NOTES

[1] If the only restrictions on the motions of the particles are that they move rectilinearly between collisions and behave like elastic billiard balls in collisions, then the motions are demonstrably non-deterministic except for some very special kinds of collisions; see Ch. III below.

[2] Laplace (1820), Preface; translation from E. Nagel (1961), pp. 281—282. Laplace seems to have given the wrong initial data problem for Newtonian gravitation; see Ch. III.

[3] There is an indirect but important connection between determinism and simplicity. Determinism (as I formulate it below) is a property of laws of nature, and simplicity is one of the features used to separate lawful from non-lawful regularities (see Ch. V).

[4] This assumes that the allowed instantaneous states are the same at every moment of time, an assumption which may fail if the laws are not time translation invariant (see Ch. VII). If the laws are time translation invariant, only the interval $t_2 - t_1$ matters and we can write $s(t_2) = F(s(t_1), t_2 - t_1)$. This is the sense in which the new Russell function need not involve time explicitly.

[5] This is the theme of most of the recent attempts to characterize natural laws; see Ch. V.

[6] Note, however, that Montague's approach is not without its potential pitfalls. Since any one of the standard formal systems of the type Montague studies is capable of representing at most a countable number of magnitudes, the possibility that there are an uncountable number of distinct physical magnitudes which interact with one another so as to produce a deterministic evolution has to be ignored. Russell's notion of determinism can be rehabilitated by requiring that there is a function which is definable in the formal system and which expresses the state at t in terms of t, t_0, and the state at t_0. Montague shows that for what he calls predicative theories this requirement is strictly stronger than determinism.

SUGGESTED READINGS FOR CHAPTER II

A fair sampling of how philosophers have sought to analyze the meaning of determinism is to be gained from Chs. 1 and 2 of Popper's (1982) *The Open Universe*, Russell's (1953) "On the Notion of Cause," E. Nagel's (1953) "The Causal Character of Modern Physical Theory," and Montague's (1974) "Deterministic Theories." The chapter on "Fate" from Taylor's (1983) *Metaphysics* and Cahn's (1967) *Fate, Logic, and Time* contain information on the standard philosophical views of fatalism.

CHAPTER III

DETERMINISM IN CLASSICAL PHYSICS

> All events, even those which on account of their insignificance do not seem to follow from the great laws of nature, are a result of it just as necessarily as the revolutions of the sun. In ignorance of the ties which unite such events to the entire system of the universe, they have been made to depend upon final causes or upon hazard, according as they occur and are repeated with regularity, or appear without regard to order; but these imaginary causes have gradually receded with the widening bounds of knowledge and disappear entirely before sound philosophy, which sees in them only the expression of our ignorance of the true causes.
>
> (P. S. Laplace, *A Philosophical Essay on Probabilities*)

This passage has been taken as a classic statement of determinism, and if it is then it is easy to appreciate how determinism came to occupy such an exalted status: if the only alternatives to determinism are final causes (e.g., divine intervention) and hazard (e.g., accident or chance), then determinism is attractive as an *a priori* truth or a methodological imperative of scientific inquiry. But some care is needed here, as already hinted in Ch. II; for Laplacian determinism as I have proposed to understand it need not be true even though all events are subject to laws that leave no room for divine intervention or accident. Classical physics would seem to be a poor choice of hunting grounds for such examples since, as we all know, the laws of classical physics are deterministic in the Laplacian sense. We know no such thing, at least if knowledge implies truth.

1. CLASSICAL WORLDS

The initial setting for the doctrine of determinism was what I called the classical world picture. It is time to be more specific about how that picture is composed. There are three features which require special emphasis. (1) All the members of the set $\mathscr{W}$ of physically possible classical worlds are assumed to have a common space-time background. This common space-time is the canvas on which the possible worlds are painted. The details of the structure of the canvas will turn out to be as crucial to the success or failure as what is painted on it: *too little structure of the right kind or too much structure of the wrong kind and determinism will never succeed no matter how furiously or cleverly we paint.* This important but largely unappreciated moral will be drawn in detail in this and succeeding chapters, but for now I will reemphasize only one element of classical space-time structure. Namely, (2) the four-dimensional space-time canvas is ruled by a family of three-dimensional hypersurfaces called the planes of absolute simultaneity; two events are simultaneous just in case they lie on the same plane. (3) The canvas is filled in by specifying the values of a collection of physical magnitudes, each of which is assumed to be a point valued quantity.

If, for sake of definiteness, we think of the physical magnitudes as geometric object fields on space-time, then classical worlds can be presented in the form of a triple $\langle M, \{G_\alpha\}, \{P_\beta\}\rangle$, $\alpha \in \mathscr{A}$, $\beta \in \mathscr{B}$ ($\mathscr{A}, \mathscr{B}$ index sets) where M is the space-time manifold (usually assumed to be $\mathbb{R}^4$), the G_α are geometric object fields on M characterizing the structure of space-time (including, of course, the simultaneity structure (2)), and the P_β are geometric object fields characterizing the physical contents of space-time. In keeping with (1), M and the G_α are common to all members of $\mathscr{W}$ while the P_β vary from world to world. Agreement of two worlds $\langle M, \{G_\alpha\}, \{P_\beta\}\rangle$ and $\langle M, \{G_\alpha\}, \{P'_\beta\}\rangle$ at a given time means agreement on a plane of absolute simultaneity of the values of the physical magnitudes.[1]

Modern physics contradicts or challenges each of the assumptions (1)–(3). The special theory of relativity contradicts (2); the general theory contradicts (1); and, according to some interpretations of quantum physics, quantum theory undermines (3). What happens to the doctrine of determinism when one or more of the props of the classical worlds is kicked out will have to be discussed in detail in later chapters. For the moment, let us assume that the props are secure. What is surprising is that even with their support, classical worlds prove to be an unfriendly environment for any form of Laplacian determinism. To the extent that determinism passes the Scylla of triviality, it appears to run a ground on the Charybdis of falsity. In Ch. II we viewed the Scylla. We must now face the Charybdis.

2. THE APPARENT FAILURE OF DETERMINISM IN LEIBNIZIAN PHYSICS

Some of the versions of Leibniz's multi-faceted principle of sufficient reason either entail or presuppose determinism. And yet, as Howard Stein (1977) has shown, Leibniz's views on the nature of space and time seem to preclude any interesting form of Laplacian determinism.

Let us recall Leibniz's version of the space-time structure of classical worlds. In accordance with the characterization of Sec. 1 above he agreed that there is an absolute notion of coexistence or simultaneity. And like all 17th century natural philosophers, he assumed that the instantaneous three-spaces have a Euclidean $\mathbb{E}^3$ structure and, further, that there is a well-defined sense of duration or temporal distance for non-simultaneous events. Finally — and this is the crucial point — he held that these elements completely exhaust the structure of space-time. The symmetry mapping of this Leibnizian space-time can be presented in the following form. Let x^α, $\alpha = 1, 2, 3$, stand for a Euclidean coordinate system; and let t stand for absolute time, i.e., $t: M \to \mathbb{R}$ is such that its level surfaces coincide with the planes of absolute simultaneity and the intervals $|t_1 - t_2|$ give the duration between the events e_1 and e_2 lying respectively on the planes $t = t_1$ and $t = t_2$. Then the symmetry maps have the form:

$$(L) \qquad x^\alpha \to x'^\alpha = R^\alpha_\beta(t)x^\alpha + a^\alpha(t)$$
$$t \to t' = t + b$$

where b is a constant, $a^\alpha(t)$ is an arbitrary smooth function of t, and $R^\alpha_\beta(t)$ is a time dependent orthogonal matrix. In words, the structure preserving maps of Leibnizian space-time onto itself are time translations and (possibly) time dependent Euclidean spatial translations and rotations.

On this canvas Leibniz painted a plenum of matter; but for ease of illustration it suffices to consider more sparsely populated worlds containing, say, three particles.

