, 3$, read

$$\frac{D\mathbf{u}}{dt} = -\nabla p + \upsilon\Delta\mathbf{u} \quad \text{(9a)}$$

$$div(\mathbf{u}) = 0 \quad \text{(9b)}$$

where $\mathbf{u}(\mathbf{x},t) = (u^1, u^2, ..., u^N)$ is the velocity of the fluid, $p(\mathbf{x},t)$ is the pressure, $\upsilon = const. \geq 0$ is the coefficient of viscosity, and $D/dt := \partial/\partial t + \sum_{j=1}^{N} u^j \partial/\partial x^j$ is the convective derivative (see Foias at al. 2001 for a comprehensive survey). If the fluid is subject to an external force, an extra term has to be added to the right hand side of (9a). The Euler equations are the special case where $\upsilon = 0$. The initial value problem for (9a-b) is posed by giving the initial data

$$\text{(9)} \quad \mathbf{u}(\mathbf{x},0) = \mathbf{u}_0(\mathbf{x})$$

[30]Another reaction to the problems of determinism posed by (8) is to postpone them on the grounds that (8) is merely a phenomenological equation; heat is molecular motion and, thus, the fate of determinism ultimately rests with the character of laws of particle motion. It will be seen below, however, that in order to guarantee determinism for particle motion the helping hand of the stipulation of boundary conditions at infinity is sometimes needed. In any case, the postponement strategy taken to its logical conclusion would mean that no judgment about determinism would be forthcoming until we are in possession of the final theory-of-everything. It seems a better strategy to do today the philosophy of today's physics while recognizing, of course, that today's best theory may be superseded by a better future theory that delivers a different message about determinism.

where $\mathbf{u}_0(\mathbf{x})$ is a smooth (C^∞) divergence-free vector field, and is solved by smooth functions $\mathbf{u}, p \in C^\infty(\mathbb{R}^N x[0,\infty))$ satisfying (9)-(10). For physically reasonable solutions it is required both that $\mathbf{u}_0(\mathbf{x})$ should not grow too large as $|\mathbf{x}| \to \infty$ and that the energy of the fluid is bounded for all time:

$$(10) \quad \int_{\mathbb{R}^N} |\mathbf{u}(\mathbf{x},t)|^2 dx < \infty \text{ for all } t > 0.$$

When $\upsilon = 0$ the energy is conserved, whereas for $\upsilon > 0$ it dissipates.

For $N = 2$ it is known that a physically reasonable smooth solution exists for any given $\mathbf{u}_0(\mathbf{x})$. For $N = 3$ the problem is open. However, for this case it is known that the problem has a positive solution if the time interval $[0,\infty)$ for which the solution is required to exist is replaced by $[0,T)$ where T is a possibly finite number that depends on $\mathbf{u}_0(\mathbf{x})$. When T is finite it is known as the "blowup time" since $|\mathbf{u}(\mathbf{x},t)|$ must become unbounded as t approaches T. For the Euler equations a finite blowup time implies that the vorticity (i.e. the *curl* of $\mathbf{u}(\mathbf{x},t)$) becomes unbounded as t approaches T.

Smooth solutions to the Navier-Stokes equations, when they exist, are known to be unique. This claim would seem to be belied by the symmetries of the Navier-Stokes equations since if $\mathbf{u}(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)$, $p(\mathbf{x},t) = g(\mathbf{x},t)$ is a solution then so is the transformed $\tilde{\mathbf{u}}(\mathbf{x},t) = \mathbf{f}(\mathbf{x} - \varepsilon\alpha(t),t) + \varepsilon\alpha_t$, $\tilde{p}(\mathbf{x},t) = g(\mathbf{x} - \varepsilon\alpha(t),t) - \varepsilon\mathbf{x}\cdot\alpha_t + \frac{1}{2}\varepsilon^2\alpha_{tt}$, where $\alpha(t)$ is an arbitrary smooth function of t alone (see Olver 1993, pp. 130 and 177 (Exer. 2.15)). Choosing $\alpha(t)$ such that $\alpha(0) = \alpha_t(0) = \alpha_{tt}(0) = 0$ but $\alpha(t) \neq 0$ for $t > 0$ results in different solutions for the same initial data unless $\mathbf{f}(\mathbf{x} - \varepsilon\alpha(t),t) + \varepsilon\alpha_t = \mathbf{f}(\mathbf{x},t)$. However, the transformed solution violates the finiteness of energy condition (11).

The situation on the existence of solutions can be improved as follows. Multiplying (9a-b) by a smooth test function and integrating by parts over $\mathbf{x}$ and t produces integral equations that are well-defined for any $\mathbf{u}(\mathbf{x},t)$ and $p(\mathbf{x},t)$ that are respectively L^2 (square integrable) and L^1 (integrable). Such a pair is called a *weak solution* if it satisfies the integral equations for all test functions. Moving from smooth to weak solutions permits the proof of the existence of a solution for all time. But the move reopens the issue of uniqueness, for the uniqueness of weak solutions for the Navier-Stokes equations is not settled. A striking non-uniqueness result for weak solutions of the Euler equations comes from the construction by Scheffer [1994] and Shnirelman [1997] of self-exciting/self-destroying weak solutions: $\mathbf{u}(\mathbf{x},t) \equiv 0$ for $t < -1$ and $t > 1$, but is non-zero between these times in a compact region of $\mathbb{R}^3$.

It is remarkable that basic questions about determinism for classical equations of motion remain unsettled and that these questions turn on issues that mathematicians regard as worthy of attention. Settling the existence question for smooth solutions for the Navier-Stokes equations in the case of $N = 3$ brings a $1 million award from the Clay Mathematics Institute (see [Fefferman, 2000]).

3.5 Continuity issues

Consider a single particle of mass m moving on the real line $\mathbb{R}$ in a potential $V(x)$, $x \in \mathbb{R}$. The standard existence and uniqueness theorems for the initial value problem of odes can be used to show that the Newtonian equation of motion

$$(11) \quad m\ddot{x} = F(x) := -\frac{dV}{dx}$$

has a locally (in time) unique solution if the force function $F(x)$ satisfies a Lipschitz condition.[31] An example of a potential that violates the Lipschitz condition at the origin is $-\frac{9}{2}|x|^{4/3}$. For the initial data $x(0) = 0 = \dot{x}(0)$ there are multiple solutions of (12): $x(t) \equiv 0$, $x(t) = t^3$, and $x(t) = -t^3$, where m has been set to unity for convenience. In addition, there are also solutions $x(t)$ where $x(t) = 0$ for $t < k$ and $\pm(t-k)^3$ for $t \geq k$, where k is any positive constant. That such force functions do not turn up in realistic physical situations is an indication that Nature has some respect for determinism. In QM it turns out that Nature can respect determinism while accommodating some of the non-Lipschitz potentials that would wreck Newtonian determinism (see Section 5.2).

3.6 The breakdown of classical solutions

Consider again the case of a single particle of mass m moving on the real line $\mathbb{R}$ in a potential $V(x)$, and suppose that $V(x)$ satisfies the Lipschitz condition, guaranteeing a temporally local unique solution for the initial value problem for the Newtonian equations of motion. However, determinism can fail if the potential is such that the particle is accelerated off to $-\infty$ or $+\infty$ in a finite time.[32] Past determinism is violated because two such solutions can agree for all future times $t \geq t^*$ (say) — no particle is present at these times anywhere in space — but disagree at past times $t < t^*$ on the position and/or velocity of the particle when it is present in space. Since the potential is assumed to be time independent, the equations of motion are time reversal invariant, so taking the time reverses of these escape solutions produces solutions in which hitherto empty space is invaded by particles appearing from spatial infinity. These invader solutions provide violations of future determinism. Piecing together escape and invader solutions produces further insults to determinism.

In the 1890's Paul Painlevé conjectured that for $N > 3$ point mass particles moving in $\mathbb{R}^3$ under their mutually attractive Newtonian gravitational forces, there exist solutions to the Newtonian equations of motion exhibiting non-collision singularities, i.e. although the particles do not collide, the solution ceases to exist

[31] $F(x)$ satisfies the Lipschitz condition in an interval $(a, b) \subset \mathbb{R}$ if there is a constant $K > 0$ such that $|F(x_1) - F(x_2)| \leq K|x_1 - x_2|$ for all $x_1, x_2 \in (a, b)$. A sufficient condition for this is that dF/dx exists, is continuous, and $|dF/dx| \leq K$ on (a, b) for some $K > 0$.

[32] See [Reed and Simon, 1975, Theorem X.5] for necessary and sufficient conditions for this to happen.

after a finite time. Hugo von Zeipel [1908] showed that in such a solution the particle positions must become unbounded in a finite time. Finally, near the close of the 20th century Xia [1992] proved Painlevé conjecture by showing that for $N = 5$ point mass particles, the Newtonian equations of motion admit solutions in which the particles do not collide but nevertheless manage to accelerate themselves off to spatial infinity in a finite time (see [Saari and Xia, 1995] for an accessible survey).

Determinism can recoup its fortunes by means of the device, already mentioned above, of supplementing the usual initial conditions with boundary conditions at infinity. Or consolation can be taken from two remarks. The first remark is that in the natural phase space measure, the set of initial conditions that lead to Xia type escape solutions has measure zero. But it is unknown whether the same is true of all non-collision singularities. The second remark is that the non-collision singularities result from the unrealistic idealization of point mass particles that can achieve unbounded velocities in a finite time by drawing on an infinitely deep potential well. This remark does not suffice to save determinism when an infinity of finite sized particles are considered, as we will see in the next subsection.

It is interesting to note that for point particles moving under mutually attractive Newtonian gravitational forces, QM cures both the collision[33] and non-collision singularities that can spell the breakdown of classical solutions (see Section 5.2). This is more than a mere mathematical curiosity since it is an important ingredient in the explanation of the existence and stability of the hydrogen atom.

3.7 Infinite collections

Consider a collection of billiard balls confined to move along a straight line in Euclidean space. Suppose that the balls act only by contact, that only binary collisions occur, and that each such collision obeys the classical laws of elastic impact. Surely, the reader will say, such a system is as deterministic as it gets. This is so, *if* the collection is finite. But if the collection is infinite and unbounded velocities are permitted, then determinism fails because even with all of the announced restrictions in place the system can seemingly self-excite itself (see [Lanford, 1974]). Pérez Laraudogoitia [2001] shows how to use such infinite collections to create an analogue of the escape solution of the preceding subsection where all of the particles disappear in a finite amount of time. The time reverse of this scenario is one in which space is initially empty, and then without warning an infinite stream of billiard balls pour in from spatial infinity.

Legislating against unbounded velocities or imposing boundary conditions at infinity does not suffice to restore determinism if the billiard balls can be made arbitrarily small [Pérez Laraudogoitia, 2001]. For then a countably infinite collection of them can be Zeno packed into a finite spatial interval, say $(0, 1]$, by placing the center of the first ball at 1, the second at $1/2$, the third at $1/4$, etc. Assume for ease of illustration that all the balls have equal mass ($\equiv 1$). A unit mass cue

[33] A collision singularity occurs when two or more of the point particles collide and the solution cannot be continued through the collision time.

ball moving with unit speed from right to left collides with the first ball and sends a ripple through the Zeno string that lasts for unit time, at the end of which all of the balls are at rest. The boring history in which all the balls are at rest for all time is, of course, also a solution of the laws of impact. Comparing this boring history with the previous one shows that past Laplacian determinism is violated.[34]

This failure of determinism carries with it a violation of the conservation and energy momentum, albeit in a weak sense; namely, in the inertial frame in which the object balls are initially at rest, the total energy and the total momentum each have different values before and after the collisions start, but in every other inertial frame there is no violation simply because the values are infinite both before and after the collisions.[35] Pérez Laraudogoitia [2005] has shown how to construct scenarios in which there is a strong violation of conservation of energy and momentum in that the violation occurs in every inertial frame.

3.8 Domains of dependence

With some artificiality one of the threats to classical determinism discussed above can be summarized using a concept that will also prove very helpful in comparing the fortunes of determinism in classical physics and in relativistic physics. By a *causal curve* let us understand a (piecewise) smooth curve in spacetime that represents the spacetime trajectory for a physically possible transfer of energy/momentum. Define the *future domain of dependence*, $D^+(S)$, of a spacetime region S as the set of all spacetime points p such that any past directed causal curve with future endpoint at p and no past endpoint intersects S. The *past domain of dependence* $D^-(S)$ of S is defined analogously. And the *total domain of dependence* $D(S)$ is the union $D^+(S) \cup D^-(S)$. If $p \notin D(S)$ then it would seem that the state in region S does not suffice to determine the state at p since there is a possible causal process that passes through p but never registers on S.

Since neither the kinematics nor the dynamics of classical physics place an upper bound on the velocity at which energy/momentum can be transferred, it would seem that in principle any timelike curve — i.e. any (piecewise) smooth curve oblique to the planes of absolute simultaneity — can count as a causal curve, and as a consequence $D(S) = \varnothing$ even when S is taken to be an entire plane of absolute simultaneity. The examples from Sections 3.4, 3.6, and 3.7 show how the "in principle" can be realized by some systems satisfying Newtonian laws of motion.

We have seen that some threats to classical determinism can be met by beefing up the structure of classical spacetime. And so it is with the threat currently under consideration. *Full Newtonian spacetime* is what results from neo-Newtonian

[34]The time reverse of the interesting history starts with all the balls initially at rest, and then subsequently the collection self-excites, sending a ripple of collisions from left to right and ejecting the cue ball. If this self-exciting history is physically possible, then future laplacian determinism is violated. However, it might rejected on the grounds that it violated Newton's first law of motion.

[35]For a comment on how the availability of an infinite amount of momentum/energy renders the indeterminism unsurprising, see [Norton, 1999, 1268].

spacetime by adding absolute space in the form of a distinguished inertial frame ('absolute space'). In this setting the spacetime symmetries are small enough that there are now finite invariant velocities (intuitively, velocities as measured relative to absolute space), and thus laws can be formulated that set a finite upper bound on the absolute velocity of causal propagation. Nor is this move necessarily *ad hoc* as shown, for example, by the fact that the formulation of Maxwell's laws of electromagnetism in a classical spacetime setting evidently requires the services of a distinguished inertial frame, the velocity of light c being the velocity as measured in this frame.

But, as is well known, such a formulation is embarrassed by the undetectability of motion with respect to absolute space. This embarrassment provides a direct (albeit anachronistic) route from classical to relativistic spacetime. Adopting for classical spacetimes the same geometric language used in the special and general theories of relativity (see [Earman, 1989, Ch. 2]), absolute space is represented by a covariantly constant timelike vector field A^a, the integral curves of which are the world lines of the points of absolute space. The space metric is represented by a degenerate second rank contravariant tensor h^{ab}, which together with A^a defines a tensor that is formally a Minkowski metric: $\eta^{ab} := h^{ab} - A^a A^b$. The unobservability of absolute motion means that there is no preferred way to split η^{ab} into an h^{ab} part and a $A^a A^b$ part, suggesting that η^{ab} is physically as well as formally a Lorentz metric. As we will see in Section 4.1, this puts determinism on much firmer ground in that domains of dependence of local or global time slices are non-empty in the spacetime setting of STR.

3.9 Determinism, predictability, and chaos

Laplace's vision of a deterministic universe makes reference to an "intelligence" (which commentators have dubbed 'Laplace's Demon'):

> We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all of the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes.[36]

[36][Laplace, 1820]. English translation from [Nagel, 1961, 281-282]. More than a century earlier Leibniz espoused a similar view: "[O]ne sees then that everything proceeds mathematically — that is, infallibly — in the whole wide world, so that if someone could have sufficient insight into the inner parts of things, and in addition has remembrance and intelligence enough to consider all the circumstances and to take them into account, he would be a prophet and would see the future in the present as in a mirror." Quoted from [Cassirer, 1956, 12].

Perhaps by taking Laplace's vision too literally, philosophers and physicists alike conflate determinism and predictability. The conflation leads them to reason as follows: here is a case where predictability fails; thus, here is a case where determinism fails.[37] This is a mistake that derives from a failure to distinguish determinism — an ontological doctrine about how the world evolves — from predictability — an epistemic doctrine about what can inferred, by various restricted means, about the future (or past) state of the world from a knowledge of its present state.

There is, however, an interesting connection between determinism and practical predictability for laws of motion that admit an initial value problem that is *well-posed* in the sense that, in some appropriate topology, the solutions depend continuously on the initial data.[38] The standard existence and uniqueness proofs for the initial value problem for the odes used in particle mechanics also furnish a proof of well-posedness, which can be traced to the fact that the existence proof is constructive in that it gives a procedure for constructing a series of approximations that converge to the solution determined by the initial data.

To illustrate the implications of well-posedness for predictability, consider the toy case of a system consisting of a single massive particle obeying Newtonian equations of motion. If a suitable Lipschitz condition is satisfied, then for any given values of the position $q(0)$ and velocity $\dot{q}(0)$ of the particle at $t = 0$ there exists (for some finite time interval surrounding $t = 0$) a unique solution: symbolically $q(t) = F(q(0), \dot{q}(0), t)$. And further, since this initial value problem is well-posed, for any fixed $t > 0$ (within the interval for which the solution is guaranteed to exist), F is a continuous function of $q(0)$ and $\dot{q}(0)$. Suppose then that the practical prediction task is to forecast the actual position $\bar{q}(t^*)$ of the particle at some given $t^* > 0$ with an accuracy of $\epsilon > 0$, and suppose that although measurements of position or velocity are not error free, the errors can be made arbitrarily small. By the continuity of F, there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that if $|q(0) - \bar{q}(0)| < \delta_1$ and $|\dot{q}(0) - \bar{\dot{q}}(0)| < \delta_2$, then $|q(t^*) - \bar{q}(t^*)| < \epsilon$. Thus, measuring at $t = 0$ the actual particle position and velocity with accuracies $\pm\delta_1/2$ and $\pm\delta_2/2$ respectively ensures that when the measured values are plugged into F, the value of the function for $t = t^*$ answers to the assigned prediction task. (Note, however, that since the actual initial state is unknown, so are the required accuracies $\pm\delta_1/2$ and $\pm\delta_2/2$, which may depend on the unknown state as well as on ϵ and t^*. This hitch could be overcome if there were minimum but non-zero values of δ_1 and δ_2 that

[37]On the philosophical side, Karl Popper is the prime example. Popper [1982] goes so far as to formulate the doctrine of "scientific determinism" in terms of prediction tasks. An example on the physics side is Reichl [1992]: "[W]e now know that the assumption that Newton's equations are deterministic is a fallacy! Newton's equations are, of course, the correct starting point of mechanics, but in general they only allow us to determine [read: predict] the long time behavior of *integrable* mechanical systems, few of which can be found in nature" (pp. 2–3). I am happy to say that in the second edition of Reichl's book this passage is changed to "[W]e now know that the assumption that Newton's equations can *predict* the future is a fallacy!" [Reichl 2004, 3; italics added].

[38]When the topology is that induced by a norm $||\cdot||$ on the instantaneous states represented by a function $s(t)$ of time, well-posedness requires that there is a non-decreasing, nonnegative function $C(t)$ such that $||s(t)|| \leq C(t)||s(0)||$, $t > 0$, for any solution $s(t)$.

answered to the given prediction task whatever the initial state; but there is no *a priori* guarantee that such minimum values exist. A prior measurement with known accuracy of the position and velocity at some $t^{**} < 0$ will put bounds, which can be calculated from F, on the position and velocity at $t = 0$. And then the minimum values can be calculated for accuracies δ_1 and δ_2 of measurements at $t = 0$ that suffice for the required prediction task for any values of the position and velocity within the calculated bounds.)