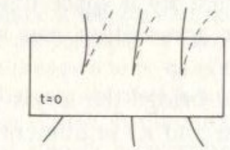

Fig. III.1

Such a world is completely described by drawing in the world lines of the three particles (the solid curves of Fig. III.1). Consider how the mappings (L) act on the particle histories. From among the mappings we can choose one which reduces to the identity for all $t \leqslant 0$ but which is non-trivial for $t > 0$. It leaves fixed the entire past history of the particles but changes their future behavior, as indicated schematically by the dashed lines. Thus, if the Leibnizian laws of motion satisfy the demand that the space-time symmetries are also symmetries of the laws (i.e., carry a physically possible history to another physically possible history), then we have a violation of even the weakest form of Laplacian determinism; for given any physically possible history of the particle trio, there will be another physically possible history which agrees with the first for all past times but disagrees in the future.

The announced demand is eminently reasonable, as is the stronger demand that the space-time symmetries and the symmetries of the laws of motion coincide. If the symmetries of the laws were more inclusive than the symmetries of the space-time, then the space-time would contain more structure than is needed to support the laws and Occam's razor would slice it away. On the other hand, the symmetries of the laws should be at least as wide as the symmetries of the space-time; for if the laws allow one history but not another, then those histories cannot be connected by a space-time symmetry — otherwise, there would be no way to express the difference between the allowed and the prohibited histories in terms of the behavior of physical magnitudes on the space-time canvas. Technically, the underlying assumption is that the set of models of the laws are closed under automorphisms of the space-time background, i.e., if $\langle M, \{G_\alpha\}, \{P_\beta\}\rangle$ is a model and d is a diffeomorphism of M onto itself such that $d^*G_\alpha = G_\alpha$ for each α, then $\langle M, \{G_\alpha\}, \{d^*P_\beta\}\rangle$ is also a model, where d^* denotes the 'drag along' by d. Conceivably, a theory of motion could postulate different lawlike behaviors in different space-time regions. But such a difference would be grounds for distinguishing the regions in terms of absolute structure; that is, for adding, if necessary, elements to $\{G_\alpha\}$ so that the regions in question are not connected by a space-time symmetry. And it seems that in this manner our assumption can always be vouchsafed (see Earman (1986) for details).

There are two ways to bridge the abyss which has opened between the vision of determinism and its fulfillment. One relies on the reinterpretation of Leibnizian space-time, the other on an enrichment of it. The two moves will be briefly reviewed in turn.

3. LEIBNIZ'S RESPONSE

Leibniz would have welcomed this challenge as an opportunity to expose the Achilles' heel of the Newtonian conception of the space-time manifold, or as he put it in the famous correspondence with Samuel Clarke, to "confute the fancy of those who take space to be a substance."

Note that the transformations (L) preserve all relative particle quantities such as relative distances, relative velocities, relative accelerations, etc. According to Leibniz, facts about the values of these relative quantities exhaust the factual content of the physical world. Thus, the 'two' world histories pictured in Fig. III.1 do not really correspond to objectively different worlds but only to different descriptions of the same world. Consequently, the alleged violation of determinism is only an illusion due to the descriptive fluff packed into our presentation of classical space-time worlds.

Leibniz's position here does not result from a question-begging desire to save determinism, but is arrived at by an independent route that passes through his meta-physics and his metaphysics. According to the former, which owes much to Descartes and Huygens, all motion must be analyzed as the relative motion of bodies. According to the latter, there would be a violation of the principle of sufficient reason if Fig. III.1 did illustrate objectively different world histories; for in deciding which of the two worlds to actualize, God would find Himself in a Buridan's ass situation, unable to choose between two worlds which are not separated by any properties that provide sufficient grounds for choice. As Leibniz put it in the third letter to Clarke:

> I say then, that if space was an absolute being, there would something happen for which it would be impossible there should be sufficient reason. Which is against my axiom. And I prove it thus. Space is something absolutely uniform; and, without the things placed in it, one point of space does not absolutely differ in any respect whatsoever from another point of space. Now from hence it follows, (supposing space to be something in itself, besides the order of bodies among themselves,) that 'tis impossible there should be a reason, why God, preserving the same situations of bodies among themselves, should have placed them in space after one certain particular manner, and not otherwise; why every thing was not placed the quite contrary way, for instance, by changing East into West. But if space is nothing else, but that order or relation; and is nothing at all without bodies, but the possibility of placing them; then those two states, the one such as now is, the other supposed to be quite the contrary way, would not all differ from one another. Their difference therefore is only to be found in the chimerical supposition of the reality of space in itself. But in truth the one would exactly be the same thing as the other, they being absolute indiscernible; and consequently there is no room to enquire after a reason of the preference of the one to the other. (Alexander (1956), p. 26)

The philosophical reaction to Leibniz's critique has tended to divide: those who share with him the notion that all motion is relative bodily motion are naturally sympathetic while those who are impressed by the fact that neither classcial *nor* relativistic physics supports this notion are less sympathetic. What both sides have failed to see (and what Leibniz himself was not clear about) is that the issue of relationism is not equivalent to the key issue Leibniz raises about our mode of presentation of space-time worlds. As I read it, his objection is first and foremost to the view that space-time is a kind of substance or container which exists over and above the events it houses. The objection can be stated in a form that is independent of the intertwined questions of whether all motion is the relative motion of bodies and what goes into the G_α. Let d be any diffeomorphism of the space-time manifold M onto itself. For fields G_α and P_β, d induces new ones d^*G and d^*P respectively (the fields 'dragged along' by d). Any two models $\langle M, \{G_\alpha\}, \{P_\beta\}\rangle$ and $\langle M, \{d^*G_\alpha\}, \{d^*P_\beta\}\rangle$ related in this way are by Leibniz's lights just different modes of presentation of the same physical reality. And this is so even if the structure of space-time, as specified by $\{G_\alpha\}$, falsifies the slogan "All motion is the relative motion of bodies," as it is falsified for orthodox Newtonian and special relativistic space-time, both of which contain inertial structure that permits the definition of absolute or invariant dynamical quantities, such as acceleration, which are *not* relative particle quantities (see below). Further, this is so even if space (or space-time) is not "absolutely uniform" but is, say, variably curved; for this curvature is represented by some appropriate object in $\{G_\alpha\}$ and is dragged along by d along with everything else so that again the original model and its image model "do not at all differ from one another" and are "absolutely indiscernible."

Thus, on my interpretation the essence of Leibniz's objection is to treating points and regions of M as real existents, as substances in the proper logical sense of objects of predication. There is a quick and cheap way to reform our presentation of space-time models so as to escape the objection; namely, take equivalence classes of "indiscernible" models and declare that each class corresponds to a single Leibnizian world. The more interesting challenge is to start from the other end and give a direct and intrinsic characterization of the Leibnizian worlds and then show that the members of an equivalence class of ordinary models arise as different but equivalent representations of the same intrinsic reality. For someone like myself who is not a relationist and who does not believe that all motion is the relative motion of bodies, the challenge takes the form of erasing the underlying manifold M of

space-time points while keeping the non-relational structure of space-time, a kind of Cheshire cat trick.[2]

Whatever the ultimate decision on the ontological status of space-time, there remains the problem of what geometric structure G_α and physical magnitudes P_β are needed in an adequate theory of motion. And here the weight of evidence goes strongly against Leibniz. From Galileo to Newton to Einstein, every successful theory of motion makes use of physical quantities which cannot be reduced to relative particle quantities. This opens up a new avenue along which determinism can move; for in order to have well-defined absolute, or non-relative, quantities of motion, the structure of Leibnizian space-time must be beefed up. Consequently, the symmetries (L) must be cut down. Such a cutting down may also cut down the counterexamples to determinism.

4. NEWTONIAN SPACE-TIME

Newton's space-time canvas is much more complex than Leibniz's. In addition to simultaneity, duration, and Euclidean space structure, it also contains a preferred family of motions, called inertial frames, and a distinguished family member called absolute space. The addition of the inertial structure makes into well-defined quantities ones which are not well-defined or invariant in Leibnizian space-time — in particular, the instantaneous (non-relative) acceleration of a particle — and it linearizes the space-time symmetries to form the familiar Galilean transformations:

$$(G) \qquad x^\alpha \to x'^\alpha = R^\alpha_\beta x^\beta + v^\alpha t + c^\alpha$$
$$t \to t' = t + b$$

where v^α and c^α are constants and R^α_β is now a constant orthogonal matrix.

The further addition of absolute space in the sense of a distinguished inertial frame makes (non-relative) velocity as well as acceleration a well-defined dynamical quantity and reduces the space-time symmetries to

$$(N) \qquad x^\alpha \to x'^\alpha = R^\alpha_\beta x^\beta + c^\alpha$$
$$t \to t' = t + b$$

The objection to full-blown Newtonian space-time, with this form of absolute space, as a setting for mechanics is that it violates the principle enunciated in Sec. 2 above connecting symmetries of space-time and symmetries of laws; for the Newtonian laws of motion are invariant under (G). It is perfectly conceivable, however, that additional laws might break the Galilean invariance, necessitating the introduction of additional space-time structure and narrowing (G) to (N); in fact, it was thought in the 19th century that the laws of optics and electromagnetism did just that. I will return to this matter in Sec. 14 below, but for the moment it can be ignored since the addition of the inertial structure to Leibnizian space-time is in itself sufficient to block the argument which threatened Leibnizian determinism; for any member of (G) which reduces to the identity for any finite interval of time, no matter how short, is the identity map everywhere. Note, however, that without the help of absolute space there are limitations to Newtonian determinism. For example, it is not possible to write a law which allows a scalar quantity Φ to vary in space at a fixed moment of time and which determines the future values of Φ from its initial value $\Phi(x, 0)$, $-\infty < x < +\infty$, at $t = 0$. For the law has to be Galilean invariant so that the application of Galilean transformation to any solution must produce a new solution. Choose the transformation so that it is the identity for $t = 0$ but not for later times. Since the initial data are preserved, we then will have two solutions which agree at $t = 0$ but differ in the future.

5. NEWTONIAN PARTICLE MECHANICS

Because the above considerations have been very abstract, it is useful to have before us some concrete examples of determinism triumphant. Since Laplace's espousal of determinism was prompted by his reflections on Newtonian celestial mechanics, it would be natural to look there for the desired example, but actually it turns out to be cleaner to envision a force law different from Newton's $1/r^2$ law.

Consider N point masses m_i ($m_i > 0$), $i = 1, 2, 3, \ldots, N$, and suppose that they attract each other in pairs with a force which acts along the line joining them and which is proportional to the product of their masses and the distance separating them.[3] Combining this force law with Newton's second law of motion yields:

$$\text{(III.1)} \quad m_k \ddot{r}_k = \sum_{\substack{j \\ j \neq k}} C m_j m_k (r_j - r_k) \qquad (C = \text{positive constant})$$

It is always possible to find an inertial system in which the center of

mass is at rest at the spatial origin. In such a system, the equations (III.1) decouple and as a result, the initial value problem with given initial data

$$\text{(III.2)} \quad r_k(0) = \mathring{r}_k,\ v_k(0) \equiv \dot{r}_k(0) = \mathring{v}_k$$

not only has a unique solution, but the general solution can be written down in closed form. Every physically possible history of the system is thus comprehended in a single analytic formula, and the possible pasts and possible futures of the system are, in Laplace's words, present before our eyes.

For Newtonian gravitation, the equation of motion is

$$\text{(III.3)} \quad m_k \ddot{r}_k = \sum_{\substack{j \\ j \neq k}} G m_j m_k (r_j - r_k)/r_{jk}^3 \qquad (r_{jk} \equiv |r_j - r_k|)$$

The initial value problem has a unique solution, at least *locally* in time. If all $r_{ij} \neq 0$, $i \neq j$, at $t = 0$, there exist unique functions r_k of t and a time interval (t_1, t_2) such that for any $t \in (t_1, t_2)$ (III.3) holds and for $t = 0$ (III.2) holds. When $N \geqslant 3$, there are initial conditions for which t_1 or t_2 (or both) are finite. If the solution cannot be extended as a smooth function of t to $t_2 = +\infty$ and $t_1 = -\infty$, the solution is said to be *singular*.

If all such singularities were due to collisions of two or more of the point particles, we could affirm a qualified doctrine of determinism:

(Q) Barring collisions, Newtonian gravitational theory of point mass particles is Laplacian deterministic.

And we can make (Q) sound more impressive by adding that the antecedent is almost always satisfied, for it is known that the set of initial conditions which lead in a finite time to collisions is of (Lebesgue) measure zero (Saari (1973)).

There are, however, some caveats about (Q). First, measure zero need not imply either insignificant or ignorable. We would not judge the set of initial conditions giving rise to collisions to be insignificant if, for example, it proved to be dense within the set of states that eventuate in strong interactions (in some appropriate sense) among the particles. Nor would we regard the measure zero set as ignorable if it loomed large within the range of cases we regard, for whatever reason, to be physically interesting. To illustrate, take the case of $N = 2$. Here it is easy to see that the set of states leading to collisions has measure zero since a collision cannot occur unless the angular momentum is zero and since for $N = 2$ the set of zero angular momentum states has measure zero. But if we are interested in zero angular momentum states, then collisions loom large — indeed, for $N = 2$ such states always lead to collisions.

The second and more important caveat is that (Q) may be false! Define $r(t) \equiv \min(r_{ij}(t))$, $i \neq j$. Then if (t_1, t_2) is the maximal interval for which the solution exists, t_2 is finite iff $r \to 0$ as $t \to t_2^-$, and similarly, t_1 is finite iff $r \to 0$ as $t \to t_1^+$. Further, for $N \leqslant 3$, $r \to 0$ iff there is a collision. But for $N \geqslant 4$ it is an open question as to whether or not $r \to 0$ implies a collision, though the evidence now available indicates a negative answer (see Sec. 7 below). How might the implication fail? One can first try to imagine that the occupant of the role of the minimum r_{ij} switches around and around. But since there are only a finite number of particles, at least one of the potential occupiers, say r_{34}, must actually occupy the role an infinite number of times as (say) t_2 is approached. Thus, we are forced to imagine an oscillatory behavior in r_{34} with $\liminf r_{34} = 0$ but $\limsup r_{34} > 0$. Such wiggling is used in constructing anomalous solutions, as we will see below in Sec. 7. But note that even if r_{34} does go to zero there need not of mathematical necessity be a collision in the proper sense that the position vectors of particles 3 and 4 both approach the same fixed point in space. For it is mathematically possible that these particles accelerate themselves off the space-time manifold and cease to exist at t_2. And a theorem of Sperling (1970) shows that such unbounded behavior must occur in non-collision singularities, should they exist.

It is shocking that determinism may break down for the very case which was supposed to serve as a paradigm example of determinism at work. But worse still, reflecting on the way determinism might break down in this case leads to a general worry about how determinism could ever be securely established in classical worlds.

Before examining the reasons behind this paradoxical worry, let us take note of the somewhat less paradoxical opposite side of this coin. We will see in Ch. VI that to the extent that determinism holds in this case, its course can be traced out by a dumb ($\neq$ stupid) digital computer, if the initial data are computable. Thus, although Laplace was overly optimistic in one way he was overly pessimistic in another; for to the extent that his demon is possible, it need not remain "infinitely remote" but can be instantiated by an uncreative mechanical calculator.