Jacques Hadamard, who made seminal contributions to the Cauchy or initial value problem for pdes, took the terminology of "well-posed" (a.k.a. "properly posed") quite literally. For he took it as a criterion for the proper mathematical description of a physical system that the equations of motion admit an initial value formulation in which the solution depends continuously on the initial data (see [Hadamard, 1923, 32]). However, the standard Courant-Hilbert reference work, *Methods of Mathematical Physics*, opines that

> "properly posed" problems are by far not the only ones which appropriately reflect real phenomena. So far, unfortunately, little mathematical progress has been made in the important task of solving or even identifying such problems that are not "properly posed" but still are important and motivated by realistic situations. [1962, Vol. 2, 230].

Some progress can be found in [Payne, 1975] and the references cited therein.

Hadamard was of the opinion that if the time development of a system failed to depend continuously on the initial conditions, then "it would appear to us as being governed by pure chance (which, since Poincaré,[39] has been known to consist precisely in such a discontinuity in determinism) and not obeying any law whatever" [1923, 38]. Currently the opinion is that the appearance of chance in classical systems is due not to the failure of well-posedness but to the presence of chaos.

The introduction of *deterministic chaos* does not change any of the above conclusions about determinism and predictability. There is no generally agreed upon definition of chaos, but the target class of cases can be picked out either in terms of cause or effects. The cause is sensitive dependence of solutions on initial conditions, as indicated, for example, by positive Lyapunov exponents. The effects are various higher order ergodic properties, such as being a mixing system, being a K-system, being a Bernoulli system, etc.[40] Generally a sensitive dependence on initial conditions *plus* compactness of the state space is sufficient to secure such properties. The sensitive dependence of initial condition that is the root cause of chaotic behavior does not contradict the continuous dependence of solutions on initial data, and, therefore, does not undermine the task of predicting with any desired finite accuracy the state at a *fixed* future time, assuming that error in measuring the initial conditions can be made arbitrarily small. If, however, there

[39] See Poincaré's essay "Chance" in *Science and Method* [1952].

[40] See Uffink, this volume, section 6.2, or [Lichtenberg and Lieberman, 1991] for definitions of these concepts.

is a fixed lower bound on the accuracy of measurements — say, because the measuring instruments are macroscopic and cannot make discriminations below some natural macroscopic scale — then the presence of deterministic chaos can make some prediction tasks impossible. In addition, the presence of chaos means that no matter how small the error (if non zero) in ascertaining the initial conditions, the accuracy with which the future state can be forecast degrades rapidly with time. To ensure the ability to predict with some given accuracy $\epsilon > 0$ for all $t > 0$ by ascertaining the initial conditions at $t = 0$ with sufficiently small error $\delta > 0$, it would be necessary to require not only well-posedness but *stability*, which is incompatible with chaos.[41]

Cases of classical chaos also show that determinism on the microlevel is not only compatible with stochastic behavior at the macro level but also that the deterministic microdynamics can ground the macro-stochasticity. For instance, the lowest order ergodic property — ergodicity — arguably justifies the use of the microcanonical probability distribution and provides for a relative frequency interpretation; for it implies that the microcanonical distribution is the only stationary distribution absolutely continuous with respect to Lebesque measure and that the measure of a phase volume is equal to the limiting relative frequency of the time the phase point spends in the volume. In these cases there does not seem to be a valid contrast between "objective" and "epistemic" probabilities. The probabilities are epistemic in the sense that conditionalizing on a mathematically precise knowledge of the initial state reduces the outcome probability to 0 or 1. But the probabilities are not merely epistemic in the sense of merely expressing our ignorance, for they are supervenient on the underlying microdynamics.

Patrick Suppes [1991; 1993] has used such cases to argue that, because we are confined to the macrolevel, determinism becomes for us a "transcendental" issue since we cannot tell whether we are dealing with a case of irreducible stochasticity or a case of deterministic chaos. Although I feel some force to the argument, I am not entirely persuaded. There are two competing hypotheses to explain observed macro-stochasticity: it is due to micro-determinism plus sensitive dependence on initial conditions vs. it is due to irreducible micro-stochasticity. The work in recent decades on deterministic chaos supplies the details on how the first hypothesis can be implemented. The details of the second hypothesis need to be filled in; particular, it has to be explained how the observed macro-stochasticity supervenes on the postulated micro-stochasticity.[42] And then it has to be demonstrated that the two hypotheses are underdetermined by all possible observations on the macrolevel. If both of these demands were met, we would be faced with a particular instance of the general challenge to scientific realism posed by underdetermination of theory by observational evidence, and all of the well-rehearsed moves and countermoves in the realism debate would come into play. But it is futile to fight these battles until some concrete version of the second hypothesis is presented.

---

[41]Stability with respect to a norm on states $s(t)$ requires that there is a constant C such that $||s(t)|| \leq C||s(0)||$, $t > 0$, for any solution $s(t)$. Compare to footnote 38.

[42]It is not obvious that micro-stochasticity always percolates up to the macro-level.

3.10 Laplacian demons, prediction, and computability

Since we are free to imagine demons with whatever powers we like, let us suppose that Laplace's Demon can ascertain the initial conditions of the system of interest with absolute mathematical precision. As for computational ability, let us suppose that the Demon has at its disposal a universal Turing machine. As impressive as these abilities are, they may not enable the Demon to predict the future state of the system even if it is deterministic. Returning to the example of the Newtonian particle from the preceding subsection, if the values of the position and velocity of the particle at time $t = 0$ are plugged into the function $F(q(0), \dot{q}(0), t)$ that specifies the solution $q(t)$, the result is a function $\mathcal{F}(t)$ of t; and plugging different values of the initial conditions results in different $\mathcal{F}(t)$ — indeed, by the assumption of determinism, the $\mathcal{F}(t)$'s corresponding to different initial conditions must differ on any finite interval of time no matter how small. Since there is a continuum of distinct initial conditions, there is thus a continuum of distinct $\mathcal{F}(t)$'s. But only a countable number of these $\mathcal{F}(t)$'s will be Turing computable functions.[43] Thus, for most of the initial conditions the Demon encounters, it is unable to predict the corresponding particle position $q(t)$ at $t > 0$ by using its universal Turing machine to compute the value of $\mathcal{F}(t)$ at the relevant value of t — in Pitowsky's [1996] happy turn of phrase, the Demon must consult an Oracle in order to make a sure fire prediction.

However, if $q(0)$ and $\dot{q}(0)$ are both Turing computable real numbers, then an Oracle need not be consulted since the corresponding $\mathcal{F}(t)$ is a Turing computable function; and if t is a Turing computable real number, then so is $\mathcal{F}(t)$. This follows from the fact that the existence and uniqueness proofs for odes gives an effective procedure for generating a series of approximations that converges effectively to the solution; hence, if computable initial data are fed into the procedure, the result is an effectively computable solution function. Analogous results need not hold when the equations of motion are pdes. Jumping ahead to the relativistic context, the wave equation for a scalar field provides an example where Turing computability of initial conditions is not preserved by deterministic evolution (see Section 4.4).

A more interesting example where our version of Laplace's Demon must consult an Oracle has been discussed by Moore [1990; 1991] and Pitowsky [1996]. Moore constructed an embedding of an abstract universal Turing machine into a concrete classical mechanical system consisting of a particle bouncing between parabolic and flat mirrors arranged so that the motion of the particle is confined to a unit

[43]The familiar notion of a Turing computable or recursive function is formulated for functions of the natural numbers, but it can be generalized so as to apply to functions of the real numbers. First, a computable real number x is defined as a limit of a computable sequence $\{r_n\}$ of rationals that converges effectively, i.e. there is a recursive function $f(n)$ such that $k \geq f(n)$ entails $|x - r_k| \leq 10^n$. Next, a sequence $\{x_n\}$ of reals is said to be computable iff there is a double sequence $\{r_{kn}\}$ such that $r_{kn} \to x_n$ as $k \to \infty$ effectively in both k and n. Finally, a function of the reals is said to be computable iff it maps every computable sequence in its domain into a computable sequence and, moreover, it is effectively uniformly continuous. For details, see [Pour-el and Richards, 1989].

square. Using this embedding Moore was able to show how recursively unsolvable problems can be translated into prediction tasks about the future behavior of the particle that the Demon cannot carry out without help from an Oracle, even if it knows the initial state of the particle with absolute precision! For example, Turing's theorem says that there is no recursive algorithm to decide whether a universal Turing machine halts on a given input. Since the halting state of the universal Turing machine that has been embedded in the particle-mirror system corresponds to the particle's entering a certain region of the unit square to which it is thereafter confined, the Demon cannot predict whether the particle will ever enter this region. The generalization of Turing's theorem by Rice [1953] shows that many questions about the behavior of a universal Turing machine in the unbounded future are recursively unsolvable, and these logical questions will translate into physical questions about the behavior of the particle in the unbounded future that the Demon cannot answer without consulting an Oracle.

The reader might ask why we should fixate on the Turing notion of computability. Why not think of a deterministic mechanical system as an analogue computer, regardless of whether an abstract Turing machine can be embedded in the system? For instance, in the above example of the Newtonian particle with deterministic motion, why not say that the particle is an analogue computer whose motion "computes," for any given initial conditions $q(0), \dot{q}(0)$, the possibly non-Turing computable function $q(t) = F(q(0), \dot{q}(0), t)$? I see nothing wrong with removing the scare quotes and developing a notion of analogue computability along these lines. But the practical value of such a notion is dubious. Determining which function of t is being computed and accessing the value computed for various values of t requires ascertaining the particle position with unbounded accuracy.

Connections between non-Turing computability and general relativistic spacetimes that are inhospitable to a global version of Laplacian determinism will be mentioned below in Section 6.6.

4 DETERMINISM IN SPECIAL RELATIVISTIC PHYSICS

4.1 How the relativistic structure of spacetime improves the fortunes of determinism

Special relativistic theories preserve the Newtonian idea of a fixed spacetime background against which the drama of physics plays itself out, but they replace the background classical spacetimes with Minkowski spacetime. This replacement makes for a tremendous boost in the fortunes of determinism. For the symmetries of Minkowski spacetime are given by the Poincaré group, which admits a finite invariant speed c, the speed of light, making it possible to formulate laws of motion/field equations which satisfy the basic requirement that the symmetries of the spacetime are symmetries of the laws and which propagate energy-momentum no faster than c. For such laws all of the threats to classical determinism that derive from unbounded velocities are swept away.

The last point can be expounded in terms of the apparatus introduced in Section 3.8. For the type of law in question, a causal curve is a spacetime worldline whose tangent at any point lies inside or on the null cone at that point, with the upshot that domains of dependence are now non-trivial. Minkowski spacetime admits a plethora of global time functions. But in contrast with classical spacetimes, such a function t can be chosen so that the domains of dependence $D(t = const)$ of the level surfaces of t are non-empty. Indeed, t can be chosen so that for each and every $t = const$ the domain of dependence $D(t = const)$ is a *Cauchy surface*, i.e. $D(t = const)$ is the entire spacetime. In fact, any inertial time coordinate is an example of a global time function, all of whose levels are Cauchy surfaces.[44] In the context of STR, the definition of Laplacian determinism given above in Section 2.1 is to be understood as applying to a t with this Cauchy property.

It is important to realize that these determinism friendly features just discussed are not automatic consequences of STR itself but involve additional substantive assumptions. The *stress-energy tensor* T^{ab} used in both special and general relativistic physics describes how matter-energy is distributed through spacetime. What is sometimes called the local conservation law for T^{ab}, $\nabla_a T^{ab} = 0$, where ∇_a is the covariant derivative determined by the spacetime metric, does *not* guarantee that the local energy-momentum flow as measured by any observer is always non-spacelike. That guarantee requires also that for any future pointing timelike U^a, $-T^{ab}U_a$ is a future pointing, non-spacelike vector.[45] Combining this requirement with the further demand that the local energy density as measured by any observer is non-negative, i.e. $T^{ab}U_aU_b \geq 0$ for any non-spacelike vector field U^a, produces what is called the *dominant energy condition.* Not surprisingly, this condition, together with the local conservation of T^{ab}, does guarantee that the matter fields that give rise to T^{ab} cannot travel faster than light in the sense that if T^{ab} vanishes on some spacelike region S, then it must also vanish on $D(S)$ (see [Hawking and Ellis, 1973, 91-94]). The dominant energy conditions is thought to be satisfied by all the matter-fields encountered in the actual world, but occasionally what are purported to be counterexamples appear in the physics literature.

4.2 Fundamental fields

In Section 3.4 examples were given to illustrate how fields have a hard time living up to the ideals of Laplacian determinism in classical spacetimes. The situation changes dramatically in Minkowski spacetime, which supports field equations in the form of hyperbolic pdes.[46] For example, the Klein-Gordon equation for a scalar field ϕ of mass $m \geq 0$ obeys the equation

(12) $\nabla_a \nabla^a \phi - m^2 \phi = 0$

[44] Exercise for the reader: Construct a global time function t for Minkowski spacetime such that none of the level surfaces of t are Cauchy.

[45] The minus sign comes from the choice of the signature $(+ + + -)$ for the spacetime metric.

[46] A standard reference on the classification of pdes relevant to physical applications is [Courant and Hilbert, 1962, Vol. 2]. See also [Beig, 2004].

which is a special case of a linear, diagonal, second order hyperbolic pde. For such equations there is a global existence and uniqueness proof for the initial value problem: given a Cauchy surface Σ of Minkowski spacetime and C^∞ initial data, consisting of the value of ϕ on Σ and the normal derivative of ϕ with respect to Σ, there exists a unique C^∞ solution of (13) throughout spacetime. Furthermore, the initial value problem is well-posed in that (in an appropriate topology) the unique solution depends continuously on the initial data. And finally the Klein-Gordon field propagates causally in that if the initial data are varied outside a closed subset $S \subset \Sigma$, the unique solution on $D(S)$ does not vary. Notice that we have a completely clean example of Laplacian determinism at work — no boundary conditions at infinity or any other enabling measures are needed to fill loopholes through which indeterminism can creep in. By contrast, giving initial data on a timelike hypersurface of Minkowski spacetime is known to lead to an improperly posed Cauchy problem; indeed, not only do solutions not depend continuously on the initial data, but there are C^∞ initial data for which there is no corresponding solution. This asymmetry between the fortunes of determinism in timelike vs. spacelike directions, could, as noted above, be turned around and used as a basis for singling out the time dimension.

It should be emphasized that only restricted classes of hyperbolic pdes are known to have well-posed initial value problems. It is a challenge to mathematical physics to show that the field equations encountered in physical theories can be massaged into a form that belongs to one of these classes. It is a comparatively easy exercise to show that, when written in terms of potentials, the source-free Maxwell equations for the electromagnetic field take the form of a system of linear, diagonal, second order hyperbolic pdes if an appropriate gauge condition is applied to the potentials. In other cases the challenge requires real ingenuity.[47]

Physicists are so convinced of determinism in classical (= non-quantum) special relativistic physics that they sort "fundamental" from "non-fundamental" matter fields according as the field does or does not fulfill Laplacian determinism in the form of global existence and uniqueness theorems for the initial value problem on Minkowski spacetime. The Klein-Gordon field and the source-free Maxwell electromagnetic field qualify as fundamental by this criterion. A dust matter field, however, fails to make the cut since singularities can develop from regular initial data since, for example, in a collapsing ball of dust the density of the dust can become infinite if the outer shells fall inward fast enough that they cross the inner shells. Such shell-crossing singularities can develop even for physically reasonable initial data for the Maxwell-Lorentz equations where the source for the electromagnetic field consists of a charged dust obeying the Lorentz force law. But no great faith in determinism is needed to brush aside the failure of determinism in such examples; they can also be dismissed on the grounds that dust matter is an idealization and, like all idealizations, it ceases to work in some circumstances. Faith in determinism, however, is required to deal with what happens when the Klein-Gordon equation is converted into a non-linear equation by adding terms to

[47] See [Beig, 2004] for examples.

the right hand side of (13), e.g.

(13) $\nabla_a \nabla^a \phi - m^2\phi = \lambda\phi^2$

where λ is a constant. It is known that solutions of (14) corresponding to regular initial data can become infinite at a finite value of t and that such data has non-zero measure (see [Keller, 1957]).

A number of attempts have been made to modify the classical Navier-Stokes equations (see Section 3.4) for dissipative fluids in order to make them consistent with STR in the sense that they become a system of hyperbolic pdes with causal propagation. A criterion of success is typically taken to be that the resulting system admits an initial value formulation, confirming once again the faith in determinism in the special relativistic setting. One difficulty in carrying out this program is that it necessitates the introduction of additional dynamical variables and additional dynamical equations, and as a result many different relativistic generalizations of the classical equations have been produced. Geroch [1995] has argued that we need not be troubled by this *embarras des riches* because the differences among the relativistic generalizations wash out at the level of the empirical observations that are captured by the Navier-Stokes theory.

4.3 Predictability in special relativistic physics

The null cone structure of Minkowski spacetime that makes possible clean examples of Laplacian determinism works against predictability for embodied observers who are not simply "given" initial data but must ferret it out for themselves by causal contact with the system whose future they are attempting to predict. Consider, for example, the predicament of an observer O whose world line is labeled γ in Fig. 1. At spacetime location p this observer decides she wants to predict what will happen to her three days hence (as measured in her proper time).

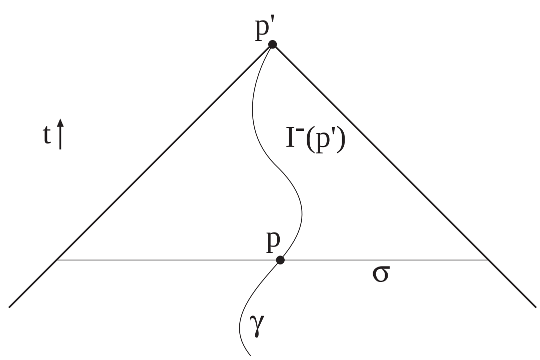

Figure 1. The failure of predictability in Minkowski spacetime

Suppose that, in fact, three-days-hence for O corresponds to the spacetime location p'. And suppose that the equations governing the variables relevant to the prediction are such that initial data on a spacelike hypersurface Σ fixes a unique solution in $D(\Sigma)$. Then to carry out the prediction by way of solving the relevant equations, O must ascertain the state on a local spacelike hypersurface that slices through the past null cone of p', such as σ in Fig. 1. As O's "now" creeps up her world line, the past light of the "now" sweeps out increasingly large portions of σ, but until her "now" reaches p' she does not have causal access to all of σ. And the same goes for any other local slice through the past cone of p'. Thus, the very spacetime structure that provides a secure basis for Laplacian determinism prevents O from acquiring the information she needs before the occurrence of the event that was to be predicted.