6. DETERMINISM AT BAY

Only a little reflection on some of the commonplaces of classical physics is needed to switch the *Gestalt* of determinism safely and smoothly at work in Newtonian worlds to puzzlement about how Laplacian determinism could possibly be true. The first commonplace is that it is hopeless to try to establish determinism for a system which is not closed to outside influences. Trying to determine the weather in Minneapolis tomorrow from even the most precise meteorological data today in Minneapolis is a thankless task since tomorrow's weather can be influenced by what is now happening in North Dakota and Wisconsin (Fig. III.2). Two remedies may be contemplated. This first is to erect imaginary boundary walls (W_1 and W_2 of Fig. III.3) to record the incoming influences as they penetrate the boundaries of the system. This gives rise to a non-Laplacian initial-boundary value problem: given the appropriate initial data on S and the appropriate boundary data on $W_1 \cup W_2$, determine the state in the interior region R. For field theories where there is action by contact such initial-boundary value problems are often well posed; a successful example will be considered below in Sec. 11. But success cannot be expected for action at a distance theories where effects are transmitted without leaving any traces on the intervening spaces. Furthermore, even when the initial-boundary value approach is successful, the success relies on a departure from pure Laplacian determinism by requiring a specification of future data. I therefore turn to a second remedy which seeks to preserve Laplacian determinism in its pristine form.

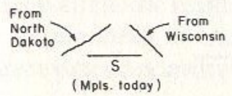

Fig. III.2

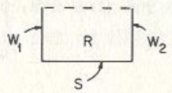

Fig. III.3

The outside influences coming from the Dakotas and Wisconsin can be co-opted by extending the boundaries of the system to take in the hinterlands. For practical purposes, a finite extension of the original initial data surface S may suffice for a pretty good determination of tomorrow's weather. But in a spatially infinite universe, S must be extended to infinity in all directions to make sure that the co-option is complete enough to rule out any possibility of a nasty surprise coming from without. In this way we are driven from the local form of Laplacian determinism to the global form.

The second commonplace threatens even the global form. The laws of classical physics place no limitations on the velocity at which causal signals can propagate. This fact is intimately related to the structure of Newtonian space-time. Without absolute space, velocity is not a Newtonian invariant; whatever the finite value of a particle velocity as measured in one inertial frame, there will always be another inertial frame in which the value is as large as you like. Thus, no law of motion invariant under the Galilean transformations can entail the existence of a fixed finite bound on particle velocity. An infinite velocity is, however, an invariant concept within the Galilean group, and this in turn leads back to the justification for absolute simultaneity: distant clocks can, in theory, be brought into absolute synchronization by means of a sequence of signals whose velocities tend towards infinity.

Signals with actually infinite velocities will be considered a little later, but for the moment it is sufficient to contemplate particle or wave motions where the velocities of propagation are everywhere finite but unbounded. Fig. III.4 illustrates the space-time history α of a particle with velocity $|\dot{x}(t)| < \infty$ for all t but with finite 'escape time' t^* = high noon on April 1, 1988.

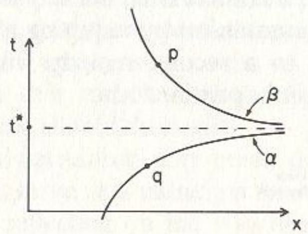

Fig. III.4

As $t \to t^*$ from below, $x(t) \to \infty$ and the particle disappears from the universe, even though $|\dot{x}(t)| < \infty$. The curve β is the temporal mirror image of and it represents the history of a particle which springs the April Fool's joke by appearing from spatial infinity. Thus, in Newtonian space-time the co-option strategy appears to be doomed to failure, for even if the system is extended to include the entire universe, it is not automatically 'closed' in the operative sense to outside influences.[4]

Please do not complain that we never have observed such disturbing disappearing and appearing acts and that, by induction, it is reasonable to expect that we never will. Determinism is a doctrine not just about the actual world but about all physically possible worlds. So even if we can safely employ induction to conclude that no such particles are actual, Laplacian determinism is still threatened if such processes are physically possible. The possibility of $\beta(\alpha)$ is a *prima facie* insult to futuristic (historical) determinism since $\beta(\alpha)$ influences points to the future (past) of $t = t^*$, e.g. $p(q)$, but does not register on $t = t^*$ and so leaves no initial data which can be projected into the future (past).

The threat can be restated by borrowing a concept used extensively in the discussion of relativistic determinism. Let S be a global or local time slice (here, a plane of absolute Newtonian simultaneity or a portion thereof). The *future domain of dependence* $D^+(S)$ of S is to consist of all points p of space-time such that (i) p lies to the future of S and (ii) the state at p depends only on the state on S. The *past domain of dependence* $D^-(S)$ of S is defined analogously. And the *total domain of dependence* $D(S)$ is then the union $D^+(S) \cup D^-(S)$. How to interpret the crucial clause (ii) turns on assumptions about the physics of the situation, but this much seems clear: $p \notin D^+(S)$ (respectively, $D^-(S)$) if there is a space-time curve, representing a physically possible causal signal, which passes through p but which never meets S no matter how far it is extended into the past (respectively, the future). The point of the preceding paragraph can now be restated thusly: Whatever the choice of S in Newtonian space-time, domains of dependence are trivial, for $D(S) = D^+(S) = D^-(S) = S$. Laplacian determinism not only doesn't get to first base, it never even has the chance to come out of the on deck circle! In relativistic space-times, as we will see in the next chapter, determinism at least can be brought to bat in that domains of dependence extend non-trivially.

7. DETERMINISM AT SEA

The threat to determinism is, so far, only a *prima facie* one. To make it palpable, it must be shown that physically possible force functions can generate the kind of behavior picture in Fig. III.4. And more, it must also be shown that the sources which generate the forces either themselves escape contact with $t = t^*$ or else that their behavior at t^* does not code up enough information to make a unique determination of the past and future.

Newtonian gravitational theory of point mass particles provides a relevant example. Mather and McGehee (1975) studied a system of four point mass particles moving colinearly under their mutual Newtonian gravitational forces. Particles 3 and 4 approach one another ever more closely, giving up potential energy in the process. Some of this energy is used to accelerate 3 and 4 and some of it is transferred to particle 1 by means of particle 2 which bounces back and forth between 1 and 3 (see Fig. III.5). Collision singularities are involved, but for the binary collisions the solutions can be extended in a physically reasonable way on the model of elastic bounces. Using this device, Mather and McGehee establish that the solution can become unbounded in a finite time t^*: as $t \to t^*$ from below, $x_1(t) \to -\infty$, $x_3(t)$, $x_4(t) \to +\infty$, while $x_2(t)$ executes an infinite number of bounces between particles 1 and 3. Since the laws of Newtonian gravitation are invariant under time reversal we can invert the Mather-McGehee scenario to produce a solution which insults futuristic determinism by presenting an empty universe up to t^* but thereafter having four particles, three of which appear from spatial infinity and the other of which oscillates infinitely back and forth.

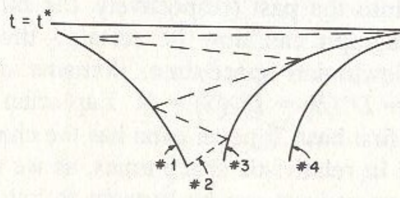

Fig. III.5

Perhaps the problem for determinism is due to collisions. If so we could retreat to the qualified form (Q) of determinism asserting that, barring collisions, Newtonian gravitational theory of point mass particles is Laplacian deterministic. However, it now seems that this retreat does not take us onto safe ground. In a recent article, Gerver (1984) presents a model with five coplanar point masses that never collide. The messenger particle 5 shuttles around the triangle, picking up energy from particle 1 and transferring part of it to 2, 3, and 4, with the

result that the triangle expands with each round trip of the messenger (see Fig. III.6). Gerver makes it plausible that the speed of the messenger and the rate of expansion of the triangle can be arranged so that within a finite time the messenger completes an infinite number of round trips while the triangle becomes infinitely large. The details of a rigorous proof remain to be given.

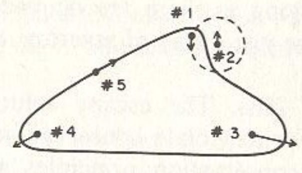

Fig. III.6

It is known that for $N = 4$ the set of initial conditions which could potentially eventuate in a noncollision singularity has measure zero (see Saari (1977)), essentially because, as in the Mather-McGehee construction, all four particles must approach a fixed line in space. But as argued in Sec. 5, measure zero does not necessarily mean insignificant or ignorable, and, moreover, cases for $N > 4$ remain to be settled.

It is not immediately clear to what extent singular but noncollision solutions, should they exist, would undermine Newtonian determinism; for it is not obvious under what conditions such solutions can be joined onto a normal solution. In the case of the heat equation, to be studied below in Sec. 10, the existence of a single solution which is null before $t = 0$ but non-null afterwards is sufficient to completely destroy futuristic determinism since, by the linearity of the equation, the self-exciting solution can be added onto any other solution to produce a new solution.

8. LIFE RAFTS

The true believer in determinism will be undaunted by the examples of Mather-McGehee and Gerver. There is, he will contend, only an apparent failure of determinism, the false appearance being due to considering a space of solutions that is too large in the sense that it encompasses solutions that are not genuinely physically possible; and once these impostors are rooted out, the triumph of determinism will again become apparent. That is the general strategy. Three concrete suggestions for implementing it come to mind.

(i) *Impose boundary conditions at infinity.* By the imposition of appropriate boundary conditions at spatial infinity we can rule out influences coming from or disappearing to God knows where. This achieves by fiat what the laws of motion were supposed to achieve on their own. Given the present state of the universe, the laws determine the future state — as long as there are no rude surprises. Boundary conditions at infinity are just a way of asserting that rude surprises will not be counteranced.

(ii) *Add additional laws.* The escape solutions discussed in the preceding section appear to violate conservation of mass and momentum, so in so far as conservation principles are sacred, the escape solutions are physically impossible. Distinguish two principles of conservation of mass: (C1) particle world lines do not have beginning or end points and mass is constant along a world line, and (C2) for all time t_1 and t_2, the total mass at t_1 = the total mass at t_2. (C1), I claim, is a fundamental principle of classical physics, and it is satisfied even in the anomalous escape solutions. Further, if the laws of motion do not allow escape solutions, then (C1) entails (C2). Some people have been misled into thinking that (C2) is a basic law of classical physics because they have not recognized the possibility of escape solutions.

A similar response is to be made to the invocation of conservation of momentum. Given that a system is closed and that the interactions among the particles satisfy certain restrictions, we can prove conservation of momentum as a theorem. But there is not the ghost of a hope of proving or securing conservation of momentum if the system is open. And the question here is precisely that of whether the universe as a whole is an open system.

(iii) *Object to the idealization of point mass particles.* There are three responses to the objection. (a) Idealizations are always involved in science, and this idealization of point mass particles moving under their mutual gravitational forces was supposed to provide the paradigm of Laplacian determinism at work. So the objection is both querulous and self-defeating. (b) Remove the idealization and consider corpulent particles. You must then say what happens when a collision takes place. Classical physics suggests that we impose laws of elastic impact. But binary collisions of unequal mass particles in two or more spatial dimensions or triple collisions of unequal masses in one spatial dimension are generally non-deterministic. (c) Consider binary collisions of equal mass particles in one spatial dimension. Each collision is deterministic. But with enough particles anomolous non-deterministic solutions can be created, as we will now see.

9. INFINITE BILLIARDS

Consider a system of billiard balls strung out in (two-dimensional) space as shown in Fig. III.7. The balls are assumed to interact only by

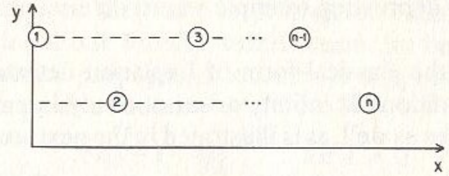

Fig. III.7

contact and then according to the Newtonian laws for perfectly elastic bodies. If for $t < t^*$ all the balls are at rest, then barring outside intervention, the balls will remain at rest for $t > t^*$. If, however, the system is infinitely expanded, letting the number n of balls increase without limit, non-uniqueness of future behavior can result. In addition to the 'normal' solution in which all the balls are at rest for $t > t^*$, Lanford (1975) shows how to construct an anomalous solution in which all the balls are at rest for $t < t^*$ but for $t > t^*$ all but a finite number of them are in motion. This solution is obtained by taking the limit of standard solutions in which the spacing of the particles and the initial direction of the nth particle are arranged so that the nth particle just grazes the n-1th particle, sending it into a grazing collision with the n-2nd particle, etc. If the velocity of the nth particle increases rapidly enough as n increases, then the limiting solution as $n \to \infty$ contains within itself an analogue of the body of Fig. III.4 which appears from spatial infinity. If we plot successively the trajectories of the n-1st particle between the time when it is hit by n and the time it hits n-2, we get a zig-zag approximation to the trajectory β. Running this scenario backwards in time produces an infinite billiard ball analogue of the curve α of Fig. III.4 and an insult to historical determinism.

We have here a very curious situation. The billiard ball example conforms to Lucretius' vision of a world composed of nothing but atoms moving in a void. It also displays a non-deterministic spontaneity but not of the sort Lucretius thought necessary for free will, for not one of the billiard balls freely or spontaneously swerves in contravention to the laws of motion.

The self-exciting feature of Lanford's anomalous solution can be ruled out and determinism restored either by imposing population control and limiting the billiard game to a finite one or by setting boundary conditions limiting the behavior of the billiard balls at spatial infinity. Unless such limitations can be independently motivated we have yet another depressing example where determinism is achieved by fiat.

The need for the classical form of Laplacian determinism to appeal to boundary conditions at infinity arises not only in particle mechanics but in field theories as well, as is illustrated in the next section.

10. HEAT

The classical heat equation in one spatial dimension has a very simple appearance which belies the wealth of peculiarities it contains; it states:

$$\text{(III.4)} \quad \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

where for convenience the thermometric conductivity coefficient has been normalized to unity.

From the remark at the end of Sec. 5 it seems to follow that the heat equation cannot support any form of determinism, for it allows the scalar quantity u to vary in space and it is first order in time so that the appropriate initial data at $t = 0$ is $u(x, 0)$. However, the remark does not apply since the heat equation is not Galilean invariant. The intended physical interpretation of $u(x, t)$ is the temperature at time t and point x of some heat conducting medium, say, an iron bar. The rest frame of the medium is thus a preferred frame for describing thermal history. For the moment I will ignore the intended application and consider (III.4) as an abstract partial differential equation.

The abstract problem of Laplacian determinism is then to find a solution $u(x, t)$ of (III.4) satisfying initial conditions $u(x, 0) = \varphi(x)$, $-\infty < x < +\infty$ and to prove uniqueness of the solution. In this abstract form the problem is not well-posed, for there are null solutions

to (III.4) which vanish at $t = 0$ but which are different from zero for $t > 0$. Since (III.4) is linear, such solutions may be added to any other solution to produce a new solution different from the original one but satisfying the same initial conditions at $t = 0$. Some null solutions are very smooth — indeed, C^∞ — so the breakdown in futuristic determinism is not due to the development of a singularity.[5]

Reflecting on the way heat is propagated according to (III.4) might make one despair of achieving any interesting uniqueness result for the initial value problem. (III.4) is a parabolic partial differential equation with characteristics coinciding with the planes of absolute simultaneity.[6] From this we deduce that heat is propagated infinitely fast so that influences coming from infinity would seem to be the norm. The deduction of infinite propagation velocity is supported by examining the fundamental source solution.

$$\text{(III.5)} \quad k(x, t) = \begin{cases} \exp(-x^2/4t) & \text{for } t > 0 \\ 0 & \text{for } t \leqslant 0 \end{cases}$$

Using the facts that for a solution u, $\int_a^b u(x, t)\,dx$ is taken to be the amount of heat contained in the medium at t between the points a and b, and that $\int_{-\infty}^{+\infty} k(x, t)\,dx = 1$, we interpret (III.5) as describing the temperature distribution of an initially cold bar into which a unit quantity of heat has been introduced at the origin. Since $k(x, t) > 0$ for any $|x| > 0$ and any $t > 0$, the heat is seen to spread instantaneously to the most remote parts of the medium.