This predictability predicament can be formalized in a way that will be useful when it comes to investigating predictability in a general relativistic spacetime $\mathcal{M}, g_{ab}$, where $\mathcal{M}$ is a differentiable manifold and g_{ab} is a Lorentz signature metric defined on all of $\mathcal{M}$, Minkowski spacetime being the special case where $\mathcal{M} = \mathbb{R}^n$ and g_{ab} is the Minkowski metric. Geroch (1977) defines the *domain of prediction* $P(q)$ of a point $q \in \mathcal{M}$ to consist of all points $p \in \mathcal{M}$ such that (i) every past directed timelike curve with future endpoint at p and no past endpoint enters the chronological past $I^-(q)$ of q,[48] and (ii) $I^-(p) \nsubseteq I^-(q)$. Condition (i) is needed to ensure that causal processes that can influence events at p register on the region $I^-(q)$ that is causally accessible to an observer whose "now" is q, and condition (ii) is needed to ensure that from the perspective of q, the events to be predicted at p have not already occurred. The predictability predicament for Minkowski spacetime can now be stated as the theorem that for every point q of Minkowski spacetime, $P(q) = \varnothing$.

Note that the predictability predicament arises not just because of the local null cone structure of Minkowski spacetime but also because of its global topological structure. To drive home this point, suppose that space in $(1 + 1)$-Minkowski spacetime is compactified to produce the cylindrical spacetime $\mathcal{C}$ pictured in Fig. 2. Now predictability is possible since $I^-(q)$ for any q contains a Cauchy surface, e.g. Σ in Fig. 2. As a result $P(q) = \mathcal{C} - I^-(q)$.

For standard Minkowski spacetime and other spacetimes for which $P(q) = \varnothing$ for every spacetime point q, one can wonder how secure predictions are possible. The answer is that if complete security is required, the only secure predictions have a conditional form, where the antecedent refers to events that are not causally accessible from q. But there will be many such conditionals, with different antecedents and different consequents, and since one will not be in a position to know which of the antecedents is actualized, the best one can do is a "prediction" (all too familiar from economic forecasts) consisting of a big set of conditionals. On the other hand, if complete security is not demanded, then unconditional predictions

[48]For a point q in a relativistic spacetime $\mathcal{M}, g_{ab}$, the chronological past $I^-(q)$ consists of all $p \in \mathcal{M}$ such that there is a future directed timelike curve from p to q. The chronological future $I^+(q)$ of a point q is defined analogously.

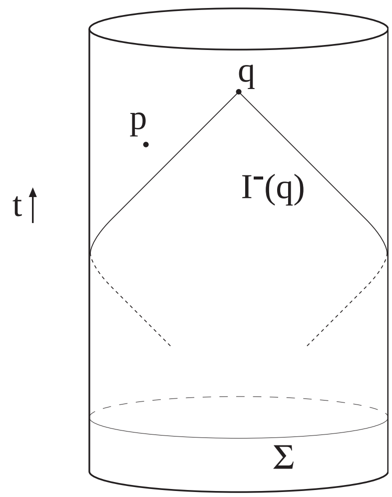

Figure 2. The improved fortunes of predictability when the spatial aspect of Minkowski spacetime is compactified

carrying probability instead of certainty are obtainable if inductive inference from past observations points to one of the antecedents of the set of conditionals as being highly likely.

If one wants predictions that are in principle verifiable, then a third condition needs to be added to the definition of the domain of prediction; namely, (iii) $p \in I^+(q)$. The point p in Fig. 2 satisfies clauses (i) and (ii) but not (iii).

4.4 Special relativity and computability

Pour-el and Richards [1981] constructed an example in which deterministic evolution does not preserve Turing computability. The equation of motion at issue is the relativistic wave equation, which in inertial coordinates is written

$$(14) \quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{\partial^2 u}{\partial t^2} = 0, \qquad c \equiv 1$$

Pour-el and Richards studied solutions corresponding to initial data at $t = 0$ of the form $u(x, y, z, 0) = f(x, y, z)$, $\partial u(x, y, z, 0)/\partial t = 0$. They showed that there is a Turing computable $f(x, y, z)$ such that the corresponding solution $u(x, y, z, t)$ is not Turing computable at $t = 1$. However, such a solution is necessarily a weak solution (in the sense of Section 3.4) to the wave equation since it must be non-differentiable. And the non-preservation result is sensitive to the norm used to define convergence. Indeed, if Turing computability is defined using the energy norm,[49] then for any Turing computable functions f and g, the solution $u(x,y,z,t)$ corresponding to $u(x,y,z,0) = f(x,y,z)$, $\partial u(x,y,z,0)/\partial t = g(x,y,z)$ is Turing computable (see [Pour-el and Richards, 1989, 116-118]).

5 DETERMINISM AND INDETERMINISM IN ORDINARY QM

The folklore on determinism has it that QM is the paradigm example of an indeterministic theory. Like most folklore, this bit contains elements of truth. But like most folklore it ignores important subtleties — in this instance, the fact that in some respects QM is more deterministic and more predictable than classical physics. And to distill the elements of truth from the folklore takes considerable effort — in particular, the folkloric notion that quantum indeterminism arises because the "reduction of the wave packet" is based on a controversial interpretation of the quantum measurement process. Before turning to these matters, I will discuss in Section 5.1 an issue that links to the some of the themes developed above and in Secs. 5.2-5.4 some issues unjustly neglected in the philosophical literature.

5.1 Determinism and Galilean invariance in QM

Here is another example of how linking determinism and symmetry considerations is fruitful in producing physical insights. Consider the motion of a single spinless particle on the real line $\mathbb{R}$, and work in the familiar Hilbert space $\mathcal{H}$ of wave functions, i.e. $\mathcal{H} = L^2_{\mathbb{C}}(\mathbb{R}, dx)$. The evolution of the state $\psi(x) \in \mathcal{H}$ of the quantum particle is governed by the Schrödinger equation

$$(15) \quad i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi$$

where $\hat{H}$ is the Hamiltonian operator. This evolution is deterministic, or so it is said. But a puzzle is quickly generated by conjoining the presumed determinism with the presumed Galilean invariance of (16).[50] Since (16) is first order in time, giving the value of the wave function $\psi(x,0)$ for all $x \in \mathbb{R}$ at $t = 0$ should fix the value of $\psi(x,t)$ for all $t > 0$. But how can this be if the Schrödinger equation is Galilean invariant? A proper Galilean transformation $x \to x' = x - vt$, $v \neq 0$, is the identity map for $t = 0$ but non-identity for $t > 0$. Assuming Galilean invariance of (16), this map must carry a solution to a solution. Since the map in question is the identity for $t = 0$ the two solutions should have the same initial data $\psi(x,0)$; but since the map is not the identity for $t > 0$ the original solution and

[49] For initial conditions f, g the energy norm is given by
$||f,g||^2 := \int\int\int[(\nabla f)^2 + g^2]dxdydz.$
And for functions u on $\mathbb{R}^4$ the norm is $||u(x,y,z,t)|| = \sup_t E(u,t)$, where
$E(u,t)^2 := \int\int\int[\nabla u + \left(\frac{\partial u}{\partial t}\right)^2]dxdydz.$
If u is a solution of the wave equation, then $E(u,t)$ is independent of t.

[50] For a treatment of the Galilean invariance of the Schrödinger equation, see [Brown, 1999].

its image under a Galilean boost should diverge in the future, violating Laplacian determinism. The resolution of this little puzzle is to reject the implicit assumption that ψ behaves as a scalar under a Galilean transformation. In fact, Galilean invariance of the Schrödinger equation can be shown to imply that the Galilean transformation of ψ depends on the mass of the particle. And this in turn entails a "superselection rule" for mass (discovered by Bargmann [1954]) which means that a superposition of states corresponding to different masses is not physically meaningful in non-relativistic QM.

5.2 How QM can be more deterministic than classical mechanics

Physics textbooks on QM offer a procedure for quantization that starts with a Hamiltonian formulation of the classical dynamics for the system of interest and produces, modulo operator ordering ambiguities, a formal expression for the quantum Hamiltonian operator $\hat{H}$ that is inserted into equation (16).[51] But to make the formal expression into a genuine operator a domain of definition must be specified since, typically, $\hat{H}$ is an unbounded operator and, therefore, is defined at best for a dense domain of the Hilbert space. Usually it is not too difficult to find a dense domain on which $\hat{H}$ acts as a symmetric operator. The question then becomes whether or not this operator is essentially self-adjoint, i.e. has a unique self-adjoint (SA) extension — which will also be denoted by $\hat{H}$.[52] If so, $\hat{U}(t) := \exp(-i\hat{H}t)$ is unitary for all $t \in \mathbb{R}$, and since $\hat{U}(t)$ is defined for the entire Hilbert space, the time evolute $\psi(t) := \hat{U}(t)\psi$ for every vector in the Hilbert space is defined for all times. (The Schrödinger equation (16) is just the "infinitesimal" version of this evolution equation.) Thus, if $\hat{H}$ is essentially SA, none of the problems which beset the deterministic evolution of the classical state can trouble the deterministic evolution of the quantum state.

What is, perhaps, surprising is that the quantum Hamiltonian operator can be essentially SA in some cases where the counterpart classical system does not display deterministic evolution. Recall from Section 3.5 the example of a particle moving on the real line $\mathbb{R}$ in a (real-valued) potential $V(x)$, $x \in \mathbb{R}$. As we saw, when the potential is proportional to $-|x|^{4/3}$ near the origin, the initial value problem for the Newtonian equation of motion does not have a unique solution. But the quantum Hamiltonian operator $\hat{H} = -\dfrac{\hbar^2}{2m}\dfrac{d^2}{dx^2} - V(x)$ is essentially SA provided that $V(x)$ is locally integrable and bounded below. And this can be satisfied by the classically non-deterministic potential by suitably modifying it away from the origin.[53]

[51] See [Landsman, this vol.] for details on various approaches to quantization.

[52] A linear operator $\hat{O}$ defined on the dense domain $D(\hat{O})$ of the Hilbert space $\mathcal{H}$ is *symmetric* just in case for all $\psi, \varphi \in D(\hat{O})$, $(\hat{O}\varphi, \psi) = (\varphi, \hat{O}\psi)$, where $(\cdot, \cdot)$ is the inner product on $\mathcal{H}$. That $\hat{O}$ is *self-adjoint* means that $\hat{O} = \hat{O}^*$, i.e. $\hat{O}$ is symmetric and $D(\hat{O}) = D(\hat{O}^*)$, where $\hat{O}^*$ is adjoint of $\hat{O}^*$. Here $D(\hat{O}^*)$ is defined to be the set of $\varphi \in \mathcal{H}$ such that there is a $\chi \in \mathcal{H}$ such that $(\hat{O}\psi, \varphi) = (\psi, \chi)$ for all $\psi \in D(\hat{O})$; then $\hat{O}^*\varphi := \chi$.

[53] An appropriate dense domain is $\{\psi \in L^2_{\mathbb{C}}(\mathbb{R}, dx) : \psi, \psi' \in AC(\mathbb{R})\ \&\ \hat{H}\psi \in L^2_{\mathbb{C}}(\mathbb{R}, dx)\}$ where

Another form of classical indeterminism occurs when the initial value problem has locally in time a unique solution but the solution breaks down after a finite time. The example given in Section 3.6 was that of a system of classical point mass particles moving under the influence of their mutual pairwise attractive $1/r^2$ force, and it was noted that the solution can break down either because of collision or non-collision singularities. Neither type of singularity occurs in the quantum analogue since again the quantum Hamiltonian operator for this case is essentially SA.[54]

QM also cures the indeterminism of the Zeno version of classical billiards discussed in Section 3.7, at least in a backhanded sense. A putative quantum analogue would mimic the Zeno construction of an infinite number of distinct particles in the unit interval $(0, 1]$ by squeezing into that interval an infinite number of wave packets with substantial non-overlap. The latter would require that the uncertainty in position Δx associated with a wave packet becomes arbitrarily small as the origin is approached. By the uncertainty principle, the uncertainty in momentum Δp would have to become unboundedly large as the origin is approached. This represents a breakdown in the putative quantum analogue if Δx and Δp both small in comparison with some specified macroscopic standard is required for mimicing classical behavior.[55]

5.3 How QM (even without state vector reduction) can be a lot less deterministic than classical mechanics

Determinism for the evolution of the quantum state is an all-or-nothing affair. If $\hat{H}$ is essentially SA then the "all" alternative applies since, as already noted, the exponentiation of the unique SA extension gives a unitary evolution operator which is defined for all times and all vectors in the Hilbert space. If $\hat{H}$ is not essentially SA there are two possibilities to consider. The first is that $\hat{H}$ has no SA extensions. This can be dismissed on the grounds that $\hat{H}$ should be a real operator, and every real symmetric operator has SA extensions. The second possibility is that $\hat{H}$ has many SA extensions. Then the "nothing" alternative applies; for the exponentiations of the different SA extensions give physically distinct time evolutions. Roughly speaking, the different self-adjoint extensions correspond to different boundary conditions at the boundary points of the configuration space. Perhaps some boundary condition can be singled out and promoted to lawlike

$AC(\mathbb{R})$ stands for the absolutely continuous functions.

[54]This result is known as Kato's theorem; see [Kato, 1995, Remark 5.6]. For a more detailed discussion of the issue of essential self-adjointness and its implications for quantum determinism, see [Earman, 2005].

[55]Mimicking a classical state in which a particle has given values of position and momentum requires a quantum state ψ that not only returns the given values as expectation values but also gives $(\Delta x)_\psi$ and $(\Delta p)_\psi$ small in comparison with the relevant macroscopic standard; for if $(\Delta x)_\psi$ (respectively, $(\Delta p)_\psi$) is large in comparison with the standard, there is an appreciable probability that the particle will be found with a position (respectively, momentum) different from the given value.

status, thus providing for a unique quantum dynamics. But restoring determinism by this route would require a justification for the hypothesized singling out and promotion. Alternatively, the effects of the non-essential self-adjointness of the Hamiltonian can be down played if it can be shown that the quantum dynamics associated with different self-adjoint extensions all have the same classical limit (see [Earman, 2005b]).

A toy example of the second possibility is given by a particle moving on the positive real line $\mathbb{R}_+$ in a potential $V(x)$, $x \in \mathbb{R}$. If the potential has the form C/x^2, with $C > 0$, then the Newtonian initial value problem has a unique solution, and the solution is defined for all times. The intuitive explanation is that no matter how much energy it has, the classical particle cannot climb over the indefinitely high potential to reach the singularity at the origin, and it cannot reach $x = +\infty$ in finite time. However, the quantum Hamiltonian operator for this case is not essentially SA on $L^2_{\mathbb{C}}(\mathbb{R}_+, dx)$ if $C < \dfrac{3}{4}$ (see [Reed and Simon, 1975, Thm X.10]). The intuitive explanation is that the quantum particle can tunnel through the barrier to reach the singularity, allowing probability to leak away. This leakage is incompatible with unitary evolution, which would obtain as the result of exponentiating the unique SA extension of an essentially SA $\hat{H}$.

The singularity in the configuration space of the toy example is wholly artificial, being created by deleting half of the real line. But analogues in the form of naked or timelike singularities occur in general relativistic spacetimes (see Section 6.4). One can ask whether a relativistic quantum particle propagating on the background of negative mass Schwarzschild spacetime can tunnel through the effective repulsive barrier that surrounds the origin $r = 0$. Horowitz and Marolf [1995] show that the answer is positive.

Essential self-adjointness of the Hamiltonian might be promoted as a selection principle to help decide what systems are "quantum possible," guaranteeing that (barring state vector collapse) the evolution of the quantum is deterministic. Those who think that determinism is an *a priori* truth may look favorably on this promotion, but otherwise its merits are hard to discern.

5.4 Chaos and predictability in QM

QM can not only be more deterministic than classical mechanics, but it can also be more predictable as well. Classical predictability is compromised or completely wrecked for observers who cannot ascertain initial conditions with complete precision if the systems whose behavior they are attempting to predict display sensitive dependence on initial conditions. But if the quantum Hamiltonian operator is essentially SA, then not only is the evolution of the quantum state completely deterministic, its predictability is not compromised by sensitive dependence on initial conditions. The point is simply that the linear, unitary nature of the evolution preserves the Hilbert space norm: $||U(t)\psi_2 - U(t)\psi_1|| = ||U(t)(\psi_2 - \psi_1)||$ $= ||\psi_2 - \psi_1||$. In words, if two states are close (as measured in the Hilbert space norm) to begin with, they remain close for all times, i.e. the evolution is stable.

This stability causes trouble for anyone seeking chaos in QM itself — they are driven to such extremes as resorting to open systems (whose evolution is not unitary) or to hidden variables whose evolution is not stable.[56] But in itself the stability of quantum evolution poses no *a priori* barrier to explaining how chaos can emerge from quantum dynamics in some appropriate classical limit. For that project only requires showing that the expectation values of relevant classical quantities can diverge rapidly enough (in some appropriate metric) to underwrite higher order ergodic properties that constitute the chaotic behavior observed on the macrolevel (see [Belot and Earman, 1997]). One obvious way to carry out this project is to use Ehrenfest's theorem to show that in the position representation the centroid of a quantum wave packet will follow a classical trajectory, as long as the mean square width of the wave packet remains sufficiently small. However, for classically chaotic trajectories the time interval in which the latter condition holds is uncomfortably short — for example, [Zurek, 1998] estimates that for the chaotic tumbling of Hyperion (a moon of Jupiter) it is of the order of 20 years. Several authors have argued that quantum decoherence comes to the rescue (see, for example, [Zurek, 1998; 2003]), but that is a topic that is beyond the scope of this chapter. Clearly, classical chaos poses a challenge to our understanding of how the classical world emerges from quantum physics.[57] Another aspect of the classical-quantum correspondence is treated in the next section.

5.5 State vector reduction, hidden variables, and all that

Showing that the Hamiltonian operator $\hat{H}$ for a quantum system of interest is essentially SA is not enough to secure the fortunes of determinism for this system, and this for two reasons. The first is that the deterministic evolution of the quantum state might be interrupted by "state vector reduction," as is postulated in some treatments of the measurement problem in QM, by which the unitary evolution $\psi(0) \mapsto \psi(t) = \exp(-i\hat{H}t)\psi(0)$ is suspended and the quantum state jumps into an eigenstate of the observable being measured. In its crudest form state vector reduction is a literal miracle — a violation of the laws of nature — making it an inappropriate topic for the present forum. But there are more sophisticated forms of state vector reduction that model the reduction as a dynamical process. Stochastic models in which the reduction occurs discontinuously and continuously have been studied respectively by Ghirardi *et al.* [1986] and Pearle [1989]. Reduction by means of a non-linear term added to the Schrödinger equation was studied by Pearle [1976]. If the stochastic models of reduction are on the right track and if the stochastic mechanisms they postulate represent irreducible randomness, then obviously determinism is breached. By contrast, the scheme of Pearle [1976]

[56] See the discussions of Kronz [1998] and Cushing and Bowman [1999]. By contrast, physicists who study "quantum chaos" do not try to find chaos in QM itself but rather study the distinguishing properties of quantum systems whose classical counterparts display chaos. For this reason Michael Berry suggested replacing "quantum chaos" with "quantum chaology." Unfortunately the suggestion did not stick.