However, the form of (III.5) also shows that the effects of the heat source are rapidly attenuated as one moves away from the spatial origin. This makes us hopeful that if the influences coming from infinity will not wreck uniqueness unless they are unboundedly strong. This hope is fulfilled with the aid of a formal condition limiting the growth of a solution u of (III.4) at infinity:

(B) There are constants C and a such that
$|u(x, t)| < C\exp(ax^2)$
for all $-\infty < x < +\infty$ and all $t > 0$

Any solutions of (III.4) which satisfy (B) and which agree at $t = 0$ must agree for $t > 0$. Condition (B) can be weakened but not substantially; e.g., replacing x^2 by $|x|^{2+\varepsilon}$, $\varepsilon > 0$, does not secure uniqueness.

The boundary condition (B) might be promoted by the argument that if the temperature grows beyond all bounds the conditions of the problem are, physically speaking, undercut since our bar of iron will vaporize. But if determinism is going to break down it might as well go down with a bang, and complaining about the breakdown of the problem situation is just a way of bemoaning the demise of determinism. In any case, (B) can be violated without having the medium become as hot as Hades at spatial infinity; u might, to the contrary, become unboundedly negative.

This observation leads to another approach to establishing uniqueness. In view of the interpretation of u as temperature, we might want $u \geqslant 0$ everywhere and always. If our desire is fulfilled, then futuristic determinism is secured in this fashion: suppose that u_1 and u_2 solve (III.4), that $u_1, u_2 \geqslant 0$, and that $u_1(x, 0) = u_2(x, 0)$ for $-\infty < x < +\infty$; then $u_1(x, t) = u_2(x, t)$ for $t > 0$.[7]

For the intended application of (III.4) we may agree that $u \geqslant 0$ is a condition *sine qua non* for physical possibility. But what we would like is for this condition to result from a single initial stipulation and thereafter from the unfolding of determinism. That is, we would like

(H) If u solves (III.4) and $u(x, 0) \geqslant 0$ for $-\infty < x < +\infty$, then $u(x, t) \geqslant 0$ for $t > 0$.

But (H) is false, as the alert reader will already have seen. For let u_1 and u_2 be any solutions of (III.4). By linearity, $\bar{u} = u_1 - u_2$ and $\bar{\bar{u}} = u_2 - u_1$ are also solutions. If u_1 and u_2 conform to the same initial conditions, then $\bar{u}(x, 0) = \bar{\bar{u}}(x, 0) = 0$, so by (H) $\bar{u}(x, t) \geqslant 0$ and $\bar{\bar{u}}(x, t) \geqslant 0$ for $t > 0$, implying that $\bar{u}_1 = \bar{\bar{u}}_2$, i.e., general uniqueness which we have seen does not hold. Thus in matters of heat once is not enough; the stipulation that $u \geqslant 0$ has to be repeated anew at each moment of time.

Finally, it may be well to note that the wringers through which heat puts determinism are not all peculiar to the heat equation. The type of field law appropriate for Newtonian space-time is a parabolic partial differential equation, the heat equation being only a special instance of the type. And for parabolic partial differential equations in general, uniqueness for the Laplacian initial value problem cannot be expected without the help of supplementary boundary conditions.

11. WALLING OUT

As an alternative to giving boundary conditions at infinity we could revert to the non-Laplacian wall strategy mentioned above. For the heat equation this would amount to specifying the function u both on the initial time slice S and on the boundary walls $W_1 \cup W_2$ (refer to Fig. III.3 again) and then trying to determine u within the interior region. This problem is well-posed under seemingly mild continuity assumptions, as follows from the maximum principle for parabolic partial differential equations. In the case of the heat equation, this principle asserts that if a solution u is uniformly continuous over the closed box of Fig. III.3, then u assumes its maximum value on the bottom or the side walls $S \cup W_1 \cup W_2$ of the box. To derive uniqueness, it suffices to take the case where $u = 0$ on $S \cup W_1 \cup W_2$; applying the maximum principle to both u and $-u$ gives $u \equiv 0$.

Instead of a portion of an infinite bar we can focus on a finite bar whose temperature at $t = 0$ is given and whose ends $x = 0$ and $x = 1$ are maintained by two stokers at prescribed temperatures over the interval from $t = 0$ to, say, $t = 1$. For this set up it is natural to require that the temperature $u(x, t)$ on the bar is continuous in x for any fixed t and continuous in t for any fixed point x of the bar. But as Hartman and Wintner (1950) note, the two-dimensional uniform continuity demanded by the maximum principle is not natural if we imagine that the stokers operate independently of one another and independently of the initial temperature distribution. But if two-dimensional uniform continuity is abandoned and what is required of a solution is ordinary continuity and the boundary conditions

$$u(x, 0^+) = \varphi(x), 0 < x < 1$$
$$u(0^+, t) = \psi(t), 0 < t < 1$$
$$u(1^-, t) = \kappa(t), 0 < t < 1$$

then the solution is not necessarily unique. Nor does the imposition of the one-sided boundedness condition $u \geqslant 0$ suffice for uniqueness for the modified problem. However, Hartman and Wintner show that ordinary continuity plus boundedness of u on the box do suffice for uniqueness. The latter condition is reasonable if we imagine that the stokers work with a finite fuel supply and stoke at a finite rate (conservationism and unionism in the service of determinism).

12. OLD HEAT

The problem of historical determinism for heat in an infinite bar has an uninteresting dissolution. In the first place, the instantaneous temperature of the bar, regarded as the *final* temperature, cannot be arbitrarily prescribed, for the assumption that heat has been diffusing according to (III.4) for any length of time, no matter how short, forces $u(x, t)$ to be analytic in x. Worse still, we know that without supplementary boundedness conditions uniqueness cannot be expected; but boundedness conditions coupled with the assumption that the heat equation holds for all past times tends to reduce the situation to an uninteresting static one. Suppose that u satisfies (III.4) for all $t < 0$. If either $|u(x, t)|$ is uniformly bounded for $t < 0$ (i.e., there is a constant C such that $|u(x, t)| < C$ for $t < 0$), or else $u(x, t) \geqslant 0$ for $t < 0$ and $u(x, 0)$ does not grow too fast as $|x| \to \infty$, then $u =$ constant for $t < 0$ (Hirschman (1952)). Interesting initial or final conditions can arise only if the system is open, either to influences coming from infinity or to home town stokers.

Introducing stokers we can formulate a backwards final-boundary value problem where the temperature distribution is known for the final time S and for the ends of the bar W_3 and W_4 for earlier times, and the temperature is sought for the interior R' of the bar at earlier times (Fig. III.8). The maximum principle which was used to prove uniqueness for the forwards looking initial-boundary problem is asymmetrical in time and does not yield uniqueness for the backwards looking problem. (Recall: It asserts that for a closed box in the x—t plane, the temperature takes its maximum either on the *bottom* or the sides.) Uniqueness is equivalent to the proposition that a solution which is 0 on $S \cup W_3 \cup W_4$ vanishes in the interior R'. Physically this would mean that a finite bar whose ends are maintained at a zero temperature cannot rid itself of all of its heat within a finite time.

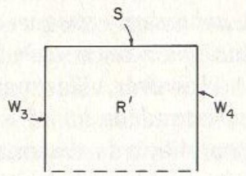

Fig. III.8

To investigate this possibility, assume that the temperature distribution at $t = 0$ is nice enough to be written as a Fourier series. For a bar extending from $x = 0$ to $x = 1$, this means that $u(x, 0) = \Sigma_{m=1}^{\infty} a_m \sin(m\pi x)$. The unique future solution is then $u(x, t) = \Sigma_{m=1}^{\infty} a_m \sin(m\pi x)\exp(-m^2 t)$. Uniqueness for the backwards final-boundary problem is then established by showing that for any finite $t > 0$, $u(x, t) = 0$ implies $u(x, 0) = 0$. Unfortunately, backwards uniqueness is of little help in practical cases of retrodiction since, as the form of the solution indicates, any error in the final data is exponentially expanded in trying to project into the past. Problems for prediction and retrodiction caused by instability will be discussed in Ch. IX.