[57] For a comprehensive survey of this problem, see [Landsman, this vol.].

achieves a deterministic state vector reduction with the help of hidden variables.[58] None of these alternatives to standard non-relativistic QM has been generalized to a viable relativistic quantum field theory, and as far as I am aware none of them play any role in the main lines of research on quantum gravity that come from string theory or loop quantum gravity (see Section 8). Thus, at present it does not seem productive to speculate about the implications for determinism of possible modifications to QM that may or may not become part of some future physics. However, the motivation for introducing state vector reduction is relevant here, for it leads to the second set of reasons why the conventional quantum state dynamics may not be sufficient to secure determinism for the quantum domain.

Classical (= non-quantum) theories wear their interpretations on their sleeves.[59] For example, for a classical theory that admits an (unconstrained) Hamiltonian formulation, observables are in one-one correspondence with functions from the phase space $\Gamma(q, p)$ to (say) the reals $\mathbb{R}$. The intended interpretation is that if f_O is the function corresponding to the observable O, then the value of O at t is o iff the state $(q(t), p(t))$ at t is such that $f_O(q(t), p(t)) = o$. This scheme can be liberalized to allow for dispositional observables which have definite values in only some states; for such an O the representing function f_O is only a partial function. Another liberalization is to allow that the range of f_O includes "fuzzy" (e.g. interval) values. To get an interpretation of QM along similar lines requires adding to the formalism at least two things: (i) an account of which SA operators correspond to quantum observables, and (ii) a semantics for quantum observables in the form of a value assignment rule that specifies what values the observables take under what conditions. I will simply assume that part (i) of the interpretation problem has been solved.

The most obvious way to supply part (ii) would be to ape the classical value assignment rule, replacing the classical state space $\Gamma(q, p)$ by the quantum state space to get a value assignment rule of the form: the value of quantum observable O at t is o iff the state vector $\psi(t)$ is such that $f_O(\psi(t)) = o$ where f_O is the representing function for the quantum observable O. If, as implicitly assumed in this formulation, the quantum state space is taken to be the the unit sphere $\mathbb{S}\mathcal{H}$ of the Hilbert space $\mathcal{H}$ of the system (i.e. $\{\psi \in \mathcal{H} : (\psi, \psi) = 1\}$), then as far as standard QM is concerned, gauge freedom is present since any two elements of $\mathbb{S}\mathcal{H}$ that differ by a phase factor correspond to the same physical state in that all expectation values of observables are the same for the two quantum states. This gauge redundancy can be removed by taking the state space to be the projective Hilbert space $\mathbb{P}\mathcal{H}$, defined as the quotient of $\mathbb{S}\mathcal{H}$ by the action of $\psi \mapsto \varsigma\psi$ where $\varsigma \in \mathbb{C}$ with $|\varsigma| = 1$; equivalently, $\mathbb{P}\mathcal{H}$ is the space of rays or one-dimensional subspaces of $\mathcal{H}$. Thus, from the point of view of conventional QM, the value assignment rule should obey the restriction that $f_O(\psi) = f_O(\psi')$ whenever the

[58]The hidden variables are the phase angles, an idea revived by Ax and Kochen [1999]; see below.

[59]But recall from Section 3 that if determinism is demanded, then the initial on-the-sleeve interpretation may have to be modified by seeing gauge freedom at work.

unit vectors ψ' and ψ belong to the same ray. Allowing the value assignment to depend on the phase would amount to introducing "hidden variables," in the terminology used below.

In any case, if the quantum value assignment rule takes the form under discussion and if the problems discussed in Section 5.3 are waived, then arguably QM is a deterministic theory, and this is so even if f_O is a partial function (i.e. is undefined for some quantum states) or even if f_O can take fuzzy values. For assuming no state vector reduction, the state $\psi(0)$ at $t = 0$ uniquely determines the state $\psi(t)$ at any $t > 0$; and assuming the implementation of (ii) under discussion, the state $\psi(0)$ at $t = 0$ uniquely determines the value assignments at any later time $t > 0$. That an observable is assigned no value or a fuzzy value at t does not represent a failure of determinism, which requires only that the laws plus the initial state determine the present and future values of all observables *to the extent that these values are determinate at all*. Thus, on the present option it is a mistake to view the Kochen-Specker theorem, and other subsequent no-go theorems, as showing that QM does not admit a deterministic interpretation. Rather, what these no-go results show is that, subject to certain natural constraints,[60] some subset of quantum observables cannot all be assigned simultaneously sharp values.[61] The same goes for the Bell-type theorems, which are best interpreted as extensions of the Kochen-Specker no-go result to an even smaller set of observables (see [Fine, 1982a; 1982b]).

One example of the type of value assignment rule at issue is the eigenvalue-eigenvector rule which says that, for an observable O whose quantum operator $\hat{O}$ has a discrete spectrum, O has a sharp value at t iff $\hat{O}\psi(t) = o\psi(t)$, in which case $O(t) = o$. But it is just this eigenvalue-eigenvector link that leads to the notorious measurement problem in QM in the form of the inability of the theory to explain why measurements have definite outcomes, and it is this problem that motivated the idea of state vector reduction. In essence the problem arises because of the insistence that "measurement" should not be taken as a primitive term but should be analyzed within QM itself as a physical interaction between the object system and a measuring instrument. But while the application of the standard linear, unitary dynamics to the composite object-system + measurement-apparatus-system can establish a one-one correlation between the eigenstates of the object observable of interest and the eigenstates of the "pointer observable" of the measuring instrument, the application of the eigenvector-eigenvalue rule to the post measurement composite state function yields the unacceptable result that the "pointer" on the measuring instrument is not pointing to any definite value

[60]For example, it is natural to require that if the quantum value assignment rule for O assigns O a definite value, that value lies in the spectrum of the operator $\hat{O}$ corresponding to O. And it is natural to require that for suitable functions g, $F_{g(O)} = g \circ F_O$.

[61]It follows from Gleason's theorem that, subject to very plausible constraints on value assignments, not all of the (uncountably infinite number of) self-adjoint operators in a Hilbert space of dimension 3 or greater can be assigned simultaneously definite values belonging to the spectra of these operators. The Kochen-Specker theorem shows that the same conclusion can be drawn for a finite set of quantum observables. See [Redhead, 1987] for an account of these no-go results.

(see [Albert, 1992] for a detailed exposition). The Schrödinger cat paradox is a cruel illustration of this conundrum in which "live cat" and "dead cat" serve as the "pointer positions."

Thus, if the eigenvalue-eigenvector link is maintained, there are compelling reasons to contemplate a modification of the standard quantum dynamics in order to ensure that in measurement interactions the quantum state evolves into an eigenstate of the relevant observables. But since the decision was made above not to treat such modifications, the discussion that follows will be confined to the other option, namely, the use of a value assignment rule that breaks the eigenvalue-eigenvector link, possibly with the help of "hidden variables" that supplement the quantum state. If hidden variables X are used, the value assignment rule takes the form: the value of quantum observable O at t is o iff the total state $(\psi(t), X(t))$ is such that $f_O(\psi(t), X(t)) = o$, where again f_O stands for the representing function for the observable O but is now a function defined on the augmented state space. If the evolution of the total state is deterministic, then by the same argument as before, the quantum domain is fully deterministic if QM is true. An example is supplied by the Bohm interpretation where $X(t)$ specifies the positions of the particles at t. The quantum component of the total state evolves according to Schrödinger dynamics, and the postulated equation of motion for particle positions guarantees that $(\psi(0), X(0))$ uniquely determines $(\psi(t), X(t))$ for $t > 0$ (see [Holland, 1993] for a survey). On the Bohm interpretation many quantum observables have a dispositional character, taking on determinate values only in adequately specified contexts (typically including a measurement apparatus together with its hidden variables). For example, in the context of a Stern-Gerlach experiment a spin $\frac{1}{2}$ particle will have spin-up (or spin down) just in case the position of the particle lies the appropriate region of the apparatus. The validity of the claim that the Bohm interpretation resolves the measurement problem thus turns on whether all measurements can be reduced to position measurements.

The family of modal interpretations of QM also attempt to resolve the measurement problem by breaking open the eigenvalue-eigenvector link wide enough to allow measurements to have definite outcomes but not so wide as to run afoul of the Kochen-Specker type impossibility results (see [Bub, 1997] and [Dickson, this vol.] for overviews), but in contrast to the Bohm interpretation the modal interpretations have no commitment to maintaining determinism. Very roughly the idea is that an observable associated with a subsystem of a composite system in state $\psi(t)$ has a definite value just in case the reduced density matrix of the subsystem is a weighted sum of projectors associated with an eigenbasis of the observable. This guarantees that in an idealized non-disturbing measurement interaction in which the pointer positions of the measuring instrument are perfectly correlated with the possible values of the object system observable being measured, both the pointer observable and the object system observable have definite values.[62]

[62]More generally, the interaction of a system with its environment will mean that "measurement" of the system is going on all the time. Thus, decoherence aids the modal interpretation by providing the conditions of applicability of the interpretation. In the other direction, deco-

Most forms of the modal interpretation supply the probabilities for an observable to have particular values, assuming that the conditions are appropriate for the observable to have a determinate value; but they are silent as to what the actual value is. Nevertheless, the actual possessed values of quantum observables can be taken to play the role of the hidden variables X, and one can ask whether the total state (ψ, X) can be given a deterministic dynamics. The answer is negative for versions of the modal interpretation discussed in the philosophy literature since these versions do not supply enough hidden variables to allow for determinism. For example, at the time $t > 0$ when an ideal measurement interaction is completed and the eigenstates of pointer position are perfectly correlated with eigenstates of the object observable, the standard modal interpretations say that both the object observable and the pointer observable have definite values. In different runs of the experiment these correlated observables have different values. But in all the runs the initial quantum state $\psi(0)$ is the same, and the experimental situation can be arranged so that modal interpretations say that the initial possessed values $X(0)$ are the same. This failure of determinism is of no concern to the modal theorist whose goal is to solve the measurement problem. To this end it is enough to show that there is a stochastic dynamics for possessed values that is compatible with the statistical predictions of QM. In fact, there is a vast array of such dynamics (see [Dickson, 1997] and [Bacciagaluppi, and Dickson, 1998]).

A different version of the modal interpretation, proposed by Ax and Kochen [1999], takes the option mentioned above of extending the standard quantum state space of rays $\mathbb{P}\mathcal{H}$ to unit vectors $\mathbb{S}\mathcal{H}$. Elements of the former are supposed to characterize statistical ensembles of systems while elements of the latter characterize individual systems. This extension allows the modal interpretation to specify what value an observable has, in circumstances when it has a definite value, and also to provide for a deterministic evolution of the augmented quantum state. It is postulated that the ensemble corresponding to a ray $\varsigma\psi$, $|\varsigma| = 1$, is composed of individual systems with phase factors ς having an initial uniformly random distribution, which accounts for the apparent failure of determinism.

Both the Bohm interpretation and the family of modal interpretations have difficulties coping with relativistic considerations. The former does not have any natural generalization to QFT, at least not one which takes seriously the lesson that in QFT fields are the fundamental entities and particles are epiphenomena of field behavior. The latter does possess a natural generalization to QFT, but it yields the unfortunate consequence that in situations that are standardly discussed, no subsystem observable has a definite value (see [Earman and Ruetsche, 2005] and the references therein).

Many worlds interpretations of QM can be given a literal or a figurative reading (see [Barrett, 1999] for an overview). On the literal reading there are literally many worlds in that spacetime splits into many branches which, from the branch

herence requires something akin to the modal interpretation, for otherwise it does not, contrary to the claims of its promoters, resolve the measurement problem. For more on decoherence, see [Landsman, this vol.].

time onwards, are topologically disconnected from one another (see, for example, [McCall, 1995]).[63] This form of many worlds can be described as a hidden variable interpretation by taking the hidden variables X to describe the spacetime branching and by taking the representing function f_O to be a mapping from the total state $(\psi(t), X(t))$ to a vector, possibly with infinitely many components labeled α, where the component α supplies the value at t of O in branch α. The fate of determinism then depends on whether or not the story of when and how branching takes place makes the evolution of the total state (ψ, X) deterministic. On the figurative reading of "many worlds" there is literally only one world, but there are many minds, contexts, perspectives, or whatever. Also there is no such thing as an observable O simpliciter but rather an observable O-in-context-α, denoted by O_α. If the representing function f_{O_α} is a function of the quantum state only, then determinism seems to be secured. However, our notation is defective in disguising the need for a specification of the contexts that are available at any given time. That specification is determined by the quantum state $\psi(t)$ alone if there is a "democracy of bases," i.e. any "branch" of $\psi(t)$ expressed as a linear combination of the vectors of any orthonormal basis of the Hilbert space of the system defines a context. Such a radical democracy seems incompatible with experience, e.g. in the Schrödinger cat experiment we either see a live cat or we see a dead cat, and we never experience a superposition of seeing a live and seeing a dead cat.[64] To overcome this difficulty some many world theorists propose to work with a preferred set of bases. The issue of determinism then devolves on the question of whether the specification of the set of preferred bases is deterministic. Even if the many worlds interpretation — on either the literal or figurative version — secures ontological determinism, the price seems to be a radical epistemic indeterminism: How do I know which branch of a splitting world or which context of a non-splitting world I am in? Being told that there is no "I" only an "I-in-branch-α" or an "I-in-context-α "is of no help when I — whichever I that is — have to make a prediction about the outcome of a measurement. Here all I can do is fall back on the statistical algorithm of QM. The many worlds interpretation seems to guarantee that even if the world is ontologically deterministic, it behaves, as far as anyone can judge, as if there is an irreducible stochasticity.

Although the discussion of the quantum measurement problem and its ramifications has been very sketchy, I trust it is sufficient to indicate why it is vain to hope for a simple and clean answer to the question of whether the truth of QM entails the falsity of determinism. To arrive at an answer to that question calls for winnowing the various competing interpretations of QM, a task that is far from straightforward, especially since judgments about how to perform the winnowing

[63]How to describe branching spacetimes within the context of conventional spacetime theories is a ticklish matter. Perhaps the most promising move is to hold on to the assumption that spacetime is a differentiable manifold but abandon the assumption that it is a Hausdorff manifold. However, non-Hausdorff manifolds can display various pathological properties that threaten determinism, e.g. geodesics can bifurcate. See Section 6.1.

[64]But how can we be sure? Perhaps momentary mental confusion is a superposition phenomenon.

are inevitably colored by attitudes towards determinism.

6 DETERMINISM IN CLASSICAL GTR

6.1 Einstein's revolution

Einstein's GTR was revolutionary in many respects, but for present purposes the initially most important innovation is that GTR sweeps away the notion — shared by all pre-GTR theories — of a fixed spacetime background against which the evolution of particles and fields takes place. In GTR the spacetime metric is a dynamical field whose evolution is governed by Einstein's gravitational field equations (EFE). Before discussing the issue of whether this evolution is deterministic, two preliminary matters need attention.

First, general relativists typically assume that the manifold $\mathcal{M}$ of a relativistic spacetime $\mathcal{M}, g_{ab}$ is Hausdorff.[65] Without this stipulation determinism would be in deep trouble. For example, non-Hausdorff spacetimes can admit a bifurcating geodesics; that is, there can be smooth mappings γ_1 and γ_2 from, say, $[0,1]$ into $\mathcal{M}$ such that the image curves $\gamma_1[0,1]$ and $\gamma_1[0,1]$ are geodesics that agree for $[0,b)$, $0 < b < 1$, but have different endpoints $\gamma_1(1)$ and $\gamma_2(1)$. According to GTR, the worldline of a massive test particle not acted upon by non-gravitational forces is a timelike geodesic. But how would such a particle know which branch of bifurcating geodesic to follow? Additionally, the (local) uniqueness of solutions to the initial value problem for EFE discussed below in Section 6.3 would fail if non-Hausdorff attachments were allowed.

Second, the reader is reminded that attention is being restricted to relativistic spacetimes $\mathcal{M}, g_{ab}$ that are temporally orientable, and it is assumed that one of the orientations has been singled out as giving a directionality to time. But even with this restriction in place, some of the spacetimes embodied in solutions to EFE are inimical to the formulation of global Laplacian determinism given in Section 2.1. For example, such spacetimes may not admit a global time function. Indeed, the spacetime of the Gödel cosmological model not only does not admit a global time function, but it does not even admit a single global time slice (spacelike hypersurface without edges) so that one cannot meaningfully speak of the universe-at-a-given-moment.[66]

One response would be to narrow down the class of physically acceptable models of GTR by requiring that, in addition to satisfying EFE, such models must also fulfill restrictions on the global causal structure of spacetime that rule out such monstrosities as Gödel's model and other models which contains closed timelike

[65] $\mathcal{M}$ is *Hausdorff* iff for any $p, q \in \mathcal{M}$ with $p \neq q$, there are neighborhoods $N(p)$ and $N(q)$ such that $N(p) \cap N(q) = \emptyset$. Of course, a manifold is (by definition) locally Euclidean and, therefore, locally Hausdorff.

[66] This is a consequence of three features of Gödel spacetime: it is temporally orientable (i.e. it admits a continuous non-vanishing timelike vector field), it is simply connected, and through every spacetime point there passes a closed future directed timelike curve. For a description of the Gödel solution, see [Hawking and Ellis, 1973, 168–170] and [Malament, 1984].

curves. This move has the independent motivation of avoiding the "paradoxes of time travel."[67] But much stronger causality conditions are needed to underwrite the global version of Laplacian determinism in the general relativistic setting.

In the first place, to carry over the classical conception of Laplacian determinism to the context of a general relativistic spacetime requires that the spacetime admit a global time function, which is not guaranteed by the absence of closed timelike curves. But even the requirement of a global time function is not strong enough because it provides no guarantee that the level surfaces of *any* such function will have the Cauchy property. To be at home, Laplacian determinism requires a spacetime $\mathcal{M}, g_{ab}$ that is *globally hyperbolic*, which is the conjunction of two conditions: first, $\mathcal{M}, g_{ab}$ must be *strongly causal* in that for any $p \in \mathcal{M}$ and any neighborhood p there is a subneighborhood such that once a future directed causal curve leaves, it never reenters (intuitively, there are no almost closed causal curves); and second, for every $p, q \in \mathcal{M}$, the causal diamond $J^+(p) \cap J^-(q)$ is compact.[68] Global hyperbolicity guarantees that $\mathcal{M}, g_{ab}$ can be foliated by Cauchy surfaces and that $\mathcal{M}$ is diffeomorphically Σ x $\mathbb{R}$, where Σ is an $n-1$ dimensional manifold if $\dim(\mathcal{M}) = n$. But simply stipulating global hyperbolicity has all the virtues of theft over honest toil. So let us see what can be achieved by honest toil.

6.2 Determinism and gauge freedom in GTR

For pre-relativistic theories a constant theme was that creating an environment friendly to determinism requires willingness to either beef up the structure of the background spacetime or else to see gauge freedom at work in sopping up the apparent indeterminism (recall Section 3.3). But in GTR there is no fixed background structure. Thus, one would expect that GTR either produces indeterminism or else that there is a non-trivial gauge symmetry at work. This expectation is not disappointed.

To see why it is necessary to be more detailed about the EFE:

$$(16) \quad R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} = \kappa T_{ab}$$

where R_{ab} and $R := R^c_c$ are respectively the Ricci tensor (which is defined in terms of g_{ab} and its derivatives) and the Ricci scalar, Λ is the cosmological constant, and T_{ab} is the stress-energy tensor. The cosmological constant can be ignored for present purposes, but it is currently the object of intense interest in cosmology since a positive Λ is one of the candidates for the "dark energy" which is driving the accelerating expansion of the universe (see [Ellis, this vol.]).

A potential model of the theory is then a triple $\langle \mathcal{M}, g_{ab}, T_{ab} \rangle$ where g_{ab}, T_{ab} satisfy (17) at all points of $\mathcal{M}$. Building such a model seems all too easy: start

---

[67] But see [Earman, Smeenk, and Wüthrich, 2005] which argues that the so-called paradoxes of time travel do not show that time travel is conceptually or physically impossible.

[68] $J^+(p)$ (respectively, $J^-(p)$) denotes the *causal future* (respectively, *causal past*) of p, i.e., the set of all points q such that there is a future directed causal curve from p to q (respectively, from q to p).

with any any general relativistic spacetime $\mathcal{M}, g_{ab}$, compute the Einstein tensor $G_{ab} := R_{ab} - \frac{1}{2}Rg_{ab}$, and define the stress-energy tensor by $T_{ab} := \kappa G_{ab}$. Thus, the understanding must be that T_{ab} arises from a known matter field. And in order to make the environment as friendly as possible for determinism, it will be assumed that the T_{ab}'s that are plugged into the right hand side of (17) fulfill the dominant energy condition (see Section 4.1) which, together with the local conservation law $\nabla^a T_{ab} = 0$ (which itself is a logical consequence of (17)), guarantees that matter-energy does not propagate faster than light.

Even with these enabling stipulations in place, it seems at first glance that determinism gets no traction, at least not if a naively realistic interpretation is given to the models of the theory. The difficulty can be made apparent by repeating a variant of the construction given in Section 3.2. Let $\langle \mathcal{M}, g_{ab}, T_{ab} \rangle$ be a model satisfying all of the above stipulations, and suppose that the spacetime $\mathcal{M}, g_{ab}$ satisfies all of the causality conditions you like, e.g. that it is globally hyperbolic. Since there is no fixed background structure to respect, save for the topological and differentiable structure of $\mathcal{M}$, one is free to choose a diffeomorphism $d : \mathcal{M} \to \mathcal{M}$ such that d is the identity map on and to the past of some Cauchy surface Σ of $\mathcal{M}, g_{ab}$ but non-identity to the future of Σ. Then $\langle \mathcal{M}, d^* g_{ab}, d^* T_{ab} \rangle$, where d^* indicates the drag along by d, will also be a model satisfying all of the stipulations imposed on $\langle \mathcal{M}, g_{ab}, T_{ab} \rangle$. By construction, $d^* g_{ab}(p) = g_{ab}(p)$ and $d^* T_{ab}(p) = T_{ab}(p)$ for all p on or to the past of Σ, but $d^* g_{ab}(p) \neq g_{ab}(p)$ and $d^* T_{ab}(p) \neq T_{ab}(p)$ for some points p to the future of Σ (unless we have inadvertently chosen a d that is a symmetry of g_{ab} and T_{ab}, which can always be avoided). The existence of this pair of models that agree for all past times but disagree in the future is a violation of even the weakest cousin of Laplacian determinism, at least if the surface structure of the theory — tensor fields on a manifold — is taken at face value.

When this sort of construction threatened to undermine determinism in a pre-GTR setting, two options were available for shoring up determinism: add more structure to the background spacetime or reject the container view of spacetime. The first option is ineffective unless the additional elements of spacetime structure are non-dynamical objects, but this represents a retreat from one of the key features Einstein's revolution. If there is to be no retreat, then the second option must be exercised. In the present context the option of rejecting the container view of spacetime takes the form of rejecting the naive realism that reads the theory as describing tensor fields living on a manifold.

Choosing the second option has a principled motivation which is not invented to save determinism in GTR but which follows in line with the treatment of gauge symmetries in pre-general relativistic theories. The field equations (17) of GTR are the Euler-Lagrange equations derived from an action principle that admits the diffeomorphism group as a variational symmetry group. Thus, Noether's second theorem applies, indicating that we have a case of underdetermination — more "unknowns" than there are independent field equations — and that arbitrary functions of the spacetime variables will show up in solutions to the field equations.

Switching from the Lagrangian to the Hamiltonian formulation, it is found, as expected, that GTR is a constrained Hamiltonian theory. There are two families of first class constraints, the momentum constraints and the Hamiltonian constraints.[69] Unfortunately the Poisson bracket algebra of these constraints is not a Lie algebra,[70] and consequently one cannot directly associate the diffeomorphism group, which acts on the spacetime, with a group which acts on the Hamiltonian phase space by finding a natural homomorphism of the Lie algebra of the diffeomorphism group into the constraint algebra. This glitch is overcome by Isham and Kuchař [1986a; 1986b] who show that if appropriate embedding variables and their conjugate momenta are used to enlarge the phase space, then the enlarged constraint algebra is a Lie algebra, and that there exists a homomorphism of the Lie algebra of the spacetime diffeomorphism group into the new constraint algebra. Thus, the standard apparatus for treating gauge symmetries can be applied, yielding the verdict that the diffeomorphism invariance of GTR is to be interpreted as a gauge symmetry. On this interpretation, the above construction does not demonstrate that GTR is indeterministic but rather produces a *faux* violation of determinism by taking advantage of the redundancy of the surface structure theory in the sense of the many-to-one correspondence between the surface structure models and the intrinsic physical situation they describe; in particular, the models $\langle \mathcal{M}, d^*g_{ab}, d^*T_{ab} \rangle$ and $\langle \mathcal{M}, g_{ab}, T_{ab} \rangle$ in the above construction cannot threaten determinism since they are to be interpreted as different descriptions of the same physical situation. Of course, the apparatus at issue has built into it a commitment to determinism, so its application to GTR cannot be taken as part of a proof that the correct interpretation of GTR makes it a deterministic theory. The only claim being made here is that this determinism-saving move for GTR is not ad hoc but is part of a systematic approach to gauge symmetries that is taken to yield the "correct" results for pre-GTR theories.[71]

What is so clear using hindsight wisdom took Einstein many years of struggle to understand. His infamous "hole argument" can be seen as a discovery of this underdetermination problem.[72] What muddied the waters was a confusion between two senses of general covariance. *Formal general covariance* demands that the laws of motion/field equations of a theory be written in a form that makes them

[69]The plural is used here since there is a momentum constraint and a Hamiltonian constraint for every point of space.

[70]The bracket of a pair of the constraints is not always a linear combination of the constraints multiplied by a "structure constant." This failure of the constraint algebra to form a Lie algebra means that GTR is not a gauge theory in the sense of Yang-Mills. But it certainly does not mean that GTR does not contain non-trivial degrees of gauge freedom.

[71]Because these matters are surrounded by so much controversy in the philosophical literature, I want to emphasize as strongly as possible that I am not proposing a new way of looking at GTR but am simply expounding what is the standard view among general relativists; see, for example, [Wald, 1984].

[72]See [Norton, 1984] and [Stachel, 1986] for accounts of how the "hole argument" figured in Einstein's search for his gravitational field equations. And see [Rovelli, this. vol, Ch. 12] for an account of how reflecting on the lessons of the "hole argument" influenced his understanding of classical GTR and his approach to quantum gravity.

covariant under arbitrary coordinate transformations. The terminology "formal" was chosen with malice aforethought since the demand of formal general covariance is a demand on the form rather than on the content of theory. For example, Newtonian and special relativistic theories can be reformulated, without change of content, so as to meet this demand. Indeed, the fact that Newtonian and special relativistic theories can be formulated in a completely coordinate-free manner already should make it clear that coordinates cannot matter.[73] *Substantive general covariance* demands diffeomorphism invariance (e.g. that for arbitrary diffeomorphism of $\mathcal{M}$, $(\mathcal{M}, d^* g_{ab}, d^* T_{ab})$ is a model of the theory if $(\mathcal{M}, g_{ab}, T_{ab})$ is) *and* that this diffeomorphism invariance is a gauge symmetry. Again the terminology "substantive" was chosen with malice aforethought since the demand of substantive general covariance is not automatically met, without change of content, for formally generally covariant Newtonian and special relativistic theories, at least not by the lights of apparatus for treating gauge symmetries that has been touted here (see [Earman, 2006]).

What invites confusion is the fact that a spacetime coordinate transformation can be taken to indicate either a relabeling of spacetime points or as indicating a (local) diffeomorphism. In the first guise these transformations are gauge transformations of an innocent kind: they relate the various coordinate representations of the intrinsic coordinate-free objects g_{ab} and T_{ab} obtained by taking the components of these objects in different coordinate systems. But there is nothing new here as regards GTR since exactly the same story holds for intrinsic coordinate-free presentations of pre-GTR theories. In the second guise, however, these transformations may or may not be gauge transformations — it depends on the content of the theory.

When he first discovered the underdetermination problem by means of the "hole argument," Einstein took it to reveal a real and intolerable form of underdetermination. To avoid it, he thought he had to reject formal general covariance as a *desideratum* for gravitational field equations. Only after wandering in the wilderness of non-covariant equations for almost three years did he re-embrace general covariance. In effecting the re-embrace Einstein did not speak the language of gauge symmetries (the terminology and the apparatus had not been invented), so he did not say that the gauge interpretation of GTR lowers the hurdle for determinism in that it requires only the uniqueness of evolution for gauge invariant quantities. But he said what amounts to the same thing; or rather he said it for a subclass of the gauge invariant quantities of GTR — what are called "point coincidences," i.e. things like the intersection of light rays.[74]

Many philosophers have traced Einstein's path in various ways. Very few of them, however, have owned up to the implications of where the path leads. If

[73] In the above presentation I have intentionally used the "abstract index" notation. Thus, g_{ab} stands for a symmetric, covariant tensor field that is defined in a coordinate-free manner as a bilinear map of pairs of tangent vectors to $\mathbb{R}$. This object can be represented by its coordinate components g_{jk} in a coordinate system $\{x^i\}$. The transformations between two such representations are gauge transformations, albeit trivial ones.

[74] See [Howard, 1999] for an account of Einstein's use of this term.

determinism in GTR is saved by treating diffeomorphism invariance as a gauge symmetry, then the only "observables" (= genuine physical magnitudes) of GTR are gauge invariant quantities. This is easy enough to say, but what exactly is the nature of the gauge invariant structure that underlies the surface structure? This is a crucial issue for those physicists who pursue a quantum theory of gravity by applying some version of the canonical quantization program to GTR, for on this program it is the "observables" of classical GTR that will be turned into quantum observables in the sense of self-adjoint operators on the Hilbert space of quantum gravity. There is no standard answer to the question of how best to characterize the observables of classical GTR. But one thing is sure: none of the familiar quantities used in textbook presentations of GTR, not even scalar curvature invariants such as the Ricci scalar appearing in the field equations (17), count as observables in the sense under discussion. And more particularly, no local quantities — quantities attached to spacetime points or finite regions — are gauge invariants. In this respect the gauge-free content of the theory has a non-substantivalist flavor. Whether this content can be characterized in a way that also satisfies traditional relationalist scruples remains to be seen.

A second closely related implication of treating the diffeomorphism invariance of GTR as a gauge symmetry concerns the nature of time and change. In the Hamiltonian formalism the implication takes the form of a "frozen dynamics." Applying to the Hamiltonian constraint of GTR the doctrine that first class constraints generate gauge transformations leads directly to the conclusion that motion in GTR is pure gauge. Put another way, the instantaneous states in the Hamiltonian formulation of the theory contain redundant structure, and any two such states, where one is generated from another by solving the Hamiltonian form of EFE, are equivalent descriptions of the same intrinsic, gauge invariant situation.[75]

For those who find these implications unpalatable, the heterodox moves mentioned in Section 3.3 may be attractive. As far as I am aware, however, such heterodoxy as applied to GTR has not been seriously pursued by the physics community.

6.3 The initial value problem in GTR

For the sake of simplicity consider the initial value problem for the source-free or vacuum EFE, i.e. (17) with $T_{ab} \equiv 0$. Since these equations are second order in time, presumably the appropriate initial data consist of the values, at some given time, of the spacetime metric and its first time derivative. The technical formulation of this idea is to take an initial data set to consist of a triple (Σ, h_{ab}, k_{ab}), with the following features and intended interpretations. Σ is a three-manifold, which is to be embedded as a spacelike hypersurface of spacetime $\mathcal{M}, g_{ab}$. h_{ab} is a smooth Riemann metric on Σ, which will coincide with the metric induced on Σ by the spacetime metric g_{ab} when Σ is embedded as a spacelike hypersurface

[75]For more on the problem of time in GTR and quantum gravity, see [Belot and Earman, 1999], [Belot, this vol.], and [Rovelli, this vol.].

of $\mathcal{M}, g_{ab}$. And k_{ab} is a smooth symmetric tensor field on Σ that coincides with the normal derivative of h_{ab} when Σ is embedded as a spacelike hypersurface of $\mathcal{M}, g_{ab}$. A spacetime $\mathcal{M}, g_{ab}$ that fulfills all of these roles is said to be a *development* of the initial data set (Σ, h_{ab}, k_{ab}). If the development $\mathcal{M}, g_{ab}$ of the initial data set (Σ, h_{ab}, k_{ab}) satisfies the source-free EFE, then h_{ab} and k_{ab} cannot be specified arbitrarily but must satisfy a set of constraint equations. The existence and uniqueness result for the source-free EFE takes the following form[76]: Let (Σ, h_{ab}, k_{ab}) be an initial value set satisfying the constraint equations; then there exits a development $\mathcal{M}, g_{ab}$ of the initial data that is the unique — up to diffeomorphism — maximal Cauchy development satisfying the source-free field equations. Furthermore, g_{ab} depends continuously on the initial data (see [Hawking and Ellis, 1973] for details of the relevant topology).

Just as the proof of the well-posedness of the initial value problem for the homogeneous Maxwell equations exploits the gauge freedom in the electromagnetic potentials (see Section 4.2), so the existence and uniqueness proof for EFE exploits the idea that diffeomorphism invariance is a gauge symmetry of GTR. When the metric potentials g_{ij} (i.e. the coordinate components of the metric g_{ab}) are subjected to a gauge condition (called the harmonic coordinate condition), the EFE take the form of a system of quasi-linear, diagonal, second order hyperbolic pdes, which are known to have locally well-posed initial value problem.

That the development $\mathcal{M}, g_{ab}$ of the given initial data is a Cauchy development means that Σ is a Cauchy surface of $\mathcal{M}, g_{ab}$ (and, thus, that this spacetime is globally hyperbolic). That it is the maximal Cauchy development means that there is no proper extension of $\mathcal{M}, g_{ab}$ which is a solution of the source-free EFE and for which Σ is a Cauchy surface. The up-to-diffeomorphism qualifier to uniqueness was to be expected from the discussion of gauge freedom in the previous subsection, and in turn the presence of this qualifier shows that the heuristic discussion given there can be given precise content. Here the qualifier means that if $\mathcal{M}', g'_{ab}$ is any other maximal development satisfying the source-free EFE, then there is a diffeomorphism $d : \mathcal{M} \to \mathcal{M}'$ such that $d^* g_{ab} = g'_{ab}$.

Curie's Principle (see Section 2.3 above) would lead one to believe that a symmetry of the initial value set (Σ, h_{ab}, k_{ab}) for the vacuum EFE should be inherited by the corresponding solution. And so it is. Let $\varphi : \Sigma \to \Sigma$ be a diffeomorphism that is a symmetry of the initial data in the sense that $\varphi^* h_{ab} = h_{ab}$ and $\varphi^* k_{ab} = k_{ab}$. Then as shown by Friedrich and Rendall [2000, 216–217], if ψ is an embedding of Σ into the maximal Cauchy development determined by (Σ, h_{ab}, k_{ab}), there exists an isometry $\bar{\psi}$ of this development onto itself such that $\bar{\psi} \circ \varphi = \varphi \circ \psi$, i.e. there is an isometry of the maximal Cauchy development whose restriction to $\varphi(\Sigma)$ is ψ. Moreover, this extension of the symmetry of the initial data is unique.

The initial value problem for the sourced EFE involves not only the stress-energy tensor T_{ab} but also the equations of motion for the matter fields that give rise to T_{ab} and, in particular, the coupling of these matter fields to gravity and to each other. Whether the coupled Einstein-matter equations admit an initial value formulation

[76]This formulation is taken from Wald [1984, Theorem 10.2.2].

and, if so, whether the initial value problem is well-posed are issues that have to be studied on a case-by-case basis. For what seem to be appropriate choices of coupling, the initial value problem for the combined Einstein-Klein-Gordon equations and the Einstein-Maxwell equations have existence and uniqueness results similar to that for the source-free Einstein equations. For other cases the results are not as nice.[77]

The results mentioned above demonstrate that substantive general covariance (in the sense that diffeomorphism invariance is a gauge symmetry) is compatible with having a well-posed initial value problem. But there is clearly a tension between the two, and so one can wonder just how tightly the combination of these two requirements constrains possible laws.[78]

6.4 Cosmic censorship and chronology protection

The positive results reported in the preceding section hardly exhaust the issue of determinism in GTR. One key concern is what happens when the maximal Cauchy development $\mathcal{M}, g_{ab}$ of initial data (Σ, h_{ab}, k_{ab}) satisfying the constraint equations is not maximal simpliciter, i.e. when $\mathcal{M}, g_{ab}$ can be imbedded isometrically as a proper subset of a larger spacetime $\mathcal{M}', g'_{ab}$ satisfying the source-free EFE. The analogous issue can also be raised for the case when $T_{ab} \neq 0$. The future boundary $H^+(\Sigma)$ of the (image of) the future domain of dependence $D^+(\Sigma)$ in the larger spacetime is called the *future Cauchy horizon* of Σ; the *past Cauchy horizon* $H^-(\Sigma)$ of Σ is defined analogously.[79] Intuitively, beyond the Cauchy horizons of Σ lie the regions of spacetime where the state of things is not uniquely fixed by the given initial data on Σ; for generally if the maximal Cauchy development $\mathcal{M}, g_{ab}$ of the initial data is not maximal simpliciter, then the larger extensions for which Σ is not a Cauchy surface are not unique (even up-to-diffeomorphism).

A relatively uninteresting reason why the maximal Cauchy development might be non-maximal simpliciter is that Σ was a poor choice of initial value hypersurface. A trivial but useful example is given by choosing Σ to be the spacelike hyperboloid of Minkowski spacetime pictured in Fig. 3. Here $H^+(\Sigma)$ is the past null cone of the point p.

Some features of this example generalize; in particular, $H^+(\Sigma)$ is always a null surface generated by null geodesics. The unfortunate case can be excluded by requiring, say, that Σ be compact or that it be asymptotically flat. Of course, these conditions exclude many cases of physical relevance, but for sake of discussion let us leave them in place. Even so, the maximal Cauchy development may fail to be maximal simpliciter for more interesting and more disturbing reasons.

---

[77] For comprehensive reviews of what is known, see [Friedrich and Rendall, 2000] and [Rendall, 2002].

[78] An analysis of gauge symmetries different from the one advertised here is given in [Geroch, 2004]. He gives only two examples of laws that have an initial value formulation and that have diffeomorphism invariance as a gauge symmetry (in his sense).

[79] More precisely, $H^+(\Sigma) := \overline{D^+(\Sigma)} - I^-(D^+(\Sigma))$, and analogously for $H^-(\Sigma)$.