It is of interest to note that if we append a non-linear term $f(u(x, t))$ to the classical heat equation equation, so that it now reads

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - f(u),$$

then the solution of the modified equation can be driven to zero in a finite time, wrecking backward uniqueness.[8]

The heat equation can be used to make inferences about the past by using predictive models retrodictively. If we assume that at some time in the distant past the earth was in a molten state and neglect heat generated by chemical reaction, radioactivity, extra-terrestrial visitors, etc., then the heat equation can be solved forwards in time to give the present temperature distribution. In the simple models studied by Fourier and Kelvin, the temperature gradient near the earth's surface turns out to be proportional to the temperature of the molten material and inversely proportional to the square root of the period T since the molten state. In this way observations of current temperature gradients can be used to estimate the 'age' T of the earth. (Aside: Kelvin's estimate of 100—200 million years and later estimates by Tait which pushed the value down to 10—20 million years caused some consternation among geologists and the followers of Darwin. Of course, we now know that Kelvin's model contained a number of false assumptions.[9])

13. DON'T FENCE ME IN

We have seen that for determinism to succeed in Newtonian particle and field theories, either the erection of boundary walls or the imposition of boundary conditions at infinity is needed. For field theories these needs would disappear if space were compactified, eliminating spatial infinity. For particle theories, however, the situation is less clear. Roll up the Euclidean x—t plane along the x-axis to produce the cylindrical version of Newtonian space-time shown in Fig. III.9.

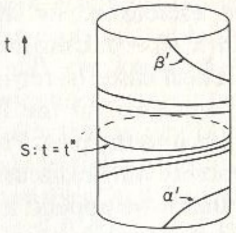

Fig. III.9

Even though there is no spatial infinity for particles to escape to or appear from, domains of dependence may still be trivial. The curves α' and β' result from α and β of Fig. III.4 when this figure is subjected to the rolling process. Since α' and β' have no end points, are everywhere time-like (i.e., are oblique to the planes of simultaneity and have finite velocities), but never meet S, $D(S) = D^+(S) = D^-(S) = S$. Similarly, for a three- or four-dimensional space-time, the initial-boundary value problem is threatened by particles which do higher dimensional analogues of the death spirals of those in Fig. III.9. Thus, in Fig. III.10, it seems that $D^+(S \cup W) = S \cup W$ and $D^-(S \cup W') = S \cup W'$.

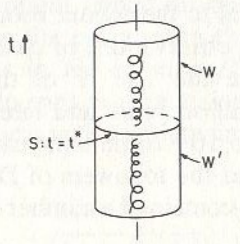

Fig. III.10

The point of erecting imaginary boundary walls was to prevent the unannounced invasion of influences from without. But if the processes illustrated in Fig. III.10 are live possibilities, then the invaders can

invade from within. Examples of such fifth-columnist invaders cannot be constructed in Newtonian gravitational theory of point masses, at least not without collisions. A result of Sperling (1970) proves that if the solution ceases to exist after a finite time and there are no collisions, then the mutual particle distances cannot remain bounded; some of the particles must escape to infinity and in so doing will register on the walls. But perhaps fifth-columnists can be created with the help of binary collisions, as in the Mather-McGehee example, or by using a different kind of force function.

By now the determinist is becoming tired of having to fight a guerrilla war against the invaders who seek to overthrow a deterministic regime; but in Newtonian worlds there is no clear-cut path towards a once-and-for-all victory. For, to repeat, to incorporate into the space-time structure an unbreachable barrier to the invaders is to break Galilean invariance, and Galilean invariance is the Newtonian expression of the well-supported principle of the equivalence of inertial frames. Only a radical change in the structure of space-time can resolve this impasse in favor of the determinist. The special theory of relativity turns out to be an answer to the determinist prayers, as will be seen in the following chapter. But before leaving the classical domain, I want tc discuss some other problems for determinism, one of which does and the other of which does not derive from very fast particles or waves.

14. CLASSICAL ELECTROMAGNETISM

The source free Maxwell equations in empty space read

$$\text{(III.6)} \quad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \nabla \cdot \mathbf{E} = 0$$

Since these equations are not Galilean invariant they require the support of a special frame of reference. In the 19th century this frame was taken to be the rest frame of a ponderable medium, the luminiferous aether, which was thought to be a necessary substratum for electromagnetic waves. However, in keeping with the dematerialization of the aether which took place at the turn of the century, I will construe the aether frame to be a special inertial frame, absolute space, which is unoccupied by a material substratum.[10]

It then follows from (III.6) that electromagnetic waves are propagated with a speed c relative to absolute space and a speed $c \pm v$ in a frame moving relative to absolute space with a speed v. This puts the theory into conflict with actual observational results, but let us imagine that Nature has spoken against Galilean invariance and then ask whether the theory provides us with an example of determinism. It does. If the values of **E** and **B**, subject to the instantaneous constraints $\nabla \cdot \mathbf{B} = \nabla \cdot \mathbf{E} = 0$, are specified at one time, then the top two of the Maxwell equations determine the future values, guaranteeing in the process that the constraint equations continue to hold.

This success for determinism becomes tainted when we attempt to add sources. Formally, the second of the top two Maxwell equations is modified by adding the current density to the right hand side and the second of the bottom two is modified by adding the charge density to the right hand side. The theory is completed by adding the Lorentz force law governing the motion of charges. The resulting formalism admits a well-posed initial value problem as long as the charges move with subluminal velocities, but there is nothing in the formalism as stated to prevent the presence of charged tachyons. With tachyonic sources it remains to be seen whether the system admits a coherent initial value problem and, if so, whether there are solutions in which the tachyons accelerate themselves off the space-time manifold. If in either of these ways classical tachyons should undermine Laplacian determinism we could consider modifying the laws of classical electromagnetism so as to prevent a charged particle from becoming a classical tachyon by accelerating itself from a sub to a superluminal velocity. But this already takes us part way towards relativity theory. And it leaves unexplained why classical charged tachyons couldn't have existed from time immemorial in a superluminal state of motion.

15. SHOCK(ING) WAVES

We made repeated use of two characteristics of the classical heat equation: linearity and infinitely fast propagation of heat. Another interesting feature is that the heat equation has a soothing effect on temperature: whatever roughness exists in the initial temperature distribution is smoothed out in ever so short a time, for solutions $u(x, t)$ become analytic in x (though not in t) for $t > 0$.

Hyperbolic partial differential equations imply finite signal velocities. But non-linear versions of these equations may not have the soothing effect of the heat equation; indeed, solutions may shed whatever smoothness exists in the initial data and become non-differentiable or even discontinuous, thus ceasing to exist as ordinary solutions.

A very simple example studied intensively by mathematicians is the first order equation.

$$\text{(III.7)} \quad \frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} = 0 \qquad (\text{where } f = f(u(x, t)).$$

Setting $f' = \mathrm{d}f/\mathrm{d}u$, (III.6) can be rewritten as

$$\text{(III.8)} \quad \frac{\partial u}{\partial t} + f' \frac{\partial u}{\partial x} = 0.$$

It follows that a solution u is constant along the characteristic curve $x(t)$ which has velocity

$$\text{(III.9)} \quad \mathrm{d}x(t)/\mathrm{d}t = f'(u(x, t))$$

To get a solution corresponding to initial data $u(x, 0) \equiv u_0(x)$, we just propagate the initial data along the characteristics thus: $u(x, t) = u_0(x - tf')$. In the linear version of (III.8), $f' =$ constant and the characteristics are independent of the particular solution. Also the characteristics radiating from the line $t = 0$ simply cover the upper half of the x—t plane, and thus the initial data at time 0 can be propagated forward in time to give a solution for all $t > 0$. But suppose that (III.8) is genuinely non-linear with, say, $f'' > 0$. If the initial data are chosen so that $u_0(x_1) > u_0(x_2)$, $x_1 < x_2$, then the characteristic radiating from $(x_1, 0)$ has a greater velocity than the one from $(x_2, 0)$. So the two must intersect at some point (x^*, t^*) with $t^* > 0$, with the result that u takes on two different values at the same point. Solutions in the ordinary sense may fail to exist after a finite time.