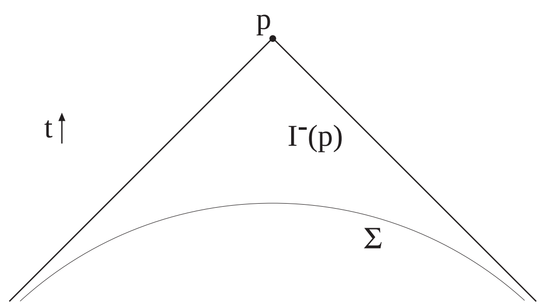

Figure 3. A poor choice of initial value hypersurface

One reason is that a spacetime can start with nice causal properties and evolve in such a way that these properties are lost. The point is illustrated by Misner's (1 + 1)-dim spacetime that captures some of the causal features of Taub-NUT spacetime, which is a solution to the source-free EFE. The Σ in Fig. 4 is a compact spacelike slice in the causally well behaved Taub portion of the spacetime, and its future Cauchy horizon $H^+(\Sigma)$ is a closed null curve. Crossing over this horizon takes one into a region of spacetime where there are closed timelike curves.

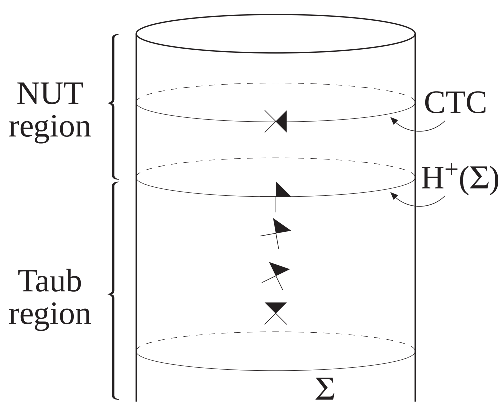

Figure 4. Misner spacetime

Another reason that the maximal Cauchy development may fail to be maximal simpliciter is illustrated in Fig. 5 which shows a non-compact asymptotically flat spacelike slice Σ on which a spherically symmetric ball of matter starts to undergo gravitational collapse. After a finite time the density of collapsing matter becomes infinite, creating a curvature singularity that is pictured as a timelike line in the figure. Strictly speaking, however, it makes no sense to call the singularity a timelike line since the singularity is not part of the spacetime.[80] But this makes no difference to the main point of relevance here; namely, a causal curve that terminates at a point to the future of $H^+(\Sigma)$ and that is extended into the past may fail to reach Σ, not because it has a past endpoint or because it gets trapped on $H^+(\Sigma)$ (as can happen in the spacetime of Fig. 4) but because it "runs into a singularity" or, better (since the singularity is not part of the spacetime), because it "runs off the edge of the spacetime."

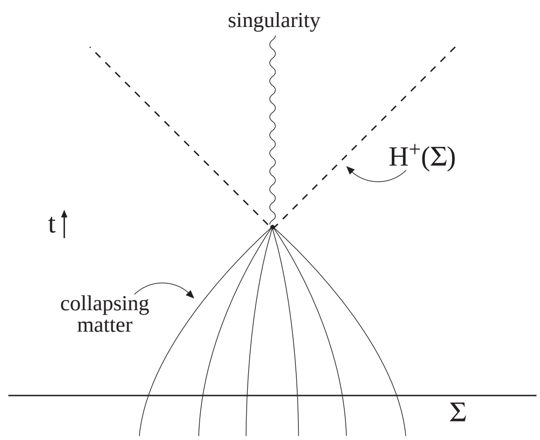

Figure 5. The development of a naked singularity in spherical gravitational collapse

It is known that EFE plus the imposition of the various energy conditions on T_{ab} discussed above do not suffice to prevent the kind of pathologies indicated by Figs. 4 and 5. But in all of such known examples there is something suspicious: either the matter fields involved are not "fundamental," i.e. even when gravity is turned off these matter fields are not well behaved in the sense that in Minkowski spacetime the initial value problem for their equations of motion do not admit global existence and uniqueness results (see Section 4.2), or else the initial conditions that eventuate in the singularity are very special, e.g. the initial configuration of matter in Fig. 5 required to produce the singularity must be perfectly spherically symmetric. One might conjecture that what holds for these examples holds generally: Consider a fundamental matter field that can serve as a

[80]One could try to attach the singular points as boundary points of the spacetime manifold. However, the extant prescriptions for doing this lead to counterintuitive features, e.g. the singular points need not be Hausdorff separated from interior points of the manifold; see [Geroch *et al.*, 1982].

source for gravitation. Then the subset of initial data for the Einstein-matter field equations for which the unique (up to diffeomorphism) maximal Cauchy development is not maximal simpliciter is of measure zero in the full space of such data, assuming poor choices of initial value hypersurfaces are excluded. To make this vague claim into a precise conjecture would require a specification of what matter fields are to be counted as fundamental, a specification of a suitable measure on the space of initial data, and a non-question begging specification of what counts as a poor choice of initial value hypersurface. The aimed-for conjecture is referred to as Penrose's *cosmic censorship conjecture*.

Less sweeping versions of the conjecture might focus specifically on one or another of the two types of pathologies illustrated in Figs. 4 and 5. Hawking's *chronology protection conjecture* aims to show that cases where closed timelike curves develop from a causally innocent past are highly non-generic among solutions to Einstein-fundamental-matter-field equations. The *weak cosmic censorship conjecture* aims to show that in generic solutions with asymptotically flat spacetimes, singularities are not "naked" in the sense of being visible to observers at infinity because whatever singularities develop (say, in gravitational collapse) are hidden inside of the event horizons of black holes which serve as one-way causal membranes that shield external observers from any of the pathologies of indeterminism that might develop within the horizon. Some progress has been made in formulating and proving precise versions of chronology protection and weak cosmic censorship, but the juries are still out on strong cosmic censorship.[81]

6.5 Predictability in general relativistic spacetimes

In Section 4.3 it was seen that the structure of Minkowski spacetimes has a double-edged quality with respect to determinism and predictability: while this structure makes possible clean examples of determinism, it also makes it impossible for embodied observers who must gather their information about initial conditions by means of causal interactions with the world to use determinism to perform genuine predictions. The point was formalized by defining the domain of predictability $P(q)$ of a point $q \in \mathcal{M}$ of a spacetime $\mathcal{M}, g_{ab}$ and noting that in Minkowski spacetime $P(q) = \varnothing$ for every q. Non-empty domains of predictability are obtained in the modified version of Minkowski spacetime with compactified space slices illustrated in Fig. 2. A feature of this case generalizes to arbitrary general relativistic spacetimes; namely, if the spacetime $\mathcal{M}, g_{ab}$ possesses a Cauchy surface Σ such that $\Sigma \subset I^-(q)$ for some $q \in \mathcal{M}$, then Σ is compact. Since a spacetime with a Cauchy surface Σ is diffeomorphically Σ x $\mathbb{R}$, the kind of complete predictability that comes with having $\Sigma \subset I^-(q)$ for some q is possible only in a spatially finite universe. The converse is not true: the existence of a compact Cauchy surface does not guarantee that there is a Cauchy surface Σ such that $\Sigma \subset I^-(q)$ for some q, de

[81] For an overview of progress on cosmic censorship, see [Chruściel, 1992; Isenberg, 1992; [Penrose, 1998; Wald, 1998]. And for an overview of progress on chronology protection, see [Earman *et al.*, 2005].

Sitter spacetime providing a relevant counterexample. Many interesting features of predictability in general relativistic spacetime are studied in [Hogarth, 1993].

6.6 Determinism and computability in general relativistic spacetimes

A *Plato machine* for gaining mathematical knowledge about an unresolved conjecture of number theory, whose prenex normal form is $(\forall n_1)(\forall n_2)...(\forall n_m)F(n_1, n_2, ..., n_m)$ or $(\exists n_1)(\exists n_2)...(\exists n_m)F(n_1, n_2, ..., n_m)$ with F recursive, can be conceptualized as an ordinary Turing machine run in Zeno fashion: enumerate the m-tuples of natural numbers and have the computer check in the first 1/2 minute whether F holds of the first tuple, check in the next 1/4 minute whether F holds of the second tuple, etc. At the end of the minute the truth of the conjecture is settled. Despite various claims to the contrary, I see no conceptual incoherency in such a device. But STR militates against the physical instantiation of such a device since the Zeno speed up would seem to require that some of the parts of the device must eventually move faster than the speed of light.[82]

General relativistic spacetimes seem to open the possibility of creating the functional equivalent of a Plato machine without Zeno tricks and without running afoul of the prohibition on superluminal propagation. Consider a spacetime with the following features. First, there is a timelike half-curve γ_1 with past endpoint, no future endpoint, and an infinite proper length. Second, there is another timelike half-curve γ_2 with past endpoint p and a point $q \in \gamma_2$ such that the proper time elapsed along γ_2 from p to q is finite and such that $\gamma_1 \in I^-(q)$. Such a spacetime has been dubbed a *Malament-Hogarth spacetime*. The theorems of any recursively axiomatizable theory — say, Zermelo-Frankel set theory — can be recursively enumerated, and a device whose worldline is γ_1 can utilize a Turing machine to effectively check each of these theorems to see one has the form "0 = 1". The device can be programmed to send out a signal — "Eureka!"— to an observer whose world line is γ_2 just in case "0 = 1" is found among the theorems. Assuming that the observer γ_2 is aware of this arrangement, she gains knowledge of the consistency/inconsistency of ZF: she knows that ZF is consistent just in case she has not received a "Eureka!" signal by the time she reaches the point q.

Similar arrangements can be used to "decide," at least in principle, Turing undecidable questions and to "compute" Turing uncomputable functions (see [Hogarth, 1994]). They, thus, threaten to falsify the physical Church-Turing thesis which asserts, roughly, that any physical computing device can be simulated by a Turing machine (see [Etsei and Németi, 2002] for a careful formulation of this thesis). In contrast to the original Church-Turing thesis which belongs to mathematical logic, the physical Church-Turing thesis lies in the borderland of mathematical logic and physics (see [Deutsch *et al.*, 2000]), and it is much harder to evaluate, especially if it is understood to require the physical realizability of the devices that implement the bifurcated supertask. Here I will confine myself to a few remarks

[82]Perhaps conflict with STR can be avoided by Zeno shrinking the parts, but this maneuver may run afoul of quantum restrictions.

on this matter and refer the interested reader to Németi and David [2005] for a fuller discussion.

Malament-Hogarth spacetimes are among the solutions of EFE — e.g. Reissner-Nordström spacetime and (the universal covering spacetime of) anti-De Sitter spacetimes. These particular spacetimes do not involve causal anomalies in the sense that they admit global time functions. However, all Malament-Hogarth spacetimes fail to be globally hyperbolic. Indeed, it can shown of such spacetimes that if $\Sigma \subset \mathcal{M}$ is any spacelike hypersurface such that the above defined γ_1 lies in $I^+(\Sigma)$, then any Malament-Hogarth point q whose chronological past contains γ_1 must lie on or beyond $H^+(\Sigma)$ (see Lemma 4.3 of [Earman, 1995, 117]). The possibility of non-deterministic influences, which might open the possibility that γ_1 receives a false "Eureka!" message, seems to undermine the use of Malament-Hogarth spacetimes for gaining knowledge in the sense of certainty. However, one should not draw hasty conclusions here since, as discussed in the following subsection, it is possible to have deterministic dynamics for fields propagating on a non-globally hyperbolic spacetime. Also it might seem that the problem is avoided by the fact that it can be arranged so that any signal from γ_1 arrives at γ_2 before the Malament-Hogarth point q and, thus, within $D^+(\Sigma)$. But since a "Eureka!" message can arrive arbitrarily close to q, the receiver must possess arbitrarily accurate discriminatory powers to separate signals that arrive before q from the potentially false signals that arrive after q.

6.7 The possibility of deterministic dynamics in non-globally hyperbolic spacetimes

For fields that propagate on a general relativistic spacetime, the failure of global hyperbolicity can undermine the initial value problem. For example, it is known that in generic two-dimensional spacetimes with closed timelike curves (CTCs) the scalar wave equation may fail to have smooth solutions or else may admit multiple solutions for the same initial data specified on a spacelike hypersurface. But remarkably, existence and uniqueness results have been proven for some four-dimensional spacetimes with CTCs (see [Friedman, 2004] for a review).

For spacetimes that do not have such blatant causal anomalies as CTCs but which nevertheless fail to be globally hyperbolic, Hilbert space techniques can sometimes be used to cure breakdowns in existence and uniqueness.[83] Consider a general relativistic spacetime $\mathcal{M}, g_{ab}$ that is static and possesses a global time function. The first condition means that there is a timelike Killing field V^a that is hypersurface orthogonal.[84] The second condition can be guaranteed by choosing

[83]The use of Hilbert space techniques to study problems in classical physics was pioneered by Koopman [1931]. However, Koopman's approach assumes determinism at the classical level and then shows how to represent this determinism as unitary evolution on a Hilbert space.

[84]The Killing condition is $\nabla_{(c}g_{ab)} = 0$ where ∇_a is the covariant derivative operator compatible with g_{ab}. Staticity guarantees that locally a local coordinate system (x^α, t) can be chosen so that the line element takes the form $ds^2 = g_{\alpha\beta}(x^\gamma)dx^\alpha dx^\beta - g_{44}(x^\gamma)dt^2$. Cf. Malament, this volume, section 2.7.

a spacelike hypersurface Σ orthogonal to V^a and by requiring that every integral curve of V^a meets Σ in exactly one point. Then every point $p \in \mathcal{M}$ can be labeled by (x, t), where $x \in \Sigma$ is the point where the integral curve of V^a through p meets Σ, and t is the value of the Killing parameter along this integral curve. Such a causally well behaved spacetime can nevertheless fail to be globally hyperbolic because, intuitively speaking, it possess a naked, timelike singularity. (To help fix intuitions, think of Minkowski spacetime with a timelike world tube removed. Or the reader familiar with GTR can think of the negative mass Schwarzschild solution to EFE, which is static and contains a timelike naked singularity at $r = 0$.) Now consider a massive $m \geq 0$ scalar field ϕ propagating on this background spacetime in accord with the Klein-Gordon equation (13). For the type of spacetime in question this equation can be rewritten in the form

$$(17) \quad \frac{\partial^2 \phi}{\partial t^2} = -A\phi$$

where t is the Killing parameter (see [Wald, 1980a], [Horowitz and Marolf, 1995]). The differential operator A can be considered a Hilbert space $\hat{A}$ operator acting on $L^2_{\mathbb{R}}(\Sigma, d\vartheta)$, where $d\vartheta$ is the volume element of Σ divided by $\sqrt{-V^a V_a}$. With the domain initially taken to be $C_0^\infty(\Sigma)$, $\hat{A}$ is a positive symmetric operator. The proposal is to replace the partial differential equation (18) with the ordinary differential equation

$$(18) \quad \frac{d^2 \phi}{dt^2} = -\hat{A}\phi$$

where the time derivative in (19) is a Hilbert space derivative. Since the Hilbert space operator $\hat{A}$ is real it has self-adjoint extensions, and since $\hat{A}$ is positive the positive square root of a self-adjoint extension $\hat{A}_e$ can be extracted to give

$$(19) \quad \phi(t) := \cos(\sqrt{\hat{A}_e} t)\phi(0) + \sin(\sqrt{\hat{A}_e} t)\dot{\phi}(0)$$

which is valid for all t and all $\phi(0), \dot{\phi}(0) \in \mathcal{H}$. Since $\phi(t)$ is a solution throughout the spacetime of the Klein-Gordon equation (13) and since it is the unique solution for given initial data $\phi(0), \dot{\phi}(0)$ on Σ, it provides (relative to the chosen self-adjoint extension) a deterministic prescription for the dynamics of the Klein-Gordon field. There are other possible prescriptions for obtaining the dynamics of ϕ, but Ishibashi and Wald [2003] have shown that the one just reviewed is the only one satisfying the following set of restrictions: it agrees locally with (18); it admits a suitable conserved energy; it propagates the field causally; and it obeys time translation and time reflection invariance. If the Hilbert space operator $\hat{A}$ is essentially self-adjoint, then the unique self-adjoint extension $\hat{A}_e$ provides *the* dynamics for the ϕ field satisfying the said restrictions. And this dynamics is fully deterministic despite the fact that the background spacetime on which the field propagates is not globally hyperbolic. Not surprisingly, however, $\hat{A}$ fails to be essentially self-adjoint for many examples of static but non-globally hyperbolic spacetimes, and unless further restrictions are added to single out one of the self-adjoint extensions, no unambiguous dynamics is specified by the above procedure. But remarkably, Horowitz and Marolf [1995] have provided examples of static, non-globally hyperbolic spacetimes where $\hat{A}$ *is* essentially self-adjoint, and in these cases the above prescription produces a dynamics of the ϕ field that is fully deterministic despite the presence of naked singularities.

7 DETERMINISM IN RELATIVISTIC QFT

Ordinary QM starts from a classical mechanical description of a system of particles — specifically, a Hamiltonian description — and attempts to produce a quantized version. Similarly, QFT starts from a classical relativistic description of a field and attempts to produce a quantized version. However, some classical fields do not lend themselves to a QFT that physicists find acceptable. Consider, for example, the non-linear wave equation (13) as a candidate for describing boson-boson interactions. A heuristic quantization procedure leads to the conclusion that there is no lowest energy state, leaving the system vulnerable to radiative collapse. On these grounds quantum field theorists have categorized the hypothetical interaction as "not physically realizable" (see [Baym, 1960]). That difficulties are encountered in QFT is perhaps not surprising when it is realized that the field in question is ill-behaved at the classical level in that regular initial data can pick out solutions that develop singularities within a finite amount of time. Is it plausible that deterministic behavior at the classical relativistic level can serve as a selection principle for what fields it is appropriate to quantize?

Determinism also plays a more constructive role in QFT. In ordinary QM, quantization involves the choice of a suitable representation of the canonical commutation relations $[\hat{x}_j, \hat{p}_k] = i\delta_{jk}$ (CCR). Since unbounded operators are involved, this form of the CCR only makes sense when the domains of the operators are specified. Such worries can be avoided by working with the Weyl, or exponentiated, form of the CCR, which also makes available the Stone-von Neumann theorem: for a finite number of degrees of freedom, the irreducible strongly continuous representations of the Weyl CCR are all unitarily equivalent — in fact, all are equivalent to the familiar Schrödinger representation. This theorem no longer applies when there are an infinite number of degrees of freedom, as in QFT, a feature of QFT that raises a number of interesting interpretational issues that are not relevant here. What is relevant is the fact that the construction of the CCR algebra for, say, the Klein-Gordon field in Minkowski spacetime, makes essential use of the deterministic propagation of this field (see [Wald, 1994]). This construction can be generalized to a Klein-Gordon field propagating in an arbitrary general relativistic background spacetime that is globally hyperbolic since the deterministic nature of the propagation carries over to the more general setting.

For a non-globally hyperbolic spacetime $\mathcal{M}, g_{ab}$ it is still the case that for any $p \in \mathcal{M}$ there is a neighborhood $\mathcal{N}(p)$ such that $\mathcal{N}, g_{ab}|_{\mathcal{N}}$, considered as a spacetime in its own right, is globally hyperbolic, and thus the field algebra $\mathcal{A}(\mathcal{N})$ for this mini-spacetime can be constructed by the usual means. One can then ask whether these local algebras can be fitted together to form a global algebra $\mathcal{A}(\mathcal{M})$ with the natural net properties (e.g. each such $\mathcal{A}(\mathcal{N})$ is a subalgebra of $\mathcal{A}(\mathcal{M})$, and if $\mathcal{N}_1 \subset \mathcal{N}_2$ then $\mathcal{A}(\mathcal{N}_1)$ is a subalgebra of $\mathcal{A}(\mathcal{N}_2)$). Kay (1992) calls the spacetimes for which the answer is affirmative *quantum compatible*, the idea being that non-quantum compatible spacetimes are not suitable arenas for QFT. A variety of non-globally hyperbolic spacetimes are not quantum compatible, e.g. 2-dim cylindrical spacetimes obtained from two-dimensional Minkowski spacetime by identifications along the time axis. But, remarkably, some acausal spacetimes have been shown to be quantum compatible (see [Fewster and Higuchi, 1996] and [Fewster, 1999]).

8 DETERMINISM AND QUANTUM GRAVITY

Arguably the biggest challenge in theoretical physics today is to combine the insights of GTR and QFT so as to produce a quantum theory of gravity (see [Rovelli, this vol.]). Some inkling of what this sought after theory will yield can perhaps be gained from semi-classical quantum gravity, which is a kind of shot-gun marriage of GTR and QFT. Semi-classical means that there is no attempt to quantize the metric of spacetime, but rather than merely treating a general relativistic spacetime as a fixed background on which quantum fields propagate (as in the preceding section), an attempt is made to calculate the back-reaction on the metric by inserting the quantum expectation value of the (renormalized) stress-energy in place of the classical stress-energy tensor on the right hand side of EFE (17). Although there may be no consistent theory underlying such a procedure, good theoretical physicists know how to extract usable information from it. Perhaps the most spectacular extraction is Hawking's conclusion that a black hole is not black but radiates exactly like a black body at temperature proportional to the surface gravity of the black hole. This *Hawking effect* is taken as confirmation that the formula for black hole entropy,[85] which had been derived by independent means, is more than a formal expression; it shows that black hole entropy is the ordinary thermodynamic entropy of a black hole (see [Wald, 1994]).[86] Theoretical physicists of different schools are in agreement that this is a stable result that has to be accommodated by an adequate quantum theory of gravity. But from this point on, the disagreements increase to the point of stridency.

The Hawking effect means that, when quantum effects are taken into account, black holes are not stable objects because the Hawking radiation must be accompanied by a diminution of the mass of the black hole. Presumably, as this process

[85] $S_{bh} = \frac{kc^3}{4G\hbar}A$, where A is the surface area of the black hole.

[86] The Hawking effect is related to, but distinct from, the *Unruh effect*. The latter effect is analyzed in terms of the apparatus of quantum statistical mechanics discussed in [Emch, this vol.]. In Minkowski spacetime the essence of the Unruh effect is that what an observer uniformly accelerated through the Minkowski vacuum experiences is described by a KMS state. The Unruh effect has been generalized to general relativistic spacetime; see [Kay and Wald, 1991].

goes deeper and deeper into the quantum regime, the semi-classical calculation will eventually break down. But *if* the continuation of the calculation can be trusted, then in the fullness of time the black hole will completely evaporate. (The estimated evaporation time for a black hole of solar mass is the order of 10^{67} years, much greater than the age of the universe. But this is no problem in a universe with an infinite future, as the latest cosmological measurements indicate is the case for our universe.) And *if* the result of the evaporation can be described by a classical general relativistic spacetime, the result is a momentarily naked singularity and a breakdown in global hyperbolicity, as is indicated in Fig. 6.[87] So even if some form of cosmic censorship holds for classical GTR, quantum effects seem to undo it.

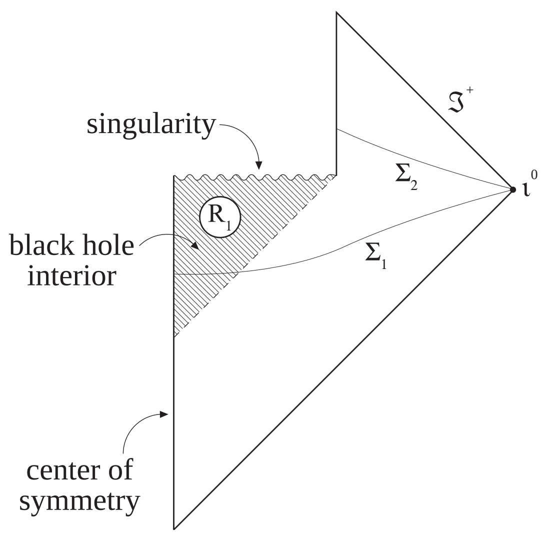

Figure 6. Conformal representation of black hole evaporation

Having gotten this far it is not difficult to establish that if at a time Σ_1 prior to the evaporation of the black hole the quantum field is in a pure state and if Hawking radiation has established correlations between relatively spacelike regions, such as the region R_1 in the black hole interior (see Fig. 6) and the region consisting of a "sandwich" about the post evaporation time Σ_2, then the state of the quantum field will be mixed at a post evaporation time Σ_2.[88] Since a pure-to-mixed state transition is necessarily non-unitary, the upshot is a loss of unitarity.[89]

[87]Following the conventions of conformal diagrams (see [Hawking and Ellis, 1973]), $\mathfrak{I}^+$ denotes future null infinity (the terminus of outgoing null rays), and ι^o denotes spatial infinity.

[88]This can be rigorously established in the algebraic formulation of QFT; see [Earman, 2002].

[89]And, incidentally, there is also a violation of time reversal invariance; see [Wald, 1986] and [Earman, 2002].

This "information loss paradox," as it is often referred to in the physics and the popular literature, has evoked an amazing variety of reactions; see [Belot *et al.*, 1999] for an overview. Most notable are the reactions from those who are so desperate to avoid the conclusion that they are willing to deploy "black hole complementarity"[90] and thereby abandon the mainstream reading of relativity theory, namely, that what the theory teaches us is that there is an intrinsic observer-independent reality — the very opposite of vulgar relativism that has it that everything is relative-to-an-observer.

But stepping back from the fray allows one to see that there is no need for such desperate measures. The pure-to-mixed evolution that is at the heart of the "paradox" need not be seen as a breakdown of quantum theory.[91] Nor is it surprising that consequences labeled 'paradoxical' flow from loss of global hyperbolicity. What needs to be questioned is whether this loss of global hyperbolicity is a plausible expectation of quantum gravity. Semi-classical quantum gravity suggests such a loss, but this way of bringing GTR and QFT together is at best a stepping stone to a full theory of quantum gravity. And just as ordinary QM showed the ability to smooth away singularities of classical mechanics, so the correct theory of quantum gravity may show the ability to smooth away the singularities of classical GTR.

Some positive indications come from the work of string theorists who are able to point to mechanisms that can smooth out singularities in classical general relativistic models; for example, Johnson *et al.* [2000] show that brane repulsion smooths out a class of naked singularities dubbed the *repulsion.* String theorists can also give a back door argument for exclusion of some types of classical singularities: postulate or prove that the sought after M-theory gives a stable ground state, and then note that this rules out analogues of the negative mass Schwarzschild solution and the like.

Other encouraging results come from loop quantum gravity (LQG), which aims to produce a quantum theory of gravity by applying to GTR a version of the canonical quantization based on a new set of canonical variables introduced by Amitaba Sen and exploited by Abay Ashtekar.[92] In the Friedmann-Robertson-Walker big bang models of classical GTR the scale factor a of spacetime goes to zero as the big bang singularity is approached, and the curvature blows up since

---

[90]Consider the case in STR of two inertial observers, O and O', who describe an ambient electromagnetic field using electric and magnetic fields $(\mathbf{E}, \mathbf{B})$ and $(\mathbf{E}', \mathbf{B}')$ respectively. There is a translation between the two descriptions which gives $\mathbf{E}'$and $\mathbf{B}'$ as functions of $\mathbf{E}$, $\mathbf{B}$, and the relative velocity of O and O' and vice versa with O and O' exchanged. The existence of such a translation follows from the fact that there an intrinsic, observer independent reality — in this case, the electromagnetic field as specified by the Maxwell tensor field. This tensor field is independent of coordinate systems, reference, frame, and observers. Contracting it with different velocity fields, representing the motions of different observers, results in different descriptions in terms of electric and magnetic fields. Whatever else it means, the "complementarity" part of "black hole complementarity" means that the different descriptions of an evaporating black hole given by two observers, one who falls through the black hole horizon and one who remains outside the horizon, are not related in the way the descriptions of O and O' are related.

[91]See [Wald, 1994, 181–182] and [Belot *et al.*, 1999].

[92]See [Rovelli, 2004] and [this vol.] for surveys of loop quantum gravity.

it scales as $1/a^2$.[93] Since there is no physically motivated way to extend such a solution through the initial singularity, the question of what happens "before" the big band belongs to theology or science fiction rather than science. The situation is startlingly different in LQG. Corresponding to the classical quantity $1/a$ there is a self-adjoint operator, acting on the Hilbert space of spatially homogeneous, isotropic quantum kinematical states, and its spectrum is bounded from above, giving a first indication that the classical singularity has been removed (see [Bojowald, 2001]). A complete proof of removal would require that the quantum dynamics gives an unambiguous evolution through the classical singularity. In LQG the "dynamics" is obtained by solving the Hamiltonian constraint equation, which restricts the physically allowed states. For the case at issue this constraint equation comes in the form of a difference equation rather than a differential equation. If the scale factor a is regarded as a "clock variable," then the constraint equation provides a "time evolution" of the quantum state through discrete steps of the clock variable. The key point is that this evolution equation does determine a unique continuation through the classical singularity.[94] However, what happens at the classical singularity is undetermined because the coefficient corresponding to this stage decouples from the other coefficients in the evolution equation (see [Ashtekar and Bojowald, 2003] for details).

From the point of view of determinism this last result means that the situation is somewhat ironic. Determinism is not threatened in classical GTR by the initial big bang singularity of the Friedmann-Robertson-Walker models because these models are globally hyperbolic, and because there is no physically motivated way to extend through the initial singularity. In LQG the initial singularity is banished both in the sense that curvature remains bounded and in the sense that there is a sensible way to extend through the classical singularity. But the price to be paid is a loss of determinism in LQG at the classical singularity, which can be seen as a Cheshire grin of the classical singularity.

Recently LQG has been used to resolve black hole singularities, leading to a new perspective on the Hawking information loss paradox in which Fig. 6 is not a valid depiction of black hole evaporation (see [Ashtekar *et al.*, 2005]). It is argued that, analogously to the FRW case, the quantum evolution continues through the classical singularity.[95] The new picture is not one in which global hyperbolicity is restored; indeed, that concept is not meaningful since what replaces the classical singularity is a region which cannot be described even approximately by the spacetime geometry of classical GTR. Nevertheless, it is argued that in the quantum

[93]The line element of FRW models can be written in the form $ds^2 = a(t)d\sigma^2 - dt^2$, where $d\sigma^2$ is the spatial line element.

[94]But see [Green and Unruh, 2004] where it is shown that in a spatially closed FRW model, the use of the scale factor as a "clock variable" is problematic. And the situation in inhomogeneous cosmologies is much more delicate and complicated; see [Brunnemann and Thiemann, 2006a; 2006b].

[95]As in the FRW case, the Hamiltonian constraint equation becomes a difference equation. The "quantum evolution" comes from this equation by choosing a suitable "clock variable" and then following the quantum state through discrete steps of the clock variable.

evolution a pure state remains pure and, in this sense, no information is lost. In its present form the argument has a heuristic character, and detailed calculations are needed to make it rigorous.

9 CONCLUSION

Is the world deterministic? Without the aid of metaphysical revelation, the only way we have to tackle this question is to examine the fruits of scientific theorizing. We can thus set ourselves the task of going through the theories of modern physics and asking for each: If the world is the way it would have to be in order for the theory to be true, is it deterministic? One of the things we discovered is that this task is far from straightforward, for the way in which theories are interpreted is colored by our attitudes towards determinism. For example, the unwillingness to see determinism fail at the starting gate in Newtonian gravitational theory militates in favor of taking gravitation to be a direct interparticle interaction and against assigning independent degrees of freedom to the Newtonian gravitational field. And an unwillingness to see determinism fail at the starting gate in GTR leads to the rejection of a naively realistic interpretation of the textbook version of the theory's description of spacetime and to the acceptance of diffeomorphism invariance as a gauge symmetry — which entails that none of the variables used in textbook presentations is counted as a genuine (= gauge invariant) physical magnitude.

The fortunes of determinism are too complicated to admit of a summary that is both short and accurate, but roughly speaking the story for classical (= non-quantum theories) is this. In Newtonian theories determinism is hard to achieve without the aid of various supplementary assumptions that threaten to become question-begging. For special relativistic theories determinism appears so secure that it is used as a selection criterion for "fundamental fields." GTR, under the appropriate gauge interpretation, is deterministic locally in time; but whether it is deterministic non-locally in time devolves into the unsettled issues of cosmic censorship and chronology protection.

Quantum physics is the strangest and most difficult case. Ordinary QM is in some respects more deterministic than Newtonian mechanics; for example, QM is able to cure some of the failures of Newtonian determinism which occur either because of non-uniqueness of solutions or the breakdown of solutions. But the fortunes of determinism in QM ultimately ride on unresolved interpretational issues. The main driving force behind these issues is the need to explain how QM can account for definite outcomes of experiments or more generally, the apparent definiteness of the classical world — an ironic situation since QM is the most accurate physical theory yet devised. Some of the extant responses to this explanatory challenge would bury determinism while others give it new life.

A new arena for testing the mettle of determinism is provided by the nascent quantum theories of gravity. There are some preliminary indications that just as ordinary QM was able to smooth out singularities of Newtonian mechanics, so quantum gravity effects may smooth out singularities of classical GTR. If this smoothing ability is broad enough it would alleviate worries that there are analogues in quantum gravity of breakdowns in determinism in classical GTR associated with failures of cosmic censorship. Quantum gravity will likely put a new face on the measurement problem and related interpretational issues that arise in ordinary QM. But it is too early to say whether this new face will smile on determinism.

AKNOWLEDGMENT

It is a pleasure to acknowledge the help and advice I have received from Gordon Belot, Michael Dickson, Klaas Landsman, Carlo Rovelli, and Laura Ruetsche. Special thanks are due to Jeremy Butterfield for numerous improvements. But, needless to say, there should be no presumption that any of the above mentioned share the opinions expressed herein.

BIBLIOGRAPHY

[Albert, 1992] D. Albert. *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press, 1992.

[Albert, 2000] D. Albert. *Time and Chance*. Cambridge, MA: Harvard University Press, 2000.

[Anderson and Thompson, 1992] I. Anderson and G. Thompson. *The inverse problem of the calculus of variations for ordinary differential equations. Memoirs of the American Mathematical Society*, No. 473. Providence, RI: American Mathematical Society, 1992.

[Ashtekar and Bojowald, 2005] A. Ashtekar and M. Bojowald. Black hole evaporation: a paradigm. 2005. *gr-qc/0504029*.

[Ashtekar *et al.*, 2003] A. Ashtekar, M. Bojowald, and J. Lewandowski. Mathematical structure of loop quantum cosmology. *Advances in Theoretical and Mathematical Physics*, 7: 233–268, 2003. *gr-qc/0304074*.

[Atmanspacher and Bishop, 2002] H. Atmanspacher and R. Bishop (eds.). *Between Chance and Choice: Interdisciplinary Perspectives on Determinism*. Thorveton: Imprint Academic, 2002.

[Ax and Kochen, 1999] J. Ax and S. Kochen. Extension of quantum mechanics to individual systems. 1999. *quant-ph/9905077*.

[Bacciagaluppi and Dickson, 1998] G. Bacciagaluppi and M. Dickson. Dynamics for modal interpretations. *Foundations of Physics,* 29: 1165–1201, 1998.

[Barbour, 1974] J. B. Barbour. Relative-distance Machian theories. *Nature*, 249: 328–329, 1974.

[Barbour, 1999] J. B. Barbour. The development of Machian themes in the twentieth century. In J. Butterfield (ed.), *The Arguments of Time*, pages 83–110. Oxford: Oxford University Press, 1999.

[Barbour and Bertotti, 1977] J. B. Barbour and B. Bertotti. Gravity and inertia in a Machian framework. *Nuovo Cimento*, 38B: 1–27, 1977.

[Bargmann, 1954] V. Bargmann. On unitary ray representations of continuous groups. *Annals of Mathematics*, 59: 1–46, 1954.

[Barrett, 1999] J. A. Barrett. *The Quantum Mechanics of Minds and Worlds*. Oxford: Oxford University Press, 1999.

[Baym, 1960] G. Baym. Inconsistency of cubic boson-boson interactions. *Physical Review*, 117: 886–888, 1960.

[Beig, 2005] R. Beig. Concepts of hyperbolicity and relativistic continuum mechanics. 2005. *gr-qc/0411092*.

[Belot and Earman, 1997] G. Belot and J. Earman. Chaos out of order: quantum mechanics, the correspondence principle, and chaos. *Studies in History and Philosophy of Modern Physics*, 28B: 147–182, 1997.

[Belot and Earman, 1999] G. Belot and J. Earman. From metaphysics to physics. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 166–186. Cambridge: Cambridge University Press, 1999.

[Belot *et al.*, 1999] G. Belot, J. Earman, and L. Ruetsche. The Hawking information loss paradox: The anatomy of a controversy. *British Journal for the Philosophy of Science*, 50: 189–229, 1999.

[Bojowald, 2001] M. Bojowald. Absence of a singularity in loop quantum cosmology. *Physical Review Letters*, 86: 5227–5230, 2001.

[Brading and Brown, 2003] C. Brading and H. Brown. Symmetries and Noether's theorems. In C. Brading and H. Brown (eds.), *Symmetries in Physics: Philosophical Reflections*, pages 89–109. Cambridge: Cambridge University Press, 2003.

[Brown, 1999] H. R. Brown. Aspects of objectivity in quantum mechanics. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 45–70. Cambridge: Cambridge University Press, 1999.

[Brunneman and Thieman, 2006a] J. Brunneman and T. Thiemann. On (cosmological) singularity avoidance in loop quantum gravity, *Classical nd Quantum Gravity*, 23: 1395-1427, 2006.

[Brunneman and Thieman, 2006b] J. Brunneman and T. Thiemann. Unboundedness of triad-like operators in loop quantum gravity, *Classical nd Quantum Gravity*, 23: 1429-1483, 2006.

[Bub, 1997] J. Bub. *Interpreting the Quantum World*. Cambridge: Cambridge University Press, 1997.

[Butterfield, 1998] J. Butterfield. Determinism and indeterminism. *Routledge Encyclopedia of Philosophy*, 3: 33–39. New York: Routledge, 1998.

[Callender, 2005] C. Callender. Time is the simplest (and strongest) thing. Pre-print, 2005.

[Carroll, 2004] J. Carroll (ed.). *Readings on Laws of Nature*. Pittsburgh, PA: University of Pittsburgh Press, 2004.

[Cassirer, 1956] E. Cassirer. *Determinism and Indeterminism in Modern Physics*. New Haven: Yale University Press, 1956.