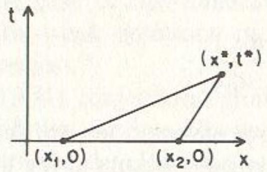

Fig. III.11

Anarchists will be happy to conclude that the law of motion breaks down and chaos reigns. Those more disposed towards law and order will seek a generalized sense of 'solution' on which u can be said to solve (III.8) even though it is not differentiable or even continuous. The mathematical theory of distributions is tailor made for this situation. u is said to be a *weak solution* of (III.7) in the sense of distributions just in case

$$\text{(III.10)} \quad \iint \left[\frac{\partial \phi}{\partial t} u + \frac{\partial \phi}{\partial x} f \right] \mathrm{d}x \, \mathrm{d}t = 0$$

for every test function $\phi(x, t)$ which is C^∞ and which vanishes outside of some compact region of the x—t plane.

Weak solutions overcome the existence problem, but at the expense of uniqueness since more than one weak solution can correspond to the same initial data.[11] The committed determinist will be convinced that the space of all weak solutions is too large, that it extends beyond the bounds of real physical possibility, and that uniqueness will be restored when the unphysical solutions are cut out. But lest he be accused of chicanery in cutting out solutions, the determinist must allow his hand to be guided by independent considerations as to what is and is not physically possible. Just such a guide comes from experience with shock waves, which provide the physical motivation for studying weak solutions in the first place. As a piston compresses a cylinder of gas it creates a wave which travels through the gas with the speed of sound s. But as the gas is compressed, s increases so that the later waves move faster than the earlier ones. In some conditions, the later waves overtake the earlier ones and in such a way that the resulting waveform develops a shock discontinuity where the velocity gradient blows up.[12] If u in our equation (III.7) is interpreted as the velocity of the gas, then it provides a simple mathematical model for the formation of shocks. The velocity gradient is

$$\text{(III.11)} \quad \frac{\partial u}{\partial x} = \frac{u_0'}{1 + u_0' t f''}$$

In keeping with the above assumptions, we set $f'' > 0$ and $u' < 0$ and find that a gradient catastrophe occurs at the positive time $t = -1/u'f''$.

Suppose then that we assume that the only way ordinary solutions

degenerate into weak ones is through the formation of shocks, which we will idealize in the following way: there is a smooth curve $y = x(t)$ across which the solution u may be discontinuous but on either side of which it is smooth. The determinist will then want to show that corresponding to any initial value problem there is a weak solution of this form. This can be done, but alas it still may not be unique. Further surgery on the class of weak solutions is required. If we think of the formation of shocks as an irreversible process, it is natural to require that matter which crosses the shock show an increase in entropy. Analytically, this amounts to requiring that the velocity $\mathrm{d}y/\mathrm{d}t$ of the shock is less than the characteristic velocity $f'(u_l)$ on the left but greater than that $f'(u_r)$ on the right; or equivalently, each point on the shock properly reflects the initial data by being connectible to the initial data surface by a characteristic. With these restrictions in place, uniqueness of weak solutions can be proved, if, as we assumed, $f'' > 0$. If f' is not an increasing function of u a more complicated form of the entropy condition is needed.[13]

For the determinist, the lesson to be drawn is clear. The apparent problem with determinism was a welcome opportunity to investigate in detail the physics of the situation and to show that when that is done determinism works its way in a more subtle and wondrous form than we could have otherwise imagined. The skeptic will complain that the determinist should have been able to say in advance what all the constraints were and should not have been allowed to cut the cloth of physical possibility to suit the needs of determinism.

16. VISCOUS FLUIDS

As a final example, I mention the Navier-Stokes equation, which is the classical equation of motion for a viscous fluid. For appropriate initial data at $t = 0$, a regular solution is known to exist at least for a finite interval $[0, t^*)$, and when it exists it is unique. A weak solution exists for all future time, and in the case of two-dimensional motion is unique. But global uniqueness for weak solutions in real three-dimensional space remains an open question.[14]

In the 1930's Leray (1934) conjectured global uniqueness does not hold for all initial data and that the breakdown of uniqueness (in weak solutions) is associated with the development of turbulence in the fluid. More recently opinion has swung away from Leray's point of view towards determinism and towards an alternative explanation of turbulence, advocated by Ruelle (1981), in terms of "strange attractors." If the evolution of the fluid is indeed deterministic, then its possible motions can be described as a flow on a phase space, each point of which represents a possible instantaneous state of the system. An attractor A is a point or more generally a compact region of the phase space such that the phase orbit uniquely determined by any point sufficiently near to A converges upon A. Conservative dynamical systems (e.g., those described by Hamiltonian mechanics) where the phase flow preserves volume in phase space cannot have attractors. But dissipative systems, such as viscous fluids, where the internal friction dissipates mechanical energy, generally do have attractors. An attractor A is strange if, roughly, the phase orbits determined by points near A are unstable. More will be said about these issues in Ch. IX. The only point I wish to convey here is that determinism in the classical description of this most earthy of processes (tea sloshing in a cup, whirlpools in rivers, etc.) is very much a live issue.

17. CONCLUSION

Several important morals can be drawn from our discussion of determinism in classical worlds. The overarching moral concerns the importance of the status and structure of space-time. On a Newtonian substantivalist conception of space-time, Laplacian determinism is not a free-standing doctrine but requires sufficiently strong space-time scaffolding to support it. A Leibnizian non-substantivalist conception of space-time may avoid the need for some of the scaffolding, but the Leibnizian alternative has never been worked out in sufficient detail to permit judgments to be made with any confidence.[15]

Newtonian space-time, whose structure is rich enough to support the possibility of Laplacian determinism, nevertheless proves to be a none too friendly environment. The principal irritant derives from the possibility of arbitrarily fast causal signals, threatening to trivialize domains of dependence. It is not surprising, therefore, to find non-uniqueness for the initial value problem for some of the most fundamental equations of motion of classical physics, both for cases of discrete particles (ordinary differential equations) and for continuous media or fields (partial differential equations). Whether such non-uniqueness entails the falsity of determinism is a difficult and delicate question, turning in

large part on the status of supplementary conditions that might be imposed on the problem. We encountered a variety of such cases, ranging from those where the supplementary conditions needed to restore uniqueness are both physically well-motivated and nonquestion-begging to others where the supplementary conditions amount to little more than a hypocritical refusal to consider the possibility of unpleasant surprises. Individual attitudes on the classification of cases is naturally influenced by one's predispositions towards determinism. Such a circularity is not unexpected; nor is it entirely unwelcome since it provides a means by which determinism can be used to probe issues about physical possibility and necessity.

Though they are perhaps obvious, there are two other points worth emphasizing. First, the trials and tribulations determinism is forced to undergo in classical physics are purely ontological. None of the ones I have described above derive from epistemological considerations, such as the ability of observers, embodied or disembodied, smart or dumb, to access and process information about the universe. Second, despite the residual and irremediable vagueness in the ontological doctrine of determinism, the threats discussed above are sharp enough to be recognizably threats. And I would add, the issues are not sharpened by yielding to the current philosophical fashion of formalization. If philosophers had spent less time fiddling with axioms, subscripts, n-tuples, and the like, and more time on physics, they would no doubt have produced a better assessment of classical determinism than I have managed.

Whatever the outcome on the substantive issues, it is clear that the long-standing confident pronouncements about classical determinism have been premature. It wasn't until quite recently that hard mathematical results on existence and uniqueness were obtained, and important questions remain open. Classical determinism is not the mummified relic that philosophical literature portrays it to be, but a living and breathing creature capable of generating surprising twists and turns.

NOTES

[1] Following the discussion in Sec. 2 below on the connection between space-time symmetries and symmetries of laws, the condition for Laplacian determinism can be formulated as follows. For any physically possible $\langle M, \{G_\alpha\}, \{P_\beta\}\rangle$ and $\langle M, \{G_\alpha\}, \{P'_\beta\}\rangle$