[Chalmers, 1970] A. F. Chalmers. Curie's principle. *British Journal for the Philosophy of Science*, 21: 133–148, 1970.

[Chruściel, 1992] P. Chruściel. On uniqueness in the large of solutions of Einstein's equations ('strong cosmic censorship'). In M. Gotay, J. E. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 235–273. Providence, RI: American Mathematical Society, 1992.

[Courant and Hilbert, 1962] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, 2 vols. New York: Interscience Publishers, 1962.

[Curie, 1894] P. Curie. Sur la symmétrie des phénomènes physiques: symmétrie d'un champ électrique et d'un champ magnétique. *Journal de Physique*, 3: 393–415, 1894.

[Cushing, 1994] J. Cushing. *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*. Chicago: University of Chicago Press, 1994.

[Cushing and Bowman, 1999] J. Cushing and G. Bowman. Bohmian mechanics and chaos. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 90–107. Cambridge: Cambridge University Press, 1999.

[Deutsch *et al.*, 2000] D Deutsch, A. Ekert, and R. Lupacchini. Machines, logic, and quantum physics. *Bulletin of Symbolic Logic*, 6: 265–283, 2000.

[Dickson, 1997] W. M. Dickson. On the plurality of dynamics: transition probabilities and modal interpretations. In R. A. Healey and G. Hellman (eds.), *Quantum Measurement: Beyond Paradox. Minnesota Studies in the Philosophy of Science*, Vol. 17, pages 160–182. Minneapolis, MN: University of Minnesota Press, 1997.

[Dickson, 1998] W. M. Dickson. *Quantum Chance and Non-Locality: Probability and Non-Locality in the Interpretations of Quantum Mechanics*. Cambridge: Cambridge University Press, 1998.

[Driver, 1977] R. D. Driver. *Ordinary and Delay Differential Equations*. New York: Springer-Verlag, 1977.

[Driver, 1979] R. D. Driver. Can the future influence the present? *Physical Review D*, 19: 1098–1107, 1979.

[Earman, 1989] J. Earman. *World Enough and Space-Time: Absolute vs. Relational Theories of Space and Time*. Cambridge, MA: MIT Press/Bradford Books, 1989.

[Earman, 1995] J. Earman. *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. Oxford: Oxford University Press, 1995.

[Earman, 2002] J. Earman. What time reversal invariance is and why it matters. *International Journal for the Philosophy of Science*, 16: 245–264, 2002.

[Earman, 2003] J. Earman. Tracking down gauge: an ode to the constrained Hamiltonian formalism. In K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, pages 140–162. Cambridge: Cambridge university Press, 2003.

[Earman, 2004a] J. Earman. Determinism: what we have learned and what we still don't know. In J. K. Campbell, M. O'Rourke, and D. Shier (eds.), *Determinism, Freedom, and Agency*, pages 21–46. Cambridge, MA: MIT Press, 2004a.

[Earman, 2004b] J. Earman. Curie's principle and spontaneous symmetry breaking. *International Studies in Philosophy of Science*, 18: 173–199, 2004b.

[Earman, 2005] J. Earman. Essential self-adjointness: Implications for determinism and the classical-quantum correspondence. Pre-print, 2005b.

[Earman, 2006] J. Earman. Two challenges to the substantive requirement of general covaraince. *Synthese*, 148: 443-468, 2006.

[Earman and Roberts, 2006] J. Earman and J. Roberts. Contact with the nomic: a challenge to the deniers of humean supervenience. *Philosophy and Phenomenological Research*, in press, 2006.

[Earman *et al.*, 2005] J. Earman, C Smeenk, and C. Wüthrich. Do the laws of phyics forbid the operation of a time machine? Pre-print, 2005.

[Earman and Ruetsche, 2005] J. Earman and L. Ruetsche. Relativistic invariance and modal interpretations. *Philosophy of Science*, forthcoming, 2005.

[Etesi and Németi,, 2002] G. Etesi and I. Németi. Turing computability and Malament-Hogarth spacetimes. *International Journal of Theoretical Physics*, 41: 342–370, 2002.

[Fefferman, 2001] C. Fefferman. Existence and smoothness of the Navier-Stokes equation. Clay Mathematics Institute, Millennium Problems, 2001. *http://www.claymath.org/*

[Fewster, 1999] C. J. Fewster. Bisolutions to the Klein-Gordon equation and quantum field theory on 2-dimensional cylinder spacetimes. *Classical and Quantum Gravity*, 16: 769–788, 1999.

[Fewster and Higuchi, 1996] C. J. Fewster and A. Higuchi. Quantum field theory on certain non-globally hyperbolic spacetimes. *Classical and Quantum Gravity*, 13: 51–62, 1996.

[Fine, 1982a] A. Fine. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23: 1306–1310, 1982.

[Fine, 1982b] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48: 291–295, 1982.

[Foias *et al.*, 2001] C. Foias, O. Manley, R. Rosa, and R. Teman. *Navier-Stokes Equations and Turbulence*. Cambridge: Cambridge University Press, 2001.

[Friedrich and Rendall, 2000] H. Friedrich and A. Rendall. The Cauchy problem for the Einstein equations. In B. G. Schmidt (ed.), *Einstein's Field Equations and their Physical Implications: Selected Essays in Honor of Jürgen Ehlers*, pages 127–223. New York: Springer-Verlag, 2000.

[Friedman, 2004] J. L. Friedman. The Cauchy problem on spacetimes that are not globally hyperbolic. In P. T. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 331–346. Basel: Birkhäuser, 2004.

[Geroch, 1977] R. P. Geroch. Prediction in general relativity. In J. Earman, C. Glymour, and J. Stachel (eds.), *Foundations of Spacetime Theories. Minnesota Studies in the Philosophy of Science*, Vol. 8, pages 81–93. Minneapolis, MN: University of Minnesota Press, 1977.

[Geroch, 1995] R. P. Geroch. Relativistic theories of dissipative fluids. *Journal of Mathematical Physics*, 36: 4226–4241, 1995.

[Geroch, 2004] R. P. Geroch. Gauge, diffeomorphisms, initial-value formulation, etc. In P. T. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 441–477. Basel: Birkhäuser, 2004.

[Geroch *et al.*, 1982] R. P. Geroch, C. B. Liang, and R. M. Wald. Singular boundaries of spacetimes. *Journal of Mathematical Physics*, 23: 432–435, 1982.

[Ghirardi *et al.*, 1986] G. C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34: 470–491, 1986.

[Green and Unruh, 2004] D. Green and W. H. Unruh. Difficulties with recollapsing models in closed isotropic loop quantum cosmology. *Physical Review D*, 103502-1-7, 2004.

[Hadamard, 1923] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, 1923. Reprinted by Dover, New York, 1952.

[Hawking and Ellis, 1973] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press, 1973.

[Henneaux and Teitelboim, 1992] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems*. Princeton, NJ: Princeton University Press, 1992.

[Hogarth, 1993] M. Hogarth. Predicting the future in general relativistic spacetimes. *Studies in History and Philosophy of Science*, 24B: 721–739, 1993.

[Hoefer, 2003] C. Hoefer. Causal Determinism. *Stanford Encyclopedia of Philosophy*. *http://plato.stanford.edu/archives/sum2005/entries/determinism-causal/*

[Hogarth, 1994] M. Hogarth. Non-Turing computers and non-Turing computability. In D. Hull, M. Forbes, and R. M. Burian (eds.), *PSA 1994*, Vol. 1, pages 126–138, 1994.

[Holland, 1993] P. R. Holland. *The Quantum Theory of Motion*. Cambridge: Cambridge University Press, 1993.

[Horowitz and Marolf, 1995] G. Horowitz and D. Marolf. Quantum probes of spacetimes singularities. *Physical Review D*, 52: 5670-5675, 1995.

[Howard, 1999] D. Howard. Point coincidences and pointer coincidences: Einstein and the invariant content of space-time theories. In H. Goenner, J. Renn, J. Ritter, and T. Sauer (eds.), *The Expanding Worlds of General Relativity, Einstein Studies*, Vol. 7, pages 463–500. Boston: Birkhäuser, 1999.

[Isenberg, 1992] J. Isenberg. Progress on strong cosmic censorship. In M. Gotay, J. E. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 403–418. Providence, RI: American Mathematical Society, 1992.

[Isham and Kuchař, 1986a] C. J. Isham and K. Kuchař. Representations of spacetime diffeomorphisms. I. Canonical parametrized field theories. *Annals of Physics*, 164: 316–333, 1986.

[Isham and Kuchař, 1986b] C. J. Isham and K. Kuchař. Representations of spacetime diffeomorphisms. II. Canonical geometrodynmamics. *Annals of Physics*, 164: 288–315, 1986.

[Ismael, 1997] J. Ismael. Curie's principle. *Synthese*, 110: 167–190, 1997.

[Ishibashi and Wald, 2003] A. Ishibashi and R. M. Wald. Dynamics in non-globally hyperbolic static spacetimes: II. General analysis of prescriptions for dynamics. *Classical and Quantum Gravity*, 20: 3815–3826, 2003.

[John, 1982] F. John. *Partial Differential Equations*, 4th ed. New York: Springer-Verlag, 1982.

[Johnson *et al.*, 2000] C. V. Johnson, A. W. Peet, and J. Polchinski. Gauge theory and the excision of repulsion singularities. *Physical Review D*, 61: 08600-1-11, 2000.

[Kato, 1995] T. Kato. *Perturbation Theory for Linear Operators*. Berlin: Springer-Verlag, 1995.

[Kay, 1992] B. S. Kay. The principle of locality and quantum field theory on (non-globally hyperbolic) curved spacetimes. *Reviews of Mathematical Physics*, Special Issue, 167–195, 1992.

[Kay and Wald, 1991] B. Kay and R. M. Wald. Theorems on the uniqueness and thermal properties of stationary, quasifree states on spacetimes with a bifurcate Killing horizon. *Physics Reports*, 207: 49–136, 1991.

[Koopman, 1931] B. O. Koopman. Hamiltonian systems and transformations in Hilbert space. *Proceedings of the National Academy of Sciences*, 17: 315–318, 1931.

[Kronz, 1998] F. M. Kronz. Nonseparability and quantum chaos. *Philosophy of Science*, 65: 5–75, 1998.

[Keller, 1957] J. B. Keller. On solutions of nonlinear wave equations. *Communications on Pure and Applied Mathematics*, 10: 523–530, 1957.

[Lanford, 1974] O. E. Lanford. Time evolution of large classical systems. In J. Moser (ed.), *Dynamical Systems, Theory and Applications. Lecture Notes in Physics*, 38: 1–207. New York: Springer-Verlag, 1974.

[Laplace, 1820] P. S. Laplace. *Théorie analytique des probabilitiés*. Paris: V. Courcier, 1820.

[Lewis, 1973] D. Lewis. *Counterfactuals*. Cambridge, MA: Harvard University Press, 1973.

[Lichtenberg and Lieberman, 1991] A. J. Lichtenberg and M. A. Lieberman. *Regular and Chaotic Dynamics*, 2nd ed. New York: Springer-Verlag, 1991.

[Malament, 1984] D. Malament. Time travel in the Gödel universe. P. D. Asquith and P. Kitcher (eds.), *PSA 1984*, Vol. 2: 91–100, 1984.

[Malament, 2004] D. Malament. On the time reversal invariance of classical electromagnetic theory. *Studies in History and Philosophy of Modern Physics*, 35B: 295–315, 2004.

[McCall, 1995] S. McCall. Time flow, non-locality, and measurement in quantum mechanics. In S. Savitt (ed.), *Time's Arrow Today*, pages 155–172. Cambridge: Cambridge University Press, 1995.

[Moore, 1990] C. Moore. Unpredictability and undecidability in dynamical systems. *Physical Review Letters*, 64: 2354-2357, 1990.

[Moore, 1991] C. Moore. Generalized shifts: Unpredictability and undecidability in dynamical systems. *Nonlinearity*, 4: 199–230, 1991.

[Nagel, 1961] E. Nagel. *The Structure of Science*. New York: Harcourt, Brace, and World, 1961.

[Németi and David, 2005] I. Németi and D. David. Relativistic computers and the Turing barrier. *Journal of Applied. Mathematics and Computation*, forthcoming, 2005.

[Norton, 1984] J. D. Norton. How Einstein found his field equations: 1912-1915. In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity. Einstein Studies*, Vol. 1: 101–159. Boston: Birkhauser, 1984.

[Norton, 1999] J. D. Norton. A quantum mechanical supertask. *Foundations of Physics*, 29: 1265–1302, 1999.

[Olver, 1993] P. J. Olver. *Applications of Lie Groups to Differential Equations*, 2nd Ed.. New York: Springer-Verlag, 1993.

[Payne, 1975] L. E. Payne. *Improperly Posed Problems in Partial Differential Equations*. Philadelphia, PA: SIAM, 1975.

[Pearle, 1976] P. Pearle. Reduction of the state vector by a nonlinear Schrödinger equation. *Physical Review D*, 13: 857–868, 1976.

[Pearle, 1989] P. Pearle. Combining stochastic dynamical state-vector reduction with spontaneous localization. *Physical Review A*, 39: 2277–2289, 1989.

[Penrose, 1998] R. Penrose. The question of cosmic censorship. In R. Wald (ed.), *Black Holes and Relativistic Stars*, pages 103–122. Chicago, IL: University of Chicago Press, 1998.

[Pérez Laraudogoitia, 2001] J. Pérez Laraudogoitia. Supertasks. *The Stanford Encyclopedia of Philosophy*, 2001. *http://plato.stanford.edu/archives/win2001/entries/spacetime-supertasks/*

[Pérez Laraudogoitia, 2005] J. Pérez Laraudogoitia. Avoiding infinite masses. Preprint, 2005.

[Pitowsky, 1996] I. Pitowsky. Laplace's demon consults an oracle: computational complexity of prediction. *Studies in History and Philosophy of Modern Physics*, 27: 161–180, 1996.

[Planck, 1932] M. Planck. *Der Kausalbegriff in der Physik*. Leipzig: Johann Ambrosius Barth, 1932.

[Poincaré, 1952] H. Poincaré. *Science and Method*. New York: Dover Publications, 1952.

[Popper, 1982] K. R. Popper. *The Open Universe*. Totowa, NJ: Roman and Littlefield, 1982.

[Pour-el and Richards, 1981] M. B. Pour-el and J. I. Richards. The wave equation with computable initial data such that its unique solution is not computable. *Advances in Mathematics*, 39: 215–239, 1981.

[Pour-el and Richards, 1989] M. B. Pour-el and J. I. Richards. *Computability in analysis and physics*. New York: Springer-Verlag, 1989.

[Prince, 2000] G. E. Prince. The inverse problem in the calculus of variations and its ramifications. In P. J. Vassiliou and I. G. Lisle (eds.), *Geometric Approaches of Differential Equations*, pages 171–200. Cambridge: Cambridge University Press, 2000.

[Radicati, 1987] L. A. Radicati. Remarks on the early developments of the notion of symmetry breaking. In M. Doncel, A. Hermann, L. Michel, and A. Pais (eds), *Symmetries in physics (1600-1980)*, pages 195–207. Barcelona: Bellatera, 1987.

[Raju, 1994] C. J. Raju. *Time: Towards a Consistent Theory*. Dordrecht: Kluwer Academic, 1994.

[Redhead, 1987] M. Redhead. *Incompleteness, Nonlocality, and Realism*. Oxford: Clarendon Press, 1987.

[Reed and Simon, 1975] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. II: Fourier Analysis, Self-Adjointness*. New York: Academic Press, 1975.

[Reichl, 1992] L. E. Reichl. *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations*. New York: Springer-Verlag, 1992.

[Reichl, 2004] L. E. Reichl. *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations*, 2nd ed. New York: Springer-Verlag, 2004,

[Rendall, 2002] A. Rendall. Theorems on existence and global dynamics for the Einstein equations. *Living Reviews in Relativity*, 2002. *http://www.livingreviews.org/Articles/Volume5/2002-6rendall*

[Rice, 1953] H. G. Rice. Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, 74: 358–366, 1953.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity.* Cambridge: Cambridge University Press, 2004.

[Saari and Xia, 1995] D. Saari and J. Xia. Off to infinity in finite time. *Notices of the American Mathematical Society*, 42: 538–546, 1995.

[Scheffer, 1993] V. Scheffer. An inivisid flow with compact support in spacetime. *Journal of Geometry and Analysis*, 3: 341-401, 1993.

[Schnirelman, 1997] A. Schnirelman. On the nonuniquenes of weak solution of the Euler equation. *Communicatins on Pure and Applied Mathematics*, 50: 1261-1268, 1997.

[Skow, 2005] B. Skow. What makes time different from space. *Noûs*, forthcoming, 2005.

[Stachel, 1986] J. Stachel. Einstein's search for general covariance: 1912-1915. In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity. Einstein Studies*, Vol. 1: 63–100. Boston: Birkhauser, 1986.

[Stöltzner, 2003] M. Stöltzner. Vienna indeterminism: from Exner to Frank and von Mises. In P. Parrini, W. Salmon, and M. Salmon (eds.), *Logical empiricism: Historical and contemporary perspectives*, pages 194–229. Pittsburgh, PA: University of Pittsburgh Press, 2003.

[Suppes, 1991] P. Suppes. Determinism or instability. Does it matter? In G. G. Brittan (ed.), *Causality, Method, and Modalities*, pages 5–22. Dordrecht: Kluwer Academic, 1991.

[Suppes, 1993] P. Suppes. The transcendental character of determinism. In P. A. French, T. E. Uehling, and H. K. Wettstein (eds.), *Midwest Studies in Philosophy*, 18:, 242–257, 1993.

[van Fraassen, 1989] B. C. van Fraassen. *Laws and Symmetry.* Oxford: Oxford University Press, 1989.

[Wald, 1980a] R. M. Wald. Dynamics in non-globally hyperbolic, static spacetimes. *Journal of Mathematical Physics*, 21: 2802–2805, 1980a.

[Wald, 1980b] R. M. Wald. Quantum gravity and time reversibility. *Physical Review D*, 21: 2742–2755, 1980b.

[Wald, 1984] R. M. Wald. *General Relativity.* Chicago, IL: University of Chicago Press, 1984.

[Wald, 1994] R. M. Wald. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* Chicago, IL: University of Chicago Press, 1994.

[Wald, 1998] R. M. Wald. Gravitational collapse and cosmic censorship. In C. V. Vishveshwara, B. R. Iyer, and B. Bhawal (eds.), *Black Holes, Gravitational Radiation, and the Universe.* Dordrecht: Kluwer Academic Publishers, 1998.

[Weyl, 1932] H. Weyl. *The Open World.* New Haven, CN: Yale University Press, 1932.

[Widder, 1975] D. V. Widder. *The Heat Equation.* New York: Academic Press, 1975.

[Xia, 1992] X. Xia. Existence of noncollision singularities in the N-body problem. *Annals of Mathematics*, 135: 411–468, 1992.

[Zeipel, 1908] H. von Zeipel. Sur les singularités du problème des n corps. *Arkiv för Matimatik, Astronomie och Fysik*, 4: 1–4, 1908.

[Zurek, 1998] W. H. Zurek. Decoherence, chaos, quantum-classical correspondence, and the algorithmic arrow of time. *Physica Scripta*, T76: 186–198, 1998.

[Zurek, 2003] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75: 715–775, 2003